Three Essays on the Macroeconomic Consequences of Prices

by

DongIk Kang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2018

Doctoral Committee:

Professor Christopher L. House, Chair
Professor Julio A. Blanco
Professor Joshua K. Hausman
Professor Miles S. Kimball, University of Colorado
Professor John V. Leahy

DongIk Kang
dikang@umich.edu
ORCID ID 0000-0003-2716-9853

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Chris House for his continued guidance and support throughout my Ph.D. I was privileged to have him as my advisor. I also thank my committee members, John Leahy, Andres Blanco, Joshua Hausman and Miles Kimball for their insightful comments and support. Alberto Arredondo, Jacob Bastian, Aakash Mohpal, Richard Ryan and Andrew Usher provided moral support as well as academic insights. Finally, I am heavily indebted to my family, in particular to my wife Naewon, for their unconditional support. The first chapter of this dissertation is co-authored with Andrew Usher. The second chapter is co-authored with Miles S. Kimball.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

To understand the workings of the macroeconomy, it is not enough to simply focus on the movements of the aggregate variables of the economy. It is necessary to also understand the behavior of its various components and their interactions. In this thesis, I study three important components of macroeconomic behavior; prices, wages, and the financial system and their connection to the behavior of the aggregate economy.

In the first chapter (joint work with Andrew Usher), we use retail scanner data to show two previously unknown empirical facts about prices. First, the probability of price adjustment increases with product revenue. Second, the absolute size of price adjustment decreases with revenue. These facts are consistent with a menu cost model where the fixed cost of adjustment does not scale with product revenue. Taken together, these facts suggest that prices of products with higher revenues respond more to monetary policy than prices of products with lower revenues. Over the business cycle, both the mean and variance of the (log) revenue distribution across goods decrease with the unemployment rate. These empirical facts imply that monetary policy should have stronger effects on the economy in recessions than in expansions. We verify this property using a quantitative menu cost model, and we provide additional evidence of the state-dependence of monetary policy using aggregate data.

In the second chapter (joint work with Miles S. Kimball), we study the optimal wage structure of a firm with imperfect monitoring of worker effort. We find that when firms can commit to (implicit) long-term contracts, imperfect monitoring leads to optimal wage profiles that reflect

worker seniority. We provide a precise definition of seniority as a measure of worker value to the firm rather than the length of service by a worker. The paper illustrates how worker seniority will evolve over the worker's tenure with the firm and how wage, effort, and separation evolve with seniority. We also show that monitoring and amenities reflect seniority as well. To solve the optimal contract problem, we present a solution technique, the "retrograde approach," of solving complex optimization problems with endogenous discounting and forward-looking state variables in a simple and intuitive way.

In chapter three, I find that system wide runs can be triggered by small shocks to fundamental asset values. Informational frictions amplify small shocks causing large contractions in the amount of credit provided to financial institutions. Asset fire sales exacerbate these effects and force a complete collapse of lending to these institutions; a system wide run. The paper identifies the incentive of healthy institutions to differentiate themselves from distressed ones as the key channel driving the contraction in credit. This contrasts with traditional bank runs that stem from the coordination failures of lenders. The findings lead to direct policy implications; including a government clearing house for loans and quantitative easing.

CHAPTER 1

Product Revenue and Price Setting: Evidence and Aggregate Implications

1.1 Introduction

Using scanner data on retail goods, we find that the probability of price adjustment increases with product revenue and the average size of adjustment decreases with product revenue. These facts are consistent with menu cost models, in which the fixed cost of adjustment does not scale with product revenue. These facts suggest that monetary policy is state-dependent; the real effect of monetary policy is larger in low output states compared to high output states.

Recent studies of price setting have found that accurately reflecting micro facts about price changes have important consequences for aggregate behavior. (See for example Golosov and Lucas 2007, Midrigan 2011, Nakamura and Steinsson 2010). This paper shows that, for many products, whether the products' price is changed or not depends importantly on how much revenue the product makes. In addition, this relationship between price adjustment and product revenue matters at the aggregate level.

First, we demonstrate that high-revenue goods adjust prices more often and by smaller amounts than low-revenue goods do and that this pattern is strong and robust in our data. The relationship between product revenue and price setting behavior remains strong even when we control for product category, stores and UPC. These relationships can arise naturally from menu cost models in which the menu cost does not scale with product revenue. Losses from charging the "wrong" price are greater if the product accounts for more revenue. As a result, firms are less likely to tolerate price discrepancies for high-revenue products. In a menu cost framework, this implies that the range of prices for which firms don't pay the menu cost is relatively small for high-revenue goods. As a result, for these goods, there is a greater likelihood of price adjustment and a smaller size of adjustment. A key component of this analysis is that menu costs do not scale with revenue. Many influential papers, however, implicitly assume that menu costs scale with revenue. This assumption is often made for technical purposes.[1] We find that not only is this at odds with our results, but that it has important implications for aggregate behavior.

The relationship between price setting behavior and revenue introduces what we refer to as a *revenue effect*, where products with higher levels of revenue are more likely to change prices then those with lower levels of revenue in response to changes in monetary policy. This has important implications for menu cost models and monetary policy transmission as the revenue effect interacts with the cross-sectional distribution of product revenue in a meaningful way. First, because the probability of price adjustment increases with the level of product revenue, the real effect of monetary policy decreases when average product revenue increases. Furthermore, the output response to a monetary policy shock decreases when the variance of the revenue distribution increases for a given level of average revenue. This is due to the fact that products with high revenue constitute a disproportionate fraction of aggregate output and as the probability of adjustment for these products increases the response of the aggregate price level increases, leading to a decrease in the response of aggregate output. Assuming that menu costs scale with revenue ignores the revenue effect and its implications.

[1] Examples of influential papers with such assumptions include Gertler and Leahy (2008), Midrigan (2011), Alvarez and Lippi (2013), and Alvarez et al. (2016). Midrigan, Alvarez and Lippi, and Alvarez et al. assume that demand shocks exactly offset productivity shocks so that firm size is normalized; this is similar to assuming that menu costs scale with revenue.

We then shift our attention to the aggregate implications of our findings. First, using variation from regional unemployment, we empirically characterize how the distribution of log revenue changes across the business cycle. Not surprisingly, the mean and the median of log revenue falls with an increase in the unemployment rate. While the change in the mean is to be expected, we also find that the variance and the spread of the log revenue distribution also decreases with unemployment. These systematic shifts in the revenue distribution, taken together with the revenue effect, suggest that aggregate prices are stickier in recessions than they are in expansions. As revenue decreases in recessions, product prices change less frequently and become less responsive to monetary shocks promoting adjustment via output.

To quantify the degree of state-dependence implied by the revenue effect, we simulate the economy using a menu cost model modified to match the behavior of the revenue distribution across the business cycle. We find considerable state-dependence in monetary policy transmission over the business cycle. We compare the response of output to a one standard deviation shock to monetary policy across two states where the initial output difference is 7 percent, reflecting a large fluctuation in output between peak and trough in the post-war era. The cumulative response of output, measured as the area under the impulse response function, is 43 percent greater in the low output state compared to the high output state. If the baseline model is augmented to include habit formation and persistent monetary shocks then the difference in reactions is even larger – roughly 81 percent greater in low output states.

Finally, using measures of monetary shocks proposed by Romer and Romer (2004) and extended to 2007 by Wieland and Yang (2017), we find strong evidence that the output effect of monetary policy is stronger during recessions than expansions, consistent with our results.[2] In addition, the differential effect of monetary policy is pronounced for both durable and non-durable

[2]For additional discussion of the evidence for the state-dependence of monetary policy see (among others) Weise (1999), Garcia and Schaller (2002), Peersman and Smets (2005), Lo and Piger (2005), Santoro et al. (2014), Barnichon and Brownlees (2016), Tanreyro and Thwaites (2016), and Jorda et al. (2017).

consumption goods. The latter result is especially noteworthy as it supports our evidence from retail goods.

This paper is most closely related to the large literature using micro data to evaluate the frequency and size of price changes. The modern literature empirically analyzing price adjustment began with Bils and Klenow (2004), who argued that the duration of prices is surprisingly short. Since that paper, there have been many studies using micro datasets to analyze price setting. Both Nakamura and Steinsson (2008) and Klenow and Kryvtsov (2008) find that, once one excludes price changes due to sales, the duration of prices increases considerably.[3] Many papers have shown that accurately reflecting price setting behavior as documented in the empirical literature has important aggregate implications for New Keynesian models. For example, Golsov and Lucas (2007) and Midrigan (2011) show that matching the distribution of the size of price changes have important implications monetary neutrality.[4] Our paper adds to this literature by studying the relationship between product revenue and price setting behavior.

Our results regarding the state-dependence of monetary policy is most closely related to Vavra (2014). Vavra finds that the cross-sectional standard deviation of the size of price changes is countercyclical. He argues that, if this is driven by volatility shocks to idiosyncratic productivity, the real effects of monetary policy will be weaker in recessions, contrary to our findings. However, Berger and Vavra (2016) find evidence that the countercyclicality in the dispersion of the size of price changes is not due to volatility shocks but rather the endogenous responsiveness of agents to different states of the economy. Our findings match more closely the results in Berger and Vavra. Also, in related work, Santoro et al.(2014) show that loss aversion could imply stronger monetary policy transmission in recessions.

[3]Other papers that study the behavior of prices using micro price data include Anderson et al. (2015), Bhattarai and Schoenle (2014), Coibion et al. (2015), and Eichenbaum et al. (2011).

[4]For other papers that study the relationship between price setting behavior at the micro level and aggregate fluctuations, see also: Alvarez and Lippi (2014), Alvarez et al. (2016), Burstein and Hellwig (2007), Caballero and Engel (2007), Caplin and Spulber (1987), Gertler and Leahy (2008), Midrigan and Kehoe (2015), and Nakamura and Steinsson (2010).

The remainder of the paper is structured as follows. Section 2 discusses our empirical results relating revenue and price setting behavior and the implications for menu cost models. Section 3 documents business cycle movements in the revenue distribution and quantifies the degree of state-dependence implied by our findings. Section 4 provides evidence for state-dependence using aggregate data and Section 5 concludes.

1.2 Revenue and Price Setting

In this section, we study the relationship between price setting and revenue. We present empirical evidence showing that (1) the probability of price adjustment increases with revenue and that (2) the average size of price adjustment is decreasing in revenue. We argue that this is supported by the menu cost framework and derive analytical expressions for the two relationships from a static menu cost model.

1.2.1 Data

We use retail scanner data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The scanner dataset they provide includes information on weekly prices, quantities sold, and various product and store characteristics beginning in the year 2006. Over 90 retail chains across all US markets participate in providing information on over 2.6 million UPCs,[5] 1,100 product categories and 125 product groups. The entire data set covers over half of the total sales volume of US grocery and drug stores and above 30 percent of all US mass merchandiser sales volume.

The dataset is very large with over a hundred billion observations, making it computationally

[5] A UPC (Universal Product Code) is a unique identification number assigned to a retail item.

Table 1.1: Summary statistics

Statistic		**Statistic**	
Adjustment probability	5.3%	Revenue (1st percentile)	\$1.59
Average adjustment size	14.7%	Revenue (10th percentile)	\$5.03
Median adjustment size	10.6%	Revenue (50th percentile)	\$25.13
Average revenue	\$64.43	Revenue (90th percentile)	\$132.00
		Revenue (99th percentile)	\$635.18

intractable to use in its entirety. For this reason, we randomly choose a sample of 30 product categories and 30 markets on which we conduct most of our analysis.[6] Nevertheless, even when using this subset of the available data, our coverage of products and markets is comparable and often greater than most studies that utilize retail scanner data as our sample covers 18,612 different stores and 31,746 different UPCs.

As is standard in studies using micro price data, we choose to focus on "regular" prices and price changes, excluding temporary sales prices. Since not all sales are directly flagged in our dataset, we use the algorithm used by Midrigan (2011) and Midrigan and Kehoe (2015) to compute regular prices. We compute the regular price as the modal price in any given window surrounding a particular week provided the modal price is used sufficiently often.[7] We define a price change as any change in regular price greater than 1% in magnitude.

Our primary unit of analysis is a product. We define a "product," as a unique UPC-store pair. For example, a two-liter bottle of Coca Cola from Kroger's on Main Street would be considered a separate product from a two-liter bottle of Coca Cola from Kroger's on State Street as well as from a two-liter bottle of Pepsi Cola from Walmart on State Street.[8] Because we treat each UPC

[6]A product category is a relatively finely defined subset of products defined by The Nielsen Company. Examples include canned tuna, canned fruit and household cleaners. A market is a designated market area defined by The Nielsen Company, which correspond approximately to a metropolitan statistical area (MSA). The full list of product categories and markets in our sample is provided in Appendix A.

[7]A detailed description of the algorithm we use to compute regular prices can be found in the appendix of Midrigan (2011).

[8]While these examples reflect what we refer to as products, the name and location of the stores are hypothetical and do not represent any actual stores in our sample.

at each store separately, there can be large random swings in revenue from week to week. In addition, in some of the smaller stores there will occasionally be zero sales over a week for some products. To address these concerns, we aggregate the data up to monthly frequency. We thus compute revenue as the total sales of each product each month. This aggregation has the added benefit of easier comparison of our results to previous studies of price setting behavior, which are usually conducted at monthly frequency.

In Table 1.1 we provide some summary statistics of our sample. The average probability of adjustment in our sample is 5.3 percent. The average size of price adjustment conditional on adjustment is 14.7 percent and the median adjustment size is 10.6 percent. These statistics are comparable to other studies using retail scanner data such as Coibion et al. (2015) who find an average frequency of price change of 5.4 percent and absolute size of change around 12 percent. The average revenue of a product in our sample is 64.43 dollars, with revenue ranging from 1.59 dollars for the lowest revenue products (1st percentile) to 635.18 dollars for the highest revenue products.

1.2.2 Empirical Findings

We now formally test for the relationship between product revenue and the probability and size of price changes. We focus on the probability of price adjustment and size of adjustment not only because they have been the focus of many previous studies, but also because as we show in Section 2.3, the menu cost model generates clear predictions about the relationship between revenue and these two variables distinct from other price setting models. Figures 1.1 and 1.2 represent our main findings.

Figure 1.1 depicts the relationship between the probability of price change and revenue in our sample. In panel (a) we compute the average monthly revenue of each product and group them into

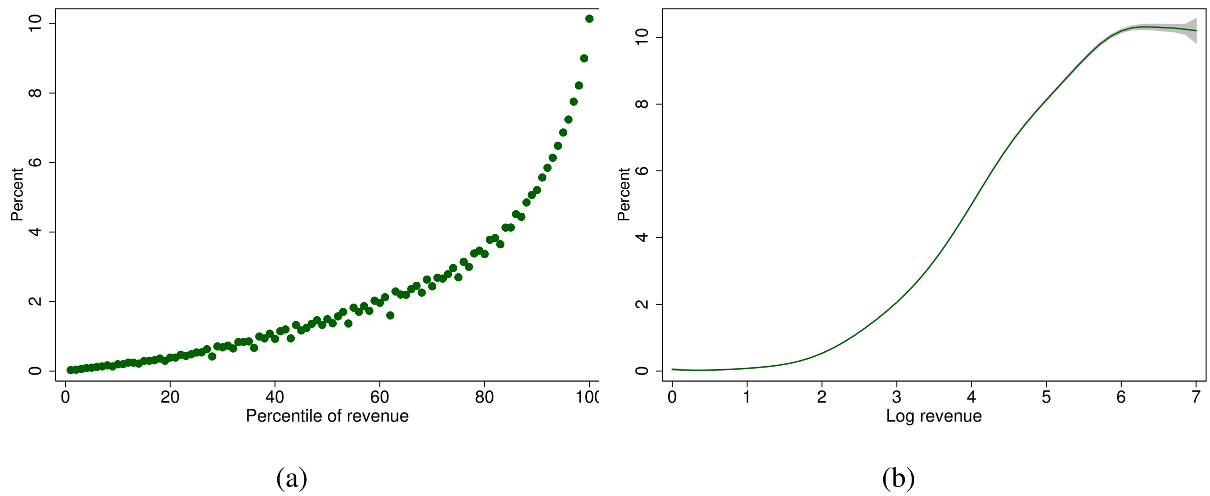


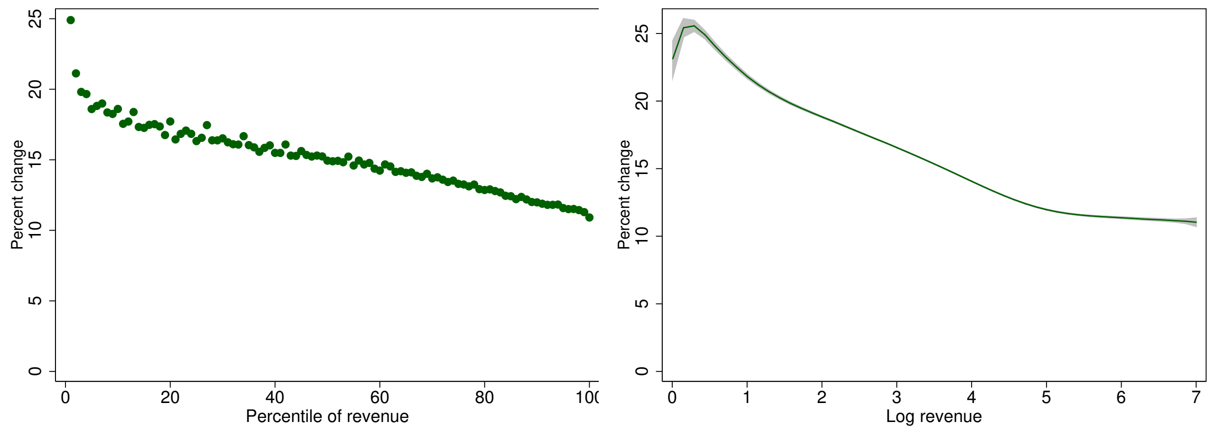
Figure 1.1: Revenue and probability of price adjustment

Note: In Panel (a) we compute the average probability of price adjustment for each product and categorize them into percentile bins by revenue. We then plot the average probability of adjustment by percentile. The sample consists of over 16 million products resulting in over 160,000 products per each percentile bin. Panel (b) shows the expected probability of price adjustment by log revenue from a local polynomial regression.

percentiles by their revenue. We then compute the average probability of price adjustment within each bin and plot this relationship. Panel (b) shows the expected probability of change by log revenue from a local polynomial regression. Panels (a) and (b) clearly show that there is a strong positive relationship between revenue and the probability of price adjustment. For example, in panel (a), a product in the tenth percentile of average revenue has a probability of adjustment of less then 1 percent, while a product at the ninetieth percentile has close to a 6 percent probability.

Figure 1.2 depicts the same relationship between revenue and absolute size of price adjustment. Panel (a) shows the average absolute size of adjustment for each percentile of revenue and panel (b) shows the relationship by log revenue. The figure shows a clear negative relationship between revenue and size of adjustment. Panel (a) shows that the average size of adjustment for a product in the tenth percentile of average revenue is approximately 18 percent, while the average size of adjustment for a product at the ninetieth percentile is approximately 11 percent.

The figures are based on simple averages and compares across different product categories and



(a) (b)

Figure 1.2: Revenue and size of price adjustment

Note: In Panel (a) we compute the average absolute size of price adjustment for each product and categorize them into percentile bins by revenue. We then plot the average size of adjustment by percentile. Panel (b) shows the expected average size of price adjustment by log revenue from a local polynomial regression.

markets. Although these simple exercises show a very strong relationship, we proceed to test the robustness of the relationship controlling for various factors. We find that the relationships remain strong even as we add controls to mitigate potential concerns.

Revenue and Probability of Price Adjustment

To test the relationship between revenue and probability of adjustment, we first construct a dummy variable D_{ijt} that takes the value of one if product i in store j has changed price in month t and zero otherwise. Then we test the relationship using the regression specification

$$D_{ijt} = \alpha_t + \beta logrev_{ijt-1} + \epsilon_{ijt} \tag{1.1}$$

where $logrev_{ijt-1}$ is the log revenue of product i at store j in month $t-1$ and α_t is a month fixed effect.[9] Equation (1.1) can be interpreted as a linear probability model, and the coefficient β

[9]Because we are concerned about changes in regular price there is concern about how to treat revenue in weeks the

represents the increase in the probability of a price change in response to increases in log revenue. For example, a β equal to 0.03 would imply that a 10% increase in revenue will result in a increase in the probability of price change by 0.3 percentage points.

Because price changes in a given month can directly affect revenue in that month, we use lagged rather than current revenue.[10] Standard errors in all specifications are clustered on month, store and UPC separately, in order to alleviate concerns about correlation in consumer preferences within localities or products, store or UPC characteristics, and macroeconomic shocks. We also estimate specifications that add product category, UPC, and store fixed effects to the baseline specification. The time fixed effect, present in every specification, mitigates concerns about spurious correlations with long-term trends such as inflation and growth, in particular with regard to product entry and exit. The inclusion of product category fixed effects rules out the possibility that the results are solely driven by differences across product categories. By adding UPC and store fixed effects, we can similarly rule out the possibility that the relationship is driven completely by differences across UPCs or stores.

We also compute the average frequency of price adjustment[11] and the log of average revenue of each product and test for a relationship between the two variables using,

$$\overline{freq}_{ij} = \alpha + \beta\overline{logrev}_{ij} + \epsilon_{ij} \tag{1.2}$$

This specification is the closest analogue to panel (b) of figure 1.1, and the coefficient β loosely reflects the average slope in the picture. We also add product category fixed effects.

product is on sale. Throughout the paper we use the average revenue of the product over a week in a month multiplied by 4 weeks as monthly revenue. Alternatively we have also tried using the average of weekly revenue of a product in weeks when the product is not on sale multiplied by 4 weeks for monthly revenue and found results that are both qualitatively and quantitatively similar (results not reported).

[10] Results (not reported) from a regression with current log revenue are qualitatively and quantitatively very similar.

[11] The average frequency of price adjustment is computed as a simple ratio of the number of price adjustments over the number of observations we have for each product.

Table 1.2: Probability of price adjustment

	(1) Baseline	(2) Category	(3) Store, UPC	(4) Cross-section	(5) CS category
log revenue	0.0276	0.0290	0.0282	0.0188	0.0208
(monthly)	(0.0007)	(0.0008)	(0.0007)	(0.0004)	(0.0004)
Observations	348,736,845	348,736,845	348,735,564	16,062,396	16,062,396
R-squared	0.029	0.041	0.057	0.173	0.291

Note: Standard errors (shown in parenthesis) are clustered on month, store, and UPC separately.

The results are reported in Table 1.2. The results confirm our earlier findings in Figure 1.1. Columns (1) through (3) reports the results from equation (1.1). Column (1) includes only the month fixed effect, column (2) adds a product category fixed effect, and column (3) includes both store and UPC fixed effects. Columns (4) and (5) show results from estimating Equation (1.2); column (5) adds product category fixed effects. The coefficients on log revenue strongly suggest that the probability of price adjustment is increasing in revenue across all specifications.

The values in columns (1), (2) and (3) are very close to the baseline value of 0.0276. The results imply that a 10% increase in revenue increases the probability of price adjustment by approximately 0.28 percentage points, which is approximately 5.2% of the average probability of price adjustment in our sample. The result is statistically significant at the 1% level for all specifications and also quite large economically. The difference between the highest revenue goods (99th percentile) and lowest revenue goods (1st percentile) in our sample is approximately 6 log units, implying that the probability of price change for the highest revenue products is about 16 percentage points higher than products with the lowest revenue.

The estimated coefficients in columns (4) and (5) are also similar, and slightly smaller than the values in columns (1) through (3). The weaker relationship is unsurprising given that it's between average revenue and average probability of adjustment, and product revenues fluctuate over time.

Nevertheless, the results are still economically significant. The difference in the average frequency of adjustment between products with the highest and lowest average revenue is approximately 11 percentage points. When considering the results in columns (1) through (5) altogether, there seems to be a fundamental relationship between revenue and the probability of adjustment that cannot be entirely attributed to differences across product categories, stores, UPCs, or any other time invariant product characteristics.

Revenue and Size of Price Adjustment

We also test for the relationship between revenue and the absolute size of price adjustment conditional on changing price. Similar to our test for probability of price adjustment, we estimate

$$|\Delta p_{ijt}| = \alpha_t + \beta logrev_{ijt-1} + \epsilon_{ijt}$$

where $|\Delta p_{ijt}|$ is the absolute size of adjustment in log price conditional on adjustment.[12] The coefficient β represents the expected increase in the size of adjustment given an increase in log revenue. As before, we cluster on month, store, and UPC, and add specifications including product category, store and UPC fixed effects.

We also compute the average absolute size of price adjustment and the log of average revenue of each product and test for a relationship between the two variables using,

$$\overline{|\Delta p_{ij}|} = \alpha + \beta \overline{logrev}_{ij} + \epsilon_{ij} \tag{1.3}$$

Table 1.3 shows the results. We find that the relationship between size of price adjustment and revenue is negative and statistically significant across all specifications. However, unlike the

[12]We compute $|\Delta p_{ijt}|$ as the absolute difference in price between the week that we observe a regular price change and the price the previous week. If there is more than one price change in a given month they are treated as two separate observations.

Table 1.3: Size of price change

	(1) Baseline	(2) Category	(3) Store, UPC	(4) Cross-section	(5) CS category
log revenue	-0.0172	-0.0081	-0.0019	-0.0226	-0.0177
(monthly)	(0.0011)	(0.0010)	(0.0009)	(0.0008)	(0.0008)
Observations	18,543,933	18,543,933	18,541,710	4,071,660	4,071,660
R-squared	0.029	0.093	0.246	0.030	0.094

Note: Standard errors (shown in parenthesis) are clustered on month, store, and UPC separately.

relationship between revenue and probability of price adjustment, we find that the coefficient varies significantly across the different specifications. This is due to the fact that the expected size of adjustment and the relationship between the expected size of adjustment and revenue both depend on the entire distribution of price changes and are sensitive to large price adjustments, making the magnitude of the effect sensitive to controls. The sign and statistical significance, however, remain stable. On the other hand, the probability of adjustment only takes into account whether the price has adjusted or not and is insulated from the distribution of price changes.

From the baseline specification, a coefficient of -0.0172 indicates that the average absolute size of price change will decrease by 0.17 percentage points when revenue increases by 10%. This is approximately 1.2% of the average size of adjustment in our sample. The implied difference in the average size of adjustment between the highest and lowest revenue products is approximately 10 percentage points. The results from the cross-section in column (4) would suggest that a product with an increase in average revenue of 10% would decrease the average size of change by 0.23 percentage points.

1.2.3 A Menu Cost Interpretation

The empirical relationships found above between revenue and probability and size of price change can be explained by a menu cost model where the menu cost does not scale with revenue. The positive relationship between revenue and probability of price adjustment and the negative relationship with size of adjustment follows from the fact that potential losses from non-adjustment increase with revenue. We illustrate this mechanism using a simple static menu cost model.

Consider a static problem of a firm with constant marginal cost and facing a demand curve with constant elasticity of demand ϵ. To derive the optimal price the firm solves the following problem.

$$\max_{p} \quad \pi(p) = (p - mc)y(p) \tag{1.4}$$

where p is the firm's price, mc is its marginal cost, and $y(p)$ is the demand schedule for the firm's product. The optimal price for this firm p^* can be solved as,

$$p^* = \frac{\epsilon}{\epsilon - 1} mc$$

A second order approximation of the profit function π around p^* yields,

$$\pi(p) \approx \pi^* + \frac{1}{2}\pi^{*''}(p - p^*)^2$$

where $\pi^* = \pi(p^*)$. We can then express the loss of a firm from having suboptimal price p as

$$L = -\frac{1}{2}\pi^{*''}(p - p^*)^2$$

where $\pi^{*''} = (1 - \epsilon)(\frac{y(p^*)}{p^*})$. Suppose that the firm inherits some price p (not necessarily equal to p^*) and the firm must pay a small menu cost b in order to change its price. The firm will choose to change its price only if the loss from suboptimal price L exceeds the menu cost b, that is, if

$$L = \frac{1}{2}(\epsilon - 1)(p^*y^*)\Big(\frac{p - p^*}{p^*}\Big)^2 > b \tag{1.5}$$

where $y^* = y(p^*)$. Rearranging equation (1.5) and utilizing the approximation that $ln(\frac{p}{p^*}) \approx (\frac{p-p^*}{p^*})$ around zero, we can solve for the inaction region of the firm to get,

$$\Big|lnp - lnp^*\Big| < \sqrt{\frac{2b}{(\epsilon - 1)(p^*y^*)}} \tag{1.6}$$

If the firm's inherited price p is adequately close to its desired price p^* as described by equation (1.6), the firm will not change its price. Note that the range of inaction, expressed as percent deviations from optimal price, is decreasing in desired revenue p^*y^*. It is also decreasing in the elasticity of demand and increasing in menu costs.[13] All else equal, a smaller range of inaction implies more frequent price changes and smaller size of changes conditional on change.

To derive analytical expressions for the relationships between log revenue and probability of adjustment and size of adjustment, suppose that there is a continuum of firms of measure one, with a distribution of inherited prices such that $lnp - lnp^*$ follows some distribution F.[14] The firms are identical otherwise. Denote $x \equiv lnp - lnp^*$. Then, we can derive the expressions for the probability of price adjustment and expected absolute size of adjustment conditional on change

[13] Burstein and Hellwig (2007) present a similar intuition where they argue the inaction region varies with demand. Their focus, however, is on the relationship between the price level and the frequency of price adjustment to infer the relative importance of cost and demand shocks in their framework.

[14] For example, suppose all firm's begin initially at their optimal price given some marginal cost mc_{-1}. The firm then receives a random shock to marginal cost such that their new marginal cost satisfies $mc = \psi mc_{-1}$ and ψ is the shock to productivity. If $ln\psi$ is drawn from the distribution F, the distribution of $lnp - lnp^*$ follows F as well.

as,

$$prob = F(-\zeta(rev^*)) + (1 - F(\zeta(rev^*))) \tag{1.7}$$

$$E[size] = \frac{1}{prob}\Big\{ \int_{-\infty}^{-\zeta(rev^*)} (-x)f(x)dx + \int_{\zeta(rev^*)}^{\infty} xf(x)dx \Big\} \tag{1.8}$$

where $prob$ is the probability of adjustment conditional on revenue and $E[size]$ is the expected absolute size of adjustment conditional on revenue. $\zeta(rev^*) = \sqrt{\frac{2b}{(\epsilon-1)rev^*}}$ is the distance between the optimal markup to the edges of the inaction region and $rev^* = p^*y^*$ is the desired (frictionless) revenue. We derive the relationship between these two statistics and log revenue by differentiating equations (1.7) and (1.8) with respect to log revenue:

$$\frac{\partial prob}{\partial ln(rev^*)} = \frac{\zeta(rev^*)}{2} \cdot [f(-\zeta(rev^*)) + f(\zeta(rev^*))] \qquad > 0 \tag{1.9}$$

$$\frac{\partial E[size]}{\partial ln(rev^*)} = \frac{\partial prob/\partial ln(rev^*)}{prob}\Big\{\zeta(rev^*) - E[size]\Big\} \qquad < 0 \tag{1.10}$$

Equations (1.9) and (1.10) represent the analytical analogues to the regression coefficients on log revenue from Tables 1.2 and 1.3. Equation (1.9) is trivially greater than zero and equation (1.10) is negative as $E[size]$ is greater than $\zeta(rev^*)$ by construction.

The inequalities in equations (1.9) and (1.10) that govern the relationship between revenue and the probability and size of adjustment cannot be derived in standard New Keynesian Calvo pricing models. Most notably, the probability of price adjustment is exogenous in these models and will not be correlated with product revenue in any way. In menu cost models, the inequalities depend on the assumption that menu costs do not scale with revenue. If the menu cost b scales with revenue then revenue cancels out in the expression for ζ, both equations would be equal to zero. Only a menu cost model where the adjustment cost does not vary with revenue is consistent with our empirical findings.

Theoretically, menu costs can consist not only of physical costs of changing prices (i.e. the cost of printing new menus) but also managerial costs and information acquisition costs associated with the decision process. If menu costs consisted primarily of decision costs, it seems unlikely that the costs of price adjustment would increase with revenue. Physical costs associated with changing price tags, on the other hand, could presumably increase with sales. However, with the advent of new technology, such as electronic shelf label systems that allow retailers to change shelf prices electronically from a central computer via a wireless communication system (Levy et al., 1997), even physical costs of changing prices may no longer increase with revenue. Studies by Zbaracki et al. (2004) and the aforementioned Levy et al. that attempt to measure menu costs directly shed some light on this point. Zbaracki et al. show that the managerial cost of price adjustment is 6 times larger than the physical cost for a U.S. industrial manufacturer.[15] Levy et al. measure large physical costs of changing prices for a U.S. supermarket chain. The prices they document, such as the cost of preparing and changing a shelf price tag and verification and supervision costs, suggest only a fixed cost associated with changing the price of a product.

However, many papers make the assumption that menu costs scale with revenue. This is primarily done for technical convenience, either to derive analytically tractable expressions or to facilitate computation. For example, Gertler and Leahy (2008) assume menu costs scale with firm size (revenue) in order to derive analytical expressions for a dynamic Phillips curve. Midrigan (2011), Alvarez and Lippi (2013), and Alvarez et al. (2016) assume that idiosyncratic demand shocks exactly offset productivity shocks such that firm size is normalized. Midrigan utilizes this assumption to reduce the computational burden of solving his model, while Alvarez and Lippi, and Alvarez et al. do so to allow for analytical solutions to their dynamic problem.

However, this assumption is not only at odds with our findings but has a substantive effect on

[15]They also find that costs of informing and negotiating with customers are even greater, approximately 20 times larger than physical costs. The relevance of customer costs in our setting, however, is less clear as the price of retail items are generally non-negotiable.

the behavior of individual firm prices and the aggregate price level. In the following section, we discuss the implications of our findings in the context of menu cost models and their implications for aggregate behavior.

1.3 State-Dependence of Monetary Policy

In this section, we study the aggregate implications of the relationships documented above. The finding that high revenue products adjust price more frequently than low-revenue ones has important implications for menu cost models and for the transmission of monetary policy in the economy. It implies that high revenue product's prices are more responsive to monetary shocks than low revenue product's prices. We refer to this effect as the *revenue effect*. Changes in the distribution of revenue in recessions, along with the revenue effect, imply that monetary policy has a larger effect on output response in recessions. We find support for this state-dependence of monetary policy using a quantitative menu cost model, and we quantify the magnitude of the effect.

1.3.1 The Revenue Effect

The dependence of price adjustment on revenue implies a *revenue effect* in which the prices of high revenue products are more responsive to monetary shocks than low revenue product prices. As shown in section 2.3 this is due to the fact that the range of prices around the frictionless optimal price for which the price is not adjusted is decreasing in product revenue meaning that high revenue products are more likely than low revenue products to adjust their prices. This is true, not only for different products, but also when a given product's revenue changes over time.

The revenue effect implies that the cross-sectional distribution of revenue is an important determinant of the real effect of monetary policy. First, if the overall level of product revenue increases, the likelihood of price adjustment will increase as well. As losses from suboptimal prices increase with greater revenue, prices will adjust to smaller deviations from the desired price induced by a monetary policy shock. Consequently the output response will be mitigated.

Furthermore, given a mean of the revenue distribution, we find that the response of aggregate output to a monetary shock decreases as the variance of the revenue distribution increases. As the variance of the revenue distribution increases, products with revenue greater than the mean become more likely to adjust prices and products with revenue lower than the mean become less likely to adjust prices. Since the products whose probability of adjustment increase also constitute a larger fraction of the aggregate output, their contribution to the response of output is more important. Thus, as the revenue of these products and their probability of adjustment become larger, more high revenue products adjust through prices instead of through output leading to an increase in the response of the aggregate price level and a fall in the response of aggregate output.

An alternative perspective is through the composition of prices that adjust in response to a monetary policy shock. Products with high revenue are more likely to respond to monetary shocks via price change, so these products constitute a disproportionate fraction of the overall number of price changes in response to monetary shocks. Products with low revenue constitute only a small fraction. When the variance of the revenue distribution increases, low revenue products become even less likely to adjust and high revenue product become more likely to adjust. Loosely speaking, the number of low revenue products that adjust price will decrease and will be replaced by high revenue products, leading to a larger response of aggregate prices and smaller response of aggregate output.[16]

As discussed in section 2.3, assuming that menu costs scale with revenue undoes the revenue

[16]An increase in the variance of revenue will lead to an increases in the heterogeneity in the frequency of price adjustment. Since Carvalho (2006), it has been long understood that heterogeneity in the frequency of price change increases the aggregate output response in time-dependent models. Extensions of this research to state-dependent models by Nakamura and Steinsson (2010) and Alvarez et al. (2016) have reached similar conclusions. However, when the heterogeneity is due to the revenue effect, an increase in the heterogeneity of the frequency of price change can mitigate the strength of the output response.

effect. The revenue effect, as discussed above, has important implications for the transmission of monetary policy. Furthermore, as we will discuss in the remainder of this section, the interaction between the revenue effect and the movements in the revenue distribution imply that monetary policy is state-dependent.

1.3.2 Revenue Distribution and Unemployment

To discuss the interaction between the revenue effect and the revenue distribution across the business cycle, we first document the movements of the revenue distribution across the business cycle. Our sample period is 10 years, which is relatively short for analyzing business cycle movements. However, we leverage the fact that we have observations in many markets and exploit variation in regional unemployment. For this exercise we expand our sample to include 112 product categories instead of the 30 we used in section 2. We expand the sample at this point for a few reasons. The fact that our unit of observation is now a moment of the cross-sectional revenue distribution in each time period (rather than a single product) greatly decreases the computational burden and allows us to handle more product categories with relative ease. Expanding the sample also gives us greater coverage of products and increased statistical power. However, our results remain qualitatively unchanged when we use the 30 product categories in section 2 (reported in appendix B).

To document the relationship between regional unemployment rate and various moments of the log revenue distribution, we estimate

$$Y_{cmt} = \alpha_t + \delta_c + \gamma_m + \beta \cdot UR_{mt} + \epsilon_{cmt} \tag{1.11}$$

where Y_{cmt} is a statistic for the distribution of revenue of product category c in market m in month

t.[17] UR_{mt} is the unemployment rate for the region m.[18] α_t is the month fixed effect, δ_c is the product category fixed effect, and γ_m is the market fixed effect. The statistics we use for the left-hand-side variable Y_{cmt} include the mean, the median, the standard deviation, and the difference in log revenue between a product at the 90th percentile in revenue and 10th percentile in revenue (the spread).

We include time fixed effects for several reasons. First, they remove the secular and nominal trends from the data, addressing concerns about spurious long-run trends driving our results. Second, the inclusion of time fixed effects suggests the relationship between unemployment and revenue is driven by demand as discussed in Coibion et al. (2015). Because most goods are produced outside the local market, aggregate productivity shocks are external to the market, implying that changes in revenue and unemployment are correlated mostly through local demand. The standard errors are clustered by market, product category, and time, to address concerns about temporal and spatial correlations.[19]

Table 1.4 shows the results. We find that not only do the mean and median of the distribution of revenue decrease in recessions as would be expected, but that both the standard deviation and the spread between the 90th and 10th percentile goods falls as well. The results show that a 1% increase in regional unemployment corresponds approximately to a 1.3% decrease in mean log revenue, a 1.4% decrease in the median log revenue, a 0.37% decrease in the standard deviation and a 1.1% decrease in the log revenue spread.

While we are focused on documenting the behavior of the revenue distribution and not on explaining the reasons behind the documented patterns, recent studies of consumer shopping behavior over the business cycle provide many potential channels through which this may occur. For example, Coibion et al. (2015) find that households reallocate consumption expenditures toward

[17]Additional results from alternative specifications are reported in appendix D.

[18]The regional unemployment rate UR_{mt} is computed as the population weighted unemployment rate of the counties that comprise market m. The unemployment rates by county are obtained from the American Community Survey.

[19]A discussion of robustness of our results is provided in appendix B.

Table 1.4: Revenue distribution over the business cycle

	(1) Mean	(2) Median	(3) Std deviation	(4) Spread
unemployment	-1.293	-1.436	-0.370	-1.115
	(0.392)	(0.435)	(0.159)	(0.408)
Observations	351,125	351,125	351,125	351,125
R-squared	0.686	0.674	0.597	0.603

Note: Standard errors (shown in parenthesis) are clustered on month, market, and product category separately.

low-price retailers when local economic conditions deteriorate. Jaimovich et al. (2017) find that people traded down in the quality of goods and services they consumed during the Great Recession. Nevo and Wong (2017) document extensive substitution behavior by consumers over the business cycle. They find that households increase coupon usage, increase purchases of goods on sale, buy larger sized products, buy generic products, and substitute purchases toward big box (discount) stores during recessions. Any one or a combination of these extensive consumer substitution patterns could potentially generate the movements in the revenue distribution that we document.

1.3.3 Quantitative Model

We build a quantitative model that matches both the pricing moments found in the micro data and business cycle variations in the revenue distribution we find in the previous section. We augment a standard menu cost model similar to those found in Golosov and Lucas (2007) and Midrigan (2011) to match the shifts in the revenue distribution across the business cycle. We use this model to quantify the degree of state-dependence of monetary policy implied by our mechanism.

Households

Households maximize discounted expected utility,

$$\max E_t \sum_{\tau=0}^{\infty} \beta^{\tau} \{\phi_{t+\tau} \frac{(C_{t+\tau})^{1-\frac{1}{\gamma}}}{1-\frac{1}{\gamma}} - \frac{N_{t+\tau}^{1+\frac{1}{\eta}}}{1+\frac{1}{\eta}}\} \tag{1.12}$$

subject to the budget constraint,

$$P_t C_t + B_t = R_{t-1} B_{t-1} + W_t N_t + \Pi_t.$$

C_t is the household consumption of the composite good, N_t is total labor supply, B_t is one-period nominal bonds, P_t is the aggregate price level, R_t is the nominal rate of return, W_t is nominal wage, and Π_t is the firm profits that are transfered to the households. ϕ_t is an aggregate preference shock that represents a shock to the discount rate and affects the intertemporal substitution of households.

Households consume a continuum of a variety of products indexed by i. The composite good C_t is a Dixit-Stiglitz index of these products,

$$C_t = \left(\int \left(\iota_{it} c_{it} \right)^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} \tag{1.13}$$

where ι_{it} is the idiosyncratic preference for product i and ϵ is the elasticity of substitution across goods. Households choose to consume variety c_{it} to maximize C_t subject to the constraint,

$$\int q_{it}^{-1} p_{it} c_{it} = P_t C_t$$

where p_{it} is the price of product i. The household's subproblem of variety choice is standard except for q_{it}. In this model q_{it} serves as an exogenous demand shifter that is needed to match the documented changes in the revenue distribution across the business cycle. We treat q_{it} as exogenous in order to focus on the firm's decision process. Furthermore, we wish to remain amenable to different household processes that may generate the shifts in demand. Nevertheless, a possible interpretation of q_{it} is as the cost of shopping effort (see Coibion et al.,2015)). Another interpretation would treat q_{it} as the household's valuation of product quality, defined as a function of the product's idiosyncratic components (z_{it}, ι_{it}); this is in line with the findings of Jaimovich et al. (2017).[20]

We model q_{it} as depending on the aggregate state of the economy as well as the idiosyncratic characteristics of the product.

$$q_{it} = q(\tilde{q}(z_{it}, \iota_{it}), \phi_t)$$

To be consistent with the empirical findings in section 4.1 we impose the condition that,

$$\frac{\partial^2 lnq_t}{\partial ln\tilde{q}_t \partial ln\phi_t} > 0$$

With a shopping cost interpretation, this condition means that in recessions, the opportunity cost of time falls and it becomes less costly to buy products that households usually purchase infrequently. A product quality interpretation would imply that household sensitivity to quality differences decreases in recessions.

The inclusion of idiosyncratic preferences ι_{it} is important in our setting. The revenue effect implies that the revenue distribution is an important determinant of output fluctuations, and the inclusion of idiosyncratic demand shocks are critical in this regard. Studies such as Burnstein and Hellwig (2007) and Kumar and Zhang (2016) find that demand shocks constitute a large fraction of the idiosyncratic shocks that firms face, and ignoring this component may lead to an understatement of revenue dispersion. In the calibration of this model, we show that this is true in our sample as well.

[20]To interpret as valuation of product quality it may be more intuitive to specify q_{it} as a part of the composite good index $C_t = (\int (q_{it}\iota_{it}c_{it})^{\frac{\epsilon-1}{\epsilon}} di)^{\frac{\epsilon}{\epsilon-1}}$ which gives almost identical results.

The demand curve for good c_{it} then follows,

$$c_{it} = \left(\frac{p_{it}}{P_t}\right)^{-\epsilon} q_{it}^{\epsilon} \iota_{it}^{\epsilon-1} C_t \tag{1.14}$$

and the expression for the aggregate price index can be derived as

$$P_t = \left\{ \int_0^1 \left(\frac{p_{it}}{q_{it}\iota_{it}}\right)^{1-\epsilon} di \right\}^{\frac{1}{1-\epsilon}} \tag{1.15}$$

Aggregate Shocks

Following much of the literature on menu cost models, we assume that the monetary authority conducts monetary policy by targeting a path of nominal GDP

$$lnM_t = g + lnM_{t-1} + \epsilon_t^m \tag{1.16}$$

where $M_t = P_t Y_t$ is nominal GDP. The aggregate preference parameter ϕ_t has two states $\{\phi^H, \phi^L\}$ that follows the Markov process,

$$\begin{bmatrix} \phi_t^H \\ \phi_t^L \end{bmatrix} = \begin{bmatrix} r_{HH} & r_{LH} \\ r_{HL} & r_{LL} \end{bmatrix} \times \begin{bmatrix} \phi_{t-1}^H \\ \phi_{t-1}^L \end{bmatrix}$$

where r_{ij} is the transition probability from state i to state j. The parameter ϕ_t determines whether the economy is in a high output state or a low output state.

Firms

Firms produce a differentiated product y_{it} with linear production technology

$$y_{it} = z_{it} n_{it}$$

where z_{it} is firm i's idiosyncratic productivity and n_{it} is labor demand. The firm's current period real profit is,

$$\pi_{it} = \Big(\frac{p_{it}}{P_t} - \frac{W_t}{z_{it} P_t}\Big) y_{it}$$

where the demand for y_{it} is given by equation (1.14).

Now consider the firm's dynamic problem. Henceforth, for notational simplicity we omit the subscript i except when explicitly necessary. It is more convenient to express the firm's dynamic problem in terms of real markup $\mu_t = \frac{p_t z_t}{W_t}$. Substituting in equation (1.14) for y_{it}, the firm's current period profit is

$$\pi_t = (\mu_t - 1)\mu_t^{-\epsilon} q_t^{\epsilon} (z_t \iota_t)^{\epsilon-1} Y_t \Big(\frac{W_t}{P_t}\Big)^{1-\epsilon}.$$

Idiosyncratic productivity follows the process

$$ln z_t = ln z_{t-1} + \epsilon_t^z$$

where ϵ_t^z is a shock to productivity. The shock ϵ_t^z follows a Poisson process,[21]

$$\epsilon_t^z \sim \left\{ \begin{array}{ll} 0 & : \text{with probability } \theta_z \\ N(0, \sigma_z^2) & : \text{with probability } 1 - \theta_z \end{array} \right\}$$

[21]Midrigan (2011) and Alvarez and Lippi (2014) argue that the process of idiosyncratic productivity shocks is important for the real effect of monetary policy as it determines the strength of the *selection effect* present in the models. In appendix C, we argue that an accurate representation of the revenue effect in the model could suggest that the selection effect in the model is accurately reflected as well.

and idiosyncratic demand evolves according to

$$ln\iota_t = ln\iota_{t-1} + \epsilon_t^{\iota}$$

where ϵ_t^{ι} follows the normal distribution $N(0, \sigma_{\iota}^2)$. We assume that log productivity and log demand follow a random walk, which allows us to define the idiosyncratic state of a firm jointly as $\omega_t = z_t \iota_t$. This reduces the number of state variables and decreases the computational burden of solving this model numerically. The idiosyncratic state ω_t then follows

$$ln\omega_t = ln\omega_{t-1} + \epsilon_t^z + \epsilon_t^{\iota}$$

Since this variable follows a random walk, we include a probability of firm exit to keep the distribution of firms stationary. We assume that firms exit with probability α. When a firm exits, it is replaced by a new firm with $z_t = \iota_t = \omega_t = 1$. We now write the firm's dynamic problem, where each period it must pay a small menu cost b to change its price as,

$$\begin{aligned}
&V(\mu_t, \omega_t; Y_t, \phi_t) \\
&= \max\Bigg\{\pi(\mu_t, \omega_t, Y_t, \phi_t) + (1-\alpha)\beta E_t \Lambda_{t+1} V\Big(\mu_{t+1}, \omega_{t+1}; Y_{t+1}, \phi_{t+1}\Big), \\
&\quad \max_{\tilde{\mu}}\Big\{\pi(\tilde{\mu}, \omega_t, Y_t, \phi_t) + (1-\alpha)\beta E_t \Lambda_{t+1} V\Big(\mu_{t+1}, \omega_{t+1}; Y_{t+1}, \phi_{t+1}\Big)\Big\} - b\Bigg\}
\end{aligned}$$

where Λ_{t+1} is the stochastic discount factor. We use log linear approximations of the relationship between labor supply and output to allow convenient aggregation as in Nakamura and Steinsson (2010). We use the Krusell and Smith (1998) algorithm to solve the model numerically. We assume that firms perceive the evolution of the price level $\frac{P_t}{P_{t-1}}$ as a linear function of variables,

$$ln\Big(\frac{P_{t+1}}{P_t}\Big) = \beta_1 + \beta_2 lnY_t + \beta_3 ln\phi_{t+1} + \beta_4 ln\phi_t + \beta_5 \epsilon^m_{t+1}.$$

1.3.4 Calibration and Results

Following our empirical results above, we set the time unit of the model to a month. The time discount factor β is set to an annual value of 0.96. Following much of the menu cost literature, we set the intertemporal elasticity of substitution γ equal to 1 and the inverse Frisch elasticity $\frac{1}{\eta}$ to zero to facilitate computation. We calibrate the elasticity of demand to $\epsilon = 3$, which is in line with the median elasticities estimated in Nevo (2001) and Broda and Weinstein (2006). The monthly growth rate for nominal GDP g is calibrated to be 0.17 percent and the standard deviation σ_m is set as 0.29 percent, the average growth rate of nominal GDP and the standard deviation of HP-filtered nominal GDP from 1990-2007 respectively.

The parameters governing the productivity process, demand process, probability of product exit, and menu cost $(\theta_z, \sigma_z, \sigma_\iota, \alpha, b)$ are jointly calibrated to match the average frequency of adjustment, the mean and median absolute size of adjustment, the strength of the relationship between probability of adjustment and revenue, and the contribution of price dispersion to the variance of revenue. We target an average frequency of adjustment of 11%, within the range of findings of Nakamura and Steinsson (2008). We target the median absolute size of adjustment to be 8.5% also following the results of Nakamura and Steinsson and target a mean absolute size of adjustment of 10% from Klenow and Kryvtsov (2008). We focus on these moments, which are computed using the micro data from the Bureau of Labor Statistics used to compute the Consumer Price Index instead of those of our sample because we are interested in the broader implications for the aggregate economy. While our dataset has the advantage that we can observe the price and quantity of each individual product at high frequency, our dataset is narrower in scope. However, for results that are

Table 1.5: Target moments

Steady state moments	**Target**	**Model**
Frequency of price adjustment	0.11	0.11
Average size of price adjustment	0.10	0.10
Median size of price adjustment	0.085	0.085
$\frac{(\epsilon-1)^2 var(price)}{var(revenue)}$	0.20	0.20
$\frac{\partial prob}{\partial lnrev} \times \frac{1}{prob}$	0.52%	0.51%
Business cycle moments		
$\frac{\Delta \overline{rev}}{\Delta N}$	1.3	1.3
$\frac{\Delta(rev_{90}-rev_{10})}{\Delta N}$	1.1	1.1
Percent difference in output (High vs low output state)	0.07	0.07
Non-target moments		
$\frac{\Delta rev_{50}}{\Delta N}$	1.4	1.1
$\frac{\Delta std(rev)}{\Delta N}$	0.37	0.23
$\frac{\partial E[size]}{\partial lnrev} \times \frac{1}{size}$	-0.12% $\sim$ -0.01%	-0.13%

Note: Model refers to the baseline model described in section 3.3. $\left(\frac{(\epsilon-1)^2 var(price)}{var(revenue)}\right)$ is the ratio of the contribution of the variance of price to the variance of revenue. $\left(\frac{\partial prob}{\partial lnrev} \times \frac{1}{prob}\right)$ is the change in the probability of adjustment as a percentage of the average probability of adjustment. $\frac{\Delta \overline{rev}}{\Delta N}$ is the percent change in mean of the log revenue distribution over the percent change in total labor supply. $\frac{\Delta std(rev)}{\Delta N}$ is the percent change in the standard deviation of the log distribution over the percent change in total labor supply. $\frac{\Delta(rev_{90}-rev_{10})}{\Delta N}$ is the percent change in the spread of the log revenue distribution over the percent change in the total labor supply. $\frac{\Delta rev_{50}}{\Delta N}$ is the percent change in the median of the log revenue distribution over the percent change in total labor supply. $\left(\frac{\partial E[size]}{\partial lnrev} \times \frac{1}{size}\right)$ is the change in the change in the expected size of adjustment as a percentage of the average size of adjustment.

unique to this paper, we have no choice but to utilize our results. Thus, we target for the probability of adjustment to increase by 5.2% of the average frequency of adjustment when revenue increases by 10% reflecting our earlier results. We target the contribution of the variance of price to the variance of revenue to be 20% in steady state. This is based on Burstein and Hellwig's (2007) findings that the variance in revenue stemming from demand shocks are 3 to 4 times larger than the variance in revenue stemming from productivity shocks in settings similar to ours.

The results of the calibration is shown in Tables 1.5 and 1.6. We calibrate the frequency of

Table 1.6: Parameter values

Parameter	Model	Description
β	$0.96^{1/12}$	discount factor
ϵ	3	elasticity of demand
α	0.07	rate of product exit
b	0.05 (0.7%)	menu cost (percentage of steady state revenue)
$1-\theta_z$	0.29	probability of productivity shock
σ_z	0.10	standard deviation of productivity shock
σ_ι	0.10	standard deviation of demand shock
g	0.17%	monthly growth rate of nominal GDP
σ_m	0.29%	standard deviation of nominal GDP
r_{ij}	0.05	probability of transition to different state
$\bar{\phi}$	0.015	magnitude of aggregate preference shock
ψ_0	0.012	slope parameter in exogenous demand
ψ_1	2.87	intercept parameter in exogenous demand

productivity shocks $1-\theta_z$ to be 29% and the standard deviation of the shock σ_z to be 10%. The standard deviation of the demand shock σ_ι is also set to 10% and the rate of product exit α is set to 7 percent. The size of the menu cost b is 0.05 which implies that the total cost of changing prices amount to 0.7% of the steady state revenue. This is approximately equal to the value of 0.7% of revenue found by Levy et al. (1997) and slightly below the 1.2% of revenue found by Zbaracki et al.(2004). We do not explicitly target the relationship between revenue and the expected adjustment size of prices. Nevertheless the calibrated model delivers a coefficient of -0.013 which is -0.13% of the average size of adjustment in the model. We find that this coefficient is between -0.01% and -0.12% of the average size of adjustment for different specifications in our sample.

We calibrate the probability of transitioning to a different state r_{ij} to be 5 percent, which implies an average business cycle length of 40 months. The size of the preference shock is assumed to be symmetric where $\phi_H = (1+\bar{\phi})$ and $\phi_L = (1-\bar{\phi})$. We adopt the following functional form for

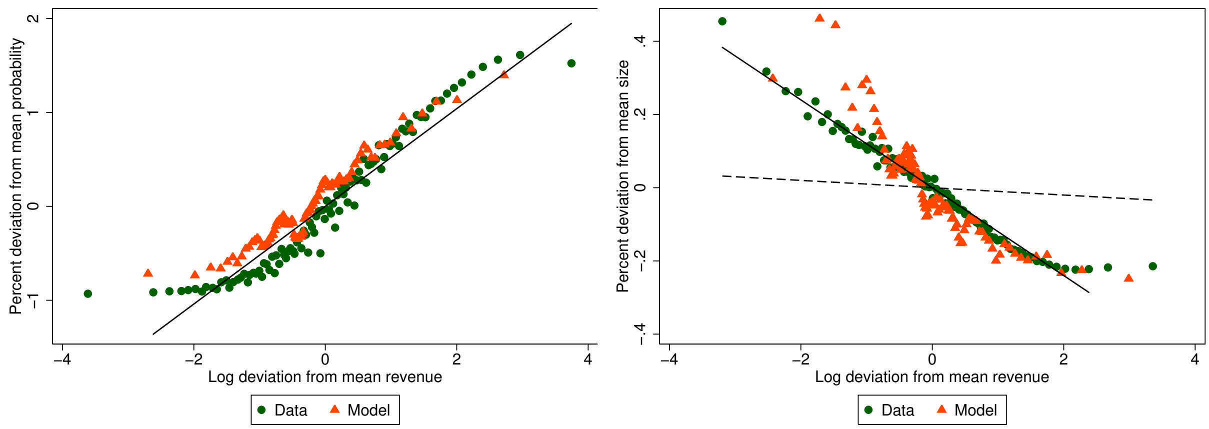
lnq_t in order to match the changes revenue distribution,

$$lnq_t = \psi_0 \cdot (ln\omega_t + \psi_1)(\frac{ln\phi_i}{ln\phi_H - ln\phi_L})$$

where $i \in \{H, L\}$. The parameters $(\bar{\phi}, \psi_0, \psi_1)$ are jointly calibrated to match the change in the mean and spread of the log revenue distribution with the state of the economy as shown in Table 1.4 and a difference in the output level between the high and low output states of 7 percent. We target the difference in output level between high and low output states to be 7 percent to reflect a large fluctuation in output between peak and trough. The model slightly underpredicts the change in the median and the standard deviation of the revenue distribution with output as shown in Table 1.5.

Figure 1.3 compares the relationship between price setting behavior in the steady state in the model, to the data. Panel (a) shows the relationship between the probability of price adjustment and log revenue. In both the data and our model, we group each product-month observation by revenue into percentile bins. We then compute the average revenue and the probability of adjustment for each bin. Then, we compute the log deviation of this revenue from the mean revenue of the data and model respectively, to facilitate comparison. We also normalize the probability of adjustment for easier comparison. We compute the deviation of the probability of adjustment from the mean probability and divide by the mean probability for the percent deviation from mean probability of adjustment. We plot, for each percentile bin, the log deviation from mean revenue and the percent deviation from mean probability of adjustment. This allows close comparison to our calibration, which targets the statistic $\left(\frac{\partial prob}{\partial lnrev} \times \frac{1}{prob}\right)$ as shown in Table 1.5.

The green dots in panel (a) represent each percentile grouping from the data and the orange triangles show the results from model. The solid line represents the calibration target, with a slope of 0.52% between log revenue and probability of adjustment. Panel (a) shows that the relationship between revenue and probability of adjustment in the model closely resembles the relationship



(a) Probability of price adjustment

(b) Size of price adjustment

Figure 1.3: Revenue and price setting: compare data and model

Note: In panel (a) we first group each product-month observation by revenue into percentile bins and compute the average revenue and the probability of adjustment for each bin for both the data and the model. We then compute the log deviation of this revenue from mean of the data and model respectively. We compute the deviation of the probability of adjustment from the mean probability and divide by the mean probability to compute the percent deviation from mean probability of adjustment. We plot, for each percentile bin, the log deviation from mean revenue and the percent deviation from mean probability of adjustment. The solid line represents the calibration target, with a slope of 0.52% between log revenue and probability of adjustment. In panel (b) we group each product-month observation by revenue into percentile bins and compute the average revenue and the average size of adjustment for each bin. Then, we compute the log deviation of this revenue from mean revenue of the data and model respectively. We normalize the size of adjustment by computing the deviation of the size of adjustment of each bin from the mean size of adjustment in the data and the model, and then divide by the mean size to compute the percent deviation from mean size of adjustment. We plot, for each percentile bin, the log deviation from mean revenue and the percent deviation from mean size of adjustment. The two lines in panel (b) show the range of the strength of the relationship between size of adjustment and product revenue that we find in our data. The solid line represents the maximum value (in absolute terms) and the dashed line represents the minimum value.

in the data. The relationship is robust throughout the support of revenue and is approximately consistent in magnitude.

Panel (b) of Figure 1.3 shows the relationship between size of price adjustment and revenue. As in panel (a), we group each product-month observation by revenue into percentile bins and compute the average revenue and the average size of adjustment for each bin. Then, also for the purposes of comparison, we compute the log deviation of this revenue from mean revenue of the data and model respectively. We normalize the size of adjustment by computing the deviation of the size of adjustment of each bin from the mean size of adjustment in the data and the model, and then divide by the mean size to compute the percent deviation from mean size of adjustment. We plot, for each percentile bin, the log deviation from mean revenue and the percent deviation from mean size of adjustment. This corresponds to the statistic $\left(\frac{\partial E[size]}{\partial lnrev} \times \frac{1}{size}\right)$ in Table 1.5.

As in panel (a), the green dots in panel (b) represent the each percentile grouping from the and the orange triangles show the results from model. The two lines in panel (b) show the range of the strength of the relationship between size of adjustment and product revenue that we find in our data. The solid line represents the maximum value (in absolute terms) and the dashed line represents the minimum value. Even though the relationship between the size of price adjustment and product revenue was not explicitly targeted in the calibration, panel (b) shows that the relationship in the model reflects the relationship in the data quite well.

Figure 1.4 depicts the impulse response function of output to a one standard deviation monetary policy shock. The figure is scaled to reflect a response to a monetary shock of 1 percentage point. The solid red line depicts the response of output to monetary policy when output is low, and the dashed blue line represents the response when output is high. The figure illustrates that the impact of monetary policy is stronger in recessions than expansions. The initial response is 10% stronger in a low output state and increases in percentage terms over time. The difference in the cumulative effect, measured as the percent difference of the area under the two impulse response

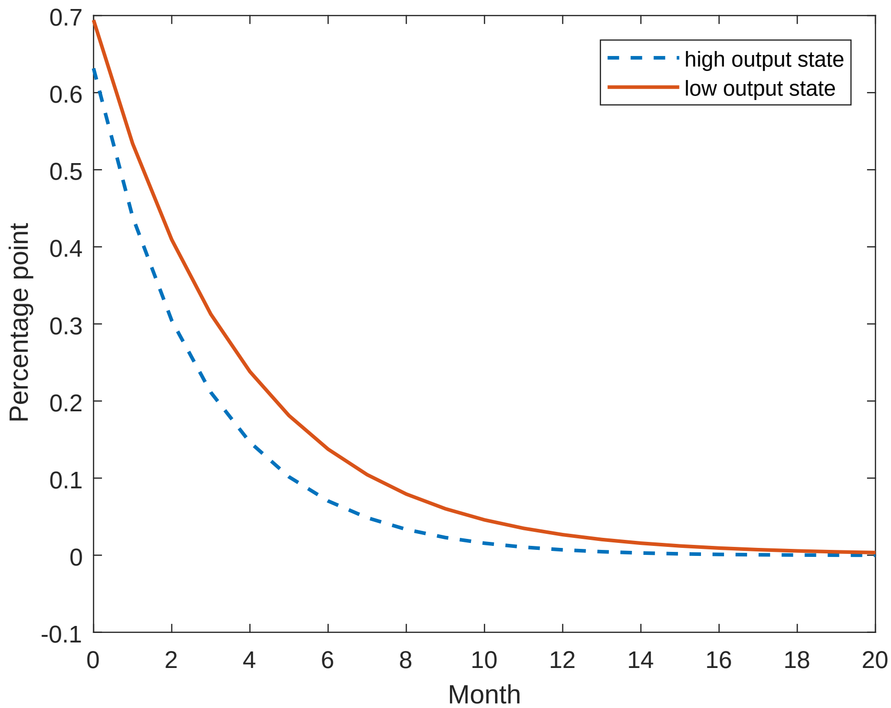


Figure 1.4: Impulse response function of output to a one standard deviation monetary shock in low versus high output states. The results are scaled to reflect a one percentage point monetary shock.

functions, implies that monetary policy has a 43 percent stronger effect on output in recessions than in expansions.

Figure 1.5 and Table 1.7 illustrates the mechanism behind the difference in the output response described in Figure 1.4. Figure 1.5 shows the impulse response for the cross-sectional frequency of adjustment. It shows that the frequency adjustment is greater in high output states than low output states as predicted. There is an initial increase in the frequency of adjustment in the initial period as a response to the monetary shock but the frequency quickly returns to the steady state level. This figure shows that the output response is greater in a low output state due, in part, to the fact that the average frequency of adjustment is lower. Because products adjust their price less often, there is a greater response in output.

Table 1.7 shows the average adjustment probability and the difference in average revenue by

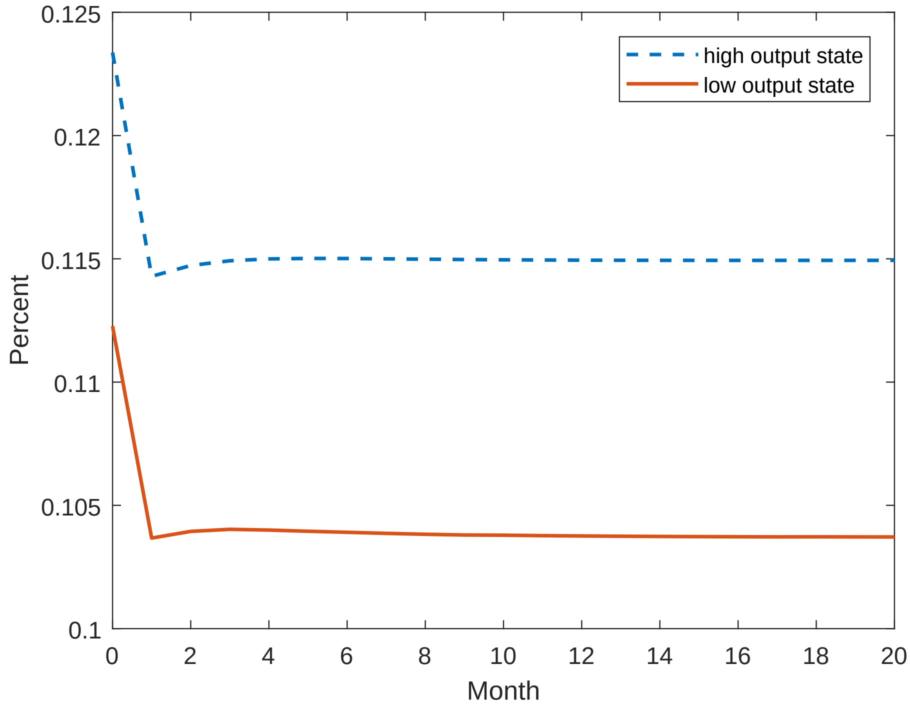


Figure 1.5: Impulse response function of the cross-sectional frequency of price adjustment to a one standard deviation monetary shock in low versus high output states.

Table 1.7: Revenue and adjustment probability by quantile (Model)

	Adj. probability (%)			Revenue (%)
Quantile	High	Low	Diff	Diff
1	6.06	5.99	0.07	6.66
2	9.55	8.69	0.86	8.88
3	11.40	9.18	2.23	8.91
4	12.82	10.29	2.53	10.54
5	14.68	13.58	1.10	13.15

Note: The table shows the average log revenue and average adjustment probability by quantile of revenue in the baseline model. Column labeled High shows the results for the high output state and the column labeled Low show the result for the low output state. Columns labeled Diff show the difference between the high and low output states.

revenue quantiles for the duration impulse response following a monetary shock. The column labeled High shows the results for the high output state and the column labeled Low show the results for the low output state. The columns labeled Diff show the difference between the two states. Table 1.7 illustrates how the distribution of revenue effects the adjustment probabilities. By construction, the difference in the average log revenue between the high and low output state are increasing by quantile. For quantiles 1 through 4, both the difference in the adjustment probability is also increasing by quantile. The difference in the adjustment probability is not strictly increasing in quantile 5 as the inaction region of highest revenue goods are very small such that these products respond to even small changes in marginal cost in both high and low output states. This is reflected in the S-shape of the relationship between log revenue and probability of adjustment in panel (a) of Figure 1.3. The marginal increase in the probability of adjustment to an increase in revenue is decreasing at high levels of revenue. The differences in revenue and adjustment probability by revenue quantile highlights the interaction of the revenue effect and the cyclical movements of the revenue distribution in determining the real effect of output in high and low output states.

1.3.5 Extending the Model

We extend the basic model to account for a more realistic response to monetary policy shocks. In particular, we add external habit formation to consumer preferences and add persistence to the monetary process. A number of papers including Fuhrer (2000) and Christiano et al. (2011) find that adding habit formation to models of monetary policy generates a more realistic hump-shaped response in output and persistent reductions in the real interest rate. We add persistence in the shocks to nominal GDP to reflect the empirical evidence showing persistence in monetary shocks such as in Coibion (2012). Vavra (2014) shows that in menu cost models persistent monetary policy shocks can also generate a hump shaped output response.

Figure 1.6 depicts the impulse response of output when we incorporate habit formation and

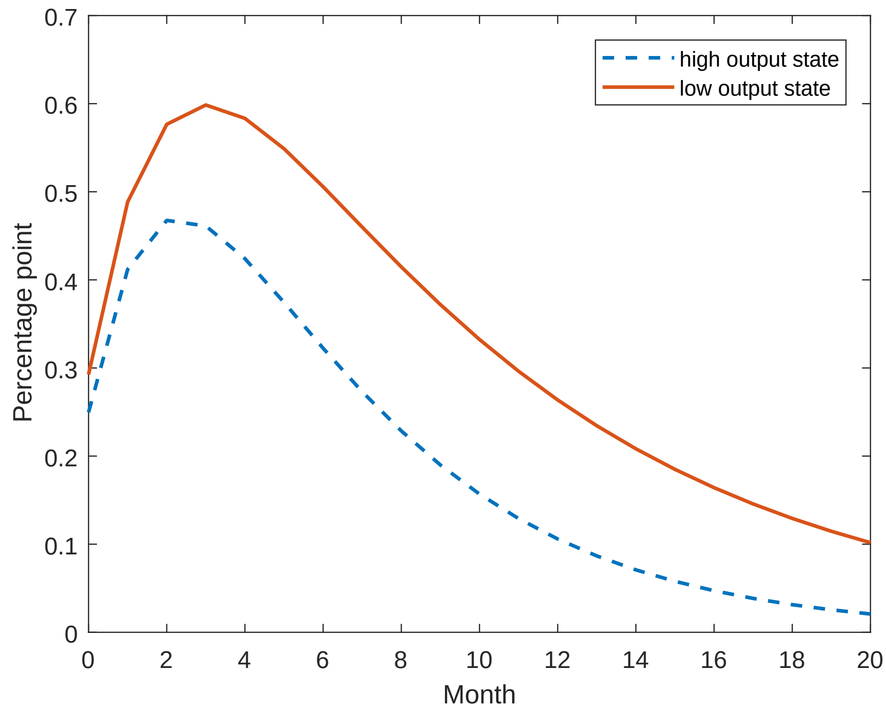


Figure 1.6: Hump shaped impulse response function of output to a one standard deviation monetary shock in a model with habit formation and persistent monetary shocks. The results are scaled to reflect a one percentage point monetary shock.

persistent monetary shocks.[22] The impulse response function has a hump shape as found in the empirical evidence. The degree of state dependence increases significantly. In the baseline model, the cumulative response of output was 43 percent greater in low output states measured as the area under the impulse response function, when the difference in output between high and low states is 7 percent. In the extended model, it is 81 percent greater for the same difference in output. The effect is greater in the extended model because of an accumulation of the differences in individual price changes with persistent monetary shocks. In addition, the change in the firm's marginal cost is greater in percentage terms due to habit formation, increasing a firm's desire to adjust in response to a monetary shock. The degree of state-dependence in the real effect of monetary policy is increasing as the responsiveness of the aggregate price level increases.

[22] A detailed description of the extended model, calibration, and its properties is given in appendix A.4.

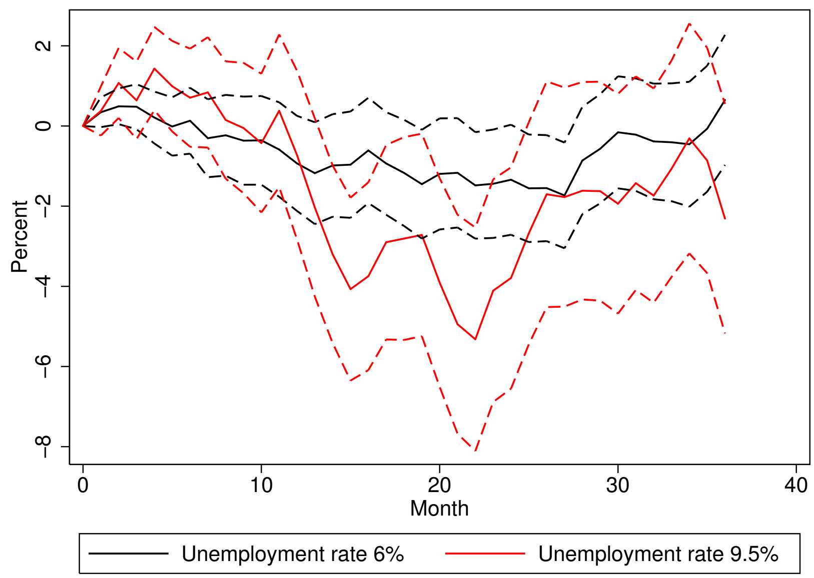


Figure 1.7: State-dependence of aggregate output

Note: The solid red line reports the cumulated response of output to a 1 percentage point shock to the Federal Funds rate when the unemployment rate is 9.5% at the time of the shock. The solid black line reports the cumulated response of output to a 1 percentage point shock to the Federal Funds rate when the unemployment rate is 6%. The dashed lines show one standard deviation bands around the point estimates.

1.4 Empirical Evidence of State-Dependence

In this section we find evidence that the real effects of monetary policy is stronger in high unemployment states than low unemployment states. While there is still debate as to whether the real effects of monetary policy is stronger when the economy is doing well or poorly, our findings add to a growing list of evidence in support of larger real effects of monetary policy in recessions compare to expansions.[23] In addition, because our sample of scanner data used to derive our micro results consist mostly of non-durable retail goods, we also test for state-dependence for durable and non-durable consumption separately to ensure that state-dependence holds for non-durable

[23]Empirical studies that find output response to monetary shocks is significantly larger in recessions than in expansions include, Weise (1999), Garcia and Schaller (2002), Peersman and Smets (2005), Lo and Piger (2005), Santoro et al.(2014), and Barnichon and Brownlees (2016). Tanreyro and Thwaites (2016) and Jorda et al.(2017), on the other hand, find that the effects monetary policy is stronger when the economy is booming.

consumption as well.[24]

We adopt a simple and intuitive approach which allows for clear interpretation. Using identified monetary shocks of Romer and Romer (2004) and extended to 2007 by Wieland and Yang (2017), we estimate the output response as a function of the state of the economy by interacting the monetary shocks with the unemployment rate at the time of the shock. We estimate the equation

$$\Delta y_t = \alpha_0 + \sum_{s=1}^{36} \alpha_s UR_{t-s} + \sum_{s=1}^{36} \delta_{t-s} \Delta y_{t-s} + \sum_{s=1}^{36} \beta_s ms_{t-s} + \sum_{s=1}^{36} \gamma_s (ms_{t-s} \times UR_{t-s}) + \epsilon_t \quad (1.17)$$

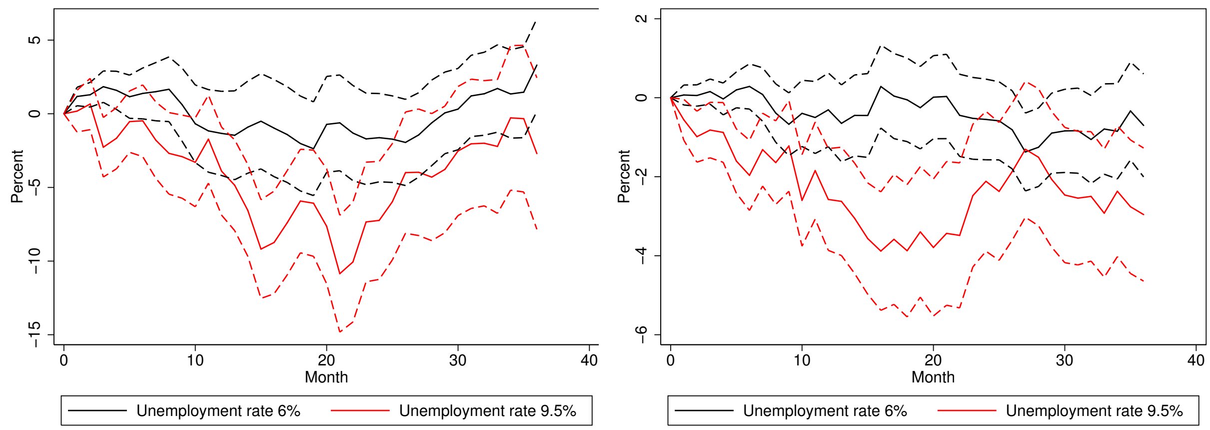
where Δy_t is the monthly growth rate of industrial production, ms_t is the monetary shock, and UR_t is the unemployment rate. The coefficient for the interaction terms γ_s pick up the state-dependent effect of the monetary shocks. The marginal effect of a monetary shock at time $t = 0$ conditional on the unemployment rate is as follows

$$\begin{aligned}
\Delta y_1 &= \beta_1 + \gamma_1 \cdot UR_0 \\
\Delta y_2 &= \delta_1(\beta_1 + \gamma_1 \cdot UR_0) + (\beta_2 + \gamma_2 \cdot UR_0) \\
\Delta y_3 &= \delta_2 \delta_1(\beta_1 + \gamma_1 \cdot UR_0) + \delta_2(\beta_2 + \gamma_2 \cdot UR_0) + (\beta_3 + \gamma_3 \cdot UR_0) \\
&\vdots
\end{aligned}$$

We fix a lag length of 36 months for the dependent variable as well as unemployment, the monetary shocks and the interaction terms. While the information criterion may suggest a shorter lag length, we fix them at the expense of efficiency in hopes of alleviating bias. To permit comparison across models we also fix the lag length across the models.[25]

[24] The data for industrial production, durable consumption, and non-durable consumption as well as the Federal Funds rate come from the Board of Governors of the Federal Reserve System. The Unemployment rate comes from the U.S. Bureau of Labor Statistics.

[25] As in both non-linear VAR and local projection methods the exogeneity of the state of the economy at the time of the shock may be questioned. In order to alleviate these concerns we include lags of relevant variables (growth rates of industrial production as well as unemployment) which is the approach often taken in the local projection methods. Nevertheless, there may be residual concerns about persistent shocks that affect both the variables indicating the state



(a) Durable consumption

(b) Non-durable consumption

Figure 1.8: State-dependence in durable and non-durable consumption

Note: Durable and non-durable consumption refer to the industrial production of durable and non-durable consumer goods respectively. The solid red line reports the cumulated response of output to a 1 percentage point shock to the Federal Funds rate when the unemployment rate is 9.5% at the time of the shock. The solid black line reports the cumulated response of output to a 1 percentage point shock to the Federal Funds rate when the unemployment rate is 6%. The dashed lines show one standard deviation bands around the point estimates.

A commonly adopted approach when studying state-dependence of monetary policy is to use the smooth transition local project method used by Auerbach and Gorodnichenko (2013) in their study of fiscal policy. Both our method and the local projection method have an advantage over VARs in that they are implementable in a single equation. Both methods require a choice of the form of the non-linear component of the response. We believe ours is reasonable insofar as unemployment proxies for the business cycle. Rather than estimating a parametric model for each state of two states, our model allows responses to depend directly on a continuous state. The benefit is clarity. Unlike the smooth transition local projection methods where the transition between states is governed by a logistic link function and a parameter governing the speed of transition, the relationship between monetary policy effects and the state of the economy in our method is simple

of the economy and future output changes that are not captured by the lagged variables. In appendix A.5 we show results including two such potential shocks: changes to TFP to capture persistent changes in aggregate productivity and the Federal funds rate changes to capture persistence in the systematic component of monetary policy. Our results are robust to these checks.

and clear. In addition, our approach requires only a slight modification to the Romer and Romer (2004) setup, allowing for a clear comparison.

In Figure 1.7 we report the cumulative effect of a monetary shock contingent on the state of the economy. The figure displays the average expected response of output to a monetary shock amounting to a 1 percentage point increase in the Federal Funds rate conditional on the state of the economy. The solid red line reports the cumulated response when the unemployment rate at the time of the monetary shock is 9.5%, and the solid black line reports the response for an unemployment rate of 6%. The dashed lines show the one standard deviation bands around the point estimates.[26] The results clearly show that monetary policy has stronger effects in high unemployment states. When the unemployment rate is 9.5% at the time of the shock, the peak effect of monetary policy is 5.8%; when unemployment is 6%, the peak response is only 1.7%.

Figure 1.8 displays the results for durable and non-durable consumption goods. Panel (a) shows the results for industrial production of durable consumption goods and panel (b) shows the results for industrial production of non-durable consumption goods. The aggregate finding that output response to monetary shocks is stronger in recessions holds for both durable and non-durable industrial production and the effect is particularly large for durables. When the unemployment rate at the time of the monetary shock is 9.5% the peak response of durable goods output is 10.8% whereas the peak output response with 6% unemployment rate is 2.4%. The difference in the effect is also large for non-durables, although smaller than for durables. The peak response of output to a monetary shock with 9.5% unemployment rate is 3.9% compared to a peak response of 1.4% when the unemployment rate is 6%.

[26]The standard errors for the cumulative responses are constructed by sampling 1000 times from the asymptotic distribution of the parameters following Romer and Romer (2004).

1.5 Conclusion

Empirically, the frequency of price adjustment is increasing in product revenue, while the average size of price adjustment is decreasing in revenue. These patterns are consistent with menu cost models where menu costs do not scale with revenue and that this finding has important implications for the workhorse menu cost models in general. To our knowledge, our paper is the first to document these patterns.

We document movements in the distribution of product revenue across the business cycle. Not only do the mean and median fall during recessions, but the variance and dispersion of log revenue fall as well. Combined with the strong connection between product revenue and the frequency of price adjustment, the cyclical variation in the distribution of revenue implies that monetary policy should be state-dependent. In recessions, prices are effectively stickier and consequently the real effects of monetary policy should be stronger. We confirm this implication using both reduced-form estimates as well as a quantitative menu cost model. Our reduced-form estimates indicate that monetary policy shocks have greater effects in recessions – this is true for durable and non-durable consumption goods. The quantitative New Keynesian menu cost model matches these predictions. In our baseline model, the cumulative output response is 43 percent larger in low output states when the difference between output in high and low states is 7 percent. When the model is extended to include habit formation and persistent monetary shocks the difference in the output response is 81 percent.

Our findings suggest that examining the interaction of price setting behavior with product characteristics may provide important insight into the behavior of aggregate prices and the real effect of monetary policy. While this paper focuses on a particular characteristic – product revenue – other characteristics such as the elasticity of demand or the degree of local market competition may also be important factors governing price setting. Our findings may also be used to advocate for monetary policy that explicitly takes account the state-dependence of its effects.

1.6 References

Alvarez, Fernando E., and Francesco Lippi, 2014. "Price Setting with Menu Costs for Multi-Product Firms." Econometrica, 82(1): 89-135

Alvarez, Fernando E., Herve Le Bilhan and Francesco Lippi, 2016. "The Real Effects of Monetary Shocks in Sticky Price Models: A Sufficient Statistic Approach." American Economic Review, 106(10): 2817-51

Anderson, Eric, Nir Jaimovich and Duncan Simester, 2015. "Price Stickiness: Empirical Evidence of the Menu Cost Channel." Review of Economics and Statistics, 97(4): 813-826

Auerbach, Alan J., and Yuriy Gorodnichenko. 2013. "Fiscal Multipliers in Recession and Expansion," Chapter in Fiscal Policy after the Financial Crisis, Alberto Alesina and Francesco Giavazzi, editors (p. 63 - 98)

Basu, Susanto, John G. Fernald, and Miles S. Kimball. 2006. "Are Technology Improvements Contractionary?" American Economic Review, 96(5): 1418-1448

Barnichon, Regis, and Christian Brownlees, 2016. "Impulse Response Estimation by Smooth Local Projections." Working paper

Bhattarai, Saroj, and Raphael Schoenle. 2014. "Multiproduct Firms and Price-setting: Theory and Evidence from U.S. Producer Prices," Journal of Monetary Economics, 66: 178-192

Berger, David, and Joseph Vavra, 2016. "Shocks vs Responsiveness: What Drives Time-Varying Dispersion?" Working paper

Bils, Mark, and Peter J. Klenow. 2004. "Some Evidence on the Importance of Sticky Prices." Journal of Political Economy, 112(5): 947-985

Broda, Christian, and David E. Wienstein. 2006. "Globalization and the Gains from Variety." Quarterly Journal of Economics, 121(2): 541-585

Burstein, Ariel, and Christian Hellwig. 2007. "Prices and Markets Shares in a Menu Cost Model." Working paper

Caballero, Ricardo J., and Eduardo Engel. 2007. "Price stickiness in Ss models: New interpretations of old results," Journal of Monetary Economics, 54: 100-121

Caplin, Andrew, and Daniel Spulber. 1987. "Menu Costs and the Neutrality of Money." Quarterly Journal of Economics, 102: 703-725

Carvalho, Carlos, 2006. "Heterogeneity in Price Stickiness and the Real Effects of Monetary Shocks." B.E. Journal of Macroeconomics Frontiers, 2(1):

Christiano, Lawrence J., Mathias Trabandt and Karl Walentin. 2011. "DSGE Models for Monetary Policy Analysis," Benjamin M. Friedman, and Michael Woodford, editors: Handbook of

Monetary Economics, Vol. 3A, The Netherlands: North-Holland.

Coibion, Olivier. 2012. “Are the Effects of Monetary Policy Shocks Big or Small?” American Economic Journal: Macroeconomics, 4(2): 1-32

Coibion, Olivier, Yuriy Gorodnichenko and Gee Hee Hong, 2015. “The Cyclicality of Sales, Regular and Effective Prices: Business Cycle and Policy Implications.” American Economic Review, 105: 993-1029

Correia, Sergio. 2017. “Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator.” Working Paper. http://scorreia.com/research/hdfe.pdf

Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo. 2011. “Reference Prices, Costs, and Nominal Rigidities,” American Economic Review, 101: 234-262

Fernald, John. 2014.“A Quarterly, Utilization-Adjusted Series on Total Factor Productivity.” FRBSF Working Paper, 2012-19

Fuhrer, Jeffrey C., 2000. “Habit Formation in Consumption and Its Implications for Monetary-Policy Models,” American Economic Review, 90(3): 367-390

Garcia, Rene, and Huntley Schaller, 2002. “Are the Effects of Monetary Policy Asymmetric?” Economic Inquiry, 40(1): 102-119

Gertler, Mark, and John Leahy. 2008. “A Phillips Curve With an Ss Foundation.” Journal of Political Economy, 116:533-572

Golosov, Mikhail and Robert E. Lucas Jr. 2007. “Menu Costs and Phillips Curves.” Journal of Political Economy, 115:171-199

Jaimovich, Nir, Sergio Rebelo and Arlene Wong. 2017. “Trading Down and the Business Cycle.” Working paper

Jorda, Oscar, Moritz Schularick, and Alan M. Taylor. 2017. “Large and State-Dependent Effects of Quasi-Random Monetary Experiments,” Working paper

Klenow, Peter J., and Oleksiy Kryvtsov. 2008. “State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation?” Quarterly Journal of Economics, 123(3): 863-904

Krusell, Per, and Anthony Smith. 1998. “Income and Wealth Heterogeneity in the Macroeconomy.” Journal of Political Economy, 106: 867-896

Kumar, Pradeep, and Hongsong Zhang. 2016. “Productivity or Unexpected Demand Shocks: What Determines Firm’s Investment and Exit Decisions?” Working paper

Levy, Daniel, Mark Bergen, Shantanu Dutta, and Rober Venable. 1997. “The Magnitude of Menu Costs: Direct Evidence From Large US Supermarket Chains,” Quartely Journal of Economics, 112: 781-826

Lo, Ming Chien, and Jeremy Piger. 2005. “Is the Response of Output to Monetary Policy

Asymmetric? Evidence from a Regime-Switching Coefficients Model." Journal of Money, Credit and Banking, 37(5): 865-886

Midrigan, Virgiliu. 2011. "Menu Costs, Multiproduct Firms, and Aggregate Fluctuations." Econometrica, 79(4): 1139-1180

Midrigan, Virgiliu, and Patrick Kehoe. 2015. "Prices are Sticky After All." Journal of Monetary Economics, 75: 35-53

Nakamura, Emi, and Jon Steinsson. 2008. " Five Facts about Prices: A Reevaluation of Menu Cost Models." Quarterly Journal of Economics, 123(4): 1415-1464

Nakamura, Emi, and Jon Steinsson. 2010. " Monetary Non-Neutrality in a Multisector Menu Cost Model." Quarterly Journal of Economics, 125(3): 961-1013

Nevo, Aviv. 2001. "Market Power in the Ready-to-Eat Cereal Industry." Econometrica, 69(2): 307-342

Nevo, Aviv, and Arlene Wong. 2017. "The Elasticity of Substitution Between Time and Market Goods: Evidence from the Great Recession." Working paper

Peersman, Gert, and Frank Smets. 2005. "The Industry Effects of Monetary Policy in the Euro Area." Economic Journal, 115(503): 319-342

Romer, Christina, and David Romer, 2004. "A New Measure of Monetary Shocks: Derivation and Implications." American Economic Review, 94(4): 1055-1084

Santoro, Emiliano, Ivan Petrella, Damjan Pfajfar, and Edoardo Gaffeo, 2014. "Loss Aversion and the Asymmetric Transmission of Monetary Policy." Journal of Monetary Economics, 68: 19-36

Vavra, Joseph, 2014. "Inflation Dynamics and Time-Varying Volatility: New Evidence and an Ss Interpretation." Quarterly Journal of Economics, 129(1): 215-258.

Weise, Charles L., 1999. "The Asymmetric Effects of Monetary Policy: A Nonlinear Vector Autoregression Approach." Journal of Money, Credit and Banking, 31(1): 85-108

Wieland, Johannes F., and Mu-Jeung Yang, 2017. "Financial Dampening," Working paper

Zbaracki, Mark J., Mark Ritson, Daniel Levy, Shantanu Dutta, and Mark Bergen, 2004. "Managerial and Customer Costs of Price Adjustment: Direct Evidence From Industrial Markets," Review of Economics and Statistics, 86(2): 514-533

CHAPTER 2

Seniority

2.1 Introduction

A large literature investigates the effects of worker tenure on wage growth, attesting to the importance and interest in this topic.[1] Much of the empirical literature has traditionally focused on firm specific human capital as the source of tenure effects on wage growth. However, some recent papers (for example, Daniel and Heywood, 2007, Heywood et al., 2010, Zwick, 2014) find deferred compensation as a driving force for upward sloping wage profiles with worker tenure. Their findings suggest that firms often back-weight their wage profiles in order to illicit worker effort.

Along with the empirical literature, much of the recent efforts in theoretical research have also focused on the firm-specific human capital explanation for seniority wages (Stevens, 2003, Lazear, 2009, and Burdett and Coles, 2010). On the other hand, there is surprisingly little work that focuses on the need for firms to motivate workers and its effect on wage profiles since Lazear (1979, 1981) first proposed deferred compensation as an efficient method to induce worker effort.

[1]The debate regarding the size and source of the effect of worker tenure on wage growth is still ongoing due in part to the fact there are many difficulties when measuring the effect of worker tenure on wage growth. Some papers that present and address the various challenges in estimating this effect include Altonji and Shakotok (1987), Topel (1991), Neal (1995), Altonji and Williams (2005), Kambourov and Manovskii (2009), Buchinsky et al. (2010), Pavan (2011), and Kwon and Milgrom (2014) among others.

While Lazear's earlier work was crucial in first illustrating the effectiveness of deferred payment as a worker motivating device, there is still much left to be understood about how it pertains to seniority and its effect on tenure earnings gradients.

An important goal of this paper is to provide a general framework in which agency frictions lead firms to value seniority and consequently offer wage profiles that reflect this valuation. We derive the optimal wage structure for the firm and provide a precise definition of seniority and its value to the firm in this context. This provides important insight as to what it means for a worker to gain seniority with tenure, which also enables a much richer interpretation of the empirical results. Our model also provides sharp predictions regarding not only the evolution of worker wages but also regarding the intensity with which firms monitor their employees and provide amenities.

We present a model in which workers receive significant returns to seniority above and beyond their productivity increase as an optimal equilibrium outcome. The model we present is an efficiency wage model with imperfectly verifiable effort, where firms hire workers by offering them implicit long-term contracts. Unlike most existing models of efficiency wages where the choice of effort is binary (work or shirk), we allow the workers to choose a desired level of effort at every instant in time.

We find that in this setup a seniority wage structure is optimal for firms. Workers are paid according to their seniority, a measure that does not directly reflect the worker's current productivity. Seniority is measured as the accumulation of marginal productivity gains gained by the firm from increased worker effort. Workers accumulate this seniority as they continuously work for a firm and their wages increase accordingly. Firms are willing to pay for this seniority because it reflects the total the benefits that a firm gets from providing additional utility to the worker.

In addition to the result for the optimal wage path, we also find that seniority can determine the level of monitoring a firm performs over the workers. If firms have the ability to vary their monitoring intensity by worker, the firms will most likely choose to monitor junior workers more intensely and senior workers less. We also consider the case in which firms provide various amenities to workers in addition to wages. We find that how much a firm provides of a particular benefit will depend not only on the amount utility it provides to the worker but also but its effect on the marginal disutility of effort for the worker.

The seniority wage structure also has interesting implications for the issue of earnings losses from job loss. Previous studies have found large and persistent losses in wages (Davis and von Wachter, 2011) that are difficult to reconcile with existing theories. Here, we show that seniority wages can provide an additional channel through which a job loss can result in large earnings losses for a worker.

As mentioned above, Lazear (1979, 1981) first showed that deferred compensation may be utilized by firms to motivate worker effort. However in his papers, the precise path of wage profiles is indeterminate as long as it provides enough motivation for the workers. In addition, he allows for the possibility that workers pay employment fees (or provide up-front bonds) which is seldom observed in the real economy. Akerlof and Katz (1989) argue that when one does not allow up-front bonds, firms should prefer to use wages higher than market clearing wages rather than choosing to defer payment. We find that when the effort choice of workers is continuous and not binary, upward sloping wage profiles will still be utilized by firms to motivate workers.

In addition to the work on seniority wages, this paper contributes to the study of efficiency wage models and their properties. Efficiency wage was once a popular area of research, providing much insight into involuntary unemployment (Shapiro and Stiglitz, 1984), the prominence of inter-industry wage differentials and the effect of industrial policies (Bulow and Summers, 1986). However for reasons that are unclear, the amount of research conducted on the topic declined drastically and there is little research currently being done on efficiency wages.

There are many reasons to believe that there should be more research focused on efficiency wages models and their implications. First, as the abundance of search models indicate, there is still a high degree of interest in the determinants of unemployment. However, the current state of research is dominated by the search and matching framework. While search frictions are undoubtedly a very important component of unemployment, other sources of unemployment also deserve attention.

The main alternative theory of unemployment in the economist's current tool box is efficiency wage theory where unemployment stems from the firms' need to motivate workers. Efficiency wage models provide an alternative prism through which to study the sources and consequences of unemployment. How much of unemployment can be attributed to the existence of search frictions and how much is motivational unemployment is ultimately a quantitative question that could potentially have important consequences for not only our understanding of unemployment but for business cycle fluctuations as well.

The nature of the employment relationship is also another area of great interest that could benefit from efficiency wage theory. Similar to search models, efficiency wage models are intrinsically about long-term labor relationships which contrasts with the traditional view of wages being determined in spot markets. We believe that efficiency wage models can add interesting perspective to the question of the employment relationship as well, including on the longstanding question of whether observed wages should be thought of as being allocative. We view our paper as contributing to this cause.

Also, this paper is a part of a literature that studies the implications of variable labor choice. Studies on this topic include Basu and Kimball (1997) who study the implications of variable labor utilization and also Basu and Kimball (2002) and Epstein and Kimball (2014) that study the implications of the stability of long-run labor supply. Nevertheless, while most of this literature focuses on variable labor choice in a neoclassical framework, this paper studies the implications of variable effort choice in an efficiency wage model.

Finally, the solution technique we utilize to solve the agents' problems in the model deserves attention. The agents in our model solve an optimization problem where the expectations of other agents are a forward-looking state variable. The structure of this problem makes it difficult to utilize traditional techniques of dynamic optimization to solve this problem. The solution technique we propose allows us to reformulate the problem in a manner such that we can utilize standard optimal control methods to solve the optimization problems. This method can be used various other situations where a forward-looking variable is a state variable of the optimization problem. Examples of situations where such problems could arise include monetary policy, capital taxation and even foreign policy.

The paper is organized as follows. In section 2, we present and solve the worker and firm problems. We also discuss the the solution technique. Section 3 analyzes the optimal contract using graphs and phase diagrams. Section 4 considers extensions and applications of the framework and section 5 discusses the steady state equilibrium. Section 6 concludes.

2.2 Model

The model is presented in continuous time and set in a partial equilibrium framework where interest rates are exogenous and are quoted in terms of the consumption good as numeraire. Capital is fixed, so output is a function of effective labor input alone. There are two types of agents in the model, firms and workers. All workers are ex-ante identical as are all firms. Each firm hires many workers to produce output.

An employed worker works for a fixed number of hours. While employed, a worker chooses the amount of effort to exert at every instant in time. We assume that this effort level is only partially observable to the firm. Firms can only verify whether or not a worker makes a bad blunder, the frequency of which will depend on the worker's effort level.

Firms hire in a competitive labor market by offering workers (implicit) long-term contracts. For each cohort of workers they hire, firms choose a wage path and a layoff policy, both of which may be conditional on macroeconomic events. They may also choose the level of monitoring intensity, which affects the accuracy with which the firm can verify workers' blunders.

The instantaneous probability of dismissal $m(1-z)$ is determined jointly by the firm's monitoring intensity m and the worker's effort level z. Together with the exogenous quit rate q this makes up the separation or attrition rate $a(m, z) = q + m(1 - z)$. In the expression for the probability of dismissal $m(1 - z)$, we implicitly assume that the probability of making an observable blunder is linear in effort. Because we assume that what matters for a firm's productivity is the total effective labor of all its workers, which is the sum of the effort of all the individual workers, this implies that there is a linear relationship between a worker's productivity and probability of dismissal. While this assumption may hold merit in its own right, the main purpose of the assumption is to simplify the analysis.

Because a firm cannot observe effort directly, it can only keep worker effort up by severely punishing any worker who gets caught. One way to punish workers who get caught would be to levy a stiff fine or to confiscate a performance bond put up by the worker when hired. In theory this could circumvent the need to fire workers. However, explicit bond posting is seldom observed in actual labor markets. There are many reasons as to why explicit bond posting is rarely observed, including moral hazard of firms and capital market imperfections.

The approach taken in this paper is to rely on capital market imperfections to prevent explicit bonding. We assume that workers cannot borrow or lend on the bond market. Being unable to borrow at reasonable rates is plausible in a world where workers have no collateral and may lose their job at any time especially since taking out a loan would reduce the worker's effort if the debt could be wiped out by bankruptcy when the worker becomes unemployed.

Being unable to save is less realistic, although the workers may be unwilling to lend because of a high rate of discount or because they receive a lower interest rate than the capitalists when they lend. We assume that the worker's discount rate ρ is greater than the interest rate through out the remainder of the paper. In any case, allowing workers to save will not drastically alter the qualitative implications of this paper.[2]

The worker's problem

A worker chooses a path of planned effort $\{z_t\}$ to maximize life time expected utility. The worker receives a flow utility of $u(w) - v(z)$ from work when employed and a flow utility of $\bar{u}$ when unemployed. The level of utility a worker receives when unemployed $\bar{u}$ is exogenously determined.

The disutility of effort function $v(z)$ should be thought of as combining elements of worker preference and elements of the production function, since effort z is measured in effectiveness of the effort in a particular job relative to the effectiveness of maximum effort in that job. Therefore if one job took more care and attention than another in order to avoid mistakes, this would be expressed as a high level of disutility necessary to get a reasonable level of effectiveness z. In a model where worker preference and the characteristics of jobs vary the two elements of $v(z)$ should be analytically separated, but for the purposes of this paper the distinction is without consequence.

Denoting the expected future utility of an unemployed worker at time t as $U_u(t)$, the expected future utility of an employed worker at time t_0 is

$$U_E(t_0) = \int_{t_0}^{\infty} e^{-\int_{t_0}^{t}(\rho + a(m, z_{\tilde{t}}))d\tilde{t}}[u(w_t) - v(z_t) + a(m, z_t)U_u(t)]dt \tag{2.1}$$

where ρ is the time discount factor of the worker. The worker gets expected utility from the chance of staying with the firm and getting $u(w_t) - v(z_t)$ at future dates t. In addition, if the worker

[2] This can be seem from the fact that the workers first order condition for effort choice (2.2), which is also critical for the firm's problem, remains unchanged even if one allows the workers to save.

is fired at date t he or she receives expected utility of $U_u(t)$ from everything after time t'. The instantaneous probability of being fired at time t depends not only on the rate of dismissal but also the chance of lasting that long as is reflected in the discount factor.

The worker chooses a path of effort $\{z_t\}$ to maximize their expected future utility as described by equation (2.1). Solving this maximization problem is not trivial, most notably due to the fact that the endogenous variable z shows up in the discount factor $e^{-\int_{t_0}^{t}(\rho+a(m,z_{\tilde{t}}))d\tilde{t}}$. The solution technique used to solve the workers problem deserves more attention and the derivation of the following first order condition is discussed in detail below. The first order condition for z is

$$-a_z(U_E - U_u) = v'(z). \tag{2.2}$$

A worker applies effort up to the point where the marginal distuility of effort is equal to the marginal reduction in the probability of job loss ($-a_z$) multiplied by the cost of job loss, the utility differential $U_E - U_u$. Denoting the utility differential by $\Delta \equiv U_E - U_u$ and since $a_z = -m$ we can rewrite equation (2.2) as

$$m\Delta = v'(z). \tag{2a}$$

Assuming $v'(z)$ is strictly increasing in z ($v''(z) > 0$) then equation (2a) can be inverted to yield

$$z = v'^{-1}(m\Delta). \tag{2b}$$

Effort depends only on the product of monitoring intensity and the cost of job loss. When m is treated as exogenous, we can write z as a function of simply the expected utility differential $z = z(\Delta)$.

Unemployed workers have no decisions of consequence to make. They receive a certain reservation utility and wait to be hired by firms choosing randomly from the pool of unemployed workers. For unemployed workers, the present value of future utility comes from two sources: the flow utility one gets when unemployed $\bar{u}$ and the probability of becoming employed again with a consequent jump in expected utility. The instantaneous probability of being reemployed is $\frac{h}{N-L}$, where h is the aggregate rate of hiring, N the whole labor force, and L the number of those employed. The jump in utility on being reemployed is the utility differential Δ for a newly hired worker which we denote Δ_0. Thus the expected future utility of an unemployed worker is,

$$U_u(t) = \int_t^\infty e^{-\rho(t'-t)}[\bar{u} + \frac{h_{t'}}{N_{t'} - L_{t'}}\Delta_{0,t'}]dt' \tag{2.3}$$

where we define,

$$\psi = \bar{u} + \frac{h}{N - L}\Delta_0 \tag{2.4}$$

as the pseudo-utility of being unemployed. This pseudo-utility accounts for the potential benefit an unemployed worker may gain from being rehired as well as the reservation utility $\bar{u}$. In the model, ψ functions as the "going wage" (or more accurately the going utility) which a firm must on average exceed to illicit effort from workers.

Given the expressions for U_E and U_u we can derive an expression for the utility differential Δ as

$$\Delta_t = \int_t^\infty e^{-\int_t^{t'}[\rho + a(m,z_{\tilde{t}})]d\tilde{t}}[u(w_{t'}) - v(z(\Delta_{t'})) - \psi_{t'}]dt'. \tag{2.5}$$

From this expression, we can see that $u(w) - v(z) - \psi$ is the worker's instantaneous surplus from employment. This surplus utility must be discounted by the attrition rate $a(m,z)$ as well as by the impatience parameter ρ, since with instantaneous probability $a(m,z)$ the worker becomes unemployed again worker surplus drops to zero.

Note that ψ is exogenous to the firm. A small firm cannot affect the value of a worker's opportunities when unemployed is seeking work. Δ_0 is also outside of a firm's control being a part of the solution to another firm's maximization problem.

Solution technique

Before proceeding to the firm's problem, the solution technique used to solve the employed worker's maximization problem needs to be discussed. The same technique is utilized to solve the firm's problem as well. We will call this method the "Retrograde approach." The definition of the word "retrograde" that we have in mind as defined by the Merriam-Webster dictionary is "moving, occurring, or performed in a backward direction." The method we propose treats the endogenous variables as if they are in retrograde motion, traveling backward in time. This interpretation allows us to reformulate the problem in a way that allows us to utilize standard Hamiltonian methods.

We demonstrate this method using the worker's problem from the previous section. As described above, at time t_0 the worker solves the problem,

$$\max_{\{z_t\}} \quad U_E(t_0) = \int_{t_0}^{\infty} e^{-\int_{t_0}^{t}[\rho + a(m,z_{\tilde{t}})]d\tilde{t}}[u(w_t) - v(z_t) + a(m,z_t)U_u(t)]dt$$

where the path of wage $\{w_t\}$ and expected future utility of unemployment $\{U_u(t)\}$ are taken as given. The problem cannot be solved using the usual optimal control techniques because the endogenous variables affect the present value of all future flow utilities. However, using the Retrograde approach we are able to reformulate the problem so that we can use the standard Hamiltonian.

The first key insight is that the function U_E itself can be thought of as a state variable. We can reformulate the problem as,

$$\max_{\{z_t\}} \quad U_E(t_0) \tag{2.6}$$

$$s.t. \quad U_E(t) = \int_{t}^{\infty} e^{-\int_{t}^{t'}[\rho + a(m,z(\tilde{t}))]d\tilde{t}}[u(w_{t'}) - v(z_{t'}) + a(m,z_{t'})U_u(t')]dt' \qquad \forall t \geq t_0$$

so that the endogenous discounting is no longer a problem. Here, we treat U_E as a *forward-looking* state variable in the maximization problem with its law of motion described by an integral equation. The objective is to maximize the value of U_E at the initial point in time t_0.

The problem, as currently formulated, still has a different structure than the usual optimization problems in economics. The optimal control problem that economists are most used to is the case with a backward-looking state variable accumulating over time according to a law of motion described by a differential equation. This is the standard case most often seen with consumption as the control variable and capital as the state variable. In this usual case, we solve the problem by selecting the path of the control variables to "control" the path of the state variable such that the two paths jointly maximize the objective function.

Thus we wish revert the problem in (2.6) into this familiar form. Note that since the state variable in (2.6) is forward-looking it accumulates backwards in time. So, now suppose that time flows backwards. Then, the state variable accumulates according to the law of motion

$$\frac{dU_E(t)}{-dt} = -(u(w_t) - v(z_t) + a(m, z_t)U_u(t)) - (\rho + a(m, z_t))U_E(t).$$

In other words, the worker's problem can be thought of as a standard optimization problem where the direction of time is reversed. $U_E(t_0)$ is the objective function and U_E is the state variable with the above law of motion.

This Retrograde interpretation provides an interesting economic context to the maximization problem. The problem can be thought of a worker trying to maximize expected utility by choosing effort z conditional on the state variable U_E. The worker accumulates this utility stock U_E by collecting the instantaneous utility benefit $u(w_t) - v(z_t) + a(m, z_t)U_u(t)$ over time with U_E depreciating each period by $\rho + a(m, z_t)$. In this interpretation, the worker is analogous to an agent in the familiar Ramsey problem of agents maximizing expected utility by choosing consumption while accumulating capital through savings. The difference is, the direction of time is reversed and the accumulation is occurring in reverse time.

The retrograde structure appears in the worker's problem because the worker's choice of effort in the present and near future affects the present value of utilities gained in all future periods. Thus the worker must always account for how effort choice today will affect the present value of future expected utility, making effort choice depend on this value. The expected utility differential Δ_t is, therefore, the appropriate state variable of the worker.

While the retrograde interpretation gives a natural insight to the workers problem, the direction of time flow itself is irrelevant to the solution of an optimal control problem. Thus we can simply write,

$$\max_{\{z_t\}} \quad U_E(t_0)$$

$$s.t. \quad \dot{U}_E(t) = -\Big((u(w_t) - v(z_t) + a(m, z_t)U_u(t)) - (\rho + a(m, z_t))U_E(t)\Big)$$

and solve the problem using the Hamiltonian equations. The Hamiltonian has the form,

$$H = -\xi\Big([u(w) - v(z) + a(m, z)U_u] - [\rho + a(m, z)]U_E\Big) \tag{2.7}$$

where z is the control variable, U_E is the retrograde state variable and ξ is the retrograde co-state variable. The objective function does not show up in the Hamiltonian because there is no flow utility that is directly affected by the values of the endogenous variables.

The Hamiltonian equations are textbook (Kamien and Schwartz, 1991). The derivative of the Hamiltonian with respect to the control variable is set to equal zero ($\frac{\partial H}{\partial z} = 0$), the derivative with respect to the state variable is set to equal $-\dot{\xi}$ ($\frac{\partial H}{\partial U_E} = -\dot{\xi}$), the derivative with respect to the co-state variable is set to equal $\dot{U}_E$ ($\frac{\partial H}{\partial \xi} = \dot{U}_E$), and the transversality condition is $\lim_{t\to\infty} \xi_t U_E(t)$. The only unorthodox condition is the condition for the free initial point, $\xi(t_0) = 1$. Finally, using the

Retrograde Hamiltonian equations we can derive the first order condition in (2.2)

$$-a_z(U_E - U_u) = v'(z).$$

The firm's problem

Firms hire workers by offering each worker an implicit long-term contract. We assume that firms can commit to a conditional wage path. This assumption is important as firms will have the incentive to renege on its promises as the value of the worker to the firm falls. As will be shown, it is beneficial for firms to defer payments to induce effort from their workers. As a result the worker's wages will exceed their productivity toward the latter parts of their careers.

It is not difficult to imagine why firms will be willing to honor their contracts even when workers are paid more than their marginal products. The firm's ability to hire workers and entice them to exert effort depends on the worker's belief that the firm will honor its promises. If a worker believes that a firm is untrustworthy after it observes a firm renege on a contract, then it may be better for the firm to honor its contracts rather than attain a bad reputation. In a game theoretical setting, it may be the firm's optimal strategy to honor all contracts if the worker's punishment strategy for broken promises is stringent enough. There can be potentially many different strategies of the workers that enforce the firm to honor all contracts in equilibrium. The approach we take here is not to explicitly model such a situation, but rather suggest that it is plausible to believe firms would indeed be sensitive to their reputation and in return assume that firms are honorable. Another possible interpretation is that these contracts, implicit as they may be, are nevertheless legally enforceable and firms are bound by law to keep their promises.

Each firm hires many workers and produce according to an identical the production function

$$pf(Z)$$

where p is productivity (or can also be thought of as the real price of output), Z is the total effective labor done for the firm by all its workers, and f is an increasing and concave function of Z ($f' > 0, f'' < 0$). A firm chooses the wage path of a worker taking the path of the marginal revenue product of effort $\mu_t \equiv p_t f'(Z_t)$ as given, and taking as given that the firm will only fire for cause. μ_t is only infinitesimally affected by what any one worker does so it can be treated as exogenous in drawing up the contract. For now the monitoring intensity m will be treated as constant and thus $z = z(\Delta)$.

For the firm, an optimal contract is a contract that maximizes the marginal profit gained from a worker over the worker's career with the firm. Denoting the expected value of discounted profit gained from a worker from time t_0 as π_{t_0}, the optimal contract is the solution to the firm's problem,

$$\max_{\{w_t'\}} \pi_{t_0} = \int_{t_0}^{\infty} e^{-\int_{t_0}^{t}[r_{\tilde{t}}+a(m,z_{\tilde{t}})]d\tilde{t}}[\mu_t z(\Delta_t) - w(t)]dt \tag{2.8}$$

such that Δ_t satisfies (2.5) for all $t > t_0$. The firm wishes to maximize the sum of the product gained from a worker μz minus wages w, discounted by the interest rate r and the probability of separation $a(m, z)$. Again, we solve this problem using the retrograde approach, and reformulate the problem as,

$$\max_{\{w_t\}} \pi_{t_0}$$

$$\begin{aligned} s.t. \qquad \dot{\pi}_t &= -\Big([\mu_t z(\Delta_t) - w_t] - [r_t + a(m, z(\Delta_t))]\pi_t\Big) \\ \dot{\Delta}_t &= -\Big([u(w_t) - v(z(\Delta_t)) - \psi_t] - [\rho + a(m, z(\Delta_t))\Delta_t]\Big) \end{aligned}$$

and proceed to solve the problem using the Hamiltonian. The Retrograde Hamiltonian for the firm's problem is,

$$H = -\lambda\Big([\mu z(\Delta) - w] - [r + a(m, z(\Delta))]\pi\Big) - \zeta\Big([u(w) - v(z(\Delta)) - \psi] - [\rho + a(m, z(\Delta))]\Big) \quad (2.9)$$

where λ is the retrograde co-state variable for π and ζ is the retrograde co-state variable for Δ. The solution to the problem is,

$$\frac{1}{u'(w_t)} = \frac{\zeta_t}{\lambda_t} \equiv s_t \quad (2.10)$$

The ratio s_t is the measure of "seniority rights" and is described by the expression,

$$s_t = \int_{t_0}^{t} e^{-\int_{t'}^{t}[\rho - r_{\tilde{t}}]d\tilde{t}}[\mu_{t'} + m\pi_{t'}]z_{\Delta,t'}dt' \quad (2.11)$$

where $\pi_{t'}$ is the value of profits obtained from the worker after time t'. Equation (2.10) shows that the wage is a monotonic function of seniority. For example, if $u(w) = \frac{w^{1-\sigma}}{1-\sigma}$, then it implies that $w_t = s_t^{\frac{1}{\sigma}}$ where $\frac{1}{\sigma}$ is the intertemporal elasticity of substitution.

The units of "seniority" as implied by (2.10) are revealing. ζ_t, the co-state variable for Δ_t, is the marginal value of the utility differential Δ_t to the firm. λ_t, the co-state variable for π_t, is the marginal value of the future expected profit. Thus s_t is the value of the utility differential in units of future expected profit and is measured in dollars per utility. The higher a worker's seniority the more a firm is willing to pay for an extra unit of utility of that worker's.

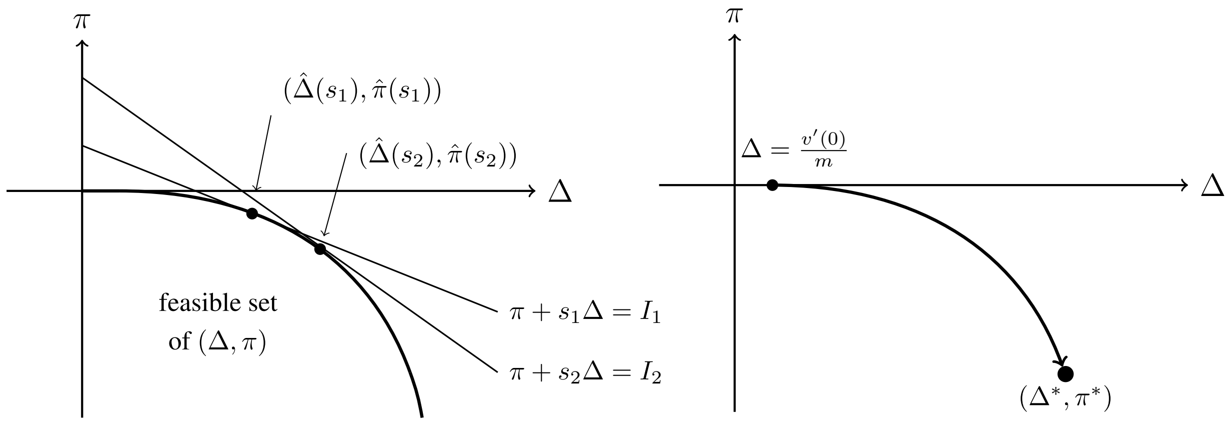
Equation (2.11) indicates that s_t is an accumulation of past marginal effects of the utility differential Δ on effort multiplied by the value of marginal effort in terms of firm profit. Worker utility at time t is valued by the firm because of its effect on effort over the interval $[t_0, t]$ as reflected in the interval of integration in equation (2.11). A firm is willing to pay for a unit of utility later on because utility off in the distance acts as a carrot for a long time since increased worker efforts comes not from utility in the present but from looking forward to the utility one will get if not fired.

The factor $\mu_{t'} + m\pi_{t'}$ in equation (2.11) is the value of increased effort to the firm. It is the combination of a direct effect of increased output per extra unit of effort plus the reduction m per unit of effort on the instantaneous probability of losing the profit $\pi_{t'}$ from the worker's activities after t'. The effect of the increased effort on the worker's utility (which the firm cares about indirectly) cancel out on the margin because of the worker's first order conditions.

The term $e^{-\int_{t_0}^{t}(\rho-r)d\tilde{t}}$ occurs because utility at time t has to be discounted at the rate $\rho + a(m,z)$ back to t_0, to reflect the effect on the cost of job loss at t', while dollars written in to the contract at time t are cheaper by a factor discounted at the rate $r + a(m,z)$. Utility at t is worth less if the worker's utility discount rate ρ is high but worth more in dollars at time t if those dollars are easy to come by because r is high. The effects of the attrition rate $a(m,z)$ cancel out.

The term $z_{\Delta,t'}$ is the easiest to understand. This is the marginal effect of a worker's expected utility on effort which translates the utility the worker is concerned with into something the firm is concerned about.

Equation (2.11) is still not fully understood until one notes that in a steady state situation the expected present value of future profits from previously hired workers will be negative. The firm will hire workers up to the point where expected profit from the time of hiring π_{t_0} is zero. But the firm tries to back-weight the path of wages so that most of the wages are off in the future which will push wages further off into the future than productivity Therefore, after the worker has been around for a while the firm will face relatively more future wages than future productivity. Because the firm promises wages off in the future and gets the benefits before it delivers on its part of the bargain, a worker ends up being a contractual burden on the firm. As a result, the fact that a worker is more likely stay with the firm if their effort is higher makes the firm less eager to elicit greater effort by promised wages later on.



(a) (b)

Figure 2.1: Optimal contract in (Δ, π)

2.3 The Optimal Contract

The optimal contract can be analyzed further with phase diagrams. The contract lies in the three dimensional space of (π, Δ, s) with the dynamics of the contract given by the law of motion of the three variables. The optimal contract will be on the three dimensional saddle path of the dynamics described by the differential equations,

$$\begin{aligned}
\dot{\pi} &= -\Big((\mu z(\Delta) - w) - (r + a(m, z(\Delta)))\pi\Big) \\
\dot{\Delta} &= -\Big((u(w) - v(z(\Delta) - \psi) - (\rho + a(m, z(\Delta)))\Delta\Big) \\
\dot{s} &= (\mu + m\pi)z_\Delta(\Delta) - (\rho - r)s
\end{aligned}$$

We depict the three dimensional diagrams in two dimensional space from three different angles. While it is difficult to directly draw a phase diagram in (Δ, π) space as $\dot{\Delta}$ is only an indirect function of π, it is also the most illuminating. Thus we begin by analyzing the contract from this perspective.

If a worker is hired at time t_0 the firm will value worker utility at $t > t_0$ at s_t dollars per utility. Since Δ_t is measured in time t utility units it can be valued on the margin at s_t dollars per utility unit. Therefore, the contract from t on will maximize expected profits from $\pi_t + s_t\Delta_t$. In other words, the contract after time t will solve,

$$\max_{\{w_t\}} \pi_t + s_t\Delta_t \tag{2.12}$$

such that Δ_t satisfies (2.5), π_t satisfies (2.8) for time t and s_t is taken as given.

To analyze this contract graphically, consider that at any time t, there will be some bounded-above set of attainable combinations of Δ and π given the expected path of the three key variables exogenous to the contract μ, ψ and r. An optimal contract will always be a point on the efficient frontier of this $\Delta - \pi$ possibility set. In Figure 2.1a the efficient frontier is shown by the thick curved line. It is a subset of the upper right-hand boundary of the convex hull of attainable (Δ, π) pairs.[3]

The set of possible (Δ, π) pairs will always be in the southeast quadrant. This is due to the fact that in equilibrium, if π is greater than zero the firm would immediately hire more workers until lower marginal product μ brought the expected marginal profit from a worker back down to zero. Δ is always greater than zero since a contract with Δ less than zero would be less attractive to workers than unemployment.

Different levels of seniority which are likely to coexist within the firm at any given time will each lead to a different choice of Δ and π. Let the functions $\hat{\Delta}(s)$ and $\hat{\pi}(s)$ denote the optimal Δ

[3]The efficient frontier can be disconnected if the region of possible (Δ, π) pairs is non-convex. Non-convexity may arise if firms are able to place workers on temporary leave. If Δ falls such that the implied level of effort z is less than zero, it is beneficial for the firms to temporarily layoff the workers (with pay) and get zero effort instead. This arrangement would not be a breach of the longterm contract as the firms would still honor the terms of their agreement. In any case, in steady state layoffs are unnecessary and the (Δ, π) possibility set will be convex.

and π for a given level of seniority s. Then the functions $\hat{\Delta}(s)$ and $\hat{\pi}(s)$ are respectively nondecreasing and non-increasing. Also, $\hat{\pi}(s) + s\hat{\Delta}(s)$ is nondecreasing.[4] This can be seen most clearly through figure 2.1a. First, note that for a given level of seniority s, the objective $\pi + s\Delta$ is shown as a straight line with the slope $-s$. Since firms maximize this objective, the optimal contract will be the tangent point of this line with the efficient frontier of the feasible set. Now, because the efficient frontier is downward sloping, we can see that as s increases the optimal point (Δ, π) will move southeast implying $\hat{\pi}(s)$ decreases and $\hat{\Delta}(s)$ increases. The value of $\pi + s\Delta$ is depicted as the π-intercept of the tangency line. The intercept also increases with seniority s. From Figure 2.1a we can see that the firm is willing to trade off more expected subsequent profit for a higher expected utility as seniority s rises.

These results for a fixed (Δ, π) frontier have two meanings. The first is cross-sectional. At any given time a firm is likely to have workers with various levels of seniority but the contacts for all of the workers must be from the feasible set. Therefore, the graph shows that more senior workers will be more of a burden on the firm (π is more negative), will value their jobs more (Δ is higher), have higher productivity ($z(\Delta)$ is higher), and are less likely to be laid off ($a(m, z)$ is lower).

The other use for these results is intertemporal. In steady state, the (Δ, π) possibility frontier will stay fixed over time. As time passes, any worker who neither quits nor is fired will gain seniority s and move down along the (Δ, π) possibility frontier.[5] Therefore the dynamic path of Δ and π for a single worker when the economy as a whole is in steady state can be shown by the (Δ, π) frontier as in Figure 2.1b. As the figure shows, the dynamic path of Δ first turns down at $\Delta = \frac{v'(0)}{m}$ which is the point at which the worker is hired. In equilibrium, the worker's effort

[4]Given $s_2 \geq s_1$, let (Δ_1, π_1) maximize $\pi + s_1\Delta$ and (Δ_2, π_2) maximize $\pi + s_2\Delta$ in the feasible set. First, note that $\pi_1 + s_1\Delta_1 \geq \pi_2 + s_1\Delta_2$ and $\pi_2 + s_2\Delta_2 \geq \pi_1 + s_2\Delta_1$ by definition. Subtracting the two conditions, we get $(s_2 - s_1)\Delta_2 \geq (s_2 - s_1)\Delta_1$. Which implies that $\Delta_2 \geq \Delta_1$ and $\hat{\Delta}(s)$ is non-decreasing. Next, from $\pi_1 + s_1\Delta_1 \geq \pi_2 + s_1\Delta_2$ we get $\pi_1 \geq \pi_2 + s_1(\Delta_2 - \Delta_1) \geq \pi_2$ which shows $\hat{\pi}(s)$ is non-increasing. Lastly, $\pi_2 + s_2\Delta_2 \geq \pi_1 + s_2\Delta_1 \geq \pi_1 + s_2\Delta_1 - (s_2 - s_1)\Delta_1 = \pi_1 + s_1\Delta_1$ which shows $\hat{\pi}(s) + s\hat{\Delta}(s)$ is non-decreasing.

[5]In steady state, once a worker is called in to work, he or she will not be placed on leave, since the worker's Δ and z will only increase. Because temporary layoffs are unnecessary in steady state, the (Δ, π) possibility set will be convex; the possibility of temporary layoffs is the only thing that introduces non-convexity into the problem.

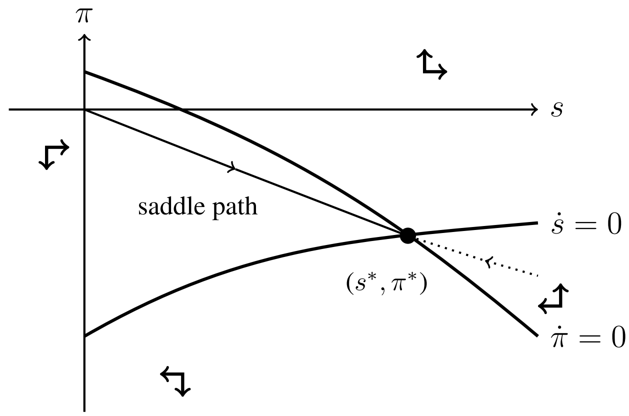


Figure 2.2: Optimal contract in (s, π)

must be zero when hired because if effort is strictly positive, then it would be possible for the firm the gain strictly positive profit by delaying all wages by a small amount of time. However, in equilibrium the marginal profit from a worker must be zero when a worker is hired.

In the analysis above we show that $\hat{\pi}(s)$ is monotonically decreasing in s and $\hat{\Delta}(s)$ is monotonically increasing in s. The graph of $\hat{\pi}(s)$ in $s - \pi$ space and $\hat{\Delta}(s)$ in $s - \Delta$ space along with figure 2.1b show the three dimensional saddle path of s, Δ and π on which the contract lies.

Figure 2.2 illustrates the evolution of seniority s and profit π. The thick lines represent the $\dot{s} = 0$ and $\dot{\pi}$ loci respectively. The direction of the dynamics are shown by the small arrows at right angles and the saddle path is as labeled. From the boundary conditions of the Hamiltonian, we know that seniority s is equal to zero when the worker is hired. Thus the contract begins on the saddle path at the point where the saddle path intersects with the π axis. Also shown in the picture is the fact that the contract must begin with $\pi = 0$ due to equilibrium effects. Thus expected marginal profit from the worker begins at zero and is decreasing with the tenure of the worker. The figure again illustrates that the expected value of an experienced worker is always negative to a firm in the steady state.

Figure 2.3 shows the evolution of seniority s and utility differential Δ.[6] Again, the contract

[6]Figure 2.3 depicts the case for when z_Δ is decreasing in Δ (effort is concave in Δ) which corresponds to the case

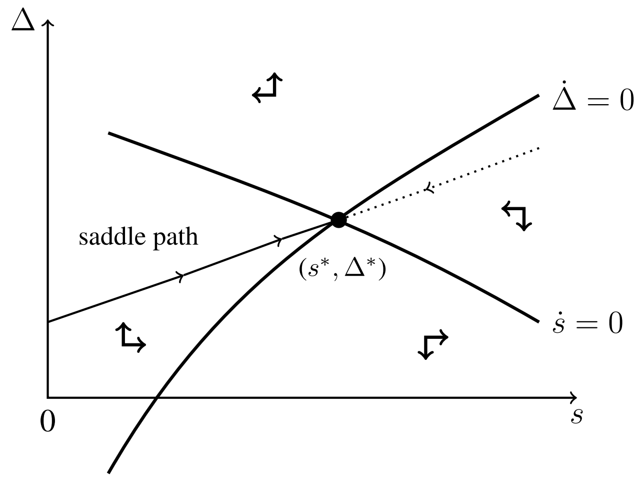


Figure 2.3: Optimal contract in (s, Δ)

begins at the point where the saddle path intersects with the Δ axis. It becomes clear that both seniority and the utility differential Δ increase over time. Since effort $z(\Delta)$ is also a monotonic function of Δ, effort will also increase as well.

The above phase diagrams reveal that there is a stationary point $\dot{s} = \dot{\Delta} = \dot{\pi} = 0$ where the contract comes to rest. Seniority s stops growing in the end because as s increases the term $m\pi$ in the $\mu + m\pi$ of the integrand for s in equation (2.11) becomes increasingly negative. At some point, promising the worker higher wages would make him or her tend to stay on too long from the firm's viewpoint, canceling out the benefit of extra productivity. This limit to how far s can rise is a reflection of diminishing returns to the firm of raising wages. Also preventing seniority s from rising indefinitely is worker impatience. When $\rho > r$, payments far in the future of a given monetary present value of less and less effect on effort now.

We also know that the wage w ends up above the marginal product of the work μz, because as $t \to \infty$ the stationary contract satisfies

$$\pi^* = \frac{\mu z^* - w^*}{r + a^*} < 0.$$

with a positive 3rd derivative of $v(z)$.

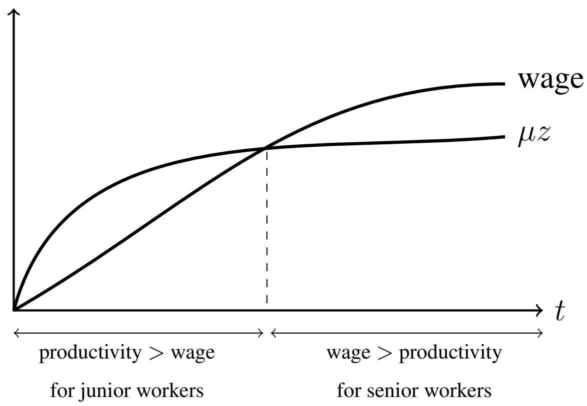


Figure 2.4: Wage vs productivity

Therefore, the marginal product at the stationary point $\mu z^*$is less than the worker's wage w^*. We also know that the wages must be below the worker's marginal product at some point to balance this, or the expected future profits due to taking on a worker at t_0 would be negative, and the firm would never agree to take the worker on those terms. Thus, the marginal product of a worker μz must rise above the wage shortly after the worker arrives on the job. Figure 2.4 shows the approximate shape of the path that wages and productivity will follow over the worker's tenure in steady state. It shows that productivity is greater than wage when the worker is junior and wage is greater than productivity for senior workers.

2.4 Extensions and Applications

The framework presented above is amenable to various extensions and applications some of which we consider in this section. First, we consider the possibility that firms can choose the intensity with which they monitor their workers. We show that firms will likely monitor junior workers more aggressively than senior workers in such a case.

Next, we allow the firm to provide amenities to workers in addition to wages to increase job pleasantness and investigate what the optimal strategy for the firms will be. We find that in general, firms will provide amenities to workers in a manner similar to wages, but some amenities will be provided more or less abundantly depending on its effect on the worker's marginal disutility of effort.

Lastly, we explore a possible application of the basic framework to the question of large earnings losses from job loss. The model provides a new perspective in reconciling the disparity between the empirical results reporting large earnings losses from job loss and theoretical models that mostly predict a smaller impact of job loss.

2.4.1 Variable Monitoring

In this section we consider the firms' choice if firms are able to vary their monitoring intensity over the worker's tenure. The firms specify their level monitoring as a part of the implicit contract. In the model described above, this amounts to firms choosing the path of the variable m, present in the attrition rate $a(m, z) = q + m(1 - z)$, over the life of the contract.

To solve for the optimal path $\{m_t\}$, we can simply solve the Hamiltonian equation for the firm's problem (2.9) such that $\frac{\partial H}{\partial m} = 0$. The resulting first order condition is,

$$\underbrace{[\mu + m\pi]z_m}_{\text{benefit of monitoring}} - \underbrace{[\pi + s\Delta]a_m}_{\text{indirect cost}} - \underbrace{c}_{\text{direct cost}} = 0 \tag{2.13}$$

where c is the marginal cost that a firm incurs every period to monitor each worker with intensity m.[7]

Equation (2.13) can be interpreted as follows. The benefits of monitoring come from the extra effort z_m extracted from the worker due to increased monitoring, with extra current effort having the value of $\mu + m\Delta$ as in equation (2.11) of the definition of seniority. On the cost side, in

[7] In addition to allowing m to be a choice variable, we also add a marginal cost c that a firm must incur to increase their monitoring intensity by an additional unit.

addition to the direct cost of monitoring c, there is a cost of shortening the average length of tenure by increased firings. The overall value of the contract that would be terminated by firing is $\pi + s\Delta$ (expected profits plus the value of keeping the utility differential high). $\pi + s\Delta$ is always positive since the firm maximizes $\pi + s\Delta$ at every point in time and can always attain at least $\pi + s\Delta = 0$ by putting the worker on temporary layoff.[8] Thus, $\pi + s\Delta$ measures the value to the firm of maintaining the contract and $(\pi + s\Delta)a_m$ is the cost of terminating through a more draconian firing policy.

Equation (2.13) predicts that junior workers will most likely be monitored more intensively than senior workers. If monitoring intensity m were unchanged by seniority, the marginal benefits of monitoring measured by $(\mu + m\pi)z_m$ would decline as π became smaller from the worker's approaching his or her upper limit of effort more closely.[9] On the cost side, the need to keep the contract secure $(\pi + s\Delta)$ grows with seniority as demonstrated in Figure 2.1a. Thus the higher attrition rate caused by more intense monitoring is a more serious problem for more senior workers.

2.4.2 Amenities

Firms may also choose to vary the amount of amenities it provides its workers in order to increase job pleasantness and worker utility. It is quite common to observe firms offering varying levels and types of amenities (such as vacation days, office space, etc) to workers depending on the amenity and the worker. In this section, we explore why and how firms provide various amenities to the workers.

[8]If the expected future utility Δ falls to the point that the worker's implied effort level is below zero, such that the firm's marginal profit from the worker coming to work is actually negative, the firm can simply ask the worker to stay home and put them on temporary layoff while still paying the worker in accordance with the longterm contract. The firm can call the worker back in once the implied effort level becomes positive.

[9]$z_{m\Delta} < 0$ if the disutility of effort $v(z)$ is such that $\frac{v''}{v'''} < \frac{v'}{v''}$ and $v''' > 0$.

We can think of amenities as being utility enhancing, or equivalently reducing the disutility of effort. To incorporate this feature into the model above, let amenities A enter the disutility of labor function as a second argument $v(z, A)$ where $v_A < 0, v_{AA} > 0$. Here, amenities is measured in units of cost to the firm of providing another unit of A.

As with variable monitoring, the optimal path of amenities can be solved by solving the Hamiltonian (2.9) such that $\frac{\partial H}{\partial A} = 0$. The optimal path of $\{A_t\}$ solves,

$$\underbrace{[\mu + m\pi]z_A}_{\substack{\text{benefit from} \\ \text{change in effort}}} + \underbrace{s(-v_A)}_{\substack{\text{benefit from} \\ \text{increase in utility}}} - \underbrace{1}_{\text{marginal cost}} = 0 \tag{2.14}$$

This optimality condition for amenities is similar to equation (2.13). $[\mu + m\pi]z_A$ is the benefit the firm reaps from the additional effort of the worker. $s(-v_A)$ is the firm's (indirect) benefit from the increase in worker utility. The marginal cost of providing an addition unit of amenities is normalized to one.

If $v(z, A)$ is additively separable $(v_{zA} = 0)$ then an increase in amenities does not effect the marginal disutility of effort $(z_A = 0)$.[10] Condition (2.14) can then be expressed as,

$$s = \frac{1}{-v_A}$$

which is identical to the optimal condition for wages in equation (2.10) with the marginal utility gain from amenities $-v_A$ replacing the marginal utility gain from wages $u'(w)$. As with wages, amenities A will increase as the worker gains seniority.

If $v(z, A)$ is not additively separable than z_A no longer is equal to zero. Then, equation (2.14)

[10]Rewriting the worker's f.o.c. (2a) with amenities becomes $m\Delta = v_z(z, A)$. Taking the total differential with variables z and A becomes $0 = v_{zz}dz + v_{zA}dA$. Rearranging, we get $\frac{dz}{dA} = -\frac{v_{zA}}{v_{zz}}$.

can be rewritten as,

$$s = \frac{-(\mu + m\pi)z_A}{-v_A} + \frac{1}{-v_A} \tag{2.15}$$

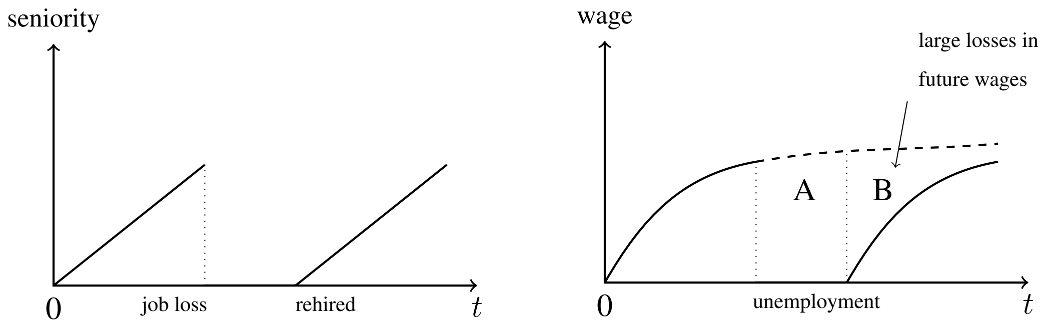
Suppose that amenities decrease the marginal disutility of effort ($v_{zA} < 0$) which implies that effort is increasing in amenities ($z_A > 0$). In this case, the first term in equation (2.15) ($\frac{-(\mu+m\pi)z_A}{-v_A}$) is negative. This implies that more of such amenities will be provided by the firm than the additively separable case all else equal. This is because in addition to the utility benefit, providing amenities makes it easier for workers to exert effort, increasing worker effort.

For amenities that increase the marginal disutility of effort ($v_{zA} > 0$) and thus decrease effort ($z_A < 0$), the term $\frac{-(\mu+m\pi)z_A}{-v_A}$ in equation (2.15) is positive. Such amenities will be much more scarcely provided, as the increase in the marginal disutility of effort will discourage workers from increasing effort.

The predictions of the model seem to be in line with what is often observed in the real world. The model predicts that amenities that do not effect the marginal disutility of effort will increase as workers gain seniority. Examples of such amenities may include vacation days or packages, luxurious office spaces, or tuition subsidies for a worker's children. These types of benefits are often provided incrementally as workers gain seniority and rank.

Amenities that decrease the marginal disutility of effort will be offered in abundance by firms. This is intuitive, as such amenities will allow the workers to increase effort more easily. Items such as air conditioning or ergonomic chairs allow workers to work longer hours with minimal physical fatigue and unpleasantness. Improved organizational techniques such as improved ways of providing workers with constructive feedback allow workers to put in the right kind of effort without increasing disutility. We can easily observe firms eagerly providing such amenities in order to increase worker effort.

On the other hand, amenities that increase the marginal disutility of effort will not be offered



(a) seniority (b) wage

Figure 2.5: Wage loss

much by firms. Amenities that tempt the worker to slack are not popular items in most workplaces. It is difficult to think of many such amenities provided by firms, consistent with the prediction that they are rarely provided, but amenities such as on-site recreational rooms or sleeping pods seem to be appropriate examples. Again, these facilities are rarely provided by most workplaces.[11]

2.4.3 Earnings Loss

Many empirical studies have found that there is a large loss in life time earnings when a worker loses their job. Most notably, they find that job loss is a highly persistent negative shock to future wages and results in a persistent increase in the likelihood of future unemployment.

Most models that attempt address earnings losses are based on search theory. However Davis and von Wachter (2011) find that current search models, for the most part, are inconsistent with these findings. They find that realistically calibrated models predict much smaller earnings losses then found in empirical studies.

The model presented in this paper provides an alternative channel in explaining large earnings

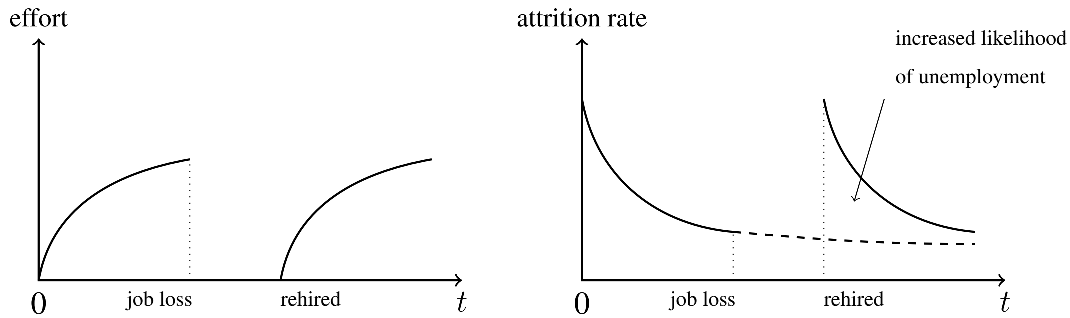
[11] One interesting counterexample is the attention some recent startups and even some larger companies such as Google have received by providing such unorthodox amenities in their work spaces. While the fact that they have garnered so much attention by doing so only emphasizes the scarcity of such policies, how such companies and industry differ from the general structure presented in this paper may be an interesting question in itself.

losses. In this model, large losses in life time earnings arise naturally from the optimality of the seniority wage structure. Wages are a reflection of seniority, and when a worker is fired they lose the seniority that they have accumulated. Since seniority is not transfered from firm to firm when a worker is fired, workers do not regain the lost seniority even when they are rehired.

Figure 2.5 illustrates this fact. Panel (a) shows the seniority that a worker accumulates as he or she works for the firm. The solid lines depict the worker's accumulated seniority. As shown, a worker loses all seniority with job loss. Even when rehired a short period later, the worker must re-accumulate seniority from scratch as would any new hire. Panel (b) shows the cumulative effect on wages. The solid lines show the workers wage and the dashed line show the counter factual wage of the worker had they not been fired. The loss is considerable. In addition to the lost wage during unemployment (the area labeled A), there is a persistent reduction in wage compared to the counter factual case due to the difference in the level of accumulated seniority (the area labeled B). This fact leads to the persistence and size of earnings losses resulting from temporary job loss.

The model can also replicate the fact that the probability of future unemployment is persistently higher after job loss. This is because senior workers have higher effort levels than junior workers in the model. The utility differential between employed and unemployed workers Δ increases as a worker gains seniority as was shown in Figure 2.3. Because effort is a monotonic function of Δ, effort increases with tenure. Thus senior workers are less likely to be fired as they work harder and make fewer mistakes as shown in Figure 2.6.

Furthermore, firms that vary monitoring intensity will choose to monitor junior workers more intensely as described in section 4.1. Thus when workers lose their accumulated seniority with job loss, the effort level is reset as is firm monitoring and this will add to the persistent increase in the probability of future job loss.



(a) effort (b) attrition rate

Figure 2.6: Probability of job loss

2.5 The Steady State Equilibrium

2.5.1 Equilibrium

At any given time a typical firm will employ many workers of varying degrees of seniority. The firm will act to maximize the expected present value of profits while keeping the promises made to workers in the past. Given a path of productivity μ_t a firm will pay workers according to equations (2.10) and (2.11).

As long as the marginal profit of a newly hired worker (π_0) is positive the firm will hire more workers. If the marginal profit is negative, there will be no hiring. Thus for markets to clear in equilibrium, the marginal profit of a newly hired worker must be zero as long as there is positive hiring. The number of workers a firm hires at time t will be pinned down by the market clearing condition $\pi_{0,t} = 0$.

Recall that workers will choose effort according to (2.2) where Δ_t is a function of the path of ψ_t. Then equilibrium of this model will be a fixed point mapping in $\{\mu_t, \psi_t\}$. The implied paths of effort, hiring and firing given the path of $\{\mu_t, \psi_t\}$ must be such that it in turn satisfies the definitions $\mu_t \equiv p_t f'(Z_t)$ and $\psi_t \equiv \bar{u} + \frac{h_t}{N_t - L_t}\Delta_{0,t}$ for all t, where total effective labor Z is defined as $Z_t \equiv \int_{-\infty}^{t} z_t(\tau) l_t(\tau) d\tau$ with $l_t(\tau)$ denoting the number of workers hired at time τ that are employed by the firm at time t and $z_t(\tau)$ is the time t effort level of a worker hired at time τ.

2.5.2 The Steady State

The equilibrium dynamics of this model can be quite complex and is beyond the scope of this paper.[12] Here, we focus on the steady state equilibrium and comparative statics to gain some insight into the general equilibrium properties of this model.

In the steady state, the aggregate variables $(\psi, \mu, r, h, \pi_0, \Delta_0, Z, L)$ must be constant. In addition, the steady state involves a constant even flow of new hirings balancing separations due to firings and exogenous quits. The hiring rate must equal the attrition rate,

$$h_t = \int_{-\infty}^{t} a(m, z_t(\tau))l_t(\tau)d\tau = \bar{a}_t L_t$$

where $\bar{a}_t \equiv \frac{\int a(m, z_t(\tau))l_t(\tau)d\tau}{\int l_t(\tau)d\tau}$ and $L_t \equiv \int_{-\infty}^{t} l_t(\tau)d\tau$.

Note that the steady state equilibrium does not imply that the contract is at the stationary point. The solution to $\dot{\Delta} = \dot{s} = \dot{\pi} = 0$ visible in Figures 2.1, 2.2 and 2.3 does not describe the steady state, but only the eventual condition of employees who by good fortune remain with the firm for a long time. Since new workers are constantly being hired, much of the employed labor force will always be far from this stationary point. The steady state of the system involves constant motion out of unemployment and, once hired, motion toward this stationary point coupled with attrition into unemployment.

In the steady state there must be positive hiring so π_0 must be zero. As discussed above, the two key variables ψ, and μ determine the steady-state path of wages, effort and utility. ψ is the flow benefit of unemployment and, loosely speaking, can be thought of as a measure of inverse wage. μ

[12]See Kimball (1994) for an exposition of the dynamics of a basic efficiency wage model.

on the other hand, can be expressed as a function of total employment L as $\mu = pf'(\bar{z}L)$ where $\bar{z}$ is the average effort of the firm's employees. We can then express the steady state equilibrium on the (ψ, L) plane in terms of the familiar supply and demand relation of the labor market.

First, consider firm demand for labor. Higher ψ increases the attractiveness of being unemployed and lowers the cost of job loss. The firm must either accept lower effort from the worker as a result of this weakened job attachment, or it must raise wages to keep the cost of job loss high. Either way the profits expected from a worker suffer. In particular, the expected present value of hiring a new worker would be negative unless the marginal revenue product of effort μ rises to make up for the reduced effort or increased wages of a worker.

The qualitative effect on the total number of employed workers L from a higher ψ, and consequently higher μ, is ambiguous. At a cursory glance it would seem that the total number of workers L should decrease in order for μ, which is a function of the marginal productivity total effective effort $f'(Z)$, to increase. However, since Z is the product of the average effort of firm's employees $\bar{z}_t = \frac{\int z_t(\tau) l_t(\tau) d\tau}{\int l_t(\tau) d\tau}$ times L, it is possible that a fall in average effort all by itself, might increase μ by enough to allow the firm expand the number of workers in order to make up for less effort per worker.

To see this, consider the following example. Let the interest rate r be zero so that the current cross-section of workers gives the same relative weights to different seniorities as does the integral giving the expected present value of the career of a newly hired worker. Since $\pi_0 = 0$,

$$\bar{w} = \bar{z}\mu = \bar{z}pf'(\bar{z}L) \tag{2.16}$$

where $\bar{w}$ and $\bar{z}$ are present value averages and cross-sectional averages as well since $r = 0$. We write (2.16) in elasticity terms,

$$\%\Delta\bar{w} = \%\Delta\bar{z} + \%\Delta\bar{p} - \gamma[\%\Delta\bar{z} + \%\Delta L] \tag{2.17}$$

where $\gamma = \frac{f''(Z)Z}{f'(Z)}$ measures the degree of diminishing returns to labor input. Solving (2.17) for $\%\Delta L$, we obtain

$$\%\Delta L = \frac{1}{\gamma}[(1-\gamma)\%\Delta\bar{z} + \%\Delta\bar{p} - \%\Delta\bar{w}]. \tag{2.18}$$

Now, suppose we are near a rounded off corner of the production function so that $\gamma = 3$ at that point and that an exogenous increase in ψ leads to a 3% decrease in effort $\bar{z}$ and a 3% increase in the wage. Then (2.18) predicts that the number of employees will increase by 1%.

As can be seen from the foregoing example, while the graph of π_0 in terms of μ against ψ is upward sloping, and the graph of ψ against Z is downward sloping, the graph of $\pi_0 = 0$ in $L - \psi$ space can be either upward or downward sloping. If effort is only moderately responsive to changes in ψ, $\bar{z}$ is relatively constant and L will decrease resulting in a downward sloping curve. However, if effort is very responsive to changes in ψ, $\bar{z}$ will fall sharply and L must increase resulting in a upward sloping curve.

A downward sloping curve corresponds to the usual intuition, since this curve is in effect the firm's demand curve for labor. If the number of workers and the amount of effective labor move in opposite direction as in the example above, the demand curve will be upward sloping. One thing to note is that the demand curve in $\psi - L$ space must be downward sloping for high enough ψ since high enough ψ would make it unprofitable to employ any workers. This could be a continuous function of L but there may be a discontinuous jump from positive to zero L as ψ rises. L can continue to decrease indefinitely as ψ decreases since a firm may be sated with a certain amount of effective labor so that any increases in the effort per worker may only reduce the number of workers needed.

The supply of labor to the firm is horizontal, since ψ is exogenous to the firm. Thus, ψ is analogous to the market wage in a competitive labor market. However, the market supply curve for labor is not horizontal. In the steady state,

$$\psi = \bar{u} + \frac{\bar{a}L}{N - L}\Delta_0$$

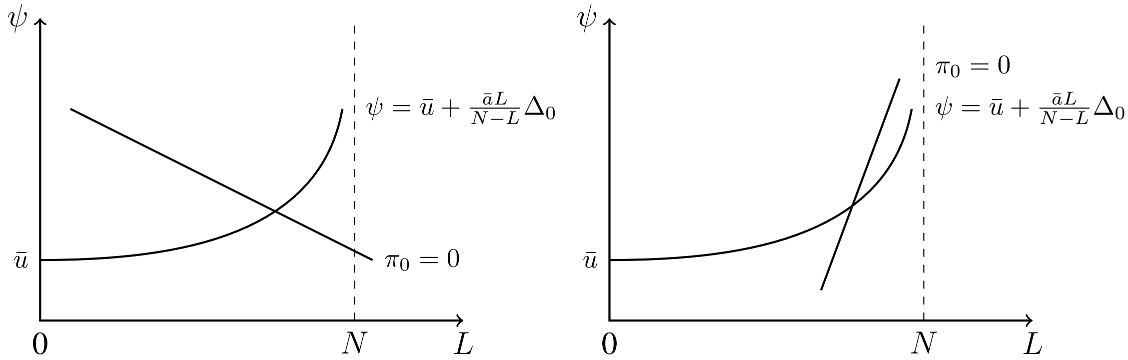
where $\bar{a}$ is the appropriately weighted average attrition rate of labor. This curve is upward sloping in general, the more so if the attrition rate $\bar{a}$ increases with ψ. For the supply curve to be downward sloping would require $\bar{a}$ to be a strong inverse function of ψ.[13] Barring such perversities, the supply curve will be upward sloping. The curve will become very steeply sloped as total labor employed approaches N.

The market demand and supply of labor can now be put together. Note that the supply curve is the analog of Shapiro and Stiglitz' "No Shirking Condition" and the demand curve to their "Marginal Product of Labor" schedule. Figure 2.7 shows the two cases in which the demand is downward and upward sloping respectively.

The comparative statics in the steady state can be analyzed from supply and demand. An increase in N will move the $\psi = \bar{u} + \frac{\bar{a}L}{N-L}\Delta_0$ curve out, while a reduction in $\bar{u}$ will move it down, both changes therefore increasing employment in Panel (a) of Figure 2.7 and decreasing employment in Panel (b), reducing ψ in both cases. An improvement in productive technology that increases the marginal product of labor will move the demand curve out. A reduction in the exogenous quit rate, an improvement in monitoring technology or exogenous worker motivation as reflected in worker utility functions will shift the $\pi_0 = 0$ curve out in $\psi - \mu$ space and thereby cause an increase in demand in $\psi - L$ space. Any increase in demand causes both L and ψ to rise.

In the intuitive case of a downward sloping labor demand curve, the supply curve will be the

[13] In a recent paper, Shimer(2012) finds that employment exit probability is largely acyclical.



(a) (b)

Figure 2.7: Steady state equilibrium

main determinant of the equilibrium unemployment rate due to the steep slope of the supply curve for reasonable levels of unemployment.[14] Therefore, the main effect of shifts in the demand curve will be on equilibrium ψ, with very little effect on equilibrium L. As a result, the effects of changes in the exogenous quit rate, monitoring technology or worker utility functions on unemployment through the supply curve overshadow their effects through the demand curve.

For example, if two distinct worker groups have different exogenous quit rates but the same utility functions, the implied lower work effort for one group will have only a small effect on the unemployment rate. The main factor leading to a higher unemployment rate for this group would be that labor discipline requires about the same average spell of unemployment for both groups, but the flow into unemployment would be greater for the group with the higher attrition rate leading the percentage of persons unemployed at any time to be greater.

Finally, as described above, the shape of the labor demand curve can potentially have important implications for the comparative statics of the labor market. This can lead to interesting policy implications as well. For example, if the effort is very responsive to changes in ψ the slope of the demand curve will become steeper and steeper and may even imply that the demand curve is upward sloping. A steep demand curve, can imply that increased unemployment benefits will have

[14]For example, the slope of the supply curve will equal $a\Delta_0 \frac{N}{(N-L)^2} = 100\bar{a}\Delta_0 N$ for 10% unemployment

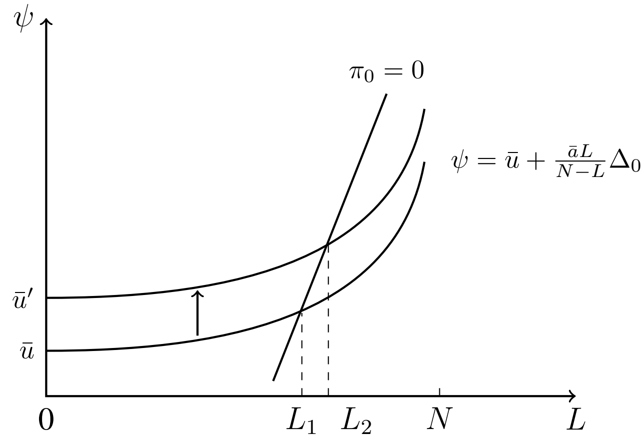


Figure 2.8: Increase in unemployment benefits

very little effect on the unemployment rate and may even lead to a decrease in the unemployment rate if the curve is upward sloping. This is because an increase in unemployment benefits will increase $\bar{u}$ which would shift the supply curve upwards. The equilibrium values of ψ and L will move along the demand curve as depicted in Figure 2.8 for the upward sloping case. Nevertheless, the analysis shows that it is unclear whether this is a desirable outcome as the increase in employment is driven by a large decrease in effort and therefore productivity.

2.6 Conclusion

This paper finds that when firms cannot monitor worker effort perfectly, they will choose to pay wages that reflect worker seniority. Seniority will be accumulated by the worker as they work for the firm, in a manner that reflects the value of the worker to the firm. Firms will monitor junior workers more closely than senior workers and also provide amenities according to seniority and its effect on the marginal disutility of effort. The steady state equilibrium is studied as well.

The findings of this paper not only enhances our understanding of the relationship between firm and worker but should also prove beneficial to empirical studies. The clear and concrete definition of seniority and a well defined relationship between seniority and wages allows for a much richer interpretation of empirical findings. It may also potentially allow for more sophisticated empirical designs. The paper also provides predictions regarding worker monitoring and amenities that can be tested empirically.

While much of this paper focused on the individual worker-firm relationship, a potentially interesting question is the business cycle consequences of the effects described in this paper. A study of the full general equilibrium dynamics of this model could contribute to our understanding of business cycles and especially the cyclical properties of unemployment.

2.7 References

Akerlof, George A., Katz, Lawrence F., 1989. Workers' Trust Funds and the Logic of Wage Profiles. Quarterly Journal of Economics 104(3), 525-536

Altonji, Joseph G., Williams, Nicolas, 2005. Do Wages Rise with Job Seniority? A Reassessment. Industrial and Labor Relations Review 58(3), 370-397

Basu, Susanto, Kimball, Miles S., 1997. Cyclical productivity with unobserved input variation. NBER Working Paper no. 5915.

Basu, Susanto, Kimball Miles S., 2002. Long-Run Labor Supply Elasticity and the Elasticity of Intertemporal Substitution for Consumption. University of Michigan working paper.

Buchinsky, Moshe, Fougere, Denis, Francis, Kramarz, Tchernis, Rusty, 2010, Interfirm Mobility, Wages and the Returns to Seniority and Experience in the United States. Review of Economic Studies 77, 972-1001

Bulow, Jeremy I., Summers, Lawrence H., 1986. A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment. Journal of Labor Economics 4(3), 376-414

Burdett, Ken, Coles, Melvyn, 2010. Tenure and Experience Effects on Wages: A Theory. CESifo Working Paper No.3203

Davis, Steven J., von Wachter, Till, 2011. Recessions and the Costs of Job Loss. Brookings Papers on Economic Activity 43(2), 1-72

Daniel, Kirsten, Heywood, John S., 2007. The Determinants of Hiring Older Workers: UK Evidence. Labour Economics 14, 35-51

Epstein, Brendan, Kimball, Miles S., 2014. The Decline of Drudgery and the Paradox of Hard Work. International Finance Discussion Papers no. 1106. Board of Governors of the Federal Reserve System

Heywood, John, Jirjahn, Uwe, Tsertsvardze Georgi, 2010. Hiring Older Workers and Employing Older Workers: German Evidence. Journal of Population Economics 23(2), 595-615

Kambourov, Gueorgui, Manovskii, Iourii, 2009. Occupational Specificity of Human Capital. International Economic Review 50(1), 63-115

Kimball, Miles S., 1994. Labor Market Dynamics when Unemployment is a Worker Discipline Device, American Economic Review 84(4), 1045-1059

Kwon, Illoong, Milgrom Eva M. Meyersson, 2014. The Significance of Firm and Occupation Specific Human Capital in Hiring and Promotion. Labour Economics 31, 162-173

Lazear, Edward P., 1979. Why is there Mandatory Retirement?. Journal of Political Economy 87(6), 1261-1284

Lazear, Edward P., 1981. Agency, Earnings Profiles, Productivity, and Hours Restrictions. American Economic Review 71(4), 606-620

Lazear, Edward P., 2009. Firm-Specific Human Capital: A Skill-Weights Approach. Journal of Political Economy 117(5), 914-940

Neal, Derek, 1995. Industry-Specific Human Capital: Evidence from Displaced Workers. Journal of Labor Economics 13(4), 653-677

Pavan, Ronni, 2011. Career Choice and Wage Growth. Journal of Labor Economics 29(3), 549-587

Shapiro, Carl, Stiglitz, Joseph E., 1984. Equilibrium Unemployment as a Worker Discipline Device. American Economic Review 74(3), 433-444

Stevens, Margaret, 2003. Earnings Functions, Specific Human Capital, and Job Matching: Tenure Bias is Negative. Journal of Labor Economics 21(4), 783-805

Topel, Robert, 1991. Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority. Journal of Political Economy 99(1), 145-175

Zwick, Thomas, 2011. Seniority Wages and Establishment Characteristics. Labour Economics 18, 853-861

CHAPTER 3

System Wide Runs and Financial Collapse

3.1 Introduction

Losses from the subprime mortgage sector that triggered the recent financial crisis was small in comparison to the catastrophic events that were to follow. According to former Fed chairman Ben Bernanke (2013),

> The problem was not just the losses. If you put together all the subprime mortgages in the United States and assumed they were all worthless, the total losses to the financial system would be about equivalent to one bad day in the stock market: they were not very big. (p.71)

How then, can one explain the dire consequences that followed the subprime shock? Gorton and Metrick (2012) characterize the financial crisis as a system wide run and argue that problems in the subprime mortgage sector triggered a run on repos resulting in the panic of 2007-2008. While this characterization of the crisis is compelling there is some disagreement over the interpretation of the events. Krishnamurthy, Nagel and Orlov (2014) also find a sharp contraction in funding to the financial sector through tri-party repos and asset-backed commercial paper. However, they argue that the phenomenon more closely resembles a credit crunch rather than traditional bank runs. The debate is not in whether there was a dramatic contraction of credit flowing to the financial sector but whether that contraction can be described as a run. In this light, it is unclear whether what Gorton and Metrick mean by a system wide run entails more then simply a collection of well timed runs on individual institutions. The purpose of this paper is to study what a system wide run is and whether it can be triggered by small shocks to fundamentals.

The paper presents a stylized model of the financial system where small shocks to fundamental asset values can cause large contractions in credit due to informational frictions. Asset fire sales can amplify these effects and lead to the complete collapse of the financial system. The large aggregate contraction of liquidity provided to the financial sector resembles what Gorton and Metrick describe as a system wide run.

A system wide run in this model is fundamentally different from a collection of individual bank runs. System wide runs stem from the incentives of high quality borrowers to distinguish themselves from low quality borrowers whereas traditional bank runs arise from coordination failures of the lenders.

Bank runs such as in Diamond and Dybvig (1983) or He and Xiong (2012) rely on an implicit lack of arbitrageurs that can intervene to mediate the run. In theory, arbitrageurs could always step in to the benefit of all relevant parties. Coordination problems could be rendered irrelevant. For individual institutions one may think of many reasons why such arbitrageurs are unavailable such as geographic segmentation or informational frictions. However, when dealing with system wide phenomena it is more difficult to imagine why and the issues of arbitrage need to be considered more carefully.

The key insight of this paper is that system wide runs can be caused by the by the incentives of financial institutions to secure better terms on their debt. Institutions with different probabilities of default have divergent preference over the combination of the amount of total debt they issue and the price of each unit of debt. Those with lower probabilities of default have a higher affinity for receiving a better price for each unit of their debt where as institutions with a high probability of default would prefer to obtain as much funding as they can.

When there is asymmetric information between borrowers (financial institutions) and lenders regarding the financial health of the borrowers, the healthier institutions can credibly signal their type through the terms they demand for their loans. They can promise to decrease the total amount of borrowing in exchange for more favorable prices. The distressed institutions will be reluctant to follow. However, to decrease the total amount of borrowing financial institutions are forced to liquidate their assets. If the amount of assets that need to be sold is too large, it may result in asset fire sales.

Asset fire sales threaten the survival of the financial sector.[1] If the fire sale discount is too large, even otherwise healthy institutions may turn insolvent. Since lenders are unwilling to lend to insolvent institutions this can result in a complete credit market collapse and consequent failure of the financial system. This is what I describe as a system wide run.

The main friction in the model is the asymmetric information between finacial institutions and their creditors regarding the institutions' financial state. This friction captures a commonly held sentiment shared by many observers during the crisis. It was not the losses per se but the uncertainty regarding who was suffering the losses that was most damaging. The model highlights the precise economic mechanism as to why this may have been so damaging. It was the desire of the healthy borrowers to differentiate themselves that paradoxically drove the crisis.

In addition to the literature on bank runs, this paper bears some resemblance to the work of Kiyotaki and Moore (1997) in that small shocks to fundamentals create credit contractions that in turn worsen assets prices which further contracts credit. However, the credit constraint in Kiyotaki and Moore is based on the inalienability of human capital and the renegotiation of debt. For

[1]For other papers that emphasize the importance of fire sales during the crisis see Shleifer and Vishny (2011), or Stein (2013) for their importance in financial regulation.

industrial firms in the real sector (which is their intended setting) this is a grave concern. For financial institutions whose assets are mainly securities, the assumption is not well justified and their mechanism does not accurately represent the phenomenon of system wide runs in the financial sector.

Understanding the source of credit contraction is policy relevant. Because the credit contraction is initially driven by the costly signaling (by decreasing the total amount of debt) that allows financial institutions to differentiate themselves, a policy that completely shuts down the possibility of signaling is effective. The paper finds that mandating participation in a government clearing house for loans garners results very similar to that of a perfect information benchmark. Also the paper finds that the effect of a widely debated policy of counter-cyclical borrowing limits is ambiguous and may not be effective in this setting.

The remainder of the paper is structured as follows. Section 2 describes the model and the main findings. Section 3 explores potential policy implications and section 4 offers a discussion on the various elements of the model. Section 5 provides a review of how the paper fits into the existing literature and section 6 concludes.

3.2 Model

The ensuing model describes a financial system where a small shock to the fundamental value of the asset can result in a collapse of the financial system. When there is a shock to the asset values, liquidity contractions occur endogenously and generate fire sales that threaten the solvency of all financial institutions. System wide runs may follow.

The financial system consists of three types of agents, broker-dealers, households and banks. They are all risk-neutral. There are three periods in the model, periods zero, one, and two. There is no discounting.

Broker-Dealers

Broker-dealers represent highly sophisticated financial institutions with the skills and expertise to exploit profitable, albeit complicated, investment opportunities. For example, they can be dealer banks investing in asset-backed securities and derivatives or hedge funds executing complex trading strategies.[2]

There is a continuum of long-lived broker-dealers of measure one who consume their entire wealth in period two. The broker-dealers have no initial wealth and must borrow from the households to finance their investment.

In period zero, each broker-dealer is given the opportunity to purchase $a^H > 1$ units of risky assets for one unit of cash. Think of these assets as complex securities or portfolios that only broker-dealers can invest in. The expected payoff of one unit of the risky asset is arbitrarily close to one (the meaning of this will become clear below) and thus the expected return from the risky asset is a^H. Investing in risky assets is very profitable and this reflects the proprietary skills of the broker-dealers that can generate excess profits. The broker-dealers will always choose to purchase the full amount (a^H units) of risky assets. This is optimal for them as long as the probability of a crisis is small (which will be assumed).

The return of the risky asset has the following structure. In period one, the state of the world is revealed. The economy can either be in a "normal" state with probability $1 - \delta$ or in a "crisis" state with probability δ. In a normal state of the world, each unit of risky asset is revealed to give a deterministic payoff of one in period two. In a crisis state the risky assets give a stochastic return of R with distribution F on $[0, \infty)$ with an expected return of one. In addition, in a crisis state a small random subset α of broker-dealers are hit with a shock that wipe out a fraction $1 - \phi$ of their

[2]See Duffie (2010) for a detailed description of dealer banks and the difficulties they faced during the financial crisis.

assets.

Thus in a crisis state, a fraction α of the broker-dealers (to be labeled the low type) hold $a^L \equiv \phi a^H$ units of assets and the remaining fraction $1 - \alpha$ of broker-dealers (to be labeled the high type) hold a^H units of risky assets, all of which have a stochastic return of R on distribution F. Furthermore, I assume that $\alpha > 0$ converges to zero. This not only simplifies the algebra but also highlights the amplification process and the fact that the magnitude of the aggregate shock can be arbitrarily small. A larger α will only reinforce the results that follow.

In period one, the broker-dealers can buy and sell the assets on a secondary market. However, the broker-dealers face an exogenous borrowing limit of d^{limit}. This borrowing limit is fixed and does not depend on the price or any other characteristic of the asset. The borrowing limit can generally be assumed to be greater than their period zero debt level and thus is not in itself a source of credit contraction.

One interpretation of the borrowing limit in this model can be as a form of regulation. For example, the Basel II accords regulate the amount of risky investments financial intermediaries could make given their quantity of capital. The Basel III accords introduce even more stringent constraints. Another potential explanation for the existence of the borrowing limit could be due to moral hazard concerns. Experts may have the incentive to steal the funds and disappear once the amount of debt becomes too large. Because the borrowing limit is fixed and essentially a free variable in the model, the source of the borrowing limit is not critical to the main arguments of the paper.

Households

The households are the source of all financing for the broker-dealers and banks in this model. The household can be thought of as any entity that provide funding to the financial system. They can literally be thought of as households or any intermediate institution (for example money market funds) that provides financing to the financial system via funds from the households.

A continuum of short-lived households of measure one are born in period zero with an endowment of ω units of cash. They consume their entire wealth one period later in period one and exit the system. Similarly another cohort of short-lived households of measure one enter in period one and consume in period two.

The households can either store cash at no discount or purchase short-term debt issued by broker-dealers or banks. They cannot invest in risky assets because they do not have the necessary expertise. Furthermore, they are only willing to lend through debt contracts as they do not have the capabilities to verify the returns of the risky assets and thus the cash flow of their counter party.[3]

The amount of resources that the households are endowed is large in aggregate. Specifically, ω is sufficiently large enough to fund all investment opportunities that arise in the model if the households so choose. Thus the households are competitive and are willing to lend as long as their expected return is greater than zero.

The structure of the households in the model imply that all lending is restricted to be through one period debt. To begin with, it is well known that financial institutions seem to be funded through disproportionately large amounts of short-term debt.[4] Diamond (1991), for example, focuses on the debt maturity structure of financial institutions and their heavy reliance on short-term debt.

In addition, recent work by Brunnermeier and Oehmke (2013) support the restriction to short-term debt in this model. They show that because of the possibility of debt dilution, creditors will

[3]While this model does not explicitly feature a cost to state verification, it is consistent with the costly state verification framework of Townsend(1979).

[4]Recent empirical evidence shows that the reliance on short-term funding is exacerbated during a crisis. For example, Krishnamurthy, Nagel and Orlov (2014) find that the maturities in tri-party repo markets decrease dramatically as the crisis unfolds.

rely entirely on short-term financing if the interim information that is released is mostly about the probability of default rather than the expected recovery in default. In this paper, all the information that is released has to do with the probability of default and does not change the recovery structure post-default.

Furthermore, the households in this model is purposely structured such that one can ex-ante rule out coordination failure among households as the source credit contractions. By ruling out coordination failures among lenders, it becomes clear that the mechanism driving system wide runs is fundamentally different from those of traditional bank runs.

Banks

Banks represent traditional commercial banks that borrow from the household sector and lend to firms. In the context of this paper, banks are important because they can participate in the secondary market for risky assets in period one. In a crisis state, banks are the buyers of risky assets that the broker-dealers seek to liquidate. The liquidity they provide in the asset market determines the price of the assets. This feature of the model is motivated by the observation by He, Khang, and Krishnamurthy (2010) and others, who find that securitized assets such as asset-backed securities liquidated by broker-dealers and hedge funds ended up on the balance sheets of commercial banks during the crisis.

There is a continuum of long-lived banks of measure one who consume in period two. The banks have no initial endowments and must borrow from the households to invest. The borrowing capacity of the banks is capped at B. In period zero, banks can invest in a continuum of safe projects that give a deterministic return in period two. The marginal return of the projects are captured by a decreasing function $I(\beta)$ where β is the amount invested in the projects. While the projects are safe, they are illiquid and cannot be sold for cash before they mature in period two. The banks do not have the opportunity to invest in risky financial assets in period zero. However, in period one they have the ability to participate in the secondary market for risky assets.

Now, consider the funding decision between banks and households. In period zero, the banks can borrow from the households at zero cost because the banks' projects have deterministic returns and the households who lend to them are competitive.

In period one, a bank's cost of borrowing may depend on the state of the world. In a normal state, the borrowing cost is zero because the return of a risky asset is deterministic going forward. However, in a crisis state, the cost of funding may no longer be zero as the banks can choose to invest in risky assets with stochastic returns. Let q_B denote the price of one unit of bank debt in period one in a crisis state.

The price of the risky asset in secondary markets is determined by supply and demand. In a normal state, the price of the risky asset must equal its expected return of one for markets to clear. In a crisis state, the price will depend on the amount of liquidity available from the banks and the amount of assets that need to be sold by the broker-dealers. Let $\rho \leq 1$ denote the price of the risky asset in a crisis state.

Because the safe projects of the banks are illiquid, banks must consider all potential future profits when choosing the amount β to invest in safe projects in period zero. The marginal return from the safe project should equal the expected return from purchasing risky assets in period one. The banks choose β such that

$$I(\beta) = \delta(\frac{q_B}{\rho} - 1) \tag{3.1}$$

The left-hand-side of equation (1) describes the marginal return from investing in a safe project. The right-hand-side describes the expected return from purchasing risky assets in period one. With probability δ the economy will be in a crisis state, and in a crisis state for each unit of debt the bank is able to issue its expected profit is $\frac{q_B}{\rho}$ minus the amount it must pay back, one. In the normal state the expected profit from investing in risky assets is zero.

For the sake of simplicity, I assume that the total return from a bank's safe project is enough to make bank debt risk-free even in a crisis state. Given β, if $\int_0^\beta I(\tilde{\beta})d\tilde{\beta} \geq B$, the bank's future return from safe projects are enough to guarantee full repayment of debt regardless of the return from risky assets. Bank debt is safe and thus banks should always be able to borrow at zero cost. In other words, I assume that $I(\beta)$ is such that in equilibrium the banks are able to borrow at the risk-free rate. This assumption does not change the qualitative implications of the results.[5]

3.2.1 The Credit Market

In this section I describe the interaction between the broker-dealers and the households. The interaction is straight forward in period zero and in a normal state period one in in period one. However, in a crisis state, the funding decision between households and broker-dealers is complicated by the fact that there are two types of broker-dealers, high types with a^H units of risky assets and low types with ϕa^H units, that may either separate or pool in equilibrium.

To borrow from the households, the broker-dealers make a take-it-or-leave-it offer (d_t, q_t) to the households specifying the total quantity of debt they plan to issue and the price per unit of that debt. That is, if accepted, a broker-dealer obtains cash $q_t \cdot d_t$ at date t and owes d_t at date $t+1$. The households can either accept and fund the broker-dealer or reject the offer and refuse to provide any funds. This feature of the credit market is designed to capture the fact that the lending market is competitive and the fact that there is no aggregate shortage of loanable funds in the model. It is also consistent with standard economic environments where agents can only earn excess profits if they have proprietary technology or skills.

In period zero, the broker-dealers need one unit of cash to invest in risky assets. They offer a price and quantity combination (d_0, q_0) such that $q_0 \cdot d_0$ equals one while allowing the households

[5] Note that this is equivalent to assuming that the banks begin with enough net worth to absorb the potential losses from investing in risky assets.

to break even. In period one, the broker-dealers must pay off their maturing debt d_0 and issue new debt to a fresh cohort of households. If the broker-dealers cannot pay off their maturing debt, they immediately default and their entire portfolio is liquidated. The proceeds from the liquidation are distributed to the households in proportion to the original debt.

In a normal state in period one, the broker-dealers are able to borrow at the risk-free price of one as the return on the risky asset is henceforth deterministic. However, if the economy in a crisis state, the equilibrium is more involved.

Suppose that the economy is in a crisis state and the price of the risky asset in the secondary market is $\rho \leq 1$. Then, a broker-dealer with a units of risky assets and d_0 units of maturing short-term debt has the expected payoff,

$$V(d_1, q_1) = \int_{\hat{R}}^{\infty} [(a - \frac{d_0 - q_1 d_1}{\rho})R - d_1] dF(R) \tag{3.2}$$

where $\hat{R} = \frac{\rho d_1}{q_1 d_1 + \rho a - d_0}$ denotes the minimum required return for the broker-dealer to be able to service their debt in full.

In the event of default, households claim the residual value of the broker-dealers. The expected return from buying one unit of debt from a broker-dealer who issues d_1 units of new short-term debt at price q_1 can be expressed as

$$\pi(d_1, q_1) = (1 - F(\hat{R}))(1 - q_1) + \int_0^{\hat{R}} [\frac{1}{d_1}(a - \frac{d_0 - q_1 d_1}{\rho})R - q_1] dF(R) \tag{3.3}$$

If equation (3) is greater than zero households accept the broker-dealer's offer and provide them cash. If its less than zero households reject the offer. If equation (3) is exactly zero households are indifferent between accepting and rejecting. In this case I assume that households always accept the broker-dealer's offer.[6]

In a crisis state, there is asymmetric information in the credit market because households cannot distinguish between a high type broker-dealer with a^H units of risky assets and low type broker-dealers with only ϕa^H units. However, the terms of the broker-dealer's offer can serve as a signal to their type. Thus this period one credit market between broker-dealers and households can be described as a signaling game where the debt contract (d, q) serves as a signal to the households.

The equilibrium in the period one market for credit between experts and lenders in a crisis state, is a perfect Bayesian equilibrium that consists of the broker-dealer's offer, the household's action of either accepting or rejecting the offer, and the household's belief of the broker-dealer's type given the offer. In addition, all equilibria must satisfy the Cho-Kreps intuitive criterion. Imposing this condition greatly reduces the set of potential equilibria and allows for interesting and insightful results.

The intuitive criterion is the standard and most commonly used refinement of perfect Bayesian equilibria and imposes some structure on the off-equilibrium beliefs of the players. It says that when an agent deviates from the equilibrium play, and the deviation is such that for a certain type of agents it implies a lower payoff than the equilibrium payoff under any belief of the opposing player, it is unreasonable for the opposing player to believe that the deviator is of that type.

The equilibrium can then be expressed as $\{(d_1^H, q_1^H), (d_1^L, q_1^L), (x, \mu)\}$ where (d_1^i, q_1^i) is the offer by the broker-dealer of type $i \in \{H, L\}$ and (x, μ) denotes the lender's action and belief respectively. The remainder of this section solves for this equilibrium.

[6]This assumption is only meaningful in the case of hybrid equilibria. In the pure-strategy equilibria of this model, households always accept when they are indifferent. However, for pure strategy equilibria this is a result, not an assumption.

The Intuitive Criterion

The equilibrium of the credit market can be either a pooling equilibrium, a separating equilibrium or a hybrid equilibrium. Remarkably, all pooling and hybrid equilibria in which the broker-dealers can borrow can be ruled out by the intuitive criterion.

First, consider pooling equilibria. In any pooling equilibrium the high type and the low type broker-dealers should both offer the same debt contract. An important result of this paper is that all pooling equilibria fail the intuitive criterion. There is always a deviation from the pooling equilibrium such that the high type receives a higher payoff then he would from his equilibrium action but the low type receives a lower payoff then he does from his equilibrium action if the lender believes that the deviator is the high type. In turn, by the intuitive criterion the lender must believe that the deviator is the high type and accept the offer.

Proposition 1. *All pooling equilibria in which the broker-dealers can borrow fail the intuitive criterion.*

A proof is given in the appendix. To see proposition 1 consider the following lemma. Recall that equation (2) describes the broker-dealers expected payoff.

Lemma 1 (Single-crossing property)**.** *Given d_0 and ρ, the slope of the indifference curve of the a broker-dealer $\frac{dq_1}{dd_1}$, is increasing in the amount of risky assets owned by the broker-dealer.*

Lemma 1 implies a single-crossing property of the indifference curves of the broker-dealers. Intuitively, the property holds because the amount of risky assets held by the broker-dealers determines their probability of default in period two. Because the low type eventually defaults for a larger range of returns, the benefit they receive from a better price for their debt is smaller than that of the high type. The difference in the benefit from borrowing another unit of debt is mitigated by the fact that it must be paid back. Therefore to decrease a unit of funding, the low type broker-dealers must be compensated with a greater increase in the price of debt than the high types.

Given lemma 1, there is always a deviation where the total amount of debt d_1 is smaller than the pooling equilibrium offer, such that if the household accepts the offer, the high type broker-dealer benefits from the deviation but the low type broker-dealer does not. Furthermore, at least one such deviation exists for all quantities of debt less than the equilibrium offer. Then at such a deviation point that is very close to the equilibrium offer, it is beneficial for the household to accept the deviation if they believe the deviator is the high type.

This implies that for any pooling equilibrium there is always a deviation that the household accepts, and the high type broker-dealer benefits but the low type does not. Thus all pooling equilibria fail the Cho-Kreps intuitive criterion. Similar logic can be used to show that all hybrid equilibria fail the intuitive criterion as well.

Proposition 2. *All hybrid equilibria in which the broker-dealers can borrow fail the intuitive criterion.*

A proof is in the appendix.

The Household's Problem

Propositions 1 and 2 imply that any equilibrium of this model in which the broker-dealers can borrow must be separating. A broker-dealer's type will be identifiable to the households in equilibrium. Thus consider the lender's problem given they are aware of the borrower's type. The interesting case occurs when the low type broker-dealers are fundamentally insolvent ($\phi a^H < d_0$) but the high types could repay their maturing debt without borrowing if the price of the risky asset was equal to its expected value ($a^H > d_0$). This is the case if $\frac{1}{1-\delta} < a^H < \frac{1}{\phi}$ which I assume.

Again suppose that the price of the risky asset in the secondary market is $\rho \leq 1$. First, note that the household's expected payoff from equation (3) is concave in q_1 for a given d_1. This is because there is a tradeoff. Increasing q_1 provides broker-dealers more money to invest in profitable risky assets decreases their probability of default in period two. On the other hand increasing q_1 decreases the gains by the household from each unit of conditional on repayment.

Due to the concavity of π in q_1 there is a price $q^{max}(d_1)$ that maximizes the household's expected profit. From the first-order condition of equation (3), $q^{max}(d_1)$ solves

$$\rho = \int_0^{R^*} RdF(R) \tag{3.4}$$

where $R^* \equiv \frac{\rho d_1}{q^{max}(d_1)d_1 + \rho a - d_0}$. If the liquidation value of the broker-dealer is positive ($\rho a^H - d_0 > 0$), the expected profit of the household will be positive at $q^{max}(d_1)$ for any d_1. However, if the liquidation value is negative, this is no longer true.

Suppose that the liquidation value of the broker-dealer is negative. Then, the expected profit of the household will be positive at $q^{max}(d_1)$ only if d_1 is greater than

$$d^i_{floor} \equiv \frac{d_0 - \rho a^i}{1 - F(R^*)} \tag{3.5}$$

Equation (5) can be derived by plugging $q^{max}(d_1)$ in to equation (3) and solving for the minimum d_1 that makes π positive. This means that any offer where the total quantity of debt is less then d^i_{floor} will not be accepted. At d^i_{floor} the only price of debt that can be accepted is $q^{max}(d^i_{floor})$.

Now, consider the maximum price per unit debt that the household will be willing to accept given d_1. Denote this price $q^i(d_1)$. Again $q^i(d_1)$ only exists if the liquidation value of the broker-dealer is positive or $d_1 > d^i_{floor}$. At the maximum price the household's expected profit should be zero. From equation (3) the household's zero profit condition is

$$(1 - F(\hat{R})) + \frac{1}{\hat{R}} \int_0^{\hat{R}} R dF(R) = q_1 \tag{3.6}$$

where $\hat{R} = \frac{\rho d_1}{q_1 d_1 + \rho a^i - d_0}$. Because π is concave in q_1, there can be multiple solutions to equation (6) for a given d_1. However, $q^i(d_1)$ is the solution to equation (6) that is greater than $q^{max}(d_1)$ and this is unique.

Lastly, there is a price of debt below which the broker-dealer will default even if the households agree to accept the offer. Let $q^i_d(d_1)$ denote the price of debt below which the broker-dealer defaults. The expression of this price can be solved to $q^i_d(d_1) = \frac{d_0 - \rho a^i}{d_1}$. At this price for their debt the broker-dealer cannot repay its maturing liabilities even if it sells its entire portfolio of risky assets at price ρ.

The Broker-Dealer's Problem

Given the range of quantities and prices of debt the households are willing to accept one can now study the broker-dealer's decision problem. To focus on the spillover of distress across institutions, I further assume that not only are the low type broker-dealers fundamentally insolvent, I assume that they would still be insolvent even if they could borrow to the full extent of the borrowing limit. Assuming $d^{limit} < (1 + \frac{1}{a^H - 1})(1 - \phi)$ satisfies this condition. This is a slightly stronger assumption than fundamental insolvency when there are fire sales of risky assets, because a broker-dealer can improve its solvency by purchasing risky assets at discounted prices and making large profits. Without a borrowing limit all broker-dealers, regardless of their current financial condition, could in theory bail themselves out by purchasing infinite amounts of risky assets at fire sale discounts.

Now, consider the broker-dealer's problem. It is not obvious that the high type broker-dealer can borrow in equilibrium when there are fire sales and asymmetric information. The only sustainable equilibrium may be a pooling equilibrium where no broker-dealer can borrow. Nevertheless, I suppose for now that the high type broker-dealer can borrow in equilibrium and solve for what the equilibrium offer must be in such a case.

In any separating equilibrium the low type broker-dealers will never be able to borrow and will always default. This implies that the low types have the incentive to pool with the high types as long as the terms of the high type's offer allows them to pay off their maturing debt. Thus any separating equilibrium offer of the high type cannot have price of debt greater than $q_d^L(d_1)$.

In addition, for any equilibrium quantity d_1^* the high type's offer will have the price $q^H(d_1^*)$. To see why, suppose that in equilibrium the high type offers q_1 that is strictly less than $q^H(d_1^*)$. Then from proposition 1 we know that the high type can deviate to a slightly smaller d_1 with a higher price for which he is strictly better off and the low type will not mimic. Because q_1 is strictly smaller than $q^H(d_1^*)$ there is always such a deviation that the household will accept. The equilibrium unravels. This deviation is only possible when the price is less than $q^H(d_1^*)$. When the price is equal to $q^H(d_1^*)$ the household may not be willing to fund such a deviation. Thus the high type will offer the maximum acceptable price $q^H(d_1)$ in equilibrium.

Thus, the equilibrium offer of the high type $(d_1, q^H(d_1))$ will be the solution to the broker-dealer's constrained optimization problem,

$$\max_{d_1} V^H(d_1, q^H(d_1)) = \int_{\hat{R}}^{\infty} [(a^H - \frac{d_0 - q^H(d_1)d_1}{\rho})R - d_1]dF(R) \tag{3.7}$$

$$s.t. \qquad q^H(d_1) \leq q_d^L(d_1)$$

The solution to this optimization problem is described by the following proposition.

Proposition 3. *In any separating equilibrium in which the high type broker-dealers can borrow, their equilibrium offer is* $(d^*, q^H(d^*))$ *where* $q^H(d^*) = q_d^L(d^*)$.

Proposition 3 states that the high type broker-dealers will choose the maximum quantity of debt that the low type broker-dealer will not have the incentive to mimic when the price of the debt is $q^H(d_1)$. In other words, $V^H(d_1, q^H(d_1))$ is weakly increasing in d_1.

First, note that $q^H(d_1)$ is a decreasing function of d_1 when the liquidation value of the broker-dealer is positive and an increasing function of d_1 if the liquidation value is negative. Since $\pi(d_1, q^H(d_1)) = 0$ by the implicit function theorem,

$$\frac{dq^H(d_1)}{dd_1} = \frac{\rho a^H - d_0}{d_1^2} \frac{\int_0^{\tilde{R}} RdF(R)}{\int_0^{\tilde{R}} RdF(R) - \rho} \tag{3.8}$$

where $\tilde{R} = \frac{\rho d_1}{q^H(d_1)d_1 + \rho a^H - d_0}$. Because $q^H(d_1)$ is greater than $q^{max}(d_1)$ this means that $\tilde{R}$ is smaller than R^* and by extension $\int_0^{\tilde{R}} RdF(R) < \int_0^{R^*} RdF(R) = \rho$. Therefore, equation (8) is negative if $\rho a^H - d_0 > 0$ and positive otherwise.

Thus, if the liquidation value of the broker-dealer is negative, V^H is increasing in d_1 because $q^H(d_1)$ is increasing in d_1. When the liquidation value of the broker-dealer is positive the same can no longer be said. Instead consider the following lemma.

Lemma 2. $V^*(d_1)$ *is increasing in* d_1 *for broker-dealers with positive liquidation value, where* $V^*(d_1) \equiv V(d_1, q^H(d_1))$.

It can be shown that $V^*(d_1)$ is increasing in d_1 by showing that $\frac{dV^*}{dd_1}(d_1) > 0$ as d_1 converges to infinity and that V^* is concave. Lemma 2 implies that high type broker-dealers prefer to maximize their borrowing quantity subject to the constraints when they are separately identified.

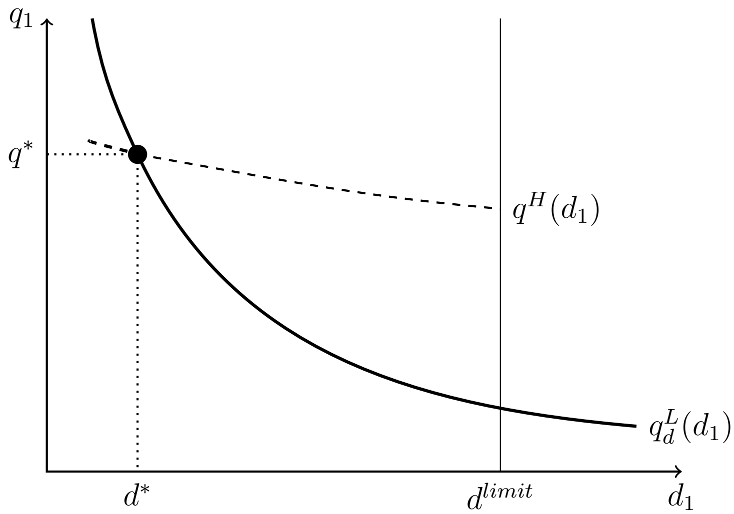


Figure 3.1: Separating equilibrium in the credit market

Thus if a separating equilibrium exists in which the high type broker-dealer can borrow, their offer will be at the point where $q^H(d^*) = q_d^L(d^*)$. Figure 1 illustrates this point. The solid point indicates the high type broker-dealer's offer in a separating equilibrium. The low type broker-dealer always defaults and thus their equilibrium point is not depicted.

3.2.2 Equilibrium

With an understanding of the credit market outcomes one can now solve for the equilibrium of the full model. The equilibrium can be described as the equilibrium in the market for risky assets and market for credit in periods zero and one. The market for risky assets is a competitive market where prices are determined to clear the market. The credit market equilibrium in period one is a perfect Bayesian equilibrium that satisfies the Cho-Kreps intuitive criterion.

Equilibrium: *The equilibrium consists of* $\{\beta, (d_0, q_0), \rho, (s^H, s^L), (d_1^H, q_1^H), (d_1^L, q_1^L), (x, \mu)\}$*; the amount invested in safe projects by banks in period zero, the broker-dealer's debt contract in period zero, the price of the risky asset in period one, the number of units of risky assets that need to sold by each type of broker-dealer, the debt contracts of high and low type broker-dealers, and* *the household's action and beliefs in the period one crisis state.*

Note that the broker-dealer and household's actions and the price of the risky asset in a period one normal state is omitted. In this state, the price of the asset is equal to its expected return of one and the price per unit of the broker-dealer's debt must also be one. The quantity of debt each broker-dealer borrows is indeterminate.

Depending on the parameters of the model, two distinct equilibrium outcomes can prevail. In one outcome, the financial suffers a large credit contraction in the crisis state but only the low type broker-dealers default in period one. The system as whole is able to withstand the crisis. In another outcome, the financial sector suffers a system wide run and all broker-dealers of both type fail, resulting in a complete collapse.

Credit Contraction

Whether or not the financial system can survive a crisis depends on the severity of the fire sales in the secondary market for risky assets. The price of the risky asset in period one is determined by the supply and demand of assets in a competitive market.

First, the demand for assets depends on the availability of liquidity by the banks. As long as the price of the assetρ is less than one, the banks will devote their entire remaining borrowing capacity to purchasing the risky assets. Let s^L and s^H denote the number of assets sold by each type of broker-dealer respectively. Then the market clearing condition for risky assets is

$$B - \beta = \rho(\alpha s^L + (1 - \alpha)s^H) \tag{3.9}$$

$B - \beta$ is the remaining borrowing capacity of the banks after investing in period zero safe projects and $\alpha s^L + (1 - \alpha)s^H$ is the total number of units of assets supplied. Price ρ must clear the market. Then combining bank's optimal investment decision described by equation (1) with the market clearing condition (9), the equilibrium price of the asset will be then determined as the solution to the following equation.

$$\delta(\frac{1}{\rho} - 1) = I(B - \rho(\alpha s^L + (1 - \alpha)s^H)) \tag{3.10}$$

A unique solution to this equation exists as long as $I(B - (\alpha s^L + (1 - \alpha)s^H))$ is greater than zero.

Now consider the supply of assets in the period one secondary market. The supply of assets will be determined by the amount of assets the broker-dealers need to sell in order for them to repay their maturing debt.

Suppose that in equilibrium the high type broker-dealers can borrow from the households. Then by proposition 1, the resulting equilibrium has to be separating. The high type broker-dealers can only separate from the low types if they offer a price of debt such that the low types default even if they pool. As described in proposition 2, the high type broker-dealers are able to secure d^* of debt at rate $q^H(d^*)$ where d^* solves $q^H(d^*) = q^L_d(d^*)$.

The high types then need $d_0 - d^* q^H(d^*)$ additional units of cash to pay off their maturing debt. Since $q^L_d(d^*) = \frac{d_0 - \rho a^L}{d^*}$, this equates to ρa^L units of cash. They must sell $s^H = a^L = \phi a^H$ units of risky assets. In order to separate the high types must liquidate a considerable amount of assets. The low types are unable to secure any funding and default. Thus they liquidate all of their remaining assets, supplying $s^L = a^L = \phi a^H$ units of risky assets. Then the total supply of assets is

$$\alpha s^L + (1 - \alpha)s^H = \phi a^H \tag{3.11}$$

Remarkably, the supply of assets into the period one market for assets does not depend on the fraction of experts α who are hit with the idiosyncratic shock $1-\phi$. Even with α converging to zero, the number of assets that need to be sold can be large. Because of the high type broker-dealer's incentive to separate, a very small number of distressed broker-dealers can induce a large sale of assets.

Combining equations (10) and (11) the price of the asset will be determined as,

$$\delta(\frac{1}{\rho}-1) = I(B-\rho\phi a^H) \tag{3.12}$$

Let ρ_c^* denote the solution to equation (12). ρ_c^* is an equilibrium price of the asset if the high type broker-dealer can borrow at the separating quantity of debt as was assumed. They can borrow if they satisfy the following condition

Condition 1. $d_{floor}^{H,c} \leq (1-\phi)a^H R_c^*$

where $d_{floor}^{H,c} = \frac{1-\rho_c^* a^H}{1-F(R_c^*)}$ and R_c^* solves equation (4) for ρ_c^*.

Condition 1 is a solvency condition for the high type broker-dealers at their separating quantity of debt. The high type broker-dealer may not be able in equilibrium if they become insolvent at the separating quantity of debt. When the high type broker-dealers are forced to liquidate some of their assets in order to separate, they suffer significant losses for each unit of asset they sell. When the losses are severe, either because the discount is high or because they are forced to liquidate a large number of assets, the high type broker-dealers may become insolvent.

Condition 1 can be thought of as governing the size of the idiosyncratic shock $1-\phi$. When the severity of losses that the distressed broker-dealers face is large, the high type broker-dealers do not need to liquidate as large a portion of their assets at a discounted price to separate and are more likely to remain solvent at the separating quantity of debt. The following proposition describes the equilibrium.

Proposition 4. *If condition 1 holds, there is an equilibrium in which, during a crisis state, the broker-dealers experience a contraction in credit but most broker-dealers are able to survive. In a crisis, the high type broker-dealer's offer is determined by proposition 3 and is accepted by the households. The low type broker-dealer defaults in period one regardless of their offer.* $s^L = s^H = \phi a^H$ *and the price of the risky asset* ρ *is determined by equation (12), and* β *by equation (1) where* $q_B = 1$. *In period zero* (d_0, q_0) *= (1,1) as* α *converges to zero.*

System Wide Run

When condition 1 fails, there is no equilibrium outcome in which the high type broker-dealers can borrow during a crisis. The only possible equilibrium outcome is where all broker-dealers default in a crisis and liquidate all their remaining assets. All lending to the broker-dealers is suspended and the broker-dealers suffer a system wide run.

Now suppose that the high type broker-dealers are insolvent at the separating quantity of debt. Then they can no longer borrow and must default. When both types of broker-dealers default they liquidate their entire portfolio of assets. Thus $s^L = a^L = \phi a^H$ and $s^H = a^H$. Furthermore, with α converging to zero equation (10) can now be written as

$$\delta(\frac{1}{\rho} - 1) = I(B - \rho a^H) \tag{3.13}$$

Let ρ_r^* denote the solution to equation (13). Because both high type broker-dealers and low type broker-dealers default in a crisis, the price of period zero debt will be determined as:

$$q_0 = 1 - \delta + \delta\frac{\rho_r^*}{d_0}a^H \tag{3.14}$$

Because the broker-dealer's purchase the risky assets for one unit of cash in period zero, the quantity of period zero debt will equal $d_0 = \frac{1}{q_0}$. Therefore,

$$d_0 = \frac{1}{1-\delta}(1 - \delta\rho_r^* a^H) \tag{3.15}$$

The high type broker-dealer is indeed insolvent at the separating quantity of debt when the price of the risky asset is ρ_r^*, if the following condition is satisfied.

Condition 2. $d_{floor}^{H,r} > (1-\phi)a^H R_r^*$

where $d_{floor}^{H,r} = \frac{1-\rho_r^* a^H}{1-F(R_r^*)}$ and R_r^* solves equation (4) for ρ_r^*. The price ρ_r^* is an equilibrium price only if condition 2 is satisfied.[7] The following proposition describes the equilibrium.

Proposition 5. *When condition 2 holds, there is an equilibrium in which the broker-dealers experience a system wide run. During a crisis, neither the high type broker-dealer or the low type broker-dealer can borrow. All offers are rejected by the households and all broker-dealers default.* $s^L = \phi a^H$ *and* $s^H = a^H$. *The price of the risky asset* ρ *is determined by equation (13), and* β *by equation (1) where* $q_B = 1$. *In period zero* d_0 *is determined by equation (15) and* q_0 *by equation (14).*

3.2.3 A Numerical Example

In this section, I provide a numerical example of an economy that suffers a system wide run during a crisis. This economy serves as a benchmark for the the remainder of this paper.

[7] Note that if conditions 1 and 2 fail simultaneously an equilibrium may not exist. A sufficient condition for the existence of an equilibrium is $\frac{f(x)}{1-F(x)} \leq \frac{1}{x}$ for values $x > R_f^*$. This condition regarding the distribution of risky asset returns guarantees that if condition 1 fails that condition 2 will hold.

Let the number of units of risky assets held by the high type experts be $a^H = 1.2$, the probability of a shock $\delta = 0.1$, and the size of the idiosyncratic shock $1 - \phi = 0.2$ implying $a^L = 0.96$. The borrowing limit is $d^{limit} = 1.05$. Suppose that the distribution of the return of the risk asset R follows a lognormal distribution with parameter values $\mu = -\frac{1}{4}$ and $\sigma^2 = 0.16$. The distribution F then has mean of one and variance of 0.188. Let the marginal return function on the bank's safe project be $I(\beta) = 20e^{-4\beta} + 0.01$ and the borrowing capacity of the bank $B = 2.4$. Assume α converges to zero.

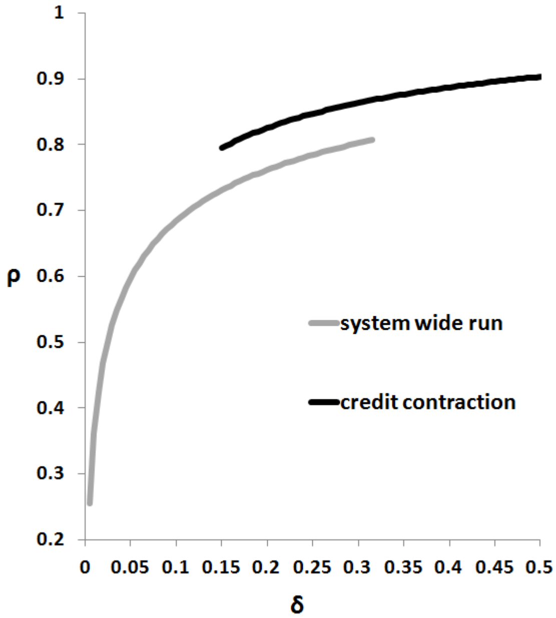
First, suppose that the high type broker-dealers can borrow in equilibrium. In this case the price of the risky asset is decided by equation (12). The implied price of the risky asset is $\rho = 0.75$. At this price the debt floor of the low type is $d^L_{floor} = 1.72$. This is higher than the debt limit, consistent with the assumption that the low type cannot borrow at any quantity and price. However, in this example, condition 1 fails. The debt floor of the high type is $d^H_{floor} = 0.63$ whereas the right side of condition 1 is equal to 0.34. The above is not an equilibrium of the model.

Instead assume that no broker-dealer can borrow in equilibrium. Then the equilibrium asset price solves equation (13). The equilibrium asset price is $\rho = 0.684$. In the debt floor of the low type broker-dealer is $d^L_{floor} = 1.73$ and the high type broker-dealer is $d^H_{floor} = 0.95$. The right-hand-side of condition 2 is equal to 0.32 and thus condition 2 holds. All broker-dealers default during a crisis. In period zero, $d_0 = 1.02$ and $q_0 = 0.98$. Banks invest $\beta = 1.58$ units of cash in safe projects in period zero.

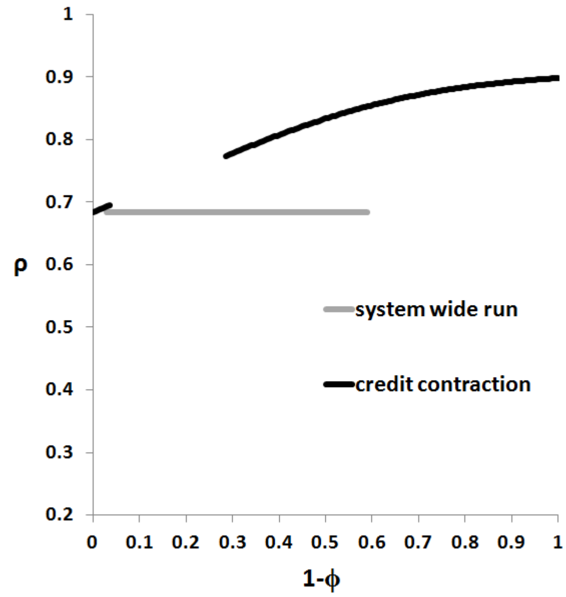
3.2.4 Properties of the Model

3.2.4.1 Comparative Statics

Various parameters determine the equilibrium of the model including the fire sale prices of the assets and which outcome, a complete financial collapse or a more moderate contraction, will



(a) Probability of Aggregate Shock



(b) Magnitude of Idiosyncratic Shock

Figure 3.2: Equilibrium fire sale prices for values of δ and $1 - \phi$

prevail. The price discount incurred by the assets in a fire sale is determined by equations (12) and (13) respectively. The comparative statics can be explicitly worked out by studying the effects of changes in various parameters to the prices implied by the equations.

Panel (a) of Figure 2 illustrates the equilibrium price of the risky asset in the secondary market as a function of the ex-ante probability of the crisis, where all other parameter values are as in the example economy. The figure shows that the price of the asset is increasing in the ex-ante probability of the crisis. Furthermore, if the ex-ante probability of the crisis is small the financial system will experience a system wide run whereas if the crisis is somewhat anticipated the system will suffer a credit contraction but the majority of broker-dealers will survive.

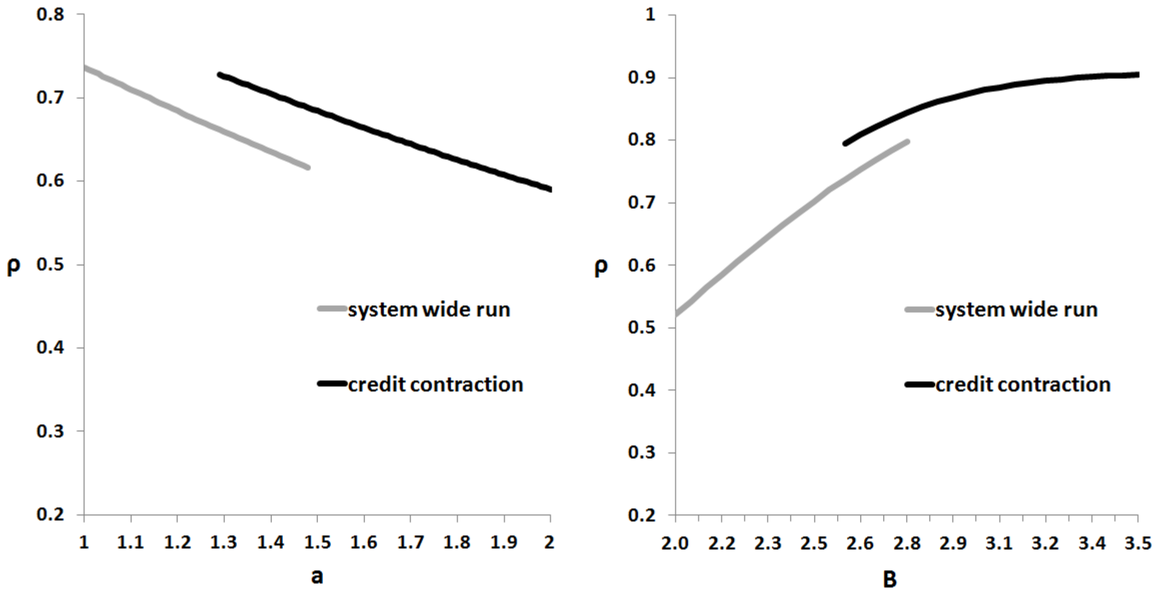
This relationship between the ex-ante probability of a crisis and the equilibrium price of the risky asset holds because the probability of the crisis determines the amount of excess borrowing capacity the banks maintain to capitalize on a fire sale. With a greater probability of a crisis, and thus a fire sale, the banks reserve a greater portion of their borrowing capacity for the possibility of a fire sale. This behavior coincides with behaviors of liquidity hoarding both reported empirically and observed in various models.[8] The focus of many models of liquidity hoarding is on the fact that this behavior decreases the lending activities of these institutions and depresses economic activity in the period leading up to the crisis. This is also true in the current model. However, this paper also highlights a potential benefit to the liquidity hoarding behavior of banks. The extra liquidity that the banks hold may allow the financial system to survive a crisis that they otherwise could not.

Panel (b) illustrates change in the equilibrium price of the asset as a function of the severity of the shock. While the aggregate magnitude of the shock is not important, as evidenced by the fact that α does not affect the outcome, the depth of the difficulties that the distressed institutions face is very important for the equilibrium outcome.

Surprising, the equilibrium outcome is worse for relatively smaller shocks. Panel (b) shows that in the example economy if the distressed institutions lose 20% of their overall value to a shock the financial system will suffer a system wide run. However, if the shock is greater than 30% of the value then there is an equilibrium outcome in which the high type broker-dealers avoid default. With shocks greater than 60% there is no longer an equilibrium in which the system fails.

When the shock is small, the high type broker-dealers must decrease their borrowing quantity significantly in order to separate from the low type broker-dealers. This results in a significant credit contraction and a greater likelihood that the financial system fails. With larger shocks, the low type broker-dealers are left with a much smaller portfolio and readily default with a smaller contraction in credit.

[8]For example, Acharya and Merrouche (2012) report evidence of liquidity hoarding by banks. Acharya and Skeie (2011) construct a model in which a bank hoards excess liquidity in anticipation of adverse asset shocks due to precautionary motives. Gale and Yorumazer (2013) construct a model in which both precautionary and speculative motives for liquidity hoarding exist.



(a) Profitability of Risky Assets (b) Borrowing Capacity of Banks

Figure 3.3: Equilibrium fire sale prices for values of a and B

Figure 3 illustrates the equilibrium price of the asset in response to varying degrees of slack in the financial system. In panel (a) the parameter of interest is the period zero expected return of risky assets. When the return on assets are low the system is vulnerable to a system wide run. When the returns are high the system is better able to withstand a crisis. The result is intuitive as the broker-dealers will have an easier time remaining solvent when their initial profit margins are large. Conditional on which equilibrium outcome prevails, the equilibrium price is decreasing in asset returns. This is due to the fact that a larger a implies that there are more assets that need to be liquidated.

In Panel (b) the parameter of interest is the total borrowing capacity of banks. As can be expected, when the total borrowing capacity of the banks increase the equilibrium price increases.

3.2.4.2 Fragility and Multiple Equilibria

As figures 2 and 3 clearly illustrate, this model exhibits both fragility and multiple equilibria. Brunnermeier and Pedersen(2009) define fragility as "the property that a small change in fundamentals can lead to a large jump in illiquidity." Illiquidity is measured as the deviation of price from its fundamental value. They find that when creditors cannot distinguish between falls in the fundamental value of the asset and deviations of prices from the fundamental value, funding markets can be fragile. This paper finds that market can be fragile even when households can make that distinction.

In addition the model features multiple equilibria for some parameter regions. In some regions an equilibrium with system wide runs and an equilibrium with only credit contractions can both exist. In these regions conditions 1 and 2 hold simultaneously. Which outcome prevails depends on the expected price of risky asset in the secondary market.

The fragility and multiplicity properties of the model have some important ramifications. It suggests that when the economy is in a region close to the boundaries of parameter regions even subtle government policy can have substantial effects. However, it also indicates that sometimes the effects of policy can be discontinuous and surprising. For example, when an economy is in a region that exhibits multiple equilibria the policy authority may be able to change the equilibrium outcome by influencing the expectation of the asset price.

3.3 Policy Implications

In this section I consider the effects of various policies on the outcome of the model. There are two distinct approaches to studying the effects of a policy. One can assume an ex-post approach and suppose that the government intervention is unexpected by the agents. In the midst of an unexpected crisis this may be the appropriate way to evaluate policy. However, if a policy is to be considered an appropriate course of action in a more general sense, it may be more appropriate to consider a situation in which the policy intervention is anticipated ex-ante. This approach can incorporate the fact that the expectation of policy intervention may alter agent behavior ex-ante. That is the approach I take here.

There are some difficulties in comparing the desirability of the outcomes. The difficulty arises from the fact that there is no natural concept of social welfare in the model. One option would be to define a social welfare function balancing the welfare considerations of the various agents in the model. That is not the approach I take. Instead I compare the equilibrium outcomes of economies in which the policies have been implemented against the equilibrium outcome of a perfect information economy. The implicit assumption is that the perfect information outcome should be efficient.

I find that the policy that mandates the participation of all broker-dealers in a clearing house for loans during a state of crisis can garner outcomes that are very close to the outcome of perfect information economy. The effect of a counter-cyclical borrowing limit is ambiguous.

3.3.1 Perfect Information Benchmark

In the perfect information case, the equilibrium outcome will depend on the borrowing limit. First, let β_0 denote the investment quantity of the bank such that $I(\beta_0) = 0$ (if $I(\beta) > 0$, $\beta_0 = B$). Then, if the borrowing limit is high, such that $(1 - d^{limit} q^H(d^{limit}))$ is less than $B - \beta_0$, the price of the risky asset on the secondary market will equal its expected value of one. Since the price of the risky asset is equal to the expected value of the asset, there is no surplus to be gained from buying or selling the assets.

Because the expected profit from purchasing risky assets is zero, the banks will fully invest in safe projects in period zero to the point that $\beta = \beta_0$. The broker-dealers are indifferent to the amount of debt they issue in a crisis. In equilibrium, the high type broker-dealers can borrow any amount $(d_1, q^H(d_1))$ where $d_1 q^H(d_1)$ is between $1-(B-\beta_0)$ and one. The low type broker-dealers will be unable to borrow in a crisis and will default.

On the other hand, if $(1 - d^{limit} q^H(d^{limit}))$ is greater than $B - \beta_0$, the high type broker-dealer will borrow to the limit d^{limit}. The price of the asset will be determined such that

$$\delta(\frac{1}{\rho} - 1) = I(B - (d_0 - d^{limit} q^H(d^{limit}))) \tag{3.16}$$

In either case the high type broker-dealers are able to fully repay their period zero debt whereas low type broker-dealers default, so the price of period zero debt is given as

$$q_0 = 1 - \delta\alpha(1 - \frac{\phi a^H}{d_0}) \tag{3.17}$$

Because the broker-dealer's purchase the risky assets for one unit of cash in period zero, the quantity of period zero debt will equal $d_0 = \frac{1}{q_0}$. Therefore,

$$d_0 = \frac{1 - \delta\alpha\phi a^H}{1 - \delta\alpha} \tag{3.18}$$

As α converges to zero (d_0, q_0) converges to (1,1).

The perfect information outcome of the example economy yields the latter case. At the debt limit of 1.05 the price of debt will be 0.88. The high type broker-dealers will be able to secure 0.925 in funds and will need to make up the difference by selling their assets. Nevertheless, the amount of funds that need to be obtained by selling assets is comparatively small and the fire sale price will be determined to be 0.894. The high type broker-dealers will be able to avoid default in period one.

Ideally, the perfect information benchmark could be achieved if the policy authority could mandate the truthful disclosure of financial status for all financial institutions. However, implementing such a policy may prove difficult. Does the authority have the expertise to swiftly and accurately assess the health of financial institutions? Recent experience during the financial crisis suggests otherwise. Without the technical capacity to promptly assess the state of the financial institutions, it is unclear as to how the governing authority can force them to truthfully reveal their financial status. It seems unlikely that the institutions would report their financial difficulties, and even less likely that potential creditors would trust the announcements.

3.3.2 Government Clearing House for Loans

While it may not be difficult to achieve the perfect information outcome through policy channels, it may be possible to attain an outcome that is very close. Because the credit contraction stems not from the household's refusal to extend credit but from the incentive of the high type broker-dealers to separate, it may be possible to improve the equilibrium outcome by preventing all attempts at signaling. If the government could enforce a pooling equilibrium by not allowing broker-dealers the opportunity to signal their type through their debt contracts, all broker-dealers would be better off.

The government could sustain pooling equilibria by mandating all broker-dealers to participate in a government clearing house for debt. During a crisis, all broker-dealers would only be allowed to borrow through a government clearing house. The government would choose a total quantity and unit price of debt that is the same for all broker-dealers. The households can either choose to provide funds to the clearing house or choose not to. As long as terms that the government chooses allow the households to break even in expectation, the households will participate willingly.

It is important that the participation in the clearing house is mandatory during from a period one perspective. If a broker-dealer is able to opt of participating in the midst of a crisis, this can serve as a signal similar to that of the original model. The pooling equilibrium will be unsustainable. Broker-dealers cannot be allowed to decide on participation after their type has been (privately) revealed.

Nevertheless, the idea of mandating all broker-dealers to participate in this clearing house is not far fetched or particularly tyrannical. From a period zero perspective, all broker-dealers benefit from the implementation of this policy. As long as they are not allowed to opt in at a future date, they would all agree to participate in this policy in period zero. Participation could be achieved voluntarily.

This policy can be very effective. In fact, if $I(B) \leq 0$ and the fraction of low type broker-dealers α converges to zero, a well implemented policy would result in an aggregate outcome that is arbitrarily close the outcome in the perfect information case. Let $\pi^H(d_1, q_1)$ be the expected profit of the households of equation (3) when a equals a^H and $\pi^L(d_1, q_1)$ the expected profit when a equals a^L.

Proposition 6. *Let $\beta^{PI}, \rho^{PI}, (d_0^{PI}, q_0^{PI}), (d_1^{H,PI}, q_1^{H,PI})$ denote the respective equilibrium outcomes in the perfect information economy and $\beta^{CH}, \rho^{CH}, (d_0^{CH}, q_0^{CH}), (d_1^{CH}, q_1^{CH})$ the equilibrium outcomes with a government clearing house for debt where (d_1^{CH}, q_1^{CH}) is the debt contract of the clearing house determined by the government.*

Let $\alpha > 0$ converge to zero. Then, if $I(B) \leq 0$, there exists an $\eta > 0$ for all $\epsilon > 0$ such that,

if $|d_1^{CH} - d_1^{H,PI}| < \eta$, $\quad |q_1^{CH} - q_1^{H,PI}| < \eta \quad$ and $\quad (1-\alpha)\pi^H(d_1^{CH}, q_1^{CH}) + \alpha\pi^L(d_1^{CH}, q_1^{CH}) \geq 0$,

then $|\rho^{CH} - \rho^{PI}| < \epsilon$, $\quad |\beta^{CH} - \beta^{PI}| < \epsilon$, $\quad |d_0^{CH} - d_0^{PI}| < \epsilon \quad$ and $\quad |q_0^{CH} - q_0^{PI}| < \epsilon$.

Even if the ideal conditions do not hold the policy seems to be reasonably effective. In other words, even when $I(B) > 0$ and (d_1^{CH}, q_1^{CH}) is not arbitrarily close to $(d_1^{H,PI}, q_1^{H,PI})$, as long as the government chooses reasonable terms the aggregate outcome should be a vast improvement over the alternative of having no policy.

For example, suppose that in the example economy of section 2.5 that the government chooses $d_1^{CH} = 1$ and $q_1^{CH} = 0.85$. Then the implied price of the asset ρ^{CH} will be 0.889, considerably higher than the 0.684 of the benchmark economy. The households will be willing to accept these terms because in equilibrium their expected profit from each unit of debt purchased is 0.02, which is strictly greater than zero. The high type broker-dealers will not default in period one and the financial system will not suffer a system wide run.

3.3.3 Counter-Cyclical Borrowing Limits

Recently there has been much debate about the role of borrowing limits, mostly in the form of capital requirements, as a regulatory measure against rapid liquidity contractions in the financial sector. Imposing capital conservation buffers has been suggested as a method of implementing counter-cyclical borrowing limits.

It turns out that, within the framework of this paper, the effects of a counter-cyclical borrowing limit is ambiguous. The counter-cyclical borrowing limit is effective only if the borrowing limit increases enough for the low type broker-dealers to borrow even when identified. An equilibrium potentially exists where the low type broker-dealer borrows up to the borrowing limit at the maximum price the household will accept. The high type broker-dealer borrows less, but at more favorable price.

Figure 4 illustrates the credit market outcome of this equilibrium. The red point represents the equilibrium contract of the low type broker-dealer and the blue point represents the contract of the high type broker-dealer. The debt contract of the high type is such that the low type is exactly indifferent between the two contracts.

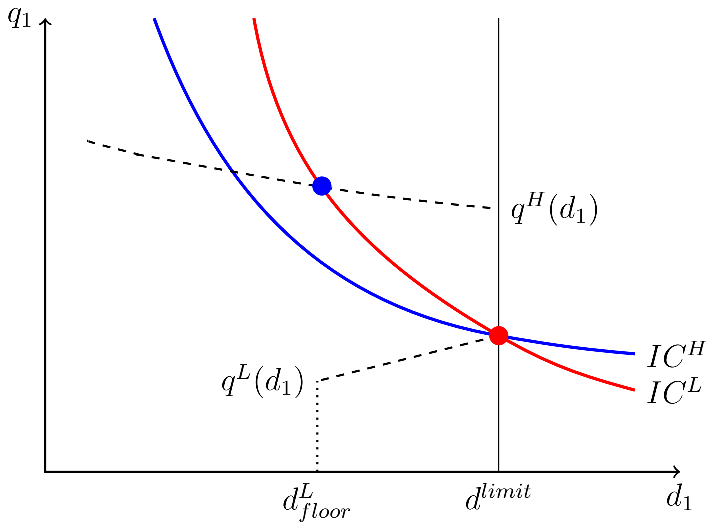


Figure 3.4: A potential credit market equilibrium with counter-cyclical borrowing limits.

However, for this equilibrium to exist, the increase in the borrowing limit must be substantial. With higher debt limits the supply of assets in fire sale markets decrease and the fire sale discount for the assets will become smaller. As the potential profit margins decrease the low type broker-dealers need to purchase still larger quantities of the assets to become solvent. If the debt limit does not increase beyond the debt floor of the low type broker-dealers the policy will have no effect.

Applying the policy to the example economy shows that the debt limit must increase significantly for the policy to be effective. Numerically solving the model shows that the debt limit must be greater than 2.1 to be effective. If the debt limit is below 2.1, the policy has no effect and the equilibrium will be as in the example economy.

When the borrowing limit is equal to $d^{limit} = 2.1$ the prevailing fire sale price is $\rho = 0.883$. The debt floor of the low type broker-dealer is 1.993 and thus they can borrow in equilibrium. They will borrow up to the debt limit of 2.1 and the price per unit debt will equal 0.666. The high type broker-dealer will borrow a nominal amount of 0.88 for the price 0.887.

The analysis above shows that a policy of counter-cyclical borrowing limits can be an effective measure against systemic failures in financial systems. However, it also cautions that for the policy to take effect the magnitude of the counter-cyclicality of the borrowing limit may need to be quite large to be successful. A moderate and mechanical approach to counter-cyclical borrowing limits may result in a completely ineffective policy.

3.4 Discussion

Commitment to Quantity and Collateralized Lending

A question that is not addressed in the main section of this paper is whether it is reasonable to believe that financial institutions can commit to their declared borrowing quantities. While this is a legitimate concern, in the end it seems plausible that the financial institutions could indeed commit. If the institution, for example, was borrowing from commercial paper markets the total quantity of debt they try to secure could be plainly visible. Admittedly, this does not preclude the institution from attempting to secretly secure funds in addition to their committed quantity.

Nevertheless, there is another method to enforce commitment that is reminiscent to transactions via collateralized lending. The financial institutions and their creditors could ensure commitment by hypothecating the assets as collateral. To ensure commitment to the quantity promised, the financial institutions must first sell some their assets to acquire the funds required to payoff their maturing debt, in excess of what they will eventually be able to borrow. Then they can provide their entire remaining portfolio as collateral and ensure commitment to their declared quantity. The implied haircut in this case would be the difference between the value of the assets and the amount of debt they issue. The interest rate would be the inverse of the price per unit of debt. Repurchase agreements and asset-backed commercial paper markets are some examples of such collateralized lending arrangements.[9]

[9] A detailed description of tri-party repos and their behavior during the crisis can be found in the paper by Krishnamurthy, Nagel and Orlov(2014).

Quantitative Easing

While not the focus of this paper, the analysis of this paper suggests an effect of quantitative easing that is not often discussed. Proposition 6 states that if banks are unconstrained in period zero $(I(B) \leq 0)$, the government can achieve a perfect information outcome by implementing a government clearing house for loans. This result is predicated on banks being unconstrained because if $I(B) > 0$ even a very small amount of risky assets needing to be liquidated can generate a discrete drop in the price of the risky asset. Quantitative easing, or the act of providing banks with excess reserves and thus excess liquidity, is equivalent to increasing B in this model. Increasing B to the point that the banks are unconstrained, can improve the outcome of this policy intervention.

The result need not be confined to this model. Providing ample liquidity to banks may be beneficial as it may allow banks to provide liquidity in various asset markets and help curtail harmful fire sales and stabilize asset prices during a crisis. Even if certain frictions inhibit an immediate and direct effect, quantitative easing may still be useful in conjunction with various other government policies geared toward stabilizing asset markets. The analysis of this paper suggests that even if quantitative easing cannot achieve much on its own, it can help maximize the effects of certain crisis management policies.

3.5 Related Literature

This paper is related to several strands of literature on financial market failures and crises. The literature on bank runs starting from Diamond and Dybvig (1983) is large and diverse. Some notable papers include Uligh (2010) who considers uncertainty aversion, Goldstein and Pauzner (2004) who use global games techniques to derive unique equilibrium predictions depending on the fundamentals and He and Xiong (2012) that extends runs to dynamic settings. Recently Martin et al.(2014) model bank runs of repo markets during the crisis.

There is a large literature regarding asset illiquidity, market freezes and funding difficulties that arise from asymmetric information problems. Stiglitz and Weiss (1981) show how credit may be rationed even with competitive lenders when there is asymmetric information between lenders and borrowers. Some recent papers in this vein include House and Masatlioglu (2010) and Kurlat (2013).

A closely related literature studies signaling and screening in credit markets with asymmetric information. Leland and Pyle(1977) present a model of capital structure where the entrepreneur's willingness to invest in their own projects serves as a signal of project quality. Bester (1985,1987) finds that when there is a cost to posting collateral, banks maybe able to screen applicants by conditioning rates as a function of collateral requirements. Milde and Riley (1987) find that banks may be able to screen borrowers when their return is a function of loan size by offering a loan schedule. These papers are similar to this paper in that they find mechanisms that allow the borrowers to be separated by their credit worthiness.

This paper is also related to the literature regarding liquidity mismatch and the interaction between credit market conditions, asset market conditions and the health of financial institutions. Brunnermeier and Pedersen (2009) analyze the complex relationship between what they term as funding liquidity and market liquidity. They find that the two can be mutually reinforcing and lead to liquidity spirals. Diamond and Rajan (2005) find that bank failures can shrink the aggregate pool of liquidity and this shortage of liquidity can lead to a systemic crisis. Holmstrom and Tirole (1998) study how the suppy and demand of liquidity interact and affect the investment decision of the firm.

In a recent paper, Shleifer and Vishny (2011) suggest fire sale risk as the most coherent mechanism in understanding systemic risk and the propagation of the financial crisis. Many papers utilize fire sales to explain financial market failures. Diamond and Rajan (2011) take this approach when they study why some banks may deliberately increase their illiquidity in a crisis. Kiyotaki and Moore (1997) show that this approach can generate large credit cycles and amplifications of the business cycle in the real economy. Stein (2012) studies how monetary policy can be utilized as a tool for financial stabilization when fire sales may occur. Shelifer and Vishny (1997), Gennaioli, Shelifer and Vishny (2013) and Brunnermeier and Pedersen (2009) generate fire sales by assuming that outside creditors cannot distinguish fire sales from deterioration of fundamental asset value.

Recent papers by Alvarez and Barlevy (2013) and Caballero and Simsek (2013) are related to this paper in that they model how uncertainty over which institutions suffer losses can lead to contagion, credit market failures, and even fire sales. However, these papers take the approach of modeling complexity in network structures where banks need to understand the cross exposures to accurately assess individual risk but find it difficult to do so, whereas in this paper the key mechanism is the signaling via debt contracts.

Lastly, the policy interventions in this paper have a connection to research on government bailouts and financial regulations. Farhi and Tirole (2012) study firm leverage and maturity decision in anticipation of distress and identify a strategic complementarity that arises due to the possibility of government intervention. Chari and Kehoe (2013) show that in a model with bankruptcy costs ex-ante regulation of debt-to-value ratio can eliminate the incentive of the authority to bail out distressed firms.

3.6 Conclusion

This paper offers a model of financial market failures and illustrates how system wide runs can be triggered by small shocks to asset values. It illustrates the channels of amplification and analyzes the effects of various policies in minimizing the consequences of a financial crisis. It sheds a new light on the ongoing discussions for prudential regulation and how they can be designed to be more effective.

In addition to the main result of system wide runs, many features and predictions of this model seem to be consistent with the empirical observations during the financial crisis. Acharya and Merrouche (2012) find that liquidity provisions increased significantly leading up to the onset of the sub-prime crisis for settlement banks in the U.K. In this paper, banks are incentivized to increase their liquidity provisions when they anticipate fire sales to take advantage of the lucrative investment opportunities they provide. Mitchell and Pulvino (2010) find that large deviations from arbitrage in the order of 10-15% were commonly observed during the crisis and such deviations persisted for long periods. They find that very similar securities that were the basis of many low risk arbitrage strategies of hedge funds experienced steep and long-lasting mispricings during the crisis due to the arbitragers' inability to borrow. The magnitude, persistence, and prevalence of the mispricings reported by Mitchell and Pulvion lend strong support for the arguments of this paper.

3.7 References

Acharya, Viral V., and David Skeie. 2011. "A Model of Liquidity Loarding and Term Premia in Inter-bank Markets." *Journal of Monetary Economics*, 58, 436-447

Acharya, Viral V., and Ouarda Merrouche. 2012. "Precautionary Hoarding of Liquidity and Interbank Markets: Evidence from the Subprime Crisis." *Review of Finance*, 1-54

Alvarez, Fernando, and Gadi Barlevy. 2014. "Mandatory Disclosure and Financial Contagion." Working paper

Bester, Helmut. 1985. "Screening vs. Rationing in Credit Markets with Imperfect Information." *American Economic Review*, 75(4), 850-855

Bester, Helmut. 1987. "The Role of Collateral in Credit Markets with Imperfect Information." *European Economic Review*, 31, 887-899

Brunnermeier, Markus, and Martin Oehmke. 2013. "The Maturity Rat Race." *Journal of Finance*, 68(2), 483-521

Brunnermeier, Markus, and Lasse Pedersen. 2009. "Market Liquidity and Funding Liquidity." *Review of Financial Studies*, 22, 2201-2199

Caballero, Ricardo J., and Alp Simsek. 2013. "Fire Sales in a Model of Complexity." *Journal of Finance*, 68(6), 2549-2587.

Chari, V.V., and Patrick J. Kehoe. 2013. "Bailouts, Time Inconsistency, and Optimal Regulation." Working paper

Diamond, Douglas W. 1991. "Debt Maturity Structure and Liquidity Risk." *Quarterly Journal of Economics*, 106, 709-737

Diamond, Douglas W., and Philip H. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy*, 91(3), 401-419

Diamond, Douglas W., and Raghuram G. Rajan. 2005. "Liquidity Shortages and Banking Crises." *Journal of Finance*, 60(2),

Diamond, Douglas W., and Raghuram G. Rajan. 2011. "Fear of Fire Sales, Illiquidity Seeking, and Credit Freezes." *Quarterly Journal of Economics*, 126(2), 557-591

Farhi, Emmanuel, and Jean Tirole. 2012. "Collective Moral Hazard, Maturity Mismatch, and Systemic Bailouts." *American Economic Review*, 102(1), 60-93

Fudenberg, Drew, and Jean Tirole. 1991. "Game Theory." The MIT Press

Gale, Douglas, and Tanju Yorulmazer. 2013. "Liquidity Hoarding." *Theoretical Economics*, 8, 291324

Gennaioli, Nicola, Andrei Shleifer, and Robert W. Vishny. 2013. “A Model of Shadow Banking.” *Journal of Finance*, 98(4), 1331-1363

Gorton, Gary, and Andrew Metrick. 2012. “Securitized Banking and the Run on Repo.” *Journal of Financial Economics*, 104, 425-451

Hart, Oliver, and John Moore. 1994. “A Theory of Debt Based on Inalienability of Human Capital.” *Quarterly Journal of Economics*, 109, 841-879

He, Zhiguo, In Gu Khang, and Arvind Khrishnamurthy. 2010. “Balance Sheet Adjustments during the 2008 Crisis.” *IMF Economic Review*, 58(1), 118-156

He, Zhiguo, and Wei Xiong. 2012. “Dynamic Debt Runs.” *Review of Financial Studies*

Holmstrom, Bengt, and Jean Tirole. 1998. “Private and Public Supply of Liquidity.” *Journal of Political Economy*, 106(1), 1-40

House, Christopher L., and Yusufcan Masatlioglu. 2015. “Managing Markets for Toxic Assets.” *Journal of Monetary Economics*, 70, 84-99

Kiyotaki, Nobuhiro, and John Moore. 1997. “Credit Cycles.” *Journal of Political Economy*, 105(2), 211-248

Krishnamurthy, Arvind, Stefan Nagel, and Dmitry Orlov. 2014. “Sizing Up Repo.” *Journal of Finance*, 69(6), 23812417

Kurlat, Pablo. 2013. “Lemons Markets and the Transmission of Aggregate Shocks.” *American Economic Review*, 103(4) 1463-89.

Martin, Antoine, David Skeie, and Ernst-Ludwig von Thadden. 2014. “Repo Runs.” *Review of Financial Studies*, 27(4), 957-989

Milde, Hellmuth, and John G. Riley. 1988. “Signaling in Credit Markets.” *Quarterly Journal of Economics*, 103(1), 101-129

Mitchell, Mark, and Todd Pulvino. 2012. “Arbitrage Crashes and the Speed of Capital.” *Journal of Financial Economics*, 104, 469-490

Shleifer, Andrei, and Robert W. Vishny. 2011. “Fire Sales in Finance and Macroeconomics.” *Journal of Economic Perspectives*

Spence, Michael. 1973. “Job Market Signaling.” *Quarterly Journal of Economics*, 87(3), 355-374

Stein, Jeremy. 2012. “Monetary Policy as Financial-Stability Regulation .” *Quarterly Journal Of Economics* 127(1), 57-95

Stein, Jeremy. 2013. “The Fire-Sales Problem and Securities Financing Transactions.” Remarks at the Federal Reserve Bank of New York Workshop on Fire Sales as a Driver of Systemic Risk in Triparty Repo and other Secured Funding Markets, October 4, 2013.

Stiglitz, Joseph E. and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review*, 71(3), 393-410

Townsend, Robert M. 1979. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory*, 21(2), 265-293

Uhlig, Harald. 2010. "A Model of a Systemic Bank Run." *Journal of Monetary Economics*, 57, 78-96

APPENDIX A

Product Revenue and Price Setting: Evidence and Aggregate Implications

A.1 List of Product Categories and Markets

A.1.1 Product Categories

The following table provides the list of the 30 product categories in our main sample. Product categories are defined and categorized by the The Nielsen Company. They were chosen randomly to the sample using product codes provided by The Nielsen Company. The full list is provided in Table A.1.

Additionally we add product categories for the results in section 4.1. They product categories we add are the following.

ANCHOVY PASTE, BABY CARE PRODUCTS-LOTIONS, BABY CARE PRODUCTS-OIL, BABY CARE PRODUCTS-OINTMENTS, BABY CARE PRODUCTS-POWDER, BARBECUE SAUCES, CANNED FRUIT - APPLE SAUCE, CANNED FRUIT - APPLES, CANNED FRUIT - BERRIES, CANNED FRUIT - FIGS, CANNED FRUIT - FRUIT MIXES & SALAD FRUITS, CANNED FRUIT - GRAPES, CANNED FRUIT - PINEAPPLE, CANNED FRUIT - PRUNES, CANNED FRUIT-APRICOTS, CANNED FRUIT-CHERRIES, CANNED FRUIT-FRUIT COCK-

Table A.1: Product categories

Product Description
PIE & PASTRY FILLING - CANNED
CANNED FRUIT - ORANGES
CANNED FRUIT - PEACHES - FREESTONE
GRAVY - CANNED
SEAFOOD-CRAB-CANNED
SEAFOOD - SARDINES - CANNED
SEAFOOD-TUNA-SHELF STABLE
CAT FOOD - WET TYPE
CAT FOOD - MOIST TYPE
DOG & CAT TREATS
EGG MIXES-DRY
CRACKERS - SPRAYED BUTTER
CRACKERS - OYSTER
COFFEE - SOLUBLE
COFFEE SUBSTITUTES
BEER
NEAR BEER/MALT BEVERAGE
GIN
VODKA
WINE-SANGRIA
WINE-SWEET DESSERT-IMPORTED
CLEANERS-METAL
CLEANERS-HUMIDIFIERS/VAPORIZERS
COOKER STEAMER AND DEHYDRATOR APPLIANCE
AIR PURIFIER AND CLEANER APPLIANCE
NUTRITIONAL SUPPLEMENTS
VITAMINS-B COMPLEX W/C
MANICURING NEEDS
HAIR SPRAY - MEN'S
BABY CARE PRODUCTS-BATH

TAIL, CANNED FRUIT-GRAPEFRUIT, CANNED FRUIT-PEACHES-CLING, CANNED FRUIT-PEARS, CANNED FRUIT-PLUMS, CAT FOOD - DRY TYPE, CATSUP, CHILI SAUCE, CLEANERS - BATHROOM, CLEANERS - DISINFECTANTS, CLEANERS - POWDERS, CLEANERS - WINDOW, COOKING SAUCE, COOKING WINE & SHERRY, CRANBERRIES - SHELF STABLE, DOG FOOD - DRY TYPE, DOG FOOD - MOIST TYPE, DOG FOOD - WET TYPE, ENTREES/SIDE DISHES - SHELF STABLE, FISH & SEAFOOD & COCKTAIL SAUCE, FLOOR CARE-CLEANERS, GLAZES - FRUIT, GLAZES-MEAT, GRAVY - CANNED, GRAVY MIXES - PACKAGED, HOT SAUCE, MEAT SAUCE, MEXICAN SAUCE, MINERALS, MUSHROOM SAUCE, MUSTARD, OVEN CLEANERS, PET CARE - DOMESTIC BIRD FOOD, PET CARE - PET FOOD, PET CARE - WILD BIRD FOOD, PIZZA SAUCE, PUMPKIN-CANNED, RUG CLEANERS, SAUCE & SEASONING MIX-REMAINING, SAUCE & SEASONING MIX-REMAINING MEXICAN, SAUCE MIX - CHEESE, SAUCE MIX - MEAT LOAF, SAUCE MIX - SOUR CREAM, SAUCE MIX - SPAGHETTI, SAUCE MIX - TACO, SAUCES - DIPPING, SAUCES - MISCELLANEOUS - SHELF STABLE, SEAFOOD - ANCHOVIES, SEAFOOD - OYSTERS - CANNED, SEAFOOD - REMAINING - CANNED, SEAFOOD - SALMON - CANNED, SEAFOOD - SHRIMP - CANNED, SEAFOOD-CLAMS-CANNED, SEASONING MIX - CHILI, SEASONING MIX - SLOPPY JOE, SPAGHETTI/MARINARA SAUCE, TABASCO/PEPPER SAUCE, TOILET BOWL - CLEANERS, TOILET BOWL - DEODORIZERS, UPHOLSTERY CLEANERS, VACUUM CLEANER BAGS & ACCESSORIES, VINEGAR, VITAMINS-MULTIPLE, VITAMINS-REMAINING, VITAMINS/TONICS-LIQUID & POWDER, WORCESTERSHIRE SAUCE.

A.1.2 Markets

We choose 32 markets on which we conduct our analysis. The designated market areas are defined by The Nielsen Company and correspond approximately (although not exactly) to a metropolitan statistical area (MSA). The full list is provided in Table A.2

A.2 Robustness: Revenue distribution

To construct log revenue moments, we find the mean, median, standard deviation, and the 10th and 90th percentiles for each market product category and month. We only consider month product category markets with more than 100 UPC stores observations, this is to ensure our results exclude potentially noisy observations. The spread of the distribution is defined as the difference between the 90th and 10th percentiles. We not only include mean and standard deviation of the log revenue distribution but also consider the mean and the spread to allow for greater robustness to outliers.

In Table A.3 we show the results for the 30 product categories used in section 2. The qualitative results remain unchanged. In Table A.4 we show the results of the same set of regressions including market-product category fixed effects instead of market fixed effects and product category fixed effects separately. The results are again robust.

A.3 Inference into the Selection Effect

At least since Golosov and Lucas (2007), it is generally believed that large idiosyncratic productivity shocks are needed to match the large average size of price changes. Golosov and Lucas find that such large productivity shocks introduce a large *selection effect* in menu cost models that mitigate the output response to monetary shocks. Also known as the extensive margin effect, the selection effect refers to the number of additional products that change price in response to a monetary shock, that would not have changed their price if there was no monetary shock. If this number is large, the aggregate price level is very responsive to aggregate shocks and the output response is

Table A.2: Designated market areas

DMA Name
PORTLAND-AUBURN ME
NEW YORK NY
PHILADELPHIA PA
DETROIT MI
BOSTON (MANCHESTER) MA-NH
FT WAYNE IN
CLEVELAND OH
WASHINGTON DC (HAGERSTOWN MD)
BALTIMORE MD
CINCINNATI OH
CHARLESTON SC
ATLANTA GA
INDIANAPOLIS IN
LOUISVILLE KY
HARTFORD & NEW HAVEN CT
TAMPA-ST PETERSBURG (SARASOTA)
RALEIGH-DURHAM (FAYETTEVILLE)
CHICAGO IL
ST LOUIS MO
MINNEAPOLIS-ST PAUL MN
KANSAS CITY MO-KS
OKLAHOMA CITY OK
OMAHA NE
NASHVILLE TN
WICHITA-HUTCHINSON PLUS KS
DES MOINES-AMES IA
LITTLE ROCK-PINE BLUFF AR
DENVER CO
PHOENIX AZ
BOISE ID
ALBUQUERQUE-SANTA FE NM
LOS ANGELES CA
SAN FRANCISCO-OAKLAND-SAN JOSE
SEATTLE-TACOMA WA

Table A.3: Revenue distribution: 30 product categories

	(1) Mean	(2) Median	(3) Std Deviation	(4) Spread
unemployment	-0.982	-1.029	-0.239	-0.978
	(0.311)	(0.359)	(0.226)	(0.385)
Observations	80,988	80,988	80,988	80,988
R-squared	0.804	0.802	0.666	0.733

Note: All specifications include market, month and product category fixed effects. Standard errors are clustered on month, market, and product category separately.

Table A.4: Revenue distribution: market-product category fixed effect

	(1) Mean	(2) Median	(3) Std Deviation	(4) Spread
unemployment	-1.710	-1.854	-0.407	-1.248
	(0.476)	(0.511)	(0.162)	(0.416)
Observations	351,115	351,115	351,115	351,115
R-squared	0.824	0.817	0.748	0.751

Note: All specifications include market, month and product category fixed effects. Standard errors are clustered on month, market, and product category separately.

small, which is what Golosov and Lucas find.

Midrigan (2011), however, argues that this effect is quite small in economies have large and infrequent idiosyncratic productivity shocks and multi-product firms. Alvarez and Lippi (2014) and Alvarez et al. (2016) develop the idea of multi-product firms further and find a relationship between micro level moments such as the frequency of adjustment, and the variance and kurtosis of the distribution of price changes to the aggregate output response of an economy to monetary policy shocks in certain settings. Nevertheless, it is still difficult to compare the strength of the selection effect implied by models to the data as it is difficult to observe directly.[1]

In our framework, it is possible to gain some insight into the empirical magnitude of the selection effect from the relationship between revenue and probability of price adjustment. The strength of both the revenue effect and the selection effect depend on the density of prices that are far away from the optimal price and near the edges of the inaction region. A stronger effect in one implies that the other will be stronger as well.

Figure A.1 depicts this relationship. The horizontal axis depicts the range of prices for a product for which p^* is the optimal price, p^H is the upper bound of the inaction region and p^L is the lower bound. If the inherited price p^* of a product is either greater than p^H or less than p^L the price is adjusted to the optimal price . The red lines labeled GL depicts the distribution of prices within the inaction region in Golosov and Lucas' (2007) setting. The solid black lines represent the distribution of prices in Midrigan's (2011) setting. In Golosov and Lucas the prices are distributed relatively evenly across the inaction region whereas in Midrigan there is a large mass of prices at the optimal price (represented by the thick vertical line) and a low density of prices else where.

Consider an expansionary monetary shock that increases marginal cost and in turn the optimal price. This moves the entire inaction region of inherited prices upward. The new range of inaction

[1] In addition to these papers Caplin and Spulber (1987) and Caballero and Engel (2007) analyze the selection effect and its aggregate consequences in detail.

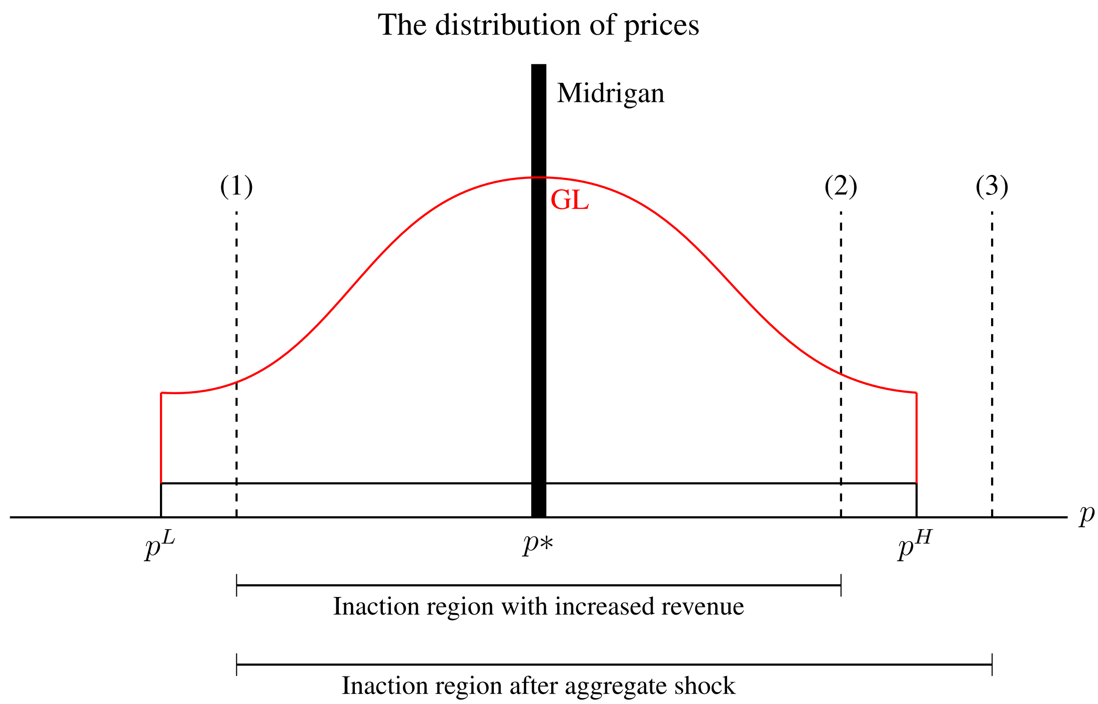


Figure A.1: Inference into the selection effect

is now between the dashed lines labeled (1) and (3). Prices previously located below the dashed line labeled (1) now must adjust to the new optimal price. Note that the mass of prices that adjust is much greater in the GL case than in Midrigan's. This difference in the density of prices that adjust is why there is a large discrepancy in the magnitude of the selection effect between the two economies.

Now consider the relationship between revenue and likelihood of price adjustment. As we show in section 2.3, an increase in revenue diminishes the inaction region. Suppose that an increase in revenue shrinks the range of inaction to areas between dashed lines (1) and (2). Prices previously located below dashed line (1) or above line (2) must now adjust. The increase in the probability of price adjustment is the result of prices outside these lines.

In this manner, the increase in the probability of price adjustment with revenue and the strength of the selection effect is closely related as both depend on the density of prices near the edges of the inaction region. An accurate representation of one should also lead to a good approximation of the other.

A.4 Extended Model

To add habit formation, we replace the household's problem in (1.12) with,

$$\max E_t \sum_{\tau=0}^{\infty} \beta^{\tau} \{\phi_{t+\tau} \frac{(C_{t+\tau} - H_{t+\tau-1})^{1-\frac{1}{\gamma}}}{1-\frac{1}{\gamma}} - \frac{N_{t+\tau}^{1+\frac{1}{\eta}}}{1+\frac{1}{\eta}}\}$$

where habit H_t is external. The stock of external habit follows the process

$$H_t = \lambda C_t$$

as commonly found in the literature. We also replace the nominal GDP process in equation (1.16) with,

$$g_{t+1} = \rho_m g_t + \epsilon_{t+1}^m$$

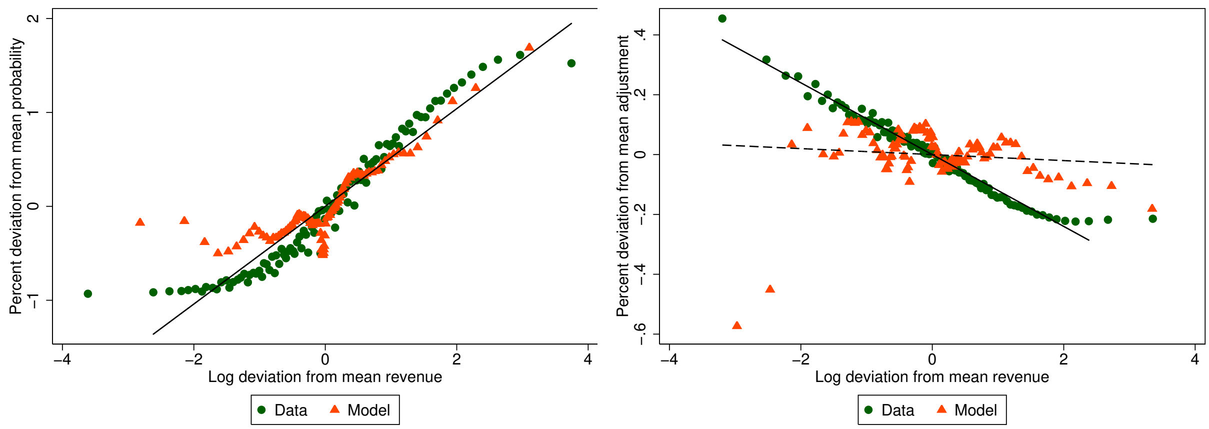
where $g_t = lnM_t - lnM_{t-1}$ is the growth rate of nominal GDP and ρ_m is the autocorrelation of the growth rate. Following Midrigan (2011), we calibrate the autocorrelation ρ_m to be 0.61 and the standard deviation of the nominal GDP shock to be 0.0037. The parameter governing habit formation λ is calibrated to 0.77 following the estimated values in Christiano et al. (2011). Table A.5 shows the target moments of the extended model and Table A.6 shows the calibrated parameters.

Figure A.2 compares the relationship between price setting behavior in the steady state in the extended model and the data. Panel (a) shows that the relationship between revenue and probability of adjustment in the model closely resembles the relationship in the data for the extended model. The green dots show the relationship between log revenue and the probability of adjustment in

Table A.5: Target moments

Steady state moments	Target	Extension
Frequency of price adjustment	0.11	0.12
Average size of price adjustment	0.10	0.11
Median size of price adjustment	0.085	0.082
$\frac{(\epsilon-1)^2 var(price)}{var(revenue)}$	0.20	0.22
$\frac{\partial prob}{\partial lnrev} \times \frac{1}{prob}$	0.52%	0.53%
Business cycle moments		
$\frac{\Delta \overline{rev}}{\Delta N}$	1.3	1.3
$\frac{\Delta(rev_{90}-rev_{10})}{\Delta N}$	1.1	1.1
Percent difference in output (High vs low output state)	0.07	0.07
Non-target moments		
$\frac{\Delta rev_{50}}{\Delta N}$	1.4	0.6
$\frac{\Delta std(rev)}{\Delta N}$	0.37	0.33
$\frac{\partial E[size]}{\partial lnrev} \times \frac{1}{size}$	-0.12% $\sim$ -0.01%	-0.04%

Note: Extension refers to the extended model including habit formation and persistent monetary shocks presented in section 4.4. $\left(\frac{(\epsilon-1)^2 var(price)}{var(revenue)}\right)$ is the ratio of the contribution of the variance of price to the variance of revenue. $\left(\frac{\partial prob}{\partial lnrev} \times \frac{1}{prob}\right)$ is the change in the probability of adjustment as a percentage of the average probability of adjustment. $\frac{\Delta \overline{rev}}{\Delta N}$ is the percent change in mean of the log revenue distribution over the percent change in total labor supply. $\frac{\Delta std(rev)}{\Delta N}$ is the percent change in the standard deviation of the log distribution over the percent change in total labor supply. $\frac{\Delta(rev_{90}-rev_{10})}{\Delta N}$ is the percent change in the spread of the log revenue distribution over the percent change in the total labor supply. $\frac{\Delta rev_{50}}{\Delta N}$ is the percent change in the median of the log revenue distribution over the percent change in total labor supply. $\left(\frac{\partial E[size]}{\partial lnrev} \times \frac{1}{size}\right)$ is the change in the change in the expected size of adjustment as a percentage of the average size of adjustment.



(a) Probability of price adjustment

(b) Size of price adjustment

Figure A.2: Revenue and Price Setting: Extended model

Note: In panel (a) we first group each product-month observation by revenue into percentile bins and compute the average revenue and the probability of adjustment for each bin for both the data and the model. We then compute the log deviation of this revenue from mean of the data and model respectively. We compute the deviation of the probability of adjustment from the mean probability and divide by the mean probability to compute the percent deviation from mean probability of adjustment. We plot, for each percentile bin, the log deviation from mean revenue and the percent deviation from mean probability of adjustment. The solid line represents the calibration target, with a slope of 0.52% between log revenue and probability of adjustment. In panel (b) we group each product-month observation by revenue into percentile bins and compute the average revenue and the average size of adjustment for each bin. Then, we compute the log deviation of this revenue from mean revenue of the data and model respectively. We normalize the size of adjustment by computing the deviation of the size of adjustment of each bin from the mean size of adjustment in the data and the model, and then divide by the mean size to compute the percent deviation from mean size of adjustment. We plot, for each percentile bin, the log deviation from mean revenue and the percent deviation from mean size of adjustment. The two lines in panel (b) show the range of the strength of the relationship between size of adjustment and product revenue that we find in our data. The solid line represents the maximum value (in absolute terms) and the dashed line represents the minimum value.

Table A.6: Parameter values

Parameter	Extension	Description
β	$0.96^{1/12}$	discount factor
ϵ	3	elasticity of demand
α	0.085	rate of product exit
b	0.21 (0.7%)	menu cost (percentage of steady state revenue)
$1-\theta_z$	0.50	probability of productivity shock
σ_z	0.10	standard deviation of productivity shock
σ_ι	0.12	standard deviation of demand shock
g	0.17%	monthly growth rate of nominal GDP
σ_m	0.29%	standard deviation of nominal GDP
r_{ij}	0.05	probability of transition to different state
$\bar{\phi}$	0.005	magnitude of aggregate preference shock
ψ_0	0.010	slope parameter in exogenous demand
ψ_1	6.0	intercept parameter in exogenous demand

Note: Extension refers to the extended model including habit formation and persistent monetary shocks presented in section 3.4.

the data and the orange triangles show the relationship in the model. The solid line represents the calibration target, with a slope of 0.52% between log revenue and probability of adjustment. Panel (b) shows the relationship between size of adjustment and log revenue. The green dots show the relationship between revenue and the size of adjustment in the data and the orange triangles show the relationship in the model. The two lines in panel (b) show the range of the strength of the relationship between size of adjustment and product revenue that we find in our data. The solid line represents the maximum value (in absolute terms) and the dashed line represents the minimum value. While the relationship between the size of adjustment and revenue is weaker in the extended model compared to the data, there is still a negative relationship. Furthermore, the magnitude of the negative relationship corresponds approximately to the slope of the dashed line, which represents the lowest value of the strength of the relationship we find in our empirical section, which we get when we control for store and UPC.

Figure A.3 shows the impulse response of the cross-sectional frequency of adjustment and Table A.7 shows the average log revenue and average adjustment probability during the impulse response

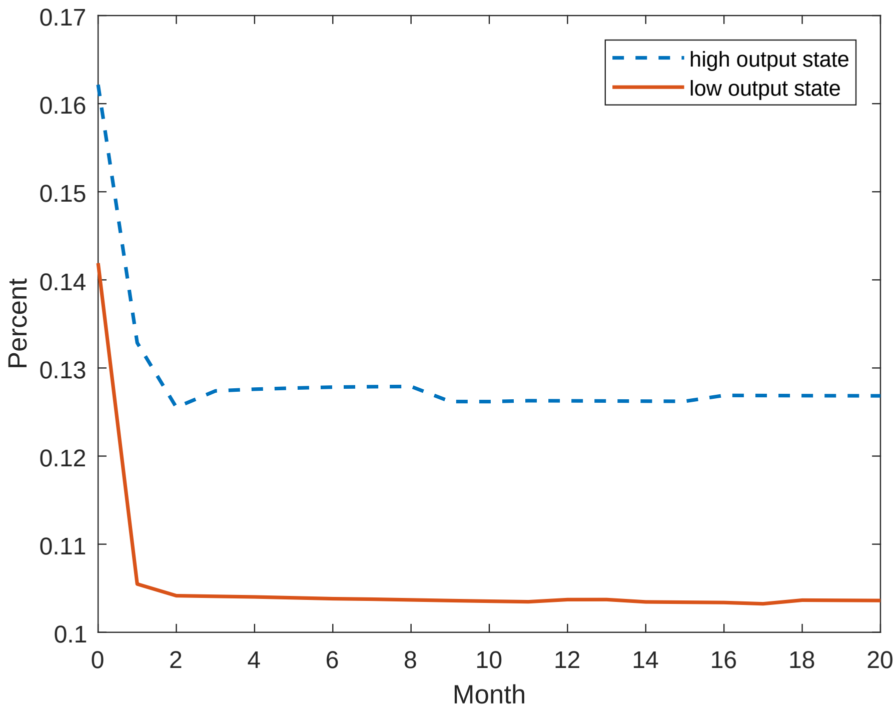


Figure A.3: Impulse response function of the cross-sectional frequency of price adjustment to a one standard deviation monetary shock in low versus high output states.

period by quantiles of revenue.

A.5 Robustness: State-Dependence

We show results that include potential controls for shocks that may effect both the state of the economy and the growth rate of industrial production. We include the utilization adjusted TFP shocks from Basu et al. (2006) and Fernald (2014). As these data are quarterly we include the 3-month through 36-month lags of the quarterly data as controls. In addition we control for persistence in the systematic component of monetary policy by including lags of changes in the Federal funds rate. We add the first 12 lags of the changes to our main specification. The impulse responses given in Figures A.4 A.5, A.6, and A.7 show that our results are robust to these controls. The error

Table A.7: Revenue and adjustment probability by quantile (Model)

	Adj. probability (%)			Revenue (%)
Quantile	High	Low	Diff	Diff
1	7.66	7.78	-0.12	7.42
2	10.29	9.65	0.64	6.95
3	13.94	9.08	4.86	3.98
4	15.10	10.01	5.10	9.46
5	17.08	16.03	1.05	15.69

Note: The table shows the average log revenue and average adjustment probability by quantile of revenue in the baseline model. Column labeled High shows the results for the high output state and the column labeled Low show the result for the low output state. Columns labeled Diff show the difference between the high and low output states.

bands, however, grow due to the loss of degrees of freedom as we include more variables in the regression.

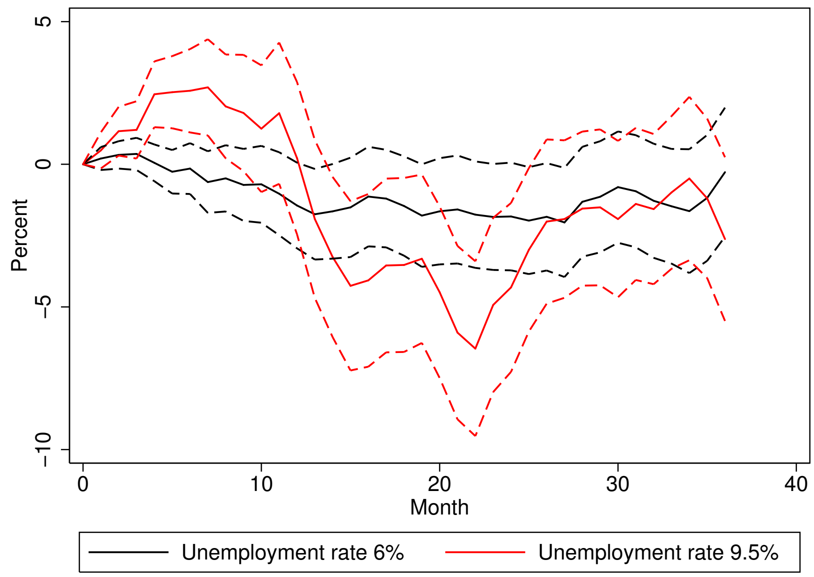
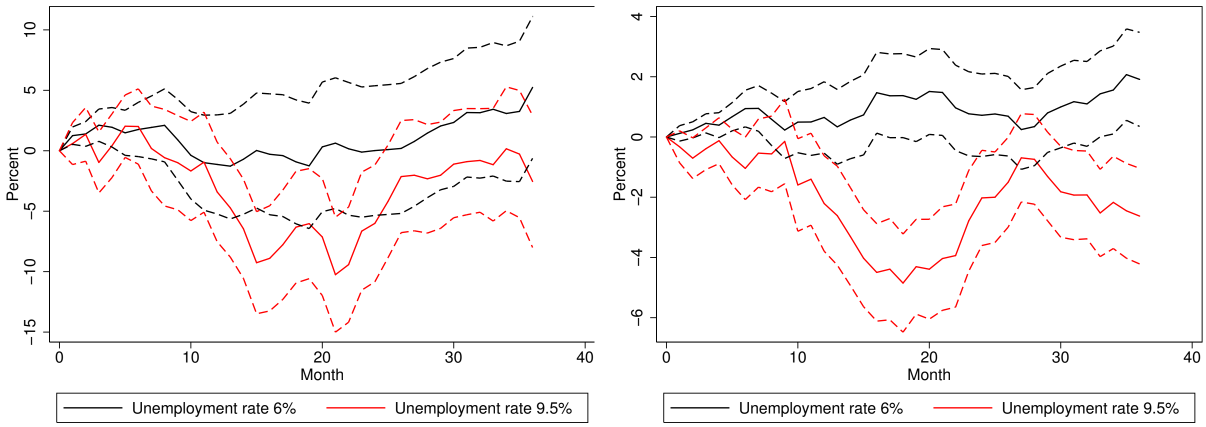


Figure A.4: TFP shocks (Industrial Production)

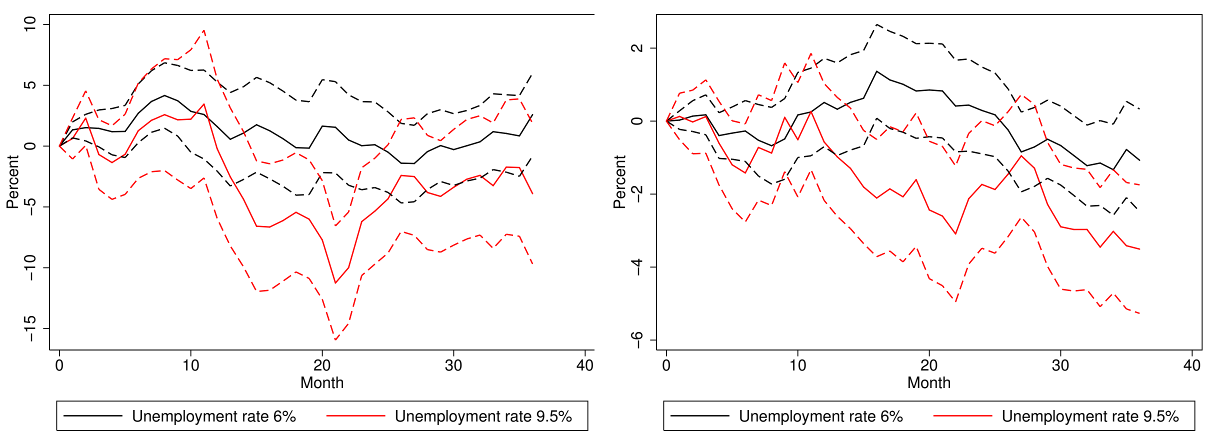
Figure A.5: Persistent Federal Funds rate changes (Industrial Production)



(a) Durable consumption

(b) Non-durable consumption

Figure A.6: Robustness check: TFP shocks



(a) Durable consumption

(b) Non-durable consumption

Figure A.7: Robustness check: Federal Funds rate changes

APPENDIX B

System Wide Runs and Financial Collapse

B.1 Proofs

Proposition 1. *All pooling equilibria in which the broker-dealers can borrow fail the intuitive criterion.*

To prove proposition 1, I first prove the following lemma. Note that the broker-dealer's expected payoff can be described by equation (2) and the household's expected payoff by equation (3).

Lemma 1. *Given d_0 and ρ, the slope of the indifference curve of the broker-dealers $\frac{dq_1}{dd_1}$ is increasing in the amount of risky assets owned by the broker-dealer.*

(proof) Let $V(d_1, q_1) = \bar{v}$ where $\bar{v}$ is some constant. By total differentiation $\frac{dV}{dd_1}dd_1 + \frac{dV}{dq_1}dq_1 = 0$. Thus the slope of the indifference curve is $\frac{dq_1}{dd_1} = -\frac{\frac{dV}{dd_1}}{\frac{dV}{dq_1}}$. I show that $\frac{\partial}{\partial a}(\frac{dq_1}{dd_1}) > 0$.

First, $\frac{\partial}{\partial a}(\frac{dq_1}{dd_1}) = -\frac{\partial}{\partial a}(\frac{q_1}{d_1}\frac{\int_{\hat{R}}^{\infty}(R-\frac{\rho}{q_1})dF(R)}{\int_{\hat{R}}^{\infty}RdF(R)}) = -\frac{1}{\int_{\hat{R}}^{\infty}RdF(R)}\frac{\partial \hat{R}}{\partial a}\{\frac{q_1}{d_1}\hat{R}[\frac{\int_{\hat{R}}^{\infty}(R-\frac{\rho}{q_1})dF(R)}{\int_{\hat{R}}^{\infty}RdF(R)} - 1] + \frac{\rho}{d_1}\}$

Note that $\{\frac{q_1}{d_1}\hat{R}[\frac{\int_{\hat{R}}^{\infty}(R-\frac{\rho}{q_1})dF(R)}{\int_{\hat{R}}^{\infty}RdF(R)} - 1] + \frac{\rho}{d_1}\}$ can be rearranged to equal $\frac{\rho}{d_1}[1 - \frac{\int_{\hat{R}}^{\infty}\hat{R}dF(R)}{\int_{\hat{R}}^{\infty}RdF(R)}] > 0$. Furthermore, $\frac{1}{\int_{\hat{R}}^{\infty}RdF(R)} > 0$ and $\frac{\partial \hat{R}}{\partial a} = -\frac{\hat{R}^2}{d_1} < 0$. $\square$

Now, let $\pi^i(d_1, q_1)$ denote the expected payoff of the household given d_1 and q_1, under the belief that the broker-dealer is of type i and $V^i(d_1, q_1)$ the expected payoff of the broker-dealer of type i. Let $C((d, q), r)$ denote a circle with radius r around the point (d, q). Then by lemma 1, for any given (d^*, q^*) and r, there exists a point $(\tilde{d}, \tilde{q}) \in C((d^*, q^*), r)$ such that $\tilde{d} < d^*$ and

(i) $V^H(d^*, q^*) < V^H(\tilde{d}, \tilde{q})$

(ii) $V^L(d^*, q^*) > V^L(\tilde{d}, \tilde{q})$

Suppose that such a point does not exist. Consider the indifference curves of the high type and low type broker-dealer that cross at (d^*, q^*). If the above point does not exist this implies that the indifference curve of the high type is above the indifference of the low type at any $d < d^*$ inside $C((d^*, q^*), r)$. However, this is a direct contradiction of lemma 1.

Now suppose that (d^*, q^*) is a pooling equilibrium offer. Then $\pi(d^*, q^*) = (1-\alpha)\pi^H(d^*, q^*) + \alpha\pi^L(d^*, q^*) \geq 0$. Since π is increasing in a,[1] $\pi^H(d^*, q^*)$ is strictly greater than $\pi^L(d^*, q^*)$ which implies that $\pi^H(d^*, q^*) > 0$. Because π is continuous in both d and q, there is an $\epsilon > 0$ small enough such that $\pi^H(d, q) > 0$ for all $(d, q) \in C((d^*, q^*), \epsilon)$.

Therefore, there is always a deviation from any pooling equilibrium offer (d^*, q^*) such that (i) and (ii) hold and the household will accept under the belief that the deviator is the high type. Thus all pooling equilibria fail the Cho-Kreps intuitive criterion $\square$.

Proposition 2. *All hybrid equilibria in which the broker-dealers can borrow fail the intuitive criterion.*

(proof) Without loss of generality, suppose that multiple actions of (d_1, q_1) are played in a

[1] $\frac{d\pi}{da} = \frac{1}{d_1}\int_0^{\hat{R}} RdF(R) > 0$

hybrid equilibrium in which the broker-dealers can borrow. Suppose (d^*, q^*) and (d^{**}, q^{**}) are two such equilibrium actions. Let $d^{**} > d^*$.

Because households always accept when they are indifferent (by assumption), the households accept all equilibrium offers with probability one. Then for households to accept, the high type broker-dealers must play all equilibrium actions with non-zero probability. Thus (d^*, q^*) and (d^{**}, q^{**}) lie on the same indifference curve of the high type broker-dealer on a $d - q$ plane.

Since, by lemma 1, the broker-dealers' indifference curves have the single-crossing property, the low type broker-dealers always prefer (d^{**}, q^{**}) over (d^*, q^*). This implies only high type broker-dealers play (d^*, q^*).

First, suppose that $q^H(d_1)$ is decreasing. Note that lemma 2 (on pg 13 and proven below) implies that the slope of the indifference curve of the high type broker-dealer is always steeper than the slope of the curve $q^H(d_1)$ at all points on $q^H(d_1)$.($|\frac{\frac{dV^H}{dd_1}}{\frac{dV^H}{dq_1}}| < |\frac{dq^H(d_1)}{dd_1}|$ for all $q_1 = q^H(d_1)$.) Otherwise $V^*(d_1)$ cannot be increasing.

Now consider the low type broker-dealer's indifference curve that crosses the point (d^{**}, q^{**}). Let $(\tilde{d}, \tilde{q})$ denote the point that this indifference curve crosses the curve $q^H(d_1)$. Because the indifference curve of the low type is steeper than the indifference curve of the high type, $\tilde{d} > d^*$.

Consider a deviation by the high type broker-dealer to the point $(d^* + 0.5 * (\tilde{d} - d^*), q^H(d^* + 0.5 * (\tilde{d} - d^*))$. If the household believes the deviator is the high type the household will accept this offer.

According to the intuitive criterion, the household should believe that the deviator is the high type. By lemma 2, the high type broker-dealer prefers $d^* + 0.5 * (\tilde{d} - d^*), q^H(d^* + 0.5 * (\tilde{d} - d^*))$ over (d^*, q^*). Because the indifference curve of the low type broker-dealer is steeper than $q^H(d_1)$, they prefer (d^{**}, q^{**}) over $d^* + 0.5 * (\tilde{d} - d^*), q^H(d^* + 0.5 * (\tilde{d} - d^*))$. The low types can always get at least $V^L(d^{**}, q^{**})$ in equilibrium. Thus the deviation is strictly dominated by the equilibrium play for the low types.

Now, suppose that $q^H(d_1)$ is increasing. Again, consider the low type broker-dealer's indifference curve that crosses the point (d^{**}, q^{**}) and let ($\tilde{d}, \tilde{q}$) denote the point that this indifference curve crosses the curve $q^H(d_1)$. Because the indifference curve of the low type is steeper than the indifference curve of the high type, $\tilde{d} > d^*$.

Again, consider a deviation by the high type broker-dealer to the point $(d^* + 0.5 * (\tilde{d} - d^*), q^H(d^* + 0.5 * (\tilde{d} - d^*))$. If the household believes the deviator is the high type the household will accept this offer. The household should believe that the deviator is the high type. The high type broker-dealer prefers $d^* + 0.5 * (\tilde{d} - d^*), q^H(d^* + 0.5 * (\tilde{d} - d^*))$ over (d^*, q^*). Because the indifference curve of the low type broker-dealer is steeper than $q^H(d_1)$, they prefer (d^{**}, q^{**}) over $d^* + 0.5 * (\tilde{d} - d^*), q^H(d^* + 0.5 * (\tilde{d} - d^*))$. The deviation is strictly dominated by the equilibrium play for the low types.

Thus all hybrid equilibria fail the intuitive criterion. $\square$

Lemma 2. *$V^*(d_1)$ is increasing in d_1 for experts with positive liquidation value, where $V^*(d_1) \equiv V(d_1, q^H(d_1))$.*

(proof) I show that $V^*(d_1)$ is increasing in d_1 indirectly by showing that $\frac{dV^*}{dd_1}(d_1) > 0$ as d_1 goes to infinity and that V^* is concave.

V^* is increasing if its first derivative is greater than zero.

$$\frac{dV^*}{dd_1} = \frac{1}{\rho} \int_{\tilde{R}}^{\infty} [(q^H(d_1) + d_1 q^{H'}(d_1))R - \rho] dF(R) \tag{B.1}$$

where $q^{H'}(d_1) = \frac{dq^H(d_1)}{dd_1}$. From equation (6) we know the expression for $q^{H'}(d_1)$. For the sake of convenience let $A \equiv \int_0^{\tilde{R}} RdF(R)$ and q_1^∞ denote the value of $q^H(d_1)$ as d_1 converges to infinity. Now, if d_1 converges to infinity $d_1 q^{H'}(d_1)$ converges to zero because $d_1 \cdot q^{H'}(d_1) = \frac{\rho a^H - d_0}{d_1}\frac{A}{A-\rho}$. Since equation (4) holds for $q^H(d_1)$, $A - \rho = -\frac{\rho}{q_1^\infty}(1 - F(\frac{\rho}{q_1^\infty}))$ which is a constant. Therefore, $\frac{dV^*}{dd_1}(d_1 \to \infty) = \int_{\frac{\rho}{q_1^\infty}}^1 [\frac{q_1^\infty}{\rho} R - 1]dF(R)$ which is greater than zero.

V^* is concave if its second derivative is less than zero.

$$\frac{d^2V^*}{dd_1^2} = \frac{1}{\rho}\int_{\tilde{R}}^{\infty}[2q^{H'}(d_1) + d_1 q^{H''}(d_1)]RdF(R) - \frac{d\tilde{R}}{dd_1}(\frac{q^H(d_1) + d_1 q^{H'}(d_1)}{\rho}\tilde{R} - 1)f(\tilde{R}) \quad \text{(B.2)}$$

$$q^{H''}(d_1) = \frac{d^2 q^H(d_1)}{dd_1^2} = -2\frac{\rho a^H - d_0}{d_1^3}\frac{A}{A-\rho} + \frac{\rho a^H - d_0}{d_1^2}\frac{-\rho}{(A-\rho)^2}\frac{dA}{dd_1} \quad \text{(B.3)}$$

where $\frac{dA}{dd_1} = \frac{d\tilde{R}}{dd_1}\tilde{R}f(\tilde{R})$ and $\frac{d\tilde{R}}{dd_1} = \frac{\rho}{q^H(d_1)d_1 + \rho a^H - d_0}(1 - \frac{d_1(q^H(d_1) + d_1 q^{H'}(d_1))}{q^H(d_1)d_1 + \rho a^H - d_0}) > 0$

From equation (23),

$$2q^{H'}(d_1) + d_1 q^{H''}(d_1) = \frac{\rho a^H - d_0}{d_1}\frac{-\rho}{(A-\rho)^2}\frac{d\tilde{R}}{dd_1}\tilde{R}f(\tilde{R}) \quad \text{(B.4)}$$

Substituting (24) into (22),

$$\frac{d^2V^*}{dd_1^2} = \frac{d\tilde{R}}{dd_1}\tilde{R}f(\tilde{R})\frac{\rho a^H - d_0}{d_1}\frac{\rho - 1}{(A-\rho)^2} < 0$$

The inequality follows because $\frac{d\tilde{R}}{dd_1} > 0$ when the liquidation value is positive. Thus since V^* is concave and increasing as d_1 goes to infinity, V^* is increasing. $\square$

Proposition 6.

Let $\alpha > 0$ converge to zero. Then, if $I(B) \leq 0$, there exists an $\eta > 0$ for all $\epsilon > 0$ such that,

if $|d_1^{CH} - d_1^{H,PI}| < \eta, \quad |q_1^{CH} - q_1^{H,PI}| < \eta$ and $(1-\alpha)\pi^H(d_1^{CH}, q_1^{CH}) + \alpha\pi^L(d_1^{CH}, q_1^{CH}) \geq 0$, then $|\rho^{CH} - \rho^{PI}| < \epsilon$, $\quad |\beta^{CH} - \beta^{PI}| < \epsilon$, $\quad |d_0^{CH} - d_0^{PI}| < \epsilon$ and $\quad |q_0^{CH} - q_0^{PI}| < \epsilon$. *Let $\{\beta^{PI}, \rho^{PI}, (d_0^{PI}, q_0^{PI}), (d_1^{H,PI}, q_1^{H,PI})\}$ denote the respective equilibrium outcomes in the perfect information economy and $\beta^{CH}, \rho^{CH}, (d_0^{CH}, q_0^{CH}), (d_1^{CH}, q_1^{CH})$ the equilibrium outcomes with a government clearing house for debt where (d_1^{CH}, q_1^{CH}) is the debt contract of the clearing house determined by the government.*

Let $\alpha > 0$ converge to zero. Then, if $I(B) \leq 0$, there exists an $\eta > 0$ for all $\epsilon > 0$ such that,

if $|d_1^{CH} - d_1^{H,PI}| < \eta$, $\quad |q_1^{CH} - q_1^{H,PI}| < \eta$ and $(1-\alpha)\pi^H(d_1^{CH}, q_1^{CH}) + \alpha\pi^L(d_1^{CH}, q_1^{CH}) \geq 0$,

then $|\rho^{CH} - \rho^{PI}| < \epsilon$, $\quad |\beta^{CH} - \beta^{PI}| < \epsilon$, $\quad |d_0^{CH} - d_0^{PI}| < \epsilon$ and $\quad |q_0^{CH} - q_0^{PI}| < \epsilon$.

(proof) First, note that $d_0^{CH} = q_0^{CH} = 1$ since $(1-\alpha)\pi^H(d_1^{CH}, q_1^{CH}) + \alpha\pi^L(d_1^{CH}, q_1^{CH}) \geq 0$ implies that all period zero debt is repaid. If $1 - d^{limit}q^H(d^{limit}) \leq B - \beta_0$, $d_0^{PI} = q_0^{PI} = 1$ as well and if $1 - d^{limit}q^H(d^{limit}) > B - \beta_0$, both d_0^{PI} and q_0^{PI} converge to zero from equations (17) and (18) as α converges to zero.

Also note that $d_1^{PI}q_1^{PI} - \xi < d_1^{CH}q_1^{CH} < d_1^{PI}q_1^{PI} + \xi$ where $\xi = \eta(d_1^{PI} + q_1^{PI}) + \eta^2$ since $d_1^{PI}q_1^{PI} - \eta(d_1^{PI} + q_1^{PI}) + \eta^2 < d_1^{CH}q_1^{CH} < d_1^{PI}q_1^{PI} + \eta(d_1^{PI} + q_1^{PI}) + \eta^2$.

(1) Suppose that $1 - d^{limit}q^H(d^{limit}) < B - \beta_0$. Then $\rho^{PI} = 1$ and $d_1^{PI}q_1^{PI} > 1 - (B - \beta_0)$. Then, there exists $\xi > 0$ small enough such that $d_1^{CH}q_1^{CH} > 1 - (B - \beta_0)$. Denote this ξ as $\hat{\xi}_1$. Then, $\rho^{CH} = 1$. If $\rho^{CH} < 1$, $\beta^{CH} < \beta_0$ and markets could not clear. Since $\rho^{CH} = 1$, $\beta^{CH} = \beta^{PI} = \beta_0$.

(2) Suppose that $1 - d^{limit}q^H(d^{limit}) \geq B - \beta_0$. Then, $\beta^{PI} = B - (d_0^{PI} - d_1^{PI}q^{PI})$ and $\beta^{CH} = B - (d_0^{CH} - d_1^{CH}q^{CH})$ implying $|\beta^{PI} - \beta^{CH}| < \xi$. Then because $I(\cdot)$ is continuous, for any $\psi > 0$, there exists a $\xi > 0$ such that $|I(\beta^{PI}) - I(\beta^{CH})| < \psi$. Then since $\rho(\frac{1}{\rho} - 1) = I(\beta)$ in this case, $|\rho^{PI} - \rho^{CH}| = |\frac{I(\beta^{PI})}{\delta + I(\beta^{PI})} - \frac{I(\beta^{CH})}{\delta + I(\beta^{CH})}| < \frac{\psi}{\delta + min(I(\beta^{PI}, I(\beta^{CH})))}$. Then there exists an η such that $\psi < \epsilon$. $\square$