A Learning-based Model of Central Bank Bailout Decision-making

Anirudh Suresh

July 2017

Abstract

In the wake of the 2008 financial crisis, conventional understanding of bailout policy grew to accept the bailout of insolvent banks due to issues of systemic risk. A model is created to simulate such decision-making on the part of a central bank in order to study commercial bank behavior and adaptation patterns and analyze relevant tradeoffs and factors in policymaking. A field of heterogeneous commercial bank agents are endowed with an endogenous reinforcement learning mechanism through which they make investment decisions and learn from the profitability of their choices. The central bank's bailout policy is represented by an objective function, and its decisions impact future commercial bank behavior. The model quantifies and produces clear visualizations of the tradeoff between moral hazard and aversion of systemic damage, and the model's relationships mirror relationships in empirical data. It also recommends bailout policy that compares favorably to empirical bailout rates. Finally, the model demonstrates how asset volatility across a financial market affects the results of a particular policy.

1 Introduction

Traditionally, in the context of bailout decision-making, the central bank (CB) has been relegated to almost exclusively considering illiquid financial institutions. Often quoted is "Bagehot's Dictum," which first appeared in Walter Bagehot's 1873 *Lombard Street: A Description of the Money Market*: "advances should be made on all good banking securities, and as largely as the public ask for them. The reason is plain. The object is to stay alarm, and nothing therefore should be done to cause alarm. But the way to cause alarm is to refuse some one who has good security to offer... No advances indeed need be made by which the Bank will ultimately lose. The amount of bad business in commercial countries is an infinitesimally small fraction of the whole business... The great majority, the majority to be protected, are the 'sound' people, the people who have good security to offer."

Bagehot's point that CBs should essentially forego lending to insolvent banks (i.e. financial firms with bad securities) was underpinned by the assumption that these firms would neither represent much of the financial activity in the market nor exhibit much potential for future success, making revival of them upon bankruptcy pointlessly costly. This argument was for many a rule of law for nearly a century and a half after its publishing. However, in the aftermath of the 2008 financial crisis, macroeconomists and policymakers realized that it was possible for an insolvent bank to be large enough in economic and financial scope to warrant resuscitation; indeed, the decision not to save such a firm could yield tremendous damage on the economy and other financial players. As a result, it has become necessary to better understand the issues behind bailout considerations for large insolvent financial firms.

This study is an effort to shed light on various issues related to bailout decision making in cases of distressed, insolvent banks. Our model aims to achieve three primary goals: 1) to quantify moral hazard in an effort to effectively address the tradeoffs inherent in this classic time-inconsistency problem; 2) to diagnose an optimal CB LOLR policy; and 3) to understand the effect of a sudden country-wide increase in investment volatility. In order to accomplish these, our model is composed of a heterogeneous set of agents representing commercial banks in the economy and a CB agent that acts as the lender of last resort (LOLR) in cases of distress among the commercial banks.

At each time step in the model, each commercial bank chooses an investment package that pays out an amount according to a distribution. Rather than being prescribed to choose certain investment packages at various time steps, the commercial banks are endowed with a reinforcement learning (RL) process that allows them to become aware of the benefit or harms of certain investment package choices. If an investment package choice has positive payout, it yields a positive reward in the RL process that provides incentive for that behavior in the future; meanwhile, if an investment package choice has negative payout, it yields a negative reward that effectively warns against future iterations of that behavior.

The CB is assumed to determine whether a distressed bank deserves bailout according to a few basic principles. First is the systemic risk issue: the CB would generally prefer not to fund an insolvent bank, as the inefficiency of that bank is likely indicative of financial mismanagement and potential to return to that state in the future. However, letting banks that are too large and interconnected with other financial firms spells danger for financial markets; such a course of action can be fuel for the implosion of the entire economy. Thus, the CB has to employ discretion in determining which banks are too interconnected to allow to fail.

The other major tenet dictating the CB's choice of action in regard to a distressed bank relates to moral hazard. In bailing out a bank, the CB demonstrates to that bank as well as the entire scope of financial firms that it is willing in at least some situations to save banks when they mess up. This incentivizes riskier behavior on the part of commercial banks in the future. For example, if the CB were to invariably bail out all distressed banks in order to ensure full protection of financial stability in the economy, all banks would learn to employ a risk-heavy strategy in order to optimize their gains over time. It would be tantamount to the notion of "go big or go home," except the underlying moral of such a situation would lack the latter option and would instead read "go big or try again." Hence, the CB must make sure to balance its protection of financial stability in the economy with discretion to discourage excessive risk-taking and waste of government funds.

The method delineated in the "Model" section quantifies these two conflicting objectives and sets out a strategy for answering the questions of interest in this study.

2 Approaches to Improve Bailout Policymaking

The topic of bailout policymaking is one of extensive research, due largely to the influence of the two aforementioned conflicting forces: moral hazard and systemic risk. Traditionally, work in this area of research focuses on one or both of these issues. Questions often addressed in the literature include the quantification of systemic risk, the issues of commitment and reputation building, and the time inconsistency of CB policy decisions due to moral hazard.

Posner and Casey (2015) serves as somewhat of a overarching, survey study of the issue and its complexities. Without getting technical, the paper provides a comprehensive discussion of bailouts, their modern political interpretation, and their real intent and features. The paper discusses the rationale behind bailing out insolvent banks, an action widely deemed poor policy since Bagehot's standards in the 19th century. The authors explain that at times, the government has to bail out insolvent banks because not doing so would present economic costs far too large: essentially the "too big to fail" and interconnectedness—symptoms of systematic risk—problems. They contend that a bailout of an insolvent bank entails a loan that may not be repaid, in contrast to a loan to a solvent but illiquid bank, which is almost certain to be paid back to the government. Posner and Casey next move to discussing moral hazard. They explain that while the U.S. government's inconsistent behavior during the 2008 financial crisis—letting Lehman fail and being harsh on AIG while being much more generous to other financial firms—was largely an effort to counter moral hazard, this inconsistency tended to favor the politically connected and went against the main intent of the government during a financial crisis—to restore confidence in the system. The authors advise that to minimize moral hazard, bailouts should be limited as much as possible to scenarios in which there is significant systemic risk and often should be accompanied by some sort of penalties on the banks in question.

Novel research in the field tries to build upon these basic understandings and broad guidelines to develop more specific or defined comprehensions of a topic or courses of action. For example, the holy grail of an optimal CB bailout policy has garnered much attention, such as from Freixas (2000), which aims to determine the optimal bailout policy for a central bank in cases where full commitment is possible as well as ones in which commitment is not possible. Freixas uses asset-based analysis to assert that the central bank should establish a policy of conditionality—a level of debt past which the central bank will not bailout the distressed bank in question. He argues that optimal policy depends largely on current state of the economy/period of the business cycle, as during a recession, a distressed bank is much more likely to be insolvent than during normal economic periods. Freixas also poses the possibility of taxing banks' uninsured debt to improve handling banks whose bankruptcies could pose tremendous systemic risk, but he acknowledges that such a policy could cause all banks to increase risk and decrease investors' incentives to closely watch banks' behavior, leading to potentially worse outcomes on the whole.

With similar intent as Freixas, Cordella and Yeyati (2003) attempts to study the negative effects of moral hazard versus the potential costs of financial instability resulting from bankruptcy. The paper proposes a model of a heterogeneous bank agent and a central bank agent, which announces a bailout policy at the initial time step. Based on the commercial bank's success or distress at subsequent time steps, the central bank agent can assign a probability of absorbing all of the liabilities of the bank and pulling it out of bankruptcy. The study determines that the optimal bailout policy should drop this probabilistic tinge, instead leading the central bank to bailout a bank with certainty if it satisfies a condition involving the parameter for probability of success of the portfolio choice at that time step. The study makes this parameter information publicly available to both agents in the model. By eliminating ambiguity from the central bank's funding decision, the study argues, an optimally minimal equilibrium risk level can be reached. The study also posits that the presence of a bailout scheme both instills some moral hazard—by creating some incentive for risky behavior—and increases the bank's probability of survival—thereby raising the stakes of failure and encouraging safer investment strategy; the authors conclude that the latter effect often dominates the former, giving ex-ante announcement and commitment a positive effect in reducing bank risk.

In addition to the trend of theoretical modeling, data analysis has been explored as a way to elucidate and quantify the true relationship between such factors as CB harshness and moral hazard. Dam and Koetter (2010) claims to be the "first [study] to use data on actual bailouts at the bank level to study moral hazard in the banking sector." It focuses on capital preservation via insurance schemes funded by commercial banks themselves rather than much rarer government bailouts. It considers a two-player game in which the commercial bank chooses a "risk level" at each time step. Larger risk level corresponds to larger potential return but also higher probability of becoming distressed and seeking capital injection from the supervisor. The supervisor determines whether to bailout the bank based on covariate data. The study finds that an increase in bailout probability by 1% leads to a 0.2% increase in distress probability; it also reports that for high levels of expected distress, management interventions/restrictions can decrease moral hazard but that for low levels of expected distress, those same restrictions can actually increase moral hazard.

Unlike the studies above, Acharya (2009) centers not on the bailout process but rather on systemic risk in the financial system. It provides a comprehensive model based on a Nash game that shows that any two banks will reach an equilibrium to decide the optimal level of correlation of risk with one another. This optimal level of correlation is determined by two conflicting externalities upon one bank's bankruptcy: "recessionary spillover," a fallout in overall investment in the economy (negative), and decrease in technological cost due to the surviving bank's absorption of the other's "human capital" (positive). In the real world, Acharya argues that the negative externality is much greater, making positive correlation of risk optimal.

Other important works relate not so much to this study's economic questions of interest but rather the methodology and techniques employed. Lu *et al* (2016) derives an optimal monetary policy from the objective of optimizing central bank reputation building among the private sector. The authors assume open inflation targeting that reaches public awareness and acknowledge a standard approach of a new central banker to stimulate output with initially high levels of inflation. Using Bayesian updates to re-center the beliefs of the private sector at each time step, the authors create a basic learning representative agent in the private sector that has the ability to, over time, learn and adapt to various central bank approaches. The study concludes that while a full-commitment approach is not optimal under all circumstances, if the actual inflation rate rises above what was originally promised, the central banker should not accommodate that deviation but rather reverse it by committing to and delivering lower-than-average inflation.

Hemmati *et al* (2010) approaches the same inflation-output optimal monetary policy problem with learning, but it does so under the condition of heterogeneity and using reinforcement learning rather than Bayesian updates. This learning process allows for a more realistic simulation of the economy with different agents having different expectations and understandings of the action space. In the model, the central bank announces an inflation target, in response to which private agents form their expectations of actual inflation. The private sector agents are posed as seeking to minimize the errors in their expectations of inflation versus the actual inflation rate. The government is assumed to know the private sector's average expectation of inflation and is set to optimize an objective function balancing deviation from target inflation and deviation from target unemployment. Whereas the central bank is designed to always optimize that function in its actions, each private agent earns rewards for his/her actions depending on how close his/her expectations of inflation were to the actual inflation rate. The study shows through an experimental case that the private sector agents learned the Nash equilibrium and converged as a population to the Nash equilibrium. Above all else, this study demonstrated the power of endogenous learning method to best reveal agents' learning patterns and how heterogeneous, individual behavior could converge to a very reasonable aggregate pattern.

This study proposes a novel approach to addressing the types of questions Freixas, Cordella and Yeyati, and Dam and Koetter tackle. Specifically, it takes on the issue of quantifying moral hazard and systemic risk (utilizing the thinking that goes into Acharya's model) in order to develop suggestions for better bailout approaches on the part of the CB—and possibly an optimal policy. Most originally, this study employs a model based on endogenous learning à la Lu *et al* and Hemmati *et al*, and it builds upon the latter in developing a heterogeneous, reinforcement learning-based approach to model financial agents' economic behavior.

3 Model

3.1 Commercial Banks

3.1.1 Basics

N heterogeneous agents occupy the commercial bank space. At time 0, agent i has a net worth of NW_{0i}. The net worths of the banks at time step t are stored in a vector $NW_t \in \mathbb{R}^N$. At any time t, each agent chooses from one of D investment package options. Given the choices of each of the N agents at t, the payouts $\mathbf{B}_t$ are calculated from a multivariate Gaussian:

$$\mathbf{B}_t \sim N(\mu, \mathbf{\Sigma}); \mathbf{B}_t \in \mathbb{R}^N, \mu \in \mathbb{R}^N, \mathbf{\Sigma} \in \mathbb{R}^{N \times N}$$

The mean vector and variance-covariance matrix are determined by the choices of each of the N agents. Given that agent i chooses package j and that agent i_2 chooses package j_2:

$$\mu_i = c_j; \mathbf{c} \in \mathbb{R}^D; c_{j_2} \geq c_{j_1}, j_2 > j_1$$

$$\Sigma_{i,i} = v_j; \mathbf{v} \in \mathbb{R}^D; v_{j_2} > v_{j_1}, j_2 > j_1$$
$$\Sigma_{i,i_2} = \rho_{j,j_2}, i \neq i_2$$

where the variance vector $\mathbf{v} \in \mathbb{R}^D$ and the variance-covariance matrix $\rho \in \mathbb{R}^{D \times D}$. The entries of $\mathbf{c}$, $\mathbf{v}$, and ρ can be manually set. For the sake of simplicity, throughout this study, individual payouts are determined to have zero covariance, so $\rho = \mathbf{0}_{N \times N}$. The distributions of the payouts of the different investment packages are unknown to the agents; in essence, $\mathbf{c}$, $\mathbf{\Sigma}$, and the normal distribution are analogous to a latent state, while $\mathbf{B}_t$ represents the observed state. The values of the payouts are used to update the net worths of agents:

$$\mathbf{NW}_{t+1} = \mathbf{NW}_t + \mathbf{B}_t$$

If $NW_{ti} < 0$, agent i is considered to be distressed/insolvent. In this case, the bank files for bankruptcy and awaits the bailout decision of the CB. If the CB decides to bail bank i out, then it pays off the negative net worth (i.e. debt) of the bank and gives it some small positive amount z so that $NW_{t+1_i} = z$. If the bank is not bailed out, it is removed from the simulation and is replaced with a new bank, while the bankruptcy track record of the simulation is updated.

3.1.2 Systemic Risk

If bank i files for bankruptcy at t, there is a notion of systemic risk that represents the damage that bank's collapse could have on the entire financial market. Based on Acharya (2009), we will model the negative effect of bank i's collapse on bank i_2 as a negative shock to the latter's net worth (at the next time period) according to the closeness of these two banks' investments. We will use correlation of package choice over time as a proxy for the closeness of the two banks' investments over time. Hence:

$$\Delta NW_{t+1_{i_2}} \propto (\max(\tau_{i,i_2}, 0))(NW_{t i_2})$$

where τ_{i,i_2} refers to the Pearson correlation coefficient between the package choices of bank i and bank j. Though not a perfect proxy how large and sprawling a bank is, a bank with high τ values for many other banks is certainly a good indicator of how well that bank's investments match up with those of other banks (and hence of how interconnected a bank is with the rest of the financial market).

3.1.3 Learning

Agent i begins the simulation with a Q-matrix $\mathbf{Q}_i \in \mathbb{R}^{1 \times D}$. This is really just a D-vector, but conventionally in reinforcement learning, $\mathbf{Q} \in \mathbb{R}^{a \times b}$, where a refers to the number of possible states the agent can take on and b refers to the number of actions the agent has access to. In each of our Q-matrices, there is only 1 row, leading us to an important assumption:

Assumption 1 *We assume that there is only* 1 *state in the state space. The implication here is that at time* t*, an agent would not choose actions differently if conditions (e.g. net worth) were different. However, this status is somewhat misleading, as an agent with a high net worth due to a certain pattern of actions is likely to have learned a tendency toward those types of actions. On the other hand, an agent with a low net worth due to a certain action must have learned a high negative penalty against action and is unlikely to attempt it again.*

Given $\mathbf{Q}_i$ at time t, agent i chooses an investment package according to the ϵ-greedy method with a softmax probabilistic operator. ϵ-greedy is an approach to allow for full exploration of the state space and in this case is constructed by determining a probability ϵ that an action is chosen purely at random. A value for

ϵ is chosen at time 0 and decreases exponentially over time steps. If an action is not chosen at random, then the softmax operator is used in conjunction with the $\mathbf{Q}_i$ to determine the probabilities of choosing each package:

$$P(action_j \mid \mathbf{Q}_i) = \frac{\exp(\frac{\mathbf{Q}_i[j]}{\theta})}{\sum_{h=1}^{D} \exp(\frac{\mathbf{Q}_i[h]}{\theta})}$$

The softmax operator is chosen to allow some probability of actions that may have negative penalty values in the Q-matrix. As θ approaches 0, the probability of agent i choosing the action with the highest value in the Q-matrix goes to 1 (essentially pure maximization in the spirit of Q-learning). On the other hand, as θ approaches ∞, the probability of each of the actions becomes equal (essentially exploration as in ϵ-greedy).

Initialization of $\mathbf{Q}_i$ is as follows:

$$\mathbf{Q}_{i0} = \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}$$

indicating a completely naive learner at the start of the simulation. Assuming agent i chooses package j and does not become distressed at t, the reward is just the payout. The update to the Q-matrix from t to $t+1$ is as follows:

$$r_{ti} = B_{ti}$$
$$\mathbf{Q}_i[j] \leftarrow (1-\eta)\mathbf{Q}_i[j] + (\eta) r_{ti}$$

where the learning rate $\eta \in [0, 1]$. If agent i becomes distressed and files for bankruptcy, then reward is not necessarily equal to payout. If the bank is bailed out, then:

$$r_{ti} = B_{ti} + q_{bailout} = z - NW_{ti}$$

where $q_{bailout}$ refers to the quantity of the bailout necessary to get the distressed bank back to $NW_{t+1_i} = z$. This in essence is a quantified version of moral hazard popping into future decision-making of the Regardless of whether bank i is bailed out, it has filed for bankruptcy, and as a result, its case is publicly apparent. Thus, we make our second assumption:

Assumption 2 *We assume that banks typically only learn from themselves and their own actions. However, in cases of bank i being in distress, the other banks in the financial market are assumed to learn from bank i's actions and reward in addition to their own respective actions and rewards.*

Let us assume that at t bank i files for bankruptcy after choosing action j. The social learning structure for bank i_2, which chose action j_2 and did not become distressed, follows in this fashion:

$$\mathbf{Q}_{i_2}[j] \leftarrow (1-\lambda\eta)\mathbf{Q}_{i_2}[j] + (\lambda\eta)(B_{ti} + q_{bailout})$$

where the social learning factor $\lambda \geq 0$. We allow this social learning process to extend to bank i as well. If bank i goes bankrupt, $q_{bailout} = 0$. Otherwise, if it receives a bailout, $B_{ti} + q_{bailout} = z - NW_{ti}$.

3.2 Central Bank

The CB's role in this model is diluted down to that of the LOLR. The CB should balance considerations of systemic risk with those of the potential for moral hazard in its bailout decision-making. In order to test out various bailout standards/policies, one option for modeling CB decision-making would be to run the simulation with various objective functions representing differing levels of lenience to distressed banks. The form of these functions would likely be logistic, thereby evaluating in each bailout case to a probability that the CB bail out the bank in question. The objective function would likely have to incorporate the CB's approximation of systemic risk, the cost of bailout, and consideration of the potential moral hazard resulting from a bailout. The last input could alternatively be expressed as a variable that determined the relative importance of moral hazard considerations to systemic risk considerations (modulating this variable from simulation to simulation or even time step to time step would alter the CB's harshness/lenience).

One possible approach to modeling the logistic function is as follows:

$$\Lambda(C, R_S, H_M) = \frac{1}{1 + e^{\beta_0 + \beta_1 C + \beta_2 R_S + \beta_3 H_M}}$$

where C refers to the cost of bailout, R_S refers to systemic risk, and H_M refers to the moral hazard index (possibly modeled by money awarded in bailout cases). The question of how to effectively come up with reasonable values for the β coefficient would need to be answered, although it is clear that increasing cost of bailout and moral hazard should push the CB away from bailing out, whereas increasing systemic risk should push the CB toward bailing out. Accordingly, we already know that $\beta_1 \geq 0$, $\beta_2 \leq 0$, and $\beta_3 \geq 0$.

3.2.1 Proxies for Variables in Objective Function

The CB is designed to know the correlation of actions between banks. Thus, its knowledge of systemic risk is perfect, equal to the actual systemic damage

$$R_S = \sum_{i_2 \neq i} \alpha \max(\tau_{i,i_2}, 0),$$

where α is the coefficient of magnitude of systemic risk. Meanwhile, let us define the proxy for moral hazard as:

$$H_M = \frac{n_{bailout}}{n_{distress}}$$

where $n_{bailout}$ and $n_{distress}$ refer to the number of bailouts and cases of distress, respectively, over the course of the simulation.

4 Calibration, Convergence, and Parameter Choice

4.1 Parameter Choice

The free parameters of interest span the space of initial conditions (e.g. initial net worth, initial number of banks), parameters controlling payout and agent decision-making, and learning hyperparameters. These parameters, their significance in the model, and their assigned values are shown in Table 1. While some parameters are varied over the course of the study, others are set to reduce ambiguity and promote model consistency.

Parameters	
Parameter Name	Model Significance
N	Initial number of commercial banks
D	Number of investment package choices
T	Number of time steps in simulation
$\mathbf{NW_0}$	Initial net worths of banks
$\mathbf{c}$	Average payouts for investment packages
$\mathbf{v}$	Variance of payouts for investment packages
z	Small net worth that bailed out banks restart with
θ	Softmax operator parameter
β_k for $k = 0, 1, 2, 3$	Coefficients in logistic objective function of the CB
η	Learning rate
λ	Social learning parameter
ϵ	ϵ-greedy initial exploration probability
ψ	ϵ-decay parameter
α	Systemic risk coefficient

Table 1: Free Parameters

Some parameters are set to balance between aggregate sufficiency and intricacy. For example, N is typically chosen to equal 15, as that value is high enough to allow for consistent aggregate behavior from simulation to simulation, keeping parameters the same. At the same time, it is not too large as to unnecessarily complicate interpretability, undermine ease of access, and exponentially increase run times. Similarly, D is often set to 3, allowing for enough investment options for interesting analysis while maintaining intelligibility of simulation output. Other parameters, such as $\mathbf{NW_0}$ and z, are designated certain values to allow for consistency and remove unnecessary variance from simulation output. Future references to model runs will include some indication of the values of the parameters above, be it in the form of table or text.

4.2 Convergence of Q

One quick check to the model's legitimacy is to verify the convergence of the RL such that the agents learn consistently and accurately. Mathematically, Q-learning can be shown to converge to the optimal Q-function, which represents the optimal action-selection policy if given sufficient time steps and interaction with the entire state-action space. Given that this method is a modified version of Q-learning—simplified to one state but with more intricacy in the learning and game-playing process—we should expect convergence here as well. An easy way to check this is to see if the average of the banks' Q-matrices experienced convergence.

Clearly, if this game were played in the typical fashion of Q-learning, with no social learning or possibility of exit from the game or external assistance in the case of negative score/net worth, we could expect that the average Q-matrix would converge to the $\mathbf{c}$ vector, which would then hold the actual average reward. Thus, as the number of time steps $T \to \infty$, given adequate exploration of the entire action space, we would expect convergence of the average Q-matrix to $\mathbf{c}$. Mathematically,

$$\lim_{T\to\infty} \overline{\mathbf{Q}} = \mathbf{c},$$

where

$$\overline{\mathbf{Q}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Q}_{iT}.$$

However, the fact that we have social learning and the possibility of exit and CB bailout means that there will be associated rewards for actions that are not found in $\mathbf{c}$. For example, a high-risk action that tends to

Parameters	
Parameter Name	Parameter Value
N	15
D	3
T	500
$\mathbf{NW_0}$	$\mathbf{1}$
$\mathbf{c}$	$\mathbf{0}$
$\mathbf{v}$	$\begin{bmatrix}0.04 & 0.25 & 1\end{bmatrix}$
z	0.3
θ	0.1
β_k for $k = 0, 1, 2, 3$	$\begin{bmatrix}0 & 0 & 0 & 0\end{bmatrix}$
η	0.05
λ	1.0
ϵ	0
ψ	0
α	0.2

Table 2: Parameter Set 1

lead to distress often leads to negative social reward for that action, meaning that in the average Q-matrix, it will likely have a value slightly smaller than the corresponding average payout in $\mathbf{c}$. Running the simulation (with parameter values shown in Table 2) 100 times leads to

$$\overline{\mathbf{Q}} = \begin{bmatrix}-0.01061627 & -0.08231045 & -0.26194413\end{bmatrix}.$$

These parameter values essentially makes bailout by the CB in every case of bank distress a 50-50 coin toss, without regard for the cost of bailout or levels of systemic risk or moral hazard. This convergence indicates a slightly negative reward for the first two investment packages and a more negative reward for the most risky option. Of course, the values are all relatively close to 0, not straying far from the expected payouts in $\mathbf{c}$. Yet, the signs of the rewards for the different levels of risk is perhaps indicative of the high negative rewards brought about by social learning. If, meanwhile, we were to leave all the coefficients the same except λ, to which we assigned a value of 0 (full parameter values shown in Table 3), we would expect the last component of the average Q-matrix to reflect a less negative reward. Indeed, under that change,

$$\overline{\mathbf{Q}} = \begin{bmatrix}-0.00501337 & -0.02717609 & -0.08141244\end{bmatrix}.$$

Compared to the other Q-matrix, this matrix has a negative reward of much lower magnitude for the riskiest package. Using the softmax operator for each of the banks in each of the simulations, we learn that for the simulation with $\lambda = 1.0$, the expected number of banks choosing each action is

$$\mathbf{A_E} = \begin{bmatrix}8.96932458 & 4.6912665 & 1.33940892\end{bmatrix},$$

whereas for the simulation with $\lambda = 0$, it is

$$\mathbf{A_E} = \begin{bmatrix}6.33709674 & 5.24428725 & 3.41861601\end{bmatrix}.$$

The significant difference here is that in the simulation with social learning, in which banks learn from others' failures, there is a greater degree of risk aversion exhibited in greater preference for the first investment package. Meanwhile, in the simulation without social learning, banks are much more risky and tend to choose the riskiest package more often. This trend can be seen in Figure 1, which shows visually the growth of moral hazard as a result of a lack of social learning in the $\lambda = 0$ case.

Parameters	
Parameter Name	Parameter Value
N	15
D	3
T	500
$\mathbf{NW_0}$	$\mathbf{1}$
$\mathbf{c}$	$\mathbf{0}$
$\mathbf{v}$	$\begin{bmatrix} 0.04 & 0.25 & 1 \end{bmatrix}$
z	0.3
θ	0.1
β_k for $k = 0, 1, 2, 3$	$\begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$
η	0.05
λ	0
ϵ	0
ψ	0
α	0.2

Table 3: Parameter Set 2

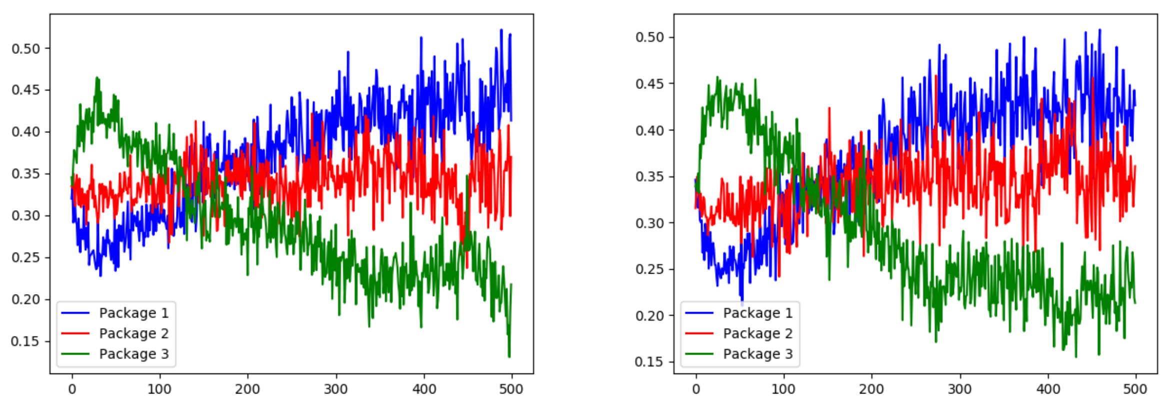

Figure 1: Distribution of investment packages over time steps. Left: $\lambda = 1.0$; right: $\lambda = 0$. Notice how the investment package 3 is chosen more often by the end in the right plot, indicating greater risk-taking and moral hazard.

Thus, the different Q-matrix and expected action convergence patterns in each of these two cases reflects a significant divergence in learning, as agents in the latter simulation learn a more lenient penalty for the riskier package. This behavior demonstrates the power of endogenous learning to reveal shifts in agent behavior depending on various factors in the model.

5 Quantification of Moral Hazard and the Pivotal Tradeoff

5.1 Moral Hazard

In order to quantify the effect of moral hazard in simulations, we adopt an approach with an analysis framework similar to the empirical analysis in Dam and Koetter (2010). Essentially, we run the simulation 50 times for each possible combination of β_k for $k = 0, 1, 2, 3$ under the constraints

$$\beta_0 \in \{-10, -9, \cdots, -1, 0\}$$
$$\beta_1 \in \{0, 0.5, \cdots, 4, 4.5\}$$
$$\beta_2 \in \{-10, -9, \cdots, -1, 0\}$$
$$\beta_3 \in \{0, 1, \cdots, 4, 5\}$$

Running the simulation 50 times for each possible combination with Parameter Set 1 (see Table 2) provides a fairly robust average of the desired output—the number of cases of distress and the bailout rate or percentage. Dam and Koetter cite a discovered relationship of a 0.2% in distress probability for a 1% increase in bailout probability. Of course, the reliance on empirical data makes causal inference challenging and limited, whereas the experimental setting provided by this model allows for a more straightforward inference of causality. Using bailout rate as a raw indicator of CB harshness, we can use the data from the 7260 combinations of β, shown in Table 4, to ascertain the existence of such a relationship.

The relationship between moral hazard—represented by the number of distress cases—and CB harshness—seen in its bailout rate—is plotted in Figure 2. The use of these proxies is especially appropriate, since the only feedback that commercial banks receive with regard to the CB's harshness comes from the policy itself—whether or not the CB decides to bail out commercial banks at the right times. Hence, bailout rate is truly the independent variable of causation with respect to commercial bank moral hazard in this case. The line of best fit can be seen plotted in red in Figure 2 and has a slope of 51.85, indicating that for a 1% increase in bailout rate, the number of distress cases rises by 51.85, on average.

The exponential model provides a fit more similar in shape to the data, and it yields us the ability to compare our results with those of Dam and Koetter, plotted in Figure 3. The functional form of the exponential model is $\hat{n}_d = 4857.655 \exp(0.006825 r_b)$, where r_b refers to the bailout rate. This model indicates that for a 1% increase in bailout rate, the number of distress cases increases by approximately $100(\exp(0.006825) - 1) \approx 0.6848\%$. This is certainly a larger growth rate than in Dam and Koetter, but it comes from a model with different assumptions about learning rate and bank behavior. The exponential form of growth itself is an important consistency with work like that of Dam and Koetter, providing a setting for calibration and further study.

5.2 Risk Tradeoff

One way to consider this question is in terms of the balance between addressing systemic risk issues and demonstrating to the field of commercial banks that failure is not an option. This tradeoff is generally

Bailout Rate (%)	Average Number of Distress Cases
7.145390	5640
8.071401	5154
...	...
100	9965

Table 4: Simulation Data on Bailout Rate-Induced Distress Inclination

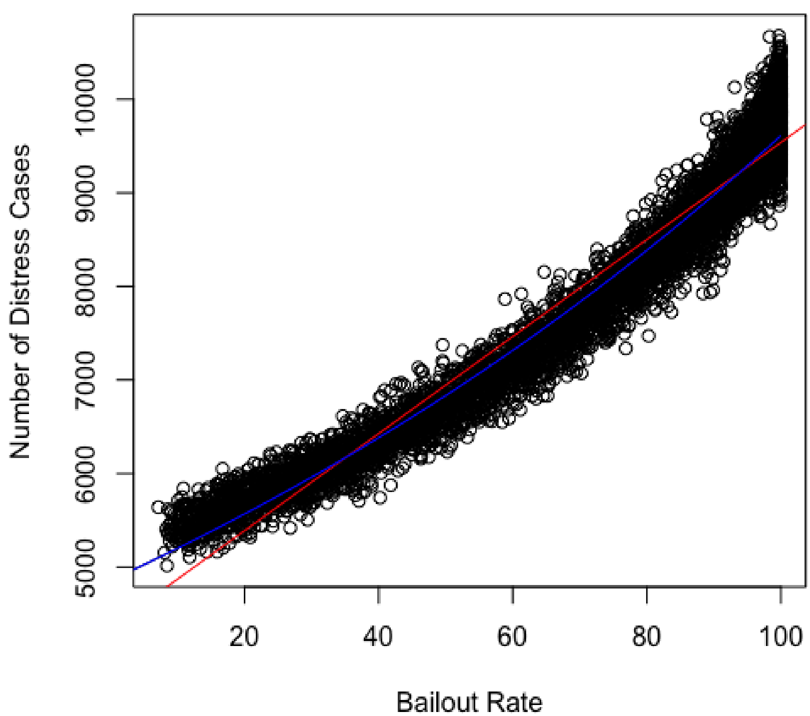

Figure 2: Number of distress cases against bailout rate. The red line of best fit has slope of 51.85, the average increase in number of distress cases per percent increase in bailout rate. The blue curve indicates the exponential model.

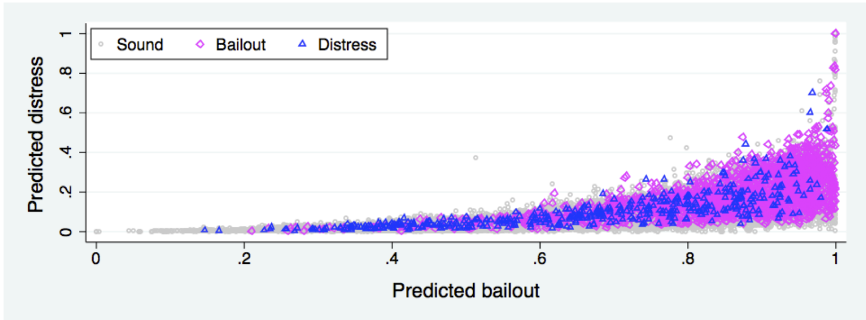

Figure 3: A plot of the relationship between distress probability and bailout probability derived from empirical data in Dam and Koetter (2010).

well-understood in the literature, yet efforts to quantify it have usually lacked robustness and replicability—empirics are not truly applicable beyond the scope of their specific context. However, from a behavioral perspective, it is possible to quantify the relationship between avoided losses due to systemic risk and incurred increases in risk via moral hazard. In order to provide such a measurement, we need to develop a probabilistic metric for investment riskiness. One possibility is

$$R_I = \frac{1}{N} \sum_{i=1}^{N} \mathbf{p}_i \bullet \sqrt{\mathbf{v}},$$

where $\mathbf{p}_i$ indicates the vector of proportions of bank i's choices of actions (that is, what proportion of bank i's actions were of each kind) and $\sqrt{\mathbf{v}}$ refers to the vector of standard deviations of the payouts of the investment packages. Measuring the relationship between averted systemic risk and incurred investment risk is now possible. Comparing R_I against averted losses due to systemic risk can shed light on the tradeoff between these two elements.

Looking at data from the β combination simulations, shown in Table 5 and Figure 4, we see the clear negative relationship between amount of systemic damage incurred and investment risk on the part of commercial banks. By nature of the bounds of $\mathbf{v}$, $0.2 \leq R_I \leq 1$, and in the data, R_I ranges from 0.4804307 to 0.5398253. The trend shown in the plot is fit with a line of best fit, which has a slope of -0.002536. This value indicates that for every increase of 1 in systemic damage, the quantity of investment risk as indicated by the metric R_I decreases by 0.002536.

Systemic Damage	Investment Risk
0	0.5348765
0	0.5311427
...	...
20.62483	0.4832411

Table 5: Simulation Data on the Tradeoff between Systemic and Investment Risks

Obviously, the fact that these metrics are less defined and ad-hoc by nature makes interpretation of this relationship more difficult than of the association between number of distress cases and bailout rate. However, both these relationships reveal the same phenomenon—the tradeoff between moral hazard and systemic damage to the financial system. Hence, they are both ways to view the same effect. One way to see this is to observe Figure 5. There is a strong negative linear relationship between the number of distress cases and the amount of systemic damage in the system. The causation pattern can be easily observed by a sort of combination of the twin effects in Figures 2 and 4: as the CB tends to bail out more often, it simultaneously reduces the amount of systemic damage in the system and increases incentive for risk-taking

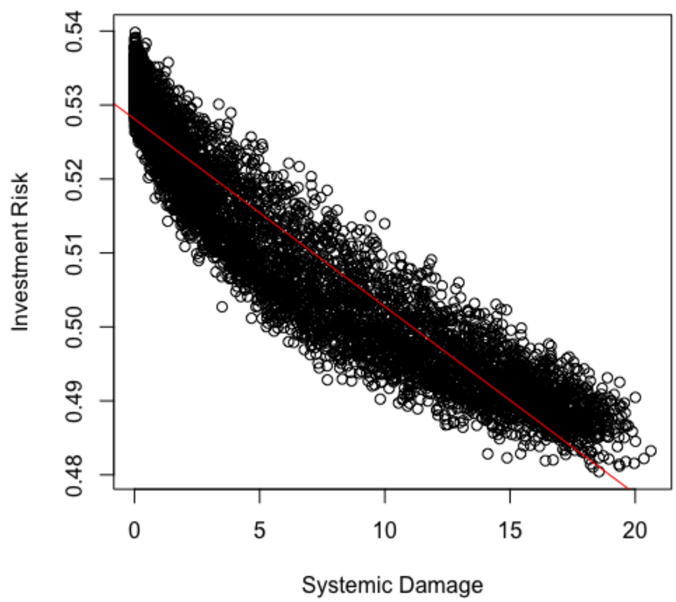

Figure 4: Investment risk against systemic damage.

by the commercial banks. Therefore, as the amount of systemic damage decreases, the number of banks taking risks and becoming distressed increases.

While the former relationship between bailout rate and number of distress cases may be easier to match up with empirical data, the latter involving the risk tradeoff offers a more direct connection to the riskiness of commercial banks and moral hazard in the system. Within a particular experimental context, it can be used to explicitly chart changes in behavior due to various initial conditions and LOLR policies.

5.3 Visualizing Moral Hazard

Several of the plots above attempt to visualize the effect of moral hazard, but other relationships can aid in this manner as well. Figure 6 illustrates how an increase in bailout rate by the CB tends to increase the riskiness of investment choices by commercial banks. Subsequently, Figure 7 shows the costs of such an increase in riskiness—as more banks tend to choose riskier actions, more of them become distressed, leading to an increase in the LOLR spending of the CB.

The specialty of running a heterogeneous, RL-based model is that each Q-matrix immediately quantifies each agent's preference for the different actions. Using this matrix, we can easily compute the probability that each agent chooses a particular investment package, and thus, the expected number of choices of each package. This is a powerful way to quantify moral hazard, as increasing incentive for risky behavior would presumably lead to higher expectation of the riskier options in the action space. Hence, just as we examined average Q-matrices and expected counts of each investment package in Section 4.2 and Figure 1, we can study the relationship between expected probabilities of investment packages and bailout rate to once again quantify and visualize moral hazard.

Figure 8 shows the relevant relationships. Interestingly, while a higher bailout rate does indeed decrease

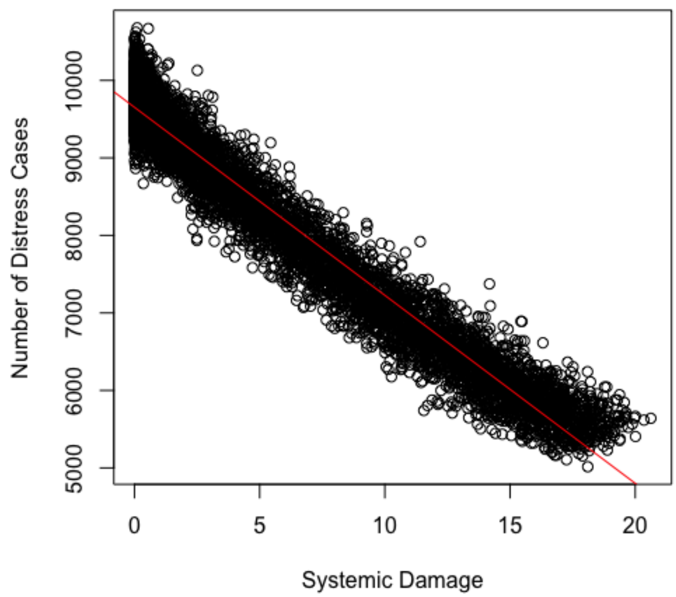

Figure 5: Number of cases against systemic damage. The line of best fit has slope of -242.6, representing the all-important tradeoff between moral hazard and systemic risk.

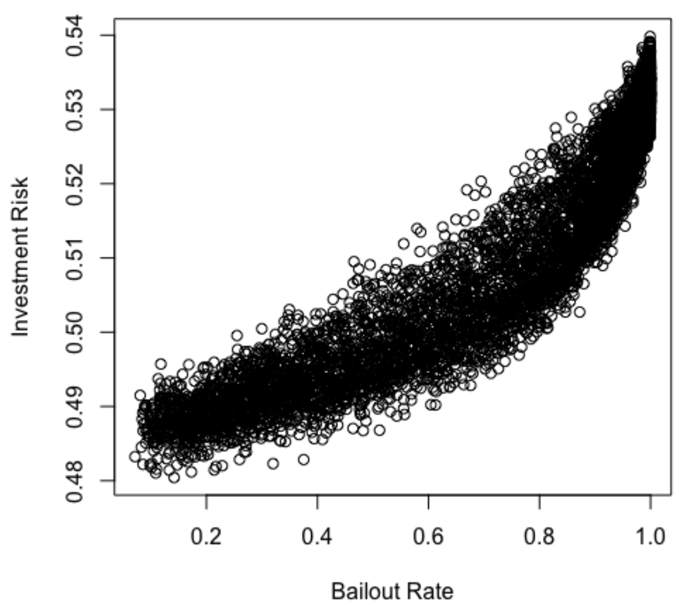

Figure 6: Investment risk against bailout rate.This plot shows the tradeoff between CB harshness and moral hazard in the form of bank investment riskiness.

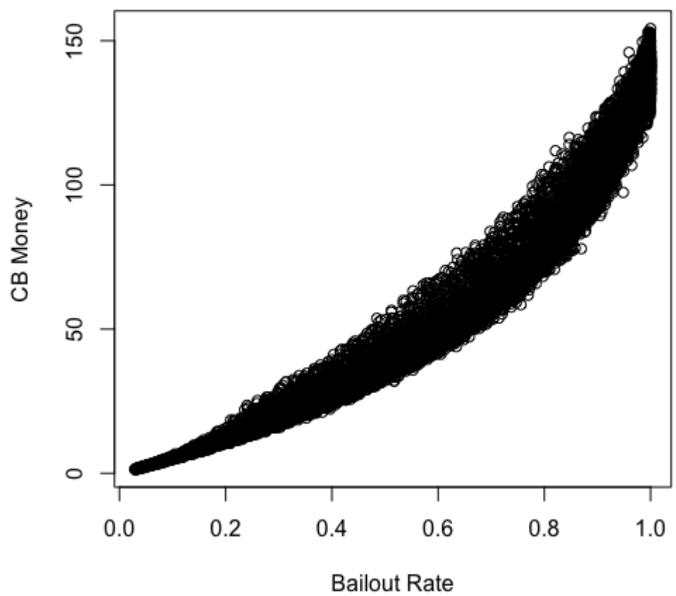

Figure 7: CB money spent on bailouts vs. bailout rate. The increase in CB bailout rate leads to increased risk-taking on the part of commercial banks, causing an increase in distress and bailout volume.

probability of the investment package of lowest risk while increasing the probability of the investment package of highest risk, it does not seemingly have much of an effect on the investment package of middle risk. However, this static distribution is likely misleading—higher bailout rate simulations likely saw some banks tend from preference of investment package 1 to investment package 2 and others move from preference of investment package 2 to investment package 3. This diffusion across the investment options ultimately results in a net increase of investment package 3, a net decrease of investment package 1, and a near-net stasis in terms of probability or count of investment package 2.

6 Optimization of Bailout Policy

Using the same approach as found the previous section, we can computationally identify an approximate optimum for the defined objective criterion. In this way, we can identify an optimal policy from the point of view of the CB. Defining some objective criterion, e.g.

$$\pi' = \arg\max_{\pi} E_0[\sum_{i=1}^{N} NW_{Ti} - \xi \sum_{bailouts} q_{bailout}]\,,$$

where π refers to CB bailout policy, π' refers to the optimal policy, T refers to a set final time step of the simulation, and ξ is the significance multiplier for government spending, we can approximately solve for the optimal policy. However, this criterion exhibits a few flaws—it places an undue amount of weight on private wealth and only measures investment risk via proxy. Moreover, setting ξ to an appropriate value would be incredibly arbitrary and ad-hoc.

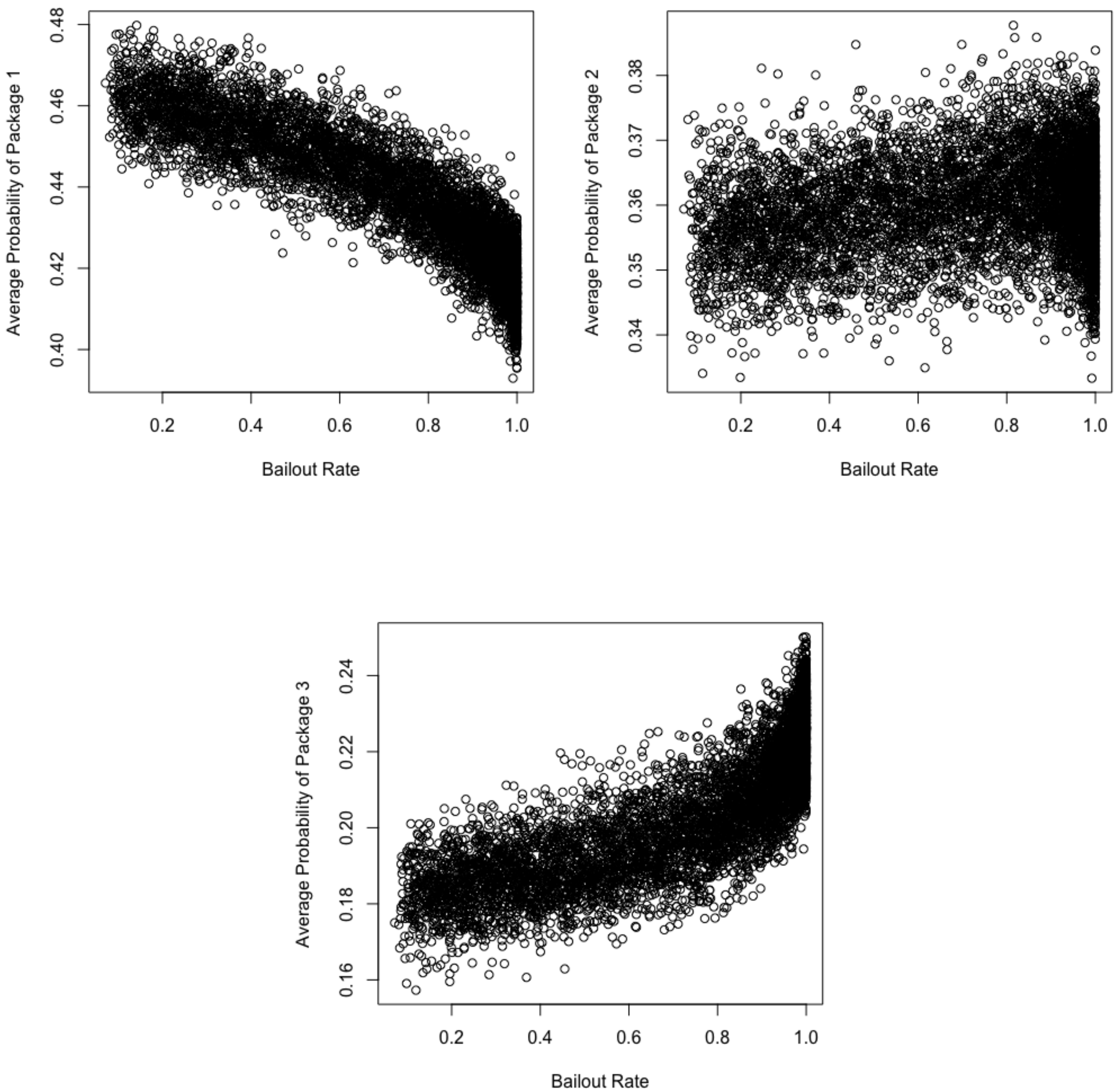

Figure 8: Distribution of probability of investment package against bailout rate. Upper left: package 1, the least risky. Upper right: package 2, the intermediately risky. Bottom: package 3, the most risky. Notice how choices of packages 1 and 3 have opposite relationships with CB harshness, while choice of package 2 is fairly steady and not very affected by bailout rate.

A better alternative is to use a criterion that truly balances investment risk and systemic risk considerations. For that purpose, it may be advisable to just adapt the investment and systemic risk metrics defined in Section 5.2. In order to combine these two metrics into one criterion that balances the conflicting effects, we need to first standardize them. After that, we can add the two together, leaving us with the objective

$$\pi' = \arg\min_{\pi} E_0[\, R_{Iz} + R_{Sz}] \,,$$

where R_{Iz} and R_{Sz} are the normalized forms of investment risk and systemic risk, respectively. Maximizing the value of this criterion over a grid search of β combinations should yield a rough estimate of the optimal policy from the perspective of this balance of moral hazard and systemic risk.

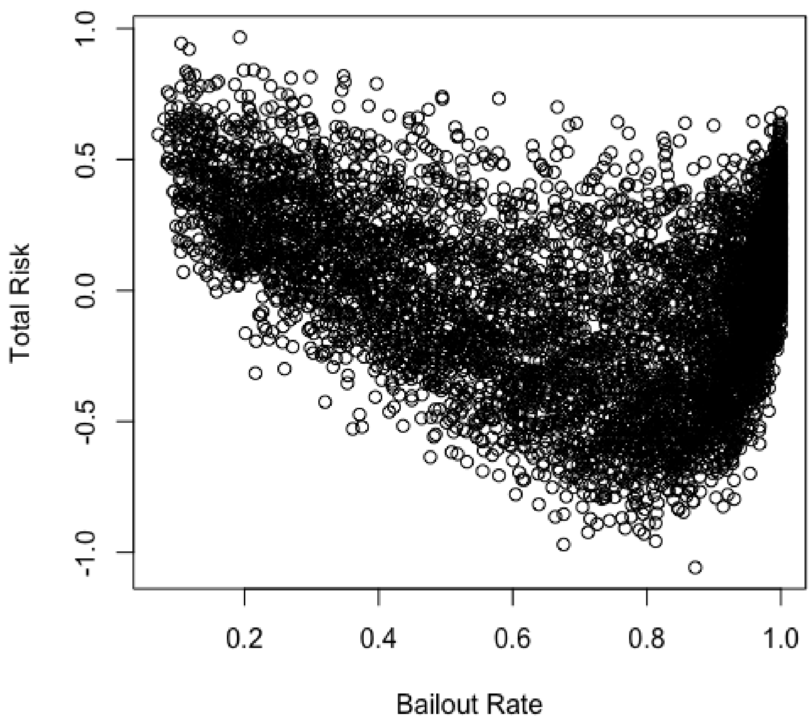

Figure 9: Total risk against bailout rate. Notice the dip in the trend between bailout rates of 60% and 90%.

The overall optimization leads to the optimal CB objective coefficients $\beta_{\pi'} = \begin{bmatrix}-6 & 4 & -7 & 1\end{bmatrix}$. This translates to a bailout rate of 87.23%. Even looking at the 50 combinations with the lowest value of total risk, we get an average bailout rate of 78.14% (with a low of 60.44% and a high of 93.02%). Even looking at local averages over all combinations run produced the results in Table 6, which support the idea that bailout rates in the 70s and 80s are optimal.

The β coefficients above are robust as well. The average β coefficients for the 50 combinations with the lowest value of total risk are $\begin{bmatrix}-5.62 & 3.73 & -6.68 & 2.24\end{bmatrix}$.

Though no external standard of comparison exists for the β coefficients, we can compare the optimal bailout rate range to conventional ones in optimal literature. Dam and Koetter do not provide an optimal rate, but it does cite that in its empirical sample data from 3500 German banks from 1994-2004, 75%-89% of banks

Bailout Rate Range	Average Criterion Value
0-10	0.55701491
10-20	0.38901921
20-30	0.26827201
30-40	0.14987271
40-50	0.01465596
50-60	−0.10424190
60-70	−0.22242445
70-80	−0.29047462
80-90	−0.27206486
90-100	0.06627877

Table 6: Average value of the total risk criterion over the different bailout rate ranges above.

in distress in a given year were bailed out. On average, 81% of banks in distress were bailed out. The fact that our optimal policy recommendation lies close to conventional behavior in the real world is indicative of its rational connection and the model's applicability.

7 Examination of Effects of Asset Volatility

7.1 Volatility's Impact on Learning and Behavior

To consider how volatility in the financial market affects outcomes of the model, such as bailout and distress volume and agent learning, we can design an experimental scenario in which we only alter measures of asset volatility—specifically, we only modulate the variance of the payouts of the investment packages. Leaving other parameters constant should allow us to see the results of the model over many runs. Hence, we use most of the parameter values shown in Table 2, with the exception that $\beta = \begin{bmatrix}-6 & 4 & -7 & 1\end{bmatrix}$ in order to study the effects of risk on CB bailout rate. We then vary the values of $\mathbf{v}$ to give us the following 3 test cases:

A $\mathbf{v} = \begin{bmatrix}0.01 & 0.0625 & 0.25\end{bmatrix}$

B $\mathbf{v} = \begin{bmatrix}0.04 & 0.25 & 1\end{bmatrix}$

C $\mathbf{v} = \begin{bmatrix}0.16 & 1 & 4\end{bmatrix}$

The results of the different test cases (averaged over 100 runs) are shown in Table 7. The only difference is that the investment risk defined in Section 5.2 is now divided by $|\sqrt{\mathbf{v}}| = \sqrt{\sum_{h=1}^{D} v_h}$ to allow for comparison among the three different test cases.

Case	Bailout Rate	Distress Count	Systemic Damage	Investment Risk	CB Money	$\mathbf{A_E}$
A	94.23%	12253	2.0316	0.4930	57.97	$\begin{bmatrix}5.32 & 5.20 & 4.48\end{bmatrix}$
B	87.59%	17478	12.6540	0.4513	92.49	$\begin{bmatrix}6.38 & 5.53 & 3.09\end{bmatrix}$
C	80.04%	22824	56.3698	0.3709	122.28	$\begin{bmatrix}9.18 & 4.42 & 1.40\end{bmatrix}$

Table 7: Results of the runs for the different test cases.

The results seem intuitive. As asset volatility increases across the financial system, cost of bailouts likely go up, leading to a decrease in bailout rate. The danger of investments also increases, leading to more distress cases. Hence, increasing asset volatility leads to an increase in systemic damage and a decrease in investment riskiness on the part of the banks.

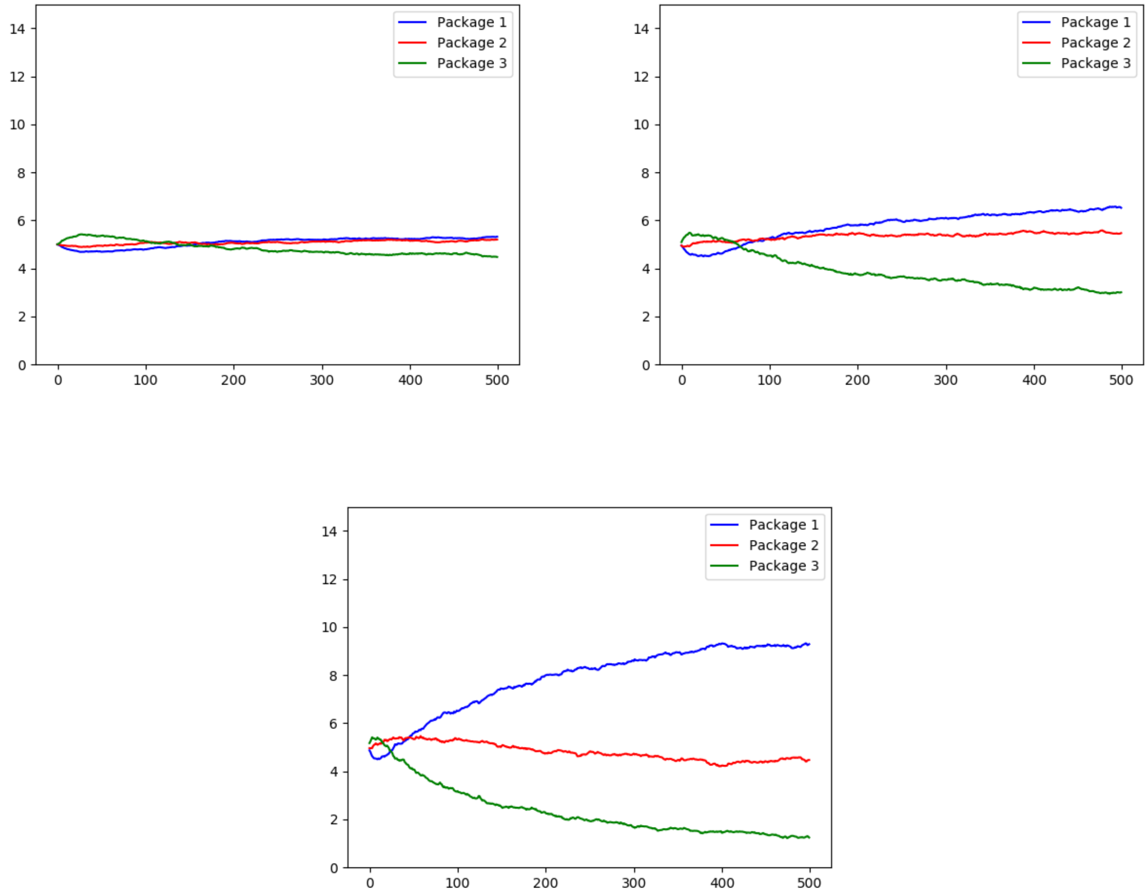

Figure 10: Progressions in expected counts of actions over time in each test case. Upper left: test case A, the least asset volatility. Upper right: test case B, intermediate asset volatility. Bottom: test case C, the greatest asset volatility.

The progressions of expected counts in the three cases is shown in Figure 10. Note how the test case with the least asset volatility displays very little difference in expected counts of each action, whereas the test case with the greatest asset volatility exhibits vast differences in expected counts of each action. When asset volatility across the financial system is very high, there are huge differences in the individual volatilities of the packages—in test case C, for instance, there is a huge difference between the standard deviations of 0.4 and 2. Meanwhile, when asset volatility is low across the system, there are more negligible differences in packages' individual volatilities—in test case A, the standard deviations of the packages range from 0.1 to 0.5. Thus, the packages in test case A are more comparable to and substitutable for each other than those in test case C.

7.2 Consideration of Industry-wide Shock to Financial Volatility

An analysis of the results and behavior in cases with different initial volatilities has been conducted above, but a volatility shock has yet to be explored. Consider a situation in which asset volatility is stable and constant for a given number of time steps but then experiences a shock at $t = t_{shock}$. How might the earlier results—bailout rate, distress count, etc.—differ, and what will the effect on learning and investment behavior be?

Let us set initial $\mathbf{v} = \begin{bmatrix} 0.01 & 0.0625 & 0.25 \end{bmatrix}$ and $t_{shock} = 500$. At t_{shock}, the asset volatilities will be multiplied by volshock^2, which represents the magnitude of the shock. Table 8 shows the results from cases with varying shock magnitudes. Figure 11 shows the progressions in expected counts of actions over time in three cases. Notice how the scenario with a decrease in asset volatility at t_{shock} experiences a convergence of the actions to approximately 5 expected counts, while the other two scenarios with increases in asset volatility see the riskiest package becoming less popular. The plots of the progressions in all 17 of the cases of volshock shown in Table 8 are shown in the Appendix section.

volshock	Bailout Rate (%)	Distress Count	Systemic Damage	Investment Risk	CB Money
0	94.59	12021	1.9870	0	57.17
0.25	94.50	13055	2.1302	0.4919	61.028
0.5	94.68	14694	2.2549	0.4908	67.89
0.75	94.45	17229	2.7853	0.4882	80.01
1	94.22	18755	3.1811	0.4842	88.64
1.25	93.42	19898	4.1968	0.4784	94.93
1.5	92.89	21434	5.0014	0.4750	103.68
1.75	91.74	23471	6.8339	0.4674	113.75
2	90.90	24681	8.9487	0.4627	119.78
2.25	90.02	25210	10.8816	0.4577	123.31
2.5	89.71	25383	12.9342	0.4541	125.12
2.75	89.21	25765	14.8085	0.4479	125.98
3	88.55	27604	18.8834	0.4418	136.16
3.25	87.64	28219	23.1195	0.4391	139.52
3.5	87.06	28688	27.2162	0.4342	141.47
3.75	86.70	29272	29.5386	0.4303	144.99
4	86.40	29921	34.9243	0.4272	148.30

Table 8: Results of the runs for different levels of volatility shock.

Table 9 shows the pre- and post-shock breakdown of the results from Table 8. One thing to note is that even for volshock = 1, the number of distress cases, systemic damage, and CB money spent are significantly lower post-shock than pre-shock. This may seem counter-intuitive, as this case represents an "identity shock," essentially a lack of a shock altogether. However, this phenomenon only demonstrates the learning process in action, as banks have learned more optimal behavior for their simulations' respective circumstances by t_{shock} and are therefore less prone to bankruptcy. Despite this, the results in the table do show some clear trends—increasing degrees of a shock to asset volatility will increase the number of distress cases but also lower the bailout rate. As a result, systemic damage will increase, while CB money spent will increase due to larger distress cases.

8 Discussion

The divergences in the convergence patterns in Section 4.2 demonstrate the role that social learning plays in commercial banks' learning and behavioral patterns. Realistically, banks are expected to learn from their peers' high-profile mistakes and failures. That model of learning skews the entire financial field away from the types of actions that led their peers toward distress in the first place—the riskiest investment packages. An important policy takeaway given this learning pattern is that CB efforts to combat moral hazard are most effective when leveled at high-profile successes and failures. Issuing policy in response to day-to-day actions

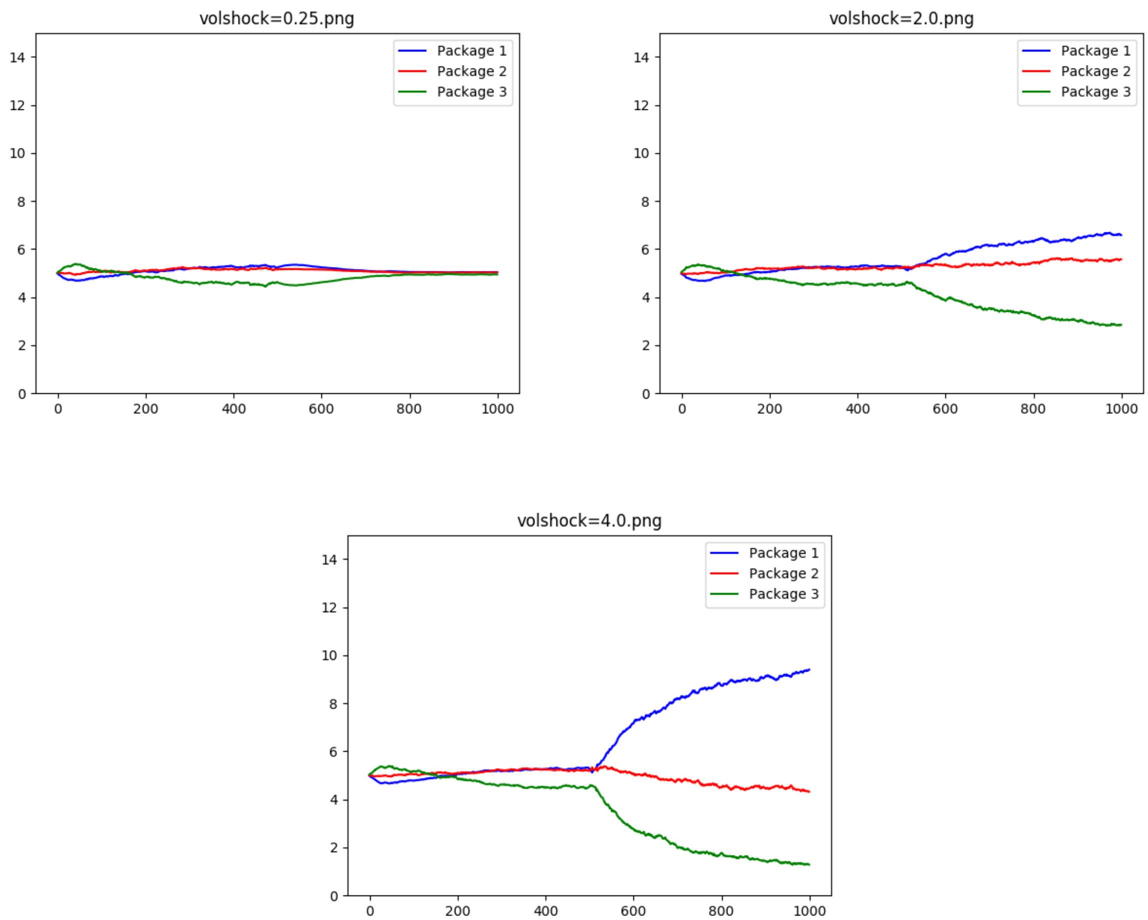

Figure 11: Progressions in expected counts of actions over time in three cases of sudden increase in asset volatility. Upper left: volshock = 0.25, a sudden decrease in asset volatility. Upper right: volshock = 2, a sudden increase in asset volatility. Bottom: volshock = 4, a sudden spike in asset volatility.

volshock	BR Pre- (%)	BR Post- (%)	n_d Pre-	n_d Post-	SD Pre-	SD Post-	CBM Pre-	CBM Post-
0	94.59	NAN	12021	0	1.9870	0	57.17	0
0.25	94.31	97.25	12219	836	2.0725	0.05761	58.10	2.93
0.5	94.39	96.32	12468	2226	2.0635	0.1914	59.06	8.83
0.75	94.20	95.18	12975	4254	2.1697	0.6156	61.33	18.68
1	94.40	93.85	12721	6034	2.0647	1.1164	60.81	27.83
1.25	94.51	91.57	12505	7393	2.0408	2.1560	59.27	35.66
1.5	94.45	90.68	12579	8855	2.0232	2.9782	59.70	43.98
1.75	94.41	88.50	12883	10588	2.0446	4.7893	61.59	52.16
2	94.38	86.85	13289	11392	2.1357	6.8130	63.21	56.57
2.25	94.64	85.35	12676	12534	1.9266	8.9550	60.25	63.06
2.5	94.68	85.04	12278	13105	2.0936	10.8406	58.47	66.65
2.75	94.65	83.93	12690	13075	1.9636	12.8450	59.99	65.99
3	94.83	83.02	12937	14667	2.0089	16.8745	61.74	74.42
3.25	94.45	82.17	12566	15653	2.0431	21.0764	60.11	79.41
3.5	94.63	81.15	12575	16113	1.9943	25.2219	59.73	81.74
3.75	94.69	80.65	12620	16652	1.9614	27.5772	59.96	85.04
4	94.51	80.26	12886	17035	2.1657	32.7585	61.60	86.70

Table 9: Pre- and post-shock breakdown of Table 8.

and results by a financial institution is unlikely to move the needle on other banks' behavior. Additionally, CBs can intelligently exploit the significant impact that social learning can have on expected future behavior of other banks. Banks' learning and behavior patterns are unlikely to be influenced if they do not see CB policy in response to a high-profile event as particularly applicable to them. In order to further its goals (e.g. to quell moral hazard in the system), a CB may do well to foster greater social learning in response to its policies—be that through more broadly applicable wording, more carefully orchestrated bailout policy, or other means. In essence, by coordinating its post-distress communication to commercial banks carefully and judiciously, a CB can channel social learning to cultivate less risky bank behavior and outcomes in the future.

The relationship between moral hazard and bailout rate described by the data and the exponential fit in Figure 2 is not necessarily definitive for real-world behavior, as it comes from a model with assumptions and defined relationships. We must investigate this value further in order to ascertain its exact accuracy and validity. However, the fact that the fit is natural and yields a growth rate similar in magnitude to the one derived in Dam and Koetter is promising for this model's ability to emulate and study real-world bailout patterns and behaviors.

A similar moral holds for the optimal policy recommendation in Section 6; we cannot definitively claim that the optimal bailout rate in the real world is a certain value based on this study's findings, but we can clearly see the resemblance of the model's optimal bailout rate to the empirical data presented in Dam and Koetter. Once again, this resemblance is promising with regard to the ability of the model to make accurate predictions and yield applicable real-world policy implications. Additionally, the average β coefficients in that ballpark of optimal bailout values indicates that at the model's optimal policy, the CB tends to pay more attention to systemic risk and less to moral hazard and cost of bailout. Further testing and validation of the results could point out whether optimal policy should indeed pay closer attention to one of these factors.

The experiments surrounding asset volatility in Section 7 reveal important information about economy-wide fluctuations in asset volatility. For one, banks' investment riskiness changes according to relative risk levels—if the economy experiences a multiplicative shock in risk, then investment risk levels diverge, and investment choices become more disparate. However, if investment risk experiences a sudden decrease economy-wide, the risk levels become more uniform relative to each other, and investment options become more interchangeable. The results in Table 7 support this finding. Yet it is not necessarily prescriptive for all such cases in the real world. The parameter set used in the asset volatility analysis gives risky investment packages lower convergence values in the final Q-matrix, meaning that risky behavior is not a generally valuable course of action in these simulations. However, there exist possible simulations in which risky behavior is often optimal, due possibly to excessively high bailout rates on the part of the CB. In such worlds, low asset volatility economy-wide would still yield relatively similar patterns of investment at each level of risk, but an increase in asset volatility would likely lead to general preference for riskier investment options. Clearly, moral hazard and CB harshness play a huge role in determining the outcome of a volatility shock.

The results in Sections 7.1 and 7.2 are similar but take on different significances. Section 7.1 can be seen as a comparison of three different markets with different, but stable, asset volatilities: an established market with low volatility, an emerging market with intermediate volatility, and a developing market with high volatility. Meanwhile, Section 7.1 poses the possibility that a single market goes through a sudden shock in which its volatility level changes dramatically. The learning patterns reveal that an instantaneous shock leads to adaptations in behavior to mirror what happens in the still cases in Section 7.1. One interesting extension might be to compare the effect of a shock that increases volatility in an established market to the results in a developing or emerging market that exhibits time-variable or Markov volatility.

8.1 Validation

The setting of the model provides a vastly superior environment for experiments, as causality can much more readily be ascertained from simulation experiments than from empirical analysis as in Dam and Koetter. If one can accurately depict real-world empirics in the context of this model, deriving causality from simulation experiments should be relatively easy. The same holds for the risk tradeoff between systemic damage and investment risk shown in Figure 4. Though the exact form of the relationship varies according to various initial conditions and circumstances, this model can serve as an excellent environment for analysis of this tradeoff when fit to empirical data. Putting the empirical data of Dam and Koetter into the context of the model, for example, one could avoid any causal inference limitations of Dam and Koetter—a semi-structural model had to be built in order to perform inference, and even then, analysis is limited to the data present. Meanwhile, this agent-based model allows for fast, controlled computer simulation experiments that can demonstrate causality and significance. Moreover, once calibration with empirical data is achieved, the model can be experimented with in multiple parameters, allowing for analyses on multiple scenarios featuring varying behavior, learning, and decision-making patterns represented by differing initial conditions and/or shocks.

The progression in investment behavior shown in Figure 11 also supports the model's validity, as the agents become aware of the instantaneous shock to asset volatility and learn to adapt their behavior. Ultimate expected counts for the volshock $= 0.25$, 4, and 16 cases, respectively, are $\begin{bmatrix} 5.07 & 5.03 & 4.90 \end{bmatrix}$, $\begin{bmatrix} 6.48 & 5.44 & 3.08 \end{bmatrix}$, and $\begin{bmatrix} 9.33 & 4.34 & 1.33 \end{bmatrix}$. The volatility shock in the last two cases leads to a scenario where the final variance vectors are the same as in test cases B and C. The final expected counts for those test cases are $\begin{bmatrix} 6.38 & 5.53 & 3.09 \end{bmatrix}$ and $\begin{bmatrix} 9.18 & 4.42 & 1.40 \end{bmatrix}$, respectively. The final expected counts are remarkably similar in both situations, demonstrating the resilience and adaptability of the agents' RL and decision-making patterns. Moreover, the results in the post- columns of Table 9 indicate the power of the learning process. In the volshock $= 1$ case, the first half of the simulation has 12721 distress cases, whereas the second half only features 7393 such cases. The amounts of systemic damage in the two halves are 2.0647 and 1.1164, respectively, and the CB spends 60.81 and 27.83 in the first and second halves, respectively. This sharp difference despite the lack of change in asset volatility is due to the learning process—as banks become more familiar with the conditions of the simulation and improve their knowledge of the true rewards of each of the investment options, they will experience fewer failures and cost the system less, a sign of the power of the learning mechanism.

9 Future Directions

The most exciting line of study using this model involves its calibration with empirical data and subsequent experimental simulation. The first step might be to compare the results of such experiments with the conclusions of well-established studies based on structural models. For example, a recommended step would be to compare the results of the semi-structural model from Dam and Koetter with those of this study's model (calibrated with the empirical data in Dam and Koetter). Seeing similar results would demonstrate the validity of the model while introducing the possibilities of behavioral and learning analysis and further experimentation with initial conditions and shocks. This model could then be used not only to corroborate the conclusions of past studies but also to build upon them in any number of directions.

Areas of theoretical analysis exist as well. Further study of the asset volatility question is in order, as this study only explores the effects of symmetric, multiplicative shocks. An interesting area of analysis would be that of nonlinear shocks. Analysis of the impact of a shock that affects different investment options to different degrees or even in inverse manners would present a more generalized form of the analysis in Section 7. To extend the question even further, one might consider the effect of fluctuating volatility—what if volatility is not constant from time step to time step, but rather changes according in some structured or stochastic way? On another note, what is the effect of a shock to asset volatility on CB optimal bailout policy and commercial bank behavior and decision-making? Understanding the answers to such questions is the first step in considering more intricate bailout situations, especially those in emerging markets where asset volatility is not stable.

Of course, the optimal policy recommendation in Section 6 demands and deserves further investigation—how robust are the model's optimal results to changes in initial conditions, size and scope of the financial field, and learning hyperparameters? Do the results from this study remain optimal with different criterions prioritizing different aspects of financial stability and health? Such analysis is important and has the potential to further bolster the model's ability to replicate real-world bailout considerations. Finally, can the endogenous learning and agent-based simulation techniques inherent in this model be applied to other areas of economic research? The topic of bank bailouts is one where studies often evade definite, quantitative results and conclusions due to the complexity of interactions and difficulty of making use of empirical knowledge. If there exist other topics whose literature mirrors this trend, they might be ripe for the introduction of the modeling techniques applied in this study.

10 Appendix

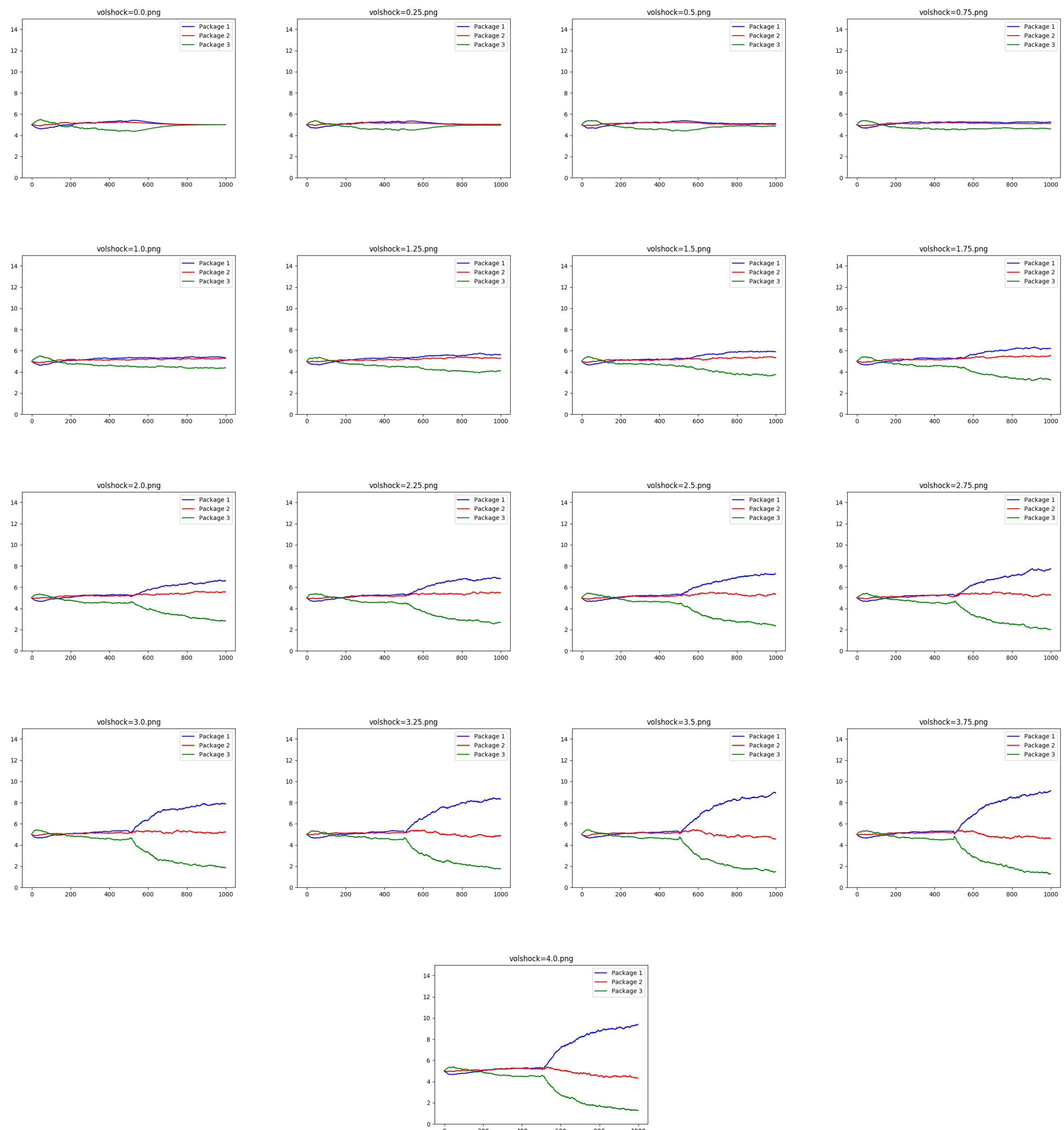

Figure 12: Progressions in expected counts of actions over time in cases of sudden increase in asset volatility. Row 1: volshock = 0, 0.25, 0.5, 0.75. Row 2: volshock = 1, 1.25, 1.5, 1.75. Row 3: volshock = 2, 2.25, 2.5, 2.75. Row 4: volshock = 3, 3.25, 3.5, 3.75. Row 5: volshock = 4.