

Article

# An agent-based approach to interbank market lending decisions and risk implications

Anqi Liu<sup>1,‡</sup>, Cheuk Yin Jeffrey Mo<sup>3,‡</sup>, Mark E. Paddrik<sup>2,‡</sup> and Steve Y. Yang<sup>3,‡,\*</sup>

<sup>1</sup> School of Mathematics, Cardiff University

<sup>2</sup> Office of Financial Research, US Department of Treasury

<sup>3</sup> School of Business, Stevens Institute of Technology

\* Correspondence: syang14@stevens.edu; Tel.: +1-201-216-3394

† Current address: 1 Castle Point Terrace, Hoboken, NJ 07030

‡ These authors contributed equally to this work.

**Abstract:** In this study, we examine the relationship of bank level lending and borrowing decisions and the risk preferences on the dynamics of the interbank lending market. We develop an agent-based model that incorporates individual bank decisions using the temporal difference reinforcement learning algorithm with empirical data of 6600 U.S. banks. The model can successfully replicate the key characteristics of interbank lending and borrowing relationships documented in the recent literature. A key finding of this study is that risk preferences at individual bank level can lead to unique interbank market structures which are suggestive of the capacity that the market responds to surprising shocks.

**Keywords:** Interbank market; Contagion risk; Multi-agent system; Reinforcement learning agents

## 0. Introduction

Prior to the financial crisis in 2008, central banks and market regulators were primarily concerned with banks' performance through a microprudential lens, examining individual banks' asset and liability portfolios to control risk. While this perspective has been commonly used to identify problematic banks, it does not consider the system-wide implications of troubled banks. During the crisis, it was evident that the interbank market suggested a heightened concern for counterparty risk, which reduced liquidity and increased the cost of financing for weaker banks [1]. Banks overall were less likely to lend liquid assets to each other. Large banks, which play a central role in this market, increased their liquidity buffers [2]. This action forced medium and small banks to seek new sources of liquidity.

As banks are highly interconnected, a macroprudential approach that incorporates interbank relationships would not only help identifying potential latent fragility within this market, but also indicate macro risk preferences and tolerances which either would support or hinder the market. Banks are autonomous decision makers to achieve individual objectives. They also respond to market changes to adapt to different risk situations that are dynamic in nature. We develop a multi-agent model in which bank agents are capable of learning using a temporal difference reinforcement learning approach based on performance goals and the changing macro environment. This model presents connections among banks from the systemic perspective, which also solves the lack of dynamics in network formation models [3–5].

We collect data from the U.S. Federal Deposit Insurance Corporation (FDIC) and build a model to represent the U.S. interbank lending system, aiming to investigate how different institutions interact, how risk preferences influence lending/borrowing decisions, and finally how these endogenous

interactions lead to converging policies of different agents. This framework reconstructs interbank exposures of autonomous banks by having each bank learn how to achieve their risk preferences through policy selection. This model captures the dynamic nature of interbank lending networks, including overnight (federal funds), short-term and long-term debts markets.

The primary contribution of this study is to develop a dynamic interbank market model with learning agents to reconstruct the banking system based on banks' behavioral patterns. From system-wide modeling perspective, we design two experiments to answer the following questions.

1. Can a multi-agent system with learning agents reconstruct the dynamics of an interbank network?
2. Does the change of agent's risk preference cause the system less prone-to contagion?

To answer these questions, we first examine the characteristics of networks generated from the learning agents. Then we design experiments to investigate the network property changes of U.S. interbank market with different bank risk profiles. We record the equilibrium interbank network of lending and borrowing that is formed under varying lending policies. We then compute and compare the network properties of the modeled interbank market under these different risk preference settings. Results show that the network degree begins to decline as bank agents continue to tighten their lending policies. Overall, our model bridges the gap in banking system contagion literature where most stress testing methods implicitly assume that banks are highly stylized with no individual optimal behavior.

The rest of the paper is organized as follows. In Section 1, we provide background literature. We present the interbank lending model framework and introduce reinforcement learning agents into the interbank framework to build a multi-agent system in Section 2. And then we discuss the data used in the study in Section 3. In Section 4, we describe the validation methods of the model and discuss the convergence of the learning process. Finally, we conduct two experiments and discuss findings accordingly in Section 5. We then conclude the paper in Section 6.

## 1. Background and related literature

### 1.1. Interbank network topology

Empirical studies about interbank network structure, especially for overnight market, have been applied to explain system-wide risk exposures. [6], [7], and [8] discovered similar network characteristics of banking system in Austria, Italy and German respectively. They suggested that interbank market presents various distinctive characteristics such as sparsity, power law degree distribution, small clustering, and small world properties. In a similar but more comprehensive study, [9] used balance sheet data to investigate Brazilian interbank network and documented that it fails to satisfy the small world conditions in terms of the arbitrary small clustering coefficient.

Other studies have explored bilateral exposures from the perspective of bank behavioral preferences. By categorizing banks into two groups, small and large banks, [10] concluded that small banks rely on large banks to borrow funds, while large banks tend to hold consistent relationships with familiar counterparties because of lower interest payments. [11] utilizes a collective dataset of an idiosyncratic and sudden shock caused by a large Indian bank. The analysis finds that interbank connections further increase the financial contagion effect and small banks tend to be more exposed to large deposit withdrawal.

### 1.2. Multi-agent systems in interbank networks

As an alternative method to the traditional general equilibrium models, multi-agent systems are better at modeling real social phenomena, such as adaptive agents' reaction to macroeconomic environment changes, and information diffusion in the interbank market [12,13]. A multi-agent system consists of autonomous agents with independent behavioral rules, connections to other agents, and a exotic environment [14]. Providing a platform for endogenously learning the network formation, multi-agent systems have been applied on network topologies and contagion risk among banks [3–5].

[15] extended this framework to a multi-layered network structure to replicate multiple types of interbank loans market. Another interesting study of an interbank trading model is to incorporate memory mechanism in agents' counterparty selection policy [16].

### 1.3. Multi-agent learning systems

In a multi-agent system, there are often two prominent types of agents: Zero-intelligence agents and learning agents. Zero-intelligence agents are described as those who “do not seek or maximize profits, and does not observe, remember, or learn” [17]. Zero-intelligence agents have been widely adapted to simulate market trading environment in which traders are represented by these agents [18].

On the other hand, learning agents can improve their performance through learning from past experiences. Reinforcement learning methods are often used in developing learning agents. The method allows agents to select an action for each state that tends to maximize the expected future value of the reinforcement signal [19]. Various learning algorithms are used in reinforcement learning and the most common approaches are temporal difference learning and Q learning. Temporal difference learning is a value prediction method that updates its estimate based on its previous estimation (bootstrap) after receiving each reinforcement signal for every action taken. Q learning is one of the most important algorithms used to optimize the expected future reward for each state-action pair. The optimal policy can be learned implicitly using Q learning.

More closely connected to our model is the work done by [5]. In this paper, [5] built a dynamic multi-agent learning system that banks within the system will experience regular deposit shocks and they have to act accordingly in order to remain solvent. They rely on the fundamental reinforcement learning concept to select their counterparties by updating the trust factor between banks.

## 2. Methodology

This section presents a multi-agent model to simulate the U.S. interbank lending market. We design an iterative dynamic framework in which banks settle debts, process payments, and update financial reports in each iteration. One iteration is equivalent to a quarter, or 3 months. During this process, reinforcement learning agent framework is utilized to determine how bank's lending and borrowing needs are satisfied based on individual banks' behaviors and preferences, and the interbank market network dynamics emerge endogenously as a result.

### 2.1. Multi-agent interbank lending framework

First the model initializes a lending network with 6600 banks linked by interbank debts. There are two types of banks and three types of debts (see Table 1). This initial network is calculated by balance sheet data in 2006 using *maximum entropy* approach. However, knowing that this algorithm solves the bilateral exposure problem by generating too many links comparing with empirical findings of the real U.S. overnight interbank market, we then allow the model to reorganize itself through the multi-agent learning process.

**Table 1.** Interbank lending network setting

Node Types	
Large Banks	Bank of America, Citibank, J.P. Morgan Chase Banks, and Wells Fargo Bank
Small Banks	Other Banks
Link Types	
Overnight Debts	Federal funds, usually expire overnight
Short-term debts	Federal securities, usually expire within 3 months
Long-term debts	Loans expire less than 1 year

In every iteration, banks first process payments to existing debts. According to expiration of the three types of interbank lending, overnight debts are paid fully, short-term debt payments and long-term debt payments are drawn from  $U(99\%, 100\%)$  and  $U(25\%, 100\%)$  respectively. We adopt Eisenburg-Noe iterative clearing vector algorithm to clear payments [20]. In this process, banks may face liquidity or solvency issue that prevents them to fulfill their obligations (see equations 1 and 2). In this case, the lenders realize write-down for debts with defaulting banks according to the Eisenburg-Noe algorithms *pro rata* loss distribution method.

Secondly banks settle debts by actively searching for counterparties. Generally, each bank sends borrowing requests to potential lenders, and lenders response with approval or rejection. Detailed decision making policies are associated with agent reinforcement learning process which will be introduced in Section 2.2. At the end of each iteration, banks process balance sheet updates to realize collected payments and latest borrowing/lending activities. Moreover, we incorporate retained earnings based on an empirically fitted distribution  $Beta(17.36, -0.1, 0.3)$ .

$$E_i(t) < A_i(t) - L_i(t) \quad (1)$$

$$C_i(t) < ON_i^p(t) + ST_i^p(t) + LT_i^p(t) \quad (2)$$

where  $ON_i^p(t)$ ,  $ST_i^p(t)$ , and  $LT_i^p(t)$  are bank  $i$ 's payments of overnight borrowing, short-term borrowing, and long-term borrowing on period  $t$ .

## 2.2. Bank lending-borrowing with reinforcement learning

In this section, we present a learning method to simulate bank's behavior in the U.S. interbank lending and borrowing market. This method utilizes reinforcement learning to guide bank decisions to lend and borrow based on the public and private information available. The rest of this section covers the reinforcement learning framework, and how bank's lending and borrowing needs are matched to form the interbank network.

Reinforcement learning is a computational technique that allows agents to learn and determine the ideal behavior based on past experience. More specifically, agents learn by receiving reinforcement signals generated through interacting with the environment. The objective of a reinforcement learning agent is to maximize its accumulative reward in the future.

There are three elements that drive the interaction between an agent and the environment: state, action, and reward. Upon evaluating the current state, the agent selects an action that maximizes their expected cumulative reward in the long run. The process of selecting an action based on a given state is typically specified in agent's policy. The environment will return a reward signal that agent can use to learn and improve their future evaluation of states. In our model, agents adopt the temporal difference approach to improve their policy in the current state based on the environment. This approach does not involve policy optimization, but instead agents have a target policy toward selecting their counterparties for new debts.

## 2.3. Banking system states

A state refers to all current information, both public and private, received by an agent to determine what to do next. The decision to select an action is based on the agent's policy and their evaluation of the current state. View on a certain state may be altered upon receiving a reinforcement signal from an action performed in the past. In other words, agent updates their prospect of how good an action is for a given state when they receive updated signal from the environment.

In the model, banks keep track of two scores for all other banks in order to pick the counterparties for new debts: a size score,  $S^{\text{size}}$ , and a relationship score,  $S^{\text{relation}}$ . The size score is calibrated through the comparison of bank size to the average size of existing counterparties (This is consistent with the preference of banks to build relationships with large banks).

$$S_{ij}^{size}(t) = \log A_j(t) - \frac{\sum_{k,kj=i} \log A_k(t-1) I_{i,k}(t-1)}{\sum_{k,kj=i} I_{i,k}(t-1)} \quad (3)$$

$$I_{i,k}(t) = \begin{cases} 1, & \text{If } i \text{ and } k \text{ are connected on period } t \\ 0, & \text{Otherwise} \end{cases}$$

150 where  $S_{ij}^{size}(t)$  is the size score of bank  $j$  evaluated by bank  $i$  on the period  $t$ ,  $A_j$  is the total assets of  
 151 bank  $j$ ,  $I_{i,k}(t)$  is a binary variable for keeping track previous debt obligations.

152 The relationship score captures the preference of continuing business with existing counterparties.  
 153 Therefore, the score will be updated upon receiving the reinforcement signal from the lending and  
 154 borrowing actions a bank makes in each iteration. In the model, temporal difference (TD) method is  
 155 adopted to update the relationship score, and the updating process is explained and illustrated by  
 156 a simplified stylized network in Section 2.5. In addition, banks also hold private information about  
 157 their own target ratios and evaluate the status of their balance sheet to help guide the lending and  
 158 borrowing policy. We assume that each bank has a settled business model that determines the fraction  
 159 of assets (or debts) to lend (or borrow) in overnight, short-term, and long-term markets. These target  
 160 ratios are set as the average on 2001-2005 balance sheet ratios.

#### 161 2.4. Actions - bank's lending and borrowing policy

162 In reinforcement learning, the objective of a policy is to map the current state to an action that  
 163 optimizes the expected sum of discounted utilities. Agents can either adopt a fixed policy to select an  
 164 action (passive learning) or learn to select an optimal action based on previous reward signals received  
 165 (active learning). In this model, banks are modeled as passive learning agents that rely on a fixed  
 166 policy to select their lending and borrowing counterparties. However, banks may learn to select better  
 167 counterparties by learning and updating the utilities of states.

168 In the model, banks first check their current balance sheet ratios and the target ratios to establish  
 169 their demand of lending and borrowing in overnight, short-term, or long-term markets. For example,  
 170 if a bank's current overnight lending ratio is lower than its target, it will look for borrowers in the  
 171 overnight market. Similarly, if a bank's current overnight borrowing ratio is lower than its target, it  
 172 will look for a lender in the overnight market. A bank will keep seeking counterparties until it reaches  
 173 all of its targets. A score-driven process is then applied to select their counterparties for initializing  
 174 new debts each period.

175 Targeting borrowing expectation, first each bank would send borrowing request to large banks  
 176 based on  $S^{relation}$  because of higher opportunity to obtain funds. As long as the targets are not met  
 177 entirely, it sends requests to highest scoring bank. If the targets are still not completed, it goes to the  
 178 highest  $S^{size}$  bank and ask for new debts. This process would continue until the requests are satisfied  
 179 perfectly or no more fund is available in the market.

180 Upon receiving a borrowing request, banks go through two key questions: 1) should they provide  
 181 new debts to borrowers? 2) how much to lend to each requester? Possessing lending preferences, each  
 182 bank with space in its overnight, short-term, or long-term lending follows a similar scoring system  
 183 as described in equation (4). Accordingly, banks go through each requests and make decisions until  
 184 lending targets are completed or no more request left to fill. However, banks may refuse to accept  
 185 requests from other banks in the market even though they have capacity. The banks utilize an S-shaped  
 186 function,  $p(S_{ij}^{total})$ , to assess the chance to lend to each specific borrower, which is a revised form but  
 187 the same assumption in [21]

$$S_{ij}^{total} = \omega S_{ij}^{relation} + (1 - \omega) S_{ij}^{size} \quad (4)$$

188 where  $s(i, j)$  is the score that borrower  $i$  assigns to lender  $j$ . It is the weighted average of relationship  
 189 score and size score of bank  $j$ . We set equal weights to these scores so that  $\omega = 0.5$ .

$$p(S_{i,j}^{\text{total}}) = \frac{1}{1 + \exp(\alpha + \beta \times S_{i,j}^{\text{total}})} \quad (5)$$

190 In this function,  $p(S_{i,j}^{\text{total}})$  presents the probability that borrower  $i$  lends to lender  $j$ , and  $\alpha$  and  $\beta$   
 191 respectively control intercept and slope.  $\alpha$  is a real number and  $\beta$  is a negative real number.

192 Following a uniform distribution, lending banks decide the fraction they want to lend out from  
 193 their lending pool. The new debt amount is set as the lower value between the one determined by  
 194 the lending bank and requested by the borrowing bank. This new debt established is observed by  
 195 both banks as a reward that their bank balance ratios move closer to their respective target ratio. In  
 summary, the flow of the bank's policy can be represented and illustrated by Figure 1.

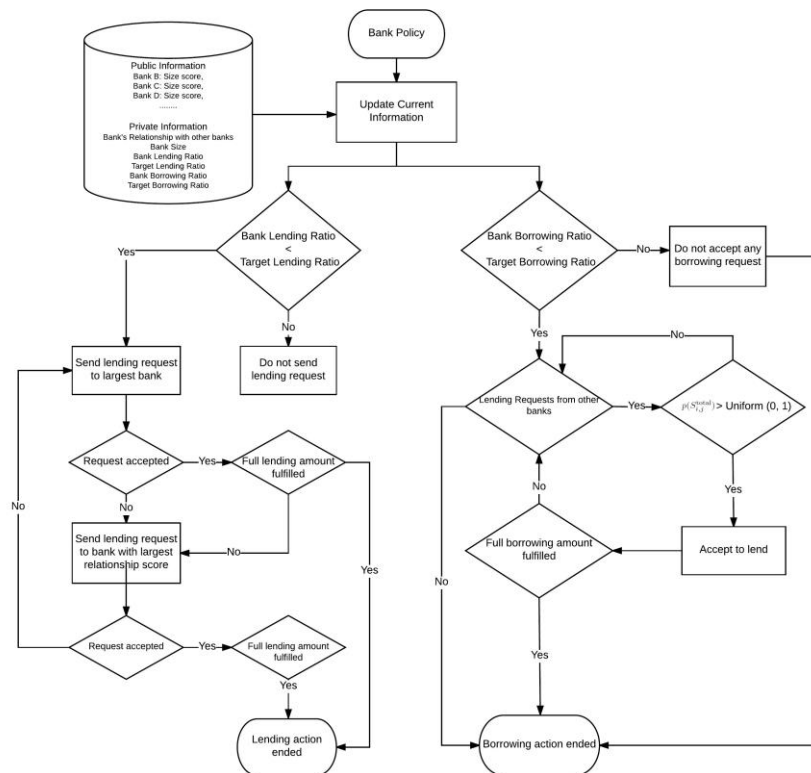


Figure 1. Bank lending and borrowing policy flowchart

196

### 197 2.5. Temporal difference learning update

The TD algorithm combines the characteristics of both the Monte Carlo method and dynamic programming. Similar to Monte Carlo method, TD algorithm is a model-free approach that is able to evaluate the value of a given state by learning directly from past experiences. In addition, it also incorporates the idea of bootstrapping from dynamic programming that the estimation of their update is based on a previous estimate. In this model, we use  $TD(0)$  method to allow banks to learn from past lending and borrowing actions.  $TD(0)$  is formulated as follows:

$$V(s_t) = V(s_t) + \alpha[r_{t+1} + \gamma V(s_{t+1}) - V(s_t)] \quad (6)$$

198 where  $V(s_t)$  is the estimate of expected sum of discounted rewards at time  $t$ ,  $r_{t+1}$  denotes the observed  
 199 reward,  $\alpha$  is the learning rate, and  $\gamma$  is a discount factor for  $V(s_{t+1})$ .

200 In the model, at the beginning of each iteration (a quarter), each bank determines its lending and  
 201 borrowing actions according to the bank policy described in previous section. Lending and borrowing  
 202 actions are decided by the size score of each bank and relationship score between two banks. Since  
 203 size score is determined by the total assets presented in the balance sheet, banks do not update the  
 204 size score using the temporal difference method. However, since the nature of relationship score is to  
 205 capture a bank's tendency to keep existing relationships, banks update their relationship score with  
 206 other banks by using the  $TD(0)$  algorithm. The updating process is described below.

207 When bank  $i$  and bank  $j$  established a new relationship, relationship score between the two banks,  
 208  $S_{i,j}^{relation}$  will be updated using equation (6). At the beginning of each quarter, TD learning is applied to  
 209 every bank to update their relationship score. In equation (6),  $V_t$  denotes the bank's evaluation of its  
 210 relationship score with all other banks.  $V_{t+1}$  is the estimation of the sum of debts established in the  
 211 future and it can be formulated as equation (7). In this model, banks assume that all debts received  
 212 from that counterparty equal to the last reward observed in the future. Therefore, equation (7) can be  
 213 expended and simplified to equation (8).

$$V(s_{t+1}) = \sum_{k=0}^{\infty} \gamma^k \times r_{t+k+2} \quad (7)$$

$$\begin{aligned} V(s_{t+1}) &= \sum_{k=0}^{\infty} \gamma^k \times D_{t+k+2} \\ &= (\gamma \times D_{t+2}) + (\gamma \times D_{t+3}) + (\gamma \times D_{t+4}) + \dots \\ &= D_{t+1}(1 + \gamma^1 + \gamma^2 + \dots) \\ &= \frac{D_{t+1}}{(1 - \gamma)} \end{aligned} \quad (8)$$

214 where  $D_t$  is the new debts between two banks in period  $t$ .

As a result, the  $TD(0)$  updating process of relationship score can be expressed as:

$$S_{i,j}^{relation}(t) = S_{i,j}^{relation}(t) + \varphi \left[ D_{t+1} + \gamma \frac{D_{t+1}}{(1 - \gamma)} - S_{i,j}^{relation}(t) \right] \quad (9)$$

We make further assumption that the learning rate  $\varphi$  is equal to  $1 - \gamma$ , resulting in:

$$\begin{aligned} S_{i,j}^{relation}(t) &= S_{i,j}^{relation}(t) + \alpha \left[ D_{t+1} + \frac{(1 - \alpha)}{(\alpha)} D_{t+1} - S_{i,j}^{relation}(t) \right] \\ &= S_{i,j}^{relation}(t) + D_{t+1} - \alpha S_{i,j}^{relation}(t) \\ &= (1 - \alpha) S_{i,j}^{relation}(t) + D_{t+1} \end{aligned} \quad (10)$$

### 215 3. Data

216 Financial reports are one of the key information sources that disclose banks' financial  
 217 fundamentals and business conditions. The Federal Reserve, FDIC, and OCC require all U.S. regulated  
 218 banks to submit quarterly reports known as *Federal Financial Institutions Examination Council Reports*  
 219 *of Condition and Income*. Those banks include national banks, state member banks and insured state  
 220 nonmember banks. Similar to regulators that rely on balance sheets to monitor banks' liquidity status  
 221 and banking system structures, we use balance sheets data from March, 2001 to December, 2014,  
 222 covering around 10,000 banks (see Table 2).

**Table 2.** Description of the bank balance sheet

Asset: A	Liability: L
Overnight lending: $ON^l$	Overnight borrowing: $ON^b$
Short-term lending: $ON^l$	Short-term lending: $ON^b$
Long-term lending: $ON^l$	Long-term borrowing: $ON^b$
Cash and balance due: C	Other liabilities: OL
Other assets: OA	<b>Equity: E</b>

223 In the model, we assume that banks target at a series of balance sheet ratios (see Table 3), associated  
 224 with banks' decisions of lending and borrowing. These targets guarantee that banks maintain their  
 225 lending-borrowing preferences for overnight, short-term, and long-term debts. Additionally, we use  
 226 the equity multiplier, the ratio of its total assets to its equity, to control a bank's expected leverage.

**Table 3.** Banks' target ratios

Equity multiplier	$E/A$
Overnight lending, borrowing ratio	$(ON^l/A)$ , $(ON^b/L)$
Short-term lending, borrowing ratio	$(ST^l/A)$ , $(ST^b/L)$
Long-term lending, borrowing ratio	$(LT^l/A)$ , $(LT^b/L)$

## 227 4. Model validation

228 We run a number of validation exercises to ensure that model can produce inter-bank markets  
 229 resembling of the real interbank markets. The model is first initialized based on 2001 financial data.  
 230 The distribution of the four selected ratios is validated according to simulation data of 20 quarters  
 231 and empirical data from 2001 to 2006. A comparison of the distributions of the four observed versus  
 232 simulated ratios shows that from a balance sheet perspective, the simulation closely resembles real  
 233 bank lending and borrowing behaviors. And then we compare the network topology features to those  
 234 described in the current literature.

### 235 4.1. Network properties validation

236 We validate overnight lending market with the U.S. federal fund market. [22] collected interbank  
 237 transactions in 2006 from Fedwire and analyzed the empirical network topology. To compare the  
 238 network structures, we conducted 100 experiments with the same number of agents (see Table 4). An  
 239 interesting finding is that the intelligent agents generate a network structure that is very close to our  
 240 previous study [21].

**Table 4.** U.S. Federal Funds Market Interbank Network Property Comparison (100 simulations)

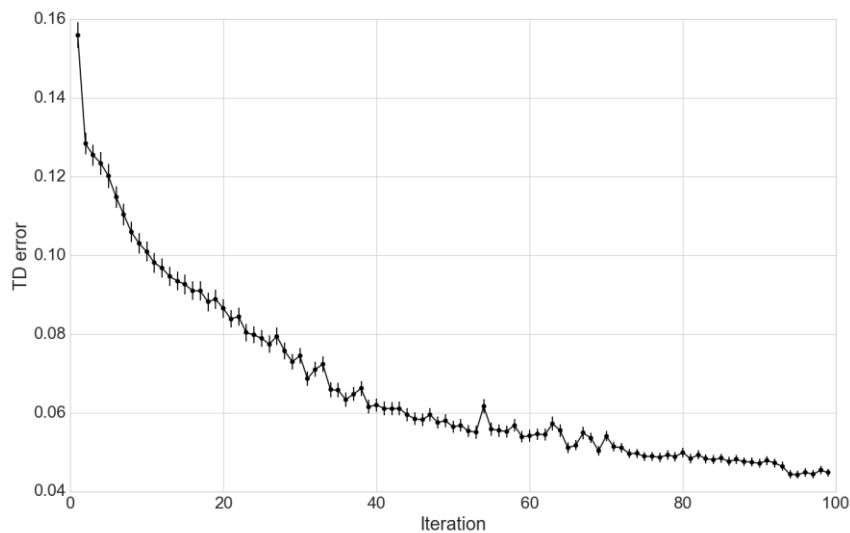
	Average In-Degree	Average Out-Degree	In-Clustering Coefficient	Out-Clustering Coefficient	Power Law
U.S. Fed. Funds Market	9.30	19.10	0.10	0.28	2.00
Model	10.39	17.14	0.03	0.21	1.94

### 241 4.2. Convergence of relationship score learning

242 Banks use temporal difference learning to update their relationship scores as they receive  
 243 reinforcement signals from borrowing and lending actions. In this experiment, we study the  
 244 effectiveness of applying  $TD(0)$  method in a complex interbank network environment. [19] provides a



245 thorough proof that  $TD(0)$  method coverages to a optimal evaluation for linear problems. However,  
246 effectiveness of applying  $TD(0)$  method in complex real-world problems remain unclear. [23]  
247 investigates the effectiveness of TD method on complex practical issues such as the game of  
248 backgammon and self-play. In this model, we validate the use of temporal difference learning by  
249 investigating the convergence of relationship score (the TD updating target) over time.



**Figure 2.** Learning convergence of TD target

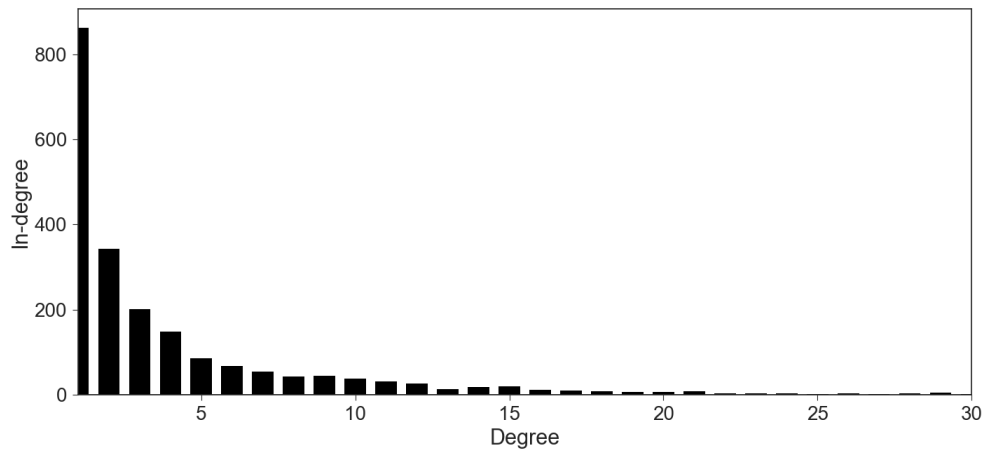
250 The experiment runs for 100 iterations of bank lending and borrowing interactions. At each  
251 iteration, the mean squared error of relationship score is recorded and computed for the entire bank  
252 population. From Figure 2, we can see that the mean squared error of relationship score converges as  
253 the system runs for 100 iterations.

## 254 5. Experiments and discussions

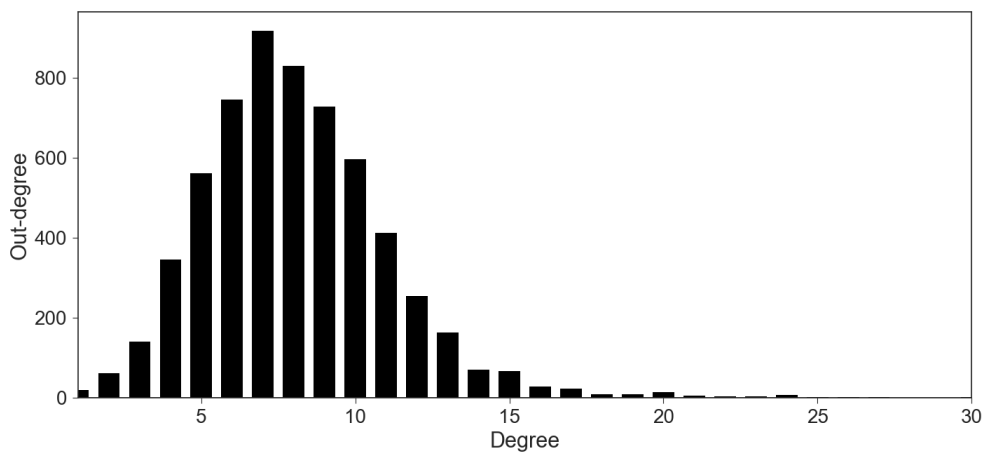
255 The multi-agent learning model simulates the banking system dynamics and provides a realistic  
256 tool to investigate problematic issues in the interbank lending system. In this section, we conduct two  
257 experiments to examine the model effects on interbank market structures due to banks' risk behaviors.

### 258 5.1. Interbank network topologies

259 Interbank network degree is asymmetric. Both in-degree and out-degree are heavily right  
260 skewed, indicating that most bank agents have few established relationships (see Figure 3 and Figure  
261 4). Moreover, we observe power-law decaying pattern in the in-degree distribution (see Figure 3).  
262 Obviously, banks prefer to maintain a lower number of lenders. Over 10% of banks only borrow from  
263 one lender. However, it is not common to minimize the number of borrowers due to consideration of  
264 risk diversification.

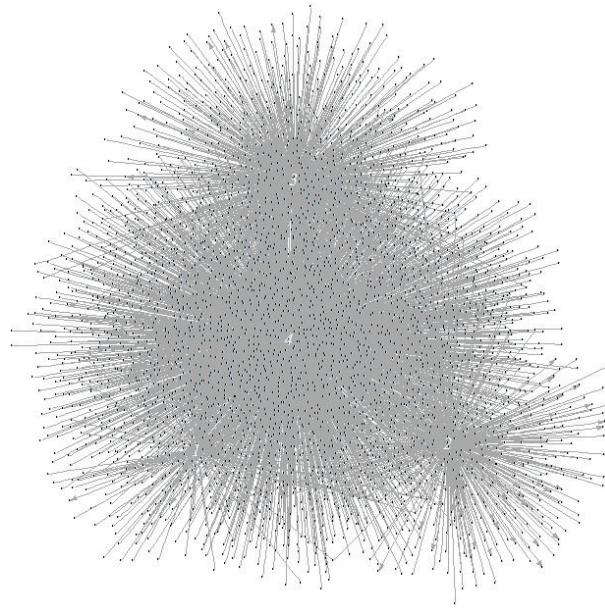


**Figure 3.** Interbank network in-degree distribution

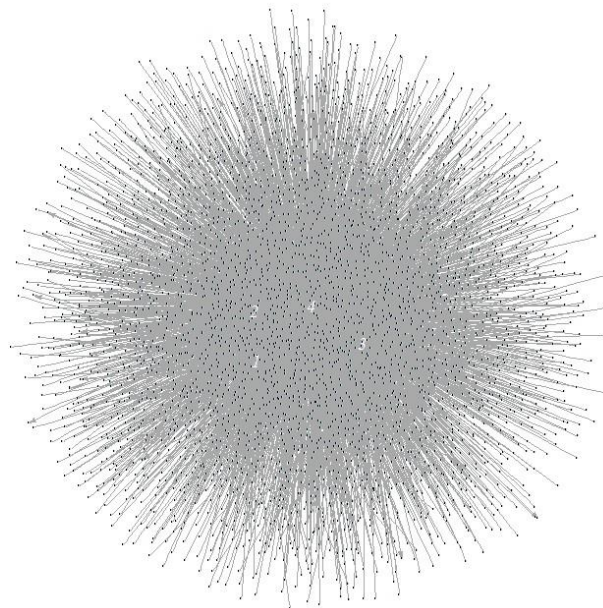


**Figure 4.** Interbank network out-degree distribution

265 In this experiment, we examine the interbank network structure after 50 iterations of the model  
 266 simulation and discuss the preliminary results that show the differences between a risk-seeking lending  
 267 policy and a risk-averse lending policy. We present two network structures that identify the lending  
 268 and borrowing relationships between banks. Figure 5 shows the interbank network when banks adopt  
 269 a risk-seeking lending policy and Figure 6 shows the network when banks become more risk averse in  
 270 accepting lending requests. In both network plots, we labeled the four large bank agents as “1,2,3,4”  
 271 because empirical studies have suggested that they tend to establish more lending and borrowing  
 272 relationships than other banks [10]. From these results we discover that large banks tend to form their  
 273 own clusters when they employ a risk seeking policy. On the other hand, this phenomenon becomes  
 274 less obvious when they adopt a more risk-averse policy.



**Figure 5.** Interbank network formation with a risk-seeking lending policy at  $\alpha = -1$



**Figure 6.** Interbank network formation with a risk-averse policy at  $\alpha = 5$

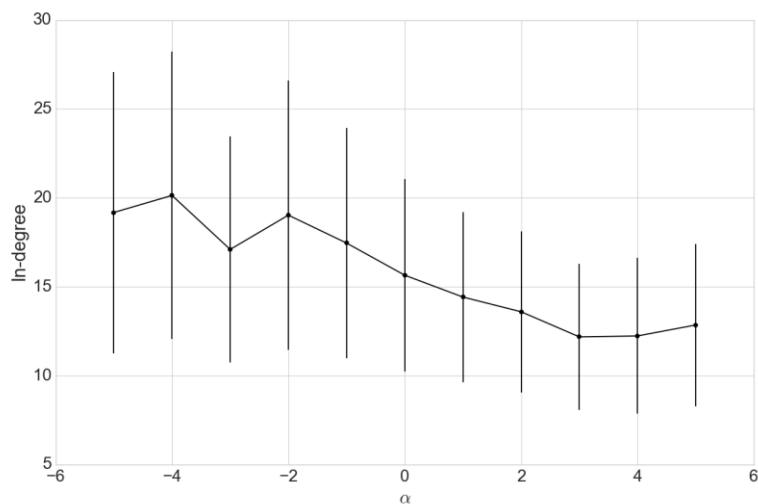
275 *5.2. Network adaptation and risk preferences*

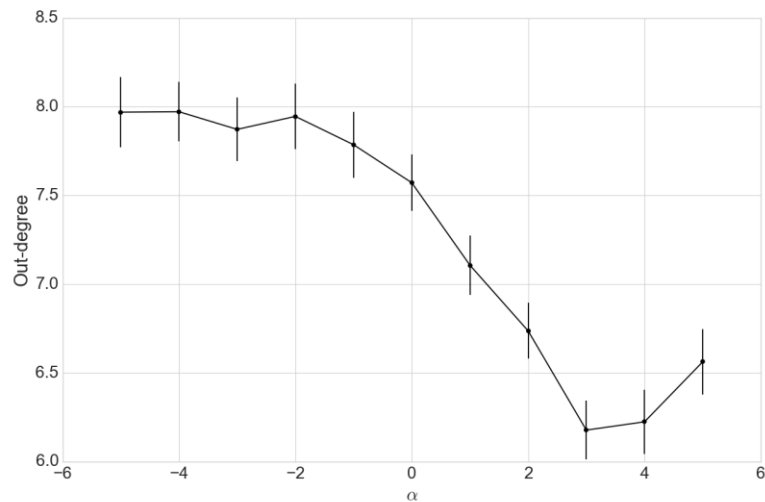
276 The bank lending policy (equation 5) shows that  $\alpha$  is the risk tolerance parameter. A small  $\alpha$   
277 represents higher risk tolerance. Put it differently, when  $\alpha$  decreases, banks risk-seeking lending policy  
278 and have higher chance to lend out. In this experiment, we study the network property adaptation  
279 by altering  $\alpha$  from  $-5$  to  $5$ . We explore changes in average degree, clustering coefficient, and average  
280 shortest path (see Table 5). These network characteristics of the interbank systems are generally inline  
281 with the empirical observations [6–8].

**Table 5.** Interbank network topology

<b>Overnight</b>	Average degree	Clustering coefficient	Power law	Average path
Risk-Seeking Policy	15.12	0.35	2.39	2.34
Risk-Averse Policy	11.51	0.19	2.42	2.66
<b>Short-term</b>	Average degree	Clustering coefficient	Power law	Average path
Risk-Seeking Policy	1.04	0.43	2.44	2.30
Risk-Averse Policy	1.04	0.53	2.29	2.21
<b>Long-term</b>	Average degree	Clustering coefficient	Power law	Average path
Risk-Seeking Policy	2.42	0.40	2.14	2.44
Risk-Averse Policy	2.42	0.57	2.15	2.28

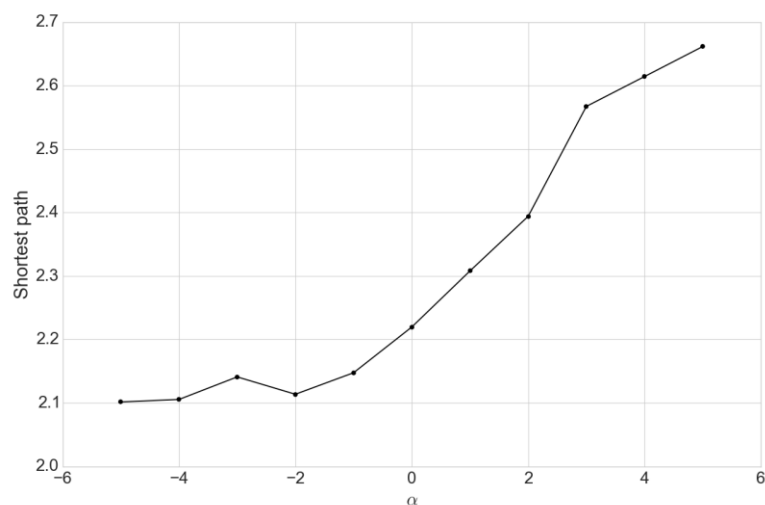
282 As the model forms a directed network in which edges are pointing from lender to borrower.  
 283 We can measure how the calibration of  $\alpha$  impacts the average number of in-degree (lending) and  
 284 out-degree (borrowing) relationships. We observe similar patterns for the two measures. The network  
 285 degree is stable when  $\alpha < -1$ , however degree decreases with increasing  $\alpha$  (see Figure 7 and Figure 8).  
 286 This result confirms that a tightened lending policy reduces the amount of debt, or counterparties, in  
 287 interbank network. Moreover, the 95% confidence interval of in-degree is much larger than that of  
 288 out-degree. These results are consistent with our finding of asymmetric distributions for lending and  
 289 borrowing. With an increasing  $\alpha$ , it becomes harder for banks to find lenders such that the in-degree is  
 290 decreasing with a tighter confidence interval (see Figure 7).

**Figure 7.** Risk preference  $\alpha$  vs. In-degree



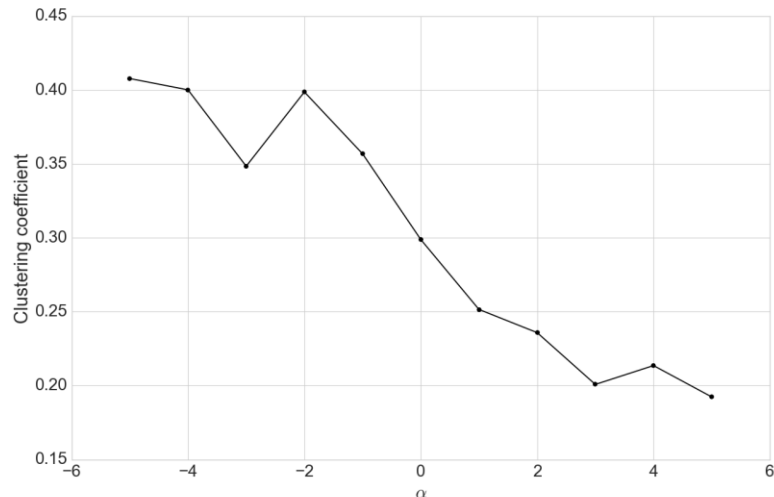
**Figure 8.** Risk preference  $\alpha$  vs. Out-degree

291 The average shortest path, as shown in Figure 9, agrees with the hypothesis that  
 292 interconnectedness of banks declining as  $\alpha$  increases. We observe longer distance among banks  
 293 when  $\alpha$  is positive, indicating that the whole system is less connected.



**Figure 9.** Risk preference  $\alpha$  vs. shortest path

294 Upon inspecting Figure 10, we find the clustering coefficient of the network becomes less clustered  
 295 as the risk aversion increases. This phenomenon shows that when banks become more averse to lending  
 296 to other banks, the network becomes less clustered thus more sparse than complete. This is validated by  
 297 also a decreased degree and increasing average shortest path, which also indicate a loosely connected  
 298 network.



**Figure 10.** Risk preference  $\alpha$  vs. clustering coefficient

## 299 6. Conclusion

300 This study proposes a multi-agent learning model to reconstruct interbank lending networks and  
 301 examine their dynamics. We incorporate a reinforcement learning framework to guide bank decisions.  
 302 Banks learn and update their choice parameters based on past interaction within the system. Our  
 303 model is able to capture the individual preferences of banks while allowing them to vary their policies  
 304 to the environmental conditions presented.

305 Two experiments are conducted in this study to investigate the network properties of the interbank  
 306 lending market as adaptive bank risk preferences are varied. First, we look at two different interbank  
 307 networks of banks adopting a risk-seeking lending policy and a tightened lending policy respectively.  
 308 We compare the network properties of the two interbank lending networks, and show that given  
 309 a certain level of risk preference, the network degree begins to decline as bank continue to tighten  
 310 their lending policies. Secondly, we examine the interconnectedness of bank agents by computing  
 311 the clustering coefficients and average shortest path amount banks. This suggests that as banks  
 312 tighten their lending policies, the network becomes sparser and thus more concentrated. As a result,  
 313 the interbank network is less at risk to concerns of contagion. Though banks are less likely to find  
 314 new counterparties, they are generally more capable of sustaining stresses, thus demonstrating how  
 315 individual choices presents beneficial overall market structure condition due to the adaptive learning  
 316 capabilities of the banks. This finding supports the general observation during the financial crisis.

317 The methodology proposed in this study offers a new perspective on how one can utilize  
 318 reinforcement learning to model bank agents. Combined with the use of financial data, we provide  
 319 guidance on how central banks and regulators may consider building more functional models  
 320 for examining the interbank systemic risks. Future research might include exploring how banks  
 321 dynamically select the optimal policy based on the observed states, and consider how monetary policy  
 322 influences competition in interbank lending.

## 323 References

- 324 1. Afonso, G.; Kovner, A.; Schoar, A. Stressed, not frozen: The federal funds market in the financial crisis. *The*  
 325 *Journal of Finance* **2011**, *66*, 1109–1139.
- 326 2. Berrospide, J.M. Bank liquidity hoarding and the financial crisis: an empirical evaluation **2012**.
- 327 3. Georg, C.P. The effect of the interbank network structure on contagion and common shocks. *Journal of*  
 328 *Banking and Finance* **2013**, *37*, 2216–2228.

- 329 4. Ladley, D. Contagion and risk-sharing on the inter-bank market. *Journal of Economic Dynamics and Control*  
330 **2013**, *37*, 1384–1400.
- 331 5. Lux, T. Emergence of a core-periphery structure in a simple dynamic model of the interbank market.  
332 *Journal of Economic Dynamics and Control* **2015**, *52*, A11–A23.
- 333 6. Boss, M.; Elsinger, H.; Summer, M.; Thurner, S. Network topology of the interbank market. *Quantitative*  
334 *Finance* **2004**, *4*, 677–684.
- 335 7. Iori, G.; Gabbi, G. A Network Analysis of the Italian Overnight Money Market. *Journal of Economic*  
336 *Dynamics and Control* **2008**, *32*, 259–278.
- 337 8. Roukny, T.; Georg, C.P.; Battiston, S. A network analysis of the evolution of the german interbank market.  
338 *Deutsche Bundesbank. Discussion Paper* **2014**.
- 339 9. Cont, R.; Moussa, A.; Santos, E.B. Network Structure and Systemic Risk in Banking Systems. In *Handbook*  
340 *on Systemic Risk*; 2013; pp. 327–368.
- 341 10. Cocco, J.F.; Gomes, F.J.; Martins, N.C. Lending relationships in the interbank market. *Journal of Financial*  
342 *Intermediation* **2009**, *18*, 24–48.
- 343 11. Iyer, R.; Peydro, J.L. Interbank contagion at work: Evidence from a natural experiment. *The Review of*  
344 *Financial Studies* **2011**, *24*, 1337–1377.
- 345 12. Gilbert, N.; Terna, P. How to build and use agent-based models in social science. *Mind & Society* **2000**,  
346 *1*, 57–72.
- 347 13. Macy, M.W.; Willer, R. From Factors to Factors: Computational Sociology and Agent-Based Modeling.  
348 *Annual Review of Sociology* **2002**, *28*, 143–166.
- 349 14. Macal, C.M.; North, M.J. Tutorial on agent-based modelling and simulation. *Journal of Simulation* **2010**,  
350 *4*, 151–162.
- 351 15. Kok, C.; Montagna, M. Multi-layered Interbank Model for Assessing Systemic Risk.
- 352 16. Iori, G.; Mantegna, R.N.; Marotta, L.; Miccichè, S.; Porter, J.; Tumminello, M. Networked relationships in  
353 the e-MID Interbank market: A trading model with memory. *Journal of Economic Dynamics and Control* **2015**,  
354 *50*, 98–116.
- 355 17. Gobe, D.; Sunder, S. Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial  
356 Substitute for Individual Rationality. *The Journal of Political Economy* **1993**, *101*, 119–137.
- 357 18. Othman, A. Zero-Intelligence Agents in Prediction Markets. *Proceedings of Seventh International Conference*  
358 *on Autonomous Agents and Multiagent Systems (AAMAS 2008)* **2008**, pp. 879–886.
- 359 19. Sutton, R.S.; Barto, A.G. Reinforcement Learning: An Introduction **1988**.
- 360 20. Eisenberg, L.; Noe, T.H. Systemic risk in financial systems. *Management Science* **2001**, *47*, 236–249.
- 361 21. Liu, A.; Paddrik, M.; Yang, S.Y.; Zhang, X. Interbank contagion: An agent-based model approach to  
362 endogenously formed networks. *Journal of Banking & Finance* **2017**.
- 363 22. Bech, M.L.; Atalay, E. The topology of the federal funds market. *Physica A: Statistical Mechanics and its*  
364 *Applications* **2010**, *389*, 5223–5246.
- 365 23. Tesauro, G. Practical Issues in Temporal Difference Learning **1992**. *277*, 257–277.