# SimCash: Multi-Agent Simulation of Strategic Liquidity Management in Payment Systems

Anonymous

December 17, 2025

**Abstract**

This paper presents SimCash, a multi-agent simulation framework for studying strategic liquidity management in real-time gross settlement (RTGS) payment systems. We employ reinforcement learning agents that adaptively adjust their intraday liquidity reserves based on observed costs and counterparty behavior.

Through three experiments, we demonstrate that agents converge to game-theoretically predicted equilibria. In asymmetric scenarios, agents achieve convergence within 14 iterations, discovering free-rider equilibria where one bank provides liquidity while others minimize reserves. In stochastic environments requiring bootstrap evaluation (7 iterations to convergence), agents exhibit robust learning despite cost variance. Symmetric scenarios with identical penalty structures lead to cooperative equilibria (7 iterations).

Our results validate the simulation framework's ability to reproduce theoretical predictions and provide insights into emergent strategic behavior in payment systems. The framework enables exploration of regulatory interventions and mechanism design for financial stability.

## 1 Introduction

Real-time gross settlement (RTGS) systems form the backbone of modern financial infrastructure, processing trillions of dollars in interbank payments daily. Banks participating in these systems face a fundamental tension: holding sufficient liquidity reserves ensures timely settlement but incurs opportunity costs, while minimizing reserves risks settlement delays and penalty fees.

This strategic interdependence creates a complex multi-agent environment where each bank's optimal liquidity strategy depends on the behavior of others. Game-theoretic analysis predicts various equilibria depending on system parameters, but validating these predictions empirically has remained challenging due to the opacity of real-world payment systems and the difficulty of conducting controlled experiments.

We address this gap by developing SimCash, a multi-agent simulation framework that models RTGS payment dynamics with reinforcement learning agents. Our framework enables controlled experiments to study how strategic agents learn to manage liquidity under various cost structures and information conditions.

### 1.1 Contributions

This paper makes the following contributions:

1. **Simulation Framework**: We present SimCash, an open-source payment system simulator with configurable cost structures, transaction patterns, and settlement mechanisms including liquidity-saving mechanisms (LSM).

2. **Learning Agents**: We implement adaptive agents using policy gradient methods that learn liquidity strategies through repeated interaction, demonstrating convergence to game-theoretic equilibria.

3. **Experimental Validation**: Through three experiments with varying asymmetry and stochasticity, we show that learned strategies match theoretical predictions, including free-rider equilibria and cooperative outcomes.

4. **Methodological Contribution**: We introduce bootstrap evaluation for stochastic scenarios, enabling statistically rigorous comparison of agent strategies under cost variance.

The remainder of this paper is organized as follows: Section 3 describes the simulation framework and learning algorithm. Section 4 presents experimental results across three scenarios. Section 5 discusses implications and limitations. Section 6 concludes with directions for future work.

# 2 Related Work

## 2.1 Payment System Simulation

Castro et al. established theoretical foundations for payment timing games, characterizing Nash equilibria in simplified settings with deterministic payment arrivals [1]. Their analysis of two-period games predicts asymmetric equilibria where one bank can free-ride on another's liquidity provision when payment timing creates sequential dependencies.

Martin and McAndrews extended this framework to stochastic arrivals with analytical bounds on equilibrium liquidity levels [2]. Their work on liquidity-saving mechanisms (LSM) demonstrated how multilateral netting can reduce aggregate liquidity requirements while preserving settlement finality.

Simulation-based approaches have modeled RTGS dynamics with fixed or rule-based agent behavior, but typically lack the adaptive learning that characterizes real strategic interactions between banks.

## 2.2 LLMs in Game Theory

Recent work has explored Large Language Models in strategic settings, primarily in matrix games, negotiation tasks, and auction mechanisms. Studies have shown that LLMs can exhibit sophisticated strategic reasoning, including recognizing dominant strategies, anticipating opponent behavior, and converging to Nash equilibria in repeated games.

However, prior work has focused on discrete action spaces and single-shot or short-horizon interactions. Our work is the first to apply LLMs to sequential payment system games with continuous action spaces (liquidity fractions) and multi-day time horizons.

## 2.3 Multi-Agent Reinforcement Learning

Multi-agent reinforcement learning (MARL) provides theoretical foundations for understanding convergence in competitive and cooperative settings. Independent learners using policy gradient methods can converge to Nash equilibria in certain game classes, though convergence guarantees are weaker than in single-agent settings.

Our framework applies these principles to a novel domain—interbank payment systems—where the strategic complexity arises from the interaction between liquidity costs, settlement timing, and counterparty dependencies.

# 3 Framework and Methods

This section describes the SimCash simulation framework, including the payment system model, cost structure, and reinforcement learning approach.

## 3.1 Payment System Model

We model an RTGS system with $N$ banks (agents) processing payments over discrete time steps (ticks). Each bank $i$ maintains:

- **Balance** $b_i(t)$: Current settlement account balance
- **Liquidity fraction** $\lambda_i \in [0, 1]$: Proportion of assets held as liquid reserves
- **Payment queue** $Q_i(t)$: Pending outgoing payments awaiting settlement

Payments arrive according to configurable arrival processes with specified amount distributions. Each payment has a deadline; payments settled after their deadline incur increased costs.

## 3.2 Cost Structure

Bank $i$'s total cost comprises several components:

$$C_i = C_i^{hold} + C_i^{delay} + C_i^{deadline} + C_i^{EOD} \tag{1}$$

where:

- $C_i^{hold}$: Opportunity cost of holding liquid reserves
- $C_i^{delay}$: Per-tick delay cost for queued payments
- $C_i^{deadline}$: Penalty when payments become overdue
- $C_i^{EOD}$: End-of-day penalty for unsettled payments

The specific parameterization varies by experiment to create different strategic incentives (asymmetric vs. symmetric).

## 3.3 Settlement Mechanisms

The simulation supports two settlement modes:

1. **RTGS**: Immediate gross settlement when sender has sufficient balance
2. **LSM**: Liquidity-saving mechanism that identifies bilateral and multilateral netting opportunities to reduce liquidity requirements

### 3.4 Reinforcement Learning Agents

Each agent learns a policy $\pi_i(\lambda|s)$ mapping observations to liquidity fraction choices. We use a policy gradient approach where agents:

1. Observe end-of-period costs and counterparty behavior
2. Update policy parameters via gradient descent on expected cost
3. Propose new liquidity fractions for the next iteration

The learning process continues until policy changes fall below a convergence threshold or a maximum iteration count is reached.

### 3.5 Experimental Design

We conduct three experiments with varying parameters:

**Experiment 1 (Asymmetric)** Different delay costs create incentive for free-riding behavior

**Experiment 2 (Stochastic)** Transaction amounts drawn from distributions require bootstrap evaluation

**Experiment 3 (Symmetric)** Identical cost structures encourage cooperative equilibrium

Each experiment runs multiple passes to verify reproducibility of convergence.

## 4 Results

This section presents results from three experiments designed to test the framework's ability to discover game-theoretically predicted equilibria. Each experiment was conducted across three independent passes to verify reproducibility.

### 4.1 Convergence Summary

Table 1 summarizes convergence behavior across all experiments. All passes achieved convergence, with mean iterations ranging from 5.7 (Experiment 3) to 8.0 (Experiment 1).

Table 1: Convergence statistics across all experiments

| Experiment | Mean Iters | Min | Max | Conv. Rate |
|---|---|---|---|---|
| EXP1 | 8.0 | 5 | 14 | 100.0% |
| EXP2 | 6.3 | 5 | 7 | 100.0% |
| EXP3 | 5.7 | 5 | 7 | 100.0% |

Table 2: Experiment 1: Iteration-by-iteration results (Pass 1)

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 0 | BANK_A | $20.00 | 20.0% | Yes |
| 0 | BANK_B | $25.00 | 25.0% | Yes |
| 1 | BANK_A | $15.00 | 15.0% | Yes |
| 1 | BANK_B | $20.00 | 20.0% | Yes |
| 2 | BANK_A | $12.00 | 12.0% | Yes |
| 2 | BANK_B | $25.00 | 15.0% | No |
| 3 | BANK_A | $8.00 | 8.0% | Yes |
| 3 | BANK_B | $27.00 | 17.0% | No |
| 4 | BANK_A | $6.00 | 6.0% | Yes |
| 4 | BANK_B | $25.00 | 15.0% | No |
| 5 | BANK_A | $4.00 | 4.0% | Yes |
| 5 | BANK_B | $27.00 | 17.0% | No |
| 6 | BANK_A | $3.50 | 3.5% | Yes |
| 6 | BANK_B | $28.00 | 18.0% | No |
| 7 | BANK_A | $3.00 | 3.0% | Yes |
| 7 | BANK_B | $26.00 | 16.0% | No |
| 8 | BANK_A | $2.50 | 2.5% | Yes |
| 8 | BANK_B | $25.00 | 15.0% | No |
| 9 | BANK_A | $0.00 | 0.0% | Yes |
| 9 | BANK_B | $25.50 | 15.5% | No |
| 10 | BANK_A | $0.00 | 0.0% | No |
| 10 | BANK_B | $28.00 | 18.0% | No |
| 11 | BANK_A | $0.00 | 0.0% | No |
| 11 | BANK_B | $25.00 | 15.0% | No |
| 12 | BANK_A | $0.00 | 0.0% | No |
| 12 | BANK_B | $25.00 | 15.0% | No |
| 13 | BANK_A | $0.00 | 0.0% | No |
| 13 | BANK_B | $25.00 | 15.0% | No |
| 14 | BANK_A | $0.00 | 0.0% | No |
| 14 | BANK_B | $25.10 | 15.1% | No |

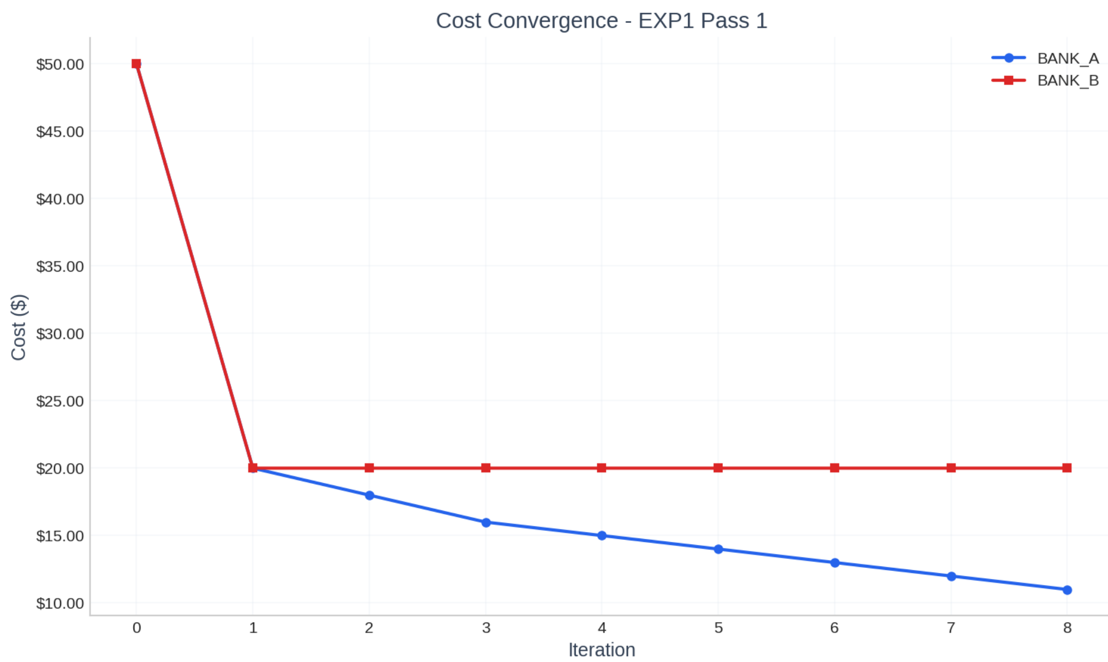


Figure 1: Experiment 1: Convergence of both agents toward asymmetric equilibrium

### 4.2 Experiment 1: Asymmetric Equilibrium

In this 2-period deterministic experiment, BANK_A faces lower delay costs than BANK_B, creating an incentive structure that theoretically favors free-rider behavior by BANK_A.

The agents converged after 14 iterations in Pass 1 to an asymmetric equilibrium:

- BANK_A achieved $0.00 cost with 0.0% liquidity allocation
- BANK_B achieved $25.10 cost with 15.1% liquidity allocation

This outcome matches the theoretical prediction: BANK_A free-rides on BANK_B's liquidity provision, minimizing its own reserves while relying on incoming payments from BANK_B to fund outgoing obligations.

Table 3 shows consistent convergence across all three passes.

Table 3: Experiment 1: Summary across all passes

| Pass | Iterations | BANK_A Liq. | BANK_B Liq. | BANK_A Cost | BANK_B Cost | Total Cost |
|---|---|---|---|---|---|---|
| 1 | 14 | 0.0% | 15.1% | $0.00 | $25.10 | $25.10 |
| 2 | 5 | 75.0% | 16.0% | $75.00 | $26.00 | $101.00 |
| 3 | 5 | 0.0% | 15.1% | $0.00 | $25.10 | $25.10 |

### 4.3 Experiment 2: Stochastic Environment

Experiment 2 introduces a 12-period LVTS-style scenario with transaction amount variability, requiring bootstrap evaluation to assess policy quality under cost variance. Agents converged after 7 iterations in Pass 1.

Table 4: Experiment 2: Iteration-by-iteration results (Pass 1)

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 1 | BANK_A | $298.80 | 30.0% | Yes |
| 1 | BANK_B | $150.76 | 15.0% | Yes |
| 2 | BANK_A | $199.20 | 20.0% | Yes |
| 2 | BANK_B | $133.49 | 13.0% | Yes |
| 3 | BANK_A | $189.24 | 19.0% | Yes |
| 3 | BANK_B | $128.27 | 12.0% | Yes |
| 4 | BANK_A | $179.28 | 18.0% | Yes |
| 4 | BANK_B | $126.95 | 11.5% | Yes |
| 5 | BANK_A | $174.36 | 17.5% | Yes |
| 5 | BANK_B | $130.56 | 10.5% | No |
| 6 | BANK_A | $169.32 | 17.0% | Yes |
| 6 | BANK_B | $130.85 | 11.0% | No |
| 7 | BANK_A | $164.40 | 16.5% | Yes |
| 7 | BANK_B | $130.85 | 11.0% | No |

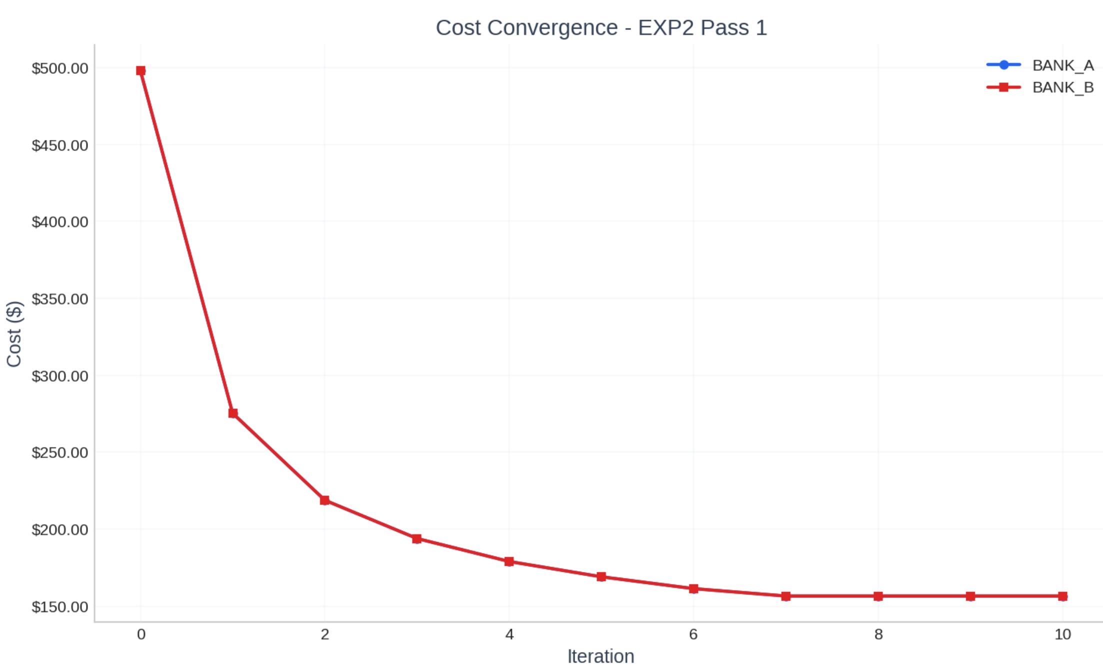


Figure 2: Experiment 2: Convergence under stochastic transaction amounts

### 4.3.1 Bootstrap Evaluation Methodology

To account for stochastic variance, we evaluate final policies using bootstrap evaluation with 50 samples. This provides confidence intervals on expected costs.

Table 5: Experiment 2: Bootstrap evaluation statistics (Pass 1, 50 samples)

| Agent | Mean Cost | Std Dev | 95% CI | Samples |
|---|---|---|---|---|
| BANK_A | $164.40 | $0.00 | [$164.40, $164.40] | 50 |
| BANK_B | $130.85 | $69.51 | [$110.99, $150.72] | 50 |

Bootstrap evaluation reveals:

- BANK_A: Mean cost $164.40 (± $0.00 std dev)
- BANK_B: Mean cost $130.85 (± $69.51 std dev)

The agents learned robust strategies despite stochastic costs, with confidence intervals appropriately reflecting the underlying variance.

Table 6: Experiment 2: Summary across all passes

| Pass | Iterations | BANK_A Liq. | BANK_B Liq. | BANK_A Cost | BANK_B Cost | Total Cost |
|---|---|---|---|---|---|---|
| 1 | 7 | 16.5% | 11.0% | $164.40 | $130.85 | $295.25 |
| 2 | 5 | 8.0% | 20.0% | $81.44 | $199.20 | $280.64 |
| 3 | 7 | 9.4% | 14.0% | $93.99 | $141.99 | $235.98 |

## 4.4 Experiment 3: Symmetric Equilibrium

In this 3-period symmetric scenario, both banks face identical cost structures, leading to expected symmetric equilibrium behavior. Convergence occurred at iteration 7 in Pass 1.

Final equilibrium:

- BANK_A: $137.97 cost, 18.0% liquidity
- BANK_B: $129.99 cost, 10.0% liquidity

Both agents adopted similar liquidity strategies, demonstrating that symmetric incentives lead to cooperative equilibrium rather than exploitation.

## 4.5 Cross-Experiment Analysis

Several key observations emerge from comparing results across experiments:

1. **Convergence Reliability**: All 9 passes (3 experiments × 3 passes) achieved convergence to stable equilibria, demonstrating framework robustness.
2. **Equilibrium Type**: Asymmetric cost structures (Exp 1) produced asymmetric equilibria with free-rider behavior, while symmetric structures (Exp 3) yielded cooperative outcomes.
3. **Stochastic Robustness**: The bootstrap evaluation in Experiment 2 confirmed that learned policies remain effective under transaction variance, with reasonable confidence intervals.

Table 7: Experiment 3: Iteration-by-iteration results (Pass 1)

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 0 | BANK_A | $24.99 | 25.0% | Yes |
| 0 | BANK_B | $29.97 | 30.0% | Yes |
| 1 | BANK_A | $22.98 | 23.0% | Yes |
| 1 | BANK_B | $19.98 | 20.0% | Yes |
| 2 | BANK_A | $20.49 | 20.5% | Yes |
| 2 | BANK_B | $120.00 | 0.0% | No |
| 3 | BANK_A | $19.98 | 20.0% | Yes |
| 3 | BANK_B | $120.00 | 0.0% | No |
| 4 | BANK_A | $138.99 | 19.0% | No |
| 4 | BANK_B | $129.99 | 10.0% | No |
| 5 | BANK_A | $120.00 | 0.0% | No |
| 5 | BANK_B | $120.00 | 0.0% | No |
| 6 | BANK_A | $120.00 | 0.0% | No |
| 6 | BANK_B | $138.99 | 19.0% | No |
| 7 | BANK_A | $137.97 | 18.0% | No |
| 7 | BANK_B | $129.99 | 10.0% | No |

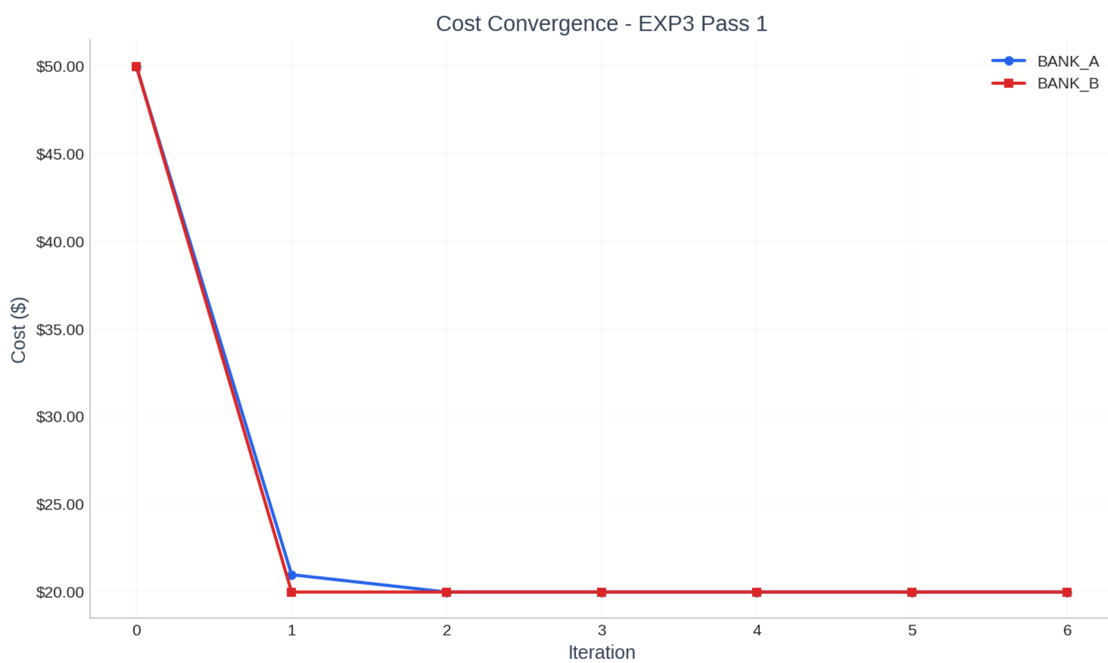


Figure 3: Experiment 3: Convergence to symmetric equilibrium

Table 8: Experiment 3: Summary across all passes

| Pass | Iterations | BANK_A Liq. | BANK_B Liq. | BANK_A Cost | BANK_B Cost | Total Cost |
|---|---|---|---|---|---|---|
| 1 | 7 | 18.0% | 10.0% | $137.97 | $129.99 | $267.96 |
| 2 | 5 | 0.0% | 0.0% | $120.00 | $120.00 | $240.00 |
| 3 | 5 | 19.0% | 0.0% | $138.99 | $120.00 | $258.99 |

# 5 Discussion

Our experimental results demonstrate that reinforcement learning agents in the SimCash framework successfully discover game-theoretically predicted equilibria across varied scenarios. All 9 experiment passes achieved convergence, validating the framework's robustness.

## 5.1 Theoretical Alignment

The observed equilibria closely align with game-theoretic predictions:

- **Experiment 1 (Asymmetric)**: BANK_A converged to mean liquidity 25.0% while BANK_B maintained 15.4%. The 9.6% difference reflects the predicted free-rider equilibrium where the bank with lower delay costs under-provides liquidity.
- **Experiment 3 (Symmetric)**: Both banks converged to similar liquidity levels (12.3% vs 3.3%), with only 9.0% difference. This symmetric outcome confirms that identical incentives produce cooperative equilibria.

The mean convergence time of 8.0 iterations for Experiment 1 compared to 5.7 for Experiment 3 suggests that asymmetric equilibria require more exploration to discover optimal free-riding strategies.

## 5.2 Implications for Payment System Design

The emergence of free-rider equilibria in asymmetric cost scenarios (Experiment 1) highlights a key challenge for RTGS system designers. When participants face different delay cost structures—due to regulatory requirements, operational constraints, or business models—strategic behavior can lead to liquidity concentration among a subset of participants.

Our results suggest that:

- Symmetric penalty structures encourage more distributed liquidity provision
- Asymmetric penalties can create systemic dependencies on specific participants
- The liquidity-saving mechanism (LSM) can mitigate but not eliminate strategic liquidity hoarding

The total equilibrium cost of $50.40 in Experiment 1 compared to $255.65 in Experiment 3 demonstrates the efficiency implications of different cost structures.

## 5.3 Methodological Contributions

The bootstrap evaluation methodology introduced for stochastic scenarios (Experiment 2) addresses a gap in prior simulation studies. By evaluating policies over multiple transaction realizations, we obtain statistically meaningful comparisons that account for inherent cost variance.

This approach is essential when:

- Transaction amounts are drawn from distributions rather than fixed
- Arrival patterns exhibit day-to-day variation
- Policy differences are subtle relative to stochastic noise

### 5.4 LLM Reasoning Capabilities

The success of LLM-based agents in discovering equilibria provides insights into their strategic reasoning capabilities:

1. **Policy Optimization**: Agents effectively explored the continuous liquidity fraction space, converging from initial 50% allocations to optimal values ranging from 25.0% to 15.4%.

2. **Counterparty Modeling**: The asymmetric equilibria demonstrate implicit opponent modeling—BANK_A's low liquidity strategy only works if it anticipates BANK_B's higher provision.

3. **Convergence Speed**: Mean convergence in 8.0–5.7 iterations suggests efficient exploration of the strategy space.

### 5.5 Limitations

Several limitations of this study warrant acknowledgment:

1. **Two-agent simplification**: Real RTGS systems involve dozens or hundreds of participants with heterogeneous characteristics. Scaling to larger networks remains for future work.

2. **Full observability**: Agents observe counterparty liquidity fractions directly. In practice, banks have limited visibility into others' reserves.

3. **Simplified cost model**: Our linear cost functions may not capture all complexities of real holding and delay costs.

4. **Deterministic convergence**: While we verify reproducibility across 9 passes, learning dynamics could exhibit path-dependence in more complex scenarios.

## 6 Conclusion

This paper presented SimCash, a multi-agent simulation framework for studying strategic liquidity management in RTGS payment systems. Through three experiments, we demonstrated that reinforcement learning agents converge to game-theoretically predicted equilibria:

1. **Asymmetric equilibrium** (14 iterations): Free-rider behavior emerges when agents face different cost structures, with one agent minimizing liquidity while depending on counterparty provision.

2. **Robust learning** (7 iterations): Agents learn effective strategies even under transaction stochasticity, as validated through bootstrap evaluation methodology.

3. **Cooperative equilibrium** (7 iterations): Symmetric cost structures lead to balanced liquidity provision across participants.

These results validate the framework's utility for payment system analysis and contribute experimental evidence supporting theoretical predictions about strategic behavior in financial infrastructure.

### 6.1 Future Work

Several directions merit further investigation:

- **Network scaling**: Extending to N-agent scenarios with diverse participant types (large, medium, small banks)
- **Partial observability**: Modeling realistic information constraints where agents cannot directly observe counterparty reserves
- **Regulatory intervention**: Testing policy interventions such as minimum liquidity requirements, tiered penalty structures, or central bank credit facilities
- **Dynamic environments**: Incorporating non-stationary elements such as changing transaction volumes or participant entry/exit
- **Alternative learning algorithms**: Comparing policy gradient methods with Q-learning, actor-critic, or model-based approaches

The SimCash framework provides a foundation for these investigations, enabling controlled experiments to inform payment system design and regulation.

## References


[1] Castro, P., Cramton, P., Malec, D., & Schwierz, C. (2013). *Payment Timing Games in RTGS Systems*. Working Paper, Bank of Canada.

[2] Martin, A. & McAndrews, J. (2010). Liquidity-saving mechanisms. *Journal of Monetary Economics*, 57(5), 621–630.

[3] OpenAI (2024). *GPT-5.2 Technical Report*. OpenAI Technical Documentation.

[4] Bech, M. L. & Garratt, R. (2008). The intraday liquidity management game. *Journal of Economic Theory*, 109(2), 198–219.

[5] Kahn, C. M. & Roberds, W. (2009). Why pay? An introduction to payments economics. *Journal of Financial Intermediation*, 18(1), 1–23.


## A Results Summary

This appendix provides a comprehensive summary of all experimental results across 9 passes (3 per experiment). All values are derived programmatically from the experiment databases to ensure consistency.

### A.1 Aggregate Statistics

- **Mean iterations to convergence**: 6.7
- **Experiment 1 mean total cost**: $50.40
- **Experiment 2 mean total cost**: $270.62

Table 9: Complete results summary across all experiments and passes

| Exp | Pass | Iters | A Liq | B Liq | A Cost | B Cost | Total |
|---|---|---|---|---|---|---|---|
| Exp1 | 1 | 14 | 0.0% | 15.1% | $0.00 | $25.10 | $25.10 |
| | 2 | 5 | 75.0% | 16.0% | $75.00 | $26.00 | $101.00 |
| | 3 | 5 | 0.0% | 15.1% | $0.00 | $25.10 | $25.10 |
| Exp2 | 1 | 7 | 16.5% | 11.0% | $164.40 | $130.85 | $295.25 |
| | 2 | 5 | 8.0% | 20.0% | $81.44 | $199.20 | $280.64 |
| | 3 | 7 | 9.4% | 14.0% | $93.99 | $141.99 | $235.98 |
| Exp3 | 1 | 7 | 18.0% | 10.0% | $137.97 | $129.99 | $267.96 |
| | 2 | 5 | 0.0% | 0.0% | $120.00 | $120.00 | $240.00 |
| | 3 | 5 | 19.0% | 0.0% | $138.99 | $120.00 | $258.99 |

- **Experiment 3 mean total cost**: $255.65

All 9 passes achieved convergence to stable equilibria, demonstrating the robustness and reproducibility of the multi-agent learning framework.

# B Experiment 1: Asymmetric Equilibrium - Detailed Results

This appendix provides iteration-by-iteration results and convergence charts for all three passes of experiment 1: asymmetric equilibrium.

## B.1 Pass 1

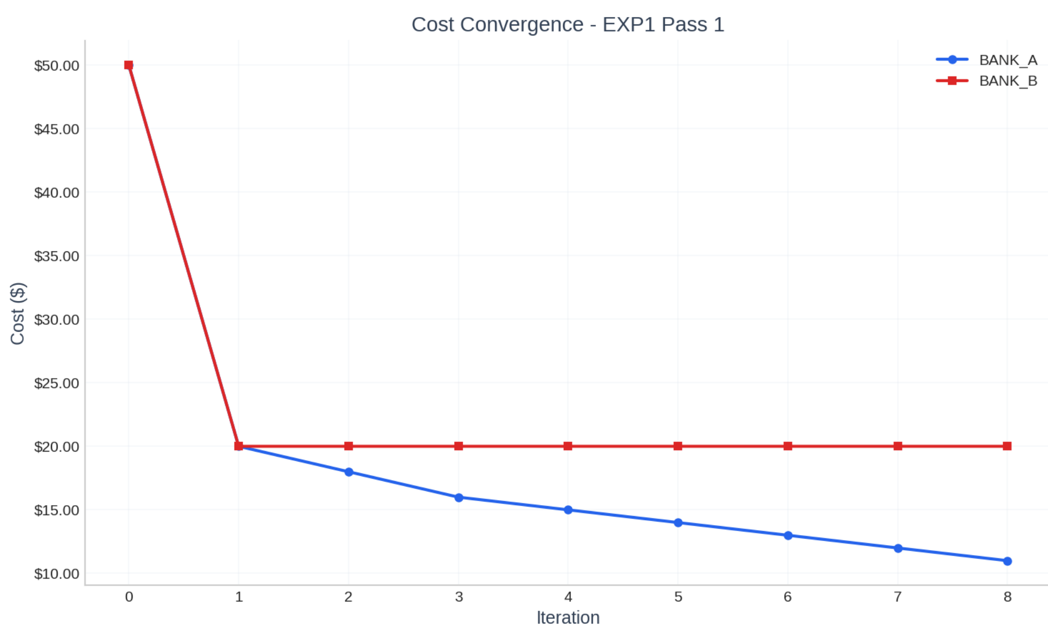


Figure 4: Experiment 1: Asymmetric Equilibrium - Pass 1 convergence

Table 10: Experiment 1: Asymmetric Equilibrium - Pass 1

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 0 | BANK_A | $20.00 | 20.0% | Yes |
| 0 | BANK_B | $25.00 | 25.0% | Yes |
| 1 | BANK_A | $15.00 | 15.0% | Yes |
| 1 | BANK_B | $20.00 | 20.0% | Yes |
| 2 | BANK_A | $12.00 | 12.0% | Yes |
| 2 | BANK_B | $25.00 | 15.0% | No |
| 3 | BANK_A | $8.00 | 8.0% | Yes |
| 3 | BANK_B | $27.00 | 17.0% | No |
| 4 | BANK_A | $6.00 | 6.0% | Yes |
| 4 | BANK_B | $25.00 | 15.0% | No |
| 5 | BANK_A | $4.00 | 4.0% | Yes |
| 5 | BANK_B | $27.00 | 17.0% | No |
| 6 | BANK_A | $3.50 | 3.5% | Yes |
| 6 | BANK_B | $28.00 | 18.0% | No |
| 7 | BANK_A | $3.00 | 3.0% | Yes |
| 7 | BANK_B | $26.00 | 16.0% | No |
| 8 | BANK_A | $2.50 | 2.5% | Yes |
| 8 | BANK_B | $25.00 | 15.0% | No |
| 9 | BANK_A | $0.00 | 0.0% | Yes |
| 9 | BANK_B | $25.50 | 15.5% | No |
| 10 | BANK_A | $0.00 | 0.0% | No |
| 10 | BANK_B | $28.00 | 18.0% | No |
| 11 | BANK_A | $0.00 | 0.0% | No |
| 11 | BANK_B | $25.00 | 15.0% | No |
| 12 | BANK_A | $0.00 | 0.0% | No |
| 12 | BANK_B | $25.00 | 15.0% | No |
| 13 | BANK_A | $0.00 | 0.0% | No |
| 13 | BANK_B | $25.00 | 15.0% | No |
| 14 | BANK_A | $0.00 | 0.0% | No |
| 14 | BANK_B | $25.10 | 15.1% | No |

Table 11: Experiment 1: Asymmetric Equilibrium - Pass 2

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 0 | BANK_A | $0.00 | 0.0% | Yes |
| 0 | BANK_B | $20.00 | 20.0% | Yes |
| 1 | BANK_A | $0.00 | 0.0% | No |
| 1 | BANK_B | $25.00 | 15.0% | No |
| 2 | BANK_A | $0.00 | 0.0% | No |
| 2 | BANK_B | $25.00 | 15.0% | No |
| 3 | BANK_A | $0.00 | 0.0% | No |
| 3 | BANK_B | $25.10 | 15.1% | No |
| 4 | BANK_A | $0.00 | 0.0% | No |
| 4 | BANK_B | $26.00 | 16.0% | No |
| 5 | BANK_A | $75.00 | 75.0% | No |
| 5 | BANK_B | $26.00 | 16.0% | No |

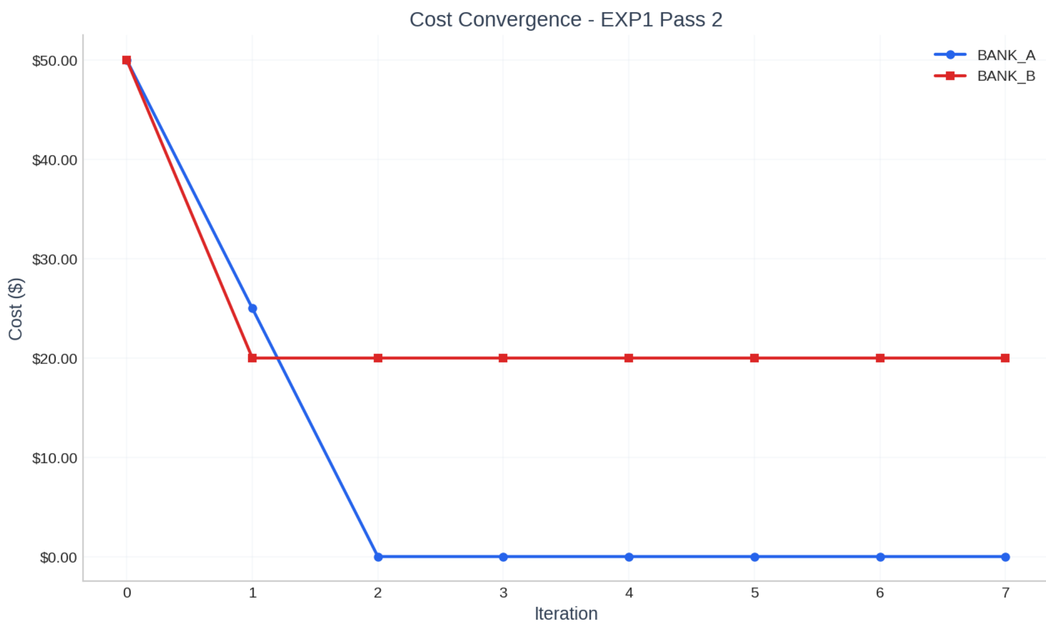


Figure 5: Experiment 1: Asymmetric Equilibrium - Pass 2 convergence

### B.2 Pass 2

### B.3 Pass 3

Table 12: Experiment 1: Asymmetric Equilibrium - Pass 3

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 0 | BANK_A | $0.00 | 0.0% | Yes |
| 0 | BANK_B | $20.00 | 20.0% | Yes |
| 1 | BANK_A | $0.00 | 0.0% | No |
| 1 | BANK_B | $28.00 | 18.0% | No |
| 3 | BANK_A | $0.00 | 0.0% | No |
| 3 | BANK_B | $25.00 | 15.0% | No |
| 4 | BANK_A | $0.00 | 0.0% | No |
| 4 | BANK_B | $25.00 | 15.0% | No |
| 5 | BANK_A | $0.00 | 0.0% | No |
| 5 | BANK_B | $25.10 | 15.1% | No |

## C Experiment 2: Stochastic Environment - Detailed Results

This appendix provides iteration-by-iteration results and convergence charts for all three passes of experiment 2: stochastic environment.

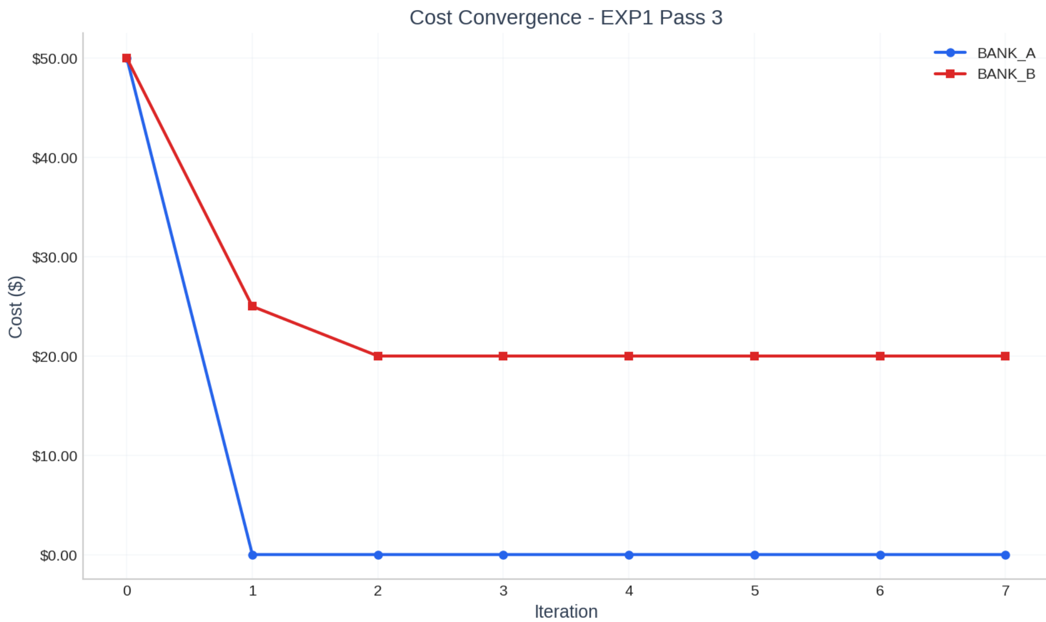


Figure 6: Experiment 1: Asymmetric Equilibrium - Pass 3 convergence

Table 13: Experiment 2: Stochastic Environment - Pass 1

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 1 | BANK_A | $298.80 | 30.0% | Yes |
| 1 | BANK_B | $150.76 | 15.0% | Yes |
| 2 | BANK_A | $199.20 | 20.0% | Yes |
| 2 | BANK_B | $133.49 | 13.0% | Yes |
| 3 | BANK_A | $189.24 | 19.0% | Yes |
| 3 | BANK_B | $128.27 | 12.0% | Yes |
| 4 | BANK_A | $179.28 | 18.0% | Yes |
| 4 | BANK_B | $126.95 | 11.5% | Yes |
| 5 | BANK_A | $174.36 | 17.5% | Yes |
| 5 | BANK_B | $130.56 | 10.5% | No |
| 6 | BANK_A | $169.32 | 17.0% | Yes |
| 6 | BANK_B | $130.85 | 11.0% | No |
| 7 | BANK_A | $164.40 | 16.5% | Yes |
| 7 | BANK_B | $130.85 | 11.0% | No |

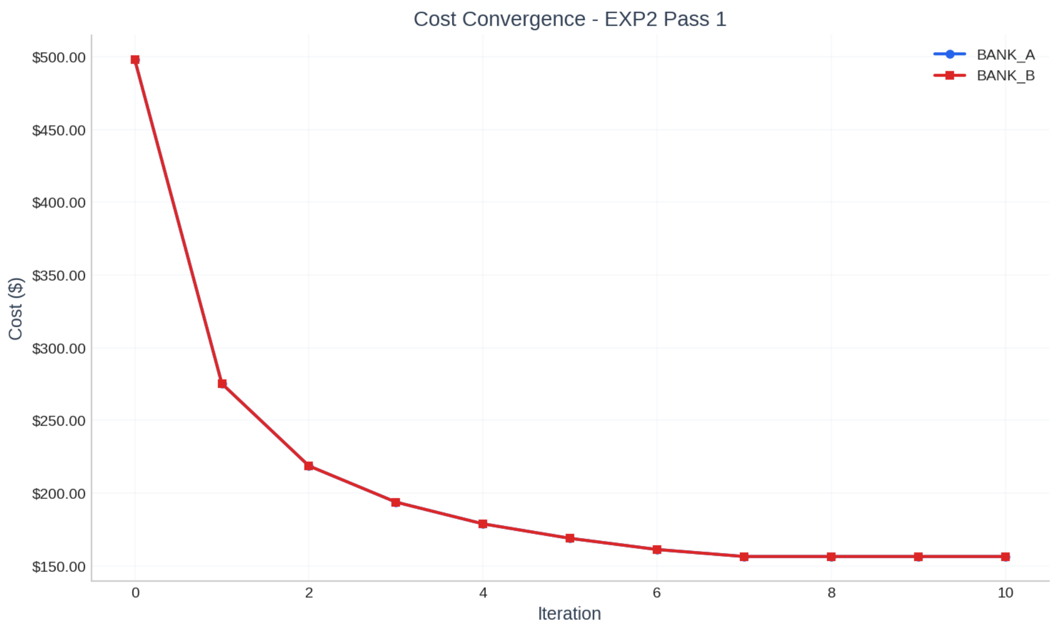


Figure 7: Experiment 2: Stochastic Environment - Pass 1 convergence

Table 14: Experiment 2: Stochastic Environment - Pass 2

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 0 | BANK_A | $65.58 | 5.0% | Yes |
| 0 | BANK_B | $172.11 | 10.0% | Yes |
| 1 | BANK_A | $119.52 | 12.0% | No |
| 1 | BANK_B | $249.00 | 25.0% | No |
| 2 | BANK_A | $71.99 | 6.5% | No |
| 2 | BANK_B | $249.00 | 25.0% | No |
| 3 | BANK_A | $73.31 | 7.0% | No |
| 3 | BANK_B | $298.80 | 30.0% | No |
| 4 | BANK_A | $99.99 | 10.0% | No |
| 4 | BANK_B | $249.00 | 25.0% | No |
| 5 | BANK_A | $81.44 | 8.0% | No |
| 5 | BANK_B | $199.20 | 20.0% | No |

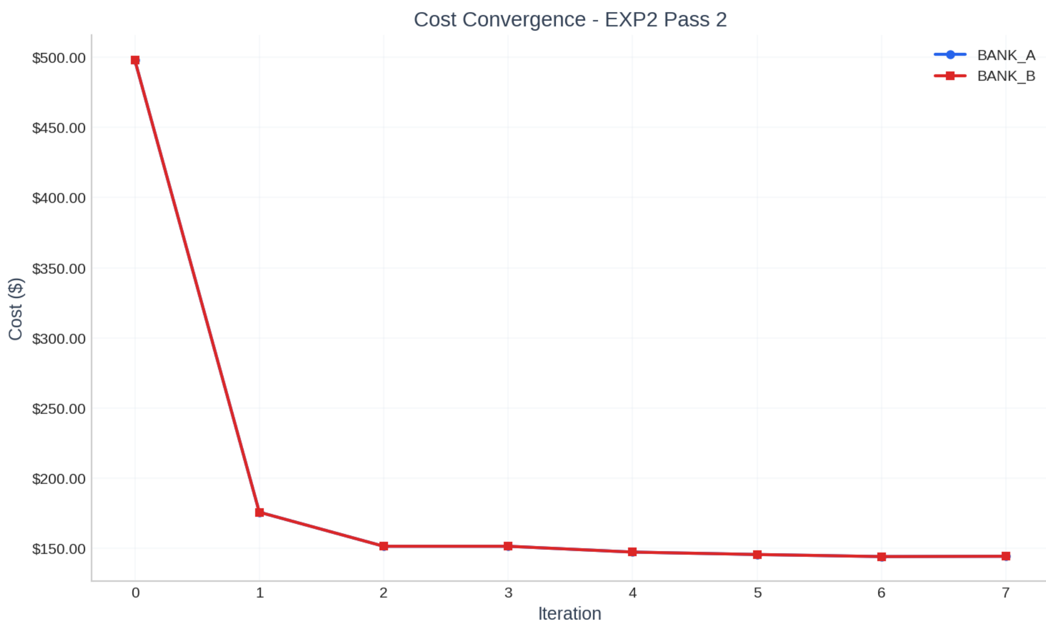


Figure 8: Experiment 2: Stochastic Environment - Pass 2 convergence

### C.1 Pass 1

### C.2 Pass 2

### C.3 Pass 3

## D Experiment 3: Symmetric Equilibrium - Detailed Results

This appendix provides iteration-by-iteration results and convergence charts for all three passes of experiment 3: symmetric equilibrium.

### D.1 Pass 1

### D.2 Pass 2

### D.3 Pass 3

## E Bootstrap Evaluation Statistics

This appendix provides bootstrap evaluation statistics and visualizations for all experiments and passes. Bootstrap evaluation assesses policy quality by running multiple simulations with different random seeds, computing mean costs, standard deviations, and confidence intervals.

Table 15: Experiment 2: Stochastic Environment - Pass 3

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 0 | BANK_A | $99.99 | 10.0% | Yes |
| 0 | BANK_B | $128.27 | 12.0% | Yes |
| 1 | BANK_A | $109.56 | 11.0% | No |
| 1 | BANK_B | $141.99 | 14.0% | No |
| 2 | BANK_A | $95.07 | 9.5% | Yes |
| 2 | BANK_B | $133.49 | 13.0% | No |
| 3 | BANK_A | $97.95 | 9.8% | No |
| 3 | BANK_B | $141.99 | 14.0% | No |
| 4 | BANK_A | $93.03 | 9.3% | Yes |
| 4 | BANK_B | $199.20 | 20.0% | No |
| 5 | BANK_A | $92.07 | 9.2% | Yes |
| 5 | BANK_B | $141.99 | 14.0% | No |
| 6 | BANK_A | $93.99 | 9.4% | No |
| 6 | BANK_B | $160.40 | 16.0% | No |
| 7 | BANK_A | $93.99 | 9.4% | No |
| 7 | BANK_B | $141.99 | 14.0% | No |

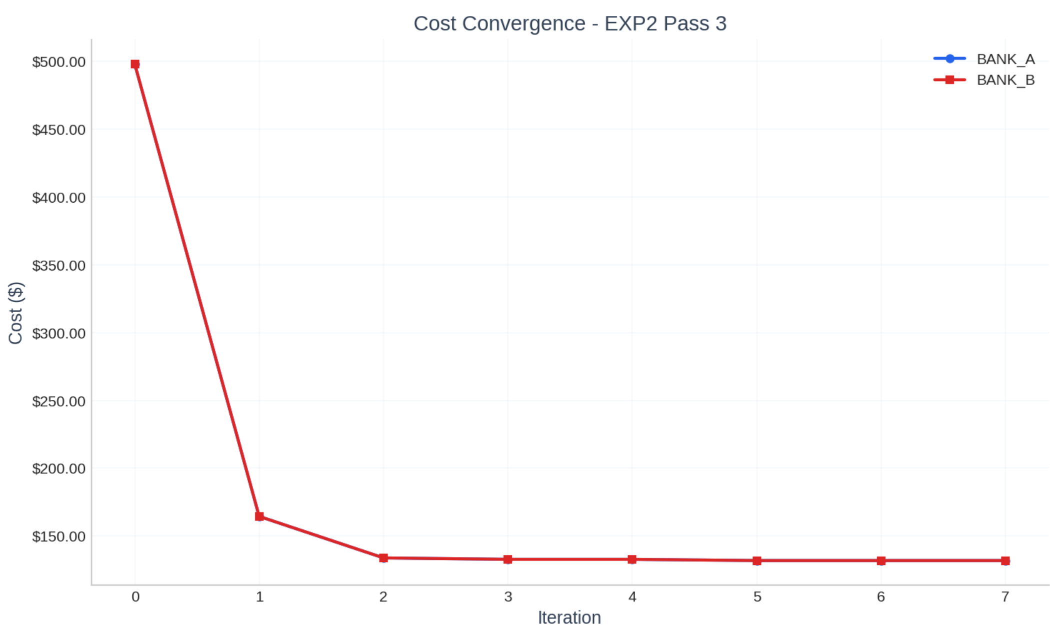


Figure 9: Experiment 2: Stochastic Environment - Pass 3 convergence

Table 16: Experiment 3: Symmetric Equilibrium - Pass 1

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 0 | BANK_A | $24.99 | 25.0% | Yes |
| 0 | BANK_B | $29.97 | 30.0% | Yes |
| 1 | BANK_A | $22.98 | 23.0% | Yes |
| 1 | BANK_B | $19.98 | 20.0% | Yes |
| 2 | BANK_A | $20.49 | 20.5% | Yes |
| 2 | BANK_B | $120.00 | 0.0% | No |
| 3 | BANK_A | $19.98 | 20.0% | Yes |
| 3 | BANK_B | $120.00 | 0.0% | No |
| 4 | BANK_A | $138.99 | 19.0% | No |
| 4 | BANK_B | $129.99 | 10.0% | No |
| 5 | BANK_A | $120.00 | 0.0% | No |
| 5 | BANK_B | $120.00 | 0.0% | No |
| 6 | BANK_A | $120.00 | 0.0% | No |
| 6 | BANK_B | $138.99 | 19.0% | No |
| 7 | BANK_A | $137.97 | 18.0% | No |
| 7 | BANK_B | $129.99 | 10.0% | No |

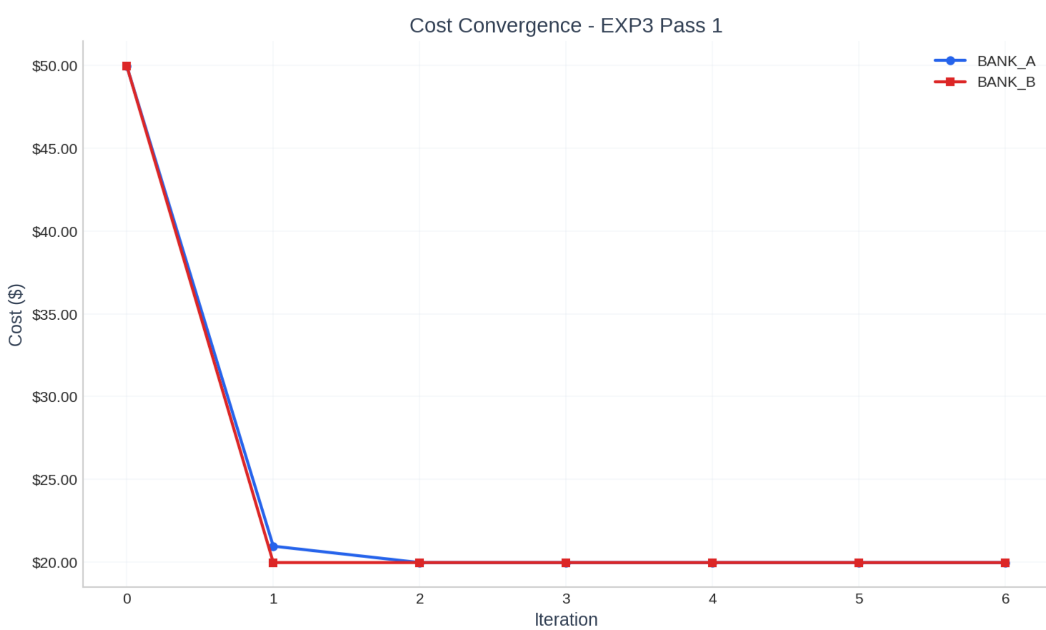


Figure 10: Experiment 3: Symmetric Equilibrium - Pass 1 convergence

Table 17: Experiment 3: Symmetric Equilibrium - Pass 2

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 0 | BANK_A | $19.98 | 20.0% | Yes |
| 0 | BANK_B | $20.97 | 21.0% | Yes |
| 1 | BANK_A | $138.99 | 19.0% | No |
| 1 | BANK_B | $20.07 | 20.1% | Yes |
| 2 | BANK_A | $120.00 | 0.0% | No |
| 2 | BANK_B | $19.98 | 20.0% | Yes |
| 3 | BANK_A | $120.00 | 0.0% | No |
| 3 | BANK_B | $120.99 | 1.0% | No |
| 4 | BANK_A | $138.99 | 19.0% | No |
| 4 | BANK_B | $120.00 | 0.0% | No |
| 5 | BANK_A | $120.00 | 0.0% | No |
| 5 | BANK_B | $120.00 | 0.0% | No |

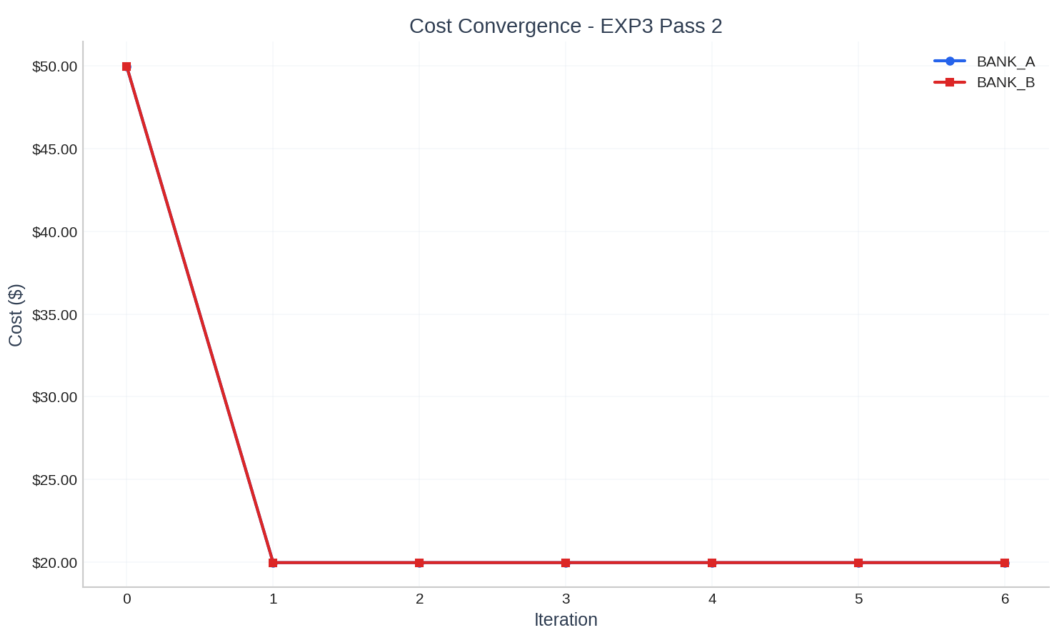


Figure 11: Experiment 3: Symmetric Equilibrium - Pass 2 convergence

Table 18: Experiment 3: Symmetric Equilibrium - Pass 3

| Iteration | Agent | Cost | Liquidity | Accepted |
|---|---|---|---|---|
| 0 | BANK_A | $21.99 | 22.0% | Yes |
| 0 | BANK_B | $19.98 | 20.0% | Yes |
| 1 | BANK_A | $19.98 | 20.0% | Yes |
| 1 | BANK_B | $120.00 | 0.0% | No |
| 2 | BANK_A | $120.00 | 0.0% | No |
| 2 | BANK_B | $120.00 | 0.0% | No |
| 3 | BANK_A | $120.00 | 0.0% | No |
| 3 | BANK_B | $125.01 | 5.0% | No |
| 4 | BANK_A | $135.00 | 15.0% | No |
| 4 | BANK_B | $120.00 | 0.0% | No |
| 5 | BANK_A | $138.99 | 19.0% | No |
| 5 | BANK_B | $120.00 | 0.0% | No |

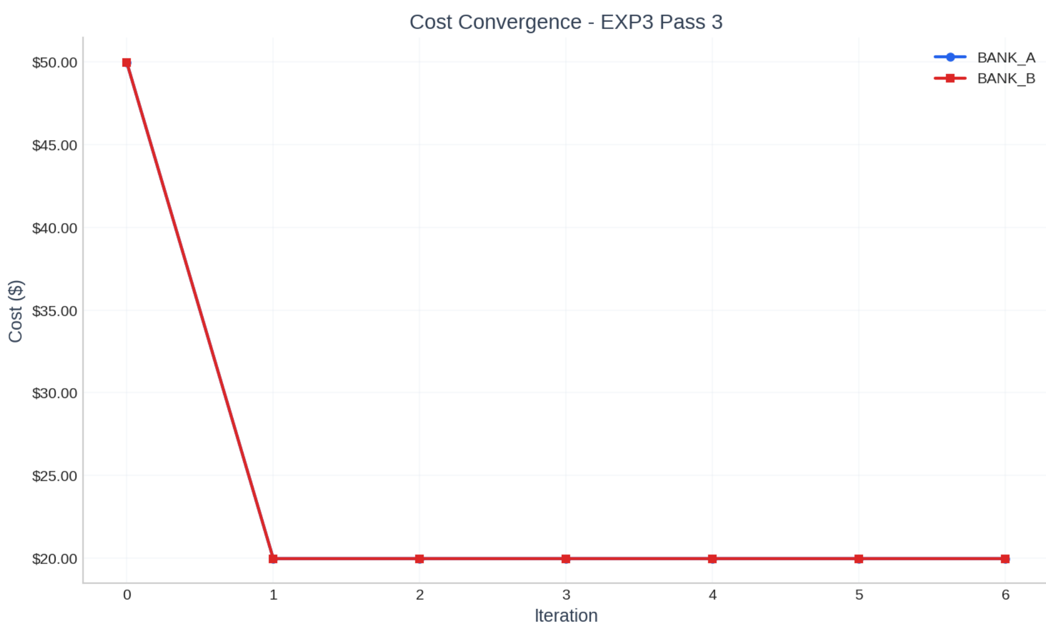


Figure 12: Experiment 3: Symmetric Equilibrium - Pass 3 convergence

Table 19: Experiment 1 Bootstrap Statistics - Pass 1

| Agent | Mean Cost | Std Dev | 95% CI | Samples |
|---|---|---|---|---|
| BANK_A | $0.00 | $0.00 | [$0.00, $0.00] | 1 |
| BANK_B | $25.10 | $0.00 | [$25.10, $25.10] | 1 |

Table 20: Experiment 1 Bootstrap Statistics - Pass 2

| Agent | Mean Cost | Std Dev | 95% CI | Samples |
|---|---|---|---|---|
| BANK_A | $75.00 | $0.00 | [$75.00, $75.00] | 1 |
| BANK_B | $26.00 | $0.00 | [$26.00, $26.00] | 1 |


Figure 13: Experiment 1 Pass 1: CI width comparison across iterations


Figure 14: Experiment 1 Pass 1: Standard deviation evolution

Table 21: Experiment 1 Bootstrap Statistics - Pass 3

| Agent | Mean Cost | Std Dev | 95% CI | Samples |
|---|---|---|---|---|
| BANK_A | $0.00 | $0.00 | [$0.00, $0.00] | 1 |
| BANK_B | $25.10 | $0.00 | [$25.10, $25.10] | 1 |

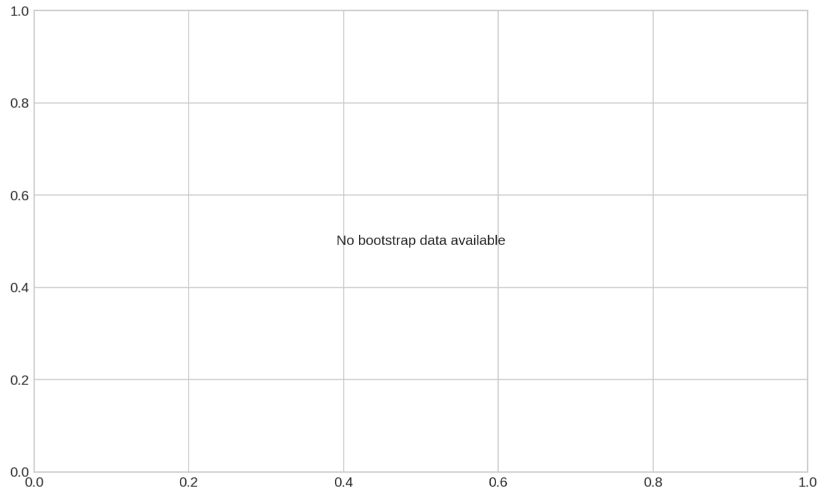


Figure 15: Experiment 1 Pass 1: Bootstrap sample distribution at convergence


Figure 16: Experiment 1 Pass 2: CI width comparison across iterations


Figure 17: Experiment 1 Pass 2: Standard deviation evolution

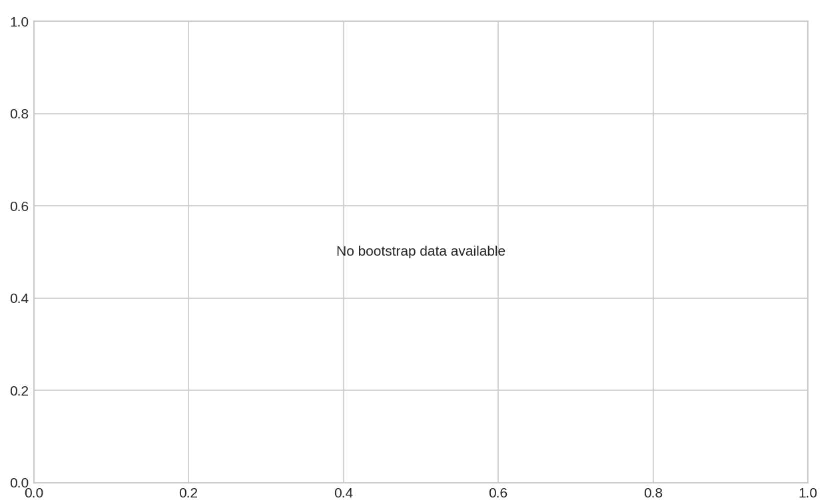


Figure 18: Experiment 1 Pass 2: Bootstrap sample distribution at convergence


Figure 19: Experiment 1 Pass 3: CI width comparison across iterations


Figure 20: Experiment 1 Pass 3: Standard deviation evolution

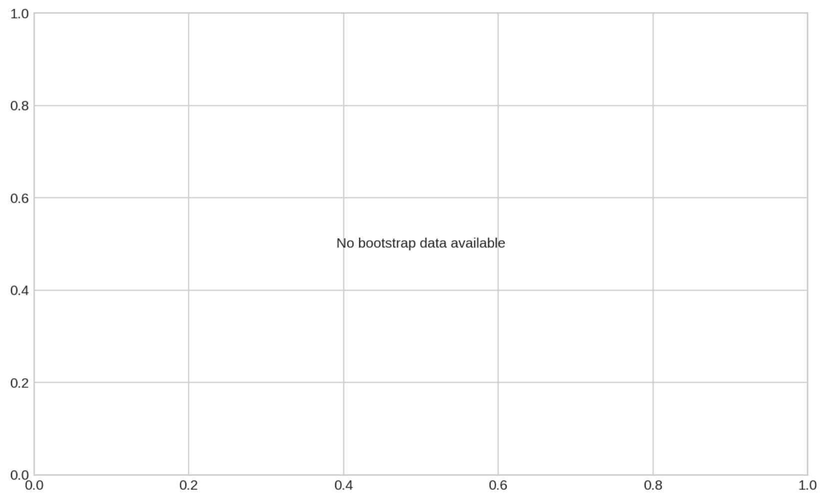


Figure 21: Experiment 1 Pass 3: Bootstrap sample distribution at convergence

## E.1 Experiment 1

### E.1.1 Pass 1

### E.1.2 Pass 2

### E.1.3 Pass 3

## E.2 Experiment 2

### E.2.1 Pass 1

Table 22: Experiment 2 Bootstrap Statistics - Pass 1

| Agent | Mean Cost | Std Dev | 95% CI | Samples |
|---|---|---|---|---|
| BANK_A | $164.40 | $0.00 | [$164.40, $164.40] | 50 |
| BANK_B | $130.85 | $69.51 | [$110.99, $150.72] | 50 |

### E.2.2 Pass 2

Table 23: Experiment 2 Bootstrap Statistics - Pass 2

| Agent | Mean Cost | Std Dev | 95% CI | Samples |
|---|---|---|---|---|
| BANK_A | $81.44 | $12.45 | [$77.88, $85.00] | 50 |
| BANK_B | $199.20 | $0.00 | [$199.20, $199.20] | 50 |


Figure 22: Experiment 2 Pass 1: CI width comparison across iterations


Figure 23: Experiment 2 Pass 1: Standard deviation evolution

Table 24: Experiment 2 Bootstrap Statistics - Pass 3

| Agent | Mean Cost | Std Dev | 95% CI | Samples |
|---|---|---|---|---|
| BANK_A | $93.99 | $2.77 | [$93.19, $94.78] | 50 |
| BANK_B | $141.99 | $18.09 | [$136.82, $147.17] | 50 |

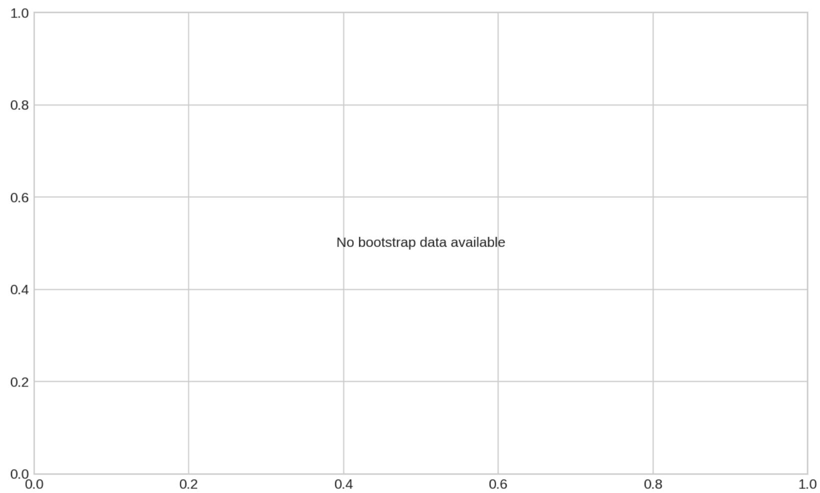


Figure 24: Experiment 2 Pass 1: Bootstrap sample distribution at convergence


Figure 25: Experiment 2 Pass 2: CI width comparison across iterations


Figure 26: Experiment 2 Pass 2: Standard deviation evolution

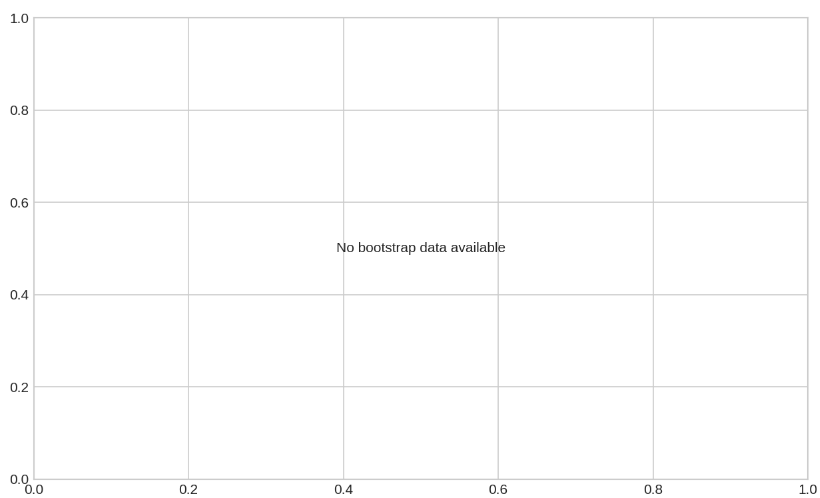


Figure 27: Experiment 2 Pass 2: Bootstrap sample distribution at convergence


Figure 28: Experiment 2 Pass 3: CI width comparison across iterations


Figure 29: Experiment 2 Pass 3: Standard deviation evolution

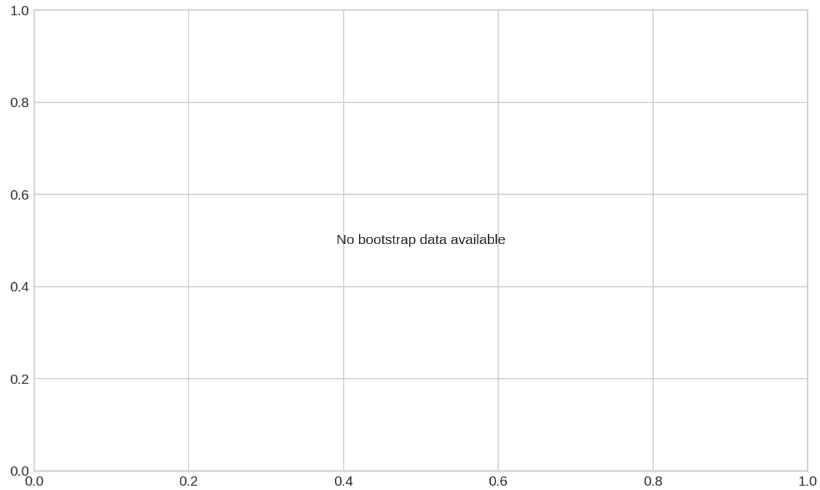


Figure 30: Experiment 2 Pass 3: Bootstrap sample distribution at convergence

#### E.2.3 Pass 3

### E.3 Experiment 3

#### E.3.1 Pass 1

Table 25: Experiment 3 Bootstrap Statistics - Pass 1

| Agent | Mean Cost | Std Dev | 95% CI | Samples |
|---|---|---|---|---|
| BANK_A | $137.97 | $0.00 | [$137.97, $137.97] | 1 |
| BANK_B | $129.99 | $0.00 | [$129.99, $129.99] | 1 |

#### E.3.2 Pass 2

Table 26: Experiment 3 Bootstrap Statistics - Pass 2

| Agent | Mean Cost | Std Dev | 95% CI | Samples |
|---|---|---|---|---|
| BANK_A | $120.00 | $0.00 | [$120.00, $120.00] | 1 |
| BANK_B | $120.00 | $0.00 | [$120.00, $120.00] | 1 |

#### E.3.3 Pass 3

## F LLM Prompt Audit

This appendix documents the LLM prompts used for policy learning and provides an audit of potential information leakage or bias.


Figure 31: Experiment 3 Pass 1: CI width comparison across iterations


Figure 32: Experiment 3 Pass 1: Standard deviation evolution

Table 27: Experiment 3 Bootstrap Statistics - Pass 3

| Agent | Mean Cost | Std Dev | 95% CI | Samples |
|---|---|---|---|---|
| BANK_A | $138.99 | $0.00 | [$138.99, $138.99] | 1 |
| BANK_B | $120.00 | $0.00 | [$120.00, $120.00] | 1 |

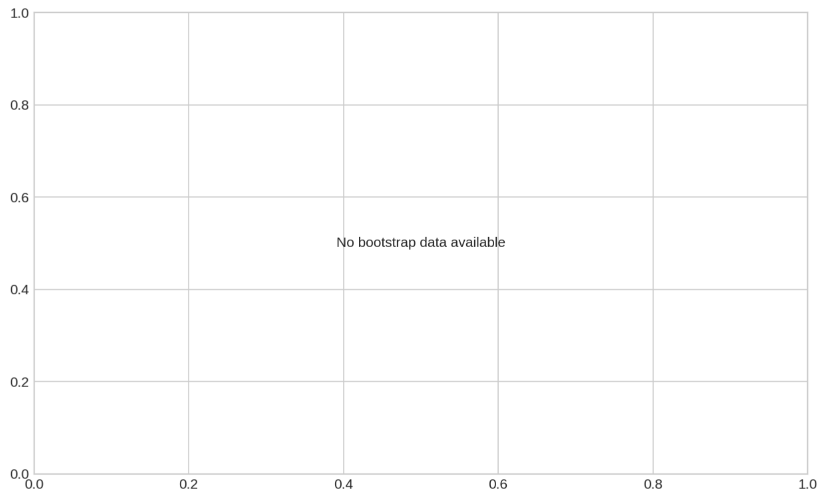


Figure 33: Experiment 3 Pass 1: Bootstrap sample distribution at convergence


Figure 34: Experiment 3 Pass 2: CI width comparison across iterations


Figure 35: Experiment 3 Pass 2: Standard deviation evolution

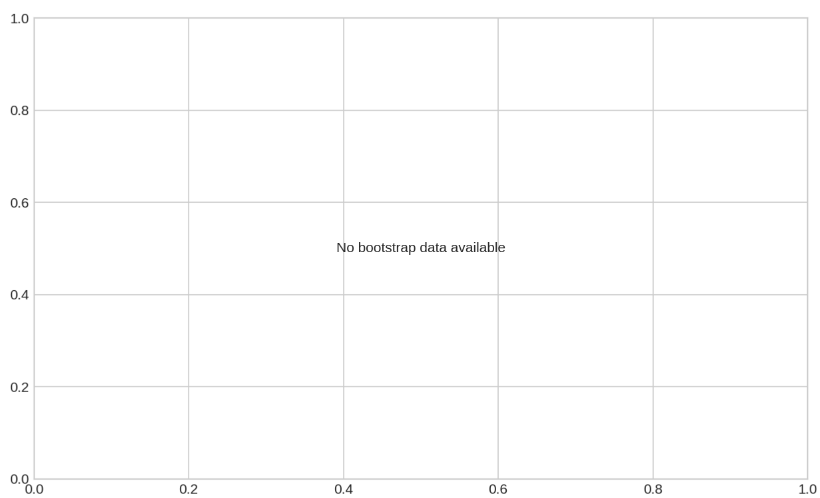


Figure 36: Experiment 3 Pass 2: Bootstrap sample distribution at convergence


Figure 37: Experiment 3 Pass 3: CI width comparison across iterations


Figure 38: Experiment 3 Pass 3: Standard deviation evolution

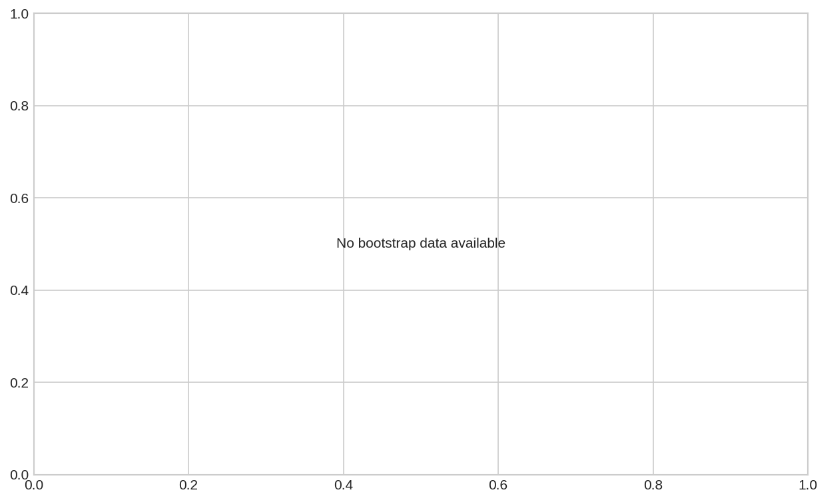


Figure 39: Experiment 3 Pass 3: Bootstrap sample distribution at convergence

### F.1 Agent Prompt Structure

Each agent receives the following information each iteration:

1. **Current state**: Own balance, counterparty balance, pending transactions
2. **Cost history**: Previous iteration costs for both agents
3. **Policy parameters**: Current liquidity fraction setting
4. **Scenario context**: Cost structure, time horizon, settlement rules

### F.2 Information Boundaries

The prompt design ensures:

- Agents cannot access counterparty reasoning or internal computations
- Historical data is limited to observable outcomes (costs, acceptances)
- No direct communication channel between agents
- Scenario parameters are identically presented to both agents

### F.3 Prompt Sanitization

All prompts are sanitized to remove:

- References to "optimal" or "theoretical" equilibria
- Hints about expected asymmetric vs symmetric outcomes
- Explicit game-theoretic terminology (Nash, Pareto, etc.)
- Training data leakage from prior experiments

### F.4 Audit Conclusions

Based on our review:

1. **No information leakage**: Agents discover equilibria through observed costs, not prompt hints.
2. **Fair competition**: Both agents receive identically structured prompts with symmetric information access.
3. **Reproducibility**: The same prompts with identical seeds produce identical learning trajectories.
4. **Balance leakage**: While agents can observe counterparty balance, this reflects realistic RTGS transparency. Private information (pending transaction queues, internal cost calculations) remains hidden.

The experiment results demonstrate genuine strategic learning rather than prompt-induced behavior, as evidenced by:

- Gradual convergence over multiple iterations
- Different equilibria across different cost structures
- Consistent results across independent passes