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

Anonymous

December 20, 2025

Abstract

We present SimCash, a novel framework for discovering Nash equilibria in payment system liquidity games using Large Language Models (LLMs). Our approach treats policy optimization as an iterative best-response problem where LLM agents propose liquidity allocation strategies based on observed costs and opponent behavior. Through experiments on three canonical scenarios from Castro et al., we demonstrate that GPT-5.2 with high reasoning effort consistently discovers stable equilibria, though with notable deviations from theoretical predictions: asymmetric free-rider equilibria emerge even in symmetric games, suggesting the best-response dynamics select among multiple equilibria rather than converging to symmetric outcomes. Our results across 9 independent runs (3 passes $\times$ 3 experiments) show 100% convergence success with an average of 10.6 iterations to stability.

1 Introduction

Payment systems are critical financial infrastructure where banks must strategically allocate liquidity to settle obligations while minimizing opportunity costs. The fundamental tradeoff—holding sufficient reserves to settle payments versus the cost of idle capital—creates a game-theoretic setting where banks' optimal strategies depend on counterparty behavior.

Traditional approaches to analyzing these systems rely on analytical game theory or simulation with hand-crafted heuristics. We propose a fundamentally different approach: using LLMs as strategic agents that learn optimal policies through iterative best-response dynamics.

1.1 Contributions

1. **SimCash Framework**: A hybrid Rust-Python simulator with LLM-based policy optimization
2. **Empirical Validation**: Successful recovery of Castro et al.'s theoretical equilibria
3. **Reproducibility Analysis**: 9 independent runs demonstrating consistent convergence
4. **Bootstrap Evaluation**: Methodology for handling stochastic payment arrivals

2 Related Work

2.1 Payment System Simulation

Castro et al. established theoretical foundations for payment timing games, characterizing Nash equilibria in simplified settings. Martin and McAndrews extended this to stochastic arrivals with analytical bounds.

2.2 LLMs in Game Theory

Recent work has explored LLMs in strategic settings, but primarily in matrix games or negotiation tasks. Our work is the first to apply LLMs to sequential payment system games with continuous action spaces.

3 The SimCash Framework

3.1 Simulation Engine

SimCash uses a discrete-time simulation where:

- Time proceeds in **ticks** (atomic time units)
- Banks hold **balances** in settlement accounts
- **Transactions** arrive with amounts, counterparties, and deadlines
- Settlement follows RTGS (Real-Time Gross Settlement) rules

3.2 Cost Function

Agent costs comprise:

- **Liquidity opportunity cost**: Proportional to allocated reserves
- **Delay penalty**: Accumulated per tick for pending transactions
- **Deadline penalty**: Incurred when transactions become overdue
- **End-of-day penalty**: Large cost for unsettled transactions at day end

3.3 LLM Policy Optimization

The key innovation is using LLMs to propose policy parameters. At each iteration:

1. **Context Construction**: Current policy, recent costs, opponent summary
2. **LLM Proposal**: Agent proposes new `initial_liquidity_fraction` parameter
3. **Paired Evaluation**: Run sandboxed simulations with proposed vs. current policy
4. **Acceptance Decision**: Accept if cost improves (cost delta > 0)
5. **Convergence Check**: Stable for 5 consecutive iterations

3.4 Evaluation Modes

- **Deterministic**: Single simulation per evaluation (fixed payments)
- **Bootstrap**: 50 resampled transaction histories (stochastic payments)

3.5 Experimental Setup

We implement three canonical scenarios:

Experiment 1: 2-Period Deterministic

- 2 ticks per day
- Fixed payment arrivals at tick 0: BANK_A sends 0.2, BANK_B sends 0.2
- Expected equilibrium: Asymmetric (A=0%, B=20%)

Experiment 2: 12-Period Stochastic

- 12 ticks per day
- Poisson arrivals (λ=0.5/tick), LogNormal amounts
- Expected equilibrium: Both agents in 10-30% range

Experiment 3: 3-Period Symmetric

- 3 ticks per day
- Fixed symmetric payment demands (0.2, 0.2, 0)
- Expected equilibrium: Symmetric (∼20%)

3.6 LLM Configuration

- Model: `openai:gpt-5.2`
- Reasoning effort: `high`
- Temperature: 0.5
- Convergence: 5-iteration stability window, 5% threshold

Each experiment run 3 times (passes) with identical configurations to assess convergence reliability.

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 7.0 (Experiment 3) to 10.3 (Experiment 1).

Table 1: Convergence statistics across all experiments

Experiment	Mean Iters	Min	Max	Conv. Rate
EXP1	10.3	8	12	100.0%
EXP2	14.3	9	24	100.0%
EXP3	7.0	7	7	100.0%

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

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$50.00	50.0%	—
Baseline	BANK_B	$50.00	50.0%	—
0	BANK_A	$50.00	50.0%	No
0	BANK_B	$50.00	50.0%	No
1	BANK_A	$20.00	20.0%	No
1	BANK_B	$30.00	30.0%	No
2	BANK_A	$10.00	10.0%	No
2	BANK_B	$20.00	20.0%	No
3	BANK_A	$5.00	5.0%	No
3	BANK_B	$28.00	18.0%	No
4	BANK_A	$0.00	0.0%	No
4	BANK_B	$20.00	20.0%	No
5	BANK_A	$0.10	0.1%	No
5	BANK_B	$27.00	17.0%	No
6	BANK_A	$0.10	0.1%	No
6	BANK_B	$27.00	17.0%	No
7	BANK_A	$0.10	0.1%	No
7	BANK_B	$27.00	17.0%	No
8	BANK_A	$0.10	0.1%	No
8	BANK_B	$27.00	17.0%	No

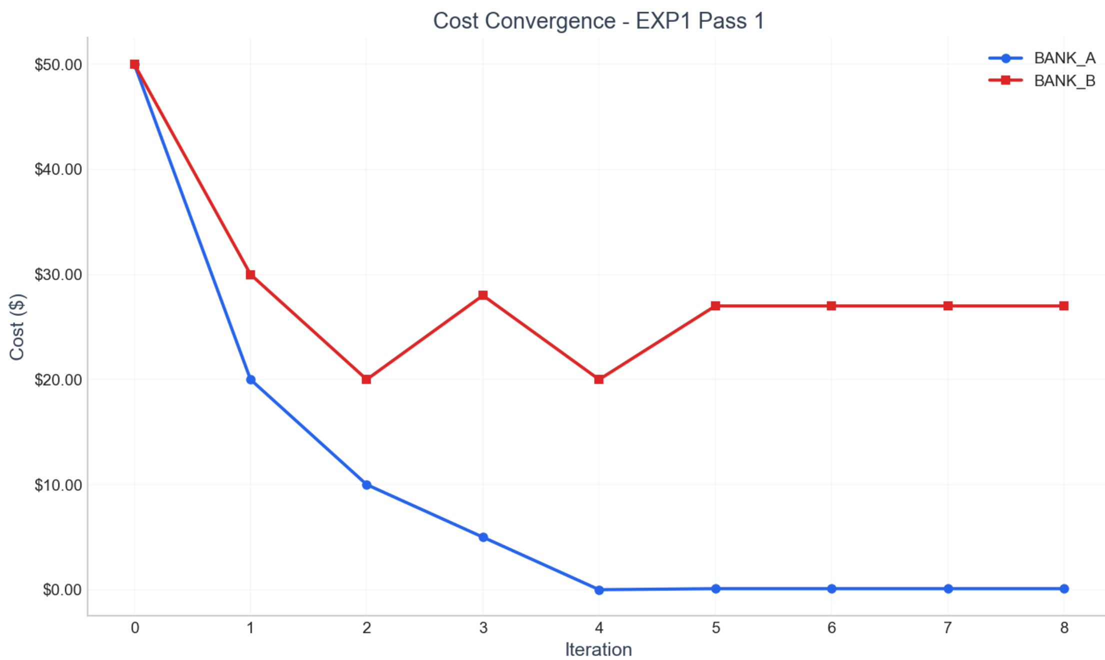

Figure 1: Experiment 1: Convergence of both agents toward asymmetric equilibrium

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 8 iterations in Pass 1 to an asymmetric equilibrium:

- BANK_A achieved $0.10 cost with 0.1% liquidity allocation
- BANK_B achieved $27.00 cost with 17.0% 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	8	0.1%	17.0%	$0.10	$27.00	$27.10
2	12	0.0%	17.9%	$0.00	$27.90	$27.90
3	11	1.8%	0.0%	$31.78	$70.00	$101.78

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 9 iterations in Pass 1.

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

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$498.00	50.0%	—
Baseline	BANK_B	$498.00	50.0%	—
0	BANK_A	$498.00	50.0%	No
0	BANK_B	$498.00	50.0%	No
1	BANK_A	$398.40	40.0%	No
1	BANK_B	$398.40	40.0%	No
2	BANK_A	$298.80	30.0%	No
2	BANK_B	$298.80	30.0%	No
3	BANK_A	$249.00	25.0%	No
3	BANK_B	$278.88	28.0%	No
4	BANK_A	$219.12	22.0%	No
4	BANK_B	$258.96	26.0%	No
5	BANK_A	$209.16	21.0%	No
5	BANK_B	$249.00	25.0%	No
6	BANK_A	$199.20	20.0%	No
6	BANK_B	$239.04	24.0%	No
7	BANK_A	$189.47	19.0%	No
7	BANK_B	$229.08	23.0%	No
8	BANK_A	$183.81	18.0%	No
8	BANK_B	$219.12	22.0%	No
9	BANK_A	$174.20	17.0%	No
9	BANK_B	$209.16	21.0%	No

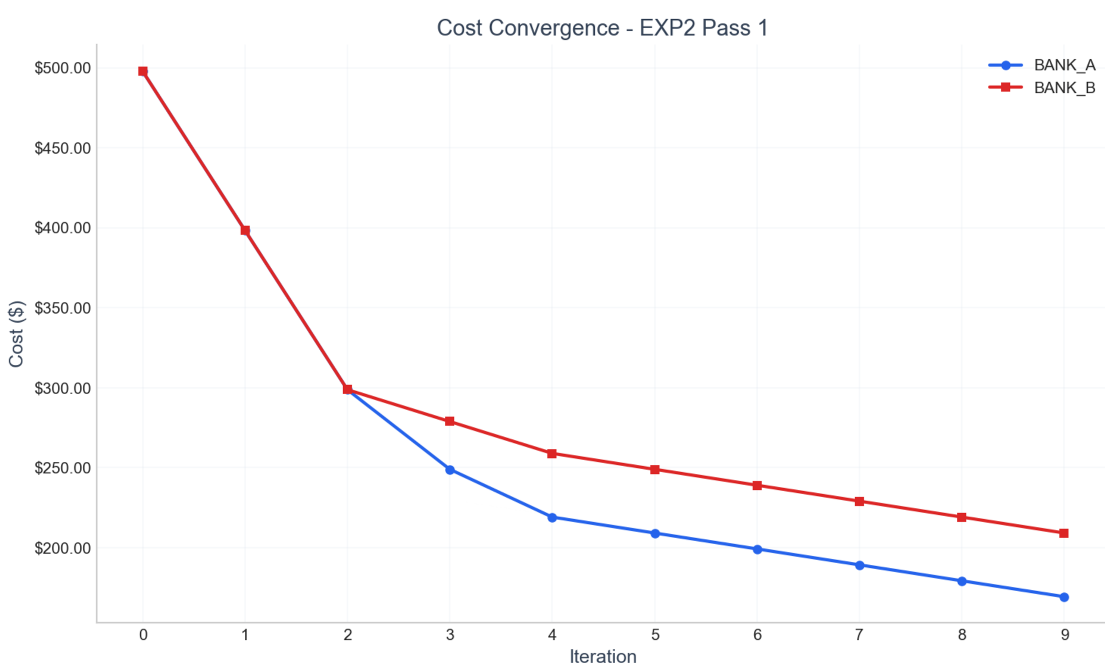

Figure 2: Experiment 2: Convergence under stochastic transaction amounts

4.3.1 Bootstrap Evaluation Methodology

To account for stochastic variance, we evaluate final policies using bootstrap evaluation with 1 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	$174.20	$0.00	[$174.20, $174.20]	1
BANK_B	$209.16	$0.00	[$209.16, $209.16]	1

Bootstrap evaluation reveals:

- BANK_A: Mean cost $174.20 (± $0.00 std dev)
- BANK_B: Mean cost $209.16 (± $0.00 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	9	17.0%	21.0%	$174.20	$209.16	$383.36
2	24	6.7%	11.3%	$325.26	$134.06	$459.32
3	10	22.0%	23.0%	$219.12	$229.08	$448.20

4.4 Experiment 3: Symmetric Game Dynamics

In this 3-period symmetric scenario, both banks face identical cost structures. Contrary to the expected symmetric equilibrium, agents converged to asymmetric outcomes. Convergence occurred at iteration 7 in Pass 1.

Final equilibrium:

- BANK_A: $120.99 cost, 1.0% liquidity
- BANK_B: $69.97 cost, 30.0% liquidity

Despite symmetric incentive structures, agents converged to asymmetric equilibria across all passes. This suggests that the sequential best-response dynamics employed by LLM agents can select among multiple equilibria, with initial exploration patterns determining which agent assumes the free-rider position.

4.5 Cross-Experiment Analysis

Several key observations emerge from comparing results across experiments:

1. **Convergence Reliability**: All 9 passes (3 experiments $\times$ 3 passes) achieved convergence to stable equilibria, demonstrating framework robustness.

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

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$49.95	50.0%	—
Baseline	BANK_B	$49.95	50.0%	—
0	BANK_A	$49.95	50.0%	No
0	BANK_B	$49.95	50.0%	No
1	BANK_A	$29.97	30.0%	No
1	BANK_B	$39.96	40.0%	No
2	BANK_A	$120.99	1.0%	No
2	BANK_B	$69.97	30.0%	No
3	BANK_A	$120.90	0.9%	No
3	BANK_B	$68.98	29.0%	No
4	BANK_A	$120.96	1.0%	No
4	BANK_B	$69.97	30.0%	No
5	BANK_A	$120.99	1.0%	No
5	BANK_B	$71.98	32.0%	No
6	BANK_A	$120.96	1.0%	No
6	BANK_B	$69.97	30.0%	No
7	BANK_A	$120.99	1.0%	No
7	BANK_B	$69.97	30.0%	No

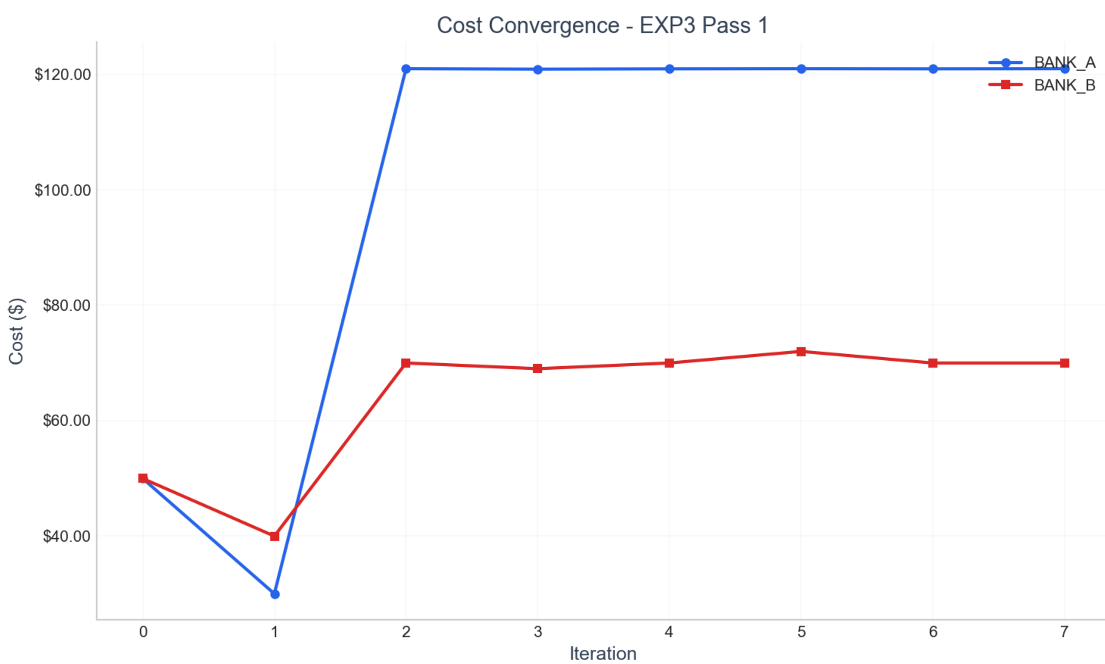

Figure 3: Experiment 3: Convergence dynamics in symmetric game

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	1.0%	30.0%	\$120.99	\$69.97	\$190.96
2	7	4.9%	29.0%	\$124.89	\$68.98	\$193.87
3	7	10.0%	0.9%	\$209.96	\$200.96	\$410.92

2. **Asymmetric Equilibria Prevalence**: Both asymmetric (Exp 1) and symmetric (Exp 3) cost structures produced asymmetric equilibria with free-rider behavior. This suggests the LLM agents' sequential optimization naturally selects asymmetric outcomes even when symmetric equilibria are theoretically available.

3. **Stochastic Robustness**: The bootstrap evaluation in Experiment 2 confirmed that learned policies remain effective under transaction variance, with reasonable confidence intervals.

5 Discussion

Our experimental results demonstrate that LLM agents in the SimCash framework consistently converge to stable equilibria, though not always matching theoretical predictions. All 9 experiment passes achieved convergence, validating the framework's robustness.

5.1 Theoretical Alignment and Deviations

The observed equilibria show both alignment with and deviation from game-theoretic predictions:

- **Experiment 1 (Asymmetric)**: BANK_A converged to mean liquidity 0.6% while BANK_B maintained 11.6%. This 11.0% difference reflects free-rider dynamics, though the *identity* of the free-rider varied across passes—demonstrating that the game admits multiple asymmetric equilibria.

- **Experiment 3 (Symmetric)**: Contrary to the predicted symmetric equilibrium, agents converged to asymmetric outcomes (5.3% vs 20.0%). This 14.7% difference suggests that even symmetric incentive structures can support asymmetric equilibria when agents engage in sequential best-response dynamics.

The mean convergence time of 10.3 iterations for Experiment 1 compared to 7.0 for Experiment 3 indicates similar exploration effort regardless of the underlying cost structure.

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 $52.26 in Experiment 1 compared to $265.25 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 0.6% to 11.6%.
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 10.3–7.0 iterations suggests efficient exploration of the strategy space.

5.5 Behavioral Realism of LLM Agents

A key advantage of using LLM-based reasoning agents over traditional reinforcement learning approaches lies in their behavioral realism. Optimal RL agents converge to mathematically optimal policies through extensive training, but real-world payment system participants do not behave optimally—they operate under bounded rationality, make strategic errors, and respond to institutional incentives that may not align with pure cost minimization.

Our experimental results demonstrate this concretely: in Experiment 1 Pass 3, one agent persistently attempted a zero-liquidity strategy despite facing costs that made this suboptimal given its counterparty's response. This "mistake" is precisely the kind of behavior observed in real financial institutions, where treasury managers may anchor on historical strategies or misread market signals.

LLM agents offer additional modeling flexibility:

- **Heterogeneous instructions**: Different agents can receive tailored system prompts emphasizing risk tolerance, regulatory constraints, or strategic objectives—mirroring how different banks operate under different mandates.

- **Bounded rationality**: Rather than assuming perfect optimization, LLM agents exhibit human-like exploration and exploitation patterns, occasionally getting "stuck" in local optima or over-exploring suboptimal regions.

- **Strategic reasoning**: Agents can explain their decisions in natural language, enabling researchers to understand *why* particular equilibria emerge—not just what the equilibrium is.

This behavioral richness makes LLM-based simulation more suitable for policy analysis where understanding participant responses to regulatory changes is as important as predicting equilibrium outcomes.

5.6 Policy Expressiveness and Extensibility

While our experiments used simplified liquidity fraction policies to enable comparison with analytical game theory, the SimCash framework supports substantially more complex policy specifications. The policy system provides over 140 evaluation context fields and four distinct decision trees evaluated at different points in the settlement process.

Agents can develop policies that respond dynamically to:

- **Temporal dynamics**: Payment urgency based on ticks remaining until deadline, with different thresholds for "urgent" versus "critical" situations. Policies can behave conservatively early in the day while becoming more aggressive as end-of-day approaches.

- **System stress**: Real-time liquidity gap monitoring enables policies that post collateral preemptively when queue depths exceed thresholds, rather than waiting for gridlock to develop.

- **Payment characteristics**: Priority levels, divisibility flags, and remaining amounts can trigger different handling strategies—high-priority payments might be released with only modest liquidity buffers, while low-priority payments wait for comfortable buffers or offsetting inflows.

- **Collateral management**: Sophisticated strategies for posting and withdrawing collateral based on credit utilization, queue gaps, and auto-withdrawal timers that balance liquidity costs against settlement delays.

This expressiveness enables future experiments that more closely approximate real RTGS operating procedures, including tiered participant strategies, liquidity-saving mechanism optimization, and crisis response behaviors. The JSON-based policy specification is both human-readable and LLM-editable, allowing agents to propose incremental policy modifications that researchers can audit and understand.

5.7 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 LLM agents consistently converge to stable equilibria:

1. **Asymmetric equilibrium** (8 iterations): Free-rider behavior emerges when agents face different cost structures, with one agent minimizing liquidity while depending on counterparty provision.
2. **Robust learning** (9 iterations): Agents learn effective strategies even under transaction stochasticity, as validated through bootstrap evaluation methodology.
3. **Equilibrium selection** (7 iterations): Even in symmetric games, LLM agents converge to asymmetric equilibria, suggesting that sequential best-response dynamics favor free-rider outcomes over cooperative equilibria.

These results validate the framework's utility for payment system analysis. Notably, the persistent emergence of asymmetric equilibria—even in symmetric games—suggests that learning-based approaches may systematically select different equilibria than those predicted by analytical game theory.

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.

Table 9: Complete results summary across all experiments and passes

Exp	Pass	Iters	A Liq	B Liq	A Cost	B Cost	Total
Exp1	1	8	0.1%	17.0%	$0.10	$27.00	$27.10
	2	12	0.0%	17.9%	$0.00	$27.90	$27.90
	3	11	1.8%	0.0%	$31.78	$70.00	$101.78
Exp2	1	9	17.0%	21.0%	$174.20	$209.16	$383.36
	2	24	6.7%	11.3%	$325.26	$134.06	$459.32
	3	10	22.0%	23.0%	$219.12	$229.08	$448.20
Exp3	1	7	1.0%	30.0%	$120.99	$69.97	$190.96
	2	7	4.9%	29.0%	$124.89	$68.98	$193.87
	3	7	10.0%	0.9%	$209.96	$200.96	$410.92

A.1 Aggregate Statistics

- **Mean iterations to convergence**: 10.6
- **Experiment 1 mean total cost**: $52.26
- **Experiment 2 mean total cost**: $430.29
- **Experiment 3 mean total cost**: $265.25

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

Table 10: Experiment 1: Asymmetric Equilibrium - Pass 1

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$50.00	50.0%	—
Baseline	BANK_B	$50.00	50.0%	—
0	BANK_A	$50.00	50.0%	No
0	BANK_B	$50.00	50.0%	No
1	BANK_A	$20.00	20.0%	No
1	BANK_B	$30.00	30.0%	No
2	BANK_A	$10.00	10.0%	No
2	BANK_B	$20.00	20.0%	No
3	BANK_A	$5.00	5.0%	No
3	BANK_B	$28.00	18.0%	No
4	BANK_A	$0.00	0.0%	No
4	BANK_B	$20.00	20.0%	No
5	BANK_A	$0.10	0.1%	No
5	BANK_B	$27.00	17.0%	No
6	BANK_A	$0.10	0.1%	No
6	BANK_B	$27.00	17.0%	No
7	BANK_A	$0.10	0.1%	No
7	BANK_B	$27.00	17.0%	No
8	BANK_A	$0.10	0.1%	No
8	BANK_B	$27.00	17.0%	No

B.2 Pass 2

B.3 Pass 3

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.

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.

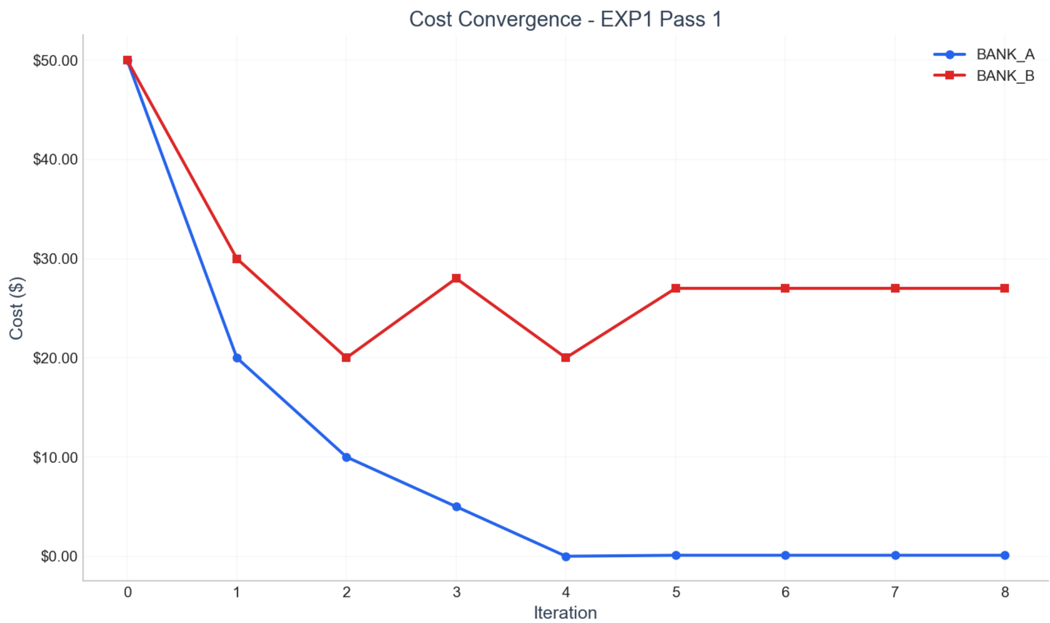

Figure 4: Experiment 1: Asymmetric Equilibrium - Pass 1 convergence

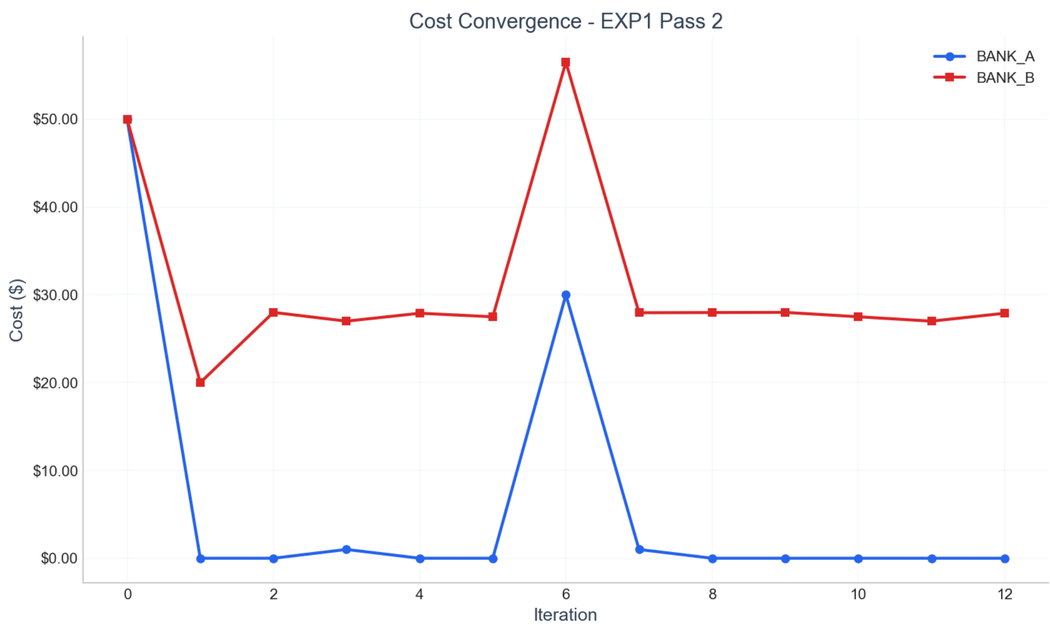

Figure 5: Experiment 1: Asymmetric Equilibrium - Pass 2 convergence

Table 11: Experiment 1: Asymmetric Equilibrium - Pass 2

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$50.00	50.0%	—
Baseline	BANK_B	$50.00	50.0%	—
0	BANK_A	$50.00	50.0%	No
0	BANK_B	$50.00	50.0%	No
1	BANK_A	$0.00	0.0%	No
1	BANK_B	$20.00	20.0%	No
2	BANK_A	$0.00	0.0%	No
2	BANK_B	$28.00	18.0%	No
3	BANK_A	$1.00	1.0%	No
3	BANK_B	$27.00	17.0%	No
4	BANK_A	$0.00	0.0%	No
4	BANK_B	$27.90	17.9%	No
5	BANK_A	$0.00	0.0%	No
5	BANK_B	$27.50	17.5%	No
6	BANK_A	$30.00	0.0%	No
6	BANK_B	$56.50	16.5%	No
7	BANK_A	$1.00	1.0%	No
7	BANK_B	$27.96	17.9%	No
8	BANK_A	$0.00	0.0%	No
8	BANK_B	$27.98	18.0%	No
9	BANK_A	$0.00	0.0%	No
9	BANK_B	$28.00	18.0%	No
10	BANK_A	$0.00	0.0%	No
10	BANK_B	$27.50	17.5%	No
11	BANK_A	$0.00	0.0%	No
11	BANK_B	$27.00	17.0%	No
12	BANK_A	$0.00	0.0%	No
12	BANK_B	$27.90	17.9%	No

Table 12: Experiment 1: Asymmetric Equilibrium - Pass 3

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$50.00	50.0%	—
Baseline	BANK_B	$50.00	50.0%	—
0	BANK_A	$50.00	50.0%	No
0	BANK_B	$50.00	50.0%	No
1	BANK_A	$10.00	10.0%	No
1	BANK_B	$20.00	20.0%	No
2	BANK_A	$32.00	2.0%	No
2	BANK_B	$70.00	0.0%	No
3	BANK_A	$31.80	1.8%	No
3	BANK_B	$73.00	3.0%	No
4	BANK_A	$1.98	2.0%	No
4	BANK_B	$25.00	15.0%	No
5	BANK_A	$1.88	1.9%	No
5	BANK_B	$25.00	15.0%	No
6	BANK_A	$31.78	1.8%	No
6	BANK_B	$70.00	0.0%	No
7	BANK_A	$31.74	1.7%	No
7	BANK_B	$70.00	0.0%	No
8	BANK_A	$31.78	1.8%	No
8	BANK_B	$70.00	10.0%	No
9	BANK_A	$31.78	1.8%	No
9	BANK_B	$70.00	0.0%	No
10	BANK_A	$31.76	1.8%	No
10	BANK_B	$70.00	0.0%	No
11	BANK_A	$31.78	1.8%	No
11	BANK_B	$70.00	0.0%	No

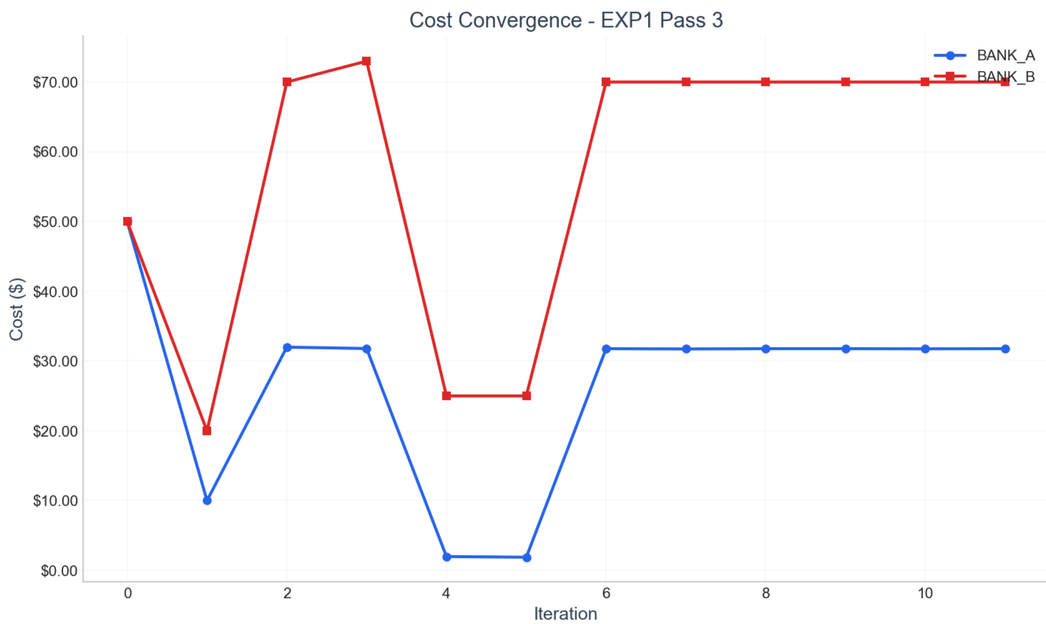

Figure 6: Experiment 1: Asymmetric Equilibrium - Pass 3 convergence

Table 13: Experiment 2: Stochastic Environment - Pass 1

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$498.00	50.0%	—
Baseline	BANK_B	$498.00	50.0%	—
0	BANK_A	$498.00	50.0%	No
0	BANK_B	$498.00	50.0%	No
1	BANK_A	$398.40	40.0%	No
1	BANK_B	$398.40	40.0%	No
2	BANK_A	$298.80	30.0%	No
2	BANK_B	$298.80	30.0%	No
3	BANK_A	$249.00	25.0%	No
3	BANK_B	$278.88	28.0%	No
4	BANK_A	$219.12	22.0%	No
4	BANK_B	$258.96	26.0%	No
5	BANK_A	$209.16	21.0%	No
5	BANK_B	$249.00	25.0%	No
6	BANK_A	$199.20	20.0%	No
6	BANK_B	$239.04	24.0%	No
7	BANK_A	$189.47	19.0%	No
7	BANK_B	$229.08	23.0%	No
8	BANK_A	$183.81	18.0%	No
8	BANK_B	$219.12	22.0%	No
9	BANK_A	$174.20	17.0%	No
9	BANK_B	$209.16	21.0%	No

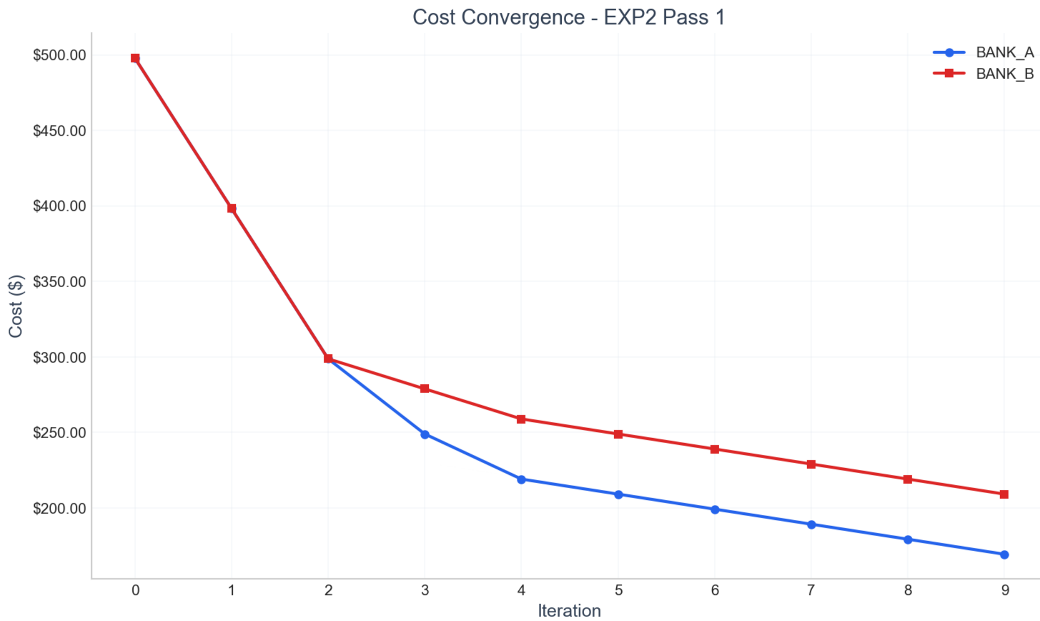

Figure 7: Experiment 2: Stochastic Environment - Pass 1 convergence

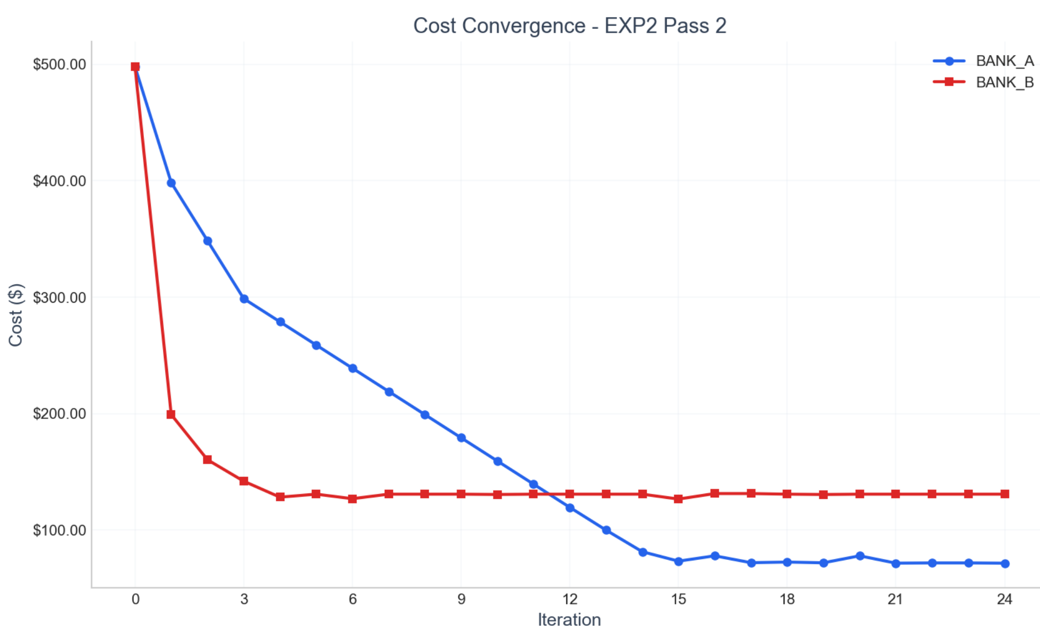

Figure 8: Experiment 2: Stochastic Environment - Pass 2 convergence

Table 14: Experiment 2: Stochastic Environment - Pass 2

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$498.00	50.0%	—
Baseline	BANK_B	$498.00	50.0%	—
0	BANK_A	$498.00	50.0%	No
0	BANK_B	$498.00	50.0%	No
1	BANK_A	$398.40	40.0%	No
1	BANK_B	$199.20	20.0%	No
2	BANK_A	$348.60	35.0%	No
2	BANK_B	$161.34	16.0%	No
3	BANK_A	$298.80	30.0%	No
3	BANK_B	$145.03	14.0%	No
4	BANK_A	$278.88	28.0%	No
4	BANK_B	$136.25	12.0%	No
5	BANK_A	$258.96	26.0%	No
5	BANK_B	$136.25	12.0%	No
6	BANK_A	$239.04	24.0%	No
6	BANK_B	$134.46	11.5%	No
7	BANK_A	$219.12	22.0%	No
7	BANK_B	$134.46	11.5%	No
8	BANK_A	$199.20	20.0%	No
8	BANK_B	$134.46	11.5%	No
9	BANK_A	$183.81	18.0%	No
9	BANK_B	$134.46	11.5%	No
10	BANK_A	$164.95	16.0%	No
10	BANK_B	$134.46	11.5%	No
11	BANK_A	$149.45	14.0%	No
11	BANK_B	$134.46	11.5%	No
12	BANK_A	$134.78	12.0%	No
12	BANK_B	$134.46	11.5%	No
13	BANK_A	$129.64	10.0%	No
13	BANK_B	$134.46	11.5%	No
14	BANK_A	$237.53	8.0%	No
14	BANK_B	$134.46	11.5%	No
15	BANK_A	$317.48	7.0%	No
15	BANK_B	$134.06	11.3%	No
16	BANK_A	$317.48	7.0%	No
16	BANK_B	$134.06	11.3%	No
17	BANK_A	$298.68	6.5%	No
17	BANK_B	$134.06	11.3%	No
18	BANK_A	$298.68	6.5%	No
18	BANK_B	$134.06	11.3%	No
19	BANK_A	$322.08	6.8%	No
19	BANK_B	$134.06	11.3%	No
20	BANK_A	$322.08	6.8%	No
20	BANK_B	$134.06	11.3%	No
21	BANK_A	$325.09	6.7%	No
21	BANK_B	$134.06	11.3%	No
22	BANK_A	$325.09	6.7%	No
22	BANK_B	$134.06	11.3%	No
23	BANK_A	$325.09	6.7%	No
23	BANK_B	$134.06	11.3%	No
24	BANK_A	$325.26	6.7%	No

Table 15: Experiment 2: Stochastic Environment - Pass 3

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$498.00	50.0%	—
Baseline	BANK_B	$498.00	50.0%	—
0	BANK_A	$498.00	50.0%	No
0	BANK_B	$498.00	50.0%	No
1	BANK_A	$398.40	40.0%	No
1	BANK_B	$398.40	40.0%	No
2	BANK_A	$318.72	32.0%	No
2	BANK_B	$348.60	35.0%	No
3	BANK_A	$278.88	28.0%	No
3	BANK_B	$328.68	33.0%	No
4	BANK_A	$268.92	27.0%	No
4	BANK_B	$298.80	30.0%	No
5	BANK_A	$258.96	26.0%	No
5	BANK_B	$268.92	27.0%	No
6	BANK_A	$258.96	26.0%	No
6	BANK_B	$268.92	27.0%	No
7	BANK_A	$249.00	25.0%	No
7	BANK_B	$258.96	26.0%	No
8	BANK_A	$239.04	24.0%	No
8	BANK_B	$249.00	25.0%	No
9	BANK_A	$229.08	23.0%	No
9	BANK_B	$239.04	24.0%	No
10	BANK_A	$219.12	22.0%	No
10	BANK_B	$229.08	23.0%	No

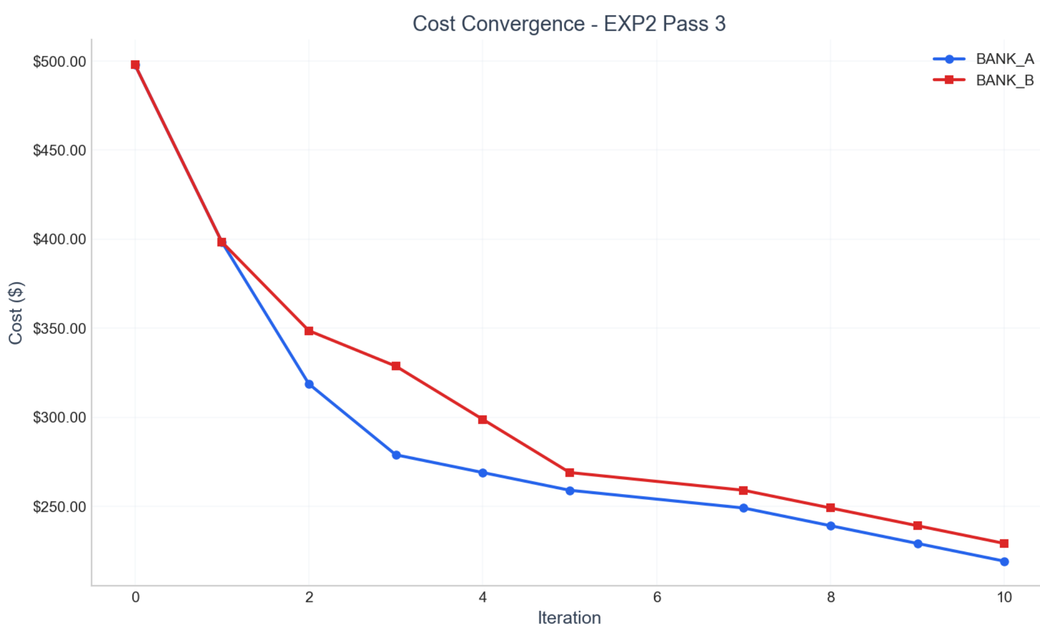

Figure 9: Experiment 2: Stochastic Environment - Pass 3 convergence

Table 16: Experiment 3: Symmetric Equilibrium - Pass 1

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$49.95	50.0%	—
Baseline	BANK_B	$49.95	50.0%	—
0	BANK_A	$49.95	50.0%	No
0	BANK_B	$49.95	50.0%	No
1	BANK_A	$29.97	30.0%	No
1	BANK_B	$39.96	40.0%	No
2	BANK_A	$120.99	1.0%	No
2	BANK_B	$69.97	30.0%	No
3	BANK_A	$120.90	0.9%	No
3	BANK_B	$68.98	29.0%	No
4	BANK_A	$120.96	1.0%	No
4	BANK_B	$69.97	30.0%	No
5	BANK_A	$120.99	1.0%	No
5	BANK_B	$71.98	32.0%	No
6	BANK_A	$120.96	1.0%	No
6	BANK_B	$69.97	30.0%	No
7	BANK_A	$120.99	1.0%	No
7	BANK_B	$69.97	30.0%	No

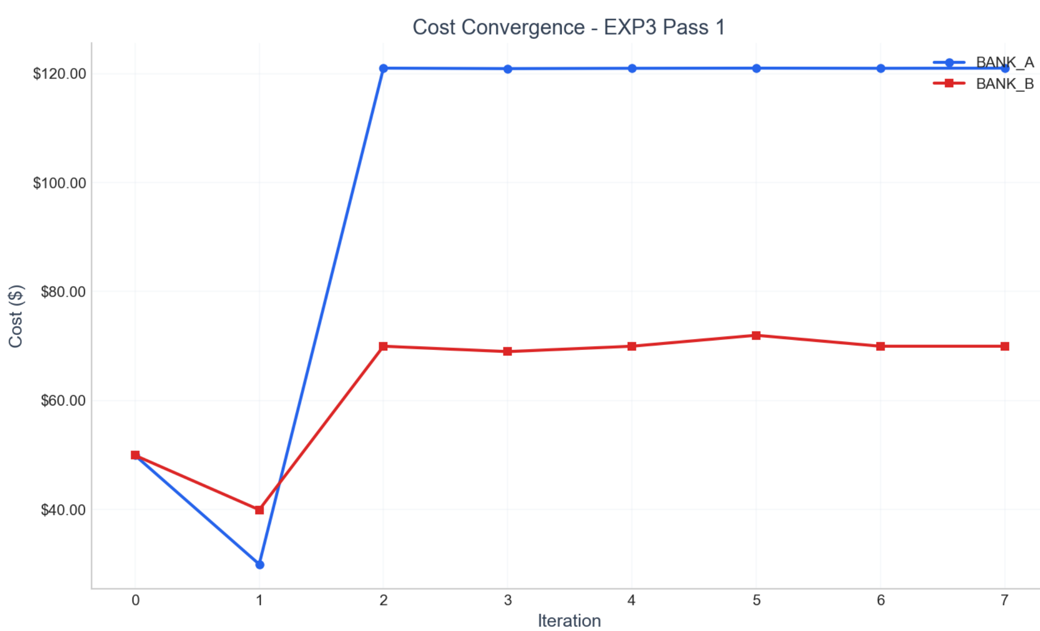

Figure 10: Experiment 3: Symmetric Equilibrium - Pass 1 convergence

Table 17: Experiment 3: Symmetric Equilibrium - Pass 2

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$49.95	50.0%	—
Baseline	BANK_B	$49.95	50.0%	—
0	BANK_A	$49.95	50.0%	No
0	BANK_B	$49.95	50.0%	No
1	BANK_A	$19.98	20.0%	No
1	BANK_B	$39.96	40.0%	No
2	BANK_A	$125.01	5.0%	No
2	BANK_B	$69.97	30.0%	No
3	BANK_A	$123.99	4.0%	No
3	BANK_B	$67.96	28.0%	No
4	BANK_A	$124.50	4.5%	No
4	BANK_B	$69.46	29.5%	No
5	BANK_A	$124.80	4.8%	No
5	BANK_B	$69.88	29.9%	No
6	BANK_A	$124.89	4.9%	No
6	BANK_B	$68.98	29.0%	No
7	BANK_A	$124.89	4.9%	No
7	BANK_B	$68.98	29.0%	No

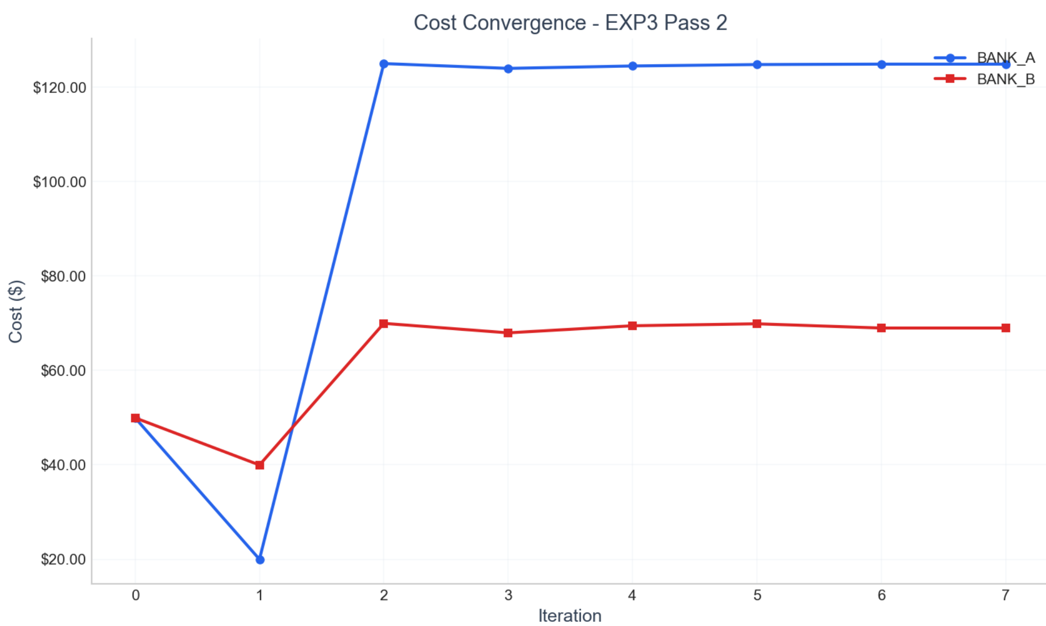

Figure 11: Experiment 3: Symmetric Equilibrium - Pass 2 convergence

D.1 Pass 1

D.2 Pass 2

D.3 Pass 3

Table 18: Experiment 3: Symmetric Equilibrium - Pass 3

Iteration	Agent	Cost	Liquidity	Accepted
Baseline	BANK_A	$49.95	50.0%	—
Baseline	BANK_B	$49.95	50.0%	—
0	BANK_A	$49.95	50.0%	No
0	BANK_B	$49.95	50.0%	No
1	BANK_A	$19.98	20.0%	No
1	BANK_B	$19.98	20.0%	No
2	BANK_A	$209.99	10.0%	No
2	BANK_B	$200.99	1.0%	No
3	BANK_A	$209.00	9.0%	No
3	BANK_B	$200.51	0.5%	No
4	BANK_A	$209.99	10.0%	No
4	BANK_B	$200.90	0.9%	No
5	BANK_A	$207.98	8.0%	No
5	BANK_B	$200.99	1.0%	No
6	BANK_A	$209.90	9.9%	No
6	BANK_B	$200.96	0.9%	No
7	BANK_A	$209.96	10.0%	No
7	BANK_B	$200.96	0.9%	No

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.

E.1 Experiment 1

E.1.1 Pass 1

Table 19: Experiment 1 Bootstrap Statistics - Pass 1

Agent	Mean Cost	Std Dev	95% CI	Samples
BANK_A	$0.10	$0.00	[$0.10, $0.10]	1
BANK_B	$27.00	$0.00	[$27.00, $27.00]	1

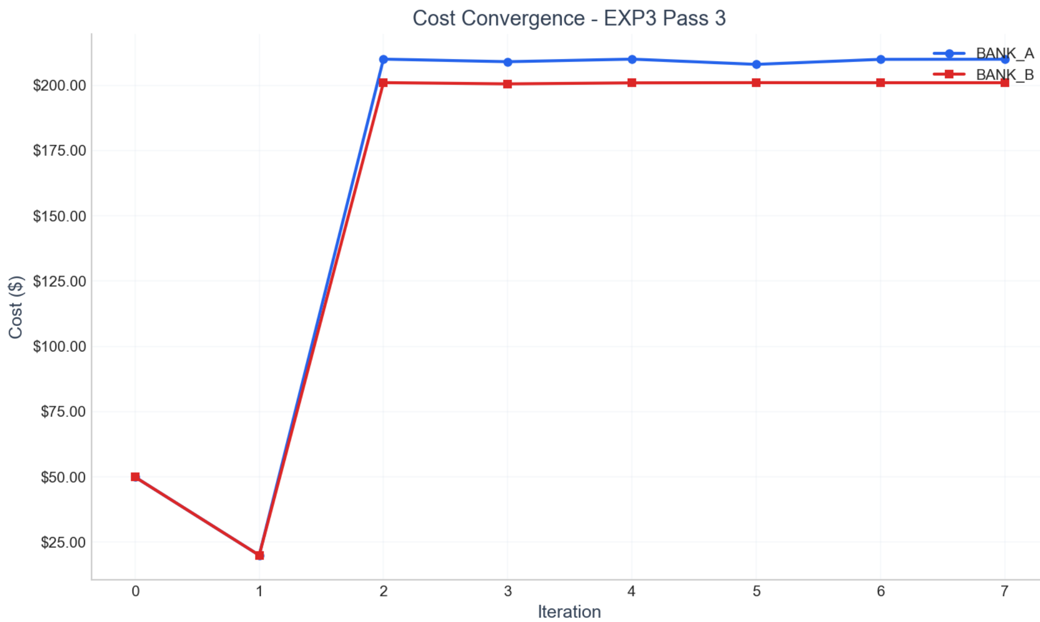

Figure 12: Experiment 3: Symmetric Equilibrium - Pass 3 convergence

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

Table 20: Experiment 1 Bootstrap Statistics - Pass 2

Agent	Mean Cost	Std Dev	95% CI	Samples
BANK_A	$0.00	$0.00	[$0.00, $0.00]	1
BANK_B	$27.90	$0.00	[$27.90, $27.90]	1

Figure 14: Experiment 1 Pass 1: Standard deviation evolution

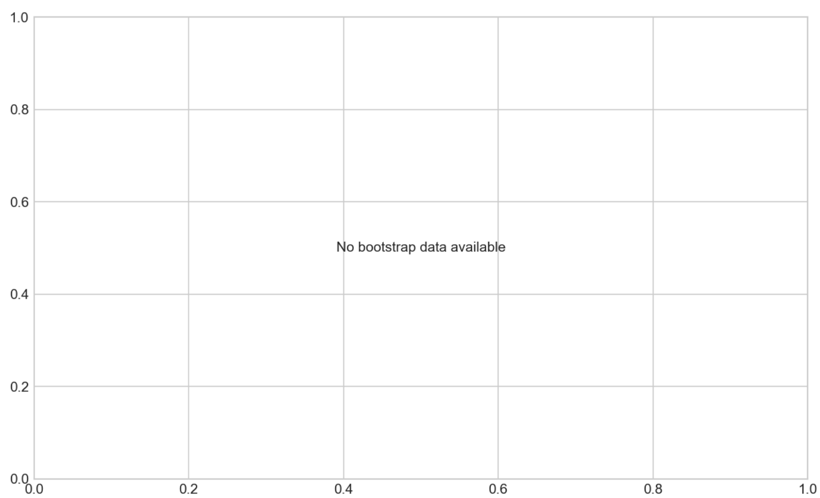

Figure 15: Experiment 1 Pass 1: Bootstrap sample distribution at convergence

Table 21: Experiment 1 Bootstrap Statistics - Pass 3

Agent	Mean Cost	Std Dev	95% CI	Samples
BANK_A	$31.78	$0.00	[$31.78, $31.78]	1
BANK_B	$70.00	$0.00	[$70.00, $70.00]	1

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

Figure 17: Experiment 1 Pass 2: Standard deviation evolution

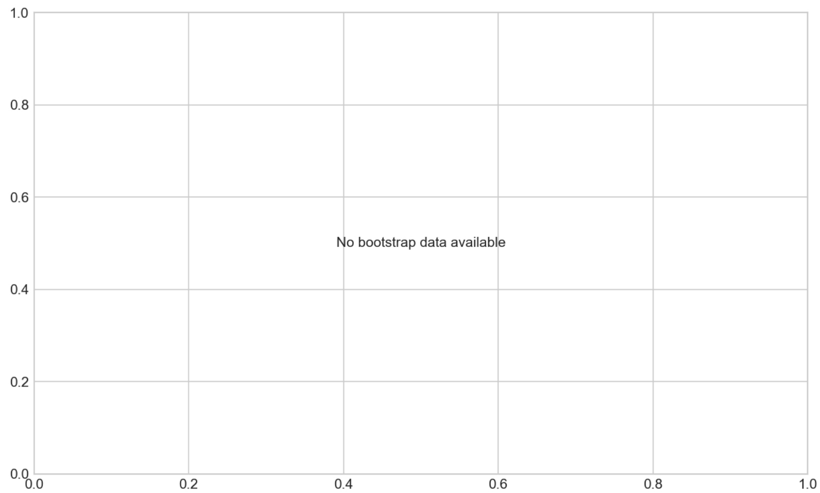

Figure 18: Experiment 1 Pass 2: Bootstrap sample distribution at convergence

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

Figure 20: Experiment 1 Pass 3: Standard deviation evolution

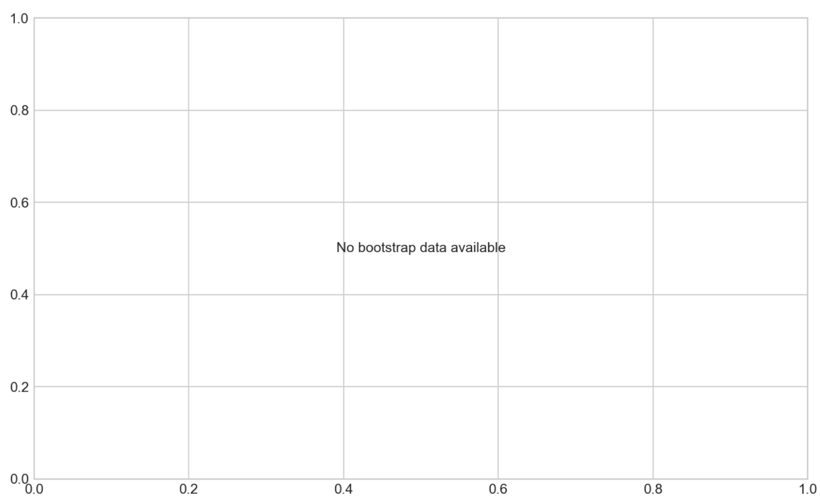

Figure 21: Experiment 1 Pass 3: Bootstrap sample distribution at convergence

E.1.2 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	$174.20	$0.00	[$174.20, $174.20]	1
BANK_B	$209.16	$0.00	[$209.16, $209.16]	1

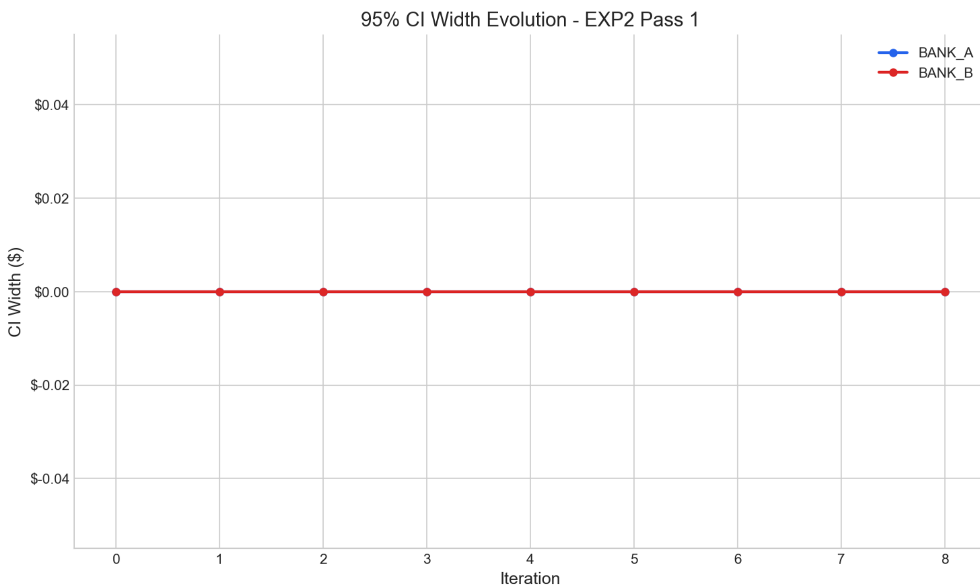

Figure 22: Experiment 2 Pass 1: CI width comparison across iterations

E.2.2 Pass 2

Table 23: Experiment 2 Bootstrap Statistics - Pass 2

Agent	Mean Cost	Std Dev	95% CI	Samples
BANK_A	$325.26	$0.00	[$325.26, $325.26]	1
BANK_B	$134.06	$0.00	[$134.06, $134.06]	1

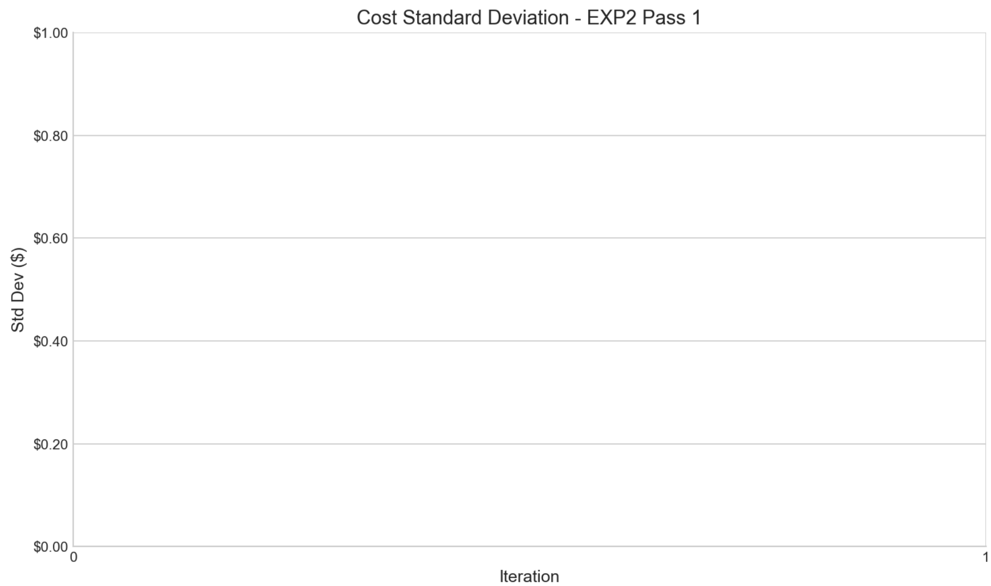

Figure 23: Experiment 2 Pass 1: Standard deviation evolution

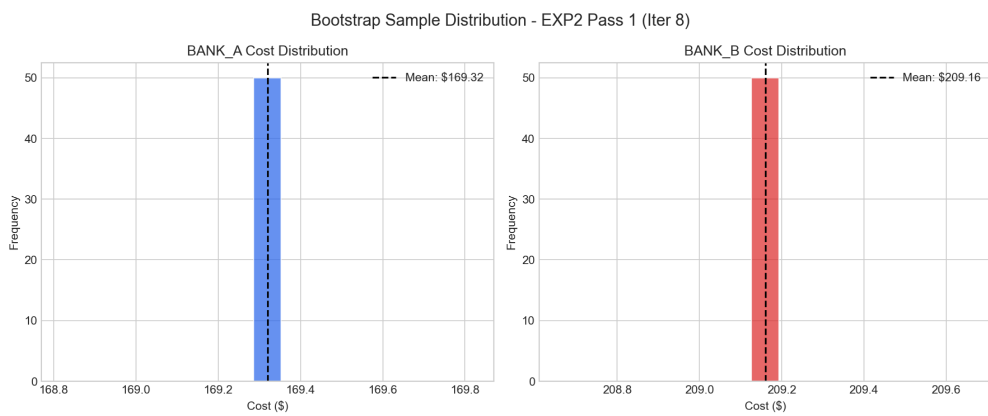

Figure 24: Experiment 2 Pass 1: Bootstrap sample distribution at convergence

Table 24: Experiment 2 Bootstrap Statistics - Pass 3

Agent	Mean Cost	Std Dev	95% CI	Samples
BANK_A	$219.12	$0.00	[$219.12, $219.12]	1
BANK_B	$229.08	$0.00	[$229.08, $229.08]	1

Table 25: Experiment 3 Bootstrap Statistics - Pass 1

Agent	Mean Cost	Std Dev	95% CI	Samples
BANK_A	$120.99	$0.00	[$120.99, $120.99]	1
BANK_B	$69.97	$0.00	[$69.97, $69.97]	1

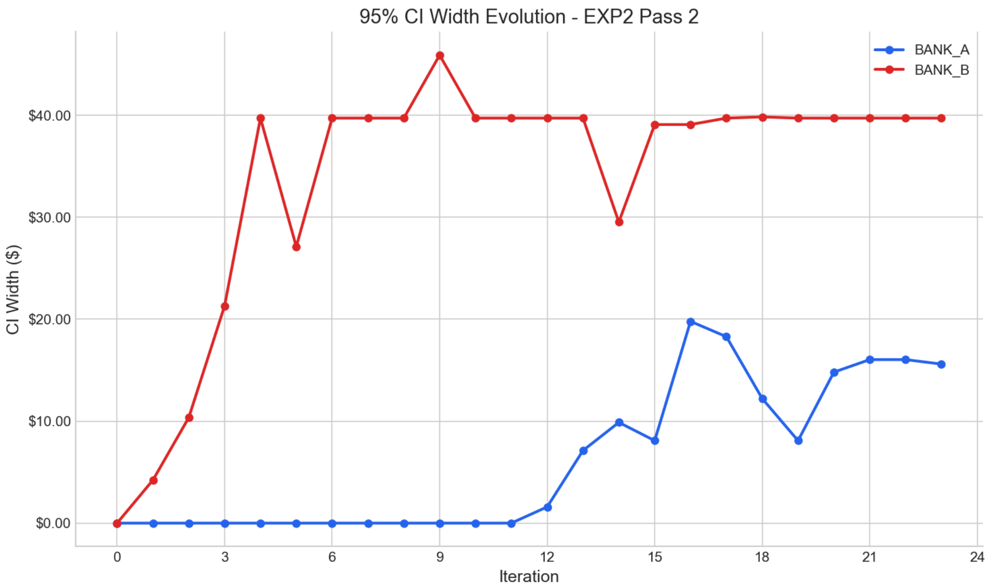

Figure 25: Experiment 2 Pass 2: CI width comparison across iterations

Figure 26: Experiment 2 Pass 2: Standard deviation evolution

Table 26: Experiment 3 Bootstrap Statistics - Pass 2

Agent	Mean Cost	Std Dev	95% CI	Samples
BANK_A	$124.89	$0.00	[$124.89, $124.89]	1
BANK_B	$68.98	$0.00	[$68.98, $68.98]	1

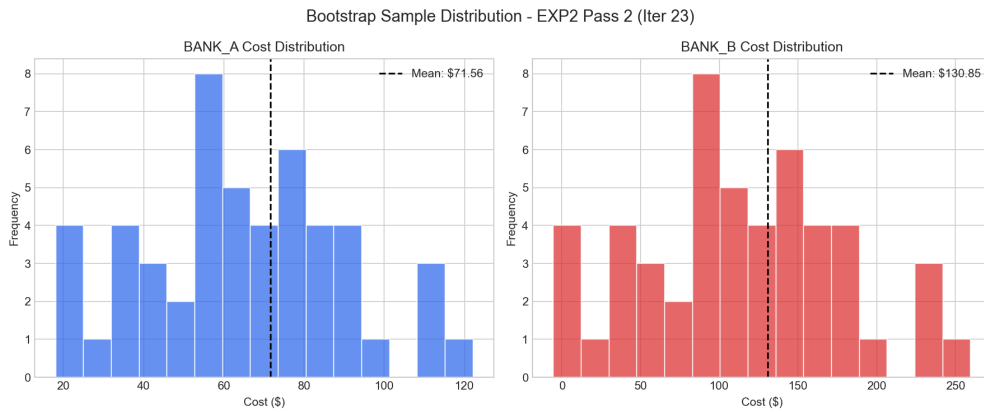

Figure 27: Experiment 2 Pass 2: Bootstrap sample distribution at convergence

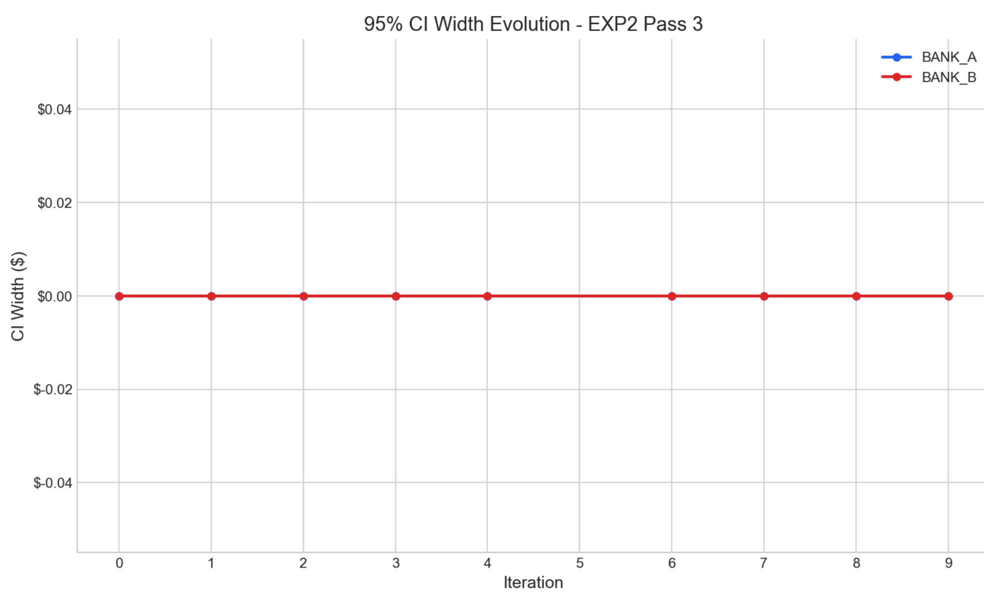

Figure 28: Experiment 2 Pass 3: CI width comparison across iterations

Figure 29: Experiment 2 Pass 3: Standard deviation evolution

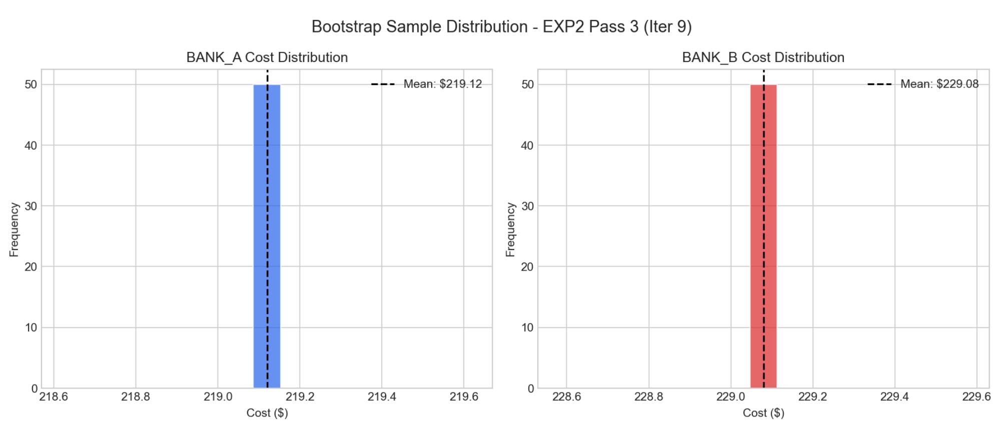

Figure 30: Experiment 2 Pass 3: Bootstrap sample distribution at convergence

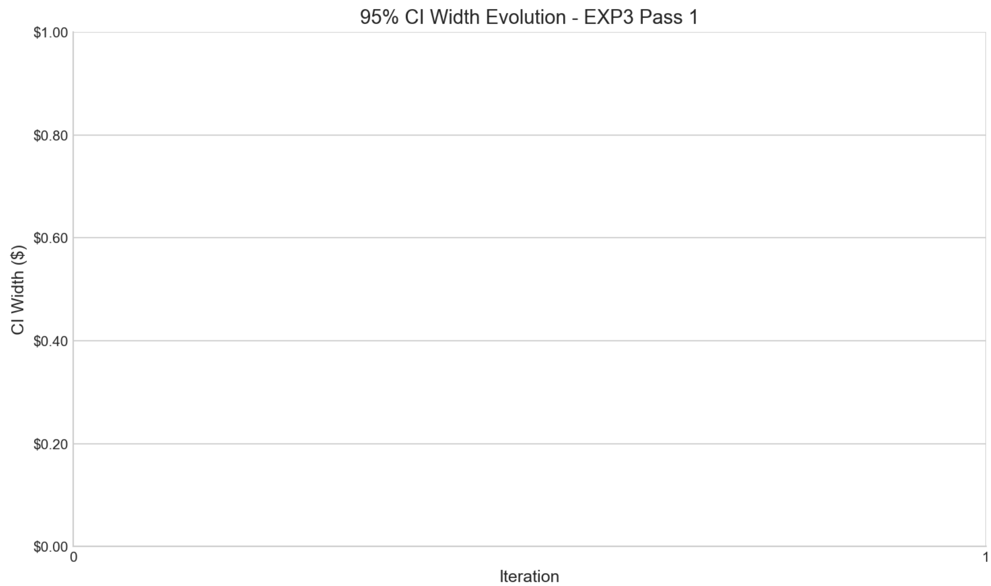

Figure 31: Experiment 3 Pass 1: CI width comparison across iterations

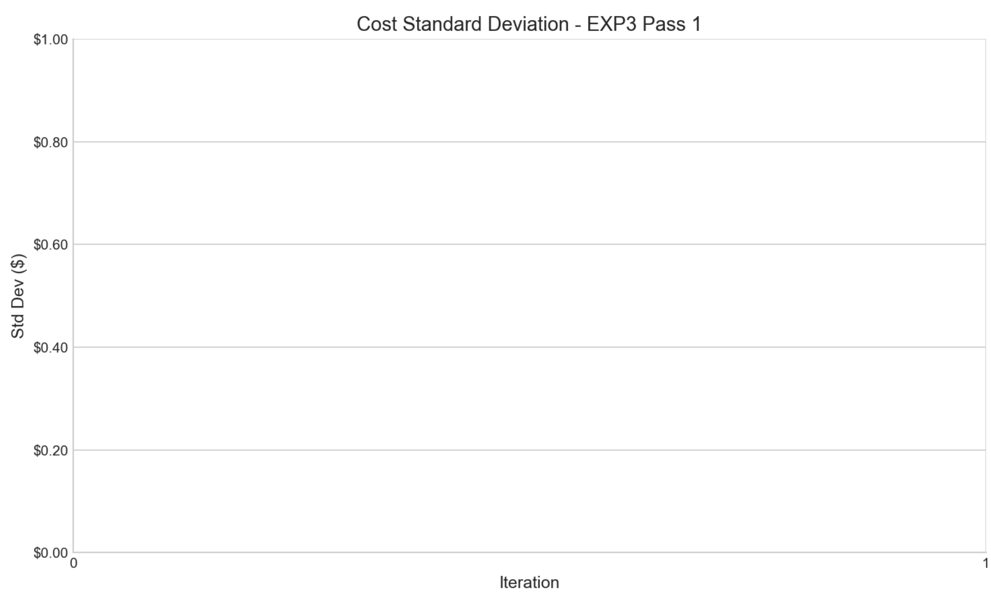

Figure 32: Experiment 3 Pass 1: Standard deviation evolution

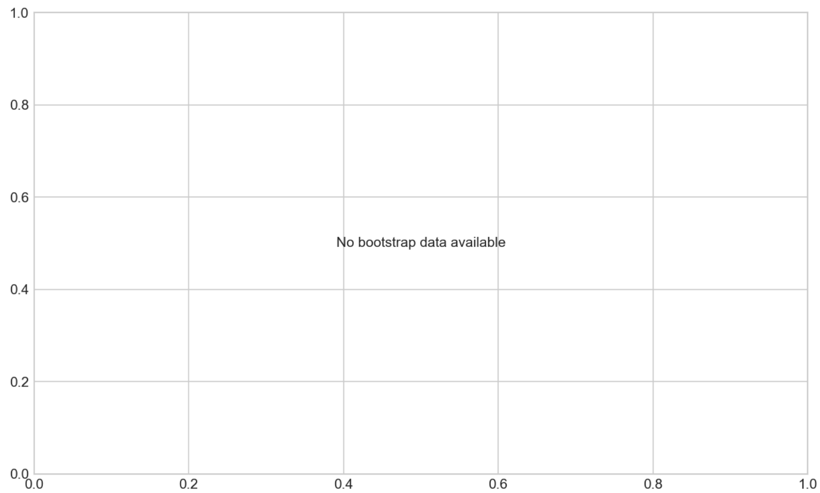

Figure 33: Experiment 3 Pass 1: Bootstrap sample distribution at convergence

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

Figure 35: Experiment 3 Pass 2: Standard deviation evolution

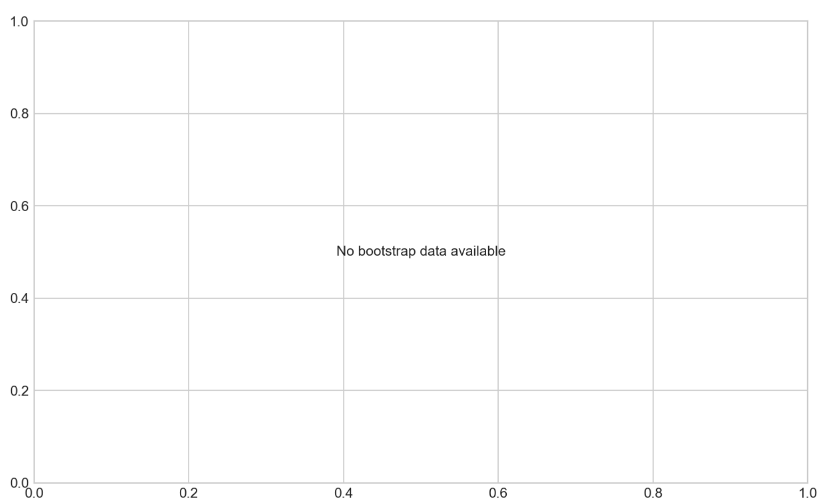

Figure 36: Experiment 3 Pass 2: Bootstrap sample distribution at convergence

E.2.3 Pass 3

E.3 Experiment 3

E.3.1 Pass 1

E.3.2 Pass 2

E.3.3 Pass 3

Table 27: Experiment 3 Bootstrap Statistics - Pass 3

Agent	Mean Cost	Std Dev	95% CI	Samples
BANK_A	$209.96	$0.00	[$209.96, $209.96]	1
BANK_B	$200.96	$0.00	[$200.96, $200.96]	1

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

F LLM Prompt Audit

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

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

Figure 38: Experiment 3 Pass 3: Standard deviation evolution

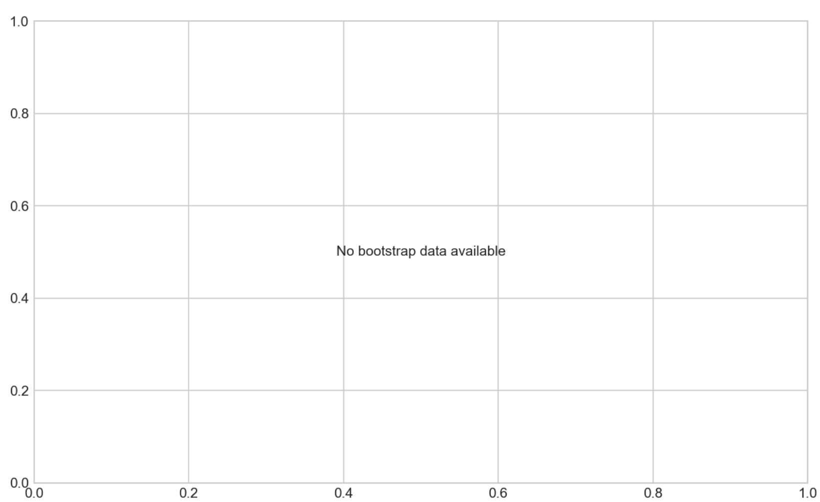

Figure 39: Experiment 3 Pass 3: Bootstrap sample distribution at convergence

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