Digital Twin-Based Simulation for Multi-Agent Financial Risk Decision Making

Luigina Sagese
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
sageseluigina@colostate.edu

Candelario Caballero
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
hellocandelario@unr.edu

Abstract

Digital twin-based simulation is emerging as a systems-level approach for understanding financial risk when markets are populated by heterogeneous algorithmic agents, human traders, institutions, and regulators. This paper examines how digital twins can support multi-agent financial risk decision making by integrating market infrastructure, agent behavior, regime detection, governance constraints, and stress-testing scenarios into a continuously updating computational representation. It argues that the central contribution of digital twins is not merely predictive accuracy but structural observability: the capacity to trace how decisions, incentives, data flows, and market mechanisms interact under normal and stressed conditions. The discussion emphasizes architecture, trust calibration, fairness, robustness, sustainability, and policy implications. It considers how multi-agent simulation can model liquidity fragmentation, herding, contagion, flash events, and regulatory intervention while acknowledging limits related to validation, data quality, adversarial behavior, and institutional adoption. Particular attention is given to market regime shifts, where historical regularities weaken and agent strategies must be recalibrated. The paper concludes that digital twin ecosystems should be governed as socio-technical infrastructures, with transparent assumptions, auditable decision pathways, and participatory oversight. Without such governance, digital twins risk becoming sophisticated but opaque instruments that concentrate power and obscure accountability. With appropriate design, however, they can improve systemic risk preparedness, support adaptive regulation, and enable more resilient financial decision making across public and private institutions.

Keywords

digital twins; multi-agent systems; financial risk; systemic risk; market regime shifts; simulation governance; fairness; resilience; financial regulation.

1. Introduction

Financial systems are complex adaptive infrastructures in which decisions are distributed across banks, asset managers, exchanges, clearinghouses, regulators, and increasingly autonomous algorithms. This distribution creates emergent behavior that cannot be inferred reliably from individual balance sheets or isolated trading strategies alone. Digital twin research has developed in manufacturing and engineering as a way to maintain a virtual representation of a physical system, often synchronized with real-time sensor data, for monitoring, simulation, and optimization [1]. As financial markets become more

computational, the same logic is being extended to market infrastructures, liquidity networks, and decision environments. The digital twin is not simply a dashboard or a historical database; it is a dynamic model that can be queried, stressed, and updated as the represented system changes [2].

Agent-based computational finance has long argued that markets should be modeled as ecologies of interacting strategies rather than as equilibria populated by a representative investor [3]. Simulation methods have been used to study herd behavior, liquidity evaporation, learning, imitation, and endogenous volatility. These approaches reveal that macro-level outcomes can shift abruptly when micro-level rules, information asymmetries, or network structures change [4]. Financial crises are often marked by such shifts, where conventional risk models fail because correlations, funding conditions, and behavioral responses mutate under stress [5]. Complexity theory reinforces this point by treating financial stability as a property of interconnected systems rather than a simple aggregate of independent risks [6].

The growth of artificial intelligence in finance intensifies these concerns. Machine learning systems can improve fraud detection, portfolio construction, and credit assessment, but they can also amplify procyclicality, obscure decision logic, and create new channels of systemic risk when many actors respond to similar signals [7]. A digital twin-based simulation environment offers a possible response: it can represent the market as a multi-layered system, allow regulators and firms to test interventions before deployment, and expose how risk emerges from interaction rather than from any single model. Yet the value of such a system depends on architecture, governance, and the willingness of institutions to treat simulation as a shared infrastructure rather than a proprietary forecasting tool. This paper develops that argument through a system-level analysis of digital twin-based simulation for multi-agent financial risk decision making.

2. Conceptual Foundations of Digital Twin-Based Financial Simulation

A financial digital twin differs from a conventional simulation because it maintains a persistent relationship with a real or hypothesized financial environment. In engineering, digital twins combine physical models, data streams, and analytical services to support lifecycle decisions [1]. In finance, the represented system includes market mechanisms, order flows, balance sheets, collateral chains, communication networks, and regulatory rules. The twin may be used for ex ante design, live monitoring, or ex post forensic analysis. Its most important property is not that it predicts a single future path but that it preserves structural relationships among actors and infrastructures so that alternative scenarios can be explored coherently [2].

The multi-agent dimension is essential because financial risk is often relational. A liquidity problem for one agent becomes a funding problem for another, and a risk-management action by one institution can become a destabilizing signal for many others. Trust-calibrated multi-agent decision intelligence under market regime shifts addresses this problem by treating trust not as a static parameter but as a dynamic property that changes with market conditions, counterparty behavior, and information quality [8]. In a digital twin, trust calibration can be represented through agent-specific confidence levels, counterparty assessments, and arbitration rules that determine when an agent follows its own model, defers to a collective signal, or reduces exposure. This is a structural issue rather than a purely mathematical one because the governance of trust determines the topology of influence within the simulated market.

A financial digital twin also requires a clear separation between the represented system and the representation. The twin may include simplified agents, approximate market mechanisms, and incomplete data, but it must document those simplifications. Otherwise, decision makers may mistake computational fluency for epistemic certainty. The twin should therefore be auditable at the level of assumptions, data provenance, model versioning, and scenario design. This aligns with broader critiques of data infrastructures, which show that data are not neutral resources but products of institutional choices, measurement regimes, and power relations [19]. In finance, those choices include what counts as risk, which counterparties are visible, and whose losses are treated as acceptable.

3. Multi-Agent Architectures and Decision Intelligence

A multi-agent financial digital twin can be organized around several architectural layers. The market layer represents exchanges, dark pools, dealers, and clearing arrangements. The agent layer represents heterogeneous decision makers, including high-frequency traders, pension funds, hedge funds, banks, central banks, and retail participants. The data layer represents prices, volumes, news, disclosures, and alternative data. The governance layer represents rules, mandates, constraints, and intervention authorities. The simulation layer coordinates time, events, and scenario execution. These layers are not independent; for example, a change in clearing rules can alter agent liquidity, which then changes market depth and price discovery. A digital twin is useful precisely because it can make these cross-layer dependencies visible.

High-frequency and algorithmic trading illustrate why architecture matters. Sociological and market microstructure studies show that electronic markets are not simply faster versions of human markets; they are materially different infrastructures in which speed, colocation, order types, and matching logic shape behavior [9]. The Flash Crash of 2010 demonstrated how automated execution, liquidity withdrawal, and cross-market linkages can produce abrupt dislocation [10]. A digital twin could help regulators and firms examine similar scenarios by replaying market conditions with alternative circuit breakers, order restrictions, or liquidity requirements. However, the same twin could also be used by private actors to optimize strategies that exploit other participants, which is why governance must accompany technical capability.

Decision intelligence in this context should not be reduced to prediction. It includes interpretation, escalation, abstention, and coordination. An agent may decide not to trade when uncertainty is high, or it may coordinate with other agents through a shared risk protocol. A digital twin can simulate these meta-decisions by representing communication channels, authority boundaries, and fallback procedures. This is especially relevant when machine behavior becomes a social phenomenon, with algorithms influencing each other through market signals and shared data [13]. If the twin ignores these feedback loops, it will underestimate systemic risk. If it represents them carefully, it can serve as a laboratory for testing how different coordination mechanisms affect stability, fairness, and efficiency.

4. Market Regime Shifts, Stress Testing, and Trust Calibration

Market regime shifts occur when the statistical and behavioral relationships that governed a period cease to hold. Volatility regimes change, correlations spike, liquidity migrates, and previously reliable signals become misleading. Traditional stress testing often relies on historical scenarios or hypothetical shocks, but it may not capture how agent strategies adapt during the transition between regimes. A digital twin can represent regime shifts as changes in the joint environment: volatility, funding costs, margin requirements, news sentiment, and

counterparty trust. Trust-calibrated multi-agent decision intelligence under market regime shifts is directly relevant because agents must decide how much to rely on past models, peer signals, and institutional guarantees when the environment has changed [8].

Stress testing within a digital twin should therefore be dynamic rather than static. Instead of applying a fixed shock to a portfolio, the twin can simulate a sequence of events in which agents observe losses, update beliefs, adjust leverage, and interact with margin calls and redemptions. This can reveal tipping points that are invisible in aggregate risk measures. For example, a moderate price decline may become a liquidity crisis if several large agents follow similar risk models and sell into the same market. The twin can test whether circuit breakers, central bank facilities, or disclosure requirements alter the path of the crisis. It can also examine unintended consequences, such as when a rule designed to protect one market pushes activity into a less transparent venue.

Trust calibration is both a technical and a governance challenge. Technically, agents need mechanisms to evaluate the reliability of data, models, and counterparties. Governance-wise, institutions need to decide which trust signals are legitimate, how they are audited, and who can override automated decisions. If trust is calibrated only by performance, agents may herd into strategies that worked in the recent past and fail collectively during a regime shift. If trust is calibrated too conservatively, markets may become fragmented and liquidity may decline. A digital twin can help explore these trade-offs, but it cannot resolve them without normative choices about stability, efficiency, and accountability.

5. Governance, Fairness, and Policy Implications

Digital twin-based financial simulation raises governance questions that cannot be answered by engineering alone. Who builds the twin, who can access it, and whose interests are represented in its assumptions? If the twin is developed primarily by large financial institutions, it may encode their risk priorities while underrepresenting households, small firms, or public authorities. If it is developed by regulators, it may lack the granularity and market expertise needed to capture private strategies. A hybrid model with public standards and private participation may be more effective, but it requires transparency, conflict-of-interest management, and independent validation. Algorithmic ethics research shows that such systems are not neutral; they shape decisions, distribute benefits and harms, and create new forms of accountability [11].

Fairness in financial digital twins has both procedural and distributive dimensions. Procedural fairness concerns whether affected parties have voice in scenario design, whether models can be contested, and whether decisions based on simulation can be appealed. Distributive fairness concerns whether the twin is used to protect systemic stability broadly or to optimize risk transfer in ways that concentrate losses on weaker participants. Machine learning fairness research has developed methods for detecting bias and assessing disparate impact, but financial fairness is complicated by the fact that risk differentiation is often legitimate [12]. A digital twin should therefore not simply enforce identical treatment; it should make the rationale for differential treatment visible and contestable.

Policy implications extend to market regulation, competition, and crisis response. Regulators could use digital twins to conduct continuous supervisory stress testing, evaluate proposed rules, and monitor emerging systemic risks. Central banks could use them to design liquidity facilities and communication strategies. Competition authorities could examine whether algorithmic coordination or data concentration creates market power. However, the use of

simulation in policy also creates risks of overconfidence, regulatory capture, and excessive reliance on quantitative models. Machine behaviour research reminds us that algorithms and their human overseers co-evolve, often in ways that are difficult to predict [13]. Governance must therefore include sunset clauses, red-team exercises, and periodic reauthorization of high-impact simulation systems.

Broader democratic concerns also arise when digital infrastructures become central to economic decision making. Big data and artificial intelligence can strengthen centralized control while weakening traditional forms of accountability if they are deployed without public oversight [14]. A financial digital twin should not become a black box that determines access to credit, liquidity, or market participation. It should be accompanied by public documentation, independent audit, and mechanisms for redress. The policy goal is not to eliminate uncertainty but to distribute the capacity to interpret and respond to uncertainty more equitably.

6. Infrastructure, Deployment, and Sustainability

Deploying a financial digital twin requires significant computational, data, and organizational infrastructure. It needs high-frequency market data, reference data, balance sheet information, network exposures, and regulatory filings. It needs simulation engines capable of running many scenarios under time constraints. It needs secure environments for sensitive data, because the twin may reveal exploitable vulnerabilities or private positions. It also needs skilled teams that can bridge finance, computer science, economics, and law. These requirements make digital twins expensive and institutionally demanding, which raises concerns about concentration. If only the largest firms and regulators can build credible twins, smaller institutions may be disadvantaged in risk assessment and regulatory negotiation.

Sustainability has at least two meanings in this context. First, computational sustainability concerns energy use, hardware lifecycle, and the carbon footprint of large-scale simulation. Running many agent-based scenarios with high-frequency data can be resource intensive. Designers should therefore consider adaptive fidelity, in which the twin uses simplified models for routine monitoring and higher-resolution models for stress events. Second, institutional sustainability concerns whether the twin can be maintained over time. Financial markets change constantly, so a twin that is not updated will decay into a historical artifact. Sustainable governance requires stable funding, clear ownership, version control, and processes for retiring obsolete components.

Data infrastructures are not neutral. They reflect decisions about what is measured, standardized, and shared [19]. A financial digital twin depends on standards for identifiers, timestamps, trade reporting, and exposure aggregation. Without such standards, the twin may be precise in one domain and blind in another. Cross-border deployment is especially difficult because legal definitions of risk, privacy, and market abuse differ across jurisdictions. A global twin may be politically fragile, while fragmented national twins may miss contagion channels. The practical path may be interoperable twins with shared protocols and negotiated data-sharing agreements rather than a single centralized model.

7. Robustness, Security, and Resilience

Robustness in a financial digital twin means that the system continues to provide useful guidance when data are missing, delayed, corrupted, or manipulated. Adversarial actors may attempt to poison the twin, probe its scenarios, or use its outputs to front-run regulatory interventions. Security therefore must include access control, anomaly detection, data

provenance, and scenario confidentiality. It must also include resilience against model failure. If the twin relies on a single machine learning component for regime detection, that component can become a single point of failure. Ensemble methods, human review, and fallback rules can reduce this risk, but they also add complexity and may slow response times.

Resilience at the market level is closely related to network structure. Banking ecosystems and financial networks can propagate shocks through direct exposures, common asset holdings, and funding channels [15]. Contagion models show that the same shock can produce very different outcomes depending on network topology, leverage, and the behavior of intermediaries [16]. The run on repo markets during the global financial crisis illustrates how funding structures can transform a valuation problem into a liquidity crisis [17]. A digital twin can represent these mechanisms, but it must be careful about network visibility. If the twin misses shadow banking linkages or off-balance-sheet exposures, it may produce false confidence.

Complex systems research also warns that crises are often driven by positive feedback, herding, and critical transitions rather than by simple exogenous shocks [18]. A digital twin can help identify early warning signals, but those signals may be ambiguous or self-defeating if actors respond to them. For example, if a regulator uses the twin to identify vulnerable institutions and imposes corrective action, markets may interpret the action as evidence of weakness and accelerate the stress. Resilience therefore requires not only better models but also better communication, credible commitment, and pre-agreed intervention protocols. The twin should support judgment, not replace it.

8. Case Illustrations and Cross-Domain Comparisons

In a hypothetical deployment, a consortium of regulators, exchanges, and large banks could maintain a shared digital twin of a multi-market trading ecosystem. The twin would simulate equities, futures, options, and repo funding, with agents representing market makers, hedge funds, asset managers, and central bank facilities. During normal periods, it would run low-fidelity monitoring to detect rising concentration or liquidity mismatch. During stress periods, it would activate higher-fidelity scenarios, including margin calls, redemption waves, and circuit breaker triggers. Regulators could test whether a proposed change to clearing rules reduces contagion without creating excessive risk-taking. Banks could test their own contingency plans within a common environment. This example shows that the twin is not merely a model; it is a coordination platform.

Cross-domain comparisons are instructive. Manufacturing digital twins emphasize sensor synchronization, predictive maintenance, and lifecycle optimization [1]. Urban digital twins emphasize infrastructure interdependencies, mobility, and public policy. Financial digital twins share these features but differ in one crucial respect: financial agents are strategic, reflexive, and capable of adapting to the twin itself. If market participants know how the twin works, they may change their behavior to influence its outputs or to avoid detection. This reflexivity means that validation cannot be a one-time exercise. The twin must be continuously red-teamed, and its use must be governed in ways that reduce manipulation and regulatory arbitrage.

The information environment also matters. Financial decisions increasingly depend on data infrastructures that combine market data, social media, geolocation, and derived analytics. The data revolution has made such information abundant but also unevenly governed [19]. A digital twin that ingests these data must address privacy, consent, and public accountability.

The fourth revolution in information and communication technologies has reshaped human reality by embedding computation into everyday life [20]. In finance, this embedding means that digital twins may become part of the market they represent, not merely observers of it. Designers must therefore ask not only what the twin can simulate but also what kind of financial system it helps to reproduce.

9. Evaluation, Validation, and Organizational Learning

Evaluating a multi-agent financial digital twin is difficult because there may be no single ground truth. Backtesting against historical episodes is useful but limited, since the twin may be calibrated to those episodes. Forward testing in sandbox environments can help, but sandboxes may not reproduce real incentives or liquidity conditions. Validation should therefore combine multiple methods: historical replay, expert review, sensitivity analysis, adversarial testing, and comparison with independent models. The goal is not to prove that the twin is correct but to establish that its limitations are known and that its outputs are used within appropriate confidence bounds.

Organizational learning is equally important. A digital twin can reveal uncomfortable truths about an institution's vulnerabilities or about the systemic effects of its strategies. If leadership treats the twin as a compliance exercise, its insights may be ignored. If it is integrated into risk committees, scenario planning, and regulatory dialogue, it can change mental models and decision routines. This requires training, incentives, and a culture that rewards the discovery of hidden risks rather than the confirmation of existing beliefs. It also requires memory: lessons from past simulations should be stored, revisited, and compared with actual outcomes.

Evaluation should also consider distributional effects. A twin may improve aggregate stability while imposing costs on particular groups, such as small banks, retail investors, or emerging markets. These effects should be modeled explicitly rather than treated as externalities. Fairness metrics, stakeholder consultations, and impact assessments can help. However, no metric can fully resolve normative disagreement. The twin should therefore support deliberation, not foreclose it. Its legitimacy depends on whether affected parties can understand its assumptions and challenge its conclusions.

10. Conclusion

Digital twin-based simulation for multi-agent financial risk decision making offers a powerful way to represent the interconnected, adaptive, and reflexive character of modern markets. By combining persistent virtual representation with heterogeneous agents, regime-sensitive stress testing, and governance layers, digital twins can improve systemic risk preparedness and support more informed regulatory and institutional decisions. They can help expose hidden dependencies, test interventions, and simulate the transition between market regimes in ways that conventional models often miss. The required trust-calibrated multi-agent approach under market regime shifts illustrates the importance of dynamic confidence, coordination, and adaptation in such environments [8].

The promise of digital twins, however, is inseparable from their governance. Architecture determines who is visible and who is excluded. Data infrastructure determines what can be known. Policy determines how simulation outputs are used. If digital twins are developed without transparency, fairness, and accountability, they may become instruments of concentrated power and false precision. If they are developed as shared socio-technical infrastructures with auditable assumptions, participatory oversight, and robust security, they

can strengthen resilience and public interest. The central challenge is not merely technical. It is to design financial digital twins that help institutions learn, cooperate, and act responsibly under uncertainty while remaining open to contestation and revision. That challenge places digital twin research at the intersection of computer science, finance, regulation, and democratic governance.

References

1. Grieves, M., & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In F.-J. Kahlen, S. Flumerfelt, & A. Alves (Eds.), Transdisciplinary perspectives on complex systems (pp. 85–113). Springer. https://doi.org/10.1007/978-3-319-38756-7_4

2. Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital twin in industry: State-of-the-art. IEEE Transactions on Industrial Informatics, 15(4), 2405–2415. https://doi.org/10.1109/TII.2018.2873186

3. LeBaron, B. (2006). Agent-based computational finance. In L. Tesfatsion & K. L. Judd (Eds.), Handbook of computational economics (Vol. 2, pp. 1187–1233). Elsevier. https://doi.org/10.1016/S1574-0021(05)02024-1

4. Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences, 99(Suppl 3), 7280–7287. https://doi.org/10.1073/pnas.082080899

5. Bookstaber, R. (2017). The end of theory: Financial crises, the failure of economics, and the sweep of human interaction. Princeton University Press.

6. Battiston, S., Farmer, J. D., Flache, A., Garlaschelli, D., Haldane, A. G., Heesterbeek, H., Hommes, C., Jaeger, C., May, R., & Scheffer, M. (2016). Complexity theory and financial regulation. Science, 351(6275), 818–819. https://doi.org/10.1126/science.aad0299

7. Financial Stability Board. (2017). Artificial intelligence and machine learning in financial services: Market developments and financial stability implications. Financial Stability Board.

8. Liu, T. (2026). Trust-Calibrated Multi-Agent Decision Intelligence for Financial Risk Management under Market Regime Shifts. Available at SSRN 7307503.

9. MacKenzie, D. (2018). Material signals: A historical sociology of high-frequency trading. American Journal of Sociology, 123(6), 1635–1673. https://doi.org/10.1086/696442

10. Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The Flash Crash: High-frequency trading in an electronic market. Journal of Finance, 72(3), 967–998. https://doi.org/10.1111/jofi.12498

11. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2). https://doi.org/10.1177/2053951716679679

12. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning. fairmlbook.org.

13. Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar,

E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A., ... Wellman, M. (2019). Machine behaviour. Nature, 568(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y

14. Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R. V., & Zwitter, A. (2019). Will democracy survive big data and artificial intelligence? In Towards digital enlightenment (pp. 73–98). Springer. https://doi.org/10.1007/978-3-319-90869-4_7
15. Haldane, A. G., & May, R. M. (2011). Systemic risk in banking ecosystems. Nature, 469(7330), 351–355. https://doi.org/10.1038/nature09659
16. Glasserman, P., & Young, H. P. (2016). Contagion in financial networks. Journal of Economic Literature, 54(3), 779–831. https://doi.org/10.1257/jel.20151228
17. Gorton, G. B., & Metrick, A. (2012). Securitized banking and the run on repo. Journal of Financial Economics, 104(3), 425–451. https://doi.org/10.1016/j.jfineco.2011.03.016
18. Sornette, D. (2003). Why stock markets crash: Critical events in complex financial systems. Princeton University Press.
19. Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. Sage.
20. Floridi, L. (2014). The Fourth Revolution: How the infosphere is reshaping human reality. Oxford University Press.