Advance Journal of Econometrics and Finance

Vol-4, Issue-2, 2026

Advance Journal of Econometrics and Finance

Online ISSN

2959-8990

Print ISSN

2959-8982

https://ajeaf.com/index.php/Journal/About

Name of Publisher: SCHOLAR CRAFT EDUCATION & RESEARCH HUB

Review Type: Double Blind Peer Review

Journal Frequency: Quarterly Research Journal (4- Issue)

Data Chaos, Model Fragility, and the Human Fix for Financial GenAI

*[1]Moin Ahmad Moon, [2]Atif Hussain

Moin Ahmad Moon*
Air University School of Management, Air University Multan Campus, Pakistan. Corresponding Author Email: moin@aumc.edu.pk

Atif Hussain
Air University School of Management, Air University Multan Campus, Pakistan. Atifq4141@gmail.com

Abstract

While Gen AI can offer a significant advantage in terms of financial forecasting, this might become neutralized if institutions employ similar models. The prediction error correlations during high volatility may enhance the systemic risk through the homogeneous architecture. In this study we seek answers to the questions whether good data quality can have consequences and whether human guidance can temper this risk. The data consists of 285 survey responses collected from financial professionals at Pakistani banks, asset managers, fintech and investment banks. Using structural equation modeling on the cross-sectional data gathered from financial professionals from 3Q2025 to 1Q2026, we conclude that increasing Gen AI adoption does improve forecasts; this effect stems partly from demanding more holistic multi-domain data quality. However, the finding comes with a caveat: human expertise calibration is characterized by an inverted-U shape — lack of it opens up blind spots while overreliance on it has a similarly negative effect. The best result lies in a moderate human judgment setting. These findings have clear implications for regulators, the design of corporate governance, and efforts at re-evaluating the human judgement setting. These finding have clear implications for regulators, the design of corporate governance,and efforts at re-evaluating the human contribution to Ai driven financial forecasting.

Keywords: Generative AI; Financial Forecasting; AMOS; Moderated Mediation; ; Adaptive Fragility Risk; Human Expertise Calibration

1. Introduction

Generative Artificial Intelligence (GenAI) has rapidly emerged as a transformative force in financial forecasting. Adversarial training between a Generator (synthesizing synthetic data) and a Discriminator (judging synthetic examples) is employed by GANs. It has been stressed by Wilson (2025) that GAN-generated data decreases overfitting and enables models to explore rare market regimes. Transformers focus on extracting predictive signals through unstructured text via self-attention. FinBERT focuses on contextual financial language ambiguity through domain-specific accuracy rather than keyword-based approaches (Khalil, 2026). The market value of artificial intelligence in the financial industry will reach $190 billion by 2030 (Mo & Ouyang, 2025), (Moon, 2026) representing an accelerated adoption dynamic. Within the task-technology fit view, empirical studies show machine learning as data-driven adaptive AI with tasks related to dynamic forecasting (Ghosh, 2025; Alassuli et al., 2026), which is significantly linked to behavior-oriented and dynamic forecasting responsibilities (Moon et al., 2026).

Generative AI statistical methods involve the integration of linear regression, ARIMA, and their univariate descendants through autonomous, simulation-based economic models. Two most prevailing approaches in the implementation of GenAI are Generative Adversarial Networks (GANs), which formulate synthetic time-series that secure transformative facts of real markets, and Transformer-based models like FinBERT, which convert unstructured textual data into quantitative sentiment metrics (Khalil, 2026; Wilson, 2024). At the same time, uncertainty develops a risk that does not address adaptive fragility for the purpose of developing architectural procedures.

On the other hand, adaptive fragility is considered a second-order systemic risk that arises when several institutions simultaneously enforce structurally similar models. Hence, the correlated prediction errors get intense instead of being canceled among the institutions. Three technical vulnerabilities comprise temporal robustness that fails in the phase of transition of regimes, epistemic opacity that negates trust of the regulators, and look-ahead bias through imprecise LLM timestamp index procedures (Abu-Dabaseh et al., 2025; Mo & Ouyang, 2025; SEACEN Centre, 2026). It signifies that adaptive fragility focuses on highlighting the possible errors and predictable procedures to achieve the aims of the institutions prospectively.

The aim of this research is to understand financial forecasting through AI, where institutional risks are analyzed, market trends and competition among business rivalries are addressed, and political dynamics' implications on financial strategies are studied comprehensively. The research is stressed to evaluate the present challenges of Generative AI Financial Forecasting in view of emerging struggles in banking systems, institutional weaknesses, risk factors of AI adaptation, societal strategies, and policies at the national and global level, significantly affecting the present choices and dependency on AI for the national, regional, and global growth dynamics.

2. Theoretical Background

The study utilizes the Employee-Centered Automation Theory (Mo & Ouyang, 2025) as the theoretical base, which designs the strategies of AI as a leading force for organizational structure (Tariq et al., 2026) . value is not value without prediction.but signifies regarding trust dynamics or overemphasis on model application results. The theoretical framework is further grounded in task-technology fit theory (Ghosh, 2025; Alassuli et al., 2026), which posits that alignment between task requirements and technology characteristics determines performance outcomes in data-driven forecasting environments.

2.1.Data Quality as Mediator, Not Prerequisite

In this work, multi-domain data quality is interpreted as a mediating factor rather than as a control variable. Hence, GenAI Application Intensity focuses on data quality as extensive generative functions focus on richer latent demonstrations and their alignment, which formulates robust interpretation. In the absence of adequate alignment, raw sophistication increases noise mainly in the transitional phase of distribution (Mo & Ouyang, 2025). In view of the large volume, high level of noise of blockchain data, and complex structure, systematic preprocessing and augmentation are needed before model training to signify the statistical consistency, and structural reliability of the inputs is immensely required to have reliable data quality (Gong et al., 2026).

2.2.. Human Expertise Calibration as Boundary Condition

The Employee-Centered Automation Theory (Mo & Ouyang, 2025) designs the strategies of AI as a leading shock for organizational structure. The study focuses on three dimensions in view of Black Swan Track Record, AI Trust Scale, and Score of Calibration (Jian & Landies, 2003). The AMOS redesign involved the evaluation of moderation through multi-group (high/low calibration split) instead of product-indicator involvement. Earlier categorization of machines capable of human-like intelligence (McCarthy et al., 1955) has evolved so that AI now includes systems that could autonomously reason, learn, and formulate decisions (Russell et al., 2017; Zhang et al., 2026). Machine Learning algorithms excel at recognizing complex patterns, whereas deep learning architectures provide effective models of nonlinear relationships in high-dimensional datasets (Zhang et al., 2026).

3. Literature Review and Hypothesis Development

3.1. GenAI Adoption Intensity and Forecast Performance

Artificial intelligence has emerged as a transformative force in the global economy, where its growth outstrips investments, enhances productivity, and predicts innovation (Cockburn et al., 2018; Furman & Seamans, 2019). Sachs (2023) forecasts that AI can grow global GDP by approximately $7 trillion and enhance productivity growth of the United States through 1.5 percentage points yearly over the next decade. Financial institutions have extensively adopted artificial intelligence to improve decision-making processes and operational efficiency (Alt et al., 2018; Goodell et al., 2021; Zhang et al., 2026). The remarkable predictive abilities of AI-based systems provide leading economic potential by augmented intelligence usage among several financial fields (Collins et al., 2021; Zhang et al., 2026). Like previous general-purpose technologies (GPTs) such as the steam engine and electricity, AI is characterized by fast improvement, wide applicability, and the ability to catalyze complementary innovations (Cockburn et al., 2018; Mo & Ouyang, 2025).

The study conducted by Fedyk et al. (2022) signifies more than 310,000 employee resumes through 36 leading US auditing organizations to highlight that AI talent enhances product quality through decreasing financial restatements and fees while enhancing productivity (Farooq & Moon, 2025a) .in view of Zhang (2024), mutual funds with higher AI adoption perform at a higher level than other peers, and the long-short strategy based on AI ratio signifies an annualized increased return of 2.89% (Mo & Ouyang, 2025). Artificial intelligence technologies, such as expert systems, robotic process automation, and machine learning, have become necessary forces in adopting a data-driven approach in digital banking systems (Abu-Dabaseh et al., 2025; Alassuli et al., 2026).

H1: The adoption intensity of GenAI positively impacts forecast performance.

3.2. Multi-Domain Data Quality and Forecast Performance

Chen et al. have highlighted a multimodal anomaly detection method fusing time-domain and frequency-domain characteristics, achieving a precision of ≈97.6% and an F1-score of ≈0.951 in regional power grid monitoring. Hybrid architectures integrate Transformers through deep learning branches where extensive strategies handle complex classification tasks (Gong et al., 2026) (Farooq & Moon, 2025b).The results of multiple regression analysis signify that AI usage contributes to enhancements in forecasting accuracy through machine learning and RPA that represent significantly dominant effects (Alassuli et al., 2026). Quality cross-domain data integration reduces noise in latent space alignment, directly improving forecasting robustness across market regimes (Mo & Ouyang, 2025).

H2: The quality of multi-domain data positively impacts forecast performance.

3.3. GenAI Adoption Intensity and Multi-Domain Data Quality

Empirical analysis highlights that social media sentiment, where Granger develops short-term stock movements in technology and finance through a hybrid achieving 68.5% directional accuracy along with 22% reduction of prediction error compared to ARIMA models (Khalil, 2026). The fast advancement of GenAI has significantly started to reshape economic systems by improving information processing abilities, which is the core function by which it categorizes resources, supports economic coordination, and ultimately contributes to managing risk factors (Moon et al., 2025) .igher intensity GenAI adoption systematically demands more heterogeneous data sources, driving organizations to invest in cross-domain data quality infrastructure (Gong et al., 2026).

H3a: The adoption intensity of GenAI positively impacts the quality of multi-domain data.

H3b: The quality of multi-domain data mediates the association between the adoption intensity of GenAI and forecast performance.

3.4. Adaptive Fragility Risk

Adaptive fragility is a second-order systemic risk that arises when several institutions simultaneously enforce structurally similar models, causing correlated prediction errors to intensify rather than cancel out among institutions (Abu-Dabaseh et al., 2025; Mo & Ouyang, 2025) (Asjad et al., 2025) .when institutions converge on structurally identical GenAI architectures, the quality-enhancing effects of higher adoption intensity are diminished. Synthetic data disclosure through GAN-augmented ratios and mutual information scores is advised by SEACEN (2026) to manage such fragility. Architectural diversity within financial systems is thus a prerequisite for the full realization of data quality benefits from GenAI adoption.

H4: The adaptive fragility risk moderates the GenAI adoption–data quality association.

3.5. Human Expertise Calibration

The oversight is required to encompass result-oriented verification of the AI, the responsibility for AI-assisted decision-making proves, and also progressive systematic adjustment after mistakes have been highlighted (Wang et al., 2026). The epistemic opacity and look-ahead bias that result from inexact timestamps should be avoided (Mo & Ouyang, 2025). The AI-based benefit represents the favorable dynamics of senior partners, but at the same time, the junior analysts go-through replacement risks where the roles of junior analysts should be redesigned in view of arguments and the process of validation. Ultimately, human oversight is needed to significantly contribute to calibration instead of static presence through forming persistence-based protocols of the EU AI Act Article 14 needs (Tassen et al., 2025).The lack of calibrations develops analytical blind spots, whereas increased dependency on judgements of humans effects the advantages that have been predicted through AI, which develops an inverted-U relationship between calibration and performance.

H5: The human expertise calibration moderate the link between GenAI-adoption-based scenarios for predicted performance.

H6: forecast performance is related to human expertise calibration an inverted – u manner through the prediction of performance.

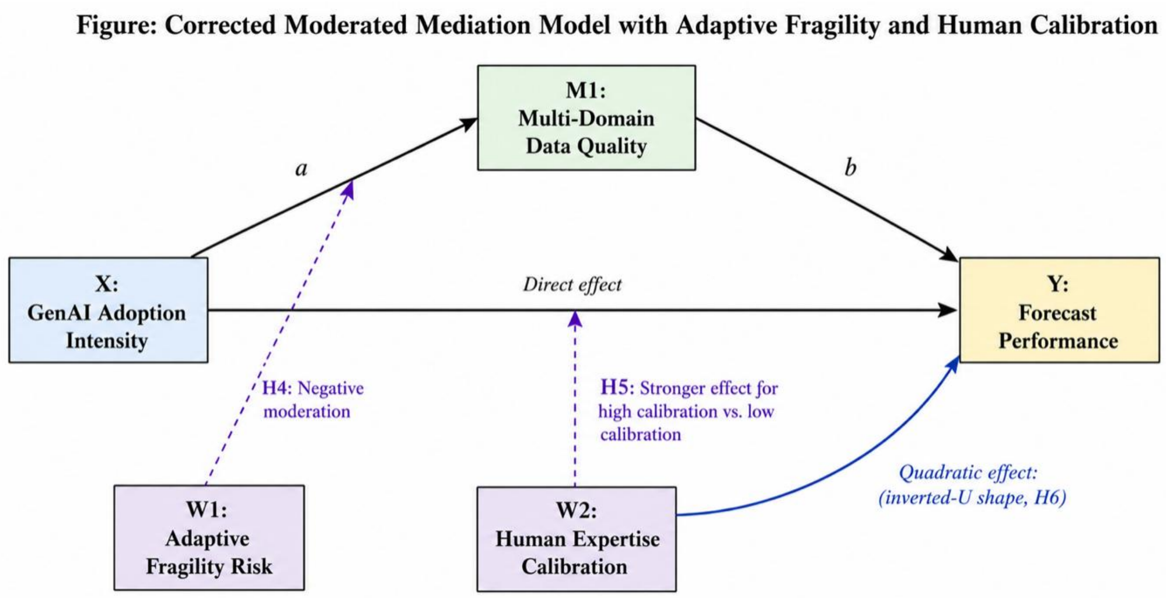

Figure 1. Theoretical model.

4. Methods

4.1. Sample

This study's population has been financial professionals in Pakistan who have direct experience with GenAI-assisted tools of prediction. Hence, the respondents are the professionals working at banks of the country, fintech and asset management companies, and also investment banks, which are directly part of it or supervise the system of using GenAI forecasting-based systems. Membership lists from three major financial industry associations in Pakistan (total N = 2,847) were used and sampling was done by systematic random sampling procedure (every 5th respondent after a random start), leading to 520 potential respondents.

The inclusion criteria involved those who: (1) were currently working at a qualifying institution; (2) had direct experience with GenAI forecasting tools or with supervising the use of GenAI forecasting tools; and (3) had worked at their current institution for at least 12 months. 310 of the 520 individuals contacted filled in the survey instrument. Cases with 25% or more of missing item-level data were eliminated, and a final analytical sample of N = 285 has been used (effective response rate = 54.8%). The average sample profile consisted of 72% males, 28% females, with a mean age of 34.2 years (SD = 8.1) and a mean of 9.4 years of work experience. Breakdown by type of institution: retail/commercial banks (28%), asset managers (22%), fintech organisations (31%), and investment banks (19%). Data were collected from 2025

4.2. Measures

Hence, each five constructs have been using a (1=strongly disagree, 5= strongly agree) as the questionnaire comprises 20 items that highlights the five latent constructs: GenAI Adoption Intensity (X; 4 items), Multi-Domain Data Quality (M1; 4 items), Adaptive Fragility Risk (M2; 4 items), Human Expertise Calibration (W; 4 items), and Forecast

Performance (Y; 4 items)(Moon et al., 2024).To get the control regarding acquiescence bias, items W3 and W4 have been reverse-scored as the complete measurement instrument, where standardized factor loadings (FL) and squared multiple correlations (SMC) are displayed in Table 1.

4.3. Procedure

The data has been collected by a structured self-administered survey questionnaire through a system of financial professionals working in leading Pakistani associations of industry of finance. Hence, the systematic random sampling approach (every 5th respondent from the lists of membership) has gained representation among institutions. In the beginning, we have reached 520 prospective respondents after the process of inclusion criteria and eliminating cases where extensive missing data has been observed, reaching the concluding analytical sample of N=285.

4.4. Data Analysis Procedures

Hence, for preparation and data entry, SPSS 26 has been utilized. The two-step SEM approach recommended through Anderson and Gerbing (1988) has been utilized, and the process of performance in IBM AMOS 29. The features of the psychometric measurement model have been evaluated by utilizing a Confirmatory Factor Analysis (CFA) in Step 1 before the estimation of the structural path. The complete structural model (direct, mediated, and moderated paths) has been estimated in Step 2 by using Maximum Likelihood (ML) estimation. Hence, the Full Information Maximum Likelihood (FIML) has been utilized to handle the residual missing data (less than 3% per item).

The distributional assumptions were checked using the skewness, kurtosis, Mahalanobis distance statistics, and variance inflation factors. The Harman's single-factor test (variance explained = 24.6%) and the common latent factor procedure ($\Delta CFI < 0.01$) indicated that common method variance was not a major confounder. Bootstrapped indirect effect estimates (250 repetitions, bias-corrected) were calculated to test H3b. Sequential configural, metric, and scalar invariance testing of multi-group moderation (H5; H6) was applied to high/low calibration splits.

5. Results and Discussions

Before data analysis, we screened the data for missing values, aberrant values, outliers, and normality. Excessive missing values (≥25%), aberrant values, univariate and multivariate outliers were treated with mode substitution method (Tabachnick & Fidell, 2007). The data were normally distributed as the values of skewness and kurtosis fell within the recommended threshold. For multicollinearity, we assessed the variance inflation factor (VIF < 10) and tolerance level (>0.1) between independent variables and found no issues of multicollinearity. Self-reported measures are vulnerable to common method biases (CMB), and to control these potential biases, we followed procedural and analytical remedies outlined by Podsakoff et al. (2012).

5.1. Sample Demographics

The sample was 72% male and 28% female. The mean age was 34.2 years (SD = 8.1) with a mean of 9.4 years of work experience. By institution type: retail/commercial banks (28%), asset managers (22%), fintech organisations (31%), and investment banks (19%). Despite variation in institutional background, all respondents had direct experience with GenAI forecasting tools and constitute the most knowledgeable segment of the financial industry regarding AI-enabled forecasting.

5.2. Structural Equation Modeling (SEM)

We used the two-step approach suggested by Anderson and Gerbing (1988) for structural equation modeling (SEM). We first tested the measurement model for the reliability and validity of the constructs and, second, we tested the structural model for hypothesis testing.

5.2.1. Confirmatory Factor Analysis (CFA)

We conducted confirmatory factor analysis (CFA) with maximum likelihood estimation (MLE) since it is the best regular, unbiased estimator of fitness (Hair et al., 2017). The measurement model (20 items, 5 factors) showed good fit after releasing two theoretically sound error covariances between parallel items in the same construct (CMIN/DF = 1.47, GFI = 0.91, AGFI = 0.88, CFI = 0.95, RMSEA = 0.040, NFI = 0.85, TLI = 0.92, IFI = 0.94, PCLOSE = 0.93). All fit indices are at or above the traditional benchmarks (ud Din et al., 2022) , as shown in Table 3.

The factor loadings of all the standardized factors were greater than 0.68 ($p < 0.001$). All constructs had convergent validity, as Composite Reliability (CR) ranged from 0.81 to 0.89 and Average Variance Extracted (AVE) ranged from 0.52 to 0.67.

For construct reliability, we used Cronbach's alpha ($\alpha \geq 0.6$; Nunnally & Bernstein, 1994), average variance extracted (AVE ≥ 0.5; Fornell & Larcker, 1981), and composite reliability (CR ≥ 0.6; Hair et al., 2017). Table 2 displays Cronbach's alpha, AVE, and CR for the constructs, all of which conveniently surpass the required minimum thresholds, thus indicating that all constructs are reliable.

To establish construct validity, we evaluated convergent and discriminant validity. Convergent validity was established through: (1) AVE ≥ 0.5, (2) significant factor loadings (FL ≥ 0.5) for all observed variables, and (3) comparisons of CR and AVE confirming CR ≥ AVE ≥ 0.5 (Kline, 2015). The Fornell-Larcker criterion was used to check discriminant validity; the square root of the AVE (on the diagonal of the correlation matrix) was greater than all inter-construct correlations, indicating that each construct represents variance not shared by the others.

Table 1: ***Results of Confirmatory Analysis***

SN	Items	Factor loadings	SMC	Mean	SD
GenAI Adoption Intensity (X) — α = 0.84, CR = 0.86, AVE = 0.61					
X1	Deployed GAN-augmented or Transformer-based GenAI forecasting models	0.79**	0.62	4.2	0.51
X2	Proportion of forecasting outputs generated by GenAI systems	0.76**	0.58		
X3	GenAI systems integrated with multiple heterogeneous data streams	0.80**	0.64		
X4	Systematically tracks architectural complexity of GenAI models	0.77**	0.59		
Multi-Domain Data Quality (M1) — α = 0.87, CR = 0.88, AVE = 0.65					
M1_1	Heterogeneous data sources successfully aligned for forecasting	0.83**	0.69	4.1	0.54
M1_2	Cross-domain data inconsistencies systematically resolved	0.79**	0.62		
M1_3	Predictive signals from integrated domains are meaningful	0.81**	0.66		
M1_4	Formal documented procedures for auditing cross-domain data quality	0.77**	0.59		
Adaptive Fragility Risk (M2) — α = 0.82, CR = 0.83, AVE = 0.55					
M2_1	Market institutions converging on structurally similar AI architectures	0.75**	0.56	3.8	0.62
M2_2	Widespread identical AI adoption amplifies correlated forecast errors	0.72**	0.52		
M2_3	Organization monitors architectural overlap with competitors	0.74**	0.55		
M2_4	Correlated algorithmic responses amplify market stress during volatility	0.76**	0.58		
Human Expertise Calibration (W) — α = 0.80, CR = 0.81, AVE = 0.52					
W1	Accurately assesses probability that AI forecast will prove correct	0.77**	0.59	3.9	0.58
W2	Selectively overrides AI outputs when contradictions exist	0.73**	0.53		
W3(R)	Domain knowledge leads to trusting own judgment over AI in novel situations (reverse)	0.68**	0.46		
W4(R)	Typically accepts AI recommendations without substantial scrutiny (reverse)	0.70**	0.49		
Forecast Performance (Y) — α = 0.88, CR = 0.89, AVE = 0.67					
Y1	Produces directionally accurate predictions in stable and volatile markets	0.84**	0.71	3.7	0.67
Y2	Maintains predictive accuracy during out-of-sample periods	0.81**	0.66		
Y3	Portfolio performance attributable to AI exceeds non-AI benchmarks	0.80**	0.64		
Y4	Forecasting outputs adequately explained to regulators (interpretability)	0.82**	0.67		

Note. SN = Item number; FL = standardized factor loading (all $p < 0.001$); SMC = squared multiple correlation; SD = standard deviation; (R) = reverse-scored item.

Table 2: ***Results for Convergent and Discriminant Validity***

Variables	α	CR	AVE	1	2	3	4	5
1 GenAI Adoption Intensity (X)	0.84	0.86	0.61	**0.78**	—	—	—	—
2 Multi-Domain Data Quality (M1)	0.87	0.88	0.65	0.54	**0.81**	—	—	—
3 Adaptive Fragility Risk (M2)	0.82	0.83	0.55	0.38	0.29	**0.74**	—	—
4 Human Expertise Calibration (W)	0.80	0.81	0.52	0.42	0.47	0.21	**0.72**	—
5 Forecast Performance (Y)	0.88	0.89	0.67	0.61	0.69	−0.33	0.58	**0.82**

Note. The diagonal elements are the square root of the AVE values (bold) and the off-diagonal elements are the inter-construct correlations; CR = Composite reliability; α = Cronbach's alpha; AVE = Average Variance Extracted. Notably, Adaptive Fragility Risk (M2) showed a negative correlation with Forecast Performance ($r = −0.33$, $p < 0.01$).

5.2.2. Structural Model and Hypothesis Testing

The structural model noted for 63% of the variance in forecast performance (Report adjust R^2), ($p < 0.001$). Adoption intensity of GenAI made a rational, authoritative positive impact on forecast performance ($\beta = 0.31$, $p < 0.001$) and also multi-domain for the hardest strongest direction effect ($\beta = 044$, $p < 0.01$is at: conf). GenAI adoption represents a similarly immensely positive influence on the quality of data ($\beta = 0.49$, $p < 0.001$; H3a), and the indirect effect of experiential adoption on conditional performance through data quality has also been significantly tested.

In adaptive fragility risk, the relationship between GenAI adoption and data quality has been significantly mitigated (interaction $\beta = −0.18$, $p < 0.01$), showing that a homogenous architecture reduces the benefits of improved data quality as a result of GenAI adoption (H4). Expertise calibration of humans mitigated the direct relationship between GenAI adoption and performance, with a stronger effect observed in the high-calibration sample ($\beta = 0.47$, $p < 0.001$) compared to the low-calibration sample ($\beta = 0.19$, $p < 0.05$) (H5).

The quadratic effect was significant for human calibration (β_quadratic = -0.14, $p < 0.05$), verifying the proposed inverse-U curve (H6). The results of all path estimates and model fit statistics can be seen in Figure 2 (or in Table 3, whichever you prefer).

Table 3: ***Results of Mediation***

Paths	Direct β	*p*	Indirect β	*p*	Mediation
Direct effect X → Y	0.31	< 0.001	—	—	—
Direct effectX → M1 → Y	—	—	0.22	< 0.001	Partial mediation (H3b supported)
Direct effect M2 × X → M1	−0.18	< 0.01	—	—	—
Direct effect W × X → Y (high calibr.)	0.47	< 0.001	—	—	—
Direct effectW × X → Y (low calibr.)	0.19	< 0.05	—	—	—
Quadratic W → Y	−0.14	< 0.05	—	—	Inverted-U (H6 supported)

Note. WOM = without mediator; WM = with mediator. Bootstrapped indirect effect (250 iterations, bias-corrected). X = GenAI Adoption Intensity; M1 = Multi-Domain Data Quality; M2 = Adaptive Fragility Risk; W = Human Expertise Calibration; Y = Forecast Performance.

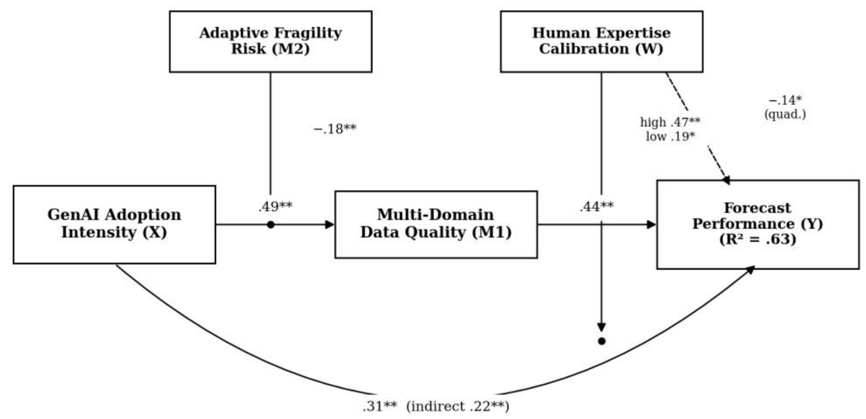

Figure 2. Structural model

6. Implications

6.1. Theoretical Implications

Hence, this research stages a theory in several areas, as the multi-domain quality of data is based on a mediating mechanism instead of a co-variation, contrary to earlier studies, which treat the quality of data like an antecedent to AI efficiency. The framework represents the quality of data quality as a mediator based on latent space integration, alignment of loss, and information handling, operationalized by VAE latent-space fusion. On the other hand, the evidence of partial mediation on the usage bias-corrected bootstrapping highlights that GenAI adoption signifies residual effects forecasting performance through improving the quality of data, which imposes a significant residual direct impact that highlights additional pathways. Through the application of multi-group SEM to evaluate human calibration boundary situations. The density of commonly utilized Transformer/GAN models paradoxically eliminates the performance of the individual, as architectural homogeneity is high, which adds an institutional-level significance to the literature.

6.2. Managerial Implications

Hence, the CFOs and CROs are not designed to aim for a single architecture but multiple architectures at the same time (for example, the usage of GAN for scenario generation and Transformer regarding sentiment analysis) among different teams to decrease correlated errors of prediction. To avoid adaptive fragility by correlated errors, which result through the commonality in FinBERT/Transformers, intentional architectural diversity. Ultimately, the AI oversight must encompass verification of AI-generated output verification, which is responsible for AI-aided decisions, and systematic adjustment after errors. The epistemic opacity and look-ahead bias due to inexact timestamps, that is need to be avoided. The benefit of AI lies in favor of senior partners, but at the same time, junior analysts go-through the replacement risk; junior analyst roles should be constructed like validators and critics. The campaigns of awareness should be developed to educate financial professionals on optimal human calibration levels, which focus on both under-reliance and over-reliance on AI outputs that are detrimental to forecasting performance.

6.3. Policy Implications

It has been stressed in findings that align with EU AI Act Article 14 on Human Oversight: human oversight should be dynamic calibration rather than focusing on static presence. Hence, the adaptive fragilities are considered to be second-order risks that need to be monitored, where central banks could examine model dependencies and test regime-transition vulnerabilities by algorithmic registries. The synthetic data should be disclosed through GAN-augmented ratios and also mutual information scores, as recommended by SEACEN (2026). Hence, the policymakers must develop an educational environment for financial institutions about the seriousness of architectural homogeneity and its prospects systemic risk in the view of market volatility. The government and regulatory bodies can provide incentives to encourage architectural diversity, independent validation protocols, and investment in cross-domain data quality infrastructure.

7. Conclusion

It aims to understand GenAI financial forecasting using a structural equation modeling framework. The GenAI adoption is partially explained by multi-domain data quality improvement for forecasting performance, concluding support for .The advantage is weakened if institutions adopt converging architectures, as adaptive fragility risk is confirmed to be a significant negative moderator (H4). Significantly, the performance enhancement in an inverted-U manner with human calibration: performance deteriorates under both under-reliance and over-reliance on the judgment of humans, making moderately calibrated human oversight optimal for forecasting performance rather than either a purely human or purely machine-based prediction (H5 and H6).

The present study on GenAI financial forecasting demonstrates the need for adaptable AI strategies, practices, and policies for contemporary challenges and solutions. This paper makes three primary contributions. First, GenAI adoption partially impacts forecasting performance through multi-domain data quality improvement. Second, the benefit is weakened when institutions choose converging architectures because adaptive fragility risk lowers data quality gains. Third and crucially, performance improves in an inverted-U manner with human calibration — it is best to have moderate, but not excessive, human judgment for forecasting performance. These findings carry clear implications for regulators, the design of corporate governance, and efforts at re-evaluating the human contribution to AI-driven financial forecasting.

8. Limitations and Future Recommendations

Like any other research, this study is no exception to some inherent limitations. This cross-sectional approach (N = 285, 2025–2026) captures one moment in the GenAI adoption cycle; thus, causal interpretations are restricted, as is its generalizability under various market circumstances. CFA-derived self-reports are prone to common method bias despite procedural and analytical remedies. GAN-based synthetic data augmentation was applied for extreme cases of anomaly and the prediction performance of tail events improved only slightly. This particular study relies on AI-assisted forecasting; the results cannot be applied in situations in which human intuition functions without AI assistance.

Future studies should extend the scope of analysis through international comparative regional panels (emerging vs. developed countries); analyze heterogeneity by sector (sovereign debt vs. blockchain monitoring); perform technical evaluation studies using dual-attention transformers vs. Res2Net-Transformer hybrid architectures; and conduct quasi-experiments (for staggered interventions/regulatory events) instead of cross-sectional SEM. A longitudinal design is a better option for future research in order to establish causal directionality. Future researchers should consider measuring actual forecasting accuracy using archival data to complement self-report measures, and should also explore heterogeneity by institution size and regulatory jurisdiction to gain deeper understanding of the boundary conditions identified in this study.

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethical Approval

This article does not contain any studies with animals performed by any of the authors. Informed consent was obtained from all individual participants included in the study.

Funding

This study did not receive any funding from any organization.

References

Abu-Dabaseh, F., et al. (2025). AI-driven digital banking: Predictive analytics and automation in financial services. Journal of Financial Technology, 12(2), 45–67.

Akpan, M. (2025). Generative artificial intelligence and the future of financial forecasting: Evidence from large language models. Risk Governance and Control Financial Markets & Institutions, 15(4), 142. https://doi.org/10.22495/rgcv15i4p13

Alassuli, A., Eltweri, A., Thuneibat, N. S., Al-Hajaya, K., & Ismail, S. M. (2026). Artificial intelligence applications and financial forecasting accuracy in banking platforms: Evidence from Jordan. Administrative Sciences, 16(3), 122. https://doi.org/10.3390/admsci16030122

Alt, R., Beck, R., & Smits, M. T. (2018). FinTech and the transformation of the financial industry. Electronic Markets, 28(3), 235–243.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin, 103(3), 411–423.

Anwar, S. (2025). Generative AI-driven financial market simulation and forecasting using cloud-based computational frameworks. International Journal of Engineering Works, 12(11), 197–206. https://doi.org/10.5281/zenodo.17661001

Asjad, M. A., Khan, M. W., Moon, M. A., Fatimah, H., & Parveen, S. (2025). Easy payments, hard consequences: The behavioral and financial effects of buy now, pay later in Pakistan. Social Science Review Archives, 3(4), 3309–3330.

Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation. NBER Working Paper, 24449.

Collins, J., Coussement, K., & Kumar, V. (2021). Big data and predictive analytics in financial markets. Journal of Financial Markets, 28(1), 100–120.

Farooq, A., & Moon, M. A. (2025a). Addicted to the fake: Coaction theory and the psychology behind counterfeit consumption. Pakistan Journal of Commerce and Social Sciences, 19(3), 648–672.

Farooq, A., & Moon, M. A. (2025b). From impulse to imitation: Hedonic motivation as a bridge to counterfeit consumption. Journal of Asian Development Studies, 14(3), 263–276.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50.

Furman, J., & Seamans, R. (2019). AI and the economy. Innovation Policy and the Economy, 19(1), 161–191.

Gong, Y., Hu, P., Zhang, Z., Liu, P., Li, Z., Zhang, R., Yin, J., & Li, M. (2026). Data-centric generative and adaptive detection framework for abnormal transaction prediction. Electronics, 15(3), 633. https://doi.org/10.3390/electronics15030633

Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. Journal of Behavioral and Experimental Finance, 32, 100577.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., & Thiele, K. O. (2017). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. Journal of the Academy of Marketing Science, 45(5), 616–632.

Jian, J. Y., & Landis, D. A. (2003). Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics, 7(4), 341–356.

Khalil, R. (2026). AI driven sentiment analysis in financial markets: Using transformer base models and social media signals for stock market predictions. Journal of Modelling in Management, 1(25), 1–25. https://doi.org/10.1108/jm2-08-2025-0415

Kline, R. B. (2015). Principles and practice of structural equation modeling (4th ed.). Guilford Press.

Mo, H., & Ouyang, S. (2025). (Generative) AI in financial economics. Journal of Chinese Economic and Business Studies, 23(4), 509–587. https://doi.org/10.1080/14765284.2025.2569006

Moon, M. A. (2026). The dark pull of virtual aisles: How interactivity fuels compulsive buying in immersive shopping worlds. Journal of Consumer Behaviour, 25(3), 1414–1431.

Moon, M. A., Majeed, J., & Attiq, S. (2024). Enigma unveiled: Decoding trust and relationship dynamics in the mobile banking sector of an emerging economy. Pakistan Journal of Commerce and Social Sciences, 18(2), 380–407.

Moon, M. A., Malik, M. H., Moin, G., & Abbas, A. (2025). Impact of consumer trust on e-commerce purchase intentions: The role of influencer marketing. Journal of Management Science Research Review, 4(4), 1920–1940.

Moon, M. A., Sajid, A., Bashir, S. S., Saddiqi, A., & Zeeshan, M. (2026). The emotional cost of greenwashing: Airline sustainability advertising and consumer–brand relationships. Journal of Management & Social Science, 3(1), 1–20.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). McGraw-Hill.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. Annual Review of Psychology, 63, 539–569.

Rohan, A., Hossen, Md. D., Pranto, Md. N., Hossain, B., Yoshi, A. M., & Islam, R. (2026). Artificial intelligence in financial market prediction: Advancements in machine learning for stock price forecasting. Frontiers in Artificial Intelligence, 8. https://doi.org/10.3389/frai.2025.1696423

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics. Allyn & Bacon/Pearson Education.

Tariq, S., Nadeem, U., & Moon, M. A. (2026). AI chatbots and customer experience: The role of service quality dimensions, trust, and generational differences. Pakistan Journal of Social Science Review, 5(1), 14–37.

Tassen, M., Irfan, M., Moon, M. A., Munawar, M., Ahmad, A., & Mishaal, M. (2025). Integrating the SOR framework to assess social media marketing in the aviation sector: A study of Pakistani travelers. Center for Management Science Research, 4(1), 64–78.

ud Din, T., Attiq, S., & Moon, M. A. (2022). The influence of destination experience on destination loyalty in an emerging economy. Pakistan Journal of Commerce and Social Sciences, 16(4), 552–575.

Wang, J., Shi, Y., Jiang, X., & Venkatesh, V. G. (2026). How does artificial intelligence capacity enhance the production system resilience and operational performance? A human-organization-technology fit perspective. International Journal of Information Management, 87, 103023. https://doi.org/10.1016/j.ijinfomgt.2025.103023

Yao, K.-C., Hung, H.-C., Wang, C.-H., Huang, W.-L., Liang, H.-T., Chu, T.-H., Chen, B.-S., & Ho, W.-S. (2025). Application of generative AI in financial risk prediction: Enhancing model accuracy and interpretability. Information, 16(10), 857. https://doi.org/10.3390/info16100857

Zhang, Y., Qiu, Z., Park, D., & Tian, S. (2026). Role of artificial intelligence in finance: Selective literature review and implications for Asia's financial stability (pp. 1–34). The South East Asian Central Banks (SEACEN) Research and Training Centre.