



Bridging the Validity Gap: Mitigating Strategic Overfitting in Management Analytics through Advanced Predictive Metrics

Richard Berimah Twum¹²³, Abdul-Aziz Ibn Musah², Ernestina Hope Turkson²

¹National Investment Bank PLC, Accra, Ghana

²TALI Graduate School, Dominion University College (now Southshore University College), Accra, Ghana

³Department of Statistics, University of Cape Coast, Cape Coast, Ghana

Email: richardberimah@gmail.com

Abstract

Purpose: In the age of data-driven management, leaders of organizations are relying more and more on structural models to predict the return on investments in knowledge. But there is a serious "validity gap": models with high explanatory power (R^2) often don't accurately predict outcomes that weren't used to train the model, which can lead to strategic overfitting and decision risk. This study shows that advanced predictive metrics (Q^2 Predict, RMSE, and CVPAT) can be used to make diagnoses in Knowledge-Based Transformation Models (KBTM).

Approach: We use Partial Least Squares Structural Equation Modelling (PLS-SEM) along with PLS-Predict and Cross-Validated Predictive Ability Testing (CVPAT) on a dataset of 310 banking professionals in Ghana. We look at how well the model predicts outcomes that weren't used to train it across nine multivariate constructs, separating internal knowledge processes from external performance outcomes.

Findings: show that there is a lot of variation at the construct level. Knowledge-intensive processes, like Knowledge Creation ($Q^2 = 0.834$), are much better at predicting outcomes than naive benchmarks ($p < 0.001$). On the other hand, performance outcomes (like Employee Performance, $Q^2 = 0.191$) don't do a good job of predicting what will happen, which means that the model is very uncertain.



Value: This study goes beyond traditional fit indices to create a set of rules for predictive validation in management analytics. We show that only using explanatory metrics hides "blind spots" in predicting performance. We suggest a framework for analytics maturity that encourages leaders to tell the difference between outcome variables that can change and process variables that can change before making strategic decisions.

Keywords

Predictive Analytics, PLS-SEM, Q² Predict, Knowledge Management, Model Validation, Decision Risk.

1. Introduction

The focus of organizational research is changing from explaining things to predicting them (Shmueli et al., 2016). Structural Equation Modelling (SEM) has helped to move theory forward by explaining variance (R^2) and confirming hypothesized relationships. However, modern strategic decision-making needs models that can accurately predict outcomes on data that hasn't been seen before (Hair et al., 2019). In the world of Knowledge-Based Transformation (KBT), where companies put a lot of money into intangible assets, being able to predict the return on these projects is not just an academic exercise; it is a strategic necessity.

Nonetheless, a significant deficiency endures in management analytics. A lot of real-world research uses in-sample fit indices (like SRMR, NFI, and R^2), but these don't guarantee that the model will be accurate when tested on new data (Lienggaard et al., 2021). Overfitting is when a model fits the current dataset perfectly but doesn't work for other datasets. This puts practitioners at risk of making bad decisions by acting on insights that seem statistically sound but don't actually predict what will happen. Also, we don't know much about how predictive metrics work with different types of latent constructs in the same multivariate model. Do internal knowledge processes predict performance outcomes as well as external ones?

This study looks at how statistically significant and useful advanced evaluation metrics are for diagnosis. These metrics are Q² Predict, Root Mean Square Error (RMSE), and the Cross-Validated Predictive Ability Test (CVPAT) within a Knowledge-Based Transformation Model. We evaluate the predictive relevance of KBT constructs utilizing data from Ghana's commercial banking sector and compare the statistical significance of these metrics to naive benchmarks. The research enhances the body of knowledge in business analytics and management science through three key contributions:

Validating Metric Utility: We show that Q² Predict and RMSE are statistically significant as important diagnostic tools for finding model overfitting in management research.



Finding Construct Heterogeneity: We show that predictive power isn't the same for all cases; knowledge processes are very predictable, but performance outcomes are less stable and more uncertain.

Reducing Decision Risk: We give you a way to measure the maturity of your analytics and tell you when a model is strong enough for strategic forecasting and when it needs to be improved.

2. Literature Review and Hypotheses Development

2.1. The Explanation-Prediction Dichotomy

Traditional SEM prioritizes parameter consistency and model fit to test theory (Henseler et al., 2020). However, high R^2 values can coexist with poor predictive power (Shmueli et al., 2019). Predictive Modeling Theory argues that model quality should be assessed by its ability to predict new observations, not just reproduce existing ones (Shmueli, 2010). In knowledge management, this distinction is vital; explaining why knowledge sharing occurs is theoretically distinct from predicting how much it will improve future performance. Relying on explanatory metrics for predictive decisions introduces systematic bias into organizational planning.

2.2. Advanced Evaluation Metrics in PLS-SEM

Recent developments in Partial Least Squares (PLS) path modelling provide reliable predictive assessment tools to address this gap:

Q^2 Predict: Evaluates predictive relevance through blindfolding or cross-validation. Predictive relevance is indicated by values that exceed 0, with thresholds of 0.02 (weak), 0.15 (moderate), and 0.35 (substantial) (Hair et al., 2019).

RMSE and MAE are error-based metrics that quantify the discrepancy between predicted and actual values. A higher degree of accuracy is indicated by lower values. (Sharma et al., 2023) **Statistical Significance (CVPAT):** Recent developments have enabled the evaluation of whether predictive metrics are significantly superior to naive benchmarks (e.g., linear model or indicator average) through the use of t-tests and p-values. This step is essential for verifying that the complex model is superior to simple heuristics.

2.3. Construct-Level Predictive Heterogeneity

Not all constructs are equally predictable. Theoretically, knowledge processes (e.g., creation, codification) are often internal and controlled organizational behaviors, potentially yielding higher predictive accuracy. In contrast, performance outcomes (e.g., employee performance) are influenced by



exogenous factors (market conditions, personal circumstances), potentially lowering predictive relevance (Mbilla et al., 2020). This suggests a Construct-Level Predictive Heterogeneity, where a single model may contain both robust and weak predictive paths.

Based on this theoretical grounding, we propose:

H1: Advanced predictive metrics (Q^2 Predict, RMSE) will show statistically significant predictive relevance for Knowledge Management constructs.

H2: Predictive relevance will be significantly higher for Knowledge Process constructs compared to Performance Outcome constructs.

H3: PLS Predict metrics will demonstrate statistically significant superiority over naive benchmarks for multivariate constructs, particularly knowledge-based ones.

3. Methodology

3.1. Research Design and Sample

The KBTM's predictive validity is examined in this study using a quantitative cross-sectional design. The Accra Metropolitan District in Ghana was the site of data collection, which involved 310 employees from 10 different commercial banks. Because of the importance of knowledge assets and data-driven performance metrics, the banking sector was chosen. To minimise selection bias, stratified random sampling was used to ensure representation across hierarchy and departments. The response rate was 98.41%, which kept the non-response bias to a minimum. The sample consists of 42% female respondents and 58% male respondents. There is enough institutional knowledge to assess the constructs because 56.1% of the participants have 5-10 years of experience. Members of lower-level management make up 26.8%, those in middle-level management 17.4%, and those in upper-level management 3.5%.

Table 4.1: Demographic Characteristics of Bank Employees

Demographics	Frequency	Percent
Job Rank		
Junior	83	26.8
Middle Management	54	17.4



Officer	162	52.3
Senior Management	11	3.5
Total	310	100.0

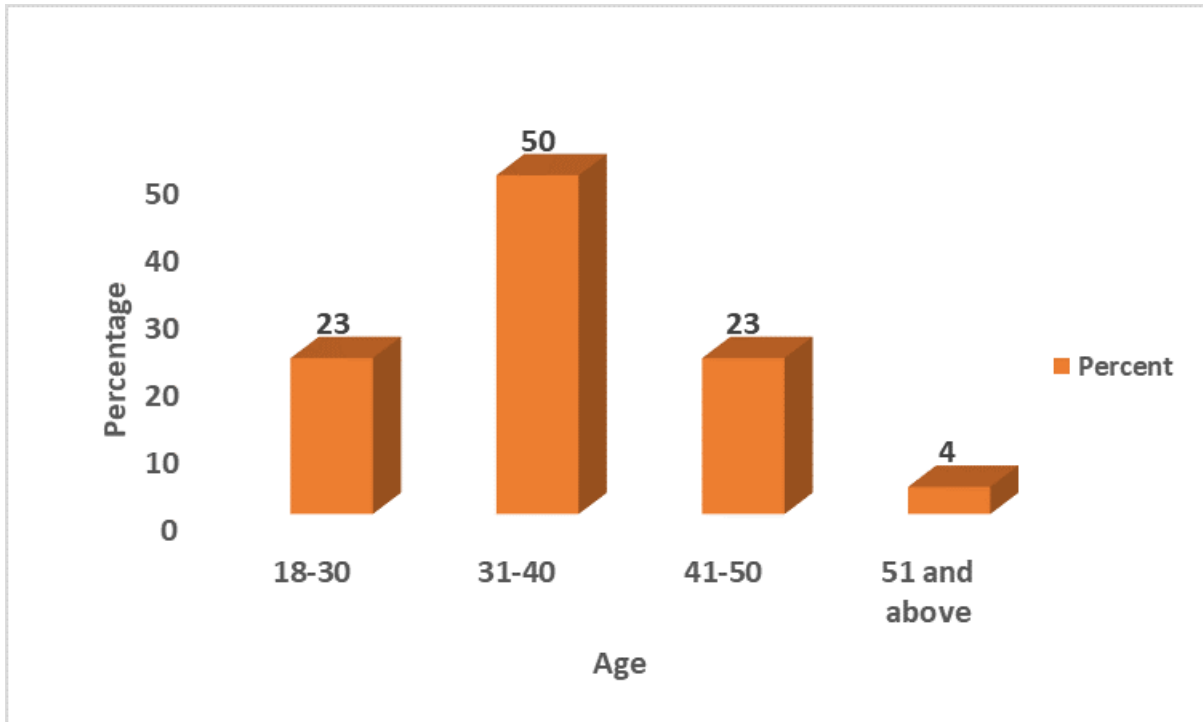


Figure 4.1: Background characteristics of respondent by age (%)

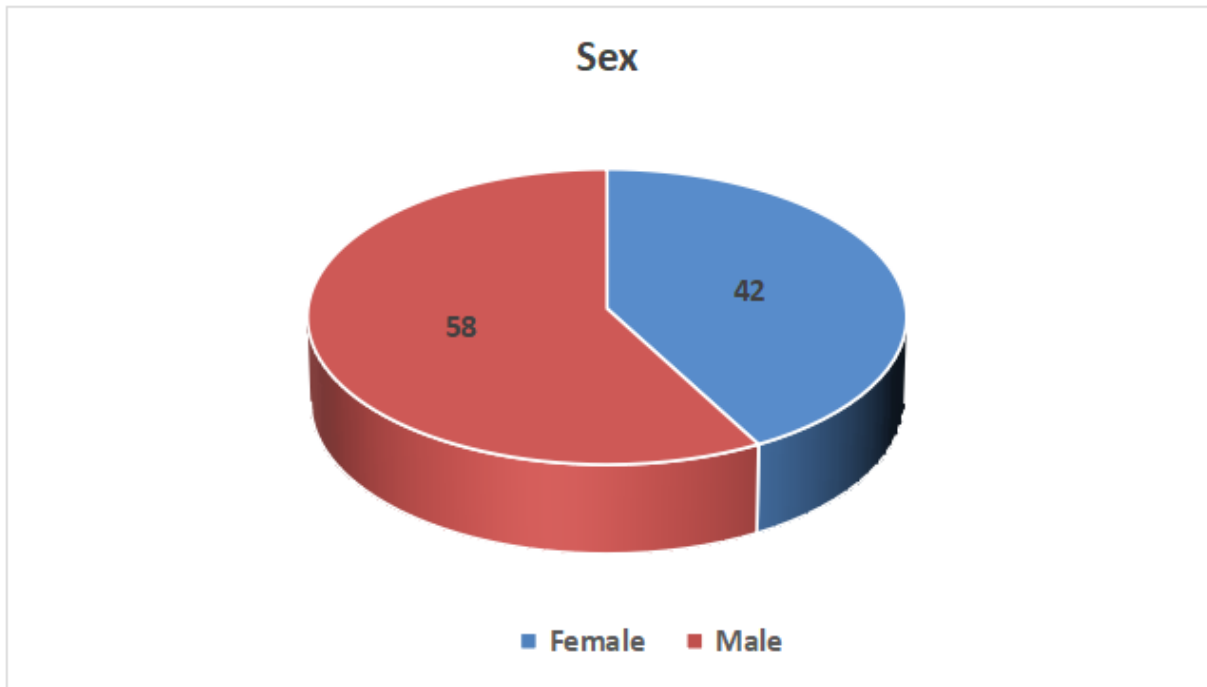


Figure 4.2: Background characteristics of respondent by gender (%)

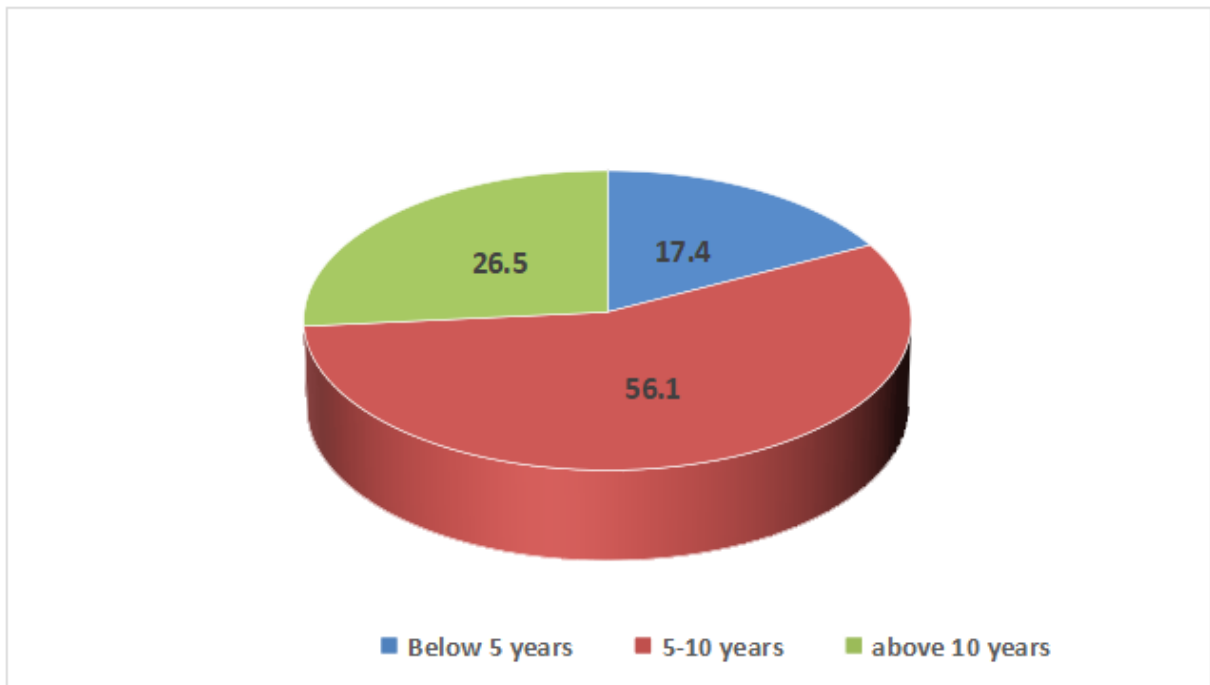


Figure 4.3: Background characteristics of respondent by Experienced

3.2. Measurement and Model

The KBTM includes nine latent constructs:

Knowledge Processes: Creation (KC), Acquisition (KAC), Sharing (KS), Application (KAP), Codification (KCO), Retention (KR).

Outcomes: Job Satisfaction (DJS), Employee Performance (EP).

Context: Government Policies (GP).

3.3. Analytical Procedure

Analysis was conducted using SmartPLS 4.

Predictive Assessment: PLS-Predict procedure with 10-fold cross-validation.

Metrics Computed: Q^2 Predict, RMSE, MAE.

Significance Testing: CVPAT was used to test the statistical significance of prediction errors against Linear Model (LM) and Indicator Average (IA) benchmarks using t-tests.

Decision Rules: $Q^2 > 0$ indicates predictive relevance; $p < 0.05$ indicates significant superiority over benchmarks.

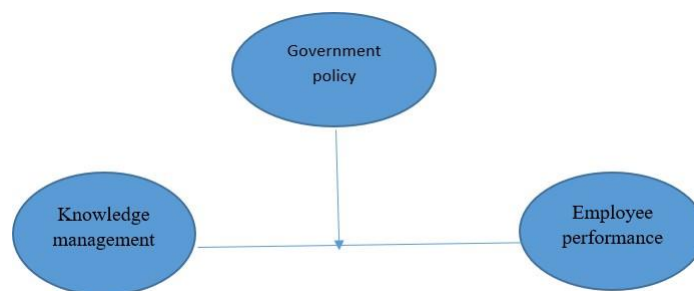


Figure 3.4: Moderating Role of Government

4. Results

4.1. Predictive Relevance (Q^2 Predict)

The analysis of Q^2 Predict values reveals distinct levels of predictive relevance across constructs.



Knowledge Constructs: All knowledge-related constructs demonstrated substantial predictive relevance. Knowledge Creation (KC) had the highest Q^2 (0.834), followed by Knowledge Retention (KR = 0.827) and Knowledge Codification (KCO = 0.803).

Performance Constructs: Employee Performance (EP) and Job Satisfaction (DJS) showed lower predictive relevance ($Q^2 = 0.191$ and 0.172 , respectively), falling into the weak-to-moderate range.

Contextual Construct: Government Policies (GP) showed moderate relevance ($Q^2 = 0.328$)

method for distinguishing between robust and weak predictive paths. This aligns with Sharma et al. (2023), who argue for significance testing in predictive assessment.

Table 1: Latent Variable (LV) Prediction Summary.

Construct	Type	Q^2 Predict	RMSE	MAE	Predictive Relevance
KC	Knowledge	0.834	0.410	0.314	Substantial
KR	Knowledge	0.827	0.418	0.316	Substantial
KCO	Knowledge	0.803	0.446	0.318	Substantial
KS	Knowledge	0.792	0.458	0.342	Substantial
KAC	Knowledge	0.744	0.508	0.379	Substantial
KAP	Knowledge	0.709	0.543	0.414	Substantial
GP	Context	0.328	0.828	0.635	Moderate
EP	Outcome	0.191	0.905	0.675	Weak/Mod
DJS	Outcome	0.172	0.916	0.737	Weak/Mod

4.2. Prediction Error Metrics (RMSE and MAE)

Error metrics corroborated the Q^2 findings. Knowledge constructs exhibited low RMSE values (e.g., KC RMSE = 0.410), indicating precise predictions. Conversely, performance constructs had significantly higher RMSE values (e.g., EP RMSE = 0.905; DJS RMSE = 0.916). This suggests that while the model explains variance in performance ($R^2 = 0.425$), the error margin for predicting individual performance outcomes remains high.

4.3. Statistical Significance of Predictive Ability (CVPAT)

Using CVPAT, we tested whether the PLS-SEM model's predictions were significantly better than naive benchmarks (Linear Model and Indicator Average). This step is critical for validating that the model complexity is justified.



Table 2: Predictive Abilities of PLS Predict VS CVPAT (Loss Difference vs Linear Model).

Construct	Average Loss Difference	t-value	p-value	Significant?
KC	0.536	19.980	0.000	Yes
KR	0.489	16.997	0.000	Yes
KCO	0.500	15.809	0.000	Yes
KS	0.552	19.096	0.000	Yes
KAC	0.576	19.191	0.000	Yes
KAP	0.562	20.199	0.000	Yes
DJS	0.114	1.990	0.048	Yes (Marginal)
EP	0.002	0.031	0.975	No
GP	-0.001	0.017	0.986	No
Overall	0.373	18.766	0.000	Yes

4.4. Hypothesis Testing

H1 Supported: Advanced metrics showed significant predictive relevance for Knowledge Management constructs ($Q^2 > 0.35$, $p < 0.001$).

H2 Supported: Predictive relevance was significantly higher for Knowledge Processes (Avg $Q^2 \approx 0.78$) compared to Performance Outcomes (Avg $Q^2 \approx 0.18$).

H3 Partially Supported: PLS Predict metrics demonstrated statistically significant superiority over benchmarks for knowledge-based constructs, but not for Performance or Government Policy constructs.

5. Discussion

5.1. The Diagnostic Value of Predictive Metrics

The study confirms the effectiveness of Q^2 Predict and RMSE in assessing model risk. The high Q^2 values for knowledge constructs (KC, KR, KCO) indicate that the KBTM accurately identifies knowledge process drivers. Internal organizational behaviors are well-defined by model variables. The lower Q^2 and non-significant CVPAT results for Employee Performance suggest that the model explains some variance (R^2) but has limited predictive power for individual performance. The KBTM may not account for external factors like personal life and market volatility that affect performance. Strategic risk: using this model to predict performance.



5.2. Theoretical Implications: Construct Heterogeneity

This study extends Predictive Modeling Theory by demonstrating that predictive relevance is construct-dependent within a single model. It challenges the assumption that a model with good fit indices is uniformly predictive. We introduce the concept of Construct-Level Predictive Heterogeneity, arguing that management models often contain "zones of high predictability" (processes) and "zones of uncertainty" (outcomes). The statistical significance tests (CVPAT) provide a rigorous method for distinguishing between these zones. This aligns with Sharma et al. (2023), who argue for significance testing in predictive assessment to prevent the deployment of overfitted models.

5.3. Practical Implications for Analytics and Management

Regarding the predictions made by bank managers and analysts concerning knowledge processes (for example, "If we invest in codification, retention will increase"), you should have faith. Before making any predictions about an individual's performance based solely on knowledge metrics, however, exercise caution. The results of the model suggest that knowledge is a necessary but not sufficient predictor of performance. The root mean square error (RMSE) on performance outcomes needs to be reduced, and additional variables, such as economic indicators and personal traits, are required. Relying solely on R² is not a good idea for researchers. Standard requirements include reporting the Q² Predict and RMSE values. In order to determine whether or not your model is significantly superior to a simple mean prediction, you can use CVPAT. It is important to recognize the limitations in predictive validity in the event that a construct fails the CVPAT test, as was the case with EP in this study.

Policymakers: Initiatives based on knowledge are levers that are both predictable and manageable. In order to acknowledge the greater degree of variability in human performance outcomes, performance initiatives need to incorporate broader interventions that go beyond basic knowledge management.

6. Conclusion and Recommendations

6.1. Conclusion

This study shows that advanced evaluation metrics like RMSE and Q² Predict are good indicators of model quality in the Knowledge-Based Transformation context. On the flip side, there is a disparity in predictive power; while knowledge processes are very predictable, performance outcomes are not. Because CVPAT employs statistical significance testing, we can validate that the KBTM is robust for knowledge constructs. But we show that performance forecasting has its limits. It is crucial to distinguish between the two in order to avoid the "illusion of validity" in management analytics.



6.2. Recommendations

Metric Adoption: Researchers and practitioners must adopt Q^2 Predict and RMSE as standard reporting requirements alongside R^2 to assess true model utility.

Model Refinement: To improve performance prediction, future models should integrate external variables (e.g., economic indicators, personal traits) to account for the unexplained variance identified by the high RMSE.

Significance Testing: Always test predictive metrics against naive benchmarks (CVPAT) to ensure the model adds value beyond simple averages before using it for strategic decision-making.

6.3. Limitations and Future Research

The study is limited by its cross-sectional nature and self-reported data, which may introduce common method bias despite procedural remedies. Future research should use longitudinal data to assess predictive stability over time and explore objective performance metrics (e.g., sales figures) to validate the self-reported findings.

References

1. Hair, J. F., et al. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*.
2. Henseler, J., et al. (2020). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*.
3. Liengard, B. D., et al. (2021). Prediction: Covariance-based or Partial Least Squares Structural Equation Modeling? *Journal of Travel Research*.
4. Mbilla, N., et al. (2020). Impact of KM practices on bank performance in Ghana. *African Journal of Business Management*.
5. Sharma, P. N., et al. (2023). Predictive model assessment and selection in composite-based modeling using PLS-SEM. *European Journal of Marketing*.
6. Shmueli, G. (2010). To explain or to predict? *Statistical Science*.
7. Shmueli, G., et al. (2019). Predictive model assessment in PLS-SEM. *European Journal of Information Systems*.