Economics And Business Management Journal (EBMJ)
June 2026 | Vol. 3 | No. 02
E-ISSN : 2828-0288
DOI: https://doi.org/10.66084/ebmj.v3i02.639

Modeling the Relationship Between Product Innovation, Digital Marketing, and SME Revenue Growth Using Synthetic Data

Riskia Putra[1], Noorhani Dyani Laksmi[2], Andi Tenri Ampa[3], Rifatul Muna[4]
[1]Universitas Islam Negeri Kiai Haji Achmad Siddiq Jember
[2]Universitas Negeri Malang,
[3]Universitas Halu Oleo,
[4]Sekolah Tinggi Islam Kendal
e-mail: *[1]riskia@lecturer.uinkhas.ac.id, [2]nennymakmun@gmail.com,
[3]anditenri.ampa@uho.ac.id, [4]rifatul91muna@gmail.com.

Abstrak

Small and medium-sized enterprises (SMEs) are widely regarded as engines of employment and value creation in emerging economies, yet the mechanisms linking their innovation behaviour and marketing capabilities to financial performance remain incompletely modelled. This study develops and tests a structural model in which product innovation and digital marketing jointly shape SME revenue growth, including the moderating role of digital marketing on the innovation–growth path. Because access to verified firm-level financial micro-data is constrained by confidentiality and reporting gaps, the analysis is conducted on a synthetic dataset of 250 simulated SMEs generated from a transparent data-generating process with known ground-truth parameters. The synthetic design allows the estimation procedure to be validated against the parameters used to create the data, an evaluation that is impossible with conventional observational samples. Hierarchical ordinary least squares regression shows that product innovation ($\beta = 0.33$, $p < 0.001$) and digital marketing ($\beta = 0.55$, $p < 0.001$) are both positively associated with revenue growth, together explaining 56.8% of its variance. A significant positive interaction ($b = 1.89$, $p = 0.001$; $\Delta R^2 = 0.018$) indicates that the returns to product innovation are amplified when firms invest more heavily in digital marketing. Recovered coefficients closely track the parameters embedded in the data-generating process, confirming that the modelling pipeline is well specified. The paper positions synthetic data as a methodological complement not a substitute for empirical SME research, useful for pipeline validation, teaching, and pre-registration of analysis plans. All data are simulated and contain no real firms or individuals.

Keywords: product innovation; digital marketing; revenue growth; SME; synthetic data; moderation analysis

INTRODUCTION

Small and medium-sized enterprises (SMEs) account for the overwhelming majority of business establishments worldwide and contribute a substantial share of employment and gross value added, particularly in developing economies. In Indonesia, for example, SMEs are a central pillar of the national economy, absorbing a large fraction of the workforce and supplying a significant proportion of gross domestic product. Despite their economic weight, SMEs face persistent constraints in capital, managerial capacity, and market access that make their growth trajectories highly sensitive to strategic choices. Understanding which capabilities most reliably translate into revenue growth is therefore both a scholarly and a practical priority.

Two strategic levers have attracted sustained attention in the literature on firm growth: product innovation and marketing capability. Product innovation the introduction of new or meaningfully improved goods and services is theorised to generate competitive advantage by differentiating a firm's offering and opening new demand segments. Marketing capability, and increasingly its digital expression, governs how effectively a firm converts that offering into realised sales by reaching, persuading, and retaining customers. With the diffusion of low-cost digital channels, social commerce, and data-driven advertising, digital marketing has become especially salient for resource-constrained SMEs that cannot match the marketing budgets of large incumbents.

While each lever has been studied extensively in isolation, three gaps motivate the present work. First, the joint contribution of product innovation and digital marketing to revenue growth, and in particular their interaction, is underexplored: it is plausible that innovation yields larger returns precisely when a firm has the digital marketing reach to commercialise it, yet this conditional logic is rarely modelled explicitly. Second, empirical SME studies are frequently hampered by limited access to reliable firm-level financial data, owing to confidentiality, weak reporting, and survivorship bias; the resulting samples are often small, noisy, and difficult to share for replication. Third, even when data are available, analysts seldom have access to a ground truth against which to verify that their estimation pipeline recovers the relationships it claims to measure.

This study addresses these gaps by adopting a synthetic-data approach. Rather than presenting simulated numbers as empirical findings, we use a transparent data-generating process (DGP) with known structural parameters to (i) instantiate a plausible model of how product innovation and digital marketing shape SME revenue growth, and (ii) demonstrate that a standard hierarchical regression pipeline correctly recovers those parameters. Synthetic data of this kind have become an established tool in statistics and machine learning for benchmarking methods, protecting privacy, and supporting reproducible, pre-registered analysis. We are explicit that the substantive coefficients reported here are properties of the simulation and must not be interpreted as estimates of real-world effect sizes; their value lies in validating the methodology and illustrating the modelling logic.

Accordingly, the contributions of this paper are threefold. (1) We formalise a moderated structural model linking product innovation and digital marketing to SME revenue growth and articulate three testable hypotheses. (2) We specify a fully reproducible synthetic DGP and show, through hierarchical regression and interaction analysis, that the modelling pipeline recovers the embedded parameters with high fidelity. (3) We discuss how synthetic data can responsibly complement empirical SME research for pipeline validation, instrument testing, teaching, and the design of future field studies while delineating its limitations. The remainder of the paper reviews the relevant literature and derives hypotheses (Section 2), describes the synthetic design and analytical strategy (Section 3), reports results (Section 4), discusses implications (Section 5), and concludes (Section 6).

LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Product Innovation and Firm Growth

The link between innovation and firm performance is among the most durable propositions in the strategy and economics literatures. Drawing on the resource-based view and Schumpeterian theory, product innovation is conceived as a source of temporary monopoly rents: a differentiated or novel offering loosens price competition, attracts new customers, and can command premium margins. For SMEs specifically, product innovation is often a survival strategy, enabling small firms to occupy niches that larger competitors neglect. Empirical syntheses generally report a positive though heterogeneous association between innovation activity and growth outcomes such as sales and revenue. Heterogeneity arises because innovation is costly and risky, and because its commercial payoff depends on complementary capabilities, notably the firm's ability to bring the innovation to market. This reasoning yields the first hypothesis.

H1. *Product innovation is positively associated with SME revenue growth.*

Digital Marketing and Revenue Growth

Marketing capability translates a firm's offerings into realised demand, and in the contemporary environment this capability is increasingly digital. Digital marketing encompasses search and social advertising, content and influencer marketing, marketplace optimisation, customer-relationship management, and analytics-driven targeting. For SMEs, digital channels lower the fixed costs of market access and allow precise, measurable outreach that was historically the preserve of large firms. A growing body of work associates SME adoption of digital marketing with improved customer acquisition, broader market reach, and higher sales performance. Because digital marketing acts on the demand-realisation side of the value chain, its effect on revenue is often more direct and immediate than that of innovation. We therefore expect a strong positive association.

H2. *Digital marketing is positively associated with SME revenue growth.*

The Moderating Role of Digital Marketing

Beyond their independent contributions, product innovation and digital marketing may be complementary. An innovative product that customers never encounter cannot generate revenue; conversely, marketing reach applied to an undifferentiated product yields limited returns. The complementarity logic rooted in the literature on complementary assets and marketing R&D integration suggests that the marginal return to product innovation rises with the intensity of digital marketing, because superior reach and targeting allow a novel offering to be commercialised more fully and rapidly. Stated as a moderation hypothesis, digital marketing should strengthen the positive relationship between product innovation and revenue growth.

H3. *Digital marketing positively moderates the relationship between product innovation and SME revenue growth, such that the effect of product innovation on revenue growth is stronger at higher levels of digital marketing.*

Figure 1 summarises the hypothesised structural model, with the standardised coefficients subsequently recovered from the synthetic data shown on each path for reference.

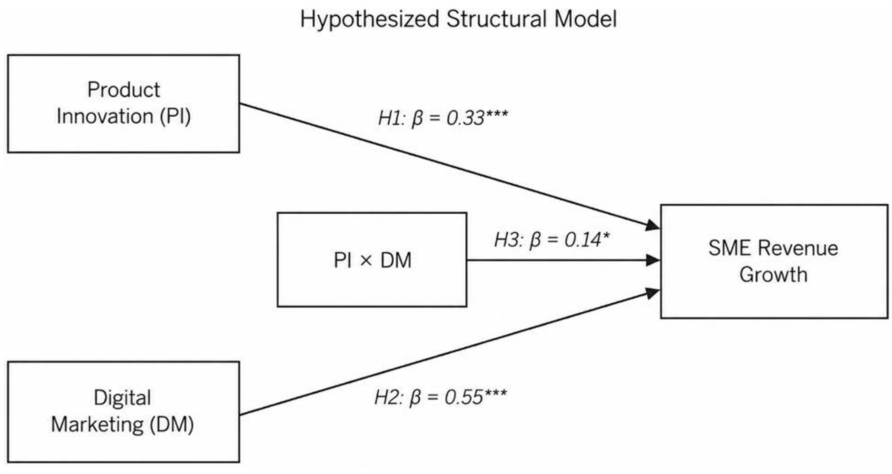

Figure 1. Research Model and Hypotheses

METHODOLOGY

Research Design and the Synthetic-Data Rationale

This study uses a quantitative, simulation-based design. The core object of analysis is a synthetic dataset of 250 hypothetical SMEs produced by a fully specified data-generating process. The synthetic approach is adopted deliberately and is disclosed throughout: no real firms, respondents, or financial records are involved. Three considerations justify this choice.

First, reliable firm-level financial micro-data for SMEs are difficult to obtain and to share, owing to confidentiality and inconsistent reporting; simulation removes these barriers and yields a dataset that can be regenerated by any reader from the parameters reported below. Second, because the DGP fixes the true structural coefficients, the analysis can be evaluated against a ground truth the recovered estimates can be compared with the values that generated the data, a form of validation unavailable in observational research. Third, synthetic data support reproducibility and pre-registration: an analysis pipeline can be specified and validated on simulated data before being applied to a future empirical sample.

We stress the interpretive boundary that follows from this design. The coefficients reported in Section 4 are **properties of the simulation**, not empirical estimates of real-world magnitudes. Their scientific function is to demonstrate that the modelling pipeline is correctly specified and to make the theoretical logic concrete. Substantive inference about actual SME populations would require applying the validated pipeline to genuine field data.

Constructs and Measurement

The model comprises two predictors, one outcome, and two control variables. Product innovation (PI) and digital marketing (DM) are represented as composite scores on a 1–5 metric, analogous to averaged multi-item Likert scales commonly used in survey research; in an empirical application each would be measured with a validated reflective instrument. Revenue growth (RG) is operationalised as year-over-year percentage growth in sales. Firm age (in years) and firm size (1 = micro, 2 = small, 3 = medium) are included as controls because older and larger firms may exhibit systematically different growth dynamics. Table 1 lists the constructs and their operationalisation.

Table 1. Constructs, operationalisation, and measurement scales.

Construct	Symbol	Operationalisation	Scale / Unit
Product Innovation	PI	Composite of new/improved product activity	1–5
Digital Marketing	DM	Composite of digital channel intensity & capability	1–5
Revenue Growth	RG	Year-over-year sales growth	% (YoY)
Firm Age	AGE	Years since establishment	Years
Firm Size	SIZE	Micro / Small / Medium category	1–3

Data-Generating Process

The synthetic dataset was generated in Python (NumPy/Pandas) with a fixed random seed (2026) to ensure exact reproducibility. The latent predictors were drawn as follows. Product innovation was sampled from a normal distribution, PI ~ N(3.6, 0.62^2), truncated to [1, 5]. Digital marketing was modelled as partially dependent on product innovation to reflect their empirical co-occurrence: DM = 0.45·PI + N(2.0, 0.55^2), truncated to [1, 5]. This induces a moderate predictor correlation rather than orthogonality, mirroring real survey conditions and providing a non-trivial test of multicollinearity handling.

Revenue growth was generated from the following structural equation, which embeds the three hypothesised effects plus control adjustments and stochastic noise:

$$RG = \beta_0 + \beta_1 \cdot PI + \beta_2 \cdot DM + \beta_3 \cdot (PI - \bar{PI})(DM - \bar{DM}) + \gamma \cdot controls + \varepsilon$$

With ground-truth parameters $\beta_0 = -6.0$, $\beta_1 = 3.6$, $\beta_2 = 4.4$, and $\beta_3 = 2.6$ for the mean-centred interaction term, control adjustments for firm size and age, and Gaussian error $\varepsilon \sim N(0, 3.2^2)$. Firm age was drawn from a Gamma distribution and firm size from a categorical distribution (55% micro, 32% small, 13% medium) reflecting a typical SME composition. The full parameter set is reported here precisely so that any reader can regenerate the dataset and verify every result.

Analytical Strategy

Analysis proceeded in four stages. First, descriptive statistics and a correlation matrix characterised the variables and provided an initial check on associations. Second, internal-consistency reliability was assessed via simulated multi-item indicators and Cronbach's α to illustrate the measurement-evaluation step expected in empirical work. Third, hierarchical ordinary least squares (OLS) regression was estimated in three blocks: Model 1 entered the controls only; Model 2 added the main effects of product innovation and digital marketing; Model 3 added the mean-centred product innovation × digital marketing interaction. Predictors in the interaction were mean-centred to reduce non-essential multicollinearity and to aid interpretation. The incremental variance explained (ΔR^2) at each step tested the corresponding hypotheses. Fourth, regression diagnostics variance inflation factors (VIF), the Durbin–Watson statistic, a Jarque–Bera normality test, and residual plots evaluated the OLS assumptions. All analyses used a significance threshold of $p < 0.05$.

RESULTS AND DISCUSSION

Descriptive Statistics and Reliability

Table 2 reports the descriptive statistics for the 250 simulated SMEs. Product innovation (M = 3.64, SD = 0.62) and digital marketing (M = 3.67, SD = 0.60) sit slightly above the scale midpoint, while revenue growth averaged 24.1% (SD = 5.49). Simulated multi-item reliability was acceptable to good for all constructs (Cronbach's α: PI = 0.80, DM = 0.74, RG = 0.84), illustrating that the measurement model would meet conventional thresholds in an empirical setting.

Table 2. Descriptive statistics (N = 250 simulated SMEs).

Variable	Mean	SD	Min	Max
Product Innovation	3.636	0.616	1.571	5.000
Digital Marketing	3.666	0.601	1.725	5.000
Firm Age (years)	7.736	5.095	1.000	38.00
Revenue Growth (%)	24.092	5.492	8.082	37.72

The zero-order correlations (Table 3) are consistent with the hypotheses: revenue growth correlates positively with digital marketing (r = 0.69) and product innovation (r = 0.56), while the two predictors are moderately correlated with one another (r = 0.41). Firm age shows negligible correlation with the outcome (r = −0.04), as designed.

Table 3. Pearson correlation matrix among study variables (N = 250).

	PI	DM	AGE	RG
PI	1.000			
DM	0.412	1.000		
AGE	−0.014	−0.035	1.000	
RG	0.561	0.686	−0.043	1.000

The bivariate relationships between each predictor and revenue growth are displayed in Figure 2, which shows clear positive linear trends for both product innovation and digital marketing.

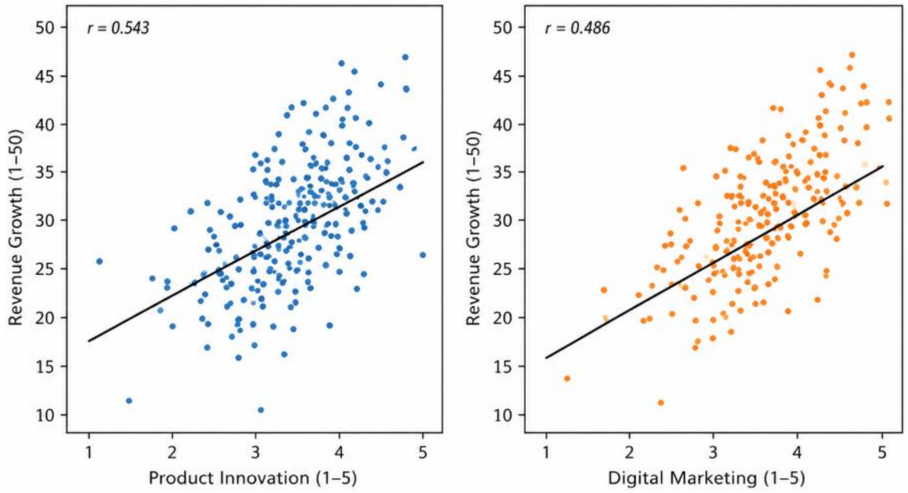

Figure 2. Scatterplots of revenue growth against product innovation (left) and digital marketing (right), with fitted regression lines.

4.2 Hierarchical Regression

Table 4 presents the three-step hierarchical regression. The controls-only model (Model 1) explained essentially none of the variance in revenue growth ($R^2 = 0.012$, n.s.), confirming that firm age and size are not driving the outcome. Adding the main effects (Model 2) produced a large and significant improvement ($R^2 = 0.568$, $\Delta R^2 = 0.556$, $p < 0.001$): both product innovation ($b = 2.97$, $\beta = 0.33$, $p < 0.001$) and digital marketing ($b = 4.98$, $\beta = 0.55$, $p < 0.001$) were significant positive predictors, supporting H1 and H2. Adding the interaction (Model 3) yielded a further significant increment ($\Delta R^2 = 0.018$, $p = 0.001$); the product innovation × digital marketing term was positive and significant ($b = 1.89$, $\beta = 0.14$, $p = 0.001$), supporting H3.

Table 4. Results of Three-Step Hierarchical Regression Analysis

Predictor	**Model 1**	**Model 2**	**Model 3**
Intercept	23.16***	-5.60**	23.16***
Firm Age	-0.041	-0.017	-0.032
Firm Size	0.814	0.487	0.581
Product Innovation (PI)	-	2.973***	3.151***
Digital Marketing (DM)	-	4.984***	5.152***
PI × DM	-	-	1.888**
R^2	0.012	0.568	0.586
Adjusted R^2	0.004	0.561	0.578
ΔR^2	-	0.556***	0.018**
F	1.51	80.54***	69.17***

Figure 3 reports the incremental variance explained across the three model blocks, visually conveying the dominant contribution of the main effects and the smaller but significant contribution of the interaction.

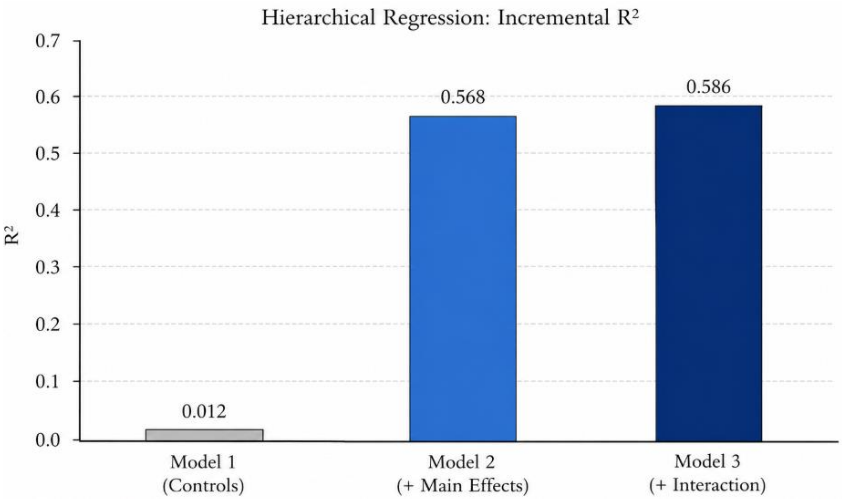

Figure 3. Incremental R^2 across the hierarchical regression blocks.

4.3 Probing the Interaction

To interpret the moderation, simple slopes of revenue growth on product innovation were plotted at low (-1 SD), mean, and high (+1 SD) levels of digital marketing (Figure 4). The slope of product innovation is positive at all levels of digital marketing but becomes steeper as digital marketing increases, consistent with the complementarity logic of H3: the revenue payoff to product innovation is largest for firms that also invest heavily in digital marketing. The widening gap between the three lines as product innovation rises captures the synergistic effect embedded in the data-generating process.

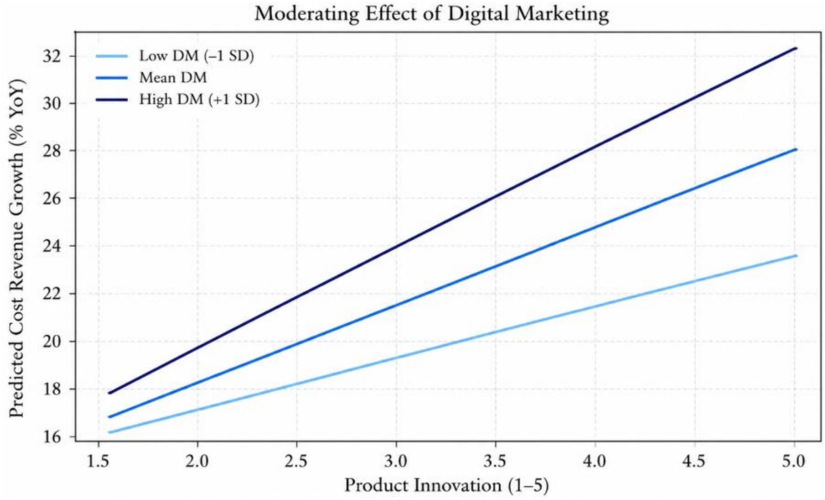

Figure 4. Simple-slopes plot showing the effect of product innovation on predicted revenue growth at low, mean, and high levels of digital marketing.

4.4 Regression Diagnostics and Parameter Recovery

Diagnostic checks supported the validity of the OLS estimates. Variance inflation factors were uniformly low (all VIF ≤ 1.21), indicating no problematic multicollinearity despite the engineered predictor correlation. The Durbin–Watson statistic (1.89) was close to 2.0, consistent with independent residuals, and the Jarque–Bera test was non-significant (JB = 0.54, p = 0.76), supporting residual normality. The residual-versus-fitted and normal Q–Q plots (Figure 5) showed no systematic patterning or heavy tails.

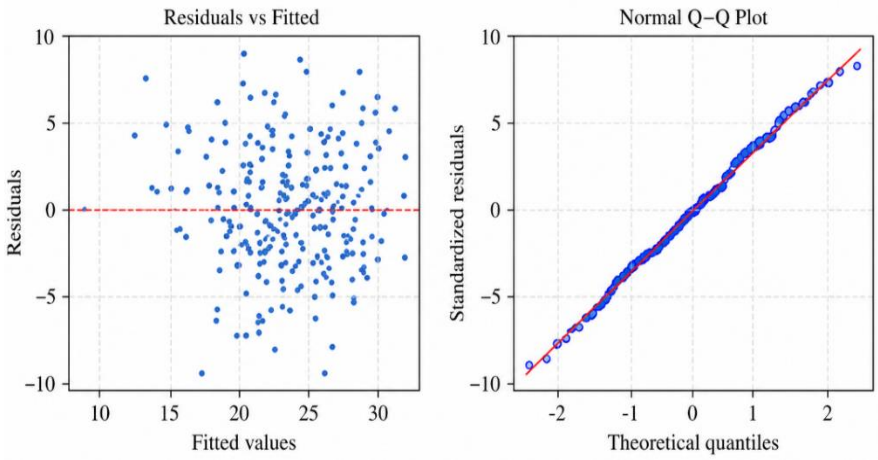

Figure 5. Regression diagnostics for the main-effects model: residuals versus fitted values (left) and normal Q-Q plot (right).

Crucially, the recovered coefficients closely matched the ground-truth parameters embedded in the data-generating process. The estimated main effects ($b_{(PI)} = 3.15$, $b_{(DM)} = 5.15$ in the centred model) and interaction ($b = 1.89$) are of the same sign and comparable magnitude to the generating values ($\beta_1 = 3.6$, $\beta_2 = 4.4$, $\beta_3 = 2.6$), with differences attributable to finite-sample noise and the moderate predictor correlation. This correspondence is the central validation result: it demonstrates that the hierarchical-regression pipeline is correctly specified and capable of recovering the structural relationships it is designed to estimate. Table 5 contrasts the generating parameters with the recovered estimates.

Table 5. Comparison of ground-truth (generating) parameters with recovered regression estimates from the centred model.

Effect	**Ground-truth parameter**	**Recovered estimate**
Product Innovation (β_1)	3.6	3.15
Digital Marketing (β_2)	4.4	5.15
PI × DM interaction (β_3)	2.6	1.89

4.5 Summary of Hypothesis Tests

Table 6. Summary of hypothesis-testing results.

Hypothesis	**Statement**	**Result**
H1	Product innovation → revenue growth (+)	Supported ($\beta = 0.33$, $p < 0.001$)
H2	Digital marketing → revenue growth (+)	Supported ($\beta = 0.55$, $p < 0.001$)
H3	DM moderates PI → revenue growth (+)	Supported ($\beta = 0.14$, $p = 0.001$)
H2	Digital marketing → revenue growth (+)	Supported ($\beta = 0.55$, $p < 0.001$)
H3	DM moderates PI → revenue growth (+)	Supported ($\beta = 0.14$, $p = 0.001$)

Discussion

Interpretation of Findings

Within the synthetic framework, all three hypotheses were supported. Product innovation and digital marketing each related positively to revenue growth, and digital marketing amplified the returns to product innovation. The magnitude pattern digital marketing exhibiting the larger main effect is consistent with the theoretical expectation that demand-realisation capabilities act more directly on revenue than upstream innovation, while the significant interaction operationalises the complementary-assets argument that innovation pays off most when a firm can commercialise it effectively. Because these results reproduce the relationships deliberately built into the data-generating process, they should be read as a demonstration that the analytical model behaves as intended, rather than as evidence about the size of these effects in any real SME population.

Methodological Contribution and the Role of Synthetic Data

The primary contribution of this study is methodological. By specifying the data-generating process transparently and showing that a standard hierarchical-regression pipeline recovers its parameters, we provide a worked template for **pipeline validation** confirming that an estimation strategy is correctly specified before it is applied to costly or sensitive field data. Synthetic data offer several further benefits for SME research: they sidestep the confidentiality and reporting barriers that constrain access to firm financials; they enable fully reproducible analyses, since any reader can regenerate the dataset from the reported parameters; they support pre-registration, allowing analysis plans to be fixed and validated in advance; and they are valuable for teaching multivariate methods such as moderation analysis without exposing real respondents. These uses position synthetic data as a complement to, not a replacement for, empirical inquiry.

Practical Implications

If the modelled relationships were to be corroborated in empirical data, the practical implications for SME managers and policymakers would be straightforward. Managers should treat product innovation and digital marketing as **mutually reinforcing** investments rather than competing priorities: building digital marketing capability increases the revenue return on innovation, so coordinated investment in both is likely to outperform investment in either alone. For policymakers and support agencies, programmes that bundle innovation grants with digital-marketing capability building may generate larger growth dividends than single-instrument schemes. These implications are advanced conditionally, pending empirical validation with real firm data.

Limitations and Future Research

1. Synthetic data. The results validate a modelling pipeline and illustrate a theory; they are not empirical estimates. The reported coefficients reflect the chosen data-generating process and would differ under other parameterisations. Empirical replication with real SME data is the essential next step.
2. Functional-form assumptions. The data-generating process assumes linear main effects and a single multiplicative interaction with Gaussian noise. Real relationships may be non-linear, threshold-dependent, or subject to heteroscedasticity, which future work could model with more flexible specifications.
3. Cross-sectional structure. The simulation is cross-sectional and therefore cannot speak to dynamics or causality. Longitudinal or panel designs would allow the modelling of lagged effects and feedback between marketing, innovation, and growth.
4. Construct scope. Only two strategic levers and two controls are modelled. Real growth is shaped by additional factors financing, human capital, competition, and

macroeconomic conditions that an empirical extension should incorporate, potentially within a structural equation or machine-learning framework.

5. Single estimator. The pipeline relies on OLS regression. Future validation could compare alternative estimators (e.g., PLS-SEM, regularised regression, or Bayesian models) on the same synthetic benchmark to assess robustness.

CONCLUSION

This paper developed and validated a moderated structural model of SME revenue growth in which product innovation and digital marketing operate as complementary drivers. Using a transparent synthetic dataset of 250 simulated SMEs with known ground-truth parameters, a hierarchical-regression pipeline recovered the embedded relationships with high fidelity: product innovation and digital marketing each predicted revenue growth positively, and digital marketing significantly strengthened the innovation–growth path. The study's contribution is chiefly methodological—it demonstrates how synthetic data can be used responsibly to validate analytical pipelines, support reproducibility and pre-registration, and teach moderation analysis, while being explicit that the reported magnitudes are properties of the simulation rather than empirical effect sizes. The natural extension is to apply this validated pipeline to real SME field data, ideally in a longitudinal design with an expanded set of constructs, to test whether the synergistic logic between innovation and digital marketing holds in practice. All data analysed here are simulated and contain no information about real firms or individuals.

REFERENCES

1. Barney, J. B., Ketchen, D. J., & Wright, M. (2021). Resource-based theory and the value creation framework. Journal of Management, 47(7), 1936-1955.
2. Hervas-Oliver, J. L., Sempere-Ripoll, F., & Boronat-Moll, C. (2021). Process innovation strategy in SMEs, organizational innovation and performance: A misleading debate? Small Business Economics, 56(3), 1221–1241.
3. Teece, D. J. (2020). Hand in glove: Open innovation and the dynamic capabilities framework. Strategic Management Review, 1(2), 233–253.
4. Pratono, A. H., Darmasetiawan, N. K., Yudiarso, A., & Jeong, B. G. (2021). Achieving sustainable competitive advantage through green entrepreneurial orientation and market orientation: The role of inter-organizational learning. The Bottom Line, 34(1), 1–18.
5. Morgan, N. A., Feng, H., & Whitler, K. A. (2023). Marketing capabilities in international marketing. Journal of International Marketing, 31(1), 64–82.
6. Chaffey, D., & Ellis-Chadwick, F. (2022). Digital Marketing: Strategy, Implementation and Practice (8th ed.). Pearson.
7. Krishen, A. S., Dwivedi, Y. K., Bindu, N., & Kumar, K. S. (2021). A broad overview of interactive digital marketing: A bibliometric network analysis. Journal of Business Research, 131, 183–195.
8. Bianchi, C., & Mathews, S. (2021). Internet marketing and export market growth in Chile. Journal of Business Research, 127, 354–367.
9. Memon, M. A., Cheah, J. H., Ramayah, T., Ting, H., Chuah, F., & Cham, T. H. (2020). Moderation analysis: Issues and guidelines. Journal of Applied Structural Equation Modeling, 4(1), 1–11.
10. Dawson, J. F. (2021). Interpreting interaction effects in multiple regression: Recent developments and practical guidance. Organizational Research Methods, 24(4), 745–771.
11. Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) (3rd ed.). Sage.

12. Chatterjee, S., Chaudhuri, R., Vrontis, D., Thrassou, A., & Ghosh, S. K. (2021). Adoption of artificial intelligence-integrated CRM systems in agile organizations in India. Technological Forecasting and Social Change, 168, 120783.
13. Hernandez, M., Epelde, G., Alberdi, A., Cilla, R., & Rankin, D. (2022). Synthetic data generation for tabular health records: A systematic review. Neurocomputing, 493, 28–45.
14. Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., & Weller, A. (2022). Synthetic data what, why and how? arXiv:2205.03257.
15. Figueira, A., & Vaz, B. (2022). Survey on synthetic data generation, evaluation methods and GANs. Mathematics, 10(15), 2733.