

## **AI-Driven Micro-Insurance for Boutique Start-Ups: Is the Gain in the Algorithm or in the Data?**

Author: Aya Maria Maalouf<sup>1</sup>

*Affiliation:*

1. *Business School, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*
2. *CIRAME Research Center, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*

### **Abstract:**

Boutique start-ups (micro-enterprises employing fewer than ten people) rank among the least insured commercial risks in the markets, thus artificial intelligence is presented as the solution. The following study poses a narrower question: when an AI-enabled scheme achieves better results than a conventional one, which element of "AI" is really responsible for the performance. A portfolio of 8,000 boutique start-ups is simulated using a compound Poisson loss model with lognormal severity along with an unobservable heterogeneity term that reflects the firm's private information. Five rating and distribution regimes are evaluated on identical data: community rating, traditional tier rating, a generalized linear model, a gradient boosting model trained on digitally observable data, and the same model combined with parametric settlement. Demand follows constant absolute risk aversion, while prices and purchase decisions converge toward a fixed point, making participation and adverse selection outcomes instead of assumptions. Three findings emerge. Community rating collapses: participation drops from 43.8% to 18.7%, and the remaining pool carries twice the risk of the population. Digital risk-based rating roughly doubles coverage, lowers the average premium from \$599 to \$310 and reduces the expense ratio by half, from 23.4% to 11.2%. Most importantly, a decomposition reveals that essentially all of the improvement in risk classification stems from the breadth of digitally observable data rather than from the learning algorithm: the data-access effect is +0.118 Gini while the algorithm effect falls between -0.015 and +0.007. Machine learning is best regarded as a replacement for the actuarial capacity that micro-insurers cannot afford, not as a superior statistical technology.

### **Keywords:**

Micro-insurance; adverse selection; risk classification; parametric insurance; machine learning.

## **Introduction:**

The boutique start-up is a familiar economic form: a two-person design studio, a small online retailer, a neighborhood salon, a single café. These firms are legally companies but economically households. They hold stock and equipment worth a few thousand dollars, keep no meaningful reserve funds, and a single fire, theft or extended power failure can put them out of business. In most markets they are almost completely uninsured.

Two forces maintain this protection gap. The first is informational: an owner knows far more about his staff, his wiring and his generator than an underwriter seeing what's on the paperwork, so under a single pooled premium the firms that know themselves to be safe exit first, the average risk of the pool increases, as well as the premium (the price spirals upward while the pool shrinks and gets riskier and riskier); the unravelling mechanism of Akerlof (1970) and Rothschild and Stiglitz (1976). The second is arithmetic: it has nothing to do with information it's cost accounting. Writing a policy and processing a claim costs roughly the same amount of staff time and paperwork whether the policy is worth \$500 or \$500,000. For a boutique start-up, the actual expected loss (the "pure premium") might only be a couple hundred dollars a year. But the fixed cost of administering that policy doesn't shrink to match so it can eat up 25% of the premium and this is a much bigger drag than it would be on a large commercial policy, where the same fixed cost is a rounding error. Biener & Eling (2012) argue this cost structure. And these two problems are the reason micro-insurance is hard to sell profitably.

Insurance technology claims AI solves both problems at once: it can spot hidden risks that a paper form misses, and it can lower costs. But "AI" is actually three separate things: the algorithm, the data it uses, and the cheaper a digital sales channel. No one has tested these three factors separately for small businesses, mainly because insurance companies never run two different systems side by side to compare them fairly and a business's true risk level is never fully known in real life.

This paper solves that problem by building a simulated market instead of a real one. When the market is simulated, the true risk level of each firm is known from the start, so each of the three factors can be tested on its own. The paper asks four questions: Does AI actually sort risky and safe firms better than the traditional method? If so, is that due to the algorithm, the data, or simply using the model correctly? What happens to insurance take-up and adverse selection once firms are free to choose whether to buy? And does the system still make financial sense once the risk of automatic, unchecked payouts is taken into account?

## **Literature Review:**

Information asymmetry forms the core theoretical starting point for this study. Akerlof (1970) showed that when product quality stays hidden, market prices adjust to average quality levels, pushing high-quality sellers out and eventually causing the market to collapse. Rothschild and Stiglitz (1976) applied this concept to insurance markets, demonstrating that a single pooled

contract cannot hold up because low-risk individuals will always favor partial coverage at a lower price. Arrow (1963) outlined the economic fallout that makes this pattern so costly: when risk-averse individuals are left with uninsurable losses, society suffers a real loss of efficiency rather than a simple shift of funds.

Whether selection problems actually surface in practice remains a subject of ongoing debate. Chiappori and Salanié (2000) found no conditional link between policy coverage and risk choices in French auto insurance data. However, Cohen and Siegelman (2010) reviewed the broader empirical evidence and concluded that selection issues routinely pop up whenever regulations or context stop insurers from using informative risk characteristics. That conditional finding is the central pivot for this study, given that micro-insurance is a setting where insurers cannot afford detailed risk screening. Highlighting this exact vulnerability, Yao et al. (2017) documented widespread adverse selection in a Pakistani micro health program because an essential risk characteristic was left unpriced.

Academic research treats micro-insurance as an economically distinct product rather than standard insurance scaled down. Biener and Eling (2012) identify the cost of evaluating risk as the main barrier here, while Biener (2013) shows that administrative overhead forces price loadings to be proportionally much heavier than in conventional insurance markets. Bringing together evidence on consumer demand, Eling et al. (2014) conclude that price tags, institutional trust, and liquidity are the primary drivers of adoption. This matches the field experiments of Cole et al. (2013), who showed that demand for rainfall insurance in rural India swings heavily with price shifts and drops under liquidity constraints. When loss adjustment costs threaten to consume small payouts, parametric policies pay out based on an observable index trigger instead (Barnett & Mahul, 2007). That trade-off brings speed and lower operating costs, but introduces basis risk, which is the gap between index payouts and actual losses. Biener et al. (2019) verified experimentally that non-performance of a contract drastically reduces willingness to pay, meaning basis risk operates as a real demand constraint rather than a minor technical footnote.

On the actuarial side, the modeling framework is well established. Aggregate losses are standardly modeled using compound distributions that combine event counts and severities, with Poisson-lognormal specifications serving as the benchmark for commercial property and business interruption risks (Klugman et al., 2019). The main open question is whether machine learning provides a genuine improvement over generalized linear models (GLMs). Blier-Wong et al. (2021) surveyed this research area and found that performance gains on standard tabular features tend to be modest, pointing to non-traditional data formats as the more promising direction. Wüthrich and Merz (2023) state the same point mathematically: a boosting model and a properly specified GLM estimate the exact same conditional expectation, so the algorithm only adds value when the analyst gets the GLM's functional form wrong. Studies focused on industry trends take a broader perspective: Eling et al. (2022) contend that AI will move insurance from paying out claims after the fact toward predicting risk in advance while narrowing information

gaps, though Eling et al. (2024) note that scoring risk in very fine detail weakens the cross-subsidies that traditional pooling offers to high-risk policyholders. More broadly, digital transformation typically reshapes both internal cost structures and external stakeholder relationships, which are the two exact channels through which an AI-enabled micro-insurance scheme operates.

Beyond pure actuarial modeling, the successful deployment of AI-driven micro-insurance must be contextualized within the broader technological and financial realities faced by boutique start-ups. In vulnerable economic environments, SMEs often lack access to traditional banking services, forcing them to rely heavily on alternative entrepreneurial intentions, business experience, and equity financing to survive (Chaanine et al., 2023; Mawad & Freiha, 2024; Rizk & Sassine, 2023). When conventional financial safety nets fail, accurately assessing the risk of financial distress requires predictive vulnerability models like the Altman Z-score (Elia et al., 2021) alongside high technology analytic adeptness to interpret shifting macroeconomic variables (Bejjani et al., 2024). Consequently, micro-insurers must adopt smart decision support systems to navigate these disruptions and offer viable products to highly exposed small enterprises (Subrahmanyam & Beainy, 2027).

Furthermore, transitioning to digital risk rating and automated parametric settlements involves profound organizational and ethical shifts. Leveraging vast digital data pools requires insurers to cultivate a robust data-driven culture and advanced knowledge management capabilities (Azoury et al., 2024; Subrahmanyam et al., 2024). However, delegating underwriting and claims to algorithms introduces complex ethical challenges regarding transparency and bias in business culture (Diab & El Hajj, 2024). As seen in the broader auditing and financial sectors, the actual adoption of AI is driven heavily by the perceived trust and behavioral profiles of both the institutions and the end-users (Jabbour Al Maalouf et al., 2026). Ultimately, establishing institutional trust—a key driver of micro-insurance demand—depends on maintaining stringent internal controls (Saber & Sassine, 2022) and ensuring that corporate and country-level governance structures can support sustainable digital financial services during severe economic shocks (Athari et al., 2023; El-Chaarani et al., 2023).

Two main research gaps follow from this literature. First, claims that artificial intelligence reduces information asymmetry center on a combined bundle of algorithm and data, yet no study separates the distinct impact of the algorithm from the data itself within a micro-enterprise portfolio. Second, existing literature mostly examines pricing models and consumer demand separately, even though answering the welfare questions raised by Arrow (1963) alongside the unraveling dynamics of Rothschild and Stiglitz (1976) requires solving both together. This paper tackles both gaps directly.

## **Methodology:**

This study relies on a Monte Carlo simulation applied to a portfolio that is synthetic yet calibrated to match structural features. The research question determines the design: to examine whether a rating technology can recover true risk; likewise, to examine whether a market deteriorates, one must observe the counterfactual scenario in which the deterioration does not occur. The paper comes with all parameters, the random seed and the estimation code, ensuring that each figure presented below can be reproduced exactly.

A population of  $N = 8,000$  boutique start-ups is created, each characterized by nine observable variables: monthly revenue, firm, employees, sector, age, stock at risk, online exposure, generator dependence digital payment share and a location risk index. The annual claim frequency of firm  $i$  is

$$\lambda_i = \exp(\beta_0 + f(X_i) + u_i - \frac{1}{2}\sigma_u^2), u_i \sim N(0, \sigma_u^2),$$

Put the expected number of claims per year for firm  $i$ ,  $\lambda_i$ , begins with a baseline rate  $\beta_0$  and is adjusted upward or downward by  $f(X_i)$ , a function of the firm's observable characteristics, including generator dependence and location risk. The additional term  $u_i$  represents the private information of the firm (what the owner knows about her own risk that none of the variables in  $X_i$  can capture). Since  $u_i$  is random and hidden from the insurer, its inclusion is precisely what permits adverse selection to emerge naturally within the simulation, rather than being imposed by assumption. The  $\exp()$  and the  $-\frac{1}{2}\sigma_u^2$  term form a conventional actuarial adjustment that leaves the population-average frequency unchanged once  $u_i$  is introduced, so adding private information alone does not alter the market's average risk. The severity of each claim when it occurs is lognormal, and the annual loss is modeled using the collective risk model (Klugman et al., 2019), after a \$50 deductible and subject to a \$2,500 annual cap. For each firm, two separate years are simulated: a historical year used to estimate the models, and a forward year used to establish what is actually paid, so that every reported figure on accuracy and profitability is truly out-of-sample.

Five regimes are estimated on 4,800 firms and then evaluated on 3,200 firms that were held out. M0 employs community rating; M1 replicates the standard sector-by-size tier rating; M2 is a Poisson-Gamma GLM built on the digital feature set; M3 is a gradient boosting model using Poisson and Gamma deviance; and M4 combines the M3 tariff with parametric settlement. Every tariff is rebalanced to the same rate level, ensuring that differences arise from discrimination among risks rather than from average price. Channel costs vary because the channels themselves differ: a broker channel involving physical inspection incurs \$28 per policy and \$45 per claim, a digital channel \$6 and \$18, and a parametric channel \$6 and \$7. The gross premium is

$$P_i = [(1 + \theta) \cdot \pi_i + c_{\text{fix}} + c_{\text{lae}} \cdot \hat{\lambda}_i] / (1 - c_{\text{var}}), \theta = 0.10.$$

Put the gross premium  $P_i$  quoted to firm  $i$  begins with the pure premium  $\pi_i$ , which is the model's estimate of the firm's expected loss, increased by a risk margin  $\theta$  of 10% to protect the insurer. Then a fixed cost  $c_{\text{fix}}$  (acquisition and administration, identical for every policy within a regime) and a per-claim handling cost  $c_{\text{lae}}$  multiplied by the firm's estimated claim frequency  $\hat{\lambda}_i$  are added,

and the resulting total is grossed up by dividing by  $(1 - c\_var)$  to account for commission and premium tax,  $c\_var$ . The values of  $c\_fix$ ,  $c\_lae$  and  $c\_var$  vary from one regime to another, since the distribution channel is not the same selling through a manual broker channel costs much more per policy than doing so through an automated digital one and Table 2 in the results of this section lists the actual dollar amounts employed in each case.

Each firm decides whether to buy by comparing the quoted price to what the coverage is worth to it. That worth is called the certainty-equivalent premium,  $\pi_i^{ce}$  (the most the firm would pay to make its risk disappear entirely)

$$\pi_i^{ce} = (1/a_i) \cdot \ln E[\exp(a_i S_i)]$$

Here  $a_i$  captures how much the firm dislikes uncertainty (higher for a more cautious owner), and  $S_i$  is the firm's own random annual loss, which includes the private information  $u_i$  that only the owner knows. A firm buys whenever the price is no higher than this value,  $P_i \leq \pi_i^{ce}$ . Prices and purchases are then repeated round by round: each round the insurer recalculates its price using only the firms that stayed, the same way real insurers set rates. Under community rating, where every firm pays the same price, this process produces the classic adverse-selection spiral safe firms leave, the price rises, and more firms leave. Solvency capital is the extra buffer an insurer holds to survive a severe, 1-in-200-year loss year, and it costs 8% annually to hold. Finally, how well a tariff tells good risks from bad ones is measured by the Lorenz-Gini of realized losses ranked against predicted premium.

## **Results:**

The simulated portfolio holds together internally: the realized claim frequency of 0.4219 lines up with the mean latent frequency of 0.4217, 30.8 per cent of firms file at least one claim, the average annual loss cost comes to \$158.37, and average willingness to pay stands at \$352.60, which is 2.07 times the true pure premium.

Table 1 presents the out-of-sample discrimination results. By construction, community rating cannot order risks in any way. Traditional tier rating attains a Gini of 0.205: sector and headcount do carry information, though its value is limited. On digital data, both models show marked improvement, with the GLM arriving at 0.425 and the boosting model at 0.352.

**Table 1. Out-of-sample risk classification (3,200 hold-out firms)**

| Rating regime              | Gini    | Spearman vs. true | RMSE vs. true (\$) |
|----------------------------|---------|-------------------|--------------------|
| M0 Community rating        | -0.0085 | 0.000             | 208.11             |
| M1 Traditional tier rating | 0.2052  | 0.358             | 196.26             |
| M2 GLM, digital data       | 0.4253  | 0.841             | 115.77             |

| Rating regime                      | Gini   | Spearman vs. true | RMSE vs. true (\$) |
|------------------------------------|--------|-------------------|--------------------|
| M3 Gradient boosting, digital data | 0.3523 | 0.733             | 166.18             |

*Note. Spearman and RMSE are computed against the true pure premium, observable only in simulation. All tariffs are rebalanced to a common mean.*

Market equilibrium is presented in Table 2. As Akerlof (1970) and Rothschild and Stiglitz (1976) set out, the community-rated market falls apart: 43.8 per cent of firms purchase at the opening quote, and because the premium is re-levelled on a pool that keeps deteriorating, participation drops through 29.0, 23.6 and 21.9 per cent before settling at 18.7 per cent. What remains is a residue of survivors — the insured pool carries 2.03 times the risk of the population, and the premium needed to sustain it stands at \$598.73, compared with a population pure premium of \$167.36. Digital rating based on risk approximately doubles participation, brings the selection index down to roughly 1.3 and reduces the premium by half. Two mechanisms work in tandem: precise relativities prevent low risks from being overcharged, while the digital channel's lower fixed costs push down the premium floor.

**Table 2. Market equilibrium and insurer technical accounts**

| Regime               | Particip. | Avg. prem. | Loss r. | Exp. r. | Comb. r. | Selection idx. |
|----------------------|-----------|------------|---------|---------|----------|----------------|
| M0 Community         | 18.7%     | \$598.73   | 73.6%   | 23.4%   | 97.0%    | 2.032          |
| M1 Traditional tier  | 20.2%     | \$563.23   | 73.2%   | 23.5%   | 96.8%    | 1.838          |
| M2 GLM               | 37.7%     | \$339.99   | 75.0%   | 19.9%   | 94.8%    | 1.316          |
| M3 Gradient boosting | 34.7%     | \$310.12   | 80.6%   | 13.2%   | 93.8%    | 1.292          |
| M4 GBM + parametric  | 35.1%     | \$312.05   | 81.4%   | 11.2%   | 92.6%    | 1.297          |

*Note. The selection index is the mean latent frequency of the insured pool divided by that of the population; 1.00 denotes no selection.*

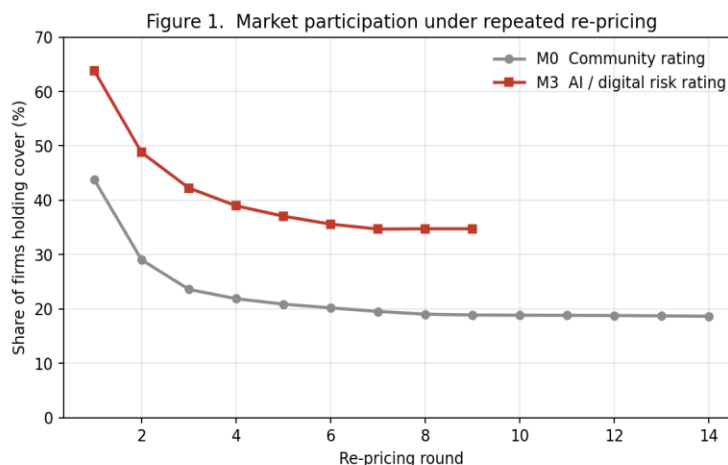


*Figure 1. Share of firms holding cover across successive re-pricing rounds. Community rating loses more than half its initial market; digital risk rating stabilizes after three rounds.*

The combined ratio declines steadily, going from 97.0 per cent down to 92.6 per cent, and the way it is composed is revealing. The loss ratio climbs from 73.6 to 81.4 per cent, and this is exactly the signal one should expect: a book that is properly rated gives back a greater portion of premium to policyholders in the form of claims instead of consuming it through acquisition and adjustment costs. All of the gain stems from the expense ratio, which is cut in half. Across every regime, solvency capital holds steady at 9 to 11 per cent of premium, which mirrors robust diversification across numerous small, capped, and largely independent exposures; consequently, risk classification does not meaningfully lower capital requirements, and its worth resides in participation and expense. Once capital is charged at 8 per cent, the economic result climbs from \$7,791 to \$23,137 — nearly a threefold gain. Total certainty-equivalent surplus grows from \$239,070 to \$302,657, even as surplus per policy drops from \$400.45 to roughly \$270 — finer classification opens the market to many moderate risks while pulling a substantial subsidy away from a handful of high risks.

The decomposition is presented in Table 3, and it constitutes the paper's key finding. If the algorithm is held constant while the feature set is broadened from a proposal form to the digital record, the Gini increases by 0.118. If the data are held constant and an expert GLM is swapped for gradient boosting, the change is  $-0.015$ ; measured against a naive GLM, the algorithm effect comes to  $+0.007$ . Proper actuarial specification contributes  $+0.009$ . Thus, the effect of data access is ten times greater than either alternative, and it carries the opposite sign from the algorithm effect when that effect is gauged against a competent benchmark.

**Table 3. Decomposition of the improvement in risk classification (Lorenz-Gini)**

| Configuration                | Gini   | Effect             | Comparison | Change  |
|------------------------------|--------|--------------------|------------|---------|
| C1 Naive GLM, proposal form  | 0.3024 | Specification      | C2 – C1    | +0.0085 |
| C2 Expert GLM, proposal form | 0.3109 | Algorithm (expert) | C6 – C5    | –0.0145 |
| C3 Boosting, proposal form   | 0.2995 | Algorithm (naive)  | C6 – C4    | +0.0069 |
| C4 Naive GLM, digital data   | 0.4109 | Data access        | C6 – C3    | +0.1183 |
| C5 Expert GLM, digital data  | 0.4323 | Total              | C6 – C1    | +0.1154 |
| C6 Boosting, digital data    | 0.4178 |                    |            |         |

*Note. Four replications, 12,000 training and 12,000 test firms each. The proposal-form set holds revenue, employees, firm age and sector; the digital set adds payment share, stock at risk, location risk, generator dependence, online exposure and prior claims.*

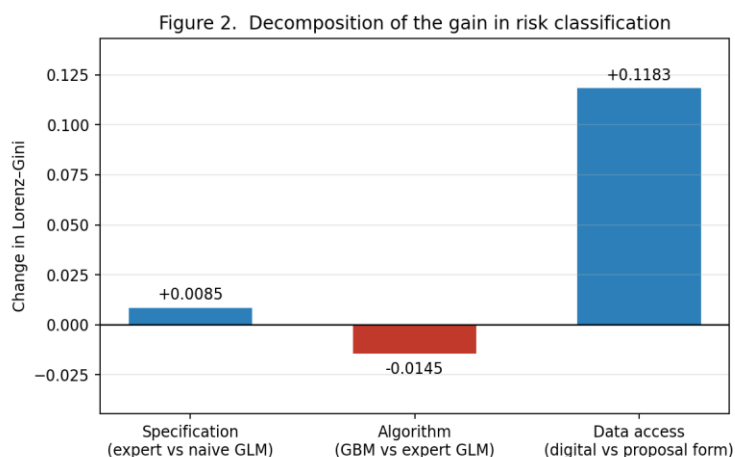


*Figure 2. The data-access effect is an order of magnitude larger than the specification effect and of the opposite sign to the algorithm effect.*

An experiment tracing the learning curve verifies that the algorithmic deficit stems from how much data is available, and it never flips direction. As the training portfolio expands, the boosting model steadily narrows the gap from a Gini of  $-0.038$  with 1,000 policies, to  $-0.009$  with 10,000, and  $-0.001$  with 100,000 but at no tested size does it surpass a properly specified GLM. The nonlinearities present in the data-generating process are genuine, yet they are mild enough for an actuary to capture simply by specifying the correct functional form.

The design is subject to two additional caveats. Parametric settlement delivers the lowest combined ratio, yet when the correlation between index and actual loss is  $\rho = 0.80$ —a level typical of a well-designed index hedge effectiveness reaches only 0.765 and 29.3 per cent of aggregate loss stays uncovered; the residual becomes modest, at 19.8 per cent, only once  $\rho \geq 0.90$ . Sensitivity analysis

reveals that increasing the fixed cost per policy from \$6 to \$34 lowers participation from 34.8 to 29.7 per cent, while the combined ratio moves only from 93.8 to 94.3 per cent—meaning distribution cost affects the market's size rather than the insurer's margin. Raising private information from  $\sigma_u = 0.20$  to 0.90 reduces participation from 23.1 to 13.4 per cent under community rating, compared with 43.8 to 23.7 per cent under digital rating: the benefit of risk-based rating is greatest exactly where private information is most acute.

### **Discussion and Conclusion:**

The numbers leave no room for doubt about commercial viability. A micro-insurance policy for boutique start-ups based on community rating simply cannot work as a business. Only 18.7 percent of eligible businesses opt in, leaving a risk pool that is twice as risky as the broader population. Keeping that pool afloat would require charging a premium more than three times the population pure premium. Moving to risk-based pricing, powered by observable digital data and delivered automatically, flips these results completely. Participation nearly doubles, premiums drop by half, expense ratios fall from 23.4 to 11.2 percent, and the combined ratio settles at 92.6 percent. The model clearly works.

Looking under the hood reveals something even more striking. Breaking down the drivers shows that broader data access alone explains +0.118 of the total +0.115 Gini improvement. The algorithm itself adds somewhere between  $-0.015$  and  $+0.007$ , depending on the baseline you measure it against. What the market calls "AI underwriting" is mostly just a data-gathering tool wearing statistical clothing. The algorithm's real perk is that it saves an insurer from needing a top-tier actuary to tune the model by hand, which is a game-changer for a small firm that cannot afford that talent, but almost meaningless for a major carrier that already has it. This is a story about substituting labor rather than breaking new analytical ground, which fits the empirical findings of Blier-Wong et al. (2021) and Wüthrich and Merz (2023) far better than the broad predictions in Eling et al. (2022).

Three clear practical takeaways follow from these findings. First, insurers entering this space should invest in payment systems and point-of-sale integrations before putting money into complex modeling, since infrastructure yields an order of magnitude higher return. Whether a product can scale up comes down to distribution and payout costs, not how smart the underwriting engine is. Second, parametric payouts should only be used when index correlation stays above roughly 0.90, because below that threshold, uncovered losses remain large enough to destroy the trust that voluntary micro-insurance relies on (Biener et al., 2019). Third, on the theoretical side, the study puts a concrete number on the welfare problem Arrow (1963) described: replacing pooled pricing with risk classification boosts total economic surplus by 27 percent, but cuts surplus per policy by a third. Efficiency and solidarity pull in opposite directions here, meaning regulators who restrict digital rating variables are actively choosing one side over the other, not avoiding the choice (Eling et al., 2024).

The study does have real limitations. Because the portfolio is built on simulated data, the findings depend on those underlying rules. Specifically, the conclusion that boosting offers no edge over a GLM assumes the true event frequency is only moderately non-linear. Losses are also assumed to be independent across firms, ignoring systemic shocks like grid failures, currency crashes, or civil unrest that drive tail risk in places like Lebanon and would require capital reserves far beyond the 9 to 11 percent reported here. On top of that, moral hazard is ignored, demand is treated as static and fully rational despite field studies showing liquidity and trust dictate actual take-up (Cole et al., 2013), competition is left out, and cost inputs rely on industry estimates rather than audited books.

These gaps point straight to the next steps for research: testing these models on actual micro-enterprise portfolios through industry data sharing, adding common shock factors to see if capital efficiency holds up during crises, modeling competitive dynamics to test for cream-skimming, and examining raw transaction data where an algorithmic edge might actually show up. Until then, the practical takeaway remains straightforward: the real path to insuring boutique start-ups runs through payment rails, not complex models.

## **References**

- Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3), 488–500.
- Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American Economic Review*, 53(5), 941–973.
- Athari, S. A., Saliba, C., Khalife, D., & Salameh-Ayanian, M. (2023). The role of country governance in achieving the banking sector’s sustainability in vulnerable environments: New insight from emerging economies. *Sustainability*, 15(13), 10538.
- Azoury, N., Subrahmanyam, S., & Sarkis, N. (2024). The Influence of a Data-Driven Culture on Product Development and Organizational Success through the Use of Business Analytics. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, 15(2), 123-134.
- Barnett, B. J., & Mahul, O. (2007). Weather index insurance for agriculture and rural areas in lower-income countries. *American Journal of Agricultural Economics*, 89(5), 1241–1247.
- Bejjani, G., Toros, E., Elia, J., & Jabbour Al Maalouf, N. (2024). Technology analytic adeptness for examining the interrelatedness between the main macroeconomic variables in Lebanon employing a vector autoregression (VAR). *Journal of Decision Systems*, 1-30.
- Biener, C. (2013). Pricing in microinsurance markets. *World Development*, 41, 132–144.
- Biener, C., & Eling, M. (2012). Insurability in microinsurance markets: An analysis of problems and potential solutions. *The Geneva Papers on Risk and Insurance – Issues and Practice*, 37(1), 77–107.

- Biener, C., Landmann, A., & Santana, M. I. (2019). Contract nonperformance risk and uncertainty in insurance markets. *Journal of Public Economics*, 175, 65–83.
- Blier-Wong, C., Cossette, H., Lamontagne, L., & Marceau, E. (2021). Machine learning in P&C insurance: A review for pricing and reserving. *Risks*, 9(1), 4.
- Chaanine, N., Mhanna, R. Y., Verdie, J. F., & Beainy, R. (2023). Leveraging business experience and knowledge to boost entrepreneurial achievement. *International Journal of Work Innovation*, 4(2), 104-115.
- Chiappori, P.-A., & Salanié, B. (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy*, 108(1), 56–78.
- Cohen, A., & Siegelman, P. (2010). Testing for adverse selection in insurance markets. *Journal of Risk and Insurance*, 77(1), 39–84.
- Cole, S., Giné, X., Tobacman, J., Topalova, P., Townsend, R., & Vickery, J. (2013). Barriers to household risk management: Evidence from India. *American Economic Journal: Applied Economics*, 5(1), 104–135.
- Diab, B., & El Hajj, M. (2024). Ethics in the age of artificial intelligence: unveiling challenges and opportunities in business culture. *Cogent Business & Management*, 11(1). <https://doi.org/10.1080/23311975.2024.2408440>
- El-Chaarani, H., Abraham, R., Khalife, D., & Salameh-Ayanian, M. (2023). Corporate governance effects on bank profits in Gulf Cooperation Council countries during the pandemic. *International Journal of Financial Studies*, 11(1), 36.
- Elia, J., Toros, E., Sawaya, C., & Balouza, M. (2021). Using Altman Z”-score to predict financial distress: Evidence from Lebanese alpha banks. *Management Studies and Economic Systems*, 6(1/2), 47-57.
- Eling, M., Gemmo, I., Guxha, D., & Schmeiser, H. (2024). Big data, risk classification, and privacy in insurance markets. *The Geneva Risk and Insurance Review*, 49(1), 75–126.
- Eling, M., Nuessle, D., & Staubli, J. (2022). The impact of artificial intelligence along the insurance value chain and on the insurability of risks. *The Geneva Papers on Risk and Insurance – Issues and Practice*, 47(2), 205–241.
- Eling, M., Pradhan, S., & Schmit, J. T. (2014). The determinants of microinsurance demand. *The Geneva Papers on Risk and Insurance – Issues and Practice*, 39(2), 224–263.
- Jabbour Al Maalouf, N., Ficani, A., & Elia, J. (2026). Drivers of AI adoption in auditing: a quantitative exploration of perceived benefits and trust across auditor profiles. *Cogent Business & Management*, 13(1). <https://doi.org/10.1080/23311975.2026.2709916>
- Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2019). *Loss models: From data to decisions* (5th ed.). Wiley.

- Mawad, J. L., & Freiha, S. (2024). Entrepreneurial intentions in the absence of banking services: The case of the Lebanese in crises. *Journal of Risk and Financial Management*, 17(7), 264.
- Replication. All results are produced by `pythonscripts.py`, supplied with this paper and depending only on `numpy`, `pandas`, `scipy` and `scikit-learn`. The random seed is fixed at 20260930; re-running the scripts reproduces every figure in Tables 1 to 3 exactly.
- Rizk, R., & Sassine, M. (2023). Equity financing and SME growth: evidence from the Eurozone. *International Journal of Work Innovation*, 4(3), 227-239.
- Rothschild, M., & Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4), 629–649.
- Saber, A. M., & Sassine, M. E. (2022). Reflections on internal control and internal audit roles for ensuring financial transparency toward Lebanese NGOs financial donors. *EuroMed Journal of Management*, 4(3), 186-198.
- Subrahmanyam, S., & Beainy, R. H. (2027). Smart Decision Support for Uncertainty and Disruptions in Industry 4.0 Operations. In *Smart Operational Research Solutions in the Era of Industry 4.0 and Sustainability* (pp. 167-198). IGI Global Scientific Publishing.
- Subrahmanyam, S., Aishwaryalaxmi, N. S., Khalife, D., Shaikh, I. A. K., Faldu, R., & Asthana, N. (2024, April). Impact of knowledge management and big data analytics capabilities on firm performance. In *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)* (pp. 1-5). IEEE.
- Wüthrich, M. V., & Merz, M. (2023). *Statistical foundations of actuarial learning and its applications*. Springer.
- Yao, Y., Schmit, J. T., & Sydnor, J. R. (2017). The role of pregnancy in micro health insurance: Evidence of adverse selection from Pakistan. *Journal of Risk and Insurance*, 84(4), 1073–1102.