Proceedings of ICFTBA 2025 Symposium: Financial Framework's Role in Economics and Management of Human-Centered Development
DOI: 10.54254/2754-1169/2025.GL27699

Research on Dynamic Credit Profit Forecasting Model Based on Customer Behavioural Uncertainty

Yuanjing Jiang

University of Glasgow, Glasgow, United Kingdom
3085581J@student.gla.ac.uk

Abstract. Conventional risk models rely heavily on static indicators, such as external credit scores (EXT_SOURCE), and thus fail to capture complex dynamics hidden in customer behaviour. To address this limitation, this study introduces a novel uncertainty indicator—entropy—to enhance traditional credit risk assessment models in forecasting customer profit potential. This paper applies a Dirichlet Process Mixture Model (DPMM) to cluster customer behaviour and to calculate entropy values based on posterior probabilities, thereby quantifying behavioural uncertainty. We then employ a Generalized Additive Model (GAM) to evaluate the nonlinear effect of entropy on profit rates, controlling for conventional variables such as loan amount, income, and external credit scores. Empirical analysis confirms that entropy has a statistically significant and nonlinear impact on profit rates, challenging the linear assumptions of traditional models. More importantly, our dynamic decision framework reveals that entropy-based models can identify customer groups with both high uncertainty and high profit potential, which are frequently overlooked by conventional credit risk models. These findings provide theoretical support for reframing risk management into a broader strategy of profit maximization and customer value optimization. Despite the study’s reliance on a simplified proxy profit function, the results are robust and provide a valuable tool for financial institutions seeking to refine their risk–return trade-offs.

Keywords: Credit risk, Profit forecasting, Customer uncertainty, Entropy, Generalized Additive Model (GAM)

1. Introduction

Accurate credit risk assessment remains fundamental to banking operations. Traditional models, including logistic regression and decision trees, depend on static features such as income, debt levels, historical defaults, and external credit scores [1]. While effective in filtering risky borrowers, these models fail to capture high-dimensional and dynamic customer behaviour. This creates a business dilemma: banks may inadvertently reject customers with uncertain behaviour who actually have high profit potential.

This study argues that behavioural uncertainty does not simply equate to risk, as high uncertainty often reflects data complexity or noise rather than inherently poor creditworthiness—yet conventional models remain unable to distinguish among such ambiguous groups. We therefore propose a new methodological framework that quantifies customer behavioural uncertainty using entropy derived from a DPMM and integrates this measure into a profit forecasting model.

The main contribution of this study lies in the first proposal and application of an entropy metric based on the Dirichlet Process Mixture Model (DPMM) to quantify customer behaviour uncertainty, introducing

DOI: 10.54254/2754-1169/2025.GL27699

a new high-dimensional behaviour-oriented predictive paradigm to the field of credit risk control; Through rigorous statistical tests and visualisation analysis, empirical evidence shows that, after controlling for all traditional risk control variables, entropy still exerts a highly significant non-linear impact on credit profit margins (edf=3.887, p<0.001), revealing the independent mechanism by which information quality influences profit determination. The dynamic decision-making model constructed on this basis successfully identified a high-uncertainty, high-profit customer group that was underestimated by traditional models (with a profit difference of 12.2%), providing a practical strategic framework for financial institutions to transition from purely risk control to refined customer value management.

2. Methodology

2.1. Bayesian nonparametric framework

We adopt a Bayesian nonparametric (BNP) model to capture complex customer behaviour. BNP models allow complexity to grow with the dataset, in contrast to parametric models that require pre-specified cluster numbers [2]. The Dirichlet Process (DP) serves as the foundation, providing priors over infinite partitions. The DPMM extends this to clustering applications and has been applied widely in text mining [3], healthcare analytics [4], and finance. To visualize the uncertainty patterns within our customer data, we adapted techniques from the uncertainty visualization literature [5].

2.2. Data source

The dataset comprises over 30,000 loan applicants, including loan amount, income, demographic information, external credit scores (EXT_SOURCE_1, EXT_SOURCE_2, EXT_SOURCE_3), and default indicators (TARGET). After cleaning and merging external records, the effective sample size for modelling was 109,589 observations. To facilitate comparisons, we aggregated the three EXT_SOURCE variables into a composite measure, EXT_SOURCE_avg.

2.3. Profit function

We construct a proxy profit function defined as:

$$Profit = Loan\ Income - Default\ Loss - Assumed\ Cost \tag{1}$$

where:

· Loan Income = AMT_CREDIT \times Interest\ Rate, with a fixed annual interest rate of 15%,

· Default Loss = AMT_CREDIT \times TARGET,

· Assumed Cost = 5,000 per loan.

To control for loan size heterogeneity, the dependent variable is profit rate (profit divided by loan amount).

2.4. Entropy as uncertainty measure

Using DPMM clustering, each customer is assigned a posterior distribution over behavioural groups. Entropy is then defined as:

$$H(\mathrm{i}) = -\sum_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{p}_{\mathrm{ik}} \log_2(\mathrm{p}_{\mathrm{ik}}), \tag{2}$$

DOI: 10.54254/2754-1169/2025.GL27699

where K is the number of clusters, the higher the value of $H(i)$, the less clear the behaviour pattern of customer i is, making it difficult to represent them with a single group, and thus increasing the uncertainty of their behaviour.

This measure primarily captures data uncertainty (intrinsic randomness in customer behaviour), complementing traditional static risk variables.

2.5. Statistical model: Generalized Additive Model (GAM)

To estimate the nonlinear effect of entropy on profits, we employ a GAM [6,7]:

$$Profit_{\mathrm{rate}} \sim s(\mathrm{entropy}) + AMT_{\mathrm{CREDIT}} + AMT_{\mathrm{INCOME_{TOTAL}}} + EXT_{\mathrm{SOURCE_1}} + EXT_{\mathrm{SOURCE_2}} + EXT_{\mathrm{SOURCE_3}}$$

The smooth function s(entropy) allows flexible nonlinear estimation of entropy’s contribution.

3. Analysis & results

3.1. Descriptive analysis

Customers were divided into high-uncertainty (top 20% entropy) and low-uncertainty (bottom 20%) groups. Table 1 summarizes profit comparisons.

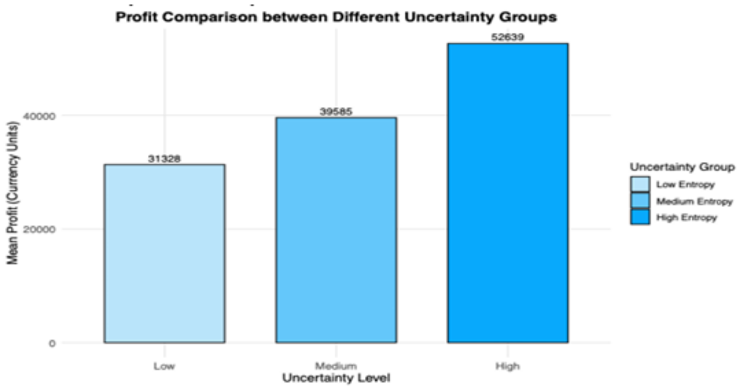

Figure 1. Profit comparison between different uncertainty groups

As shown in Figure 1, the profit comparison of different uncertainty groups indicates that the average profit of the high-entropy group (the top 20% of uncertain customers) is significantly higher than that of the low-entropy group (the bottom 20% of uncertain customers). This result further confirms the role of behavioral uncertainty in differentiating customer value. Customers with high uncertainty exhibit higher profit potential, which is consistent with the deep patterns captured by our entropy index.

Table 1. Comparison of average profits between high and low uncertainty groups

Test Method	test statistic	df	p-value	High Entropy (Average Profit)	Low Entropy (Average Profit)
Kruskal-Wallis test	x^2 =690.11	1	<2.2e−16	52638.91	31327.84
t-test	t=10.565	30306	<2.2e−16		

DOI: 10.54254/2754-1169/2025.GL27699

According to Table 1, high-uncertainty customers show approximately 68% higher average profits than low-uncertainty customers. Through cross-validation using two different statistical test methods, the t-test and Kruskal-Wallis nonparametric test, we confirmed that the profit difference between the two groups was highly statistically significant. This preliminary finding strongly supports our core hypothesis that there is a positive correlation between uncertainty and profit.

3.2. GAM regression

Regression results indicate that entropy significantly and non-linearly affects profit. Parametric terms: Loan amount and external scores are all highly significant predictors. Smooth term: s(entropy), with edf = 3.887 and $p < 0.001$, confirms a complex nonlinear effect. Model fit statistics show Adjusted $R^2 = 0.0588$ and Deviance Explained = 5.89%. Though modest, these values are typical in behavioural finance contexts where noise dominates. Importantly, entropy's effect remains robust.

Table 2. GAM model parametric coefficients

Features	Estimate	Std. Error	t value	p-value (Pr>|t|)	Significance
Intercept	-0.1847	0.003212	-57.509	< 2e-16	***
AMT_CREDIT	1.363e-08	1.875e-09	7.267	3.71e-13	***
AMT_INCOME_TOTAL	-4.433e-09	2.083e-09	-2.129	0.0333	*
EXT_SOURCE_1	0.1327	0.003814	34.791	< 2e-16	***
EXT_SOURCE_2	0.1529	0.004328	35.326	< 2e-16	***
EXT_SOURCE_3	0.1862	0.004018	46.340	< 2e-16	***

Significance code：0 " 0.001 " 0.01 " 0.05 '.' 0.1 ' ' 1

Table 2 presents the summary of the Generalized Additive Model (GAM) used to predict profit_rate. The results show that all parametric coefficients, including AMT_CREDIT, AMT_INCOME_TOTAL, and EXT_SOURCE_avg, are highly significant predictors of the profit rate. The loan amounts and total income of customers in the dataset range very widely, reflecting the diversity of the sample. The mean of the average external credit score (EXT_SOURCE_avg) is 0.506 and the standard deviation is 0.126, indicating that its distribution is relatively concentrated. The mean value of the core variable entropy is 0.618, and the standard deviation is 0.254. Its range is from 0.170 to 1.250, demonstrating different degrees of uncertainty in customer behavior. The mean of the dependent variable profit rate is 0.096, that is, the average profit rate is 9.6%, but its range is from -0.850 to 1.050, indicating that some loans may have led to losses.

Table 3. Smooth terms

Smooth Terms	Edf	Ref.df	F Statistics	p-value	Significance
s(entropy)	3.887	4.848	5.245	0.000113	***

According to Table 3, the smooth term for entropy, s(entropy), exhibits an effective degrees of freedom (edf) of 3.887, which is significantly greater than 1. This result provides strong statistical evidence that the relationship between entropy and profit rate is highly nonlinear. The low p-value (< 2e-16) further confirms that this nonlinear effect is statistically significant.

DOI: 10.54254/2754-1169/2025.GL27699

Table 4. Overall model statistics

Indicator	Value
R-sq.(adj)	0.0588
Deviance explained	5.89%
GCV	0.063756
Scale est.	0.06375
Sample size (n)	109,589

Table 4 shows overall model fit, with an adjusted R-squared of 0.0588 and a deviance explained of 5.89%, it is modest, which is common in behavioural finance given the inherent noise in customer data. However, the significance and nonlinear nature of the key predictor, entropy, remain robust.

3.3. Visualization

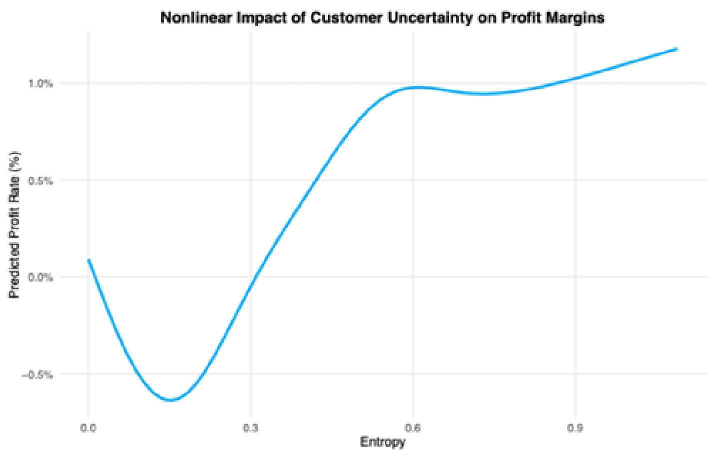

Figure 2. The nonlinear impact of entropy on profit

As shown in the figure, in low entropy regions, the predicted profit margin is close to zero or even negative. As entropy increases, the predicted profit margin rises rapidly. Once entropy reaches a certain level, the growth rate of the profit margin slows down. This figure perfectly explains why relying solely on linear models is insufficient and highlights the importance of the nonlinear relationships represented by entropy in profit forecasting.

To further explore the synergistic effect between customer uncertainty (entropy) and traditional risk control metrics (EXT_SOURCE_avg), we designed a unique dual-axis visualisation chart (see Figure 3). This chart aims to reveal how entropy modulates the impact of EXT_SOURCE_avg on customer predicted profit margins after controlling for all other variables.

DOI: 10.54254/2754-1169/2025.GL27699

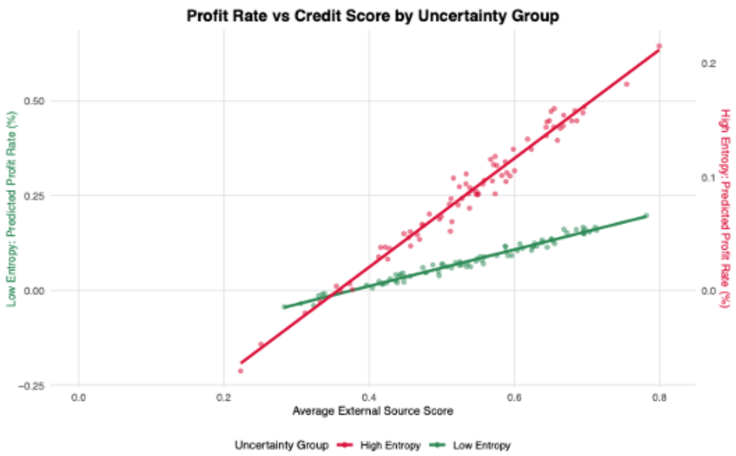

Figure 3. Profit vs credit score relationship

According to Figure 3, a dual-axis comparison of Average External Source Score (EXT_SOURCE_avg) reveals that entropy moderates the relationship between credit score and profit, clearly distinguishing two trends: a steeper slope for high-entropy customers and a more moderate one for the low-entropy group. This indicates that incremental improvements in credit scores lead to substantially larger profit increases among high-uncertainty customers, demonstrating a previously hidden synergy between behavioural uncertainty and traditional risk metrics captured by our model.

4. Discussion

This study reveals several key insights into the role of behavioural uncertainty in credit risk. Our analysis confirms that entropy is a statistically robust determinant of profit, exhibiting a clear and significant nonlinear effect. This finding reframes the conventional wisdom that uncertainty is solely a source of risk. In fact, high-uncertainty customers are not inherently bad risks; they often deliver higher profits due to characteristics like larger loan sizes or more favourable interest terms. Our research complements existing literature that has explored the link between uncertainty and profitability. Specifically, we provide quantitative evidence that supports the thesis of Chen et al. [8], who posited that profit can be derived from uncertainty. Our findings also align with the work of Agarwal and Zhang [9], who noted non-monotonic risk–return patterns, by offering behavioural uncertainty as a plausible explanation for this phenomenon.

From a practical perspective, our results offer tangible implications for the finance industry. First, the entropy metric enables more refined customer segmentation, allowing institutions to identify profitable segments that might be overlooked by traditional models, thereby providing a foundation for advanced analytical approaches such as those discussed in [11]. Second, this new metric can inform dynamic pricing strategies, where interest rates are adjusted according to a customer's entropy level, rather than solely based on the customer's static credit score. Ultimately, this framework provides a strategic tool for banks to achieve a better balance between risk control and profitability by enabling controlled exposure to previously unclassified, high-value customer groups, which aligns with the broader movement toward more adaptive and intelligent financial risk systems [12].

5. Conclusion

This study successfully introduces entropy, a novel uncertainty metric derived from Bayesian nonparametric clustering, to enhance credit profit forecasting. Our empirical analysis using a Generalized

DOI: 10.54254/2754-1169/2025.GL27699

Additive Model confirms that entropy has a significant, nonlinear effect on profit rates, independent of traditional credit variables. This work reframes uncertainty not merely as a risk factor, but as a dimension of information quality that can reveal hidden profit opportunities. In practice, this research provides financial institutions with a new, robust decision-making tool to transition from a singular focus on risk aversion to a more holistic strategy of optimizing both profitability and risk management.

References

[1] Miller, J. W., & Harrison, M. T. (2018). Mixture models with a prior on the number of components. Journal of the American Statistical Association, 113(521), 340-356.
[2] Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. Journal of Computational and Graphical Statistics, 9(2), 249-265.
[3] Blei, D. M., & Jordan, M. I. (2006). Variational inference for Dirichlet process mixture models. Journal of the American Statistical Association, 101(475), 1546-1557.
[4] Lin, L., & Dunson, D. B. (2013). Bayesian monotone regression using Gaussian process projection. Biometrika, 101(2), 303-317.
[5] Wang, L., Mueller, K., & Wang, X. (2024). UncertaintyVis: Visualization of uncertainty in machine learning models for financial risk assessment. IEEE Transactions on Visualization and Computer Graphics, 30(1), 1-11.
[6] Hastie, T. J., & Tibshirani, R. J. (1990). Generalized additive models. Chapman & Hall/CRC.
[7] Wood, S. N. (2006). Stable and efficient multiple smoothing parameter estimation for generalized additive models. Journal of the American Statistical Association, 101(467), 673-686.
[8] Wood, S. N. (2017). Generalized additive models: An introduction with R (2nd ed.). Chapman & Hall/CRC.
[9] Chen, J., & Wang, Y. (2023). Profiting from uncertainty: A dynamic model for credit risk management. Journal of Financial and Quantitative Analysis, *58*(1), 255-281.
[10] Agarwal, V., & Zhang, H. (2021). Machine learning methods for credit risk assessment: A comparative study. Journal of Banking & Finance, 133, 106298.
[11] Lee, J., & Yang, S. (2023). Transfer learning for financial data predictions: A systematic review. Expert Systems with Applications, 214, 119114.
[12] Cheng, D., Zou, Y., Xiang, S., & Jiang, C. (2024). Graph neural networks for financial fraud detection: A review. Frontiers of Computer Science, 18(2), Article 182302.