



Credit Scoring with Machine Learning Supported by E-Commerce Data

Sema Işık Çalışkan^{1*}, Tuncer Cem Uğurluer², Emre Arıkan³, Sinan Uzun⁴, Muhammet Alper Aydın⁵,
Handan Derya Ercan⁶, Yavuz Selim Hindistan⁷

¹ Hepsipay, Orcid ID: <https://orcid.org/0000-0001-6796-278X>,
e-mail: sema.caliskan@hepsipay.com

² Hepsipay, <https://orcid.org/0009-0004-0699-4330>,
e-mail: tuncer.ugurluer@hepsipay.com

³ Hepsipay, <https://orcid.org/0009-0007-1127-6992>,
e-mail: emre.arikan@hepsipay.com

⁴ Hepsifinans, Orcid ID: <https://orcid.org/0000-0001-5274-259X>,
e-mail: sinan.uzun@hepsifinans.com

⁵ Hepsifinans, Orcid ID: <https://orcid.org/0009-0000-1543-9065>,
e-mail: muhammet.aydin@hepsifinans.com

⁶ Bogazici University, Orcid ID: <https://orcid.org/0000-0003-1193-6560>,
e-mail: handan.ercan@std.bogazici.edu.tr

⁷ Ozyegin University, Orcid ID: <https://orcid.org/0000-0001-9031-8167>,
e-mail: yavuzselim.hindistan@ozyegin.edu.tr

* Correspondence: sema.caliskan@hepsipay.com; (+90 534 010 5317)

Received: 18 June 2025

Revised: 21 August 2025

2nd Revised: 28 September 2025

3rd Revised: 24 October 2025

Accepted: 02 November 2025

Published: 31 December 2025

This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Reference: Çalışkan, S. I., Uğurluer, T. C., Arıkan, E., Uzun, S., Aydın, M. A., Ercan, H. D., & Hindistan, Y. S. (2025). Credit scoring with machine learning supported by e-commerce data. *Orclever Proceedings of Research and Development*, 7(1), 105–116.

Abstract

With the rapid growth of e-commerce, the need for credit in e-commerce has increased. E-commerce platforms require high performance as a competitive advantage in their activities. Traditional credit risk models need improvement to sustain the performance expected by e-commerce platforms. In this study, we investigate alternative behavioral and transactional variables obtained from an e-commerce platform. We examine whether these variables improve the predictive



performance of credit risk models beyond traditional financial data. Our research is based on a real e-commerce environment where a machine learning based credit scoring system was implemented. The study focuses on developing and evaluating a credit risk system that integrates platform specific behavioral data, such as shopping frequency, payment methods, Buy Now Pay Later (BNPL) repayment behavior, and wallet usage, with traditional financial and Credit Bureau(CB) indicators. Our findings demonstrate a significant improvement in model discrimination and Gini performance. The localized AI-driven credit scoring system achieved a low-cost, fast, and more accurate credit assessment.

Keywords: AI-driven credit scoring, Buy Now Pay Later (BNPL), financial risk assessment, alternative data, fintech, credit risk, machine learning



1. Introduction

With the rapid growth of e-commerce, the need for credit in online shopping has increased. Credit scoring on e-commerce platforms requires high performance due to the competitive nature of online markets. There is a strong need for flexible and fast credit systems. Online customers increasingly prefer deferred payment methods such as Buy Now, Pay Later (BNPL) because those payment methods allow them to complete purchases without immediate payment [5]. These developments have increased the pressure on both financial institutions and fintech companies to assess credit risk in real time and with higher precision.

Traditional credit scoring models were developed for banking environments. They rely on static and backward-looking data such as income, liabilities, and Credit Bureau (CB) history [1]. The Credit Bureau of Türkiye is a centralized institution that collects, aggregates, and shares individual and corporate credit information from financial institutions to support credit risk assessment, financial transparency, and informed lending decisions. These models are slow to update and are not well suited to digital consumers who have short or incomplete credit histories [2]. Consequently, many potentially creditworthy e-commerce users remain underserved, while risk prediction errors remain common.

On the other hand, there are vast amounts of behavioral and transactional data in e-commerce platforms, such as purchase frequency, product category, payment type, repayment patterns, and wallet activity [1], [2]. This information reflects both consumer's financial discipline and behavioral reliability. These variables can strengthen credit risk prediction and reduce default uncertainty in hybrid models that combines them with traditional financial indicators [5].

Large and rapidly changing data sources create complex frameworks, so big data approaches are required to support their scalability [3]. Processing such data requires Machine Learning (ML) methods. ML algorithms can also identify patterns that traditional statistical techniques may miss. They allow credit scoring systems to adapt dynamically and deliver more accurate predictions [2], [5].

There are several studies highlighting the potential of alternative data in credit scoring. Although non-traditional data and ML techniques improve prediction quality, empirical evidence comparing these models with traditional lending systems remains limited.



Current applications (such as e-commerce transactions and digital wallet histories) are still in early stages, particularly in local markets where automated decision systems are uncommon [6]–[8]. Huang et al. note that further empirical work is needed to validate the reliability and generalizability of such models across diverse populations and economic contexts [8].

This study is based on a project of ML-based credit evaluation system that integrates traditional financial indicators (e.g., CB data, debt ratios, income levels) with behavioral and transactional variables from e-commerce platforms (e.g., BNPL activity, wallet usage, and purchase behavior). The project's innovations include creating an in-house solution, applying advanced ML and data mining techniques on real data, and continuously improving models through performance monitoring. While prior studies show that alternative data enhances credit risk prediction, empirical validation in operational e-commerce systems remains limited [2], [5], [8], [9]. To bridge this gap, this study implements an ML-based Credit Scoring Decision Support System and examines whether alternative behavioral and transactional variables from e-commerce platforms improve model performance compared to traditional financial data.

2. Literature Review

Credit evaluation and its risk are a major concern in the financial sector. Supervising both liquidity and credit risk remains a core part of banking operations [6]. Although traditional credit scoring models have been used to estimate default probability, they are limited in digital finance environments [5]. These models depend on structured factors such as income, payment history, and CB data. Many consumers with little or no formal credit history remain excluded because of this dependency on structured factors [8].

The growth of FinTech and Big Data has expanded the scope of risk assessment [3], [8]. Alternative data introduces non-traditional indicators that reflect digital behavior, and it offers new insights for more accurate and inclusive credit evaluation. Alternative data includes financial information such as e-commerce transactions and utility payments, or non-financial details such as social media activity [7]. E-commerce platforms generate detailed behavioral data, including shopping patterns, repayment habits, and wallet usage. Platforms such as Alibaba's Ant Financial and Mercado Libre use this data to allocate credit and manage portfolio risk [7]. Xu et al. [2] show that analyzing seller transaction data on Taobao supports sustainable and trustworthy e-commerce financing.



Cai and Yan [4] add that e-commerce platform financing enables faster, and collateral-free loans compared to banks.

The volume and variety of alternative data require ML techniques for analysis [8]. ML models can identify patterns that conventional linear models cannot. They enhance both the speed and accuracy of credit scoring [5]. Gambacorta et al. [9] research showed that incorporating non-traditional data, such as information from mobile phone applications and e-commerce platforms, improved the discriminatory power of credit scoring models by 2.2%. The use of ML techniques provided an even greater boost, adding an extra 5.3% to the models' overall predictive power. The study demonstrated that ML-based models perform better during periods of economic instability or stress because they are better equipped than linear models to uncover the complex, non-linear relationships between variables when the financial environment changes suddenly.

The Gini index is a global measure of a scoring function's quality and discriminatory power measure in credit scoring models. It is defined as twice the area between the Lorenz curve and the diagonal, with values ranging from 1 (for an ideal model that perfectly separates clients) to 0 (for a random model) [10]. The Gini index uses the following quality levels:

- Ideal/Perfect Model (Successful): An ideal model, which perfectly separates good and bad clients, has a Gini index equal to 1 (100%).
- Random Model (Weak): A model that assigns a random score to a client has a Gini index equal to 0.

3. Methodology

This study was designed to empirically answer the question that integrating behavioral and transactional data from an e-commerce platform can significantly enhance the predictive performance of a credit risk model beyond traditional financial data alone. The methodology was structured around a four-stage development lifecycle, as summarized in Table 1, encompassing data preparation, model development, evaluation, and implementation.

To conduct this analysis, two primary datasets were utilized. The first consisted of traditional financial data, sourced from a CB, and included variables such as income, existing loan obligations, credit card usage, and historical delinquency records. The



second dataset comprised alternative data from an e-commerce platform, which was categorized into specific behavioral and transactional variables. These were selected to capture a customer's financial discipline and reliability directly from their digital footprint:

- **Transaction Frequency & Volume:** Metrics such as the number of purchases per month and total spending amount.
- **Payment Method Patterns:** The frequency of using different payment options, including credit cards, debit cards, and digital wallets.
- **BNPL Repayment Behavior:** Critical indicators such as the number of active BNPL loans, historical on-time repayment rate, and instances of late or missed payments.
- **Digital Wallet Usage:** Intensity of wallet use, including the linkage of bank cards to the wallet and the average balance maintained.
- **Product Category Affinity:** The categories a customer most frequently shops in (e.g., electronics, childcare products, cosmetics), which serve as proxies for lifestyle and financial stability.

The initial data universe contained 15.4 million shared customers, from which 1.4 million were successfully matched using masked identifiers to ensure anonymity. All data handling and processing adhered to GDPR and local banking regulations.

The project workflow began with “Data Preparation”. In this stage, the financial and e-commerce datasets were cleaned, merged, and filtered to include only active customers. The final dataset, now enriched with the behavioral variables, was partitioned into a 70% training set, a 30% test set, and an out-of-time validation set to ensure the model's robustness and temporal stability. During the “Model Development” phase, the target variable was defined as the occurrence of either legal action or a delinquency of 90 days or more within a 12-month performance window. A diverse set of machine learning algorithms were employed, including Light Gradient Boosting Machine (LGBM), Multilayer Perceptron (MLP), Random Forest, Decision Tree, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Logistic Regression, to capture the complex, non-linear relationships inherent in the new behavioral data. Rigorous data preprocessing was applied, involving imputation for missing values, encoding of categorical variables, and the removal of features with poor predictive quality. The “Model Evaluation” stage involved training and assessing three distinct baseline modules to isolate the predictive contribution of different data types:

- **Module 1:** Consumer Finance CB, built solely on traditional financial data.



- **Module 2:** E-commerce CB, which combines CB data with the e-commerce behavioral variables.
- **Module 3:** E-commerce Internal, relying exclusively on the platform's internal behavioral data.

Table 1: Credit Scoring Project Development Summary

Source: Authors' compilation based on project workflow and model development records.

| Phase | Goal/Process | Key Data & Methods |
|----------------------------|--|--|
| 1. Data Preparation | Merge, clean, and segment datasets for model training. | 15.4M shared customers - 1.4M matched (masked ID). |
| | | Financial + e-commerce data (loans, transactions, shopping). |
| | | 70/30 Train-Test split + Out-of-Time set. |
| 2. Model Development | Define default criteria and train algorithms. | Default: legal action or 90+ days delinquency within 12 months. |
| | | Algorithms: LGBM, MLP, Random Forest, Logistic Regression, SVM, XGBoost, Decision Tree |
| | | Feature selection via missing-rate and performance filtering. |
| 3. Evaluation | Build and assess three standalone modules. | Module 1: Consumer Finance CB |
| | | Module 2: e-Commerce CB |
| | | Module 3: Internal Data |
| | | Ensemble via Logistic Regression (Stacking). |
| 4. Implementation | Deploy automated decision system. | Final Ensemble 1 (Module 1 + Module 3) |
| | | Final Ensemble 2 (Module 2 + Module 3) |
| | | Deciding Weights for CB and Internal Data. |

To enhance predictive stability and accuracy, these three modules were combined using a stacking ensemble technique, with a Logistic Regression model as the meta-learner. This approach allowed the system to optimally weigh the predictions from each base module. Finally, for “Implementation and Monitoring”, the ensemble model was deployed into an automated decision system capable of rendering credit decisions in five seconds or less.

4. Result

The implemented ML credit scoring system successfully integrated traditional financial data with behavioral variables from the e-commerce platform. It made significant gains in predictive accuracy and operational efficiency. The end-to-end development process, summarized in Figure 1, ensured a robust and scalable system.



The predictive framework was built on three core scoring modules, each leveraging different data sources, as detailed in the Methodology. The standalone performance of these modules was evaluated to isolate the contribution of each data type:

- **Module 1 (Consumer Finance CB)**, using only traditional CB data, established a strong performance baseline (Gini = 76.75%). The detailed algorithm-level performance is shown in Table 2. Out-of-Time validation refers to model testing on data from a later time period not used during training or testing, ensuring temporal robustness. LGBM and Random Forest achieved the highest discrimination and generalization capacity, confirming that tree-based ensemble methods outperform simpler linear or kernel-based models in the CB dataset.

Table 2: Module 1 (Consumer Finance CB)

Source: Authors' calculations based on model evaluation metrics

| Model Name | Train Gini | Test Gini | Out-of-Time Gini |
|---|---------------|---------------|------------------|
| Light Gradient Boosting Machine (LGBM) | 84.40% | 76.75% | 75.50% |
| Multilayer Perceptron (MLP) | 78.00% | 76.20% | 74.30% |
| Random Forest | 85.50% | 75.80% | 72.60% |
| Logistic Regression | 75.40% | 75.50% | 72.20% |
| Support Vector Machine (SVM) | 73.00% | 72.10% | 69.90% |
| Extreme Gradient Boosting (XGBoost) | 72.70% | 72.50% | 68.20% |
| Decision Tree | 73.70% | 70.00% | 66.00% |

- **Module 2 (E-commerce CB)**, which combined CB and behavioral data, showed similar performance (Gini = 75.87%). The detailed algorithm-level performance is shown in Table 3. The hybrid structure benefited from alternative data: LGBM and XGBoost again ranked highest, indicating robustness against time-shift and data-drift effects.

Table 3: Module 2 (E-commerce CB)

Source: Authors' calculations based on model evaluation metrics

| Model Name | Train Gini | Test Gini | Out-of-Time Gini |
|--|---------------|---------------|------------------|
| Extreme Gradient Boosting (XGBoost) | 81.97% | 75.87% | 72.97% |
| Light Gradient Boosting Machine (LGBM) | 83.07% | 75.27% | 72.47% |
| Multilayer Perceptron (MLP) | 74.37% | 74.37% | 70.57% |
| Random Forest | 84.07% | 74.07% | 70.07% |
| Logistic Regression | 73.27% | 73.57% | 69.77% |
| Support Vector Machine (SVM) | 70.67% | 70.57% | 67.27% |
| Decision Tree | 72.07% | 69.77% | 64.27% |



- **Module 3 (E-commerce Internal)**, relying solely on platform-specific behavioral data, had limited discriminatory power on its own (Gini = 40.43%), indicating that behavioral data alone is insufficient for robust credit assessment.

The key innovation was combining these modules into a stacking ensemble. The final ensemble models achieved a Gini of 78.14% for the Consumer Finance segment and 77.57% for the E-commerce segment. This represents an overall improvement of approximately four Gini points over the initial baseline, demonstrating that the behavioral data from Module 3 provides a unique predictive signal that significantly enhances the core financial models when combined intelligently.

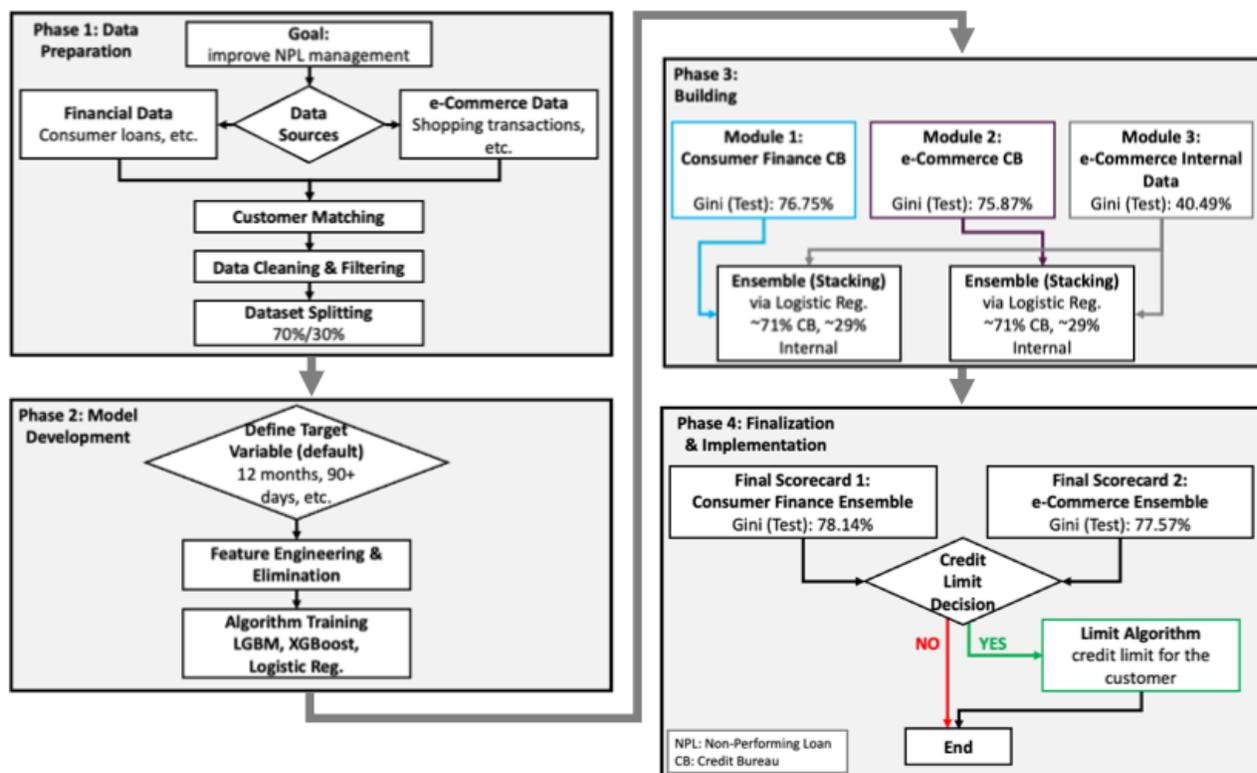


Figure 1: Credit Scoring Project Development Steps

The deployment of the ensemble model transformed the credit evaluation process. As summarized in Table 4, the operational results were substantial.

- **Efficiency:** The system fully automated credit decisions, slashing the average decision time from ~30 seconds (manual) to under 5 seconds, an 83% reduction.
- **Portfolio Growth and Quality:** This speed and accuracy fueled a 62.5% increase in monthly average credit transaction volume. Crucially, this expansion was



achieved while improving portfolio quality, as the NPL rate fell from 4.0% to 2.9%, a 27.5% reduction.

- **Customer Satisfaction:** The enhanced experience was reflected in a 6% increase in the Net Promoter Score (NPS) for BNPL products.

Table 4: Module and Operational Performance Summary

Source: Authors' calculations based on model evaluation and operational performance data.

| Metric | Before ML | After ML (Ensemble) | Improvement / Observation |
|----------------------------------|--------------------------------|-----------------------|----------------------------------|
| Decision Time per Application | ~30 sec (manual) | ≤ 5 sec (automated) | -83 % time reduction |
| Credit Transaction Volume | 400M TRY/month | 650M TRY/month | + 62.5% volume increase |
| Non-Performing Loan (NPL) Rate | 4.00% | 2.90% | -1.1 pp (27 % drop) |
| Customer Satisfaction (NPS) | Pre-deployment Baseline | +6% | Positive user responses for BNPL |

An in-depth analysis revealed which behavioral variables drove this performance improvement. Feature importance analysis confirmed that digital wallet usage with a linked card was a particularly strong predictor, associated with a 60% lower risk profile. Furthermore, behavioral segmentation revealed that purchase categories served as reliable proxies for financial discipline: customers frequently purchasing childcare products exhibited lower default risk, while frequent cosmetic buyers showed higher spending volatility and slightly elevated credit risk.

5. Discussion and Conclusion

This study provides clear empirical evidence that incorporating behavioral and transactional data from e-commerce platforms significantly enhances credit risk prediction accuracy. The findings directly address the central research question, revealing that integrating these alternative variables into a hybrid machine learning model yields an approximately four-point increase in the Gini coefficient, thereby validating their added predictive value beyond traditional financial data.

The project's primary innovation lies in its execution as a proprietary, in-house solution. As highlighted in the introduction, this approach directly addresses the limitations of traditional systems that are often "high-cost, time-consuming, and error-prone." By developing a local system, we achieved a low-cost, fast, and more accurate credit assessment system, reducing decision times by 83% and decreasing the NPL rate by 27.5%. This shift from rule-based logic to a dynamic ML ensemble allowed for the



nuanced capture of financial discipline through variables like BNPL repayment behavior and digital wallet usage, which emerged as top predictors.

While the ensemble model proved highly effective, the limited standalone power of the internal behavioral module (Gini = 40.43%) is an important finding. It underscores that e-commerce data is most powerful as a complement to, not a replacement for, traditional data, revealing complex patterns that enhance a core financial model.

Future work will focus on maintaining this competitive edge through continuous ML operations monitoring and extending the framework to new products such as micro-loans and seller financing. Additionally, a new project is being developed for credit allocation to merchants operating on the e-commerce company platform. Among these merchants, individual sellers will be evaluated using the developed scorecard model to ensure consistent and data-driven risk assessment. This project establishes a validated, operational blueprint for leveraging digital behavior to build more inclusive, responsive, and reliable credit systems in the digital economy.

6. Acknowledge

This work was supported by Hepsipay as part of an internal Research & Development (R&D) project. The authors thank the company's data science and engineering teams for their invaluable collaboration and for providing the data and computational resources essential to this study.



References

- [1] L. Jia, G. Xue, Y. Fu, and L. Xu, "Factors Affecting Consumers' Acceptance of E-Commerce Consumer Credit Service," *International Journal of Information Management*, vol. 40, pp. 103–110, Feb. 2018, doi: 10.1016/j.ijinfomgt.2018.02.002
- [2] Y.-Z. Xu, J.-L. Zhang, Y. Hua, and L.-Y. Wang, "Dynamic Credit Risk Evaluation Method for E-Commerce Sellers Based on a Hybrid Artificial Intelligence Model," *Sustainability*, vol. 11, no. 19, p. 5521, Oct. 2019, doi: 10.3390/su11195521
- [3] Y. S. Hindistan, B. Y. Kiyakoğlu, A. M. Rezaeinazhad, H. E. Korkmaz, and H. Dağ, "Alternative Credit Scoring and Classification Employing Machine Learning Techniques on a Big Data Platform," in *Proc. 2019 International Conference on Information and Communication Technologies (ICT)*, Istanbul, Turkey, 2019, pp. 1–7.
- [4] S. Cai and Q. Yan, "Online Sellers' Financing Strategies in an E-Commerce Supply Chain: Bank Credit vs. E-Commerce Platform Financing," *Electronic Commerce Research*, vol. 23, pp. 2541–2572, Apr. 2023, doi: 10.1007/s10660-022-09552-w
- [5] M. R. Mahmud, M. R. Hoque, T. Ahammad, M. N. H. Hasib, and M. M. Hasan, "Advanced AI-Driven Credit Risk Assessment for Buy Now, Pay Later (BNPL) and E-Commerce Financing: Leveraging Machine Learning, Alternative Data, and Predictive Analytics for Enhanced Financial Scoring," *Journal of Business and Management Studies*, vol. 6, no. 2, pp. 180–189, Mar. 2024, doi: 10.32996/jbms.2024.6.2.19
- [6] G. Altan and S. Demirci, "Credit Scoring on Cash Flow Table with Machine Learning: XGBoost Approach," *Journal of Economic Policy Researches*, vol. 9, no. 2, pp. 397–424, 2022.
- [7] World Bank Group, *The Use of Alternative Data in Credit Risk Assessment: Opportunities, Risks, and Challenges*, Washington, DC, USA, 2024.
- [8] Y. Huang, Z. Li, S. Pan, and X. Wen, "A Review of Alternative Data in Credit Risk," in *Proceedings of the 7th International Conference on Economic Management and Green Development*, 2023, pp. 212–220, doi: 10.54254/2754-1169/31/20231545.
- [9] L. Gambacorta, Y. Huang, H. Qiu, and J. Wang, "How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm," *Journal of Financial Stability*, vol. 73, p. 101284, 2024, doi: 10.1016/j.jfs.2024.101284.
- [10] M. Řezáč and F. Řezáč, "How to measure the quality of credit scoring models," *Finance a úvěr – Czech Journal of Economics and Finance*, vol. 61, no. 5, pp. 486–507, 2011.