# AI-Enhanced Satellite Imagery in Institutional Lending Decisions: A Difference-in-Differences Study of Retail Firm Performance

Research Proposal

Winter 2024/25 - Research Topics in Finance I

Melisa Mazaeva

May 12, 2025

**Abstract**

This proposal explores whether advanced artificial intelligence techniques applied to satellite imagery can enhance retail firm performance measurement and influence institutional lending behavior. Building on Kang's (2024) findings that satellite-derived car counts reduce lenders' reliance on private information, this study introduces a richer performance metric using deep learning methods—including Convolutional Neural Networks for feature extraction, Faster Region-based Convolutional Neural Network for object detection, and Temporal Convolutional Networks for temporal analysis. The AI-enhanced metric is incorporated into a difference-in-differences framework to test whether superior predictive signals alter lenders' behavior compared to traditional car count data. Syndicated loan records from DealScan (2020–2024) will be matched with Compustat/CRSP firm data and high-resolution satellite-derived parking lot metrics. Improved measures—such as vehicle turnover, density, and contextual indicators—are expected to reduce reliance on private information and lower institutional participation in loan syndicates. The findings are expected to inform institutional decision-making and regulatory standards, contributing to finance literature by reshaping information asymmetry in credit markets while addressing prior data limitations.

# Table of contents

# 1 List of Abbreviations

**AI**: Artificial Intelligence
**CNN**: Convolutional Neural Network
**CRSP**: Center for Research in Security Prices
**DiD**: Difference-in-Differences
**DV**: Dependent Variable
**Faster R-CNN**: Faster Region-based Convolutional Neural Network
**IV**: Independent Variable
**TCN**: Temporal Convolutional Network
**US**: United States
**WRDS**: Wharton Research Data Services

## 2 Introduction

Institutional lenders are known to gain an information advantage by accessing borrowers' private data through lending relationships and using it for informed trading (Ivashina and Sun 2011). However, the rise of alternative data, particularly satellite imagery, can offer real-time insights into firm performance that can erode this advantage (Kang 2024). Kang (2024) provided evidence that once daily car-count data from satellite images became available for retail firms, institutional investors were less inclined to participate in those firms' syndicated loans. The intuition is that accurate, timely public signals of performance undermine the exclusive value of private information. The main research question addressed in this study is: If institutional lenders have access to AI-enhanced performance signals from satellite imagery, do they further reduce their reliance on private information (and thus participation in loans) compared to the era of simpler satellite metrics? This research question is important because it investigates how advances in Artificial Intelligence (AI) can fundamentally transform the competitive dynamics of credit markets, making public alternative data nearly as informative as proprietary private data, which has significant implications for regulatory policy and lending practices.

Advancements in computer vision now allow a reexamination of satellite imagery. Whereas traditional methods like Orbital Insight's car-count algorithm used by Kang (2024) provide useful but limited measures (e.g. excluding covered parking), modern AI models can overcome these gaps. Convolutional Neural Networks (CNNs) can extract high-level features from raw images (Krizhevsky, Sutskever, and Hinton 2017); Faster Region-based Convolutional Neural Network (Faster R-CNN) improves vehicle detection and context (Ren et al. 2017); and Temporal Convolutional Networks (TCNs) capture temporal patterns to yield a more stable performance indicator (Lea et al. 2017). This research examines whether refined public signals from AI-enhanced imagery further reduce institutional lenders' reliance on private information. The structure of this proposal is as follows: the next section reviews related literature, followed by details on data and methodology, and concludes with anticipated results and contributions.

## 3 Related Literature

This study builds on and extends existing research on alternative data in capital markets and institutional lending by exploring the evolving role of AI-enhanced satellite imagery in reducing information asymmetry. Prior work indicates that the introduction of satellite data coverage for major retailers enables sophisticated investors to design profitable trading strategies (Katona et al. 2024). Feng and Fay (2022) further find that parking lot traffic predicts forward-looking performance (Tobin's Q) for retailers, underlining the value of such data even for future expectations. Kang (2024) showed that the availability of daily satellite data reduced institutional participation in loans, suggesting that robust public signals can diminish the need for private data. However, these basic metrics miss nuances such as covered parking and the distinction between short- and long-term trends. It is unclear whether advanced AI techniques can yield more informative public signals.

This study employs sophisticated AI to extract richer features, positing that these enhanced signals will provide deeper insights into firm performance and more strongly influence investor behavior.

Based on these insights, the primary hypothesis is that the integration of AI-enhanced performance signals from satellite imagery will lead to a further reduction in institutional lenders' participation in syndicated loans relative to traditional car count measures, as more precise signals reduce the marginal benefit of acquiring private information for informed trading.

# 4 Data

Data will focus on U.S. retail firms during 2020–2024—a period when deep learning for image analysis has matured, following the commercialization of car-count data since 2015 (Kang 2024). High-resolution raw satellite images from providers such as Planet Labs or Maxar are expected to yield tens of thousands of daily observations across roughly 500 firms. These images will be processed using a fine-tuned Faster R-CNN for object detection and a TCN to capture temporal variations, yielding a multidimensional performance index that overcomes limitations outlined by Kang (2024) (e.g., exclusion of covered parking lots).

The AI-derived metrics will be merged with syndicated loan data from DealScan and firm financials from Compustat and Center for Research in Security Prices (CRSP) using common identifiers; all three datasets are available through Wharton Research Data Services (WRDS) (Wharton Research Data Services 2025). Key variables include vehicle count, parking lot utilization rate, turnover (entries/exits), and spatial distribution of vehicles, which will be aggregated into a performance index. The study will use the same syndicated loan data from DealScan as Kang (2024) and match loan issuance details with firm financial data from Compustat and CRSP. The treatment group consists of retail firms with AI-enhanced metrics from 2020 onward, while control firms are the ones that did not experience significant improvements in data measurement, with 2015–2019 as the baseline and 2020–2024 as the post-treatment period. A difference-in-differences (DiD) approach will be employed to isolate the effect of AI-enhanced signals on lending behavior. Data limitations—such as missing imagery due to poor weather or resolution variations—will be addressed via cleaning, imputation, and robustness checks. With unlimited funds, the cost of these datasets is not a constraint, and the data are readily available.

# 5 Methodology

The empirical strategy employs a DiD design to compare changes in institutional lending behavior for treated versus control firms before and after the introduction of AI-enhanced performance metrics. Conceptually, Kang (2024) implemented a DiD where the "treatment" was the introduction of basic satellite car-count coverage. This study maintains a similar framework but redefines the treatment in terms of AI-enhanced performance signals. An indicator variable (`AI_Treat_i`) is defined for each firm such that it equals 1 if the firm is measured using AI-enhanced satellite metrics, and 0 otherwise. A post-period indicator (`Post_t`) is set to 1 for observations from 2020 to 2024 and 0 for those from 2015 to 2019. The interaction term (`AI_Treat_i × Post_t`) thus captures the effect of the enhanced performance signal on institutional lending outcomes.

The baseline DiD regression model is specified as follows:

$$\mathrm{InstPart}_{i,t} = \alpha + \beta_1 \left(\mathrm{AI\_Treat}_i \times \mathrm{Post}_t\right) + \beta_2\, \mathrm{AI\_Treat}_i + \beta_3\, \mathrm{Post}_t + \gamma X_{i,t} + \varepsilon_{i,t},$$

where $\mathrm{InstPart}_{i,t}$ represents the measure of institutional lender participation (e.g., a dummy for institutional presence or the percentage share of the loan) and serves as the dependent variable (DV). The key independent variable (IV) is the interaction term $\mathrm{AI\_Treat}_i \times \mathrm{Post}_t$, which measures the impact of AI-enhanced performance signals on lending behavior. Additional IVs include the binary treatment indicator $\mathrm{AI\_Treat}_i$, the post-period indicator $\mathrm{Post}_t$, and control variables $X_{i,t}$ such as firm size, credit rating, and loan maturity. Fixed effects at the industry×time level will be incorporated to account for broader trends. A significantly negative $\beta_1$ is anticipated, indicating that improved public signals reduce the need for private information.

The AI pipeline involves several deep learning models to construct the enhanced performance index. First, Convolutional Neural Networks (e.g., ResNet) will extract features from raw satellite images. Next, Faster R-CNN will be applied for precise vehicle detection and to capture spatial arrangements, hence, overcoming limitations such as the exclusion of covered parking data as outlined by Kang (2024). Finally, TCNs will aggregate daily observations into a smooth, temporally robust index. This AI-derived performance index will then be merged with syndicated loan data from DealScan and firm financials from Compustat and CRSP using common identifiers.

The AI component of the methodology is implemented in Python using libraries such as PyTorch and torchvision. Initially, satellite images are pre-processed and converted into numerical arrays that the deep learning model can work with efficiently. A pre-trained Faster R-CNN model (specifically, fasterrcnn_resnet50_fpn) is then used to detect vehicles within these images, thereby estimating the number of cars present. The detected counts are aggregated into daily metrics, which are subsequently fed into a TCN to capture temporal patterns such as seasonal trends and fluctuations. The resulting time series signals—representing refined measures of retail firm performance, such as weekly average parking lot occupancy or turnover rates—serve as the key independent variables in the difference-in-differences regression analysis.

Robustness checks will test the parallel trends assumption by examining pre-2020 lending patterns for both groups, and additional falsification tests (e.g., using a false treatment date) will be conducted to ensure that the observed effects are not driven by other unrelated factors.

## 6 Proposed Contribution

This study offers several contributions. First, it advances the literature on alternative data by demonstrating how AI-driven methods can extract a richer, multidimensional signal from satellite imagery. While Kang (2024) relied on basic car counts—which had limitations such as exclusion of images from covered parking lots and limited contextual detail—the proposed AI approach captures additional dimensions like spatial distribution, vehicle turnover, and temporal patterns. Second, by integrating this refined metric with established syndicated loan data within a DiD framework, the study contributes to the understanding of how improved public information affects institutional lending behavior. The results are anticipated to provide policy-relevant insights on the shifting balance between public and private information in credit markets. Finally, from a methodological perspective, the study presents a blueprint for incorporating innovative AI techniques into empirical finance research, offering a model that can be adapted for other applications.

# 7 Anticipated Results

It is expected that institutional lenders will reduce their participation in syndicated loans for firms with AI-enhanced performance metrics relative to those measured by traditional car counts (Kang 2024). In a DiD framework, a negative and significant coefficient on the interaction term (`AI_Treat × Post`) would indicate that superior public signals lessen the need for private information, with a larger effect than observed with simple car counts. Treated firms are also anticipated to face less favorable loan terms (e.g., higher spreads, stricter covenants). Robustness checks will assess whether the effect is especially strong for opaque firms, and the study will verify that the AI-driven index more accurately predicts firm performance (e.g., sales growth) than traditional metrics.

Following approval and completion, the findings will be shared in academic journals and policy briefings to influence regulatory frameworks and lending practices. The results may inform policymakers that advanced public data can substitute for private information, reshaping credit market transparency. Study limitations—such as the reliance on high-quality satellite imagery and difficulty isolating AI effects—will be critically examined. Future research may expand to other geographic regions to examine whether similar AI-enhanced metrics can improve public signals of firm performance and influence lending behavior in diverse market environments, thereby shedding light on the broader applicability of these methods in different regulatory and economic contexts.

# References


Feng, Cong, and Scott Fay. 2022. "An Empirical Investigation of Forward-Looking Retailer Performance Using Parking Lot Traffic Data Derived from Satellite Imagery." *Journal of Retailing* 98 (4): 633–46. https://doi.org/10.1016/j.jretai.2022.03.004.

Ivashina, Victoria, and Zheng Sun. 2011. "Institutional Stock Trading on Loan Market Information." *Journal of Financial Economics* 100 (2): 284–303. https://doi.org/10.1016/j.jfineco.2010.11.004.

Kang, Jung Koo. 2024. "Gone with the Big Data: Institutional Lender Demand for Private Information." *Journal of Accounting and Economics* 77 (2-3): 101663. https://doi.org/10.1016/j.jacceco.2023.101663.

Katona, Zsolt, Marcus O. Painter, Panos N. Patatoukas, and Jean Zeng. 2024. "On the Capital Market Consequences of Big Data: Evidence from Outer Space." *Journal of Financial and Quantitative Analysis*, April, 1–29. https://doi.org/10.1017/S0022109023001448.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2017. "ImageNet Classification with Deep Convolutional Neural Networks." *Communications of the ACM* 60 (6): 84–90. https://doi.org/10.1145/3065386.

Lea, Colin, Michael D. Flynn, Rene Vidal, Austin Reiter, and Gregory D. Hager. 2017. "Temporal Convolutional Networks for Action Segmentation and Detection." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1003–12. Honolulu, HI: IEEE. https://doi.org/10.1109/CVPR.2017.113.

Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2017. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (6): 1137–49. https://doi.org/10.1109/TPAMI.2016.2577031.

Wharton Research Data Services. 2025. "Get Data - Data Products and Vendors at WRDS." *WRDS*. https://wrds-www.wharton.upenn.edu/pages/get-data/.