



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS

Master's Thesis Presentation

Brian Park

Department of Statistics
The University of Chicago

“Separating Predictive Gains from Incremental
Information: Text Embeddings and Analyst
Forecasts of Corporate Earnings”

July 22, 2026, at 10:30 AM
Jones 111, 5747 S. Ellis Avenue

Abstract

Using 54,700 firm-years that match the Management's Discussion and Analysis (MD&A) section of 10-K filings from 1994 to 2024 to the IBES analyst consensus, we ask whether mandated narrative disclosure carries forward-looking information about next-year earnings per share beyond what the consensus already reflects, and whether any forecasting gain is associated with abnormal returns. Combining MD&A embeddings with the consensus improves out-of-sample forecasts: R^2 rises from 0.405 to 0.497 (OpenAI) and 0.479 (FinBERT), and directional AUC from 0.705 to 0.715 and 0.721, gains that remain significant after Bonferroni correction. The evidence suggests, however, that the improvement does not reflect information the consensus lacks. Text predicts neither the consensus forecast error, nor the error of a consensus recorded before the filing, nor the revision of the consensus around the filing, and the consensus barely moves when the 10-K arrives. Long/short portfolios formed on the text-augmented forecast earn no risk-adjusted return distinguishable from that of consensus-sorted portfolios, in any holding window we examine; an apparent premium in the first post-filing months reflects sample composition. Mandated narrative disclosure can thus yield large predictive gains without conveying incremental information or generating incremental returns.