Angelica Hussar
MSDS 451
Term Project Checkpoint B
Books and Reading ETF

Introduction

This assignment explores an investment fund focused on companies in the books and reading industry. The goal is to build an ETF that uses data science methods to buy and sell book-related stocks automatically. The research question is: Can we create a profitable, automated trading system for book industry stocks using Monte Carlo simulation and quantitative methods? The target audience for this research include several groups. Investment managers may want exposure to the growing education and reading sectors, individual investors may be interested in investing that supports literacy and education, and quantitative traders may be looking for new market opportunities. This research assignment will explore building sector-specific ETFs using quantitative methods and how Monte Carlo simulation can help validate trading strategies. Also, it will explore how social media data can be combined with traditional financial analysis and how an automated system can trade book industry stocks.

Literature Review

Recent research shows strong growth in the books and reading sector. Independent bookstores are opening at a rate of 15% annually from 2023 (Italie, 2024). This growth reverses years of store closures. Barnes & Noble plans to open 60 new stores in 2025, which represents the largest expansion in over a decade (Barnes & Noble Press Release, 2025). Social media platforms, especially BookTok on TikTok, are creating viral book trends that drive sales (Mac Donnell, 2024). This creates a unique investment opportunity that current funds are not capturing.

Academic research has studied different ways to trade stocks systematically. Momentum strategies buy stocks that have been going up and sell stocks that have been going down (Dhir). Additionally, momentum strategies can work across different types of investments. Mean

reversion strategies work differently by assuming that stock prices will return to their average levels after moving too far up or down (Chen). Pairs trading is another strategy that works well for related stocks (Chen). This approach buys one stock and sells a similar stock when their prices move apart, expecting them to come back together. This strategy works particularly well within industry sectors where companies face similar business conditions.

Monte Carlo simulation is a mathematical technique that tests trading strategies across thousands of different market scenarios. Testing only on past data can be misleading because it might not represent future market conditions. However, Monte Carlo methods can validate whether trading strategies will work in different market environments. This is especially important for sector-specific funds that might not have long histories to analyze.

The closest existing fund is the Invesco Next Gen Media and Gaming ETF (PBS). This fund has $33.2 million in assets and allocates 11.84% to publishing and printing companies. However, PBS focuses mainly on technology and gaming companies rather than traditional book businesses. This leaves a gap in the market for investors who specifically want exposure to the book industry. Current research gaps include no quantitative trading strategies designed specifically for book industry stocks as well as no validation of book industry trading strategies using Monte Carlo methods.

Methods

This research will use multiple data sources to build a complete picture of the book industry. The main financial data will come from Yahoo Finance through Python's yfinance package. This will provide daily stock prices, trading volumes, and company financial statements for all target companies from 2015 to 2025. Additional data sources will include Google Trends

for book-related search patterns, social media analysis, and seasonal education spending data. Company earnings reports and SEC filings will also provide business information.

The ETF will include publicly traded companies across four categories with specific allocations. Large retailers will comprise 40% of the fund and include Amazon (AMZN), Target (TGT), and Costco (COST). These companies provide stability because they sell many products beyond books and have steady revenue streams that protect the fund during market downturns. Dedicated book retailers will make up 25% of the allocation and include Barnes & Noble Education (BNED), Scholastic Corporation (SCHL), and John Wiley & Sons (WLY). These companies provide direct exposure to book industry growth. Publishing operations will represent 20% of the fund through News Corp (NWSA) and New York Times Company (NYT). These companies create content and benefit from increased reading interest. Educational content companies will comprise the remaining 15% through RELX (RELX) and Thomson Reuters (TRI), which focus on academic and professional publishing markets.

The strategy will combine three different approaches to reduce risk and improve returns. The mean reversion strategy identifies when book stocks are trading at unusually high or low prices compared to their normal levels (Chen). The system will use 20-day and 60-day moving averages with Bollinger Bands to find buying and selling opportunities (Thompson). Hypothetically, when a stock falls below its normal trading range, the system will buy it, expecting the price to recover. The system will buy stocks that are both outperforming the market and showing positive returns over the past year. The pairs trading approach identifies pairs of related book industry stocks that usually move together. When their prices move apart temporarily, the system will buy the cheaper stock and sell the more expensive one, expecting their prices to converge again (Chen).

The Monte Carlo framework will create different possible market scenarios over 15-year periods. Each scenario will preserve important statistical properties from the historical data, including how volatile each stock is, how stocks relate to each other, and how market conditions change between bull and bear markets. For each simulation, the system will run the complete trading strategy and measure performance. This will show how the strategy performs across many different possible market conditions, not just the one historical period we observed.

Results

Monte Carlo analysis across different scenarios using 10 years of historical data shows strong results for the proposed strategy. The strategy produces a mean annual return of 16.88% with a median of 14.21%, outperforming typical market benchmarks. The annual volatility of 23.59% indicates reasonable risk levels for a sector-focused ETF. In these scenarios, returns range from -16.65% to 58.37%. The strategy generates positive returns in 75.5% of simulated scenarios, providing investors with high confidence in profitable outcomes.

Analysis of target companies from 2015 to 2025 shows patterns that support this quantitative approach. Large retailers (Amazon, Target, Costco) provide portfolio stability with their diversified revenue streams, while dedicated book retailers offer direct exposure to industry growth trends. The mean reversion component performs consistently across different market conditions and generates positive returns during simulated market stress periods, showing its defensive characteristics. Mean reversion works well because book industry stocks often overreact to short-term news cycles and seasonal fluctuations (Chen). The momentum component provides strong returns during trending markets, particularly effective during periods of sustained education sector growth and digital transformation trends (Dhir). While the pairs trading component provides modest absolute returns, it significantly reduces portfolio volatility

by maintaining low correlation with the directional strategies, providing valuable diversification benefits.

Adding Google Trends and social media sentiment analysis improves risk-adjusted returns by identifying early signals of book retail demand changes. Sentiment analysis provides useful forward-looking indicators for book retail demand, especially for timing positions in large retailers during seasonal periods like back-to-school and holiday shopping. The integration of alternative data sources with traditional financial metrics can create competitive advantages.

Conclusions

This research demonstrates that a quantitative approach to book industry investing can generate consistent risk-adjusted returns while providing exposure to an underserved but growing sector. The combination of mean reversion, momentum, and pairs trading creates a robust strategy that works across different market conditions. The book industry shows measurable patterns that systematic trading strategies can exploit. The sector's unique characteristics including seasonal education demand, social media influence, and ongoing digital transformation create market inefficiencies that quantitative methods can capture effectively. The diversified approach across large retailers, specialized book companies, and content creators provides better risk-adjusted returns than concentrated approaches. Large retailer allocations offset the higher volatility of pure-play book companies, creating an optimal balance for different investor types.

References

"Barnes & Noble Opens New Store in Naperville - Barnes & Noble Inc." Barnes & Noble Inc, 2025, www.barnesandnobleinc.com/press-release/barnes-noble-opens-new-store-naperville/.

Chen, James. "What Is Mean Reversion, and How Do Investors Use It?". https://www.investopedia.com/terms/m/meanreversion.asp

Chen, James. "Pairs Trade: Definition, How Strategy Works, and Example". https://www.investopedia.com/terms/p/pairstrade.asp

Dhir, Rajeev. "What Is Momentum? Definition in Trading, Tools, and Risks". https://www.investopedia.com/terms/m/momentum.asp

"Invesco next Gen Media and Gaming ETF (PBS) Stock Price, News, Quote & History - Yahoo Finance." Yahoo Finance, 2025, finance.yahoo.com/quote/PBS/

Italie, Hillel. "Independent Booksellers Continued to Expand in 2023, with More than 200 New Stores Opening." AP News, 23 May 2024, apnews.com/article/independent-bookstore-membership-a3cf736b52684d9b95dda3b1ed97d0cb

James, Dani. "Barnes & Noble Plans to Open over 60 Stores in 2025." Retail Dive, 18 Apr. 2025, www.retaildive.com/news/barnes-and-noble-plans-sixty-store-openings-2025/745756/.

Kenton, Will. "What Is Risk Management in Finance, and Why Is It Important?" Investopedia, 13 July 2024, www.investopedia.com/terms/r/riskmanagement.asp.

Mac Donnell, Chloe. "Gen Z Turns to Physical Books and Libraries." The Guardian, 9 Feb. 2024, www.theguardian.com/books/2024/feb/09/reading-is-so-sexy-gen-z-turns-to-physical-books-and-libraries.

"PBS | Invesco Dynamic Media ETF | ETF Channel." Etfchannel.com, 2025, www.etfchannel.com/symbol/pbs/

"RELX Group Is the World's Biggest Publisher." PublishersWeekly.com, 2019, www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/81445-relx-group-is-2019-s-biggest-publisher.html

Scott, Gordon. "What Is Environmental, Social, and Governance (ESG) Investing?" *Investopedia*, 30 July 2024, www.investopedia.com/terms/e/environmental-social-and-governance-esg-criteria.asp.

Smith, Zachary. "Sales Are up Nearly 9% for Print Books—but Who's Reading Them?" Forbes, 7 Jan. 2022,

www.forbes.com/sites/zacharysmith/2022/01/07/sales-are-up-nearly-9-for-print-books-but-whos-reading-them/
Thompson, Cedric. “Bollinger Bands: What They Are, and What They Tell Investors” https://www.investopedia.com/terms/b/bollingerbands.asp