Predictive Modeling of Equity Settlement Failures: A Quantitative Research Framework

Section 1: The Microstructure of Settlement Failures

1.1 The U.S. Equity Settlement Lifecycle: From Trade to Finality

To construct a predictive model for settlement failures, a granular understanding of the U.S. equity settlement lifecycle is a prerequisite. A failure is not an isolated event but the culmination of a breakdown within a complex, high-speed, multi-stage process. This process, recently compressed by the transition to a one-day settlement cycle (T+1) on May 28, 2024, involves several key entities and systems that collectively ensure the orderly transfer of securities and funds.

The lifecycle begins at the point of **Trade Execution and Capture**. When a trade is executed on an exchange like the New York Stock Exchange (NYSE) or Nasdaq, its details are transmitted to the National Securities Clearing Corporation (NSCC) through the Universal Trade Capture (UTC) system. The UTC system serves as a central hub, validating and reporting equity transactions from various market centers throughout the trading day.

Once a trade is captured and validated, the NSCC, a subsidiary of the Depository Trust & Clearing Corporation (DTCC), steps in to act as the **Central Counterparty (CCP)**. Through a process known as novation, the NSCC interposes itself between the original buyer and seller, becoming the buyer to every seller and the seller to every buyer. This critical step mitigates counterparty risk; the original trading parties are no longer exposed to each other's potential default, but rather to the NSCC, which guarantees the completion of all matched trades under its rules.

The core of the NSCC's operational efficiency lies in the **Continuous Net Settlement (CNS) System**. Rather than settling each trade individually, the CNS system aggregates all of a clearing member's trades in a particular security for a given day and nets them down to a single net long (obligation to receive) or net short (obligation to deliver) position. This netting process dramatically reduces the number of security movements and the associated costs required for settlement. A crucial implication of this netting is that a Fail-to-Deliver (FTD) reported at the DTCC level is not attributable to a specific trade or trader. Instead, it represents a clearing member's failure to meet its *net* delivery obligation for that security on that day, an obligation that is the aggregate result of potentially thousands of individual long and short sales from its various clients.

The final stage of the process is managed by The Depository Trust Company (DTC), another DTCC subsidiary that functions as the **Central Securities Depository (CSD)**. The DTC holds the vast majority of U.S. equity securities in an immobilized, book-entry format, eliminating the need for physical certificate movement. Based on the final, netted instructions received from the NSCC's CNS system, the DTC processes the electronic transfer of securities between the accounts of its participating clearing members. The cycle achieves **Settlement and Finality** at the end of the settlement day (T+1), when the corresponding net cash obligations are settled through the Federal Reserve's National Settlement Service. This is the point at which the transfer of securities and funds becomes final and irrevocable. The recent transition to T+1 has compressed this entire intricate sequence of events, heightening the need for operational efficiency and increasing the potential for friction-related failures.

1.2 Anatomy of a Fail: Defining FTD and FTR

A settlement failure is formally defined as a situation where one party to a transaction does not meet its delivery obligation by the settlement date. A **Fail-to-Deliver (FTD)** occurs when the seller's clearing member does not deliver the required securities to the DTC's CNS system. The corresponding position for the buyer's clearing member, which does not receive the expected securities, is a **Fail-to-Receive (FTR)**. These are two sides of the same contractual breach. The causes of these failures are multifaceted and can be broadly categorized.

Operational and Administrative Failures: These are unintentional errors arising from friction within the settlement process. They can include simple human errors, such as incorrect data entry for a trade; technical problems within a broker's or the clearinghouse's systems; or delays in communication between a firm's front-office trading desk and its back-office operations team. While the dematerialization of securities has made them less common, delays in processing physical stock certificates can also contribute to such failures.

Long Sale Failures: FTDs are not exclusively the domain of short sellers. A failure can arise from a "long" sale—a sale of duly-owned stock. A common scenario is when an institutional investor sells shares that it has simultaneously lent out via the securities lending market. If the investor is unable to recall the loaned shares in time to meet the T+1 settlement deadline, a delivery failure will occur.

Short Sale Failures and "Naked" Shorting: A significant portion of FTDs originates from short sales. In a standard short sale, a trader borrows a security and sells it, hoping to buy it back later at a lower price. An FTD occurs if the short seller's broker fails to borrow the security in time for delivery. This category includes the highly scrutinized practice of "naked" short selling, where a security is sold without the seller having first borrowed or arranged to borrow it. While often portrayed as inherently manipulative, some naked short selling can occur due to legitimate market-making activities or unintentional delays in locating shares that a broker reasonably believed were available to borrow.

Strategic Failures: This category elevates the analysis of FTDs from a mere operational issue to a sophisticated economic decision. A strategic fail is a deliberate choice by a market participant, typically a short seller, to fail on a delivery obligation because the economic cost of failing is lower than the cost of settling. This decision is most often driven by conditions in the securities lending market. When a stock is "hard-to-borrow," the fee to borrow it can become prohibitively expensive. In such cases, particularly when rebate rates turn negative (meaning the lender pays less interest on the cash collateral than the prevailing risk-free rate), a rational short seller may find it more profitable to incur any penalties associated with failing to deliver rather than pay the high cost of borrowing. This transforms the FTD from a simple error into a powerful economic signal about scarcity and demand in the securities lending market.

1.3 The Economic Role and Market Impact of FTDs

The public discourse and regulatory focus on FTDs have often been dominated by a narrative of manipulation, where "naked" short sellers use fails to create "phantom shares" and artificially depress stock prices. However, a substantial body of academic research challenges this simplistic view, suggesting a far more nuanced and, at times, beneficial role for FTDs in market function.

The seminal study by Fotak, Raman, and Yadav (2014) provides a robust counter-narrative to the manipulation hypothesis. Analyzing 1,492 NYSE stocks from 2005 to 2008, the study found no evidence that FTDs caused price distortions or were responsible for the failure of major financial firms during the 2008 financial crisis. On the contrary, the research demonstrated that for crisis-stricken firms like Bear Stearns and Lehman Brothers, abnormally large spikes in FTDs occurred *after* major price collapses and the public dissemination of negative news, not before. This indicates that FTDs were a *response* to new, negative fundamental information, consistent with the price discovery role of informed short selling, rather than being the trigger for the price declines.

More strikingly, the study found that, on aggregate, increases in FTDs were associated with statistically significant *improvements* in market quality. An increase in the FTD ratio equivalent to 0.1% of a company's shares outstanding was found to lead to a 1.7% reduction in bid-ask spreads, a 0.2% decline in intraday volatility, and a 3% reduction in pricing errors. These results suggest that traders who fail to deliver are often acting as sophisticated value arbitrageurs and liquidity providers. When a stock is overpriced, these traders sell it short, contributing to pricing efficiency. When liquidity is scarce (i.e., bid-ask spreads are wide), they step in to provide it. The ability to fail on delivery is what enables them to perform these functions more effectively, especially when the formal borrowing market is constrained.

This leads to the **"Release Valve" Hypothesis**. The securities lending market, while crucial, is an over-the-counter (OTC) market that can be opaque and illiquid, particularly for hard-to-borrow securities. When short-selling demand for a stock spikes, the available lendable supply can be quickly exhausted, causing borrowing costs to soar. In this scenario, the ability to fail on a delivery serves as a critical release valve. It allows informed trading and liquidity provision to continue even when the formal lending market is seized up. Forcing all short-selling activity through a constrained lending channel by strictly banning FTDs can paradoxically harm market quality by creating severe frictions and raising costs for all short sellers.

This effect was empirically demonstrated during the SEC's temporary Emergency Order in July-August 2008, which mandated pre-borrowing for short sales in 19 financial stocks, effectively eliminating FTDs arising from short sales in those names. The research shows that during the period the order was in force, the affected securities experienced a significant *worsening* of market quality, with wider bid-ask spreads, higher intraday volatility, and larger pricing errors compared to a control group of unaffected financial firms. This natural experiment provides strong evidence that restricting the ability to fail can have predictably negative consequences on market liquidity and price discovery.

The FTD ratio, therefore, should not be viewed as a simple metric of operational inefficiency. It is a complex, latent variable that reflects the aggregate friction and strategic positioning within the interconnected ecosystem of securities trading, clearing, and lending. A high FTD rate can simultaneously signal operational bottlenecks, intense short-selling pressure on a security perceived to be overvalued, and a constrained, expensive securities lending market. A successful predictive model cannot treat FTDs as a monolithic event. It must use a rich feature set—incorporating variables like borrowing costs, short interest, and trading volume—to attempt to deconstruct this composite signal. Predicting the *likely driver* of an FTD is as important as predicting the FTD itself, as a strategic fail driven by high borrowing costs has vastly different market implications than an FTD caused by a random administrative error.

Furthermore, the recent market-wide transition to T+1 settlement represents the most significant structural break in the FTD data-generating process in years. This compression of the settlement cycle inherently increases the probability of operational and administrative failures, as all parties have half the time to perform post-trade processes like allocation, confirmation, and affirmation. This is particularly challenging for international market participants operating across different time zones. For strategic fails, the decision window for a short seller to either source a borrow or decide to fail is also halved, which may alter the economic calculus of that decision. Consequently, any predictive model trained predominantly on pre-T+1 data will likely face performance degradation and require significant recalibration. The T+1 transition date must be treated as a major structural break in any backtesting framework, as the relative importance of operational friction versus strategic intent as a driver for FTDs may have shifted fundamentally.

Section 2: Identifying Drivers of Settlement Failure Rates

2.1 Security-Level Characteristics Prone to High FTD Rates

Certain securities are inherently more susceptible to high rates of settlement failure due to their fundamental and market characteristics. Building a predictive model requires identifying and quantifying these attributes.

Market Capitalization: There is a strong inverse relationship between a company's size and its propensity for high FTD ratios. The analysis in Fotak et al. (2014) shows that firms in the highest decile of FTDs had an average market capitalization of $1.6 billion, whereas firms in the lowest decile averaged $9.1 billion. Smaller-cap stocks typically have lower liquidity, a smaller public float (the number of shares available for trading), and are often subject to greater speculative interest, making them more difficult to borrow and more likely targets for concentrated short-selling campaigns that can lead to delivery failures.

Liquidity and Trading Volume: The relationship between liquidity and FTDs is nuanced. On one hand, high aggregate trading volume is naturally correlated with a higher absolute number of FTDs, as more transactions create more opportunities for settlement friction. However, the more critical metric is the FTD rate relative to volume. Highly illiquid, thinly-traded securities are often cited by regulators as being particularly vulnerable to manipulative schemes involving naked short selling, which can result in persistent fails. The academic evidence suggests that FTDs themselves can act as a source of liquidity, narrowing spreads. Therefore, a model must consider both absolute volume and liquidity measures like the bid-ask spread to capture this complex dynamic.

Volatility: Securities with high price volatility tend to exhibit higher FTD rates. Volatility creates more opportunities for value arbitrage and speculative short selling, thus increasing the demand to borrow shares. Furthermore, rapid price movements can strain the inventories of market makers who are obligated to provide liquidity, potentially leading them to sell short without an immediate borrow and fail on delivery.

Short Interest and "Days-to-Cover": These are direct measures of short-selling pressure and are highly indicative of potential settlement issues. A high short interest ratio (the percentage of a company's shares sold short) signifies strong demand to borrow the stock, which strains the available lendable supply and increases both the cost of borrowing and the probability of fails. The "days-to-cover" ratio, calculated by dividing the number of shares sold short by the average daily trading volume, measures how long it would take for all short positions to be closed. A high

days-to-cover ratio (e.g., over 8-10 days) suggests a crowded short trade that is difficult to exit, making it more susceptible to "short squeezes" and delivery failures as borrowers struggle to locate shares to close their positions.
Regulation SHO Threshold List: A security's inclusion on the Regulation SHO Threshold Security List is a definitive red flag for persistent settlement failures. The criteria for inclusion are precise: a security must have aggregate FTDs at a registered clearing agency of 10,000 shares or more for five consecutive settlement days, and this level of fails must represent at least 0.5% of the issuer's total shares outstanding. Once a security is placed on this list, it becomes a "threshold security," which triggers stricter mandatory close-out requirements under Rule 204 of Regulation SHO. Broker-dealers must immediately purchase shares to close out any fail-to-deliver position in a threshold security that has persisted for 13 consecutive settlement days. This regulatory designation is therefore a critical feature for any predictive model.

2.2 The Securities Lending Market Nexus

The securities lending market is the arena where the supply of and demand for borrowable shares meet, and its dynamics are arguably the most powerful predictors of strategic FTDs. This OTC market is where institutions lend their securities to borrowers (typically hedge funds and market makers) in exchange for a fee and collateral.
Stock Borrow Cost & Rebate Rates: The price of borrowing a stock is the most direct driver of strategic fails. This price can be quoted in two ways. For "hard-to-borrow" or "special" stocks, it is an explicit annualized fee (e.g., 5% per year) paid by the borrower. For "easy-to-borrow" or "general collateral" stocks, the transaction is structured around a rebate rate. The borrower posts cash collateral (e.g., 102% of the stock's value) and receives interest on this cash from the lender. The rebate rate is the interest rate paid. If the rebate rate is below the prevailing risk-free rate (like the Fed Funds rate), the difference represents the implicit cost of the borrow. When demand for a stock is extremely high, the rebate rate can become negative, meaning the borrower earns less than zero interest on their collateral—a direct and often substantial cost. High explicit fees or negative rebates create a powerful economic incentive for a short seller to choose to fail delivery and avoid this cost.
Lendable Supply & Utilization: The supply side of the market is defined by the "lendable inventory" or "lendable supply"—the total pool of shares that institutional investors have made available for lending. A key metric derived from this is **Utilization**, defined as the number of shares on loan divided by the number of shares in the lendable inventory. A high utilization rate (e.g., approaching 100%) indicates that the available supply is nearly exhausted. This scarcity is a leading indicator of rising borrow costs and an increased probability of "fails" as new borrowers are unable to locate shares.
Data Providers: Given the OTC nature of this market, comprehensive data is not freely available. Commercial data providers are essential for gaining transparency. The dominant provider in this space is **S&P Global Market Intelligence**, which acquired IHS Markit, which had previously acquired Data Explorers. This firm aggregates data directly from the source—the systems of prime brokers, custodian banks, and asset managers—to provide a near-complete view of the market. Their datasets provide critical variables for modeling, including daily figures for lendable quantity, on-loan quantity, utilization rates, and various proprietary metrics for borrowing costs, such as the "Indicative Fee" and the "Daily Cost of Borrow Score (DCBS)," a 1-10 scale of borrow difficulty.

2.3 Counterparty Dynamics and Behavior

Ultimately, FTDs are the result of decisions and actions taken by specific market participants. Understanding the roles and incentives of these key players is crucial for building a predictive model.

Market Makers: These firms are critical liquidity providers, standing ready to buy and sell securities to maintain an orderly market. To perform this function, they must often sell shares they do not immediately own, creating short positions. Recognizing this vital role, regulations have historically provided market makers with exceptions from certain locate and close-out requirements for FTDs resulting from "bona fide" market-making activities. Academic research confirms that market makers behave strategically, choosing to fail to deliver when borrowing costs are high, and that this behavior can influence the pricing of related options contracts. Using the introduction of Rule 204T in late 2008—which effectively banned FTDs for most participants but maintained the market maker exception—Fotak et al. (2014) estimated that market-making activity was responsible for approximately 29% of all FTDs in the pre-Rule 204T era.

Prime Brokers: These divisions within large investment banks serve as the operational hub for hedge funds and other large institutional clients. Their bundled services include trade clearing and settlement, custody of assets, financing (leverage), and, most critically for this analysis, facilitating securities lending. A prime broker acts as an intermediary, sourcing hard-to-borrow securities from a network of institutional lenders to meet the short-selling needs of its hedge fund clients. The efficiency and risk appetite of a prime broker's securities lending desk directly determine whether its clients' short sales will settle on time or result in a fail. The DTCC's CNS Prime Broker Interface is a specialized mechanism that allows an executing broker to seamlessly hand off the settlement obligation of a hedge fund's trade to the fund's prime broker, centralizing the settlement risk and responsibility with that single entity.

Hedge Funds: As active and often aggressive market participants, hedge funds are a primary source of demand for short selling across a wide range of strategies, including long/short equity, statistical arbitrage, and event-driven investing. A hedge fund that establishes a large, leveraged short position in a small-cap, illiquid stock that is already heavily shorted is a prototypical generator of FTDs. The resulting fails could be unintentional, stemming from the prime broker's inability to source a sufficient quantity of shares to borrow, or they could be a deliberate, strategic decision to avoid exorbitant borrowing costs.

The placement of a security on the Regulation SHO Threshold List should be viewed as more than a simple lagging indicator of past failures; it is a dynamic event that can act as a forward-looking causal factor. The process begins when a security experiences persistent fails for five consecutive days, at which point it is flagged and added to the list. This initial listing is a reflection of past events. However, the designation triggers a significant change in the regulatory environment for that specific stock. Mandatory buy-in requirements under Rule 204 become much stricter for any new or existing fails. This regulatory shift dramatically increases the cost and risk of maintaining a short position. Short sellers who were previously failing strategically are now faced with the high probability of a forced buy-in, compelling them to either locate shares in an already-tight lending market (likely at a very high cost) or buy shares in the open market to cover their positions. This forced buying from multiple short sellers can create a powerful, self-reinforcing feedback loop known as a "short squeeze," driving the stock price sharply higher, which in turn financially pressures even more short sellers to cover, amplifying the upward price movement. A predictive model should therefore not treat "OnThresholdList" as

a simple binary feature. More sophisticated features, such as the number of consecutive days a stock has been on the list or the rate of change of FTDs while on the list, could capture the escalating pressure and predict the likelihood of a subsequent price squeeze or a sudden resolution of the fail positions.
Furthermore, the highly concentrated nature of the prime brokerage industry introduces a systemic risk vector for settlement failures. A small number of large, global investment banks service the vast majority of hedge fund clients. These funds are deeply reliant on their prime brokers for the core functions of trade execution, clearing, and, critically, the sourcing of securities for borrowing. This concentration means that an idiosyncratic shock affecting a single major prime broker—such as a sudden credit downgrade, a significant operational failure, or a management decision to abruptly curtail its risk appetite—could have market-wide ramifications. Such an event could lead to a mass recall of loaned securities or an inability to source new borrows for a large swath of the hedge fund industry simultaneously. This would trigger a sudden, correlated demand shock in the securities lending market, causing borrow fees to spike across the board and potentially leading to widespread settlement failures. This cascade would not be limited to the clients of the distressed prime broker but would affect all participants competing for a now-scarcer pool of lendable shares. While data on the internal health of a specific prime broker is not public, a model could incorporate proxy variables to capture this systemic risk. For instance, the credit default swap (CDS) spread of a prime broker's parent bank holding company could serve as a real-time indicator of its perceived credit risk. A sharp, significant widening in the CDS spread of one or more major prime brokers could be a powerful predictive feature for an impending increase in market-wide FTD rates, shifting the analytical focus from the individual security level to the systemic, counterparty-risk level.

Section 3: A Methodological Framework for FTD Prediction

3.1 Data Architecture and Acquisition

The foundation of any robust quantitative model is a comprehensive and meticulously curated data architecture. Developing a predictive model for FTDs requires assembling a diverse array of datasets from multiple sources, each capturing a different facet of market activity.
FTD Data (Dependent Variable): The target variable for the model is the FTD level itself. This data is publicly available directly from the U.S. Securities and Exchange Commission (SEC). The SEC publishes files on a semi-monthly basis that contain the settlement date, CUSIP, ticker symbol, issuer name, closing price, and the total number of FTD shares for each security with a non-zero fail balance on that day. A critical data processing step involves accounting for the historical reporting threshold. Prior to September 16, 2008, the SEC only reported FTD figures for securities with a balance of 10,000 shares or more. For this earlier period, a non-reported value for a given security-day does not signify zero fails but rather a value between 0 and 9,999. This is a form of left-censored data, and treating these observations as true zeros would introduce significant bias into the model.
Securities Master and Corporate Actions Data: A reliable securities master file is essential for normalization and adjustment. Data providers such as the Center for Research in Security Prices (CRSP) or S&P Compustat are standard sources. This data provides the daily number of shares outstanding, which is necessary to convert the absolute number of FTD shares into a

normalized FTD_Ratio. It also provides CRSP share codes, allowing for the filtering of the sample to include only ordinary common shares (e.g., codes 10 and 11) and exclude ETFs, ADRs, and other security types that may have different settlement dynamics. Corporate actions data is also vital for creating accurate, continuous time series by adjusting historical price and volume data for events like stock splits, dividends, and mergers.

Market and Microstructure Data: High-frequency data is required to compute daily market quality and activity metrics. The NYSE's Trade and Quote (TAQ) database is the canonical source for academic research, providing tick-by-tick records of all trades and quotes on U.S. exchanges. From this data, one can construct daily time series for variables such as trading volume, intraday volatility (e.g., standard deviation of 5-minute returns), relative bid-ask spreads, and order imbalance. Commercial data vendors like FactSet and LSEG (formerly Refinitiv) also provide access to this data in a more processed format.

Securities Lending Data: This is the most critical proprietary dataset for modeling strategic fails. As the market is OTC, this data must be sourced from a specialized vendor that aggregates information from market participants. The leading provider is S&P Global Securities Finance. Their data provides daily security-level metrics on the supply (Lendable Quantity), demand (On Loan Quantity), and price (various fee metrics) in the lending market. Key variables include Utilization, Indicative Fee, and the Daily Cost of Borrow Score (DCBS), which are indispensable for capturing lending market tightness.

Short Interest Data: Official short interest figures are reported by the exchanges (e.g., NYSE, Nasdaq) but are released with a significant lag and only on a bi-weekly basis. While this data is a useful, albeit low-frequency, indicator of market sentiment, the higher-frequency data from securities lending providers can be used to construct more timely daily estimates of shorting activity.

The following table provides a consolidated view of the essential data architecture for this research project.

Table 1: Key Data Sources for FTD Modeling

Data Category	Specific Dataset	Provider(s)	Key Variables	Model Role
Settlement Fails	Fails-to-Deliver Data	SEC	Settlement Date, CUSIP, Ticker, FTD Shares	Dependent Variable
Security Master	CRSP Daily Stock File	CRSP / WRDS	Shares Outstanding, Share Code, Price, Return	Feature Engineering, Normalization
Microstructure	NYSE TAQ	NYSE	Trades, Quotes	Independent Variables (Volume, Volatility, Spread, Order Imbalance)
Securities Lending	Securities Finance Data	S&P Global	Lendable Quantity, On Loan Quantity, Indicative Fee, DCBS	Independent Variables
Short Interest	Exchange Short Interest	NYSE, Nasdaq, shortsqueeze.com	Short Volume, Total Shares Short	Independent Variables
Regulatory	Threshold Security List	NYSE, Nasdaq	CUSIP, Date Listed	Independent Variable

3.2 Defining and Engineering Predictive Features

Raw data is rarely in a form suitable for direct input into a predictive model. The process of feature engineering involves transforming these raw inputs into a set of informative, structured variables that capture the underlying drivers of FTDs.

Dependent Variable Construction: The raw FTD share count from the SEC must be transformed into a standardized target variable. Several formulations are possible, and the choice depends on whether the modeling goal is regression or classification.

- **FTD Ratio (for regression):** The most common dependent variable is the number of FTD shares normalized by the total shares outstanding. This creates a scale-free measure of failure intensity. $FTD\_Ratio_{i,t} = \frac{FTD\_Shares_{i,t}}{Shares\_Outstanding_{i,t}}$ where i indexes the security and t indexes the settlement date.
- **Change in FTD Ratio (for regression):** As demonstrated in the academic literature, the daily change in the FTD ratio is often more informative and has better statistical properties (e.g., stationarity) than the level itself. $\Delta FTD\_Ratio_{i,t} = FTD\_Ratio_{i,t} - FTD\_Ratio_{i,t-1}$
- **FTD Event (for classification):** The problem can be simplified to predicting whether a significant failure event will occur. This involves defining a threshold (τ) and creating a binary variable. $FTD\_Event_{i,t} = 1 \text{ if } FTD\_Ratio_{i,t} > \tau, \text{ else } 0$ The threshold τ could be a fixed value (e.g., 0.5%, mirroring the Reg SHO criteria) or a dynamic value (e.g., the 95th percentile of the FTD ratio distribution).

Feature Engineering: This stage involves creating a rich set of independent variables (predictors) from the source data.

- **Lags and Moving Averages:** The impact of market events and conditions on settlement is not instantaneous. Therefore, it is essential to create lagged versions of predictor variables (e.g., Volatility_t-1, BorrowCost_t-1). Furthermore, using rolling moving averages (e.g., 5-day or 20-day moving average of trading volume) can smooth out daily noise and capture underlying trends.
- **Interaction Terms:** The effect of one variable may depend on the level of another. For example, high utilization in the lending market is far more likely to lead to fails if the cost to borrow is also high. Creating interaction terms (e.g., Utilization * BorrowCost_DCBS) allows the model to capture these non-linear relationships.
- **Regime Indicators:** The regulatory and market environment has changed significantly over time. Creating binary dummy variables for distinct periods—such as pre/post-Regulation SHO (Jan 2005), the 2008 financial crisis, pre/post-Rule 204T (Sep 2008), and pre/post-T+1 settlement (May 2024)—is crucial. These variables allow the model to fit different parameters for each regime, accounting for structural breaks in the data-generating process.

The following table presents a comprehensive, though not exhaustive, library of potential features to be constructed and tested in the modeling process.

Table 2: A Universe of Potential Predictive Variables

Category	Variable Name	Definition	Rationale / Source Snippet
Market Microstructure	Volatility_5D	5-day rolling standard deviation of daily returns.	High volatility can strain liquidity and increase shorting demand.
	Spread_Rel_Avg	Daily average of (Ask	High spreads indicate

Category	Variable Name	Definition	Rationale / Source Snippet
		Price - Bid Price) / Midpoint Price.	illiquidity; FTDs are dynamically related to liquidity measures.
	OrderImbalance	(Dollar value of buyer-initiated trades - Dollar value of seller-initiated trades) / Total dollar volume.	Order imbalances can predict short-term price pressure and signal liquidity demand.
	Volume_Turnover_20D	20-day moving average of (Daily Shares Traded / Shares Outstanding).	High turnover can indicate heightened speculative interest and potential for settlement friction.
Securities Lending	Utilization	Shares On Loan / Shares in Lendable Inventory.	High utilization signals scarcity in the lending market, driving up borrow costs and fail likelihood.
	BorrowCost_DCBS	Daily Cost of Borrow Score (1-10 scale from S&P Global).	A direct, standardized measure of the cost to borrow; high cost incentivizes strategic fails.
	LendableSupply_Ratio	Shares in Lendable Inventory / Shares Outstanding.	Low relative supply indicates a smaller pool of shares available to meet borrowing demand.
Security/Sentiment	ShortInterest_Ratio	Total Shares Held Short / Shares Outstanding.	High short interest indicates high aggregate borrowing demand and negative sentiment.
	DaysToCover	Total Shares Held Short / 30-day Average Daily Volume.	A high value indicates a crowded short trade that is difficult to exit, increasing fail risk.
	MarketCap_Log	Natural logarithm of market capitalization.	Smaller firms are empirically shown to have higher FTD rates.
Regulatory	OnThresholdList	Binary variable (1 if the security is on the Reg SHO Threshold List on day t, 0 otherwise).	A direct indicator of persistent, ongoing settlement issues that trigger mandatory buy-ins.

Category	Variable Name	Definition	Rationale / Source Snippet
	DaysOnThreshold	A continuous count of consecutive days the security has been on the Threshold List.	Captures the duration and severity of the persistent fail condition.

Section 4: Constructing and Interpreting the Predictive Model

4.1 Econometric and Time-Series Approaches

Classical econometric and time-series models provide a strong foundation for analyzing FTDs, offering interpretability and a framework to test specific economic hypotheses. These methods are particularly well-suited for capturing the dynamic, interdependent nature of settlement failures and market quality metrics.

Vector Autoregressive (VAR) Models: As effectively demonstrated by Fotak et al. (2014), VAR models are a powerful tool for this problem. A VAR framework treats a set of variables as an interconnected system, where each variable is modeled as a function of its own past values (lags) and the past values of all other variables in the system. For FTD prediction, a VAR system could include variables such as \Delta FTD_Ratio, intraday volatility, relative spread, and order imbalance. This approach explicitly acknowledges and models the endogenous feedback loops present in the market: for example, an increase in FTDs may lead to a subsequent decrease in spreads (improved liquidity), while an initial widening of spreads (poor liquidity) might itself contribute to a higher likelihood of FTDs. By analyzing the system's impulse response functions, one can trace the dynamic impact of a shock to one variable (e.g., a sudden increase in FTDs) on all other variables over time.

GARCH (Generalized Autoregressive Conditional Heteroskedasticity) Models: A well-documented characteristic of financial time series is volatility clustering—periods of high volatility tend to be followed by more high volatility, and vice versa. The time series of FTDs likely exhibits similar behavior. GARCH models are designed specifically to capture this time-varying volatility. While a standard GARCH model focuses on predicting the conditional variance rather than the mean, it can be integrated into a broader modeling framework (e.g., an ARMA-GARCH model) to account for the fact that the uncertainty around FTD predictions is not constant over time.

ARIMA (Autoregressive Integrated Moving Average) Models: As a foundational time-series technique, an ARIMA model serves as an excellent baseline for comparison. It predicts future values of a series based on a linear combination of its own past values (the Autoregressive 'AR' component) and its past forecast errors (the Moving Average 'MA' component). A univariate ARIMA model applied to the FTD_Ratio series can capture its persistence and mean-reverting tendencies. More advanced versions, such as ARIMAX, can incorporate exogenous predictor variables, providing a bridge between simple time-series models and more complex multivariate systems.

4.2 Machine Learning for Classification and Prediction

While econometric models are strong in interpretability and handling system dynamics, machine

learning (ML) models excel at capturing complex, non-linear patterns in large, high-dimensional datasets. They are particularly effective when the underlying relationships between predictors and the target variable are not well-defined by economic theory.
Problem Formulation: The FTD prediction task can be framed in two primary ways for ML models:

1. **Regression:** The model is trained to predict a continuous value, such as the FTD_Ratio or ΔFTD_Ratio for the next settlement day.
2. **Classification:** The model is trained to predict a categorical outcome. This could be a binary classification task (e.g., predicting whether the FTD_Event will be 1 or 0) or a multi-class classification task (e.g., predicting whether the FTD level will be 'Low', 'Medium', or 'High'). Classification is often more robust when dealing with noisy financial data and rare events.

Model Selection:

- **Tree-Based Models:** Ensembles of decision trees, such as **Random Forests** and **Gradient Boosting Machines (GBMs)** like XGBoost and LightGBM, are often the state-of-the-art for tabular data problems in finance. They can automatically capture complex, non-linear interactions between features (e.g., the effect of high utilization is only present when borrow cost is also high) without requiring them to be specified in advance. They are also relatively robust to outliers and the scale of input features. RandomForestClassifier is a strong starting point, while GBMs often provide higher predictive accuracy with careful tuning.
- **Deep Learning:** For datasets of sufficient size and complexity, deep learning models can be employed. **Long Short-Term Memory (LSTM)** networks, a type of Recurrent Neural Network (RNN), are specifically designed to learn long-term dependencies in sequential data, making them a natural fit for time series forecasting. More recent architectures like Transformers, which use attention mechanisms, could also be applied to learn which past events are most relevant for predicting future FTDs.

The commercial application of these techniques is already underway. For instance, the clearinghouse Clearstream has launched an AI-powered tool to predict settlement fails for its clients, demonstrating the industry's recognition of the value of ML in this domain.

4.3 From Prediction to Insight: Model Interpretation

A highly accurate "black box" model has limited practical utility if its decision-making process is opaque. For risk management and strategy development, understanding *why* a model is making a particular prediction is as important as the prediction itself.
Feature Importance: For tree-based models, it is straightforward to calculate global feature importance scores. Metrics like "Gini Importance" or "Permutation Importance" provide a ranking of all input variables based on their overall contribution to the model's predictive power. This analysis can confirm which drivers—such as borrow cost, utilization, and market cap—are most influential in predicting FTDs.
SHAP (SHapley Additive exPlanations): To move beyond global rankings and understand individual predictions, SHAP is a powerful, model-agnostic technique rooted in cooperative game theory. For any single prediction (e.g., a high FTD forecast for stock XYZ on a specific day), SHAP values decompose the prediction, showing how much each individual feature value (e.g., BorrowCost_DCBS = 9, MarketCap_Log = 19.5) contributed to pushing the forecast away from the baseline average. This allows an analyst to construct a narrative for each prediction, such as: "The model predicts a high FTD risk for XYZ primarily because of its extremely high

borrow cost and its recent inclusion on the Threshold List, despite its high liquidity." This level of interpretability is crucial for building trust in the model and integrating its outputs into a decision-making workflow.
The following table provides a high-level comparison to guide the selection of an appropriate modeling technique.

Table 3: A Comparative Analysis of Modeling Techniques

Model Class	Key Strengths	Key Weaknesses	Best Use Case
VAR	Captures feedback loops and system dynamics; strong theoretical grounding.	Assumes linearity; can become unwieldy and difficult to interpret with many variables.	Modeling the dynamic, interdependent system of FTDs and market quality metrics.
GARCH	Explicitly models the time-varying nature of volatility (volatility clustering).	Focuses on the conditional variance of the series, not the prediction of the mean level of FTDs itself.	As a component within a larger model to handle heteroskedasticity in the FTD time series.
Random Forest	Robust to outliers; effectively captures non-linearities; provides a strong baseline performance.	Can be less accurate than gradient boosting; interpretation can be more complex than linear models.	An excellent initial ML model for establishing a performance baseline and assessing feature importance.
Gradient Boosting	Often achieves state-of-the-art predictive performance; handles complex interactions and large feature sets.	Highly prone to overfitting if not carefully tuned; can be a "black box" without interpretability tools.	The primary candidate for the final predictive model where accuracy is the paramount goal.

Section 5: Robust Model Validation and Backtesting

5.1 The Principles of Financial Backtesting

A predictive model's true value is not its performance on data it has already seen, but its ability to generalize and make accurate predictions on new, unseen data. Backtesting is the disciplined process of simulating a model's historical performance to estimate its future efficacy. A rigorous backtesting framework is non-negotiable and must adhere to several core principles.
Avoiding Overfitting: This is the most critical challenge in financial modeling. Overfitting occurs when a model becomes too complex and learns the random noise and idiosyncratic patterns in the training data, rather than the true underlying signal. Such a model will exhibit excellent performance on the historical data it was trained on but will fail dramatically when deployed on new data. Rigorous out-of-sample testing is the primary defense against this pitfall.
Avoiding Lookahead Bias: A backtest must strictly simulate real-world conditions, meaning that at any point in historical time *t*, the model can only use information that would have been genuinely available to a market participant at or before time *t*. For FTD modeling, this is particularly subtle. The SEC releases FTD data for a given settlement date *t* several days later (on a semi-monthly schedule). Therefore, a model predicting the FTD level for date *t* cannot use the actual FTD data from date *t* as a feature. All input data must be correctly time-stamped to reflect its actual availability to the market.

Accounting for Transaction and Borrowing Costs: While the primary goal is prediction, if the model's output is intended to inform a trading strategy (e.g., shorting stocks with a high predicted FTD rate), the backtest of that strategy must incorporate realistic costs. This includes trading commissions, bid-ask spreads (slippage), and, most importantly, the cost of borrowing the security. A strategy that appears profitable before costs can easily become unprofitable once these real-world frictions are included.

5.2 In-Sample vs. Out-of-Sample Validation

The cornerstone of a valid backtest is the strict separation of data used for model development (in-sample) from data used for final evaluation (out-of-sample). Evidence based on out-of-sample performance is considered far more trustworthy and better reflects the real-time forecasting challenge.
Data Splitting: The most basic approach is to partition the historical dataset chronologically. For example, if the dataset spans from 2005 to 2023, the period from 2005-2019 could be designated as the **in-sample** or **training set**, used for model fitting, feature selection, and parameter tuning. The period from 2020-2023 would be held aside as the **out-of-sample** or **test set**. The model is trained only once on the in-sample data, and its final performance is judged solely on its predictions on the untouched out-of-sample data.
Walk-Forward Analysis: For time-series data, a more robust and realistic validation technique is walk-forward analysis (or rolling-window backtesting). This method better simulates how a model would be periodically retrained and deployed in a live environment. The process is as follows :

1. Train the model on an initial window of data (e.g., 2005-2010).
2. Use the trained model to make predictions for the next period (e.g., the year 2011).
3. Record the performance for 2011.
4. Roll the window forward: add the 2011 data to the training set (now 2005-2011) and retrain the model.
5. Use the newly trained model to make predictions for 2012.
6. Repeat this process until the end of the dataset. The final performance is the aggregated result from all the out-of-sample prediction periods. This method tests the model's stability and adaptability over time.

Cross-Validation: For tuning the hyperparameters of ML models (e.g., the depth of trees in a Random Forest), k-fold cross-validation should be used exclusively on the in-sample training data. This involves splitting the training data into 'k' folds, training the model on k-1 folds, and validating on the held-out fold, rotating until each fold has been used for validation. This allows for hyperparameter selection without "contaminating" the final, out-of-sample test set.

5.3 Stress Testing and Structural Breaks

A model's average performance is important, but its behavior during periods of market stress and structural change is often what truly determines its reliability. The backtesting process must explicitly test the model's robustness under adverse conditions.
Identifying Structural Breaks: The historical data for FTDs is not stationary; its underlying statistical properties have changed over time due to significant events. A model that ignores these breaks is likely to be misspecified and produce biased inferences. Key structural breaks that must be accounted for in the analysis include:

- The introduction of Regulation SHO (January 2005).

- The global financial crisis (2008-2009).
- The implementation of SEC Rule 204T, which drastically curtailed non-market-maker FTDs (September 2008).
- The recent transition to T+1 settlement (May 2024).

Regime-Specific Backtesting: The model's performance should be analyzed separately across different market regimes. For example, the backtest results should be segmented into periods of high market volatility (e.g., 2008, 2020) versus low volatility. A robust model should either maintain stable performance across regimes or exhibit performance changes that are explainable and consistent with economic intuition. If a model trained on pre-2008 data fails completely in the post-2008 environment, it indicates that it has learned spurious correlations rather than fundamental, enduring relationships.

The Basel Framework for Backtesting: While originally designed for evaluating Value-at-Risk (VaR) models in banking, the conceptual framework from the Basel Committee on Banking Supervision is highly relevant. It involves counting the number of "exceptions"—instances where the actual outcome exceeded the model's forecast. Based on the number of exceptions over a given period (e.g., 250 trading days), a model's performance is classified into zones: a "green zone" (acceptable performance), an "orange zone" (warrants scrutiny), or a "red zone" (model is likely flawed). This provides a structured, statistically-grounded approach for monitoring model performance and triggering a formal review or recalibration when necessary.

To quantitatively evaluate the model's performance during backtesting, a clear set of metrics must be defined. The choice of metrics depends on whether the problem is framed as classification or regression.

Table 4: Performance Metrics for FTD Prediction Models

Task Type	Metric	Formula / Definition	Interpretation
Classification	Precision	\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}	Of all the instances the model predicted as a high-FTD event, what fraction actually were? Measures the accuracy of positive predictions.
	Recall (Sensitivity)	\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}	Of all the actual high-FTD events that occurred, what fraction did the model successfully identify? Measures completeness.
	F1-Score	2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}	The harmonic mean of Precision and Recall. It is a crucial metric for imbalanced datasets where simply maximizing accuracy is misleading.
	AUC-ROC	Area Under the Receiver Operating Characteristic Curve	Measures the model's overall ability to discriminate between

Task Type	Metric	Formula / Definition	Interpretation
			the positive and negative classes across all possible classification thresholds.
Regression	RMSE	Root Mean Squared Error: $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	The standard deviation of the prediction errors (residuals). It heavily penalizes large errors.
	MAE	Mean Absolute Error: $\frac{1}{n}\sum_{i=1}^{n}$	$y_i - \hat{y}_i$
	R-squared	Coefficient of Determination	The proportion of the variance in the FTD ratio that is predictable from the independent variables.

Section 6: Conclusion and Avenues for Further Research

6.1 Synthesis of Findings

This report has outlined a comprehensive, data-driven framework for developing, validating, and interpreting a mathematical model to predict Fails-to-Deliver in U.S. equity markets. The analysis reveals that FTDs are not merely operational noise but are a complex signal reflecting the interplay of market microstructure, counterparty behavior, and, most critically, frictions within the securities lending market.

The key determinants of high FTD rates are multifaceted. At the security level, low market capitalization, high price volatility, illiquidity, and significant short interest are strong indicators of elevated risk. However, the most potent predictive information comes from the securities lending market. Metrics such as high utilization of lendable shares and high borrowing costs (or negative rebate rates) provide a direct economic incentive for market participants to engage in "strategic fails," transforming the FTD from a passive error into an active, information-rich decision. Counterparty analysis further reveals the distinct roles of market makers, who use fails to provide liquidity, and prime brokers, who act as the central nervous system for hedge fund short-selling activity.

For modeling, a hybrid approach is recommended. Vector Autoregressive (VAR) models are well-suited for capturing the systemic, endogenous feedback loops between FTDs and market quality variables like spreads and volatility. For achieving maximum predictive accuracy, however, non-linear machine learning techniques such as Gradient Boosting Machines are likely to be superior, capable of identifying complex interactions within a large feature set. Crucially, any model must be subjected to a rigorous backtesting protocol that includes walk-forward analysis and stress testing across different market regimes and structural breaks, such as the recent shift to a T+1 settlement cycle.

6.2 Practical Implications and Recommendations

The output of a well-validated FTD prediction model has significant practical applications for various market participants:

- **For Risk Managers:** The model can function as a sophisticated early warning system. By screening for securities with a high predicted probability of settlement failure, risk departments can proactively manage counterparty exposure, adjust collateral requirements, and anticipate potential liquidity drains associated with settlement gridlock.
- **For Traders and Portfolio Managers:** The model's predictions can be a valuable input for alpha-generating strategies. A high predicted FTD rate, when decomposed using interpretability tools like SHAP and identified as likely "strategic" (i.e., driven by high borrow costs and shorting demand), serves as a strong, quantifiable bearish signal on a stock's future performance. Conversely, identifying a security that is on the Reg SHO Threshold List with escalating FTDs can signal the potential for a short squeeze, offering a tactical long opportunity.
- **For Compliance and Operations:** With settlement discipline regimes like Europe's CSDR imposing direct financial penalties for settlement fails, the cost of operational inefficiency is rising. A predictive model can help operations teams anticipate and allocate resources to trades that are at high risk of failing, allowing them to intervene proactively to secure borrows or resolve issues before the settlement deadline, thereby avoiding penalties and reducing operational overhead.

6.3 Future Directions

While the framework presented here is comprehensive, several advanced avenues exist for future research that could further enhance the predictive power and utility of FTD models.

- **Network Analysis of FTDs:** Settlement failures do not occur in isolation. A single large fail can trigger a cascade or "daisy chain" of subsequent fails as the receiving party, now short of expected securities, fails on its own onward delivery obligations. Advanced research could employ network graph theory to model the financial system as a network of counterparty obligations. This would allow for the analysis of contagion risk and the identification of systemically important participants whose failure could trigger a market-wide settlement crisis.
- **Integration of Options Market Data:** The equity options market is deeply intertwined with the underlying stock's settlement and lending dynamics. Options market makers constantly hedge their positions by buying and selling the underlying stock. This hedging activity is a major source of trading volume and borrowing demand, and market makers often use FTDs as a tool to manage their inventory, especially when borrowing costs are high. Incorporating features from the options market—such as put/call open interest ratios, implied volatility surfaces, and the trading activity of options market makers—could significantly improve the model's ability to predict FTDs.
- **Intraday Predictive Modeling:** The current framework is based on a daily prediction cycle. However, as data providers like S&P Global begin to offer higher-frequency, intraday securities lending data , the next frontier will be the development of real-time or near-real-time FTD prediction models. Such a model could update its risk assessment throughout the trading day, providing traders and risk managers with far more timely and actionable intelligence to manage their settlement exposures before the end-of-day cutoff.

Works cited

1. Trade Settlement: Know Your T+1 Blind Spots - DTCC, https://www.dtcc.com/dtcc-connection/articles/2024/april/29/trade-settlement-know-your-t1-blind-spots 2. Understanding Settlement Cycles: What Does T+1 Mean for You? | FINRA.org, https://www.finra.org/investors/insights/understanding-settlement-cycles 3. 8 Things to Know About T+1 Settlement | Charles Schwab, https://www.schwab.com/learn/story/7-things-to-know-about-t1-settlement 4. Equities Clearing - DTCC Learning Center, https://dtcclearning.com/products-and-services/equities-clearing.html 5. Efficient Netting & Settlement with CNS - DTCC, https://www.dtcc.com/clearing-and-settlement-services/equities-clearing-services/cns 6. Understanding the DTCC Subsidiaries Settlement Process | DTCC, https://www.dtcc.com/understanding-settlement/index.html 7. Settlement Period: Definition, Process, and SEC Rules - Investopedia, https://www.investopedia.com/terms/s/settlement_period.asp 8. Settlement Service Guide - DTCC, https://www.dtcc.com/globals/pdfs/2018/february/27/service-guide-settlement 9. Tackling Post-Trade Friction - Supporting a Global Shortened Settlement Cycle - Clearstream, https://www.clearstream.com/resource/blob/4506066/6a111fd83f89b7a196723b331cd9b900/tackling-post-trade-friction-t-1-data.pdf 10. What Is Failure to Deliver, and What Happens With FTDs? - Investopedia, https://www.investopedia.com/terms/f/failuretodeliver.asp 11. Failure to deliver - Wikipedia, https://en.wikipedia.org/wiki/Failure_to_deliver 12. Fail to Deliver in the Stock Market - CEOBLOC, https://ceobloc.com/blog/fail-to-deliver-in-the-stock-market 13. Fails to Deliver: A Comprehensive Guide - Number Analytics, https://www.numberanalytics.com/blog/ultimate-guide-fails-to-deliver-securities-law 14. All About Failing Trades | Nasdaq, https://www.nasdaq.com/articles/all-about-failing-trades-2021-04-08 15. Volume of fails to deliver and daily trade volume for stocks listed on... - ResearchGate, https://www.researchgate.net/figure/olume-of-fails-to-deliver-and-daily-trade-volume-for-stocks-listed-on-the-New-York-Stock_fig1_228260887 16. Bangers, Bashers & Short Sellers I Regulation SHO - Securities Lawyer 101, https://www.securitieslawyer101.com/2020/short-seller-regulation-sho/ 17. Fail to Deliver - IBKR Campus, https://www.interactivebrokers.com/campus/glossary-terms/fail-to-deliver/ 18. GAO-09-483, Regulation SHO: Recent Actions Appear to Have Initially Reduced Failures to Deliver, but More Industry Guidance Is Needed, https://www.gao.gov/assets/a289483.html 19. What Short Interest Tells Us - Investopedia, https://www.investopedia.com/articles/01/082201.asp 20. Staff Working Paper No. 821 - Securities settlement fails network and buy-in strategies, https://www.icmagroup.org/assets/documents/Regulatory/Secondary-markets/Bond-Market-Liquidity-Library/BoESeptember-2019securities-settlement-fails-network-and-buy-in-strategies-230919.pdf 21. Failure Is an Option: Impediments to Short Selling and Options Prices | Request PDF - ResearchGate, https://www.researchgate.net/publication/272242625_Failure_Is_an_Option_Impediments_to_Short_Selling_and_Options_Prices 22. Over-the-Counter Market Liquidity and Securities Lending, https://www.icmagroup.org/assets/documents/Regulatory/Secondary-markets/Bond-Market-Liquidity-Library/BISwork768OTC-market-liquidity-and-securities-lendingFeb-2019-260219.pdf 23. SEC Requires T+1 Settlement Cycle | Insights & Resources - Goodwin, https://www.goodwinlaw.com/en/insights/publications/2023/02/02_17-sec-requires-t1-settlement

-cycle 24. Market Capitalization: What It Means for Investors - Investopedia, https://www.investopedia.com/terms/m/marketcapitalization.asp 25. Understanding high-risk investments | FCA, https://www.fca.org.uk/investsmart/understanding-high-risk-investments 26. (PDF) On the relationship between trading volume and stock price volatility in CASE, https://www.researchgate.net/publication/46545462_On_the_relationship_between_trading_volume_and_stock_price_volatility_in_CASE 27. The Relationship between Trading Volume, Volatility and Stock Market Returns - Business Perspectives, https://www.businessperspectives.org/images/pdf/applications/publishing/templates/article/assets/1266/imfi_en_2005_03_Ahmed.pdf 28. Market microstructure - Wikipedia, https://en.wikipedia.org/wiki/Market_microstructure 29. What Is the Threshold Securities List? Definition and Criteria, https://www.investopedia.com/terms/t/thresholdlist.asp 30. Regulation SHO: Definition, What It Regulates, and Requirements - Investopedia, https://www.investopedia.com/terms/r/regsho.asp 31. The Essentials - Reg SHO Compliance | Oyster Consulting, https://www.oysterllc.com/what-we-think/the-essentials-reg-sho-compliance/ 32. Division of Market Regulation - SEC.gov, https://www.sec.gov/rules-regulations/staff-guidance/trading-markets-frequently-asked-questions-8 33. What Is a Stock Loan Fee (Borrow Fee)? Definition and Example - Investopedia, https://www.investopedia.com/terms/s/stock-loan-fee.asp 34. Hard-To-Borrow List: What it Means, How it Works - Investopedia, https://www.investopedia.com/terms/h/hardtoborrowlist.asp 35. In Short Supply: Short-Sellers and Stock Returns - Wharton Accounting, https://accounting.wharton.upenn.edu/wp-content/uploads/2015/04/BeneishLeeNichols.pdf 36. Thresholds in Securities Lending Metrics, https://cdn.ihsmarkit.com/www/pdf/0321/MSF-Research---Thresholds-in-Securities-Lending-Metrics--16-Sep-2015_Updated.pdf 37. Data Explorers - Wikipedia, https://en.wikipedia.org/wiki/Data_Explorers 38. Securities Finance: Short Interest Data Dataset | S&P Global Marketplace, https://www.marketplace.spglobal.com/en/datasets/securities-finance-short-interest-data-(1704372663) 39. Global securities financing data, providing market leading analytics on short seller demand, supply and borrow costs., https://www.spglobal.com/marketintelligence/en/mi/info/0920/ecr-bm-si.html 40. S&P Global Market Intelligence - Short Interest Data | Data Analytics - LSEG, https://www.lseg.com/en/data-analytics/financial-data/pricing-and-market-data/equities-market-data/sp-global-market-intelligence 41. Securities Finance - S&P Global, https://www.spglobal.com/market-intelligence/en/solutions/products/securities-finance 42. I believe that the failing to deliver and feeling to receive that are plaguing the stock market are an enormous risk to our econ - SEC.gov, https://www.sec.gov/comments/4-848/4848-581275-1670903.pdf 43. The Role of a Prime Broker - Investopedia, https://www.investopedia.com/articles/professionals/110415/role-prime-broker.asp 44. Foreign Exchange Prime Brokerage - Federal Reserve Bank of New York, https://www.newyorkfed.org/medialibrary/microsites/fxc/files/2005/fxc051219a.pdf 45. Prime Brokerage Services, Example, and Requirements for an Account - Investopedia, https://www.investopedia.com/terms/p/primebrokerage.asp 46. JP Morgan Securities LLC ("JPMS") Guide to Investment Banking Services and Prime Brokerage Services, https://www.jpmorgan.com/content/dam/jpm/global/disclosures/by-regulation/prime-brokerage-services-jpms.pdf 47. CNS Prime Broker Interface - DTCC, https://www.dtcc.com/clearing-and-settlement-services/equities-trade-capture/cns-prime-broker-i nterface 48. The 9 Biggest Hedge Fund Failures - Investopedia, https://www.investopedia.com/articles/investing/101515/3-biggest-hedge-fund-scandals.asp 49. Hedge Fund Strategies | CFA Institute, https://www.cfainstitute.org/insights/professional-learning/refresher-readings/2025/hedge-fund-strategies 50. Economic Issues No. 19 -- Hedge Funds: What Do We Really Know?, https://www.imf.org/external/pubs/ft/issues/issues19/ 51. Hedge Funds, Financial Intermediation, and Systemic Risk - Federal Reserve Bank of New York, https://www.newyorkfed.org/medialibrary/media/research/epr/07v13n3/0712kamb.pdf 52. The prime broker–hedge fund nexus: recent evolution and implications for bank risks, https://www.bis.org/publ/qtrpdf/r_qt2403y.htm 53. Basel III E: Synthetic Financing by Prime Brokers - EliScholar, https://elischolar.library.yale.edu/cgi/viewcontent.cgi?article=1042&context=journal-of-financial-crises 54. Fails-to-Deliver Data - SEC.gov, https://www.sec.gov/data-research/sec-markets-data/fails-deliver-data 55. Data and Research - SEC.gov, https://www.sec.gov/data-research 56. Fails-to-Deliver Data - Dataset - Catalog, https://catalog.data.gov/dataset/fails-to-deliver-data 57. LSEG Data & Analytics | Financial Technology & Data, https://www.lseg.com/en/data-analytics 58. FactSet | Leading Financial Data, Market Analysis & Insights, https://www.factset.com/ 59. Structural Breaks in Financial Time Series | Request PDF - ResearchGate, https://www.researchgate.net/publication/226038515_Structural_Breaks_in_Financial_Time_Series 60. Market Volatility and Trading volume - UWI St. Augustine, https://sta.uwi.edu/conferences/financeconference/Conference%20Papers/Session%208/Does%20Trading%20Volume%20really%20explain%20Stock%20returns%20Volatility.pdf 61. A hybrid model for financial time‐series forecasting based on mixed methodologies - NSF-PAR, https://par.nsf.gov/servlets/purl/10227918 62. Time-Series Analysis: What Is It and How to Use It | TigerData, https://www.tigerdata.com/blog/time-series-analysis-what-is-it-how-to-use-it 63. What Is a Time Series and How Is It Used to Analyze Data? - Investopedia, https://www.investopedia.com/terms/t/timeseries.asp 64. Predicting failed trade settlements - Splunk Lantern, https://lantern.splunk.com/Splunk_Platform/UCE/Financial_Services_and_Insurance/Predicting_failed_trade_settlements 65. Estimation of Machine Learning–Based Models to Predict Dementia Risk in Patients With Atherosclerotic Cardiovascular Diseases: UK Biobank Study - JMIR Aging, https://aging.jmir.org/2025/1/e64148 66. Predictive Analytics and Time Series Forecasting - IGI Global, https://www.igi-global.com/chapter/predictive-analytics-and-time-series-forecasting/355224 67. Clearstream launches AI tool to predict settlement fails - Finadium, https://finadium.com/clearstream-launches-ai-tool-to-predict-settlement-fails/ 68. Backtesting - Wikipedia, https://en.wikipedia.org/wiki/Backtesting 69. What Is Backtesting? Definition, Benefits, and Limitations - Investopedia, https://www.investopedia.com/terms/b/backtesting.asp 70. Quantitative Investment Strategies: Backtesting, P... | FMP - Financial Modeling Prep, https://site.financialmodelingprep.com/education/financial-ratios/Quantitative-Investment-Strategies-Backtesting-Performance-Attribution-and-Machine-Learning-Applications 71. Why do probabilistic clinical models fail to transport between sites - PMC - PubMed Central, https://pmc.ncbi.nlm.nih.gov/articles/PMC10907678/ 72. Glossary:In-sample vs. out-of-sample forecasts - Statistics Explained - Eurostat, https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:In-sample_vs._out-of-sample_forecasts 73. What is the difference between in-sample and out-of-sample forecasting? - Milvus, https://milvus.io/ai-quick-reference/what-is-the-difference-between-insample-and-outofsample-forecasting 74. supervisory framework for the use of “backtesting” in conjunction with the internal models approach to market risk capital requirements, https://www.bis.org/publ/bcbsc223.pdf 75. The Settlement Discipline Regime of CSDR - LSEG, https://www.lseg.com/content/dam/post-trade/en_us/documents/lch/white-papers/lch-the-settlement-discipline-regime-of-cdsr-0.pdf 76. Informed short selling, fails-to-deliver, and abnormal returns | Request PDF - ResearchGate, https://www.researchgate.net/publication/303537618_Informed_short_selling_fails-to-deliver_and_abnormal_returns