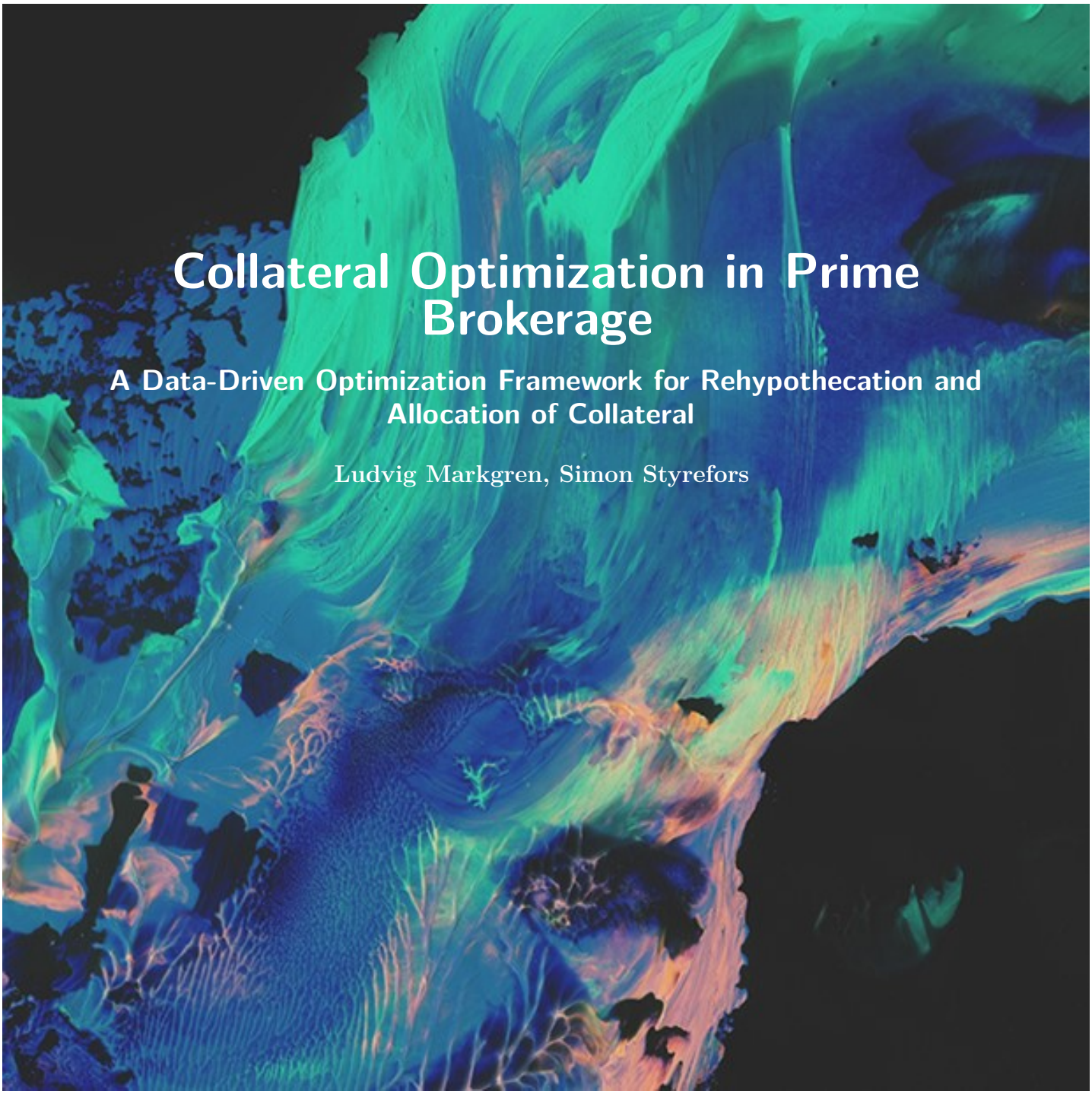
UMEÅ UNIVERSITY



Collateral Optimization in Prime Brokerage

A Data-Driven Optimization Framework for Rehypothecation and Allocation of Collateral

Ludvig Markgren, Simon Styrefors

Master's Thesis, 30 ECTS
M.Sc. in Industrial Engineering and Management, 300 ECTS
Spring term 2026

Collateral Optimization in Prime Brokerage
A Data-Driven Optimization Framework for Rehypothecation and Allocation of Collateral

Copyright © 2026 Ludvig Markgren & Simon Styrefors
All rights reserved

Ludvig Markgren[†], luma0054@student.umu.se
Simon Styrefors[†], sist0064@student.umu.se

Supervisors: Johan Hagenbjörk, Clear Street
Peter Anton, Department of Mathematics and Mathematical Statistics

Examiner: Per Arnqvist, Department of Mathematics and Mathematical Statistics

Degree Project in Industrial Engineering and Management, 30 ECTS
Department of Mathematics and Mathematical Statistics
Umeå University
SE-901 87 Umeå, Sweden

[†]Equal contribution.

Abstract

Collateral allocation in prime brokerage has become a complex optimization problem driven by regulatory constraints, funding costs, and market conditions. This thesis develops a data-driven optimization framework for the daily allocation of securities and collateral across competing uses, including internalization and secured funding through external lending. The problem is formulated as a linear programming model that integrates rehypothecation limits, segregation requirements, and key economic drivers such as lending fees, borrowing costs, and balance sheet constraints. The framework captures both inventory flows and their associated cash dynamics, enabling a consistent evaluation of allocation decisions.

The model is evaluated through historical backtesting against the firm's current heuristic approach. The results show consistent improvements in economic performance, achieved through more efficient allocation rather than increased volume. To further understand these gains, a time-series regression analysis is conducted to identify the conditions under which the model outperforms. The findings indicate that performance improvements are driven by market rates and the availability of rehypothecatable collateral.

Overall, the results demonstrate that a systematic optimization approach can significantly improve collateral allocation efficiency in prime brokerage.

Keywords: *Prime Brokerage, Collateral Optimization, Linear Programming, Rehypothecation, Securities Lending.*

Sammanfattning

Allokering av tillgångar och säkerheter inom prime brokerage har utvecklats till ett komplext optimeringsproblem drivet av regulatoriska begränsningar, finansieringskostnader och marknadsförhållanden. Denna studie utvecklar ett datadrivet optimeringsramverk för daglig allokering av säkerheter mellan olika användningsområden, såsom intern matchning eller extern utlåning. Problemet formuleras som en linjär optimeringsmodell som integrerar rehypotekeringsgränser, segregeringskrav och ekonomiska drivkrafter såsom utlåningsavgifter, upplåningskostnader och avvägningar kopplat till företagets balansräkning. Ramverket fångar både flöden av värdepapper och tillhörande kassaflöden, vilket möjliggör konsekvent utvärdering av allokeringsbeslut.

Modellen utvärderas genom en historisk backtesting mot företagets nuvarande heuristiska metod. Resultaten visar konsekventa förbättringar i ekonomisk prestation, vilket uppnås genom mer effektiva allokeringsbeslut snarare än ökade volymer. För att ytterligare förstå dessa förbättringar genomförs en tidsserieanalys i form av regressionsmodellering för att identifiera under vilka förhållanden modellen presterar bäst. Resultaten indikerar att förbättringarna drivs av marknadsräntor samt tillgången på säkerheter att rehypotekera.

Sammanfattningsvis visar studien att ett systematiskt optimeringsbaserat angreppssätt kan förbättra effektiviteten i säkerhetsallokering inom prime brokerage.

Svensk titel: *Säkerhetsoptimering inom Prime Brokerage: Ett datadrivet optimeringsramverk för rehypotekering och allokering av kundsäkerheter*

Acknowledgements

We would like to express our gratitude to Clear Street for the opportunity to conduct this master's thesis in collaboration with them. We are especially thankful for the warm welcome, access to their facilities, and the resources that made this project possible. We also extend our appreciation to everyone at the company for their support and willingness to always take the time to offer their help and guidance.

A special thanks goes to our supervisor Johan Hagenbjörk at Clear Street for his continuous support, guidance, and for sharing his knowledge in the field. We would also like to thank Peter Anton at Umeå University for his valuable support and helpful feedback throughout the thesis work.

Ludvig Markgren
Simon Styrefors
Umeå, June 10, 2026

Abbreviations

AAPL Apple Inc.

ACF Autocorrelation Function

ADF Augmented Dickey-Fuller

AI Artificial Intelligence

AMPL A Mathematical Programming Language

AR Autoregressive

CCP Central Counterparty

DW Durbin-Watson

GC General Collateral

HAC Heteroskedasticity and Autocorrelation Consistent

ILP Integer Linear Programming

LP Linear Programming

MILP Mixed-Integer Linear Programming

OBFR Overnight Bank Funding Rate

OLS Ordinary Least Squares

PACF Partial Autocorrelation Function

PnL Profit and Loss

Repo Repurchase Agreement

SEC Securities and Exchange Commission

TU Totally Unimodular

Contents

1. Introduction **1**

1.1. About Clear Street . . . 1
1.2. Project Description . . . 1
1.3. Aim of the Thesis . . . 2
1.4. Related work . . . 2
1.5. Delimitations . . . 3
1.6. Disposition . . . 4
1.7. Use of AI . . . 4

2. Background **5**

2.1. Prime Brokerage . . . 5
2.2. Core Concepts . . . 6
2.2.1. Long and Short Positions . . . 6
2.2.2. Margin Lending and Collateral Management . . . 6
2.2.3. Rehypothecation . . . 6
2.2.4. Segregation . . . 7
2.2.5. Internalization . . . 7
2.2.6. General Collateral and Specials . . . 9
2.3. Sources and Uses . . . 10
2.3.1. Sources of Inventory . . . 11
2.3.2. Uses of Inventory . . . 12
2.4. Regulatory Framework . . . 12
2.4.1. SEC Rule 15c3-3 . . . 13
2.4.2. SEC Rule 15c3-1 . . . 13
2.5. Balance Sheet Costs . . . 14
2.6. Economic Costs of External Borrowing . . . 15
2.7. Value of Internalization . . . 15

3. Theory **17**

3.1. The Rehypothecation Limit . . . 17
3.2. Rebate Rate Pricing . . . 17
3.3. Optimization Theory . . . 18
3.3.1. Network Flows and Side Constraints . . . 18
3.3.2. Linear programming . . . 20
3.3.3. Integer Programming and LP Relaxation . . . 21
3.3.4. The simplex method . . . 22
3.3.5. The dual simplex method . . . 22
3.4. Backtesting and Comparative Analysis . . . 23
3.5. Statistical Dependencies in Financial Time-Series . . . 24
3.5.1. Autocorrelation and the i.i.d. Assumption . . . 24
3.5.2. Testing for Autocorrelation . . . 24
3.5.3. The Augmented Dickey-Fuller test . . . 25

3.5.4. HAC Robustness: The Newey–West Estimator . . . 26
3.5.5. Time-Series Regression . . . 26

4. Methodology **27**
4.1. Data . . . 27
4.1.1. Core Position and Inventory Data . . . 27
4.1.2. Market Data and Rates . . . 27
4.1.3. Regulatory and Operational Constraints . . . 28
4.1.4. Allocation Data for Backtesting and Benchmarking . . . 29
4.1.5. Data Processing . . . 29
4.1.6. Input Data Example . . . 29
4.2. Modeling . . . 29
4.2.1. Model Sets, Parameters, and Variables . . . 30
4.2.2. Flow Architecture . . . 31
4.2.3. Economic Interpretation . . . 33
4.2.4. Objective Function . . . 33
4.2.5. Constraints . . . 34
4.3. Implementation and Solver . . . 36
4.4. Evaluation Framework . . . 36
4.4.1. Backtesting Design . . . 37
4.4.2. Performance Driver Analysis . . . 37
4.4.3. Interpretation . . . 38

5. Results and Discussion **39**
5.1. Performance Overview . . . 39
5.2. Allocation Mechanisms and Efficiency . . . 40
5.2.1. Rehypothecation Budget Utilization . . . 40
5.2.2. Internalization Efficiency . . . 41
5.2.3. External Borrowing and Funding Strategy . . . 42
5.3. Time-Series Diagnostics and Robustness . . . 44
5.3.1. Autocorrelation Analysis . . . 44
5.3.2. Stationarity Analysis . . . 45
5.4. Drivers of Performance . . . 45
5.4.1. Regression Results . . . 46
5.4.2. Performance Drivers Model Validation . . . 47

6. Conclusion **50**
6.1. Summary and Answers to Research Questions . . . 50
6.2. Limitations and Future Research . . . 51

References **54**

A. Pseudocode **56**

1. Introduction

This section introduces the context, motivation, and objectives of the thesis. It presents the industry partner, Clear Street, and outlines the scope of the project. The research questions and aims are then defined, followed by a review of relevant literature. Finally, the section describes the study's delimitations, outlines the structure of the report, and discloses the use of Artificial Intelligence (AI) in the research process.

1.1. About Clear Street

Clear Street LLC is a U.S.-based financial services firm that provides clearing and prime brokerage services, including securities lending, margin lending, and collateralized financing solutions (Clear Street LLC, 2024). The firm operates a cloud-native clearing platform designed to support complex and global markets and serves a broad base of institutional clients across the capital markets (Clear Street, 2026).

This thesis is conducted in cooperation with the Engineering team at Clear Street, based in New York, USA, and supervised from Clear Street's Umeå office in Sweden.

1.2. Project Description

This thesis investigates the design, construction, and evaluation of a dynamic optimization model for collateral and inventory allocation within a prime brokerage environment. The primary objective is to develop a mathematical framework capable of determining, on a daily basis, the optimal allocation of inventory sourced from internal and external channels to competing uses, such as internalization and funding through external securities lending.

The core of the project involves modeling the economic trade-offs inherent in daily collateral management. By structuring the problem as a source–use allocation optimization, the model seeks to maximize economic value while minimizing unnecessary funding and borrowing costs. A particular focus is placed on capturing the revenue potential from securities lending and the cost-saving benefits of internalization, which are key drivers of prime brokerage profitability.

The optimization model is designed to be data-driven and dynamic, utilizing daily end-of-day client positions, inventory data, and market information as inputs. Crucially, the model operates under a set of realistic operational and regulatory constraints, ensuring that allocations respect inventory balances, eligibility requirements, and mandatory asset segregation levels without compromising client protection.

Finally, the proposed model is evaluated through a comprehensive backtesting framework. By simulating the optimization model on historical data, its performance is benchmarked against the heuristic collateral management approach currently employed by the prime broker. This evaluation assesses the model's impact on key economic metrics, specifically borrowing costs,

funding efficiency, and net profitability.

1.3. Aim of the Thesis

The primary aim of this thesis is to develop and evaluate a mathematical optimization model for daily inventory allocation within a prime broker. The study seeks to quantify the economic efficiency of replacing current heuristic methods with a data-driven strategy, while still complying with U.S. regulations.

The following research questions are addressed:

RQ1 How can prime brokerage inventory flows be modeled to balance the economic trade-offs between competing uses?

RQ2 How can complex regulatory mandates regarding asset segregation and rehypothecation limits be integrated as binding constraints within an optimization framework?

RQ3 To what extent does the proposed optimization model improve overall economic performance relative to the firm's current heuristic allocation approach?

RQ4 Which market and inventory characteristics explain when the optimization model outperforms the heuristic benchmark?

1.4. Related work

Research relevant to this thesis can be broadly divided into two categories: (i) economic studies describing how financial regulation has turned inventory management from a routine task into a strategic issue, and (ii) technical studies that apply mathematical optimization to collateral decision-making. The first category establishes the economic motivation for treating collateral as a scarce resource, while the second provides the algorithmic methods that this thesis extends.

Studies following the 2008 financial crisis emphasise that collateral has shifted from being primarily an operational margining task to a strategic decision problem with balance sheet, liquidity, and funding implications. Java et al. highlights how post-crisis regulatory requirements related to collateral, segregation, and operational frictions reduce collateral mobility and increase the importance of systematic allocation (Java et al., 2017). Similarly, Aydin characterizes this development as a transition from collateral *management* to collateral *optimization*, arguing that firms must balance multiple cost and risk drivers, such as liquidity usage, leverage exposure, and capital impacts, rather than simply posting the cheapest eligible asset (Aydin, 2016).

In the specific context of prime brokerage, Eren provides empirical evidence that matching opposing client flows internally, rather than borrowing externally, is a primary driver of profitability. He estimates significant cost savings for major banks and argues that post-crisis

regulations have made this type of inventory efficiency a strategic necessity rather than just an operational convenience (Eren, 2015). Although these contributions are largely conceptual, they provide the strong economic motivation for the formal optimization models developed in this thesis.

Formal optimization models are less common in the academic literature and most studies take the perspective of a bank allocating collateral to external Central Counterparty (CCP). Orrsveden and Tarukoski formulate collateral allocation in securities lending using Linear Programming (LP), Integer Linear Programming (ILP) and Mixed-Integer Linear Programming (MILP) models that incorporate opportunity costs, CCP-related costs, and transaction effects. Using real institutional data, they show that optimization can reduce allocation costs relative to heuristic approaches, but the decision setting studied remains a bilateral posting problem rather than an internal inventory management problem (Orrsveden and Tarukoski, 2019).

Bylund studies collateral allocation in a central clearing setting and formulates the problem as a MILP that minimizes collateral costs subject to eligibility, haircut and concentration constraints (Bylund, 2017). The modeling structure again reflects a posting-to-CCP perspective, where the main question is how to select collateral to meet margin requirements at minimum cost, rather than how to allocate a shared inventory across competing internal and external uses.

More recently, Giron et al. formulate collateral allocation as a MILP and then reformulate it as a Quadratic Unconstrained Binary Optimization to study quantum and hybrid solution approaches. Their framework allocates assets to CCP under exposure, haircut, and eligibility constraints, providing an alternative computational approach while maintaining a traditional posting setting (Giron et al., 2023).

Overall, prior work shows that collateral allocation can be expressed as LP, ILP and MILP-type models under eligibility and haircut constraints, and that optimization can outperform rule-based allocation in bank-style posting problems. However, limited attention has been given to a prime brokerage setting, where the allocation problem is multi-sided: inventory must be split across segregation, rehypothecation, and revenue-generating activities, often with client-level constraints. This thesis extends the optimization literature to that setting by modelling client-level inventory, reuse limitations and segregation requirements within a dynamic allocation framework.

1.5. Delimitations

This thesis investigates the design and evaluation of a daily optimization model for inventory allocation within a U.S. prime brokerage. The analysis is limited to end-of-day allocation decisions based on historical position and market data provided by Clear Street. Intraday

dynamics, real-time liquidity shocks, and extreme stress scenarios are not explicitly modeled.

The optimization framework incorporates key regulatory constraints related to asset segregation and rehypothecation under the U.S. regulatory regime. International regulatory frameworks and cross-border collateral considerations are outside the scope of this study. The economic performance of the model depends on the accuracy of input parameters such as borrowing costs, lending rebates and regulatory constraints. When complete or up-to-date data are unavailable, certain parameters may be estimated or approximated using prior-day information, which introduces uncertainty into the backtesting results.

Finally, the empirical evaluation is conducted through historical backtesting against the firm's current heuristic allocation approach. The results therefore reflect past market conditions and do not constitute a fully implemented production system for live deployment.

1.6. Disposition

This thesis is structured as follows. Section 2 provides the background to prime brokerage, collateral management, and the regulatory framework governing inventory allocation.
Section 3 outlines the theoretical framework, detailing the core financial mechanics of prime brokerage and the principles of linear optimization that form the study.
Section 4 describes the data, parameters, and implementation methodology, as well as the design of the backtesting framework.
Section 5 presents the empirical results and evaluates the model's performance relative to the existing heuristic allocation approach. Finally, Section 6 discusses the findings, answers the research questions, and outlines conclusions and directions for future research.

1.7. Use of AI

Throughout this thesis project, AI tools were utilized to streamline the data engineering phase, specifically by assisting in the formulation of SQL queries for data extraction. Additionally, AI helped troubleshooting and debugging within the A Mathematical Programming Language (AMPL) modeling environment. During the drafting of this report, AI was employed to generate LaTeX code for tables and mathematical equations, as well as to refine the grammatical flow and overall formatting of the report.

2. Background

To establish the contextual and economic foundation for the thesis, the background section introduces the operational mechanics of prime brokerage. It describes core concepts such as rehypothecation, segregation, and internalization, alongside the strict U.S. regulatory frameworks governing inventory flows. Furthermore, the section establishes the economic costs associated with balance sheet capacity and external financing.

2.1. Prime Brokerage

Prime brokerage is a specialized form of financial intermediation that provides institutional clients, such as hedge funds, with services including trade execution, clearing, custody, margin financing, and securities lending. A core element of the prime brokerage business model is the management and allocation of client collateral across these services in order to support financing activities and generate economic value (Financial Stability Board, 2012).

A key feature of this model stems from the broker's role as a liquidity provider. When extending margin loans to finance client positions, prime brokers may obtain contractual rights to reuse the assets pledged as collateral. This practice, known as rehypothecation, allows brokers to reuse pledged securities for their own purposes, including securities lending, subject to regulatory and contractual constraints (Financial Stability Board, 2012). As discussed by Bottazzi et al., collateral reuse can enhance liquidity and funding efficiency in securities markets, while also introducing important trade-offs related to risk management and the protection of client assets (Bottazzi et al., 2012).

By contrast, asset segregation refers to the regulatory requirement that certain client assets must be held separately from the broker's own inventory and cannot be rehypothecated, thereby ensuring client asset protection but reducing the broker's funding and revenue capacity. For prime brokers, these opposing mechanisms make the allocation of collateral between segregated and reusable pools a central operational decision with direct economic implications. Efficient collateral and inventory management is therefore critical to a prime broker's profitability and balance sheet optimization. Available inventory may originate from multiple sources, including client margin positions, internal inventory and external markets. These can be deployed across competing uses such as client segregation, internalization or external lending. The allocation of collateral across these uses involves fundamental economic trade-offs between funding costs, liquidity, regulatory capital usage, and revenue generation (Financial Stability Board, 2012).

At the same time, prime brokers operate under strict regulatory frameworks designed to safeguard market stability and protect client assets. Regulatory requirements impose minimum segregation levels for client assets, restrict the extent of collateral reuse, and mandate robust risk management practices related to rehypothecation and funding activities. These constraints significantly limit the feasible allocation space and introduce additional complex-

ity into daily collateral management decisions. As a result, prime brokers must continuously evaluate and optimize how inventory and collateral are allocated across sources and uses in order to minimize unnecessary funding and transaction costs while remaining compliant with regulatory and operational constraints. This creates a natural setting for a formal, data-driven optimization framework, which can support systematic and transparent allocation decisions and improve economic outcomes relative to heuristic or rule-based approaches (Financial Stability Board, 2012).

2.2. Core Concepts

To fully comprehend the mechanics of prime brokerage inventory management, it is necessary to first define the fundamental market positions and the regulatory mechanisms that drive inventory flows. The daily operations of a prime broker rely on a continuous cycle of these interacting states.

2.2.1. Long and Short Positions

The basis of the inventory equation consists of two fundamental states: long positions (supply) and short positions (demand). A long position represents ownership of a security, where the client anticipates an increase in value. In a prime brokerage context, these positions are often financed through leverage, meaning the client borrows cash from the broker to purchase the asset. Conversely, a short position involves the sale of a security that the client does not own, typically to profit from a price decline. To settle this sale, the client must borrow the security from the prime broker, creating a delivery obligation (Financial Stability Board, 2012).

2.2.2. Margin Lending and Collateral Management

Central to the prime brokerage relationship is the concept of margin lending. Institutional clients, such as hedge funds, frequently borrow cash from their prime broker to leverage their investment positions. To secure these loans, clients pledge securities as collateral. This collateral serves as the primary protection for the broker against counterparty credit risk. As noted by the Dutch Central Bank, efficient collateral management has evolved beyond simple risk mitigation into a sophisticated optimization process known as collateral transformation, where lower-quality assets are exchanged for higher-quality liquid assets to meet regulatory requirements (Capel and Levels, 2014).

2.2.3. Rehypothecation

When a client pledges collateral to secure a margin debit, the prime broker obtains the right to use these assets for its own purposes. This practice, distinct from a standard security interest, is defined by the International Capital Market Association as rehypothecation. It involves the prime broker reusing the pledged collateral to secure its own funding or to cover the short positions of other clients. Unlike a Repurchase Agreement (Repo) transaction which typically involves a transfer of title, rehypothecation in the U.S. context relies on the contractual

right to use the client's asset as if it were the broker's own, subject to regulatory limits. Specifically, Federal Reserve Regulation T and Securities and Exchange Commission (SEC) Rule 15c3-3 limit the amount of a client's assets which a prime broker may rehypothecate to the equivalent of 140% of the client's net liability to the prime broker (International Capital Market Association, 2024).

2.2.4. Segregation

Segregation can be viewed as the inverse of rehypothecation. Assets that are fully paid for by the client, or that represent collateral value in excess of regulatory limits, must be physically separated from the broker's own assets to ensure client protection. However, holding excess segregated assets reduces the broker's flexibility. These assets cannot be reused, pledged, or invested in other activities. In practice, this means the broker earns little or no return on assets that could otherwise generate income. Under Basel III frameworks, efficient management of this liquid asset buffer is critical for meeting the Liquidity Coverage Ratio, balancing the need for safety against the opportunity cost of holding idle assets (Basel Committee on Banking Supervision, 2013). Figure 1 illustrates the complete lifecycle of a margin trade, demonstrating how a client's pledged collateral is mechanically split between the broker's usable inventory and locked custody based on the 140% rule.

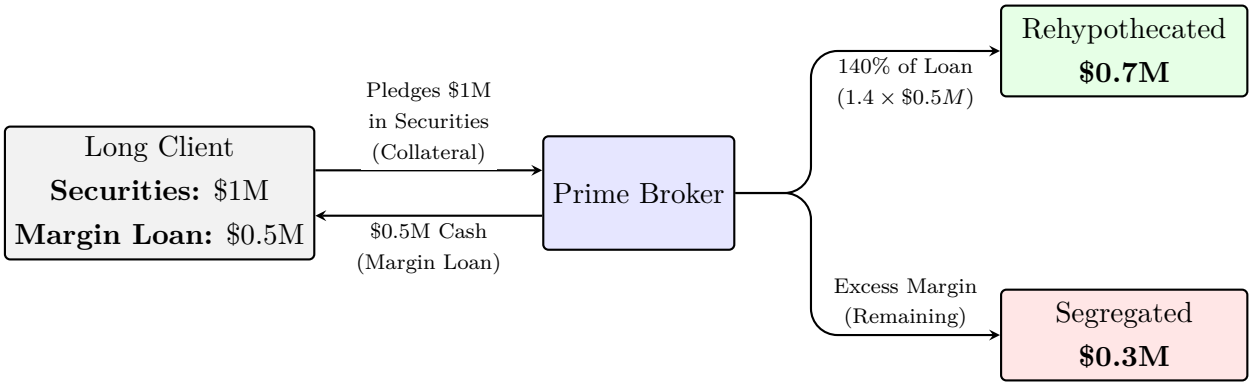


Figure 1: The relationship between margin lending, rehypothecation, and segregation. A client finances 50% of a $1M position via a margin loan. The prime broker provides a $0.5M cash loan and receives the $1M position as collateral. The broker can rehypothecate securities up to 140% of the loan ($0.7M) for its own operations, while the remaining $0.3M is legally locked in a segregated account.

2.2.5. Internalization

The interplay between client demand for short positions and the available supply of rehypothecated collateral drives the process of internalization. Internalization occurs when a prime broker utilizes rehypothecable collateral from a client with a long position to cover the delivery obligation of another client with a short position in the same security.

To illustrate the mechanics of this process, Eren describes a scenario involving a high-volume security such as Apple Inc. (AAPL) (Eren, 2015). Suppose a hedge fund client wishes to

take a leveraged long position in AAPL. To finance the trade, the client borrows cash from the prime broker (a margin loan) and pledges the purchased AAPL shares as collateral. The prime broker, holding the right to rehypothecate these assets, now holds AAPL inventory. Simultaneously, a second client wishes to execute a short sale of AAPL and requires the securities for delivery.

In an internalized transaction, the prime broker intermediates directly between these two clients. The broker utilizes the AAPL shares pledged by the long client and lends them to the short client to settle the sale. The cash proceeds from the short sale are then held by the broker as collateral. Through this process, the broker facilitates the flow of securities from the long hedge fund to the short hedge fund, and the flow of cash from the short hedge fund to the long hedge fund, without needing to access external markets (Eren, 2015). Figure 2 illustrates this complete mechanism, demonstrating how the prime broker acts as the central node for both the cash and security flows.

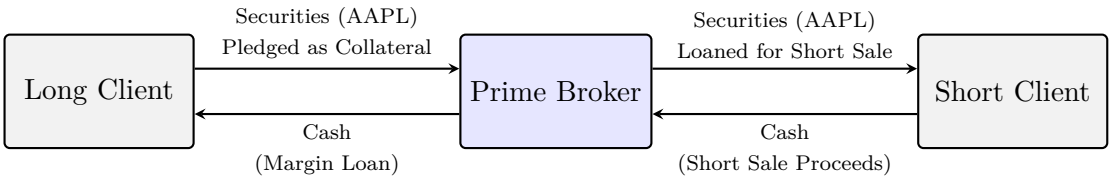


Figure 2: Mechanics of an internalized transaction, illustrating the simultaneous flow of cash and securities between clients and the prime broker without external market intermediation.

By matching these opposing flows internally, the broker avoids the potential costs associated with borrowing securities from external lenders. Eren highlights that this mechanism is a primary driver of prime brokerage profitability, estimating significant annual savings for major banks derived specifically from reduced search frictions and balance sheet efficiency. This economic incentive drives a phenomenon known as *specialization*, where prime brokers actively curate client bases with correlated strategies to maximize the probability of netting internal flows. In the post-crisis regulatory environment, internalization has evolved from an operational efficiency into a strategic necessity for minimizing balance sheet usage (Eren, 2015).

Complementing this strategic perspective, Kirk et al. provide the operational framework for understanding these balance sheet benefits. They establish a hierarchy of funding sources where internalization is ranked as the most efficient method due to its ability to net positions on the balance sheet. Under accounting standards such as U.S. GAAP, internalized trades allow the broker to offset the asset (the margin loan) against the liability (the short sale proceeds), thereby reducing reported leverage exposure and regulatory capital requirements (Kirk et al., 2014). Furthermore, Kirk et al. introduce the critical concept of *internalization capacity*: the theoretical maximum amount of client obligations that can be netted given the broker's inventory. They argue that operational frictions often prevent banks from fully uti-

lizing this capacity, creating a gap between potential and actual efficiency that necessitates advanced allocation logic (Kirk et al., 2014).

2.2.6. General Collateral and Specials

Within the mechanics of securities lending, the specific identity and market demand of an asset heavily dictate a prime broker's financing costs. Consequently, securities are broadly classified into two fundamental categories: General Collateral (GC) and Specials.

General Collateral refers to ordinary, highly liquid assets, such as major sovereign bonds or large-cap equities, that are widely available. In a GC transaction, the party receiving the collateral is relatively indifferent to the specific security delivered; the transaction is fundamentally a cash-driven financing arrangement (International Capital Market Association, 2024).

Specials, by contrast, are specific securities subject to exceptional borrowing demand, typically driven by intense short-selling interest or a structural scarcity of supply in the cash market. Transactions involving Specials are securities-driven, meaning participants actively compete to borrow that exact asset (Duffie, 1996).

To understand the economic impact of this classification, it is necessary to examine the mechanics of collateral and rebates. In standard U.S. securities lending, the borrower of a security must post cash collateral to the lender, typically at a margin of roughly 102% of the security's market value. The lender of the security holds this cash for the duration of the trade and reinvests it to earn interest at a prevailing benchmark overnight rate. Rather than paying a separate reinvestment rate and charging a separate borrowing fee, the U.S. market combines these two components into a single, netted *rebate rate* paid by the lender to the borrower. The intrinsic fee for borrowing the security is therefore the difference between the prevailing benchmark interest rate and this quoted rebate rate (Duffie, 1996).

For a GC security, the market is competitive and the asset is easy to borrow. Therefore, the lender rebates the vast majority of the interest earned back to the borrower, resulting in a positive rebate. Because a lender will not rebate more interest than the cash collateral actually generates, this positive rebate rate is strictly capped by the prevailing macroeconomic benchmark rate.

However, as a security becomes harder to borrow and transitions into a Special, the lender gains pricing power and decreases the rebate offered. For extremely scarce Specials, this rebate rate can drop to zero or become heavily negative. A negative rebate implies that the borrower abandons all interest returns on their cash collateral and must additionally pay a direct fee to the lender to secure the asset (Duffie, 1996). Unlike positive rebates, the implicit borrowing fee for an ultra-Special is theoretically uncapped. Consequently, negative rebate

rates can reach hundreds or even thousands of percent annualized.

This rebate mechanism forms the economic foundation of prime brokerage inventory management. Borrowing Specials from the external market to cover client shorts incurs severe, potentially uncapped financing costs due to deep negative rebates. To avoid these punitive costs, a prime broker's primary objective is to prioritize the internalization of Special securities. By matching internal short demands with available internal long positions, the broker eliminates the need for expensive external borrowing. Consequently, the firm captures the massive scarcity premium within the house rather than paying it to the street, while any remaining excess long inventory can still be lent out externally for profit.

2.3. Sources and Uses

The operational framework of prime brokerage inventory management can be conceptualized as a flow system between sources and uses. On any given trading day, a prime broker aggregates a pooled inventory of available assets from various entities. This supply (the sources) must be efficiently allocated to meet diverse financing needs and operational obligations (the uses).

While this aggregated pool offers significant flexibility, the allocation process is not entirely discretionary, it is governed by strict regulatory and contractual constraints. Because of these limitations, heuristic or manual allocation methods may fail to capture the economic trade-offs for each specific security and entity. To achieve operational efficiency, it is necessary to quantify the marginal value generated, or the cost saved, by assigning a specific unit of inventory to a specific use. This quantification allows for a global optimization of inventory flows, with the objective of balancing direct Profit and Loss (PnL) against the implied value of balance sheet capacity and funding requirements.

As illustrated in Figure 3, the allocation engine must balance the available supply of inventory against competing demands, specifically short coverage, external financing activities, and regulatory segregation.

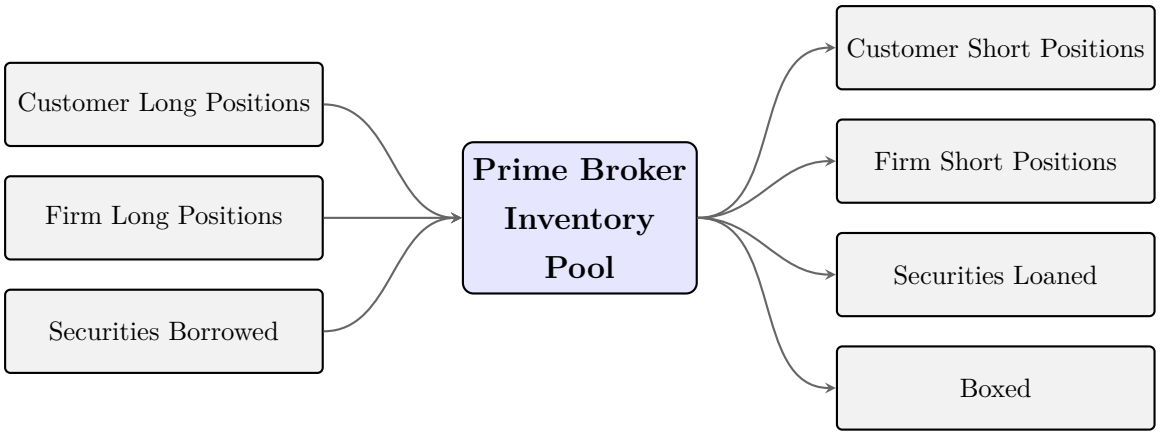


Figure 3: The flow of inventory in a prime brokerage. The optimization framework determines the optimal routing from available sources (supply) to required uses (demand) subject to regulatory constraints.

The following sections provide a detailed taxonomy of the specific inventory sources and funding uses that comprise this optimization ecosystem.

2.3.1. Sources of Inventory

Inventory sources represent the supply of securities available to the prime broker. These can be categorized based on their origin and the cost associated with utilizing them.

Customer Long Positions Customer long positions represent securities held on behalf of the prime broker's clients. Under the rehypothecation agreements described in Section 2.2.3, the broker may utilize a portion of these assets as internal inventory. This is typically the most cost-efficient source of inventory, as it does not involve external borrowing fees.

Firm Long Positions Proprietary positions held by the prime broker itself for hedging, market making, or investment purposes. Unlike customer assets, which are subject to segregation limits (Reg 15c3-3), firm inventory is fully unrestricted.

Securities Borrowed When internal sources such as Customer and Firm longs are insufficient to meet short demand, the broker must borrow securities from external counterparties, such as other banks or agent lenders. To secure these borrowed assets, the prime broker is required to post cash collateral, making this activity a direct use of cash that consumes firm liquidity. The economic impact of this borrowing is dictated by the rebate rate on the posted cash. When borrowing GC, the broker receives a positive rebate rate, effectively earning interest on its cash collateral. Conversely, when borrowing scarce Specials, the broker is charged a negative rebate rate, resulting in a direct out-of-pocket cost. Due to this liquidity consumption and the risk of steep negative rebate fees, external borrowing is generally a less efficient source of inventory compared to internal matching.

2.3.2. Uses of Inventory

Inventory uses represent the demand side of the equation, where the securities are ultimately deployed or required to be held.

Customer Short Positions The primary driver of inventory demand in a prime brokerage. When a client executes a short sale, the broker is obligated to deliver the securities to the clearing house for settlement. Covering these positions requires allocating inventory from one of the available sources.

Firm Short Positions Proprietary short positions established by the broker's own trading desks. Similar to customer shorts, these create a delivery obligation that consumes available inventory.

Securities Loaned When the broker holds excess inventory (long positions) that is not needed to cover internal short demand, these securities can be lent to external counterparties. Because securities lending transactions are collateralized with cash, this activity serves as a critical source of secured funding. However, the direct profitability of this activity depends heavily on the scarcity of the underlying asset. While lending out scarce Specials generates direct revenue via negative rebate rates, lending GC incurs an interest expense, as the broker must pay a positive rebate rate on the cash collateral received.

Boxed Inventory that is not actively used for settlement or financing is held in custody. A critical component of this use category is the segregation requirement. Regulatory rules mandate that fully paid customer securities and excess margin securities must be physically segregated and cannot be used for any other purpose. Optimizing the allocation to this use is key to minimizing the trapping of valuable inventory while remaining compliant.

2.4. Regulatory Framework

The regulatory landscape for prime brokerage and securities lending was fundamentally reshaped by the Global Financial Crisis of 2007–2008. Prior to the crisis, the management of collateral and inventory was often characterized by non-transparent rehypothecation chains and high degrees of leverage within the shadow banking system. The collapse of major financial institutions highlighted the systemic risks associated with these practices, where a loss of confidence led to a rapid withdrawal of funding and liquidity (Financial Stability Board, 2012). In response, global regulatory bodies implemented stringent frameworks designed to safeguard market stability, protect client assets, and ensure the solvency of financial intermediaries.

For a modern prime broker, this post-crisis regulatory environment constitutes the primary set of constraints within the optimization framework. Compliance is no longer merely a legal

obligation but a quantitative factor that directly influences inventory availability, funding costs, and balance sheet capacity (Clear Street LLC, 2024).

2.4.1. SEC Rule 15c3-3

The foundational constraint regarding inventory management begins with the Federal Reserve Regulation T, which defines the specific conditions under which a prime broker is permitted to rehypothecate client collateral. Under this framework, a broker may rehypothecate customer securities that are pledged as collateral for margin debits, but only up to a value equivalent to 140% of the client's debit balance (Board of Governors of the Federal Reserve System, 2024). This limit effectively creates a dynamic upper bound on the supply of available inventory from any given client entity.

Any collateral value exceeding this 140% threshold is classified as excess margin and automatically falls under the stringent segregation requirements of SEC Rule 15c3-3, commonly referred to as the Customer Protection Rule. This regulation mandates that broker-dealers maintain physical possession or control over all fully paid and excess margin securities carried for customer accounts. These assets must be held in a good control location and cannot be commingled with the broker's proprietary assets or utilized for financing activities (U.S. Securities and Exchange Commission, 2024b). Consequently, the optimization model must treat these segregated assets as strictly unavailable for internalization or external lending, regardless of the potential economic utility they might otherwise offer.

Beyond the physical segregation of assets, SEC Rule 15c3-3 imposes a parallel cash segregation requirement through its Reserve Formula computation. While Regulation T dictates which physical securities are eligible for rehypothecation, the Reserve Formula governs the cash dynamics resulting from their utilization. Broker-dealers must periodically calculate the difference between *credits* (monies owed to customers, including cash collateral generated by lending) and *debits* (monies customers owe the firm, such as debit balances in margin accounts). If total credits exceed debits, the firm is legally required to deposit the net excess into a locked Special Reserve Bank Account (U.S. Securities and Exchange Commission, 2024b). For the optimization framework, this introduces a critical funding cost mechanic. To calculate this cost accurately, the model must track both the origin of each security (such as a customer margin account) and its specific application (such as lending to the street versus internalizing a short). These factors dictate whether a transaction increases the firm's cash lockup requirement. Consequently, the objective function must weigh the gross revenue of lending or internalizing assets against the punitive funding cost of triggering a cash reserve requirement.

2.4.2. SEC Rule 15c3-1

Finally, beyond the physical availability of securities and cash lockup dynamics, the allocation decision is further constrained by capital adequacy requirements, specifically SEC Rule 15c3-

1. This rule, known as the Net Capital Rule, requires broker-dealers to maintain a minimum level of liquid capital to cover potential market risks. A key component of this regulation is the application of haircuts, which are percentage deductions from the market value of securities held in inventory (U.S. Securities and Exchange Commission, 2024a). Holding excess inventory in the box or engaging in specific lending transactions consumes regulatory capital through these haircuts. Therefore, the optimization framework must account for the opportunity cost of regulatory capital usage, incentivizing allocations that minimize unnecessary inventory accumulation and prioritize capital-efficient uses.

2.5. Balance Sheet Costs

A central economic concept underlying this thesis is that funding is not costless merely because the borrowing rate is low. Expanding a prime broker's balance sheet imposes implicit costs arising from regulatory capital requirements, leverage constraints, and operational frictions. These costs are commonly referred to as *balance sheet costs* or *shadow funding costs*.

Fleckenstein and Longstaff provide a framework for identifying and measuring such costs. Their key insight is that holding assets on balance sheet generates indirect costs beyond the explicit financing rate. When a financial intermediary purchases and holds a security, the asset increases total balance sheet size and leverage exposure. Under modern regulatory regimes, this typically requires the institution to allocate additional regulatory capital and comply with leverage and liquidity requirements (Fleckenstein and Longstaff, 2018).

Consequently, ownership of an asset generates not only a direct financing cost (e.g., the nominal borrowing fee), but also indirect costs associated with regulatory capital usage, leverage constraints, funding illiquidity, and compliance requirements. For clarity, these indirect frictions are collectively referred to as balance sheet constraints, and the economic burden they impose is referred to as the balance sheet cost.

The total cost of holding a security can therefore be decomposed as:

$$\text{Total Cost of Ownership} = \text{Direct Financing Rate} + \text{Balance Sheet Cost}.$$

Fleckenstein and Longstaff empirically demonstrate that these balance sheet costs are economically significant and time-varying. Importantly, they show that these costs increase during periods of financial stress and are related to measures of capital tightness and intermediary funding conditions. This suggests that the costs are caused by real limits on balance sheet capacity, not just temporary market inefficiencies (Fleckenstein and Longstaff, 2018).

For a prime broker, this insight has direct implications. Extending margin loans, borrowing securities to cover client shorts, or financing securities through external lending all expand total balance sheet exposure and consume scarce regulatory capital. Even if the explicit lending or borrowing rate appears low, the true economic cost of funding includes the opportunity

cost of balance sheet capacity.

2.6. Economic Costs of External Borrowing

As established in Section 2.2.6 and Section 2.5, sourcing securities externally incurs costs that extend far beyond the nominal borrow rate. By combining the mechanics of the rebate market with balance sheet constraints, Lou (2021) shows that the true economic cost of external borrowing consists of three distinct components:

1. **Market borrowing fee.** The direct, market-determined cost of borrowing the asset. As discussed in Section 2.2.6, this is represented by the intrinsic fee embedded within the netted rebate rate, driven by the scarcity of the underlying security.

2. **Funding cost of over-collateralization.** Because U.S. market conventions require posting 102% cash collateral, the prime broker must finance a position larger than the asset's actual market value. Lou (2021) highlights that maintaining this excess margin generates a distinct and mandatory funding charge, even when expected counterparty losses are negligible.

3. **Capital (balance sheet) charge.** External exposures require the allocation of economic capital. Crucially, this shadow cost applies to both borrowing and lending transactions. Whether the broker is posting cash to borrow a security or receiving cash collateral when lending one, any expansion of the balance sheet consumes scarce regulatory capacity. Building on the framework of Fleckenstein and Longstaff (2018) in Section 2.5, Lou (2021) emphasizes that this consumed capacity carries an explicit opportunity cost.

Consequently, the 102% collateral convention has direct structural implications. While it effectively neutralizes counterparty credit risk, it systematically inflates the notional amount that must be funded and placed on the balance sheet. The total economic burden of external intermediation is therefore never limited solely to the market fee. For borrowing, it inherently includes the liquidity drain of the posted collateral. Furthermore, since any external transaction expands the balance sheet, the shadow cost of capital penalizes both external borrowing and external lending equally, thereby reinforcing the immense strategic value of internalization.

2.7. Value of Internalization

The previous section established that the economic cost of external borrowing is driven by three compounding frictions: the direct market borrowing fee, the funding cost of over-collateralization, and the shadow cost of balance sheet expansion. These costs occur whenever the broker must borrow securities in the external market to cover a client's short position.

As detailed in Section 2.2.5, internalization allows a prime broker to match these opposing client positions within its own network rather than sourcing securities externally. The

key economic mechanism, as highlighted by Eren, is that internalization reduces reliance on this costly external intermediation (Eren, 2015).

Connecting these insights, the value of internalization can be understood as the avoided cost of external securities borrowing. When a client short position is matched internally against another client's long position, the broker does not need to borrow securities in the market, does not need to post 102% collateral to an external lender, and does not need to allocate additional balance sheet capacity for that transaction. As a result, the broker saves on borrowing fees, funding costs, and capital charges.

Internalization creates value precisely because borrowing securities externally is expensive. When borrowing fees are high, funding is costly, or balance sheet capacity is tight, the benefits of matching positions internally become larger. This reasoning is consistent with Eren's framework, where internalization lowers overall costs (Eren, 2015).

3. Theory

The theory section presents the key technical and theoretical components of the thesis. It outlines the regulatory mechanisms governing collateral reuse, the pricing structures of securities lending, and the mathematical principles of linear optimization utilized to model the allocation problem. Furthermore, it details the backtesting methodology and introduces the time-series statistical frameworks necessary for robust performance evaluation.

3.1. The Rehypothecation Limit

As introduced in Section 2.2.3, the capacity to reuse client assets is bounded by regulatory constraints. Mathematically, the maximum allowable rehypothecation value for a given client entity is defined as a fixed scalar multiple of that client's liability. The upper bound is expressed as:

$$R_e^{\max} = 1.40 \cdot D_e$$

where $D_e \geq 0$ is the margin debit balance of client entity $e \in E$, and $R_e^{\max}$ represents the theoretical maximum value of securities that can be legally rehypothecated.

However, the feasible inventory available to the prime broker is constrained not only by this regulatory bound but also by the physical inventory pledged. The strictly feasible rehypothecation capacity is therefore defined as:

$$B_e = \min(V_e, R_e^{\max})$$

where $V_e \geq 0$ is the total aggregate market value of all eligible collateral held in entity e's account. B_e thus represents the effective maximum value of collateral extractable from entity e for internal or external routing.

Any collateral value satisfying $V_e > R_e^{\max}$ constitutes excess margin ($V_e - R_e^{\max}$). As detailed in Section 2.2.4, these specific assets are mathematically treated as strictly segregated and must be excluded from the usable inventory pool. The optimization model must therefore simultaneously map allocations against the value-based bound ($R_e^{\max}$) and the physical quantity constraints (V_e).

3.2. Rebate Rate Pricing

Building upon the mechanics of GC and Specials described in Section 2.2.6, the cost of sourcing or funding inventory is formalized through a netted rate formulation rather than separate cash flows.

The rebate rate, denoted as R^{rebate}, is defined as the benchmark reinvestment rate on the posted cash collateral minus the intrinsic securities lending fee:

$$R^{\text{rebate}} = R^{\text{benchmark}} - F^{\text{SL}}$$

where the benchmark rate $R^{\text{benchmark}}$ is an observable overnight reference rate (e.g., the Overnight Bank Funding Rate (OBFR)), and the intrinsic fee $F^{SL} \in [0, \infty)$ represents the market scarcity premium of the specific asset. By utilizing R^{rebate}, the model integrates both the interest generated by cash collateral and the scarcity cost of the asset into a single optimization parameter.

3.3. Optimization Theory

This subsection introduces the key theoretical concepts and mathematical frameworks utilized for the optimization. It covers the principles of network flows, linear programming, and the specific algorithms used to identify optimal solutions.

3.3.1. Network Flows and Side Constraints

The problem of allocating available inventory across competing internal and external uses naturally maps to a network flow formulation. In operations research, network flow models are utilized to route entities, such as data, water, or financial assets, through a system of nodes and arcs in the most efficient manner possible (Ahuja et al., 1993, Ch. 9).

Pure Network Flow and Conservation A standard network flow problem is defined on a directed graph $G = (N, A)$, where N represents the set of nodes and A represents the set of directed arcs (i, j) connecting the nodes. Let x_{ij} denote the quantity of flow routed along arc (i, j), and let c_{ij} represent the cost per unit of flow on that arc.

The fundamental principle governing any network model is the conservation of flow, which dictates that for any intermediate node, the total inflow must exactly equal the total outflow. Formally, for each node $i \in N$, the flow balance equation is defined as:

$$\sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = b_i \quad \forall i \in N$$

where b_i represents the external supply or demand at node i. If $b_i > 0$, the node is a supply node (source); if $b_i < 0$, it is a demand node (sink); and if $b_i = 0$, it is a pure transshipment node.

One of the most important theoretical properties of a pure network flow model is that its constraint matrix is Totally Unimodular (TU). If all supplies b_i and arc capacities are integers, the TU property guarantees that the optimal solution found by standard LP will automatically be strictly integer-valued, without the need for computationally expensive integer programming techniques (Ahuja et al., 1993, Ch. 11).

Multi-Commodity Network Flow In many real-world applications, including prime brokerage inventory allocation, multiple distinct types of goods must share the same network. This is formalized as a multi-commodity flow problem (Ahuja et al., 1993, Ch. 17), as conceptually illustrated in Figure 4.

Let K be the set of distinct commodities. In a financial context, each specific security ticker $k \in K$ (e.g., AAPL) constitutes a unique commodity that must flow through the system independently; one security cannot be substituted for another to cover a specific delivery obligation.

The flow variable is extended to x_{ij}^k, representing the flow of commodity k on arc (i, j). The conservation of flow must now hold strictly for each independent commodity:

$$\sum_{j:(i,j)\in A} x_{ij}^k - \sum_{j:(j,i)\in A} x_{ji}^k = b_i^k \quad \forall i \in N, \forall k \in K$$

where b_i^k is the supply or demand of security k at node i.

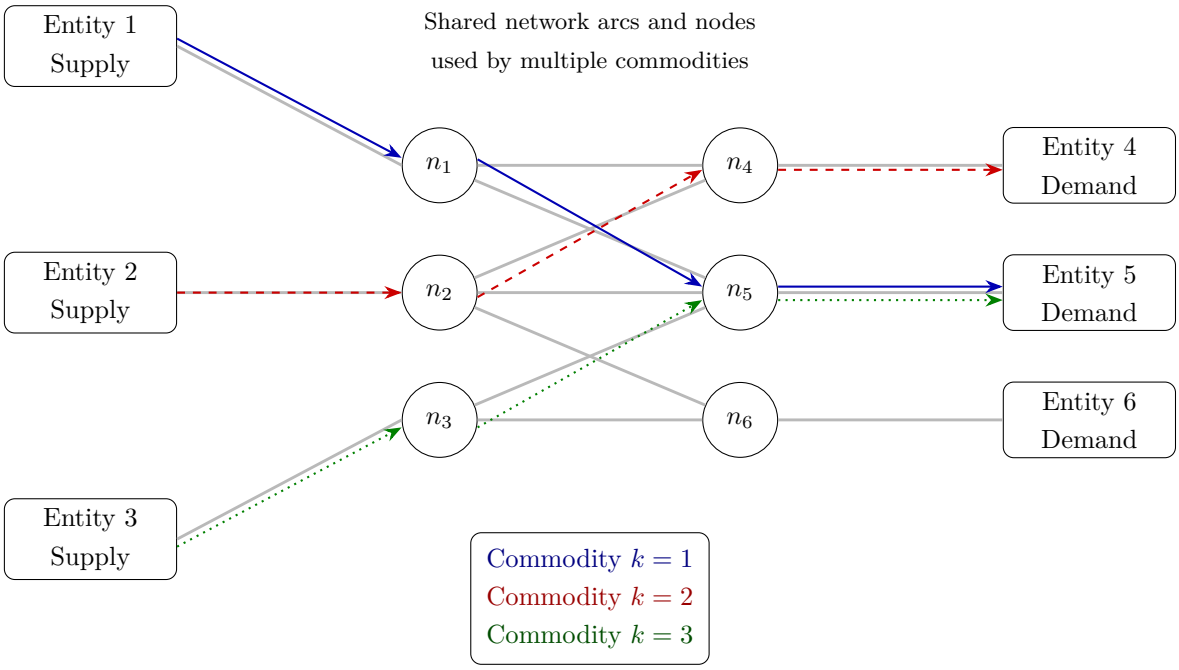


Figure 4: Illustration of a multi-commodity network flow. Multiple distinct commodities move simultaneously through the same shared network of nodes and arcs, while flow conservation must hold separately for each commodity.

Side Constraints and the Loss of Unimodularity While the flow of each security is tracked independently, the broker must also adhere to aggregate financial limits. In optimization theory, constraints that bundle or sum across multiple distinct commodities are known as side constraints or knapsack constraints (Ahuja et al., 1993, Ch. 17).

In a prime brokerage context, a primary example of a side constraint is an aggregate utilization cap. Regulatory frameworks or internal risk policies often dictate that the total market value of all client securities allocated to external lending and internalization cannot exceed the firm's aggregate margin budget. If each unit of security $k \in K$ utilizes a shared financial resource equal to its market price p^k, the side constraint bounding the total utilization across all securities against an aggregate margin budget M can be conceptually expressed as:

$$\sum_{k \in K} p^k \left(\sum_{(i,j) \in A_{utilization}} x_{ij}^k \right) \leq M \tag{1}$$

The introduction of side constraints fundamentally alters the mathematical nature of the problem. Equations linking multiple commodities such as Equation (1) break the TU property of the underlying network matrix. Consequently, the problem loses the guarantee of natural integer solutions and can no longer be solved using highly specialized, fast network algorithms. Because distinct assets must compete for shared financial limits, the physical routing of inventory effectively transforms into a generalized resource allocation problem. To manage this, the model transitions into a LP, ILP or MILP framework, requiring robust solvers capable of handling coupled constraints and binary variables (Ahuja et al., 1993, p. 649-653).

3.3.2. Linear programming

Linear programming is a mathematical method used to analyze resource allocation problems in which limited inputs must be allocated across activities in order to maximize or minimize an objective function. The model assumes linear relationships between variables and that parameters are known and constant. Under these assumptions, the objective function and constraints can be expressed as linear equations or inequalities (Panik, 2019, p. 21-23).

Canonical form The basic structure of a linear programming problem can be written in *canonical form* as

$$\begin{aligned} \max_{x_1,\ldots,x_p} \quad & f = c_1x_1 + c_2x_2 + \cdots + c_px_p \\ \text{s.t.} \quad & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \leq b_1 \\ & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \leq b_2 \\ & \vdots \\ & a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mp}x_p \leq b_m \\ & x_1, x_2, \ldots, x_p \geq 0 \end{aligned}$$

The inequalities define the structural constraints while the non-negativity conditions restrict the decision variables to non-negative values. Together, they define the feasible region containing all admissible solutions.

In matrix notation, the problem can be written as

$$\begin{aligned} \max_{X} \quad & f = C^\top X \\ \text{s.t.} \quad & AX \leq b \\ & X \geq 0 \end{aligned}$$

A vector X satisfying these constraints is called a feasible solution. A feasible solution X_0 is optimal if $f(X_0) \geq f(X)$ for all other feasible solutions X.

Standard form Instead of expressing the structural constraints as inequalities, they can be converted into equalities by introducing non-negative *slack variables*. For each inequality constraint, a slack variable is added to represent unused capacity. The resulting formulation is called the *standard form* of a linear program:

$$\begin{aligned} \max_{X, X_s} \quad & f = c_1 x_1 + \cdots + c_p x_p + 0 x_{p+1} + \cdots + 0 x_{p+m} \\ \text{s.t.} \quad & AX + I_m X_s = b \\ & X \geq 0 \\ & X_s \geq 0 \end{aligned}$$

where X_s is the vector of slack variables and I_m is the $m \times m$ identity matrix. Defining

$$\bar{X} = \begin{bmatrix} X \\ X_s \end{bmatrix}, \quad \bar{C} = \begin{bmatrix} C \\ 0 \end{bmatrix}, \quad \bar{A} = [A \; I_m],$$

the standard form can be compactly written as (Panik, 2019, pp. 52–53):

$$\begin{aligned} \max_{\bar{X}} \quad & f = \bar{C}^\top \bar{X} \\ \text{s.t.} \quad & \bar{A}\bar{X} = b \\ & \bar{X} \geq 0. \end{aligned}$$

3.3.3. Integer Programming and LP Relaxation

A standard LP formulation assumes that decision variables are continuous, meaning they can take any fractional value within the feasible region. However, many real-world optimization problems require variables to represent indivisible quantities. When integrality constraints are imposed on some or all of the decision variables, the model is classified as an ILP or a MILP (Wolsey, 1998, p. 1–3).

The introduction of integrality constraints fundamentally alters the mathematical properties and the computational complexity of the problem. While continuous LP problems can be solved efficiently in polynomial time, ILP problems are generally NP-hard. This implies that the computational effort required to find and verify an optimal integer solution can grow exponentially with the size of the problem (Wolsey, 1998, p. 12–14).

A foundational technique to manage this computational complexity is to formulate the continuous LP relaxation of the integer problem. The LP relaxation is obtained by simply dropping the integrality constraints on the variables. Formally, a requirement that a variable x_j must be a non-negative integer ($x_j \in \mathbb{Z}^+$) is relaxed to a standard continuous non-negativity constraint ($x_j \geq 0$) (Wolsey, 1998, p. 40).

Because the feasible region of the relaxed problem is a larger superset that includes all integer solutions, the optimal objective value of the LP relaxation always provides an *optimistic estimate* of the true integer optimum. Specifically, the relaxation yields a valid upper bound for maximization problems and a valid lower bound for minimization problems. Since the solver searches across a less constrained solution space, it establishes an idealized mathematical benchmark that the true discrete system can at most equal, but never exceed. While this optimistic estimate is a powerful analytical tool, it also represents a practical limitation, as the resulting fractional solutions may be overly idealized and not directly implementable in a real-world discrete system (Wolsey, 1998, p. 91).

3.3.4. The simplex method

The simplex method is an algorithm for solving linear programs in standard form. It relies on the geometric property that the feasible region has a finite number of extreme points, each corresponding to a basic feasible solution. Because the objective function is linear, the Fundamental Theorem of Linear Programming guarantees that if an optimal solution exists, at least one optimal solution must be an extreme point (Panik, 2019). Therefore, the algorithm can find the optimum by searching exclusively among these vertices.

Starting from an initial basic feasible solution, the method moves from one basic feasible solution to an adjacent one, improving or maintaining the objective function at each step. This process continues until no further improvement is possible, at which point an optimal solution has been reached.

The decision of whether the current solution is optimal is determined by the so-called optimality evaluators, which measure how the objective function would change if a nonbasic variable were to enter the basis. If none of these values indicate an improvement of the objective, the current solution is optimal. Otherwise, the algorithm performs a pivot operation to move to a new basic feasible solution (Panik, 2019, p. 60–65). The mathematical steps governing these pivot operations are formalized in Appendix A.

3.3.5. The dual simplex method

While the primal simplex method maintains primal feasibility throughout the iterations and improves the objective value until optimality is reached, the dual simplex method proceeds in the opposite way. It starts from a solution that satisfies the optimality conditions (dual feasibility) but may violate the nonnegativity constraints of the basic variables.

Starting from this dual feasible but primal infeasible solution, the algorithm performs a sequence of pivot operations that restore primal feasibility while preserving dual feasibility. At each iteration, a basic variable that violates the nonnegativity constraint is selected to leave the basis, and a suitable nonbasic variable enters the basis in a way that maintains the optimality condition.

The process continues until all basic variables satisfy the nonnegativity constraints. At that point, both primal and dual feasibility hold simultaneously, implying that the solution is optimal for both the primal and dual problems (Panik, 2019, p. 115–117).

The main steps of the dual simplex method can be summarized as follows (Panik, 2019, p. 120–122):

Algorithm (Dual Simplex Method)

1. **Initialize.** Start from a basic solution that satisfies the dual feasibility (optimality) conditions.
2. **Check primal feasibility.** If all variables in the current solution are nonnegative, the solution is optimal.
3. **Select leaving variable.** Choose a basic variable with a negative value to leave the basis.
4. **Select entering variable.** Choose a nonbasic variable that restores feasibility while preserving dual feasibility.
5. **Pivot and iterate.** Perform a pivot operation and repeat the procedure until all basic variables are nonnegative.

While the steps above outline the general framework of the algorithm, a more formal and detailed mathematical representation of the dual simplex method, including the specific criteria for pivot selection, is provided in Appendix A.

3.4. Backtesting and Comparative Analysis

In the fields of operations research and financial modeling, backtesting is a quantitative framework used to empirically validate the performance and robustness of a mathematical model using historical data. Instead of deploying an untested model directly into a live operational environment, backtesting reconstructs past states to observe how the system's decision rules would have behaved under historical conditions. To ensure the mathematical validity of the simulation, the framework must strictly respect the chronological availability of information. The model is restricted to utilizing data that was historically known at each specific simulated point in time, thereby preventing look-ahead bias and ensuring that the simulated

performance is a realistic representation of real-world deployment (Harvey and Liu, 2015).

Furthermore, backtesting provides the foundation for comparative analysis, where the outputs of a proposed model are systematically benchmarked against an existing baseline framework. By executing the proposed model on identical historical inputs, researchers can isolate the marginal impact of the altered decision logic. Comparing the simulated performance metrics directly against the historical outcomes provides an objective, data-driven assessment of whether the alternative framework yields statistically or economically meaningful improvements.

3.5. Statistical Dependencies in Financial Time-Series

This subsection examines the presence of temporal dependencies in financial time-series data, where current observations are influenced by past values. Such behavior violates the classical assumption of independently and identically distributed (i.i.d.) observations and has important implications for statistical modeling and inference.

3.5.1. Autocorrelation and the i.i.d. Assumption

Autocorrelation arises when the residuals of a regression model are correlated across time. In contrast to cross-sectional data, time series observations are often influenced by prior observations, meaning that values observed at different points in time are not independent. This is common in economic and financial data, where current outcomes are affected by past values.

The presence of autocorrelation violates a key assumption of classical linear regression, namely that the error terms are independent. While Ordinary Least Squares (OLS) estimates remain unbiased, autocorrelation leads to inefficient estimates and biased standard errors. As a result, standard hypothesis tests, such as t-tests and F-tests, may produce misleading conclusions.

Consequently, the assumption that observations are i.i.d. is generally not valid in time series settings, motivating the use of inference methods that are robust to temporal dependence (Sanghro, 2024).

3.5.2. Testing for Autocorrelation

To assess whether the assumption of independent errors is violated, both visual diagnostics and formal statistical tests are commonly used to detect autocorrelation.

ACF Plots A common graphical tool for detecting autocorrelation in time series data is the Autocorrelation Function (ACF). The ACF measures the correlation between observations at time t and their lagged values $t - k$, formally defined as

$$\rho_k = \mathrm{Corr}(y_t, y_{t-k}), \quad k = 1, 2, \ldots$$

where k denotes the lag. By plotting the estimated correlations against different lags, an ACF plot provides a visual diagnostic of temporal dependence. In the absence of autocorrelation, the plot should exhibit no systematic pattern, with correlations close to zero. In contrast, large values or persistent patterns across lags indicate that observations are correlated over time, suggesting that past values influence current outcomes and that the independence assumption is violated (Pennsylvania State University, 2026a).

PACF Plots While the ACF measures the total correlation between y_t and y_{t-k}, the Partial Autocorrelation Function (PACF) is used to measure the specific association between these two variables by filtering out the linear influence of the random variables that lie in between (i.e., $y_{t-1}, y_{t-2}, \ldots, y_{t-(k-1)}$). This process requires a transformation of the time series, after which the correlation of the transformed series is calculated to obtain the partial autocorrelation.

The PACF is most useful for identifying the appropriate order of an autoregressive model. Sample partial autocorrelations that are significantly different from zero indicate specific lagged terms of y that serve as useful predictors of y_t. However, it is important that the choice of order indicated by the PACF remains logically sound within the context of the data being modeled (Pennsylvania State University, 2026a).

The Durbin–Watson Test In addition to graphical methods, formal statistical tests can be used to detect autocorrelation. A formal test for first-order autocorrelation is the Durbin-Watson (DW) test, which evaluates whether the regression residuals are correlated across adjacent time periods. The test is based on the hypotheses $H_0 : \rho = 0$ against $H_A : \rho \neq 0$, where ρ denotes the autocorrelation of the error terms. The test statistic is given by

$$D = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2},$$

where e_t are the OLS residuals. The statistic typically takes values between 0 and 4, with values close to 2 indicating no autocorrelation, while values substantially below or above 2 suggest positive or negative autocorrelation, respectively. This test provides a formal complement to ACF-based diagnostics when assessing the validity of regression assumptions (Pennsylvania State University, 2026b).

3.5.3. The Augmented Dickey-Fuller test

To assess whether a time series is stationary, the Augmented Dickey-Fuller (ADF) test is commonly used. The test evaluates the null hypothesis that the series contains a unit root (i.e., is non-stationary) against the alternative of stationarity.

The ADF test is based on the regression

$$\Delta y_t = \alpha + \beta y_{t-1} + \delta t + \sum_{j=1}^{k} \zeta_j \Delta y_{t-j} + \varepsilon_t,$$

where Δy_t is the first difference of the time series y_t, α is a constant intercept, β is the coefficient testing for a unit root, δ is the coefficient on a deterministic time trend t, k is the lag order, ζ_j are the coefficients for the lagged differences, and ε_t is the white noise error term. The lagged differences are included to control for serial correlation. The null hypothesis $H_0 : \beta = 0$ corresponds to a unit root, while $\beta < 0$ implies stationarity (StataCorp, 2013).

3.5.4. HAC Robustness: The Newey–West Estimator

When the i.i.d. assumption is violated due to heteroskedasticity or autocorrelation, the standard variance formula for the OLS estimator is no longer consistent. In such cases, conventional standard errors become unreliable, leading to invalid statistical inference.

To address this, Heteroskedasticity and Autocorrelation Consistent (HAC) estimators are used to obtain a consistent estimate of the covariance matrix. A commonly used HAC estimator is the Newey–West estimator, which extends the standard approach by incorporating autocovariances of the error terms across different time lags.

By weighting these lagged covariance terms, the Newey–West estimator provides a consistent estimate of the variance-covariance matrix even when autocorrelation of unknown form is present. This allows for valid standard errors in time series regression models (Ao, 2009).

3.5.5. Time-Series Regression

A commonly used class of models in time-series analysis is the Autoregressive (AR) model. These models have a strong intuitive appeal, as they closely resemble the structure of a classical linear regression model. In a standard regression framework, the dependent variable is explained by a set of external predictors. In contrast, an AR model replaces these predictors with past (lagged) values of the time series itself.

Formally, an AR model of order p, denoted AR(p), expresses the current value of a time series as a linear function of its previous p observations. The model can be written as

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + Z_t,$$

where $\phi_1, \ldots, \phi_p$ are parameters and Z_t is an error term.

Because of their structural similarity to classical regression models, many statistical results developed for regression analysis can be extended to AR models with only minor modifications (Chan, 2010).

4. Methodology

The practical implementation of the optimization framework and the methodology used for its empirical evaluation are detailed in this section. It begins by outlining the structure and processing of the daily input data, including core inventory positions, market rates, and regulatory parameters. The mathematical optimization model is then formally defined, translating the theoretical concepts into explicit sets, parameters, objective functions, and constraints. Furthermore, the section details the computational implementation and the chosen solver algorithm. Finally, it describes the historical backtesting design and the regression-based evaluation framework utilized to analyze the model's economic performance and its underlying drivers.

4.1. Data

The optimization model is built on two primary categories of inputs: (i) daily end-of-day client position and inventory data, and (ii) daily market and regulatory parameters. Together, these inputs describe the supply of usable inventory and the demand created by client shorts, while providing the economic and regulatory conditions that determine the value and feasibility of different allocation choices.

4.1.1. Core Position and Inventory Data

The main source of data is Clear Street's daily end-of-day client position and inventory snapshots. These datasets describe long and short positions across securities and legal entities and form the foundation of the optimization problem.

For each trading day, the data are aggregated to the level of

$$(\text{security},\ \text{entity}),$$

and includes, for each security and entity, quantities representing:

- **Long inventory:** securities available from client longs and firm inventory where applicable.
- **Short demand:** client short positions that must be covered for settlement.

This position-driven view defines the feasible allocation universe on each day: inventory must be allocated across competing uses such as internalization and external lending.

4.1.2. Market Data and Rates

To evaluate the economic trade-offs of each allocation, the optimization model utilizes daily market data and interest rates. Crucially, the rates incorporated into the dataset reflect the U.S. market convention of netted rebate rates, which embed both the cash reinvestment rate and the intrinsic securities lending fee. The primary data inputs include:

- **Security Prices:** Utilized to convert share quantities into market values, which is strictly necessary for computing entity-level budgets and value-based cash constraints.
- **Street Borrowing Rates:** Representing the external sourcing revenue or costs generated when the prime broker must borrow securities from the market to cover short positions. As outlined in Section 2.3.1, this rebate rate can be either positive or negative, yielding interest on posted cash for GC or incurring direct fees for Specials, based on the mechanics detailed in Section 2.2.6.
- **Street Lending Rates:** Representing the external funding revenues or costs generated when the prime broker pledges long inventory to the market. As established in Section 2.3.2, the quoted rebate rate can similarly be positive or negative, incurring an interest expense on cash collateral received for GC, or generating direct revenue when lending scarce Specials at negative rates, strictly governed by the mechanics detailed in Section 2.2.6.
- **Client Billing Rates:** Representing the revenues or costs generated from clients executing short sales. Reflecting the external market, this client rebate rate can also be positive or negative, incurring an interest expense when the broker pays the client on their generated cash balances for GC shorts, or generating direct revenue by charging premium fees to clients shorting scarce Specials.
- **Benchmark Funding Rates:** Such as the OBFR, which serves as the baseline funding cost.

4.1.3. Regulatory and Operational Constraints

The optimization model is bounded by both regulatory limits on client assets and operational limits on market liquidity. The key regulatory input is *extended customer margin*, which determines how much client collateral can legally be reused for funding, while operational caps ensure realistic market depth.

- **Entity-level rehypothecation capacity** is calculated from customer margin and defines a per-entity upper bound on how much collateral can be reused for internalization or funding.
- **Reserve lock-up requirement** arises when the value of collateral deployed exceeds the available margin base. The excess generates a cash reserve requirement, reducing the amount of capital available for other uses.
- **Security-level lending caps** act as an operational control, restricting the funding volume per security to the maximum amount historically loaned out in a single day over the previous year. This ensures the model does not assume unrealistic street liquidity for any single asset.

While customer margin acts as the central regulatory constraint linking inventory availability and funding capacity, the security-level lending caps ensure that the resulting allocations remain practically executable in the external market.

4.1.4. Allocation Data for Backtesting and Benchmarking

In addition to the model input data, Clear Street provides a daily dataset describing the realized position-level allocation decisions made by the current production process. This data contains observed allocations across entities, securities, and funding channels for each trading day (e.g., quantities internalized, borrowed or pledged for funding).
This allocation dataset is not a driver of the optimization inputs. Instead, it is used for backtesting and evaluation. The baseline allocations are treated as the benchmark solution against which the model-generated allocations are compared. This enables a day-by-day assessment of whether the optimization improves efficiency and economic outcomes relative to the firm's existing heuristic allocation method.

4.1.5. Data Processing

All inputs are aligned to a daily end-of-day frequency with quantities aggregated to entity-security level. Observations with no relevant supply or demand are excluded from the daily optimization universe. The resulting dataset provides, for each trading day, a consistent representation of (i) inventory supply, (ii) short demand, (iii) economic rates and prices, and (iv) regulatory feasibility bounds.

4.1.6. Input Data Example

To illustrate the granular structure of the dataset fed into the optimization model, Table 1 provides a representative cross-section for a single entity-security pair.

The data captures the financial capacity, available physical inventory, prevailing market rates, and operational lending limits required to evaluate the economic feasibility of an allocation on a given trading day.

Table 1: Representative input data for a single entity-security pair. The complete dataset contains one row for each unique client-security combination.

Security	Entity	Margin ($)	Budget ($)	Price ($)	Long Qty	Short Qty	Fee (%)	Borrow (%)	Loan (%)	OBFR (%)	Lend Cap
AAPL	1234	1,000,000	1,400,000	260	50,000	0	2.93	3.78	3.78	3.63	2,500,000

4.2. Modeling

This section presents the mathematical optimization model used to determine the daily allocation of prime brokerage inventory. The objective of the model is to maximize the economic value generated from client inventory while respecting regulatory constraints, inventory availability, and firm-level funding limitations.

4.2.1. Model Sets, Parameters, and Variables

Sets The allocation problem is structured over the following index sets, which map the dimensions of the client base, the available asset universe, and the tiered cost structure:

- $\mathcal{E}$ — set of client entities.
- $\mathcal{A}$ — set of securities (assets).
- $\mathcal{P} \subseteq \mathcal{E} \times \mathcal{A}$ — set of entity–security pairs for which positions exist.
- $\mathcal{K}$ — set of balance sheet capacity tiers for piecewise linear penalization.

Parameters To capture both customer and firm inventory dynamics, along with piecewise linear capacity costs, the model defines the following input parameters:

- $L_{e,a}^{\text{cust}}$ — customer long inventory of security a held by entity e.
- L_a^{firm} — aggregate firm long inventory of security a available for internal use.
- $S_{e,a}^{\text{cust}}$ — customer short inventory of security a held by entity e. Let $S_a^{\text{cust}} = \sum_e S_{e,a}^{cust}$ denote the aggregate customer short demand for asset a.
- S_a^{firm} — aggregate firm short demand for security a.
- M_e — margin balance (customer debit) of entity e.
- P_a — market price of security a.
- C_a — contract price used in securities lending transactions, strictly defined as 102% of market value ($C_a = 1.02P_a$).
- r_a^B — external borrow rebate rate for security a.
- r_a^L — external lending rebate rate for security a.
- $f_{e,a}^c$ — client borrowing fee for shorting security a.
- r^{OBFR} — benchmark overnight funding rate.
- s_m — spread applied to client margin balances.
- L_k — capacity limit (maximum volume) for balance sheet tier k.
- c_k — marginal cost (in percentage points) associated with utilizing balance sheet tier k.
- L_a^{cap} — maximum allowable lending quantity for security a, based on historical peak volumes.
- γ — rehypothecation ratio, equal to 1.4 under Regulation T.
- $\omega_{\text{hypo}}, \omega_{\text{firm}}$ — small positive tie-breaker penalties (e.g., 10^{-5}) to favor rehypothecation and internal firm short coverage when economically neutral.

Decision Variables The optimization engine allocates inventory and manages the firm's global cash pool through the following non-negative, continuous variables:

- $h_{e,a}$ — number of shares of security a rehypothecated from customer entity e.
- i_a — number of customer long shares internalized to cover shorts.
- i_a^{firm} — number of firm long shares of security a used for internal short coverage.
- b_a — number of shares of security a borrowed from the external street market.
- f_a — number of shares of security a pledged externally for funding.
- x_a — number of rehypothecated shares of security a that remain unused (boxed).
- c_a^{firm} — number of firm short shares of security a covered internally.
- T_{cash} — amount of residual unsecured treasury cash required to fund operations.
- I_{cash} — amount of excess cash generated and invested at the benchmark rate.
- R_{lockup} — mandated cash reserve lock-up required.
- V_{ext} — total gross market value of external balance sheet activity (borrowing and funding).
- v_k — market value of external balance sheet activity allocated to tier k.

4.2.2. Flow Architecture

To conceptually bridge the mathematical parameters and decision variables defined above, Figure 5 illustrates the complete flow architecture of the optimization model. The diagram maps the dual circulation of physical securities and capital. It highlights how long inventory is routed through the firm's central hubs to either cover internal short obligations or generate external funding, while simultaneously tracking the flowing impact of these decisions on the global cash pool.

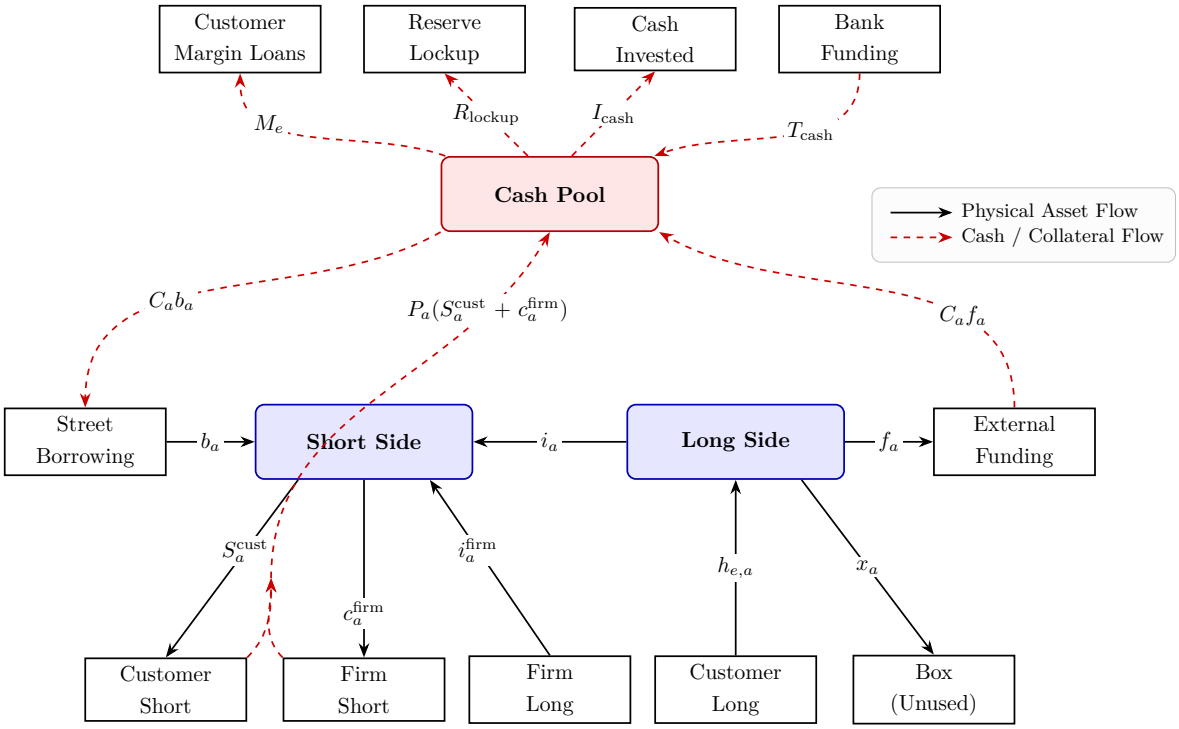


Figure 5: Conceptual flow architecture of the prime brokerage system. The framework maps how physical inventory (solid black lines) is routed from long sources to either cover short demand, fund externally, or remain unused. Crucially, the model captures how these physical routing decisions automatically trigger corresponding cash and collateral flows (dashed red lines), which govern the liquidity available in the central Cash Pool for margin loans, investments, and reserve requirements.

As illustrated by the dashed paths in Figure 5, physical inventory allocations directly generate corresponding cash flows. These cash flows are derived as the product of the allocated share quantities and their respective valuation prices:

- **Short Proceeds ($P_a(S_a^{\text{cust}} + c_a^{\text{firm}})$):** When the broker settles a short sale for a client or its own firm account, it receives cash equal to the market value of the sold securities (P_a). This acts as a substantial inflow of liquidity into the cash pool.
- **Collateral Received ($C_a\ f_a$):** When the broker pledges available long inventory to the external market to raise funding, it receives cash collateral in return. In U.S. markets, this collateral is typically sized at a standard 102% margin (C_a) of the market value of the pledged shares (f_a).
- **Collateral Posted ($C_a\ b_a$):** Conversely, when the broker must borrow shares from the external market to cover a short position, it must post cash collateral to the lender. This represents a significant liquidity outflow from the cash pool, again sized at the standard 102% margin (C_a) of the borrowed shares (b_a).

4.2.3. Economic Interpretation

The model fundamentally represents a daily source-and-use allocation problem within a global system of physical securities and cash liquidity. The prime broker holds a finite supply of scarce resources: rehypothecatable customer collateral and firm-owned inventory, which must be routed to satisfy mandatory internal short demand or pledged to external markets to generate funding.

The optimization balances several competing economic forces simultaneously. It must evaluate the scarcity premium of individual assets, such as the revenue generated from lending Specials, against the costs of executing those trades. These costs include external borrowing fees, internal balance sheet utilization charges modeled through tiered leverage costs, and the cash implication of triggering a regulatory reserve lock-up. Ultimately, the solver determines the exact routing of every asset for every entity to maximize the firm's net profitability, ensuring that all physical delivery obligations are met without violating the liquidity and regulatory constraints of the prime brokerage.

4.2.4. Objective Function

The objective is to maximize the daily net economic value (Π) generated from the allocation state. All annualized percentage rates are converted to daily terms by applying a standard $\frac{1}{360}$ day-count convention. To ensure mathematical clarity, the overall objective function is decomposed into five distinct economic components: net fees, balance sheet costs, structural tie-breakers, capital funding, and margin income.

$$\max \Pi = \frac{1}{360}\Big(\Pi_{\text{fee}} + \Pi_{\text{bs}} + \Pi_{\text{tie}} + \Pi_{\text{cap}} + \Pi_{\text{margin}}\Big) \tag{2}$$

(I) Fee Component This component captures the core PnL generated from physical inventory movements. It includes the net revenue from street borrowing (b_a) and funding (f_a), as well as the constant fee revenue collected from customer short positions (S_a^{cust}).

$$\Pi_{\text{fee}} = \sum_{a\in\mathcal{A}} \left[b_a C_a r_a^B + f_a C_a(-r_a^L) + \sum_{e\in\mathcal{E}} S_{e,a}^{cust} P_a(-f_{e,a}^c) \right]$$

(II) Balance Sheet Cost Component External transactions expand the firm's gross balance sheet, requiring regulatory capital allocation. In practice, the cost of leverage is generally considered non-linear; as the firm utilizes more external capacity and approaches regulatory or internal risk limits, the marginal cost of capital is expected to escalate. To capture this progressive penalty within a linear programming framework, a piecewise linear approximation is employed. The total external balance sheet volume is partitioned into a set of sequential tiers $\mathcal{K}$. Each tier has an escalating marginal cost c_k, ensuring that the solver naturally

exhausts cheaper base capacity before expanding further into more expensive capacity tiers.

$$\Pi_{\text{bs}} = -\sum_{k \in \mathcal{K}} v_k c_k$$

(III) Tie-Breaker Terms Small artificial premiums (ω_{hypo}, ω_{firm}) are introduced to guide the solver toward desirable secondary behaviors when multiple routing paths are economically equivalent. They encourage the model to maximize customer hypothecation ($h_{e,a}$) and prefer covering optional firm shorts (c_a^{firm}) over leaving inventory unused in the box, but their magnitudes are deliberately kept negligible to prevent any distortion of the primary economic objective.

$$\Pi_{\text{tie}} = \sum_{a \in \mathcal{A}} \left[\omega_{hypo} \sum_{e \in \mathcal{E}} h_{e,a} P_a + \omega_{firm} c_a^{firm} P_a \right]$$

(IV) Capital Component This term captures the cost of unsecured funding and the return on excess liquidity. Treasury borrowing (T_{cash}) incurs a funding cost at the OBFR, while excess invested cash (I_{cash}) earns interest at the same rate.

$$\Pi_{\text{cap}} = -r^{OBFR} T_{cash} + r^{OBFR} I_{cash}$$

(V) Margin Component The final component calculates the interest income generated from the margin balances extended to clients (M_e), incorporating a predetermined margin spread (s_m). Because the total margin extended is fixed for a given day, this term acts as a constant in the optimization problem but is included to ensure complete and accurate PnL reporting.

$$\Pi_{\text{margin}} = (r^{OBFR} + s_m) \sum_{e \in \mathcal{E}} M_e$$

4.2.5. Constraints

The optimization problem is constrained by a set of operational and regulatory requirements that govern how inventory and cash can be allocated. These constraints ensure feasibility and reflect the key mechanics of the prime brokerage environment.

Rehypothecation Limits and Availability Rehypothecation ($h_{e,a}$) is strictly capped by value (140% of margin, γ) and physical quantity available ($L_{e,a}^{\text{cust}}$):

$$\sum_{a:(e,a) \in \mathcal{P}} h_{e,a} P_a \leq \gamma M_e \qquad \forall e \in \mathcal{E}$$

$$h_{e,a} \leq L_{e,a}^{cust} \qquad \forall (e,a) \in \mathcal{P}$$

Inventory Flow and Exempt Limits All extracted customer inventory must be strictly accounted for across uses, and firm internalization (i_a^{firm}) cannot exceed firm supply (L_a^{firm}):

$$\sum_{e:(e,a)\in\mathcal{P}} h_{e,a} = i_a + f_a + x_a \qquad \forall a \in \mathcal{A}$$

$$i_a^{firm} \leq L_a^{firm} \qquad \forall a \in \mathcal{A}$$

Short Coverage Dynamics Customer shorts (S_a^{cust}) are mandatory delivery obligations and must be fully covered. Firm shorts (S_a^{firm}) are opportunistic and can only be covered up to their maximum demand without forcing external borrowing:

$$c_a^{firm} \leq S_a^{firm} \qquad \forall a \in \mathcal{A}$$

$$S_a^{cust} + c_a^{firm} = i_a + i_a^{firm} + b_a \qquad \forall a \in \mathcal{A}$$

$$b_a \leq S_a^{cust} \qquad \forall a \in \mathcal{A}$$

The restriction $b_a \leq S_a^{cust}$ explicitly prevents the broker from borrowing externally solely to cover internal firm shorts.

External Lending Caps To reflect realistic market depth and avoid over-allocating to the street, the volume of a security pledged externally for funding (f_a) is strictly bounded by its historically observed maximum daily lending volume (L_a^{cap}):

$$f_a \leq L_a^{\text{cap}} \qquad \forall a \in \mathcal{A}$$

Regulatory Cash Lockup If the total value of customer inventory deployed externally (f_a) or internally utilized (i_a) exceeds the aggregate customer margin pool, the difference creates a punitive cash reserve requirement (R_{lockup}):

$$R_{lockup} \geq \sum_{a\in\mathcal{A}} P_a(f_a + i_a) - \sum_{e\in\mathcal{E}} M_e$$

Balance Sheet Leverage and Tiering To dynamically evaluate the piecewise linear cost of balance sheet expansion, the total gross external volume (V_{ext}) must be calculated and distributed across the defined tiers. The total volume is the sum of all external borrowing and funding, valued at the respective contract prices (C_a):

$$V_{\text{ext}} = \sum_{a\in\mathcal{A}} C_a \left(b_a + f_a\right)$$

This aggregate volume is strictly partitioned into the specific tier variables v_k, each bounded by its respective capacity limit L_k:

$$V_{\text{ext}} = \sum_{k\in\mathcal{K}} v_k$$

$$v_k \leq L_k \qquad \forall k \in \mathcal{K}$$

Because the objective function maximizes profit (thereby minimizing the penalty term $-v_k c_k$), the linear solver intrinsically fills the capacity of the tiers sequentially from the lowest marginal cost to the highest.

Global Cash Balance Finally, the firm must maintain a globally balanced cash pool. The sources of cash (collateral received from funding, unsecured treasury funding T_{cash}, and cash proceeds from short sales) must equal the uses of cash (margin loans extended, collateral posted for borrowing, mandated lockups, and any uninvested excess cash I_{cash}):

$$\begin{aligned} &\sum_{a \in \mathcal{A}} C_a f_a + T_{cash} + \sum_{a \in \mathcal{A}} P_a (S_a^{cust} + c_a^{firm}) \\ = &\sum_{e \in \mathcal{E}} M_e + \sum_{a \in \mathcal{A}} C_a b_a + R_{lockup} + I_{cash} \end{aligned}$$

By tightly coupling physical inventory constraints with this global cash equation, the optimization framework dynamically evaluates the true funding cost of every allocation decision, utilizing R_{lockup} and the tiered capacities as regulatory and economic controls.

4.3. Implementation and Solver

The optimization model was implemented using AMPL, which allows optimization problems to be defined using algebraic expressions for sets, parameters, variables, objective functions, and constraints. For each trading day in the backtesting framework, the processed inventory and market data are provided as input to the AMPL model, which then computes the optimal allocation.

The model is solved using the *HiGHS* optimization solver. The problem is formulated as a LP, and *HiGHS* applies the dual simplex method to efficiently compute the optimal solution. It should be noted that this LP formulation serves as a continuous relaxation of the true discrete problem, given that the underlying allocation variables physically represent integer quantities of shares. However, because the primary objective of this thesis is to investigate whether there is potential value in an optimal allocation framework compared to the current heuristic approach, utilizing this continuous relaxation is highly appropriate. Furthermore, initial test runs with strict integer constraints showed nearly identical results but required significantly longer time to solve. Thus, the relaxation improves computational efficiency while still accurately capturing the economic potential of the system.

4.4. Evaluation Framework

This subsection describes how the optimization model is evaluated against the firm's current heuristic allocation process. The empirical evaluation is based on a historical backtesting framework, where both methods are applied to identical daily market environments and compared using the same economic metrics. The framework is designed to assess both the magnitude and statistical significance of performance differences, as well as to identify the

conditions under which the optimization model delivers the greatest improvements.

4.4.1. Backtesting Design

The evaluation is conducted as a historical backtest over a sequence of trading days $t = 1, 2, \ldots, T$. For each day, the optimization model is solved using the exact information that would have been available at that point in time, including daily positions, inventory, prices, rates, and regulatory constraints.

To formalize this comparative framework, let X_t^{opt} denote the optimal allocation matrix generated by the optimization model on day t, and let X_t^{heur} denote the corresponding realized heuristic allocation executed by the firm. Both allocations are evaluated under the same market inputs and pricing assumptions through the objective function $f(\cdot)$, previously formulated in Equation (2). Here, $f(\cdot)$ captures the net economic profitability of a given allocation state.

The daily economic improvement, or marginal value added by the optimization model, is defined as:

$$\Delta P_t = f(X_t^{opt}) - f(X_t^{heur}) \tag{3}$$

A positive value of ΔP_t implies that the optimization model yields a higher economic value than the heuristic benchmark on day t. Since the optimization is evaluated daily on historical data, this framework provides an estimate of the value the model could have generated had it replaced the heuristic process during the sample period.

4.4.2. Performance Driver Analysis

To address **RQ4**, the evaluation framework is extended to analyze which market and inventory conditions explain variation in model outperformance.

Since both allocation methods are evaluated under identical daily conditions, the resulting performance differences form a paired time series. However, as discussed in Section 3, financial time series typically exhibit autocorrelation and heteroskedasticity, violating the classical i.i.d. assumption.

To ensure valid statistical inference, all regression-based analyses are conducted using HAC standard errors. This provides consistent standard errors and valid hypothesis testing in the presence of autocorrelation and heteroskedasticity.

Let ΔP_t denote the daily performance difference. The theoretical relationship between performance and explanatory variables is modeled as

$$\Delta P_t = \beta_0 + \sum_{k=1}^{K} \beta_k Z_{k,t} + \varepsilon_t,$$

where $Z_{k,t}$ represents day-level characteristics derived from the market environment, inventory composition, and funding conditions, and ε_t is the error term. The parameters β_k represent the true marginal effect of each explanatory variable on the daily performance difference, holding other factors constant.

To account for persistence and dynamic dependence in performance differences, a lagged dependent variable ΔP_{t-1} is included in the specification. The initial set of explanatory variables included a broad range of market, inventory, and funding characteristics motivated by the economic structure of the allocation problem. To obtain a more interpretable model, a backward elimination procedure was applied, sequentially removing statistically insignificant variables based on their p-values.

The theoretical model is estimated using OLS, yielding the empirical estimates $\hat{\beta}_k$. These parameters represent the marginal effect of each explanatory variable on the daily performance difference, holding all other factors in the model constant. Based on these estimates, statistical inference is conducted by testing the null hypothesis $H_0 : \beta_k = 0$ for each explanatory variable, where rejection indicates that the corresponding variable has a statistically significant association with model outperformance, holding all other variables constant. All statistical significance is evaluated using HAC standard errors.

This regression framework provides an interpretable mapping between economic conditions and model performance, allowing identification of the primary drivers of optimization gains.

4.4.3. Interpretation

Taken together, the evaluation framework addresses two core empirical questions. Firstly, it evaluates the economic performance of the optimization model relative to the heuristic benchmark (**RQ3**). Secondly, it identifies the conditions under which these improvements are strongest (**RQ4**).

By combining backtesting, statistical inference, and regression-based analysis, the framework provides both a robust performance comparison and a deeper understanding of the underlying economic mechanisms.

5. Results and Discussion

The empirical results of the backtesting framework and an evaluation of the proposed optimization model's performance relative to the current heuristic allocation approach are presented in this section. The analysis is structured in three parts.

Firstly, the overall economic performance of the model is assessed by comparing daily objective values and relative improvements. Secondly, the underlying allocation mechanisms are examined through key operational metrics, including internalization efficiency and external funding behavior. Finally, statistical analyses are conducted to evaluate the time-series properties of the performance differences and to identify the primary drivers of model outperformance.

5.1. Performance Overview

The primary performance metric is the daily outperformance of the optimization model. Figure 6 illustrates the total economic value of the objective function for both approaches, along with the daily improvement.

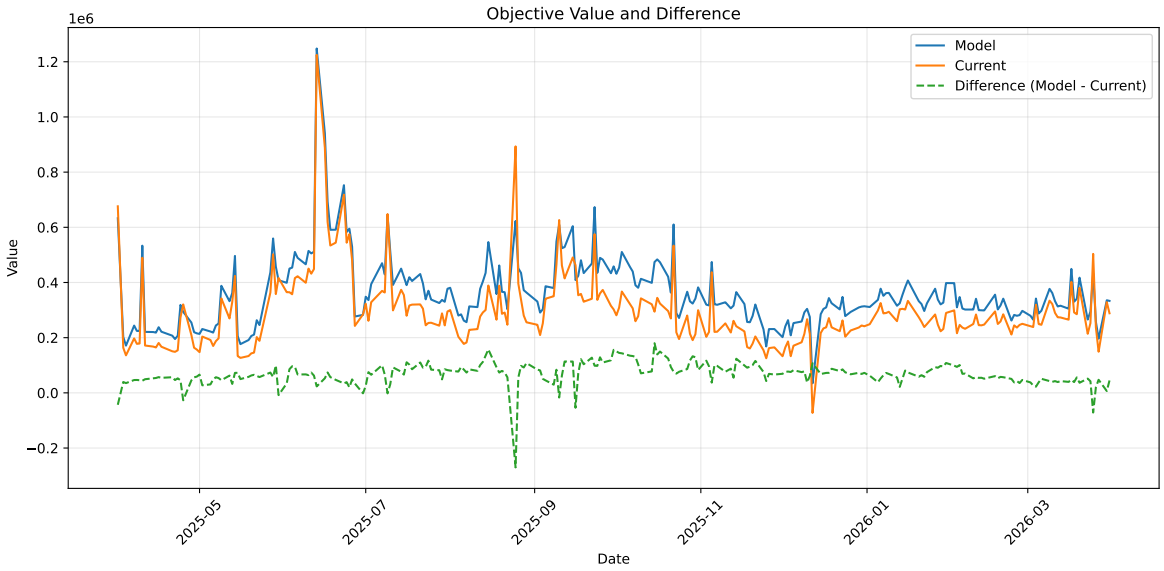


Figure 6: Daily economic value of the objective function for the Optimization Model vs. the Current Heuristic Approach (April 2025 – April 2026).

As shown in Figure 6, the optimization model consistently generates a higher objective value than the heuristic benchmark throughout the backtesting period, achieving an average increase of approximately **20.3%** in net economic value. While both series exhibit similar movements over time, reflecting the same exposure to underlying market conditions, the model maintains a persistent positive spread relative to the current approach. This indicates that the performance improvement is systematic rather than driven by individual events.

Despite this overall performance advantage, there are occasional days where the heuristic

benchmark outperforms the optimization model. In theory, such outcomes should not occur if both approaches are evaluated under the same constraints and inputs, as the optimization model is designed to identify the optimal feasible allocation. These deviations therefore likely reflect implementation frictions or data inconsistencies, and are examined in more detail in the discussion and conclusion section.

The daily performance difference, illustrated by the dashed line, remains positive for the majority of the sample but exhibits noticeable temporal structure. In particular, periods of strong outperformance tend to cluster, suggesting that the gains are not independent across time. Visually, the difference series appears to display autocorrelated behavior, where the magnitude of today's improvement is related to that of previous days. This observation motivates the formal time-series diagnostics conducted later in this section.

Overall, the figure provides strong visual evidence that the optimization framework delivers consistent economic gains while closely tracking the underlying market dynamics captured by the benchmark.

5.2. Allocation Mechanisms and Efficiency

While the aggregate profit and loss results demonstrate the model's economic superiority, the underlying drivers of this outperformance are found in the specific routing of inventory. In a prime brokerage environment, value is not generated simply by increasing volume, but by optimizing the source–use pairings to minimize explicit costs and shadow balance sheet charges. This section analyzes the physical and cash flows that differentiate the optimization framework from the heuristic benchmark.

5.2.1. Rehypothecation Budget Utilization

The model's capacity to generate value is fundamentally bounded by the regulatory rehypothecation limit. Figure 7 illustrates the total market value of customer collateral rehypothecated from client entities.

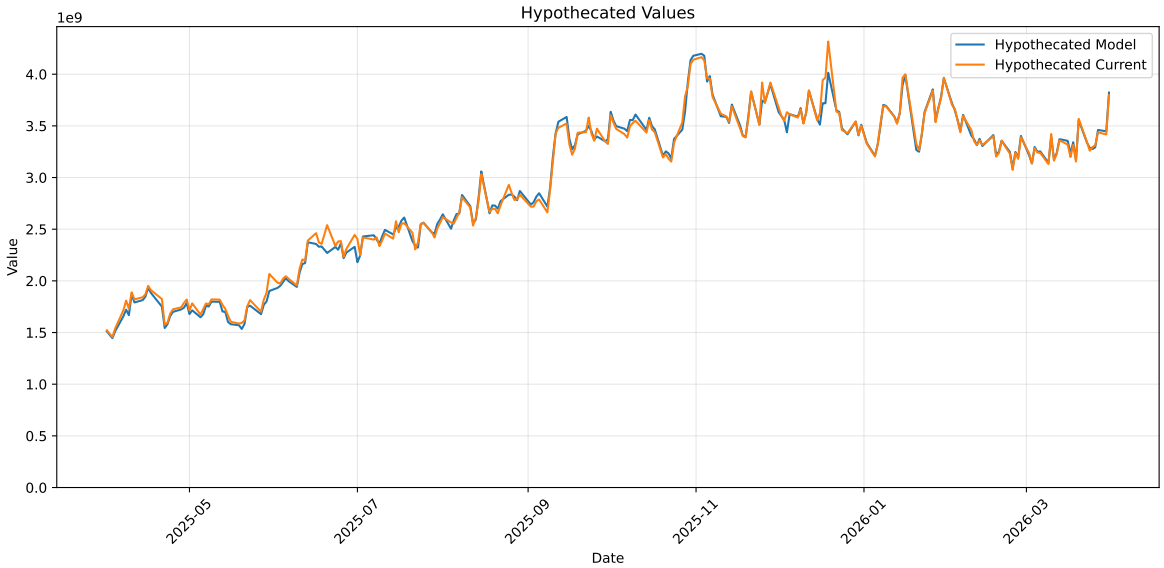


Figure 7: Total Market Value of hypothecated assets over the backtesting period. The plot compares the aggregate value of rehypothecated inventory between the optimization model and the current heuristic allocation.

As observed, the optimization model tracks the heuristic benchmark utilization of the rehypothecation budget. The solver systematically pulls all legally available collateral into the firm's central pool, making it available for reuse. This increases the probability of finding internal matches for short positions and reduces the firm's overall reliance on the external market.

Interestingly, Figure 7 highlights specific days where the heuristic benchmark appears to exceed the optimization model's value. In theory, these instances of outperformance should not occur if both the optimization model and the current heuristic approach were restricted by the same rehypothecation budget. However, these slight deviations have been traced back to an intraday data mismatch: the margin budget data reflects start-of-day balances, while the positions and realized allocations reflect end-of-day states. These anomalies, which are a byproduct of the backtesting environment rather than a flaw in the model's logic, are discussed in more depth in Section 6.

5.2.2. Internalization Efficiency

Internalization serves as a critical efficiency mechanism within a prime broker, functioning as the primary tool for both explicit cost avoidance and balance sheet optimization. By matching a client's long position with another client's short position in the same security, the broker eliminates the need to source that security from the external lending market. This provides significant economic benefits; firstly, it avoids the explicit borrowing fees and negative rebates associated with external sourcing, and secondly, it reduces the expansion of the balance sheet by bypassing the over-collateralization requirements (typically 102%) required when borrowing from the street.

Figure 8 illustrates the daily total market value of internalized assets, comparing the optimization model against the heuristic benchmark.

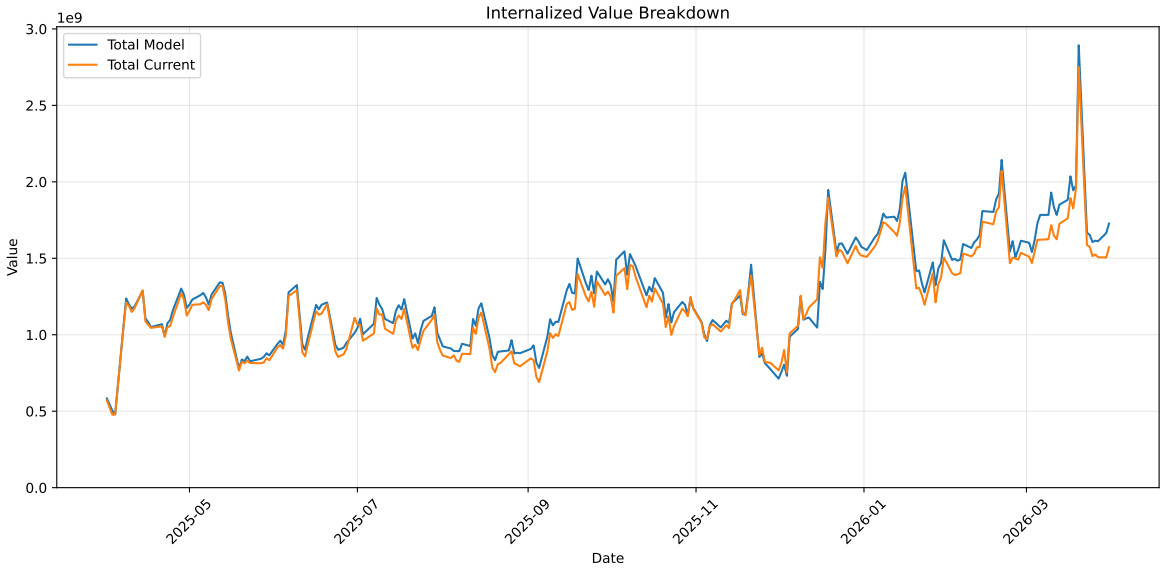


Figure 8: Total Market Value of internalized inventory. This metric represents the aggregate value of delivery obligations settled via internal matching rather than external borrowing.

As observed in Figure 8, the total market value of internalized inventory is remarkably consistent between the two approaches, although the optimization model achieves a modest increase of approximately **4.1%** on average. This similarity in overall volume suggests that the capacity for internal matching is largely bounded by the physical overlap of client positions and the 140% rehypothecation limit, rather than the specific allocation logic employed.

However, when viewed together with the economic performance noted in Section 5.1, these results indicate that the model's value is derived from selection quality rather than strict volume. While the heuristic approach may internalize securities indiscriminately, the optimization model utilizes the objective function to prioritize the internalization of high-fee Specials. By keeping the most expensive-to-borrow securities inside the firm, if there is a short demand, and routing lower-cost GC to the street, the model achieves a higher rate of cost avoidance for the same total value of internalized assets. This demonstrates that the optimization framework provides "smarter" inventory routing within the existing physical constraints of the portfolio.

5.2.3. External Borrowing and Funding Strategy

To analyze the impact on balance sheet composition and external financing, we examine the total market value of street borrowing and external funding alongside the resulting net cash position. Street borrowing (b_a) consumes firm liquidity as cash collateral must be posted to external lenders, while external funding (f_a) involves pledging long inventory to raise cash.

Efficient inventory management requires minimizing expensive external borrows while strategically utilizing available collateral for funding only when market rates are favorable relative to the firm's internal funding costs.

Figure 9 displays the daily total market value of securities borrowed externally versus those utilized for external funding, comparing the optimization model against the heuristic benchmark.

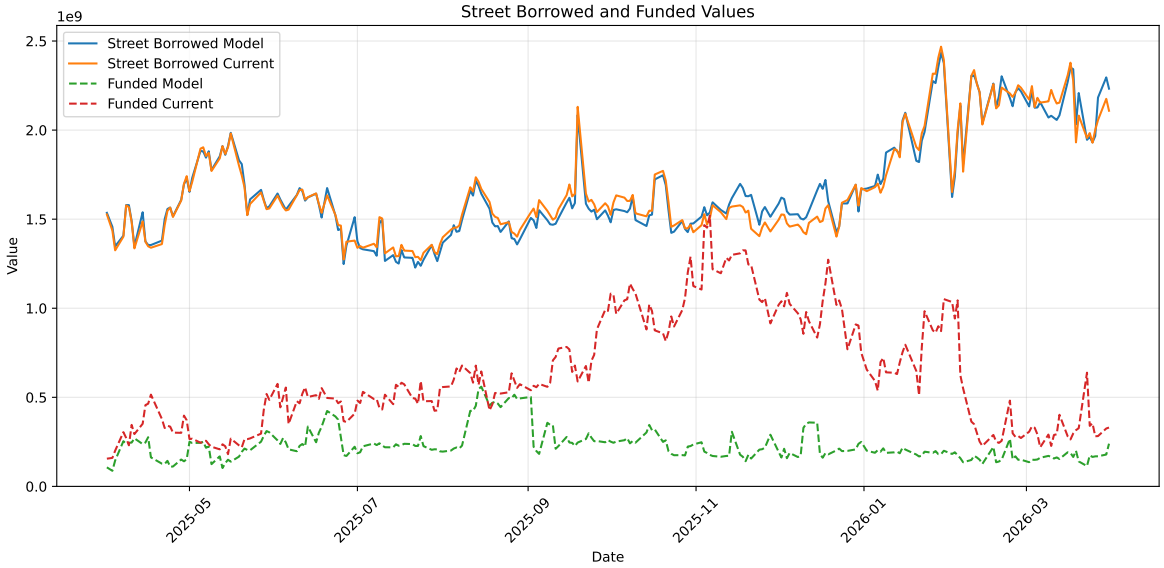


Figure 9: Total Market Value of external inventory flows. The plot compares the aggregate value of securities sourced from the street versus those pledged for external funding.

As illustrated in Figure 9, the total market value of street borrowing for the optimization model follows the current heuristic approach closely, suggesting that both approaches require similar gross volumes to satisfy client short demand. However, a distinct difference is observed in the external funding volumes, particularly during the period from September 2025 to March 2026. While the current approach exhibits significant volatility and a notable spike in funding during this window, the optimization model maintains a much more stable funding level. This suggests a more systematic evaluation of whether pledging collateral is economically justified relative to other funding sources. This stability in asset allocation is directly linked to the firm's broader liquidity management and cash dynamics.

To further understand the economic trade-offs driving these allocation decisions, we examine the firm's net cash position as illustrated in Figure 10. This metric represents the net liquidity remaining after accounting for all cash-consuming activities, such as funding client margin loans, posting collateral for borrowed securities, and satisfying reserve requirements. Effectively, it indicates whether the firm has a cash surplus to reinvest or a deficit that must be covered through alternative treasury funding.

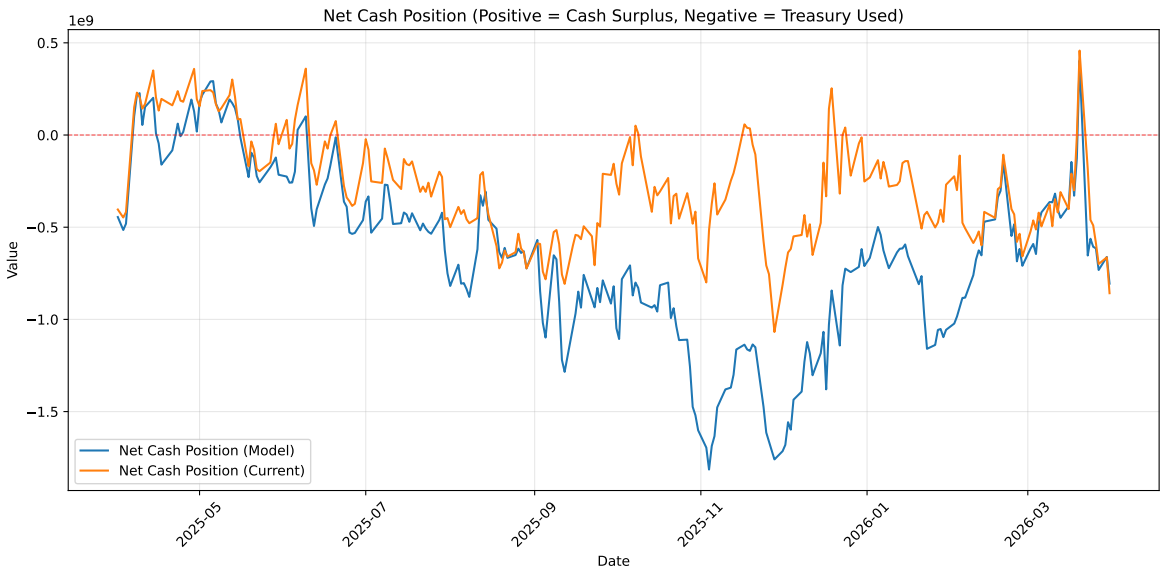


Figure 10: Net cash position over the backtesting period. Positive values indicate excess cash reinvested, while negative values represent treasury usage.

The relationship between these two metrics is clarified by the net cash position shown in Figure 10. During the period from September 2025 to March 2026, where the current approach utilized more external funding, the optimization model instead relied more heavily on unsecured treasury funding, reflected in a deeper negative net cash position. This indicates that the model dynamically evaluates the marginal cost of funding alternatives; it avoids pledging securities when external lending rates are unfavorable, opting instead for treasury usage to minimize funding costs. This behavior further illustrates the model's ability to dynamically select the most cost-efficient funding strategy.

5.3. Time-Series Diagnostics and Robustness

Before conducting formal hypothesis testing on the economic improvements generated by the optimization model, it is necessary to examine the statistical properties of the daily performance data. Standard inferential statistics assume that observations are independent and identically distributed (i.i.d.). However, financial time-series, such as daily PnL and allocation deltas, often violate these assumptions through temporal dependence and non-constant variance.

From this point onward, the analysis is based on the daily performance difference, ΔP_t, as defined in Equation (3). This measure captures the incremental economic value generated by the optimization model relative to the heuristic benchmark.

5.3.1. Autocorrelation Analysis

To evaluate whether autocorrelation is present, the DW test was applied to the daily performance differences ΔP_t. A DW statistic of **0.2845** was obtained, indicating strong positive first-order autocorrelation.

This finding is reinforced by the ACF and PACF diagnostics shown in Figure 11. The ACF exhibits a very slow, persistent decay, with significant positive autocorrelation across a large number of lags, suggesting a high degree of temporal dependence in the series. In contrast, the PACF shows a dominant spike at lag 1, while other lags are relatively small and largely within the confidence bounds.

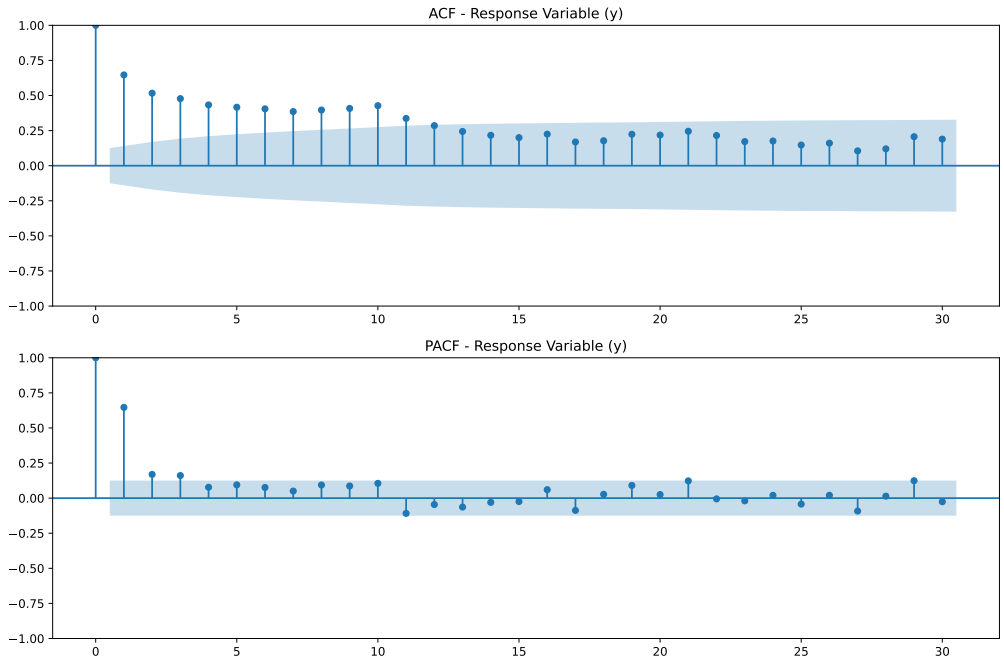


Figure 11: ACF and PACF functions for the daily performance differences ΔP_t. The ACF shows a slow decay with significant positive autocorrelation across multiple lags, while the PACF displays a dominant spike at lag 1, indicating strong first-order dependence.

5.3.2. Stationarity Analysis

To assess whether the performance difference series ΔP_t is stationary, an ADF test was conducted. The test yields a statistic of **-5.612** with a corresponding p-value **less than 0.01**, allowing for the rejection of the null hypothesis at both the 5% and 1% significance levels.

This indicates that the series ΔP_t is stationary, implying that its statistical properties, such as mean and variance, remain stable over the backtesting period.

5.4. Drivers of Performance

This section investigates which market and inventory characteristics explain when the optimization model outperforms the heuristic benchmark. To this end, a regression framework is employed where the dependent variable is the daily performance difference, ΔP_t, as established in Equation (3).

The model is estimated using OLS with a backward elimination procedure to retain only

statistically significant explanatory variables. To account for the strong persistence identified in the time-series diagnostics, a lagged dependent variable, ΔP_{t-1}, is included.

5.4.1. Regression Results

The estimated coefficients and key model statistics are summarized in Table 2. The final specification explains a substantial share of the variation in performance differences, with an R^2 of **0.619**. All retained explanatory variables are statistically significant at the 1% level, holding all other variables constant, and economically interpretable. All standard errors are computed using HAC estimators with three lags, ensuring robust inference given the time-series properties of the data.

Table 2: Regression Results: Drivers of Performance

Variable	Coefficient	Std. Error	p-value	Significance
Intercept	-35,380	18,400	0.055	*
ΔP_{t-1}	0.2172	0.073	0.003	***
VW Avg. Loan Rate	-16,040	3,251	0.000	***
VW Avg. Borrow Rate	17,830	6,416	0.005	***
Total Margin Extended	3.339e-05	6.30e-06	0.000	***

Notes: HAC (Newey-West) standard errors with 3 lags are reported. *** and ** denote significance at the 1% and 5% levels, respectively.

The model includes the following key drivers:

- **Lagged performance (ΔP_{t-1}):** The coefficient of 0.2172 indicates strong persistence in the performance difference. A substantial share of the model's outperformance carries over from one day to the next, suggesting that the benefits of optimized allocations tend to persist for several days rather than being limited to a single period.
- **Volume-weighted average lending rate:** The estimated coefficient is negative and highly significant. Under the adopted sign convention, more negative lending rates correspond to higher lending income. The result therefore implies that the optimization model performs better when lending opportunities are more profitable, and worse when lending becomes relatively costly.
- **Volume-weighted average borrowing rate:** The positive coefficient indicates that higher borrowing rates are associated with increased model outperformance. Given that positive borrowing rates correspond to income and negative rates to costs, this suggests that the model is particularly effective in environments where borrowing conditions are favorable.
- **Total margin extended:** The positive and statistically significant coefficient indicates that higher levels of margin lending are associated with greater performance differences.

Since total margin extended reflects the amount of financing provided to clients, it also determines the size of rehypothecation budgets. Higher margin levels therefore allow a larger share of client collateral to be reused, increasing the flexibility to reallocate and internalize securities. This suggests that the model's relative performance tends to be stronger in environments where more collateral is available for reuse.

The intercept is negative and not statistically significant at the 5% level ($p = 0.055$), suggesting that the baseline model outperformance is primarily explained by the active explanatory variables rather than a constant fixed effect.

5.4.2. Performance Drivers Model Validation

The validity of the regression model is assessed using both goodness-of-fit measures and residual diagnostics. The model achieves an R^2 of **0.619**, indicating that approximately 62% of the variation in the performance difference ΔP_t is explained by the included explanatory variables.

To evaluate the model specification, residual diagnostics are examined. Figure 12 presents the residuals plotted against fitted values. The residuals are generally centered around zero with no clear systematic pattern, suggesting that the linear functional form provides a reasonable approximation of the relationship between the dependent and explanatory variables. The distribution of observations is more concentrated at higher fitted values, but the spread of residuals in this region remains relatively stable. At lower fitted values, a smaller number of observations exhibit greater dispersion, including a few notable outliers.

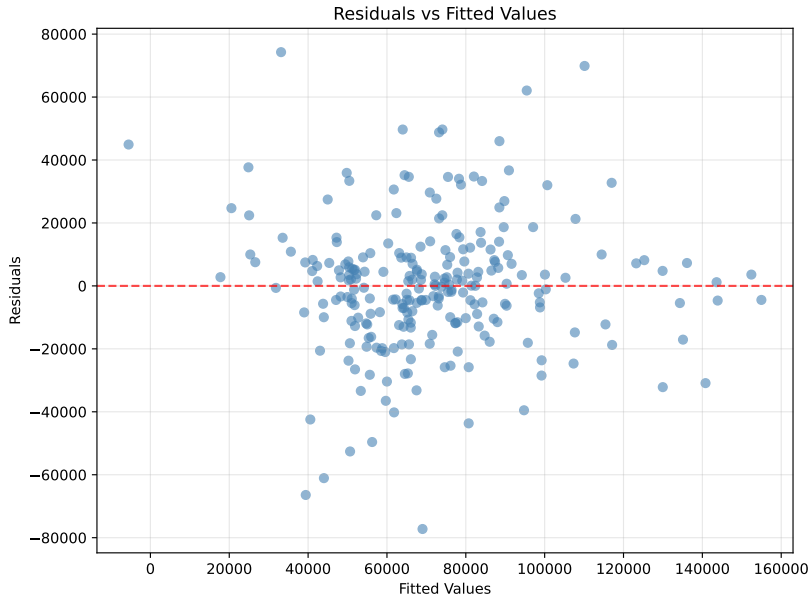


Figure 12: Residuals versus fitted values for the final regression model. Residuals are centered around zero with no clear systematic pattern. A few observations at lower fitted values display higher dispersion, while the spread remains relatively stable for larger fitted values.

Overall, there is limited evidence of strong heteroskedasticity, although some mild variation in dispersion is present. In addition, earlier time-series diagnostics identified autocorrelation in the raw performance series. To account for both potential heteroskedasticity and serial dependence, HAC standard errors are employed throughout the analysis.

To further assess remaining temporal dependence, both the ACF and PACF of the residuals are examined in Figure 13. The autocorrelations and partial autocorrelations are generally small and lie within acceptable bounds across all lags, indicating that no meaningful correlation remains. The PACF closely mirrors the ACF, reinforcing the absence of any clear lag structure. Although the correlation at lag 3 slightly exceeds the confidence interval, the deviation is minor and does not suggest a systematic pattern.

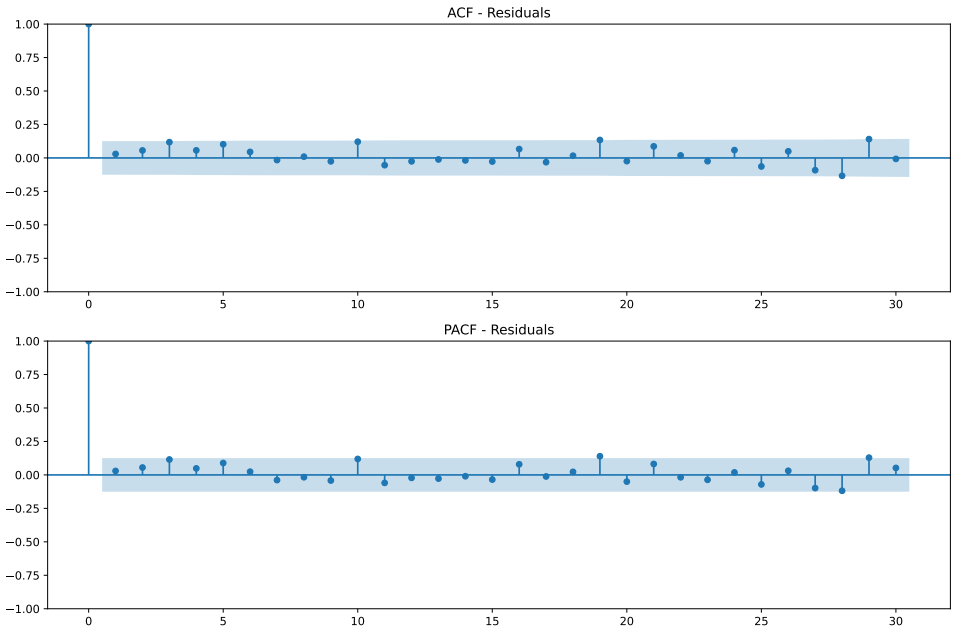


Figure 13: ACF and PACF of the regression residuals. The autocorrelations and partial autocorrelations are small and largely within the confidence bounds, indicating no material remaining serial dependence.

This is consistent with the inclusion of the lagged dependent variable, which captures persistence in the original series. The DW statistic of **1.80**, close to the benchmark value of 2, further supports the conclusion that first-order autocorrelation has been effectively addressed.

Finally, the stationarity of the residuals is evaluated using the ADF test. The test yields a statistic of **-13.81** with a corresponding p-value of **less than 0.01**, strongly rejecting the null hypothesis of a unit root. This indicates that the residuals are stationary, providing additional support for a well-specified model and reducing concerns that the observed relationships are driven by non-stationary behavior.

Overall, the diagnostics suggest that the model provides a well-specified representation of the data, with no major violations of the underlying assumptions affecting statistical inference.

6. Conclusion

This section summarizes the main findings, answers the research questions, and outlines both model limitations and future research opportunities.

6.1. Summary and Answers to Research Questions

This thesis has investigated the design, implementation, and empirical evaluation of a dynamic optimization model for collateral and inventory allocation within a prime brokerage environment. The primary focus was on transitioning from a heuristic, rule-based allocation process to a data-driven optimization framework. This framework was formulated to maximize net economic value by simultaneously evaluating the trade-offs between internalization, external lending, and borrowing, while strictly adhering to U.S. regulatory requirements.

Through historical backtesting and statistical analysis, the study evaluated how well the proposed model works in practice and its economic performance. These efforts were designed to answer the following research questions:

RQ1 How can prime brokerage inventory flows be modeled to balance the economic trade-offs between competing uses?

RQ2 How can complex regulatory mandates regarding asset segregation and rehypothecation limits be integrated as binding constraints within an optimization framework?

RQ3 To what extent does the proposed optimization model improve overall economic performance relative to the firm's current heuristic allocation approach?

RQ4 Which market and inventory characteristics explain when the optimization model outperforms the heuristic benchmark?

Regarding the modeling of inventory flows (**RQ1**), the study demonstrated that prime brokerage operations can be successfully formulated as a daily source-and-use resource allocation problem. Central to this formulation was the development of a comprehensive flow architecture, detailed in Section 4.2.2, that mapped the dual, interconnected circulation of physical securities and cash liquidity. By directly tying every physical inventory movement to its impact on the cash pool, the model provided a complete and unified view of the firm's resources. Using a continuous linear programming relaxation, this architecture allowed the optimization engine to efficiently quantify the economic trade-offs between competing uses, balancing explicit financing revenues against the implicit costs of capital and balance sheet utilization.

In addressing regulatory integration (**RQ2**), the framework translated U.S. mandates into binding mathematical constraints. Rehypothecation was capped via physical inventory limits and the 140% margin value threshold derived from Federal Reserve Regulation T. Furthermore, SEC Rule 15c3-3 requirements were incorporated by enforcing the segregation of fully paid and excess margin securities, ensuring that these assets are excluded from the reusable

inventory pool. In addition, the framework accounts for cash reserve requirements by capturing the economic impact of increased reserve balances when customer collateral usage exceeds regulatory thresholds.

The empirical backtesting confirmed the strong economic performance of the proposed framework (**RQ3**). The optimization model consistently generated higher daily objective values compared to the heuristic benchmark. It achieved this not by simply increasing the total volume of internalized or funded assets, but through more effective allocation decisions. By prioritizing the internal matching of high-cost Specials and routing lower-cost GC to external markets, the model systematically reduced costs and stabilized external funding levels, leading to more efficient use of its assets. These results should, however, be interpreted with some caution, as several limitations of the backtesting framework and data assumptions are discussed in the later section.

Finally, the time-series regression analysis identified the specific market conditions that drive this outperformance (**RQ4**). The results show that the model's relative economic value increases significantly with the total margin extended, holding other variables constant, indicating that optimization becomes more important as client debit balances grow.

From an operational perspective, this finding provides some indication of how the model may behave as the business scales. As a prime broker expands its client base, total margin extended typically increases, leading to larger client debit balances and a greater pool of rehypothecatable collateral. The results suggest that the economic benefits of an optimization-based allocation approach may increase in such settings. In other words, higher levels of client debits appear to be associated with greater potential gains from optimization.

Furthermore, prevailing volume-weighted borrowing and lending rates were found to be significant drivers, holding other variables constant, indicating that the model adapts effectively to changing rate environments and responds more systematically to market conditions than the heuristic approach.

Overall, all intended contribution areas were addressed. The findings conclude that replacing heuristic inventory management with a mathematical optimization framework provides a robust, scalable, and economically effective solution for prime brokerage operations.

6.2. Limitations and Future Research

While the proposed optimization framework demonstrates significant economic improvements, several limitations should be acknowledged. These limitations not only highlight simplifying assumptions made in the current model but also point toward directions for future research and practical implementation.

Firstly, to deploy this optimization framework in a live production environment, future implementations must transition from the current continuous linear programming relaxation to a strict MILP formulation. The present model allows for fractional allocations of securities, which is computationally efficient and suitable for benchmarking purposes. However, in practice, securities are indivisible units and must be allocated in integer quantities. Incorporating integer constraints would ensure operational feasibility but comes at the cost of increased computational complexity, potentially requiring more advanced solution techniques or heuristic approximations to maintain tractability at scale.

Secondly, the model relies on simplified approximations regarding both internal cost structures and external market dynamics. Internally, the cost of balance sheet utilization is modeled using a static piecewise linear function, and the OBFR serves as a direct proxy for the alternative funding cost of the treasury department. In reality, leverage costs and internal funds transfer pricing are highly intricate, depending on shifting liquidity and regulatory risk perceptions. Externally, the model uses historical peak volumes as static caps to represent market demand, assuming perfectly elastic pricing up to this limit. This implies that any quantity can be lent without price impact, which may not hold true for low-volume or hard-to-borrow securities. Future extensions could replace these linear approximations and static caps with non-linear cost structures and dynamic, volume-dependent pricing mechanisms to better reflect real-world behavior.

Thirdly, the current framework assumes frictionless execution and does not explicitly account for operational risks. In practice, issues such as settlement failures (fails-to-deliver) or funding mismatches (fails-to-fund) can affect the real performance. Because the optimization model systematically exploits all mathematically feasible opportunities, the resulting performance estimates may represent an upper bound under ideal conditions. Incorporating probabilistic execution risks or reliability constraints into the model could simulate more conservative and realistic performance estimates.

Fourthly, the simulation framework evaluates and optimizes allocations one day at a time, treating each trading day as an independent problem. This single-period approach does not capture longer-term temporal dependencies such as multi-day collateral lock-ins or term funding structures common in prime brokerage operations. While modeling these dependencies through a multi-period or stochastic optimization framework would yield a more holistic and strategically robust allocation policy, it would also significantly increase the state space and computational complexity of the problem.

Fifthly, the backtesting analysis revealed a data-driven limitation where the heuristic benchmark appeared to outperform the optimization model in specific instances. From a theoretical standpoint, the optimization framework guarantees a globally optimal solution within the defined constraints. However, a detailed investigation revealed a timing mismatch in the historical input data: margin constraints were calculated using start-of-day balances,

while positions and allocations reflected end-of-day states. Intraday trading activity caused these datasets to diverge, leading to an imperfect comparison where the observed heuristic allocations occasionally violated the strict regulatory limits enforced in the model. Future research should address this by using time-synchronized, intraday data snapshots to ensure consistency.

Finally, the framework is subject to a form of design-level in-sample bias. Because the historical dataset was fully visible during the formulation of the model, certain structural choices may have been subconsciously guided to fit this specific period. For instance, the calibration of tie-breaker penalties and the exact framing of constraints could have been inadvertently tuned to optimize performance on this historical window. Consequently, the generalizability of these exact model settings remains a caveat until they can be tested against an entirely separate and unseen market regime.

References

Ahuja, Ravindra K., Thomas L. Magnanti and James B. Orlin (1993). *Network Flows: Theory, Algorithms, and Applications*. Englewood Cliffs, New Jersey: Prentice Hall. ISBN: 978-0136175490.

Ao, Xiang (Jan. 2009). *Heteroskedasticity and Autocorrelation Consistent Standard Errors*. URL: https://www.hbs.edu/research-computing-services/Shared%20Documents/Training/hac.pdf (visited on 04/07/2026).

Aydin, Bilgehan (2016). "Evolution of collateral 'management' into collateral 'optimisation'". In: *Journal of Securities Operations & Custody* 8.3, pp. 259–271.

Basel Committee on Banking Supervision (Jan. 2013). *Basel III: The Liquidity Coverage Ratio and liquidity risk monitoring tools*. Tech. rep. Bank for International Settlements. URL: https://www.bis.org/publ/bcbs238.pdf (visited on 02/11/2026).

Board of Governors of the Federal Reserve System (2024). *12 CFR Part 220 – Credit by Brokers and Dealers (Regulation T)*. URL: https://www.ecfr.gov/current/title-12/part-220 (visited on 02/11/2026).

Bottazzi, Jean-Marc, Jaime Luque and Mário R. Páscoa (2012). "Securities Market Theory: Possession, Repo and Rehypothecation". In: *Journal of Economic Theory* 147.2, pp. 477–500. DOI: 10.1016/j.jet.2010.11.004.

Bylund, Johanna (2017). "Collateral Optimization". Department of Physics, Master of Science in Engineering Physics. Master's thesis. Umeå, Sweden: Umeå University.

Capel, Jeannette and Anouk Levels (2014). *Collateral optimisation: re-use and transformation*. Occasional Studies 5. De Nederlandsche Bank (DNB). URL: https://www.dnb.nl/media/rzlp1obc/201407_nr-_5_-2014-_collateral_optimisation-_re-use_and_transformation.pdf (visited on 02/11/2026).

Chan, Ngai Hang (2010). *Time Series: Applications to Finance with R and S-Plus*. Hoboken, NJ: John Wiley & Sons. ISBN: 9780470414354.

Clear Street (2026). *About*. URL: https://www.clearstreet.io/about (visited on 01/20/2026).

Clear Street LLC (2024). *Consolidated Statement of Financial Condition*. URL: https://www.clearstreet.io/legal (visited on 01/20/2026).

Duffie, Darrell (1996). "Special Repo Rates". In: *The Journal of Finance* 51.2, pp. 493–526. URL: https://www.jstor.org/stable/2329369.

Eren, Egemen (Dec. 2015). *Matching Prime Brokers and Hedge Funds*. SIEPR Discussion Paper 16-014. Job Market Paper. Stanford Institute for Economic Policy Research (SIEPR).

Financial Stability Board (Apr. 2012). *Securities Lending and Repos: Market Overview and Financial Stability Issues*. Interim Report. Financial Stability Board. URL: https://www.fsb.org/uploads/r_120427.pdf (visited on 01/27/2026).

Fleckenstein, Matthias and Francis A. Longstaff (Jan. 2018). *Shadow Funding Costs: Measuring the Cost of Balance Sheet Constraints*. NBER Working Paper 24224. National Bureau of Economic Research. DOI: 10.3386/w24224. URL: https://www.nber.org/papers/w24224.

Giron, Megan C., Georgios Korpas, Waqas Parvaiz, Prashant Malik and Johannes Aspman (2023). "Approaching Collateral Optimization for NISQ and Quantum-Inspired Computing". In: *IEEE Transactions on Quantum Engineering* 4, pp. 1–18. DOI: 10.1109/TQE.2023.3314839.

Harvey, Campbell R. and Yan Liu (2015). "Backtesting". In: *The Journal of Portfolio Management* 42.1, pp. 13–28. URL: https://www.proquest.com/docview/1730656355/fulltextPDF/B32F03B62D5946A9PQ/1?accountid=14581&sourcetype=Scholarly%20Journals.

International Capital Market Association (2024). *Frequently Asked Questions on Repo: What is rehypothecation of collateral?* URL: https://www.icmagroup.org/market-practice-and-regulatory-policy/repo-and-collateral-markets/icma-ercc-publications/frequently-asked-questions-on-repo/10-what-is-rehypothecation-of-collateral/ (visited on 02/11/2026).

Java, Mayur, Julianne Carman, Max Thoman and Manan Shah (2017). "Collateral management: A changing landscape". In: *Journal of Securities Operations & Custody* 9.1, pp. 25–29.

Kirk, Adam, James McAndrews, Parinitha Sastry and Phillip Weed (Dec. 2014). "Matching Collateral Supply and Financing Demands in Dealer Banks". In: *FRBNY Economic Policy Review*, pp. 127–151.

Lou, Wujiang (2021). "Securities Lending Haircuts and Indemnification Pricing". In: Updated version, September 2021. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3874679.

Orrsveden, Magnus and Emil Tarukoski (2019). "Optimization of Collateral Allocation for Securities Lending: An Integer Linear Programming Approach". Master's thesis. Stockholm: KTH Royal Institute of Technology.

Panik, Michael J. (2019). *Linear Programming and Resource Allocation Modeling*. English. 1st. Hoboken, NJ: Wiley. ISBN: 9781119509448.

Pennsylvania State University (2026a). *Autocorrelation and Time Series Methods*. URL: https://online.stat.psu.edu/stat462/node/188/ (visited on 04/07/2026).

Pennsylvania State University (2026b). *Regression with Autoregressive Errors*. URL: https://online.stat.psu.edu/stat462/node/189/ (visited on 04/07/2026).

Sanghro, Majid Ali (Oct. 2024). *Understanding and Dealing with Autocorrelation in Time Series Econometrics*. Accessed: 2026-04-07. URL: https://maseconomics.com/understanding-and-dealing-with-autocorrelation-in-time-series-econometrics/.

StataCorp (2013). *dfuller — Augmented Dickey–Fuller unit-root test*. Stata Time-Series Reference Manual. StataCorp LLC. College Station, TX. URL: https://www.stata.com/manuals13/tsdfuller.pdf.

U.S. Securities and Exchange Commission (2024a). *17 CFR §240.15c3-1 – Net capital requirements for brokers or dealers (SEC Rule 15c3-1)*. URL: https://www.ecfr.gov/current/title-17/section-240.15c3-1 (visited on 02/11/2026).

U.S. Securities and Exchange Commission (2024b). *17 CFR §240.15c3-3 – Customer protection—reserves and custody of securities (SEC Rule 15c3-3)*. URL: https://www.ecfr.gov/current/title-17/section-240.15c3-3 (visited on 02/11/2026).

Wolsey, Laurence A (1998). *Integer Programming*. New York: John Wiley & Sons.

A. Pseudocode

Algorithm 1 Primal Simplex algorithm (high-level steps)

Require: Constraint matrix A, right-hand side b, objective coefficients c, initial basis B

Initialize basic solution $x_B := B^{-1}b$
while optimality condition not met **do**
    Compute reduced costs $\bar{c}_N := c_N^\top - c_B^\top B^{-1}N$
    if $\bar{c}_N \leq 0$ **then then**
        break (optimal solution found)
    end if
    Select entering variable $j \in N$ such that $\bar{c}_j > 0$
    Compute search direction $d := B^{-1}A_j$
    if $d \leq 0$ **then then**
        Abort: Problem is unbounded
    end if
    Select leaving variable $i \in B$ minimizing $\frac{(x_B)_i}{d_i}$ for $d_i > 0$
    Update basis B by swapping i and j
    Update $x_B := B^{-1}b$
end while
return optimal basic feasible solution x_B

Algorithm 2 Dual Simplex algorithm (high-level steps)

Require: Constraint matrix A, right-hand side b, objective c, initial dual-feasible basis B (where $\bar{c}_N \leq 0$)

Initialize basic solution $x_B := B^{-1}b$
while primal feasibility not met **do**
    if $x_B \geq 0$ **then then**
        break (optimal and feasible solution found)
    end if
    Select leaving variable $i \in B$ such that $(x_B)_i < 0$
    Compute dual search direction $w^\top := e_i^\top B^{-1}N$
    if $w \geq 0$ **then then**
        Abort: Primal problem is infeasible
    end if
    Select entering variable $j \in N$ minimizing $\left|\frac{\bar{c}_j}{w_j}\right|$ for $w_j < 0$
    Update basis B by swapping i and j
    Update $x_B := B^{-1}b$ and reduced costs $\bar{c}_N$
end while
return optimal basic feasible solution x_B