Statistical methods with applications to pairs trading and equipment lifetime modeling

by

Allan Vieira de Castro Quadros

Lic., State University of Campinas (UNICAMP), Brazil, 2008

M.S., State University of Campinas (UNICAMP), Brazil, 2012

B.S., University of Brasilia (UnB), Brazil, 2018

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Abstract

This dissertation develops innovative statistical methods with applications to financial markets and risk-based management, presenting two main contributions: first, we introduce a family of distribution-based approaches to pairs trading that enhance the traditional cointegration strategy; second, we address the critical question of estimating the time-to-failure of industrial equipment when no failure data are available. In our first contribution, the family of pairs trading methods uses the distribution of the hedge ratio as a confirmatory layer for trading signals. Under the parametric method, we use a Bayesian hierarchical model that derives the full conditional distribution of the hedge ratio, whereas a Non-Overlapping Block Bootstrap is adopted as the nonparametric strategy. Both methods demonstrate significant outperformance across U.S. and Brazilian markets through 2025 with improved returns, reduced volatility, and enhanced risk-adjusted metrics. In the second part, we develop a methodology under the hierarchical Bayesian model framework with simulated likelihood to compensate for the lack of failure data. The model relies on carefully constructed prior distributions that can incorporate manufacturer-specified and expert information. We used the case of the National Bio and Agro-Defense Facility (NBAF) to validate our approach.

Statistical methods with applications to pairs trading and equipment lifetime modeling

by

Allan Vieira de Castro Quadros

Lic., State University of Campinas (UNICAMP), Brazil, 2008

M.S., State University of Campinas (UNICAMP), Brazil, 2012

B.S., University of Brasilia (UnB), Brazil, 2018

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Approved by:

Co-Major Professor
Dr. Brian Silverstein

Approved by:

Co-Major Professor
Dr. Michael J. Higgins

Copyright

Abstract

This dissertation develops innovative statistical methods with applications to financial markets and risk-based management, presenting two main contributions: first, we introduce a family of distribution-based approaches to pairs trading that enhance the traditional cointegration strategy; second, we address the critical question of estimating the time-to-failure of industrial equipment when no failure data are available. In our first contribution, the family of pairs trading methods uses the distribution of the hedge ratio as a confirmatory layer for trading signals. Under the parametric method, we use a Bayesian hierarchical model that derives the full conditional distribution of the hedge ratio, whereas a Non-Overlapping Block Bootstrap is adopted as the nonparametric strategy. Both methods demonstrate significant outperformance across U.S. and Brazilian markets through 2025 with improved returns, reduced volatility, and enhanced risk-adjusted metrics. In the second part, we develop a methodology under the hierarchical Bayesian model framework with simulated likelihood to compensate for the lack of failure data. The model relies on carefully constructed prior distributions that can incorporate manufacturer-specified and expert information. We used the case of the National Bio and Agro-Defense Facility (NBAF) to validate our approach.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This journey would not have been possible without my unwavering faith in God, the boundless love and support of my wife, Beatriz, my family, and the exceptional guidance from my advisors, Dr. Michael Higgins and Dr. Brian Silverstein. I am deeply grateful to the Statistics Department chairs, Dr. Perla Reyes and Dr. Christopher Vahl, for their invaluable support throughout my academic journey. A special thanks to my former professor, Dr. George von-Borries, whose encouragement inspired me to pursue my Ph.D. at K-State. Finally, my sincere appreciation goes to my friends Madhav Dhital and Kingsley Darko, whose constant encouragement made this challenging path brighter.

Dedication

I dedicate this work to my wife Beatriz, my sons Benjamin and Bernard†, my grandmother Lourdes†, my sister Karine, my parents Cristina and Airton, my brother-in-law Gustavo, and my parents-in-law Sueli and Beto.

Chapter 1

Introduction to Pairs Trading

Pairs trading is a market-neutral investment strategy based on statistical arbitrage that exploits temporary mispricings between two assets that historically move together. When the price spread between the two assets diverges from equilibrium, it creates a potential arbitrage opportunity where investors aim to profit from the transitory divergence (Gatev et al., 2006). This strategy assumes that the spread will revert to its historical long-term mean due to the underlying economic forces holding the pair together (Avellaneda and Lee, 2010). Pairs trading is attractive to investors because of its ability to generate low volatility returns that are largely uncorrelated with broader market movements. Furthermore, the strategy can be implemented across various time horizons, with holding periods ranging from intraday trades lasting minutes to longer-term positions that may span several days, weeks, or even months. The combination of low volatility, minimal exposure to systemic market risk, and adaptability across asset classes has made pairs trading a versatile tool in the portfolio of quantitative traders.

While the fundamental strategy behind pairs trading is conceptually simple, its practical implementation faces three key challenges that can significantly affect its effectiveness. First, accurately identifying pairs of stocks with a stable relationship over the desired timeframe can be difficult. Market conditions may disrupt assumptions of mean reversion, particularly during periods of stress when historical correlations break down. Traditional models often

fail to account for non-Gaussian return distributions, fat tails, or volatility clustering, all of which are common in asset return data (Guillaume et al., 1997; Lux and Marchesi, 2000; Cont, 2001).

Second, allocating shares for the pairs strategy is problematic because the appropriate balance between long and short positions is the foundation of the self-financing strategy, which can be sensitive to price movements and market volatility (Alexander and Dimitriu, 2002). In practice, fixed allocation strategies, such as equal dollar amounts or a simple ratio, may not account for time variations in market dynamics or the differing volatilities of the two assets. This can lead to disproportionate exposure to one asset, increasing the risk of loss if the asset moves unfavorably.

The third challenge involves determining optimal entry and exit points for trades. Strategies such as the distance method or standard cointegration rely on historical thresholds or statistical deviations to trigger buy and sell signals. However, these signals may be unreliable when faced with market regime shifts or structural changes (Engelberg et al., 2009). Identifying the correct time to exit a trade, particularly when the spread fails to converge as expected, can result in significant losses if the market behaves unpredictably. Additionally, transaction costs and liquidity issues can further erode the strategy's profitability, making the precise timing of trades critical to maintaining profitability.

1.1 Literature Review on Pairs Trading

The academic and practical exploration of pairs trading strategies encompasses five main methodologies: distance method, cointegration, copulas, Ornstein-Uhlenbeck, and machine learning approaches.

Most of these methods are based on the foundational concept of mean reversion, which operates on the principle that certain stocks tend to exhibit synchronous movements over time, occasionally diverging due to temporary market inefficiencies or external shocks, before eventually converging back to their historical mean relationship.

1.1.1 Distance Method

The distance method was first formally documented in academic literature by Gatev et al. (2006). This method identifies pairs of stocks that exhibit similar price movements by selecting those with the lowest sum of squared differences in normalized prices during a formation period (typically 12 months). Trading signals are subsequently generated when the normalized spread between prices exceeds a threshold of $\pm 2\sigma$ during the trading period (typically 6 months). Since the selection methodology prioritizes pairs with minimal price differences, there is no theoretically normative approach for determining the optimal number of shares to trade; when a trading signal occurs, investors typically establish a long position of d dollars in stock Y and simultaneously short an equivalent d dollars of stock X. Positions are closed when the spread reverts to zero, and the entire procedure is reinitiated after the conclusion of the 6-month trading period.

The authors evaluated this method using daily data from 1962 to 2002 for the U.S. market based on the CRSP index[1] under two distinct scenarios: one without industry or sector restrictions in pair selection, and another forming pairs exclusively within the same sector. Their findings indicated that the unrestricted approach yielded higher returns.

Despite the method's initial success, its effectiveness has been increasingly questioned due to evolving market conditions. Subsequent analyses by Do et al. (2006) and Do and Faff (2008) documented a decline in profitability, primarily attributable to the diminishing arbitrage opportunities and the growing impact of trading costs, rendering the strategy largely ineffective in the post-2002 period. Further research by Jacobs and Weber (2015) confirmed that the strategy's profitability exhibits significant temporal variability, as demonstrated by their application of the method to data from 34 international markets. The distance method remains the most extensively studied and empirically tested pairs trading strategy in the literature. Additional studies implementing this method across various markets, time periods, and asset classes were conducted by Gatev et al. (2006); Engelberg et al. (2009); Do and Faff (2008); Andrade et al. (2005); Papadakis and Wysocki (2007); Broussard and Vaihekoski

[1]The Center for Research in Security Prices (CRSP) provides historical stock market data and indexes that serve as benchmarks for the investment community and academic research.

(2012).

After fitting a linear model relating Y to X, i.e., $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$, if the residuals (e_t) are stationary, i.e., $e_t \sim I(0)$, where $I(0)$ denotes a stationary process, then the pair is said to be cointegrated.

1.1.2 Copulas

Copulas have gained increased attention in finance over the last decade thanks to research indicating the asymmetric dependency structure of financial assets. In simple terms, this means that we tend to observe higher correlation among assets during financial crises or euphoric market conditions than during regular market regimes. This illustrates the tail dependence phenomenon, which measures how strongly significant movements in the upper or lower extremes of each asset price's distribution occur together. It becomes crucial, therefore, to select a copula structure (Gumbel, Student t, Normal, Clayton, etc.) that accurately represents one's expectations for these joint extreme movements in tail events. When applied to pairs trading, copulas allow us to relax linearity and symmetry constraints on pair dependencies, improving the modeling of asset dependencies and potentially yielding excess returns.

The copula method in pairs trading consists of finding the following conditional cumulative distributions:

$$g_X(u_X|u_Y) = P(U_X \leq u_X|U_Y = u_Y) = \frac{\partial C(u_X, u_Y)}{\partial u_Y}$$

$$g_Y(u_Y|u_X) = P(U_Y \leq u_Y|U_X = u_X) = \frac{\partial C(u_X, u_Y)}{\partial u_X}$$

If X and Y are random variables respectively representing the prices of assets X and Y, their copula function C can be denoted as the joint distribution of their CDFs, i.e.,

$C(u_X, u_Y) = P(U_X \leq u_X, U_Y \leq u_Y)$, where $P(U_X \leq u_X) = F_X(x)$ and $P(U_Y \leq u_Y) = F_Y(y)$. Note that both $F_X(x)$ and $F_Y(y) \sim \text{Uniform}(0, 1)$, which means the inputs u_X and u_Y are quantiles from data mapped by their respective marginal CDFs. Therefore, we can use functions g_X and g_Y in pairs trading to calculate the likelihood of one stock in the pair moving higher or lower than its current price given the other stock's price. The type of copula chosen to model the underlying dependence of the two assets directly affects the resulting probability, thus influencing the entire pairs trading strategy under this framework. Clearly, the copulas method represents more of a probabilistic approach than a traditional mean-reversion statistical arbitrage algorithm.

The pairs trading literature still lacks comprehensive empirical evidence supporting copula-based strategies across a broad set of stocks and markets. The existing studies show contradictory performance when this framework is applied in pairs trading. The first studies appeared in 2013 with Liew and Wu (2013) and Stander et al. (2013), indicating promising results for the copulas method, even considering it profit-wise superior compared to the distance and cointegration methods. On the other hand, more recent studies in copula pairs trading have shown considerably poorer performance compared to the distance and cointegration methods (Rad et al., 2016).

Moreover, all studies have consistently shown that the number of trades is considerably higher under the copulas approach. Although this implies more trading opportunities, it also results in higher transaction costs (Rad et al., 2016; Wu, 2018). Empirical results also demonstrate significantly higher drawdowns under copula pairs trading (Wu, 2018), making this strategy less attractive to investors given the trade-off between implementation complexity and its risk-return ratio[2]. Another common criticism of the copulas approach is that the framework does not adequately account for the time-varying characteristics of financial data when modeling pair-dependence structures.

An important aspect to highlight is that the copula pairs trading approach primarily addresses only the last of the three pairs trading challenges identified earlier: how to generate

[2]Rad et al. (2016) begin their paper by stressing the importance of balancing sophistication and performance in algorithmic trading strategies such as pairs trading.

trading signals. Thus, practitioners must resort to different techniques to perform pair selection or balance trade sizes for the underlying assets. For pair selection, for example, one can search for cointegrated pairs and/or assets with the lowest squared distances in prices (as in the distance method). As for the optimal balance of traded shares, investors typically adopt either the dollar-neutral strategy from the distance method or the estimated slope $\hat{\beta}_1$ from the model $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$.

1.1.3 Ornstein-Uhlenbeck

An Ornstein-Uhlenbeck process is a stochastic process characterized by a mean-reversion pattern, formulated within the context of optimal-stopping theory (Ferguson, 2007). Its application in pairs trading involves modeling the price spread using an Ornstein-Uhlenbeck process, which enables the deployment of a probabilistic methodology to determine the optimal price intervals for entering and exiting the market.

The application of the Ornstein-Uhlenbeck process to pairs trading was first introduced in academic theses by Rampertshammer (2007) and Qiang (2009). Both studies represented initial efforts to link pairs trading with optimal-stopping theory. However, neither study analyzed empirical data. It was only in 2015 that Leung and Li (2015) provided a broader and more in-depth study on the topic, including the application of the Ornstein-Uhlenbeck pairs trading framework to real stock data. Several studies on the topic have followed since then. Endres and Stübinger (2017) derived an alternative strategy using a Lévy-driven Ornstein-Uhlenbeck[3] process with outstanding performance when applied to intraday data of the S&P500. Zhu et al. (2021) focused on a mean-variance optimization of the regular Ornstein-Uhlenbeck process. Despite these efforts, comprehensive comparisons between this approach and more traditional pairs trading methods remain limited.

The Ornstein-Uhlenbeck pairs trading method as described by Leung and Li (2015) is based on the following stochastic differential equation (SDE):

[3]The most common Ornstein-Uhlenbeck pairs trading strategy uses regular Brownian motion in the stochastic differential equation (SDE).

$$dX_t = \mu(\theta - X_t)dt + \sigma dB_t,$$

$$\text{with } \mu, \quad \sigma > 0, \quad \text{and } \theta \in \mathbb{R},$$

where:

- θ is the long-term mean level around which all future trajectories of X will eventually revolve.
- μ denotes the reversion speed, indicating how quickly these trajectories will converge back to θ over time.
- σ represents the instantaneous volatility, quantifying the immediate degree of randomness affecting the system. Higher σ values signify greater randomness.

The parameter estimates are obtained through Maximum Likelihood estimation (MLE) using historical data. The second step involves setting up the optimal stopping problem and subsequently solving it. To achieve this goal, Leung and Li (2015) incorporate equations that account for transaction and opportunity costs, maximizing the expected discounted values of entry and liquidation to find the optimal levels for signal generation. The optimal allocation is based on the MLE of β and established according to the following spread X_t formed by the differences of stocks $S_t^{(1)}$ and $S_t^{(2)}$, following the authors' notation:

$$X_t^{\alpha,\beta} = \alpha S_t^{(1)} - \beta S_t^{(2)}$$

Similar to other approaches[4], α is fixed so that the optimal allocation is β dollars of asset $S_t^{(2)}$ for each 1 dollar of asset $S_t^{(1)}$.

[4]Although the Ornstein-Uhlenbeck model is the most widespread stochastic approach in pairs trading, it is important to mention the existence of relevant studies based on the stochastic control framework (Jurek and Yang, 2007; Mudchanatongsuk et al., 2008; Xing, 2022), which also use stochastic differential equations to determine optimal entry and exit points.

1.1.4 Machine Learning Strategies

In recent years, we have witnessed a remarkable surge in the application of machine learning algorithms to quantitative trading strategies, with particular emphasis on statistical arbitrage and pairs trading.

The evolution of pair selection methodologies represents a significant area of advancement for methods based on machine learning models. While traditional approaches such as distance metrics, mean reversion modeling and copula methods have provided the foundation, machine learning has introduced powerful new dimensions to this process. Han and Wei (2023) pioneered the application of unsupervised learning for pair selection, employing clustering algorithms that can process multidimensional company characteristics to identify stocks with genuine co-movement patterns. This represents a substantial leap beyond conventional statistical measures, allowing for richer data incorporation and more robust pair identification.

Comprehensive comparative research by Krauss and Stübinger (2017) has established important benchmarks in this domain. Their evaluation of deep neural networks, gradient-boosted trees, and random forests across S&P 500 constituents revealed that ensemble approaches, which integrate multiple algorithms, consistently generate superior results compared to both individual techniques and traditional statistical methods.

Building on these foundations, Fischer and Krauss (2018) demonstrated the power of temporal modeling in financial series through deep learning. Their implementation of LSTM neural networks effectively captured complex temporal dependencies across large stock universes, generating significant risk-adjusted returns even after accounting for transaction costs. Their methodology particularly excelled in feature extraction for pairs trading signals, addressing one of the most challenging aspects of the strategy.

Reinforcement learning has emerged as a particularly innovative approach within the machine learning domain. Initially proposed by Fallahpour et al. (2016), this methodology was substantially advanced by Brim (2019) and Kim and Kim (2019), who successfully implemented deep reinforcement learning frameworks for pairs trading. Brim (2019) developed

a sophisticated model that selects from S&P 500 stocks based on cointegration and variance parameters, ensuring pairs exhibit statistically significant mean-reverting relationships with sufficient volatility for signal generation. Their model intelligently determines position actions (long, short, or hold) based on spread characteristics, employing a training mechanism that penalizes negative returns to foster more conservative trading approaches.

Further innovations in reinforcement learning came from Wang et al. (2021), who introduced reward-shaping methods incorporating baseline policies to guide RL agents. Their comprehensive testing on NASDAQ Nordic markets using five years of intraday data confirmed that RL models deliver both higher returns and superior profit-risk ratios compared to conventional approaches.

The integration of complex neural network architectures has pushed the boundaries of what's possible in modeling market relationships. Zhang et al. (2019) developed an innovative hybrid architecture combining convolutional and recurrent neural networks specifically designed to capture both cross-sectional and temporal dependencies in pairs trading. Their approach demonstrated remarkable effectiveness in identifying regime shifts in cointegration relationships, outperforming traditional statistical tests in dynamic market environments.

Despite these advances, significant challenges remain in the application of machine learning to pairs trading. David H. Bailey and Zhu (2014) highlighted the critical issue of backtest overfitting. Model complexity presents another obstacle, as noted by Shihao Gu and Xiu (2020), who demonstrated that highly parameterized deep learning models often underperform simpler approaches out-of-sample due to the low signal-to-noise ratio inherent in financial data.

Furthermore, empirical evidence suggests that machine learning models trained on historical data often struggle to adapt when confronted with structural breaks and regime changes in market dynamics (Suárez-Cetrulo et al., 2023). This adaptability challenge is compounded by interpretability concerns, which remain a significant barrier to institutional adoption. Risk management frameworks typically demand transparent decision processes, creating a natural resistance to black-box models whose internal operations cannot be easily explained or audited. Additionally, the computational complexity of deploying sophisticated machine

learning algorithms on high-dimensional financial data in time-sensitive trading environments has driven researchers to develop increasingly optimized implementation strategies, balancing model sophistication with execution efficiency.

In conclusion, while machine learning has undoubtedly transformed the landscape of pairs trading by enhancing each stage of the strategy's implementation, its successful application requires careful consideration of financial data's unique characteristics. The future of this field lies in developing approaches that balance the power of advanced algorithms with the practical realities and constraints of financial markets.

1.1.5 Alternative Approaches to Pairs Trading

Beyond traditional cointegration or distance-based methods, several alternative approaches have emerged to address inherent challenges in pairs trading, often combining elements from multiple analytical frameworks or introducing novel perspectives. Kalman filtering, for instance, dynamically tracks the evolving relationships between paired assets, adapting to gradual parameter changes rather than assuming stability. The first applications were designed to model the hedge-ratio as a time-dependent variable, aiming to enhance resilience to structural market shifts. Later implementations, despite also trying to identify structural breaks in the long-term equilibrium, have focused on modeling the spread between the two analyzed assets (de Moura et al., 2016).

Wavelet analysis offers another methodologically robust alternative, employing multi-resolution decomposition of price series to simultaneously capture short-term trading opportunities while accounting for longer-term structural relationships (Eroğlu et al., 2023). Pioneering applications of wavelet analysis to pairs trading have successfully identified mispricing at specific temporal scales while considering broader asset relationships. More advanced studies have emerged recently, such as the work combining wavelet transform with convolutional neural networks (CNN), long-short term memory (LSTM) networks, and deep reinforcement learning to enable real-time detection of stationary relationship breakdowns that could lead to significant trading losses Lu et al. (2022).

1.1.6 Cointegration

Despite the emergence of numerous innovations in pairs trading, cointegration remains one of the widest adopted strategies in the field, valued for its strong theoretical foundation and straightforward implementation. The core principle of the cointegration strategy involves identifying two stocks (Y and X) whose price series (y_t and x_t) move together over time, despite each individual series being non-stationary (Vidyamurthy, 2004). While each stock price may wander extensively, their relationship maintains a stable equilibrium in the long run. This property enables traders to capitalize on temporary deviations from this equilibrium with statistical confidence that the spread will revert to its mean.

In technical terms, we need $y_t \sim I(1)$ and $x_t \sim I(1)$, where $I(1)$ denotes an integrated process of order 1, meaning that the series becomes stationary only after taking the first difference. After fitting a linear model relating Y to X, i.e., $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$, if the residuals ($e_t$) are stationary, i.e., $e_t \sim I(0)$, where $I(0)$ denotes a stationary process, then the pair is deemed cointegrated.

Stationarity implies a strong mean-reverting relationship between Y and X in the observed period, making the residuals' behavior predictable in the long run. Once suitable pairs are identified, position sizing is determined by the regression coefficient ($\hat{\beta}_1$), which represents the optimal hedge ratio. This creates a naturally balanced portfolio where for each dollar invested in stock Y, approximately $\hat{\beta}_1$ dollars of stock X are traded in the opposite direction.

The standardized spread $y_t - \hat{\beta}_1 x_t$, which is also stationary as a direct consequence of the residuals' stationarity (Vidyamurthy, 2004), is then used to generate trading signals. Positions are taken when the standardized spread deviates significantly from its historical mean (typically beyond ± 2 standard deviations). When one of these thresholds is crossed, a trading opportunity emerges. investors go long or short on $\hat{\beta}_1$ dollars of stock X for each dollar of stock Y. Positions are initiated with the expectation that the spread will revert to its mean, and they are typically unwound when the spread returns to its equilibrium value or when a predetermined stop-loss threshold is reached.

Testing for Cointegration

The cointegration test was first introduced in the seminal paper by Engle and Granger (1987). The authors proposed using the Augmented Dickey-Fuller (ADF) unit root test to assess the presence or absence of stationarity in the original series and its residuals when regressing y_t on x_t. As previously mentioned, cointegration is present when two non-stationary series, after being combined in a linear model, produce stationary residuals.

The original ADF procedure tests the null hypothesis that a time series has a unit root, implying that the series is non-stationary and follows a stochastic trend. A time series y_t is said to have a unit root if shocks to the series have a permanent effect, indicating that the series does not revert to a mean over time. The ADF test fits the following time series regression model:

$$\Delta y_t = \alpha + \theta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \ldots + \delta_p \Delta y_{t-p} + \epsilon_t$$

where $\Delta y_t = y_t - y_{t-1}$ represents the first difference of the time series, and $\Delta y_{t-m+1} = y_{t-m+1} - y_{t-m}$, $m = 2, \ldots, p+1$, denotes the m^{th} difference between consecutive observations in the series. In this model, γ is the key parameter for testing the presence of a unit root. If $\gamma = 0$, then y_t has a unit root, i.e., the series is non-stationary and follows a random walk. In contrast, if $\gamma < 0$, the series is stationary and reverts to a long-term mean. The θ term represents the deterministic trend, and the δ_i terms capture the auto-regressive dynamics of the first differences of the series.

Subsequent methodologies were proposed as more robust alternatives to the original Engle-Granger framework. Among several methods (Johansen, 1988; Toda and Yamamoto, 1995; Gregory and Hansen, 1996; Saikkonen and Lütkepohl, 2000; Banerjee et al., 1998; Pesaran et al., 2001; Westerlund, 2007; Pedroni, 1999, 2004; Bracegirdle and Barber, 2012), the Phillips-Perron (PhP) test (Phillips and Perron, 1988) is one of the most commonly adopted alternatives, having the notable advantage of providing a nonparametric approach to handling nuisance parameters. The test consists of a variation of the Augmented Dickey-

Fuller (ADF) procedure that takes a different approach to handling autocorrelation and heteroscedasticity when testing for the presence of a unit root in the series.

In the ADF test, lagged difference terms (Δy_{t-k}) are included in the regression to correct for autocorrelation in the residuals. These lags help ensure that the regression residuals are approximately white noise (i.e., uncorrelated and homoscedastic). However, the effectiveness of the ADF test depends on the proper selection of the number of lagged terms, which may require additional model selection criteria. Even with this correction, there are cases where the residuals remain autocorrelated or heteroscedastic, violating the assumption of white noise and potentially leading to biased results.

In contrast, the PhP test does not include lagged terms in the regression model. Instead, it directly adjusts the test statistic using heteroscedasticity and autocorrelation consistent (HAC) standard errors. The method builds on the foundational work of Huber (1967) and White (1980), who developed techniques for robust covariance estimation to handle heteroscedasticity. Newey and West (1987) later extended these ideas to create a unified approach that addresses both heteroscedasticity and autocorrelation.

The regression model used in the PhP test mirrors the basic Dickey-Fuller framework but omits the lagged terms typically included to address autocorrelation:

$$\Delta y_t = \alpha + \theta t + \gamma y_{t-1} + \epsilon_t,$$

where $\Delta y_t = y_t - y_{t-1}$ represents the first difference of the series, and ϵ_t is the error term, which may exhibit autocorrelation or heteroscedasticity. The PhP procedure modifies the test statistic associated with γ (the parameter used to test for the presence of a unit root) by applying nonparametric corrections to account for these issues. This approach eliminates the need to explicitly model the residual structure using lagged terms, making the test less sensitive to user-defined parameters.

By applying robust variance and covariance estimators, the Phillips-Perron test also accommodates a broader range of time series models, including those with heterogeneously distributed innovations. These characteristics make the Phillips-Perron test a more flexible

and reliable tool for practical applications, particularly for economic and financial time series data. However, like any other hypothesis test, the PhP test is also susceptible to the occurrence of false positives.

1.1.7 Advantages and Challenges

The cointegration approach offers several advantages over simpler methods like the distance approach. It provides a statistically rigorous framework for pair selection, incorporates an optimal hedge ratio for position sizing, and accounts for the underlying economic relationship between securities. These features have made it particularly popular among institutional investors and sophisticated quantitative trading firms.

However, the methodology also faces challenges, including sensitivity to structural breaks in the relationship between securities and the need for careful statistical verification to avoid spurious cointegration. Various refinements to the basic framework have emerged over time, including error correction models, threshold cointegration, and regime-switching approaches that accommodate changing market dynamics (Banerjee et al., 1998; Westerlund, 2007).

1.2 Proposal

Building upon this foundation, our research introduces novel distribution-based approaches that significantly enhance traditional cointegration methods. By focusing specifically on the challenge of signal generation, we develop methodologies that provide more robust trading signals through a comprehensive characterization of the hedge ratio relationship between paired assets.

We propose two methodologies that derive the distribution of hedge ratios between cointegrated assets and then use specific quantiles of these distributions to create a confirmatory layer upon the standard cointegration strategy. This two-layer decision process refines both entry and stop-loss signals, substantially improving trading performance while maintaining the fundamental mean-reversion premise of traditional pairs trading.

In Section 2.2, we introduce a parametric approach based on hierarchical Bayesian models that derives the full conditional distribution of hedge ratios. By utilizing quantiles of this distribution as confirmation thresholds for trading signals generated within the standard cointegration framework, our method significantly enhances signal precision and adaptability. This Bayesian approach not only improves trading performance and risk management compared to traditional strategies but also enhances the pair selection process by effectively filtering out false positives in cointegration tests. Through simulation studies, we demonstrate how this approach identifies pairs with genuine long-term equilibrium relationships, addressing a critical limitation of conventional cointegration testing procedures that often select spurious relationships.

In Section 3, we develop a Non-Overlapping Block Bootstrap (NBB) approach that generates the empirical distribution of the hedge ratio through resampling techniques specifically designed to preserve temporal dependencies in financial time series. Unlike standard bootstrap methods that may disrupt the autocorrelation structure of financial data, our NBB methodology maintains these critical dependencies, resulting in hedge ratio distributions that better reflect actual market dynamics. This approach demonstrates more consistent performance with fewer parameter sensitivities compared to the Bayesian method, while maintaining operational characteristics closer to traditional trading methods. These features make the NBB approach particularly accessible for implementation while still providing substantial performance enhancements.

Empirical testing across both U.S. and Brazilian markets from 2023 to 2025 demonstrates that our proposed methodologies significantly outperform standard cointegration-based pairs trading strategies. Both approaches deliver enhanced cumulative returns, reduced return volatility, and improved risk-adjusted performance metrics. The NBB method in particular achieves substantially higher returns while significantly reducing volatility and maximum drawdowns for all analyzed pairs. Risk-adjusted metrics improved dramatically, with many pairs transitioning from negative to strongly positive Sharpe and Sortino ratios.

Together, these approaches constitute a family of probabilistic strategies for pairs trading that represent a significant advancement in statistical arbitrage techniques. By incorporating

distributional information about the hedge ratio, these methods provide a versatile toolkit that can be adapted to various market environments and trading objectives. The NBB method offers a more accessible implementation with consistent performance enhancements across diverse market conditions, while the Bayesian approach provides similar performance with a more sophisticated implementation, but improved pair selection.

Chapter 2

Bayesian approach to distribution-based trading signals

2.1 Introduction

In this study, we propose a novel approach that improves the pair selection process and optimizes entry and exit points for pairs trading utilizing Bayesian hierarchical modeling. Our key contribution lies in adopting a truncated normal prior for hedge ratios (β_1) and using posterior quantiles derived from its conjugate distribution as a confirmation signal to enhance the precision of trading signals within the cointegration framework. In the proposed method, we can also improve pair selection by introducing a tolerance threshold for consecutive negative β_1 values in the Gibbs sampling process. When the number of negative β_1 values surpasses this threshold, the algorithm discards the pair, ensuring that only pairs with stable hedge ratios are selected for trading.

Using Bayesian hierarchical modeling provides several advantages over the traditional pair selection and trading signal systems employed in the standard cointegration strategy. By incorporating a truncated normal prior and its full conditional conjugate for hedge ratios, we impose a structure that ensures hedge ratios remain within a realistic range, reducing the risk of extreme or unstable values. This contrasts with traditional approaches that

rely solely on historical averages or static thresholds, which may not adapt well to changing market dynamics. Introducing a tolerance threshold for consecutive negative β_1 values serves as an additional safeguard, helping filter out pairs that are likely to break down in practice. This approach ensures that our method is adaptive and resilient to regime shifts and market anomalies, ultimately improving trading performance.

We show that our Bayesian method consistently outperformed the traditional cointegration strategy across most pairs analyzed in this study. This underscores its ability to generate more accurate and reliable trading signals. The adaptive nature of the Bayesian model allows it to dynamically adjust to market conditions, providing a more flexible and precise response compared to static approaches based on fixed standardized spread (z-score) rules. Moreover, our simulations reveal that the Bayesian approach is particularly effective in detecting false positives in cointegration tests, a crucial factor in enhancing the reliability and robustness of pairs trading strategies. This heightened accuracy reduces the risk of selecting unstable pairs and mitigates the potential for losses in highly volatile or regime-shifting environments.

In addition, the ability of the Bayesian framework to incorporate uncertainty into the model further contributes to its superior performance. The method adapts to short-term fluctuations and long-term structural shifts in asset behavior by accounting for variations in hedge ratios and market volatility. This flexibility enables the algorithm to better capture asset pairs' mean-reversion properties while minimizing exposure to false signals. Consequently, our results suggest that the proposed algorithm has the potential to set a new standard in the field of statistical arbitrage, offering practitioners a more resilient and sophisticated tool for pairs trading.

The remainder of this chapter is organized as follows. Section 2.2 introduces the Bayesian hierarchical model and details the derivation of the full conditional distribution for hedge ratios, completing the theoretical foundation for our method. Section 2.3 presents the backtesting methodology and the algorithm's performance compared to the traditional cointegration strategy in selected pairs of the US and Brazilian markets. Section 2.4 consists of simulation studies, focusing on the method's robustness in identifying false positives in cointegration tests. Finally, Section 2.5 concludes the chapter, summarizing our findings and

proposing avenues for future research.

2.2 Methodology

2.2.1 Introduction

Our new Bayesian approach to pairs trading builds upon the cointegration framework by dynamically enhancing the reliability of pair selection and refining the trading signals through probabilistic evaluation of the hedge ratio (denoted by β_1). We contend that β_1 contains critical information about the pair relationship, such that any breakdown in the "co-moving" assumption will first manifest as changes in the pair's hedge ratio. By quantifying deviations of the current β_1 from its expected distribution, we implement an additional confirmatory layer for entry and exit signals generated by the standard cointegration strategy.

The core of the strategy lies in generating the full conditional distribution of the parameter vector, $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix}^\top$, through a hierarchical Bayesian model. Here, β_0 represents the intercept, and β_1 is the hedge ratio, which defines the relationship between the cointegrated assets. An additional benefit of our approach is that sampling from the full-conditional distribution of β_1 serves as an enhanced pair selection filter, significantly reducing false positives in the cointegration test—a result we demonstrate empirically in Section 2.3 and more comprehensively through simulation studies in Section 2.4.

With the full conditional distribution derived, we identify two quantiles (e.g., 5^{th} and 95^{th}) from the distribution of β_1 and utilize them as confirmatory triggers for placing both entry and exit orders. If the z-score of the spread $y_t - \hat{\beta}_1 x_t$ at decision period t crosses[1] the predefined boundaries (e.g. $\pm 2\sigma$, where σ represents the standard deviation of the spread), and the current $\hat{\beta}_1$ is well-behaved[2], i.e., within the chosen quantiles of the hedge ratio's distribution, the entry signal is confirmed and the orders are executed. Otherwise, the entry

[1] Assuming a discrete time span in which $t+1$ is the trading time, $\hat{\beta}_1$ is estimated using data from the start of the analyzed period until $t-1$.

[2] The current $\hat{\beta}_1$ is estimated using data from the start of the analyzed period until decision time t, just before the trading time $t+1$.

signal is canceled.

Stop-loss thresholds are also established using standard deviations of the standardized spread, e.g. $\pm|2\sigma + \xi|$, where ξ is a pre-established loss margin measured in standard deviations. Once an operation is opened, if the standardized spread crosses the defined stop-loss boundaries in the next periods, the algorithm emits a stop-loss signal that is only confirmed if the updated $\hat{\beta}_1$ is outside the defined quantiles of the hedge ratio's distribution. Otherwise, the exit signal is canceled and the operation continues.

In our method, the take-profit threshold is set using a prespecified margin of gain based on the standardized spread and will not depend on the current value of the hedge ratio. The common choice is $z = 0$, given that it represents the long-term equilibrium of the stationary process.

We hypothesize that the proposed method is profit-wise superior to the regular cointegration framework because it may

1. Enhance the pair selection step;

2. Generate more timely trading signals.

If a pair is cointegrated and the two shares truly move together, it is reasonable to assume that the slope of the linear model $y_t = \beta_0 + \beta_1 x_t + \epsilon_i$ is positive. Based on this assumption, we assign a normal distribution truncated at $\beta_1 > 0$ as the prior for $\boldsymbol{\beta}$ to reflect the expected behavior of the pair. The full conditional will also follow a truncated normal distribution, as demonstrated in Section 2.2.2. If, when drawing realizations from the full conditional of $\boldsymbol{\beta}$, we obtain a sequence of m (e.g., $m = 30$) negative β_1 values, we opt to discard the pair due to the strong contradiction between the sampled values and our prior belief. This approach is primarily heuristic, grounded in the observation that cointegrated pairs that move together will have positive hedge ratios ($\beta_1 > 0$). While this is not a strict theoretical result, it aligns with the practical behavior of pairs trading strategies. In this way, our method can perform pair selection by confirming or discarding the results of the cointegration test, adopting this heuristic to guide decision-making.

Another essential component of our strategy is improving the precision of trading signals. Relying exclusively on the standardized spread for signal generation can result in entering trades just as the pair is about to lose its long-term equilibrium (cointegration) status or when the pair balance given by β_1 is excessively distorted. These two scenarios can lead to losses because the pair will not return to the long-term mean or because of an unbalanced share allocation. Furthermore, it is not uncommon for pairs trading practitioners to observe early stoppages that considerably affect the strategy's performance. We hypothesize that the adoption of the distribution of $\boldsymbol{\beta}$ as an extra layer of confirmation contributes to generating more trustworthy and timely trading signals, thus mitigating the aforementioned risks and improving the strategy's profitability.

We empirically evaluate the performance of our method against the standard cointegration strategy in terms of profitability and risk metrics in Section 2.3. The patterns of cumulative returns and drawdowns provide insights into the differences in trading signals between the two methods, aiding in the assessment of hypothesis (ii). In Section 2.4, we further investigate hypothesis (i) by applying the proposed method to simulated cointegrated and non-cointegrated time series. This allows us to evaluate whether the Bayesian approach can correctly discard false positive indications from the cointegration test.

2.2.2 Theoretical Background

The Bayesian representation of the regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ consists of a hierarchical model comprised of prior distributions for the model parameters and a likelihood function representing our data model. When combined, these distributions allow us to compute the full conditional distribution for $\boldsymbol{\beta}$ which is the distribution of interest in our proposed method of generating dynamic trading signals.

We suppose that $\boldsymbol{y}$ follows a multivariate normal distribution

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I})$$

with $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix}^\top$, and analytical form for the likelihood given by

$$p(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right). \tag{2.1}$$

Since our approach focuses on cointegrated pairs, the residuals are expected to be stationary by construction. This property justifies the assumption of homoscedasticity and uncorrelated errors ($\boldsymbol{\Sigma} = \sigma^2 I$) within the Bayesian framework. Furthermore, it simplifies computation and aligns with our goal of dynamically updating posterior distributions for $\boldsymbol{\beta}$ without imposing unnecessary complexity.

As for our priors, we propose an Inverse Gamma distribution for σ^2 to model its strictly positive domain and accommodate skewness

$$\sigma^2 \sim \mathcal{IG}(q, r),$$

where $q > 0$ is the shape parameter and r^{-1} is the scale parameter with $r > 0$. The analytical form for its probability density function (pdf) is

$$p(\sigma^2 \mid q, r) = \frac{1}{r^q \Gamma(q)} (\sigma^2)^{-(q+1)} \exp\left(-\frac{1}{r\sigma^2}\right). \tag{2.2}$$

For $\boldsymbol{\beta}$, we adopt a truncated multivariate normal prior to enforce positivity on its first component, β_1:

$$\boldsymbol{\beta} \sim \mathcal{TN}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}},\ 0, +\infty).$$

The pdf of this truncated normal distribution is proportional to a standard multivariate normal pdf, but with support restricted to $\beta_1 > 0$:

$$p(\boldsymbol{\beta} \mid \boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \propto |\boldsymbol{\Sigma}_{\boldsymbol{\beta}}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}_{\boldsymbol{\beta}}{}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}})\right), \tag{2.3}$$

where $\beta_1 > 0$ ensures the positivity constraint on the first component of $\boldsymbol{\beta}$.

We initialize $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ using the sample regression estimator and the sample variance-covariance matrix (Rencher and Schaalje, 2007) based on all data collected up to time $[t-1]$. Specifically, we set $\boldsymbol{\mu}_{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \hat{\sigma}_0^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$, where $\hat{\sigma}_0^2 = \frac{1}{n-2}\boldsymbol{y}^\top(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}$ and $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$. These choices follow standard Bayesian linear regression practices, using ordinary least squares (OLS) estimates as informative priors. In addition, based on empirical evidence from our backtesting of the pairs trading strategy, we fix the variance $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ at $\hat{\sigma}_0^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$. This approach enhances model stability and identifiability by reducing the uncertainty associated with $\boldsymbol{\beta}$, a technique that has proven effective in similar contexts (see, e.g., Chib et al. (2006) and Aguilar and West (2000) for applications in factor models). For further details on initialization methods in Bayesian regression, see Hooten and Hefley (2019).

Full Conditional Distributions

The choice of a truncated normal distribution for $\boldsymbol{\beta}$ is justified by the co-moving assets assumption and the adoption of the cointegration framework as the foundation of our pairs trading strategy. If two assets are predominantly moving in cooperation and exhibit cointegration, it is expected that their hedge ratio β_1 will remain positive throughout the period analyzed. This is aligned with Gelfand et al. (1992) suggestion, where the authors indicate the truncated normal as the prior distribution in cases where the parameter of interest cannot be negative.

However, generating distributions directly from a truncated normal distribution can be computationally challenging. Following the recommendation by Gelfand et al. (1992), we instead derive the unconstrained full conditional distribution and impose truncation through rejection sampling.

The full conditional distribution for $\boldsymbol{\beta}$ will thus have the form

$$p(\boldsymbol{\beta} \mid \boldsymbol{y}, \sigma^2) \propto \begin{cases} p(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}) & \boldsymbol{\beta} \in S \\ 0 & \boldsymbol{\beta} \notin S \end{cases}$$

Where $S = \left\{(\boldsymbol{y}, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{S_Y^k}\right\}$, and $\boldsymbol{S_Y^k}$ is the constraint set given by $\beta_1 > 0$ (Gelfand et al., 1992).

Therefore, the analytical form of the unconstrained full conditional for $\boldsymbol{\beta}$ when $\beta_1 > 0$ is

$$\begin{aligned} p(\boldsymbol{\beta} \mid \boldsymbol{y}, \sigma^2) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right) \\ \times |2\pi\boldsymbol{\Sigma}_{\boldsymbol{\beta}}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}})^\top {\boldsymbol{\Sigma}_{\boldsymbol{\beta}}}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}})\right). \end{aligned}$$

After discarding terms that do not depend on $\boldsymbol{\beta}$ and using $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$, we have

$$p(\boldsymbol{\beta} \mid \boldsymbol{y}, \sigma^2) \propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\beta}^\top \left(\boldsymbol{X}^\top (\sigma^2 \boldsymbol{I})^{-1} \boldsymbol{X} + {\boldsymbol{\Sigma}_{\boldsymbol{\beta}}}^{-1}\right) \boldsymbol{\beta} - 2\boldsymbol{\beta}^\top \left(\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y} + (\sigma^2 \boldsymbol{I})^{-1} \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\right]\right). \tag{2.4}$$

We can further simplify quadratic and linear terms by adopting $\boldsymbol{A} = \left(\boldsymbol{X}^\top (\sigma^2 \boldsymbol{I})^{-1} \boldsymbol{X} + {\boldsymbol{\Sigma}_{\boldsymbol{\beta}}}^{-1}\right)$ and $\boldsymbol{b} = \left(\boldsymbol{X}^\top (\sigma^2 \boldsymbol{I})^{-1} \boldsymbol{y} + {\boldsymbol{\Sigma}_{\boldsymbol{\beta}}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)$. Therefore, the posterior can be written as

$$p(\boldsymbol{\beta} \mid \boldsymbol{y}, \sigma^2) \propto \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}^\top \mathbf{A} \boldsymbol{\beta} - 2\boldsymbol{b}^\top \boldsymbol{\beta}\right)\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(-2\boldsymbol{b}^\top \boldsymbol{\beta} + \boldsymbol{\beta}^\top \boldsymbol{A} \boldsymbol{\beta}\right)\right). \tag{2.5}$$

After using further manipulation (see Hooten and Hefley (2019)), we observe that the full conditional distribution of $\boldsymbol{\beta}$ has the kernel of a Normal distribution with mean $\boldsymbol{A^{-1}b}$

and variance $\boldsymbol{A^{-1}}$.

$$\boldsymbol{\beta} \mid \boldsymbol{y}, \sigma^2 \equiv \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{y},\sigma^2} = \boldsymbol{A^{-1}b}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{y},\sigma^2} = \boldsymbol{A^{-1}}).$$

According to Gelfand et al. (1992), once we know the form of the unconstrained full conditional for the parameter of interest, we can then draw realizations from this distribution using a Gibbs sampler. This procedure is straightforward and can be done by simply generating from the unconstrained full conditional and retaining the variate value only if it falls in the cross-section constraint region.

The conjugate full conditional distribution for $\boldsymbol{\beta}$ truncated at $\beta_1 < 0$ is then

$$\boldsymbol{\beta} \mid \boldsymbol{y}, \sigma^2 \equiv \mathcal{TN}(\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{y},\sigma^2} = \boldsymbol{A^{-1}b}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{y},\sigma^2} = \boldsymbol{A^{-1}}, \ 0, +\infty).$$

It is important to note that the full conditional distribution for σ^2 will also be affected by the constraints imposed to $\boldsymbol{\beta}$. Nonetheless, once we have the form of the unconstrained full conditional, we can use the same procedure as before retaining the variate value only if it falls in the cross-section constraint region.

The analytical form of the full conditional for σ^2 is

$$p(\sigma^2 \mid \boldsymbol{y}, \beta) \propto \begin{cases} p(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2)p(\sigma^2) & \boldsymbol{\beta} \in S \\ 0 & \boldsymbol{\beta} \notin S \end{cases}$$

Where $S = \left\{(\boldsymbol{y}, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{S_Y^k}\right\}$, and $\boldsymbol{S_Y^k}$ is the constraint set given by $\beta_1 > 0$.

Therefore, when $\beta_1 > 0$, we have:

$$p(\sigma^2 \mid \boldsymbol{y}, \beta) \propto \left(|2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right) \right) \frac{1}{r^q \Gamma(q)} (\sigma^2)^{-(q+1)} \exp\left(-\frac{1}{r\sigma^2} \right)$$

$$\propto (\sigma^2)^{-(q+\frac{n}{2}+1)} \exp\left(-\frac{1}{\sigma^2} \left(\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \frac{1}{r} \right) \right). \tag{2.6}$$

The result above clearly depicts the kernel of an Inverse Gamma distribution. Note that q is the location parameter whereas r^{-1} is the scale parameter of the prior Inverse Gamma distribution[3]. By letting $\tilde{q} = q + \frac{n}{2}$ and $\tilde{r} = \left[\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \frac{1}{r} \right]^{-1}$, we have the unconstrained full conditional distribution for σ^2 as a conjugate Inverse Gamma (Hooten and Hefley, 2019)

$$\sigma^2 \mid \boldsymbol{y}, \boldsymbol{\beta} \equiv \mathcal{IG}(\tilde{q}, \tilde{r}).$$

2.3 Results

This section presents the empirical analysis of the proposed Bayesian approach. We compared its performance against the traditional cointegration method (referred to as the 'Reference' strategy) in terms of cumulative returns, volatility, maximum drawdown, Sharpe ratio (Sharpe, 1966), and Sortino ratio (Sortino and van der Meer, 1991). Our dataset consists of equity pairs from the U.S. and Brazilian markets, selected based on historical price co-movement and suitability for statistical arbitrage. We obtained intraday price data at a 30-minute frequency from two primary sources: Alpha Vantage Inc. for the U.S. market and the MetaTrader 5 database of XP Investiments for the Brazilian market. The dataset spans a two-year period from February 1, 2023, to January 31, 2025. In cases where data

[3]In this work, we assign $q = 1/2$ and $r = 1$ as initial values for the Markov Chain - Monte Carlo simulation.

availability was constrained for one or both assets in a pair, we restricted our analysis to the overlapping period to ensure consistency in comparison.

For the trading strategy implementation, we utilized adjusted close prices to identify entry and exit signals, while simulating trade executions at the midpoint between the open and adjusted close prices of each 30-minute interval. This approach provides a more realistic representation of achievable execution prices compared to alternative methods such as using daily close-to-close prices, as it better accounts for intraday price dynamics and execution slippage that traders would encounter in real-world scenarios.

2.3.1 Backtesting Methodology

To test the proposed method, we implemented an algorithm using a walk-forward framework that continuously adjusts strategy parameters based on new data. To achieve that, we used a sliding window that moves incrementally throughout the price time series. The sliding window comprises three stages: formation, decision, and trading periods. The formation period has length w and is used to estimate the trading parameters.

Assuming the decision period in which we opt to trade or stay out of the market occurs at time t, the formation period spans from $t - w + 1$ to $t - 1$, whereas the trading period is expressed as $t + 1$. If the pair is not cointegrated in the formation period, we slide the rolling window one discrete time unit forward and re-evaluate the cointegration status. Alternatively, if the pair is cointegrated in the formation window, we draw random values from the full conditional of $\boldsymbol{\beta}$ using a Gibbs sampler, and the mean of the distribution for β_1, i.e. $\mathbb{E}[\beta_1 \mid \boldsymbol{y}, \beta_0, \sigma^2] = \int_0^\infty \beta_1 \; p(\beta_1 \mid \boldsymbol{y}, \beta_0, \sigma^2)\, d\beta_1$, is used as our hedge ratio coefficient.

A pre-defined α is used to obtain two quantiles q_α and $q_{1-\alpha}$ under the full conditional of β_1, which will serve as our entry and stop-loss confirmation thresholds. We then compute the standardized spread for the given period and define the standardized spread levels (L and U) to start an operation, i.e. sell the spread at U and buy the spread at L.

Once the trading parameters are defined, we proceed to the decision and trading stages. We start by fitting a linear regression model $\hat{y}_{[t-w+1:t]} = \hat{\beta}_{0,[t-w+1:t]} + \hat{\beta}_{1,[t-w+1:t]} \;\; x_{[t-w+1:t]}$ to

the period comprised of the entire formation window plus an additional data point, i.e. from $t-w+1$ to t, and store $\hat{\beta}_{1,[t-w+1:t]}$. If the standardized spread has crossed any of the defined boundaries and $\hat{\beta}_{1,[t-w+1:t]}$ is well behaved, i.e. $q_\alpha < \hat{\beta}_{1,[t-w+1:t]} < q_{1-\alpha}$, an entry signal is emitted for $t+1$. Once an operation is initiated in $t+1$, we start monitoring for exit signals from $t+1+k$ onward, where $k = \{1, 2, 3, \ldots\}$.

The take-profit signal is activated when the standardized spread reverts to its theoretical mean of 0 within subsequent k periods. A stop-loss signal is triggered at time $t+1+k$ whenever the standardized spread crosses a pre-established loss margin ξ (either $L-\xi$ or $U+\xi$), and the updated hedge ratio $\hat{\beta}_{1,[t-w+1:t+1+k]}$ derived from a linear regression fitted from the start of the formation period to the current time falls outside the specified quantiles, i.e. $\hat{\beta}_{1,[t-w+1:t+1+k]} \leq q_\alpha$ or $\hat{\beta}_{1,[t-w+1:t+1+k]} \geq q_{1-\alpha}$. This process is repeated periodically until a take-profit or stop-loss signal is emitted, and the operation is consequently terminated. Once a trade is closed, the algorithm is restarted in the next period. All procedures are summarized using pseudocode in Algorithm 1.

Algorithm 1 Backtesting the Bayesian approach for distribution-based trading signals

Require: Time series observations $\{(x_t, y_t)\}_{t=1}^{n}$, sliding window length w, MCMC replications M, negative $\hat{\beta}_1$ tolerance m, lower bound entry threshold L, upper bound entry threshold U, stop-loss margin ξ, quantile parameter α

Require: $w \leq n$ ▷ Sliding window at most the same size as the backtest data

flag $\leftarrow FALSE$ ▷ Flag indicating the existence of an active operation

for t in $[(w+1) : n]$ **do**

if flag is FALSE **then**

Run cointegration test on data $\{(x_t, y_t) : t \in [t-w+1 : t-1]\}$

if pair is cointegrated **then**

Run Gibbs sampler on data $\{(x_t, y_t) : t \in [t-w+1 : t-1]\}$ and obtain the full conditional $\boldsymbol{\beta}|\boldsymbol{y}, \sigma^2$

if $\boldsymbol{\beta_1}[1:m] < 0$ **then** ▷ First m drawn β_1's were all negative

Discard the cointegration status for the period ▷ Using the algorithm to perform pair selection

else

Compute $\hat{\beta}_{1,[t-w+1:t-1]} = \mathbb{E}[\beta_1 \mid \boldsymbol{y}, \beta_0, \sigma^2]$

Compute q_α and $q_{1-\alpha}$

Compute spread $\delta_{[t-w+1:t-1]} = y_{[t-w+1:t-1]} - \hat{\beta}_{1,[t-w+1:t-1]}\ x_{[t-w+1:t-1]}$ on data $\{(x_t, y_t) : t \in [t-w+1 : t-1]\}$

Compute standardized spread at time t:

$z[t] = \frac{\delta[t] - \bar{\delta}_{[t-w+1:t-1]}}{s_{\delta_{[t-w+1:t-1]}}}$

Fit $\hat{y}_{[t-w+1:t]} = \hat{\beta}_{0,[t-w+1:t]} + \hat{\beta}_{1,[t-w+1:t]}\ x_{[t-w+1:t]}$ to data $\{(x_t, y_t) : t \in [t-w+1 : t]\}$

Store $\hat{\beta}_{1,[t-w+1:t]}$

if $(z[t] \geq U$ or $z[t] \leq L)$ and $q_\alpha < \hat{\beta}_{1,[t-w+1:t]} < q_{1-\alpha}$ **then** ▷ Look for entry triggers

ENTRY Signal on $t+1$

flag $\leftarrow$ TRUE

end if

end if

end if

Algorithm 1 Backtesting the Bayesian approach for distribution-based trading signals (continued)

else ▷ Look for exit triggers

Fit $\hat{y}_{[t-w+1:t+1+k]} = \hat{\beta}_{0,[t-w+1:t+1+k]} + \hat{\beta}_{1,[t-w+1:t+1+k]}\ x_{[t-w+1:t+1+k]}$ to data $\{(x_t, y_t) : t \in [t-w+1 : t+1+k]\}$ where $t+1+k$ represents current time

Store new $\hat{\beta}_{1,[t-w+1:t+1+k]}$

Compute spread $\delta_{[t-w+1:t+1+k]} = y_{[t-w+1:t+1+k]} - \hat{\beta}_{1,[t-w+1:t+1+k]}\ x_{[t-w+1:t+1+k]}$ on data $\{(x_t, y_t) : t \in [t-w+1 : t+1+k]\}$

Compute standardized spread at current time $t+1+k$:

$z[t+1+k] = \frac{\delta[t+1+k] - \bar{\delta}_{[t-w+1:t+1+k]}}{s_{\delta_{[t-w+1:t+1+k]}}}$

if ($z[t+1+k] \geq U+\xi$ or $z[t+1+k] \leq L-\xi$) and ($\hat{\beta}_{1,[t-w+1:t+1+k]} \leq q_\alpha$ or $\hat{\beta}_{1,[t-w+1:t+1+k]} \geq q_{1-\alpha}$) **then**

EXIT Signal ▷ Stop-loss signal

flag $\leftarrow$ FALSE

else if ($z[t+1+k] < 0$ or $z[t+1+k] > 0$) **then**

EXIT Signal ▷ Take-profit signal

flag $\leftarrow$ FALSE

end if

end if

end for

2.3.2 Parameter Specification and Sensitivity Analysis

Our model requires the specification of several parameters, which can be categorized into two groups. The first group includes parameters common to both our Bayesian approach and the standard cointegration strategy. These parameters include the formation window length (w), which determines the historical period used for estimating the hedge ratio, the standardized spread thresholds (L and U), which define entry and exit conditions for trades, and the stop-loss margin (ξ), which prevents excessive losses by triggering an exit when the spread exceeds a predefined limit. The take-profit margin is implicitly defined at the equilibrium point (0), ensuring that positions are closed when the spread reverts to its historical mean.

The second group consists of parameters unique to the Bayesian framework. The quantile parameter (α), is unique and has a key role in defining probabilistic entry and exit signals. Unlike the standard approach, which relies on fixed spread levels, our Bayesian method utilizes the posterior distribution of the hedge ratio to dynamically adjust trade execution. The quantile parameter determines the degree of certainty required before executing a trade, thereby influencing both trading frequency and risk exposure.

To assess the robustness of our model, we conduct sensitivity analysis on key parameters,

examining their impact on performance metrics such as Sharpe ratio, maximum drawdown, and cumulative returns. We systematically vary the formation window length, adjusting the historical data used for estimation, and analyze how changes in spread thresholds and quantile levels affect trading outcomes. The results of this analysis provide insights into the optimal parameter configurations and highlight the adaptability of our Bayesian framework in different market environments.

Formation Window (w)

For intraday data in the 30-minute timeframe, our analysis indicates that the appropriate length w for the formation period ranges between 65 and 130 observations, representing approximately two and four trading weeks of data, respectively. These values were selected for testing as they provide sufficient data points for robust statistical analysis while remaining responsive to market conditions.

The sensitivity analysis revealed no consistent advantage between using $w = 65$ or $w = 130$ across all trading pairs. Some pairs performed better with the shorter window ($w = 65$), while others benefited from the extended formation period ($w = 130$). This variation likely reflects differences in individual pairs' underlying price dynamics and cointegration characteristics.

The absence of a clear optimal window length aligns with theoretical expectations. Excessively small windows can increase the chances of an incorrect rejection of the cointegration hypothesis (Type II errors) due to low statistical power, higher estimator variance, and insufficient capture of long-run relationships (Cheung and Lai, 1993; Otero and Smith, 2000). In theory, larger samples should improve cointegration detection by increasing statistical power; however, in practice, too large time series tend to result in a rejection of cointegration status due to the inclusion of structural breaks or regime changes that disrupt long-term equilibrium relationships (Gregory and Hansen, 1996; Hakkio and Rush, 1991). This dual challenge in window selection has been well documented in the econometric literature (Engle and Granger, 1987; Johansen, 1988, 1991), highlighting the importance of finding an appro-

priate balance in window sizing: large enough to ensure statistical robustness but constrained enough to avoid capturing multiple structural changes that could mask existing cointegrating relationships.

Standardized Spread Thresholds (L and U)

Our analysis reveals that thresholds of ± 1 standard deviation yield optimal results for he proposed strategies. This configuration strikes an effective balance between identifying trading opportunities and filtering out market noise. Notably, the method maintains robust performance and consistently outperforms the reference cointegration approach across all analyzed pairs, even when using the more conservative ± 2 standard deviation thresholds.

This resilience to threshold adjustments demonstrates the stability of the proposed method across varying entry levels. The consistent performance across different threshold settings suggests that the Bayesian framework effectively captures the underlying dynamics of pair relationships regardless of the specific entry criteria employed, highlighting its adaptive capability in diverse market conditions for the analyzed pairs.

Stop-loss Margin (ξ)

The stop-loss margin ϵ in the cointegration method defines the level of deviation in the standardized spread that, when exceeded, triggers the automatic closure of the position to limit potential losses. Typically, this parameter requires different specifications depending on the chosen levels for the standardized thresholds and the take-profit margin, allowing investors to implement proper risk management.

In our study, we tested ξ in two configurations: half of the take-profit margin or equal to it. This approach was applied to both algorithms. Although the best results were obtained using a 1/1 ratio for the take-profit/stop-loss margin, the proposed Bayesian method performed well across all evaluated scenarios among the analyzed pairs. From a practical standpoint, it is well established that adopting smaller entry thresholds, such as ± 1, increases the number of operations but may also introduce considerable noise. We hypothesize that the Bayesian

method's superior performance in high-risk scenarios stems from its enhanced capability to discriminate between noise and genuine trading signals. This characteristic, combined with an increased operational frequency, enables the algorithm to capitalize on more opportunities while maintaining precision, ultimately generating amplified returns in contexts where the traditional method is less effective.

Quantile Parameter (α)

The quantile parameter α exerts significant control over the responsiveness of the proposed model. This parameter effectively regulates the occurrence of entry and stop-loss signals in the following manner:

A smaller α value (e.g., 0.05) expands the area in which β_1 is assumed to be well-behaved, and make the Bayesian method less selective for entry signals and more selective for stop-loss signals. A larger α value (e.g., 0.15) creates a narrower middle band, which filters more potential entries compared to the reference method, but makes the algorithm less selective and thus more responsive to exit signals from the standard cointegration layer (Figure 2.1).

2.3.3 Aggregate Performance Evaluation

The empirical results presented in Table 2.1 demonstrate that the Bayesian strategy consistently outperforms the traditional cointegration approach in terms of both profitability and risk management across the tested asset pairs. The Bayesian method exhibits superior risk-adjusted performance, characterized by lower volatility, reduced drawdowns, and higher cumulative returns over the evaluation period. This improved performance is achieved despite an increased number of trades, as the Bayesian framework allows for greater accuracy in trade execution, optimizing both entry and exit signals.

The proposed method not only enhances profitability but also provides more reliable risk management, making it a superior alternative to the standard cointegration strategy. Even in cases where absolute returns are not significantly higher, incorporating the Bayesian approach as a foundational algorithm within a broader trading strategy can substantially

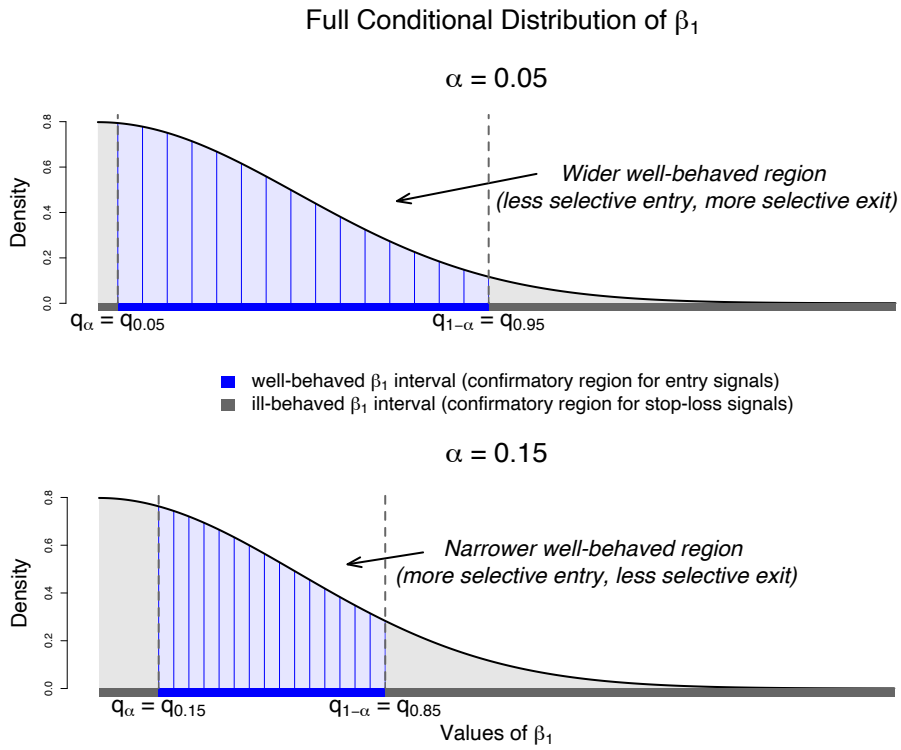

Figure 2.1: Quantile parameter α and 'well-behaved' intervals for β_1 under the full conditional distribution. The upper and lower panels illustrate how different values of α affect the confirmatory regions for entry and exit signals. For $\alpha = 0.05$ (upper panel), the well-behaved region is wider, resulting in less selectivity for entry signals and more selectivity for exit signals. For $\alpha = 0.15$ (lower panel), the well-behaved region is narrower, resulting in more selectivity for entry signals and less selectivity for exit signals. The blue regions represent intervals where β_1 is considered well-behaved, while the gray regions indicate intervals where β_1 falls outside the quantiles defined in the full conditional distribution.

improve overall performance.

2.3.4 US Performance

Equity Pairs: Stock Performance Analysis

In the equities segment, the comparison between Lennar Corporation (LEN) and D.R. Horton, Inc. (DHI) highlights the Bayesian strategy's effectiveness in cyclical sectors. While the reference strategy declined approximately 20% below its initial value, the Bayesian approach achieved a cumulative return of 1.662 compared to 0.812, demonstrating its ability to capitalize on mean-reversion opportunities. Additionally, the Bayesian model exhibited

Table 2.1: Comparison between the Bayesian ("Bayes") and the standard cointegration methods ("Ref.") for cumulative returns as a multiple of the initial investment and risk-return metrics from 02/01/2023 to 01/31/2025 on selected pairs in the U.S. and Brazilian markets. The lower (L) and upper (U) standardized spread thresholds were set at ± 1, with a stop-loss margin $\xi = 1$ and a take-profit threshold at equilibrium (0).

Parameters			Cum. Returns		Trade Count		Volatility		Max. Drawdown		Sharpe		Sortino	
Quantile	Window	Pair	Ref.	Bayes	Ref.	Bayes	Ref.	Bayes	Ref.	Bayes	Ref.	Bayes	Ref.	Bayes
US														
$\alpha = .05$	$w = 65$	REMX-LIT	0.572	2.625	178	313	0.22	0.138	0.45	0.093	-0.127	0.722	-0.179	1.107
		GLD-IAU	0.973	1.264	720	1204	0.042	0.02	0.094	0.019	-0.039	1.084	-0.054	2.064
		KO-PEP	0.856	1.179	232	337	0.088	0.027	0.165	0.03	-0.121	0.58	-0.172	0.839
	$w = 130$	IWM-VTWO	1.17	1.904	303	709	0.08	0.051	0.113	0.093	0.223	1.203	0.316	1.769
		VOO-IVV	1.024	1.111	525	924	0.063	0.061	0.132	0.129	0.066	0.193	0.094	0.274
		IEF-TLT	0.932	1.11	93	372	0.047	0.031	0.092	0.037	-0.116	0.327	-0.162	0.468
		SPY-IVV	0.991	1.11	578	919	0.062	0.059	0.147	0.128	0.017	0.196	0.024	0.277
$\alpha = .15$	$w = 65$	LEN-DHI	0.812	1.662	123	502	0.25	0.142	0.343	0.264	0.048	0.406	0.068	0.588
		VUG-IWF	0.611	2.016	107	280	0.316	0.113	0.356	0.13	0.017	0.633	0.024	1.295
		VNQ-IYR	0.793	2.108	546	693	0.083	0.055	0.304	0.039	-0.217	1.305	-0.282	2.334
		QQQ-QQQM	0.955	1.288	147	972	0.084	0.028	0.164	0.039	-0.008	0.868	-0.012	1.31
	$w = 130$	V-MA	0.873	1.209	208	370	0.105	0.056	0.17	0.108	-0.068	0.344	-0.098	0.493
Brazil														
$\alpha = .05$	$w = 65$	ELET3-ELET6	0.836	1.287	32	203	0.046	0.026	0.158	0.051	-0.34	0.927	-0.471	1.462
		GGBR4-GOAU4	0.975	1.312	114	227	0.053	0.029	0.142	0.025	-0.019	0.894	-0.027	1.389
		SPXI11-IVVB11	0.988	1.336	478	494	0.025	0.02	0.077	0.012	-0.032	1.343	-0.045	2.2
	$w = 130$	PETR4-PETR3	0.972	1.144	61	142	0.024	0.021	0.107	0.022	-0.099	0.618	-0.137	0.939
		ITSA4-ITUB4	0.93	1.181	30	127	0.04	0.025	0.101	0.021	-0.151	0.635	-0.213	0.944

lower volatility (0.142 vs. 0.25) and a significantly reduced maximum drawdown (0.264 vs. 0.343). This supports the Bayesian method's superior precision in detecting and exploiting temporary price distortions in cyclical industries.

In the financial sector, the performance of Visa Inc. (V) and Mastercard Incorporated (MA) further validates the Bayesian approach for highly correlated financial equities. The Bayesian strategy generated a cumulative return of approximately 20%, while the reference strategy declined by nearly 10%. Furthermore, the Bayesian model reduced volatility by nearly 50% (0.056 vs. 0.105), reinforcing its ability to stabilize returns while maintaining positive performance. The V-MA pair exemplifies how Bayesian filtering techniques effectively capture short-term pricing inefficiencies in highly correlated stocks.

Similarly, in consumer staples, the Coca-Cola Company (KO) versus PepsiCo, Inc. (PEP) pair demonstrated consistent positive performance under the Bayesian approach, while the reference strategy exhibited a steady decline. The Bayesian method reduced volatility by

69% (0.027 vs. 0.088) and maximum drawdown by 82% (0.03 vs. 0.165), confirming its effectiveness in stabilizing returns in defensive equity pairs.

Index and Sector Pairs: Index ETF Performance Analysis

The SPDR S&P 500 ETF Trust (SPY) versus iShares Core S&P 500 ETF (IVV) and Vanguard S&P 500 ETF (VOO) versus iShares Core S&P 500 ETF (IVV) pairs saw modest but meaningful outperformance under the Bayesian approach. The strategy's ability to extract value in highly liquid instruments suggests that its sophisticated signal processing remains effective even in markets with minimal arbitrage opportunities.

In the technology sector, the analysis of Invesco QQQ Trust (QQQ) versus Invesco NASDAQ 100 ETF (QQQM) demonstrates that the Bayesian method extracts meaningful alpha despite the high correlation between these Nasdaq-100 tracking instruments. The Bayesian strategy achieved approximately 30% cumulative returns, whereas the reference strategy experienced a slight decline. Additionally, the Sharpe ratio improved from -0.008 to 0.868, and the Sortino ratio from -0.012 to 1.31, reinforcing the Bayesian model's superior entry and exit timing for risk-adjusted performance.

The large-cap growth ETF pair of Vanguard Growth ETF (VUG) versus iShares Russell 1000 Growth ETF (IWF) highlights the Bayesian strategy's ability to capitalize on market inefficiencies. The Bayesian model effectively captured pricing dislocations, yielding a Sortino ratio of 1.295 compared to the reference method's 0.024. The strong risk-adjusted returns observed across growth and S&P 500 ETFs further confirm the Bayesian strategy's ability to dynamically adjust in shifting market conditions.

Similarly, in small-cap equity ETFs, the iShares Russell 2000 ETF (IWM) versus Vanguard Russell 2000 ETF (VTWO) pair confirmed the Bayesian strategy's robustness. The

Bayesian method achieved a Sharpe ratio of 1.203 versus 0.223 under the reference strategy, with comparable drawdowns but significantly higher returns, demonstrating its ability to capture fundamental arbitrage opportunities rather than merely exploiting sector-specific anomalies.

The real estate and commodity sectors exhibited some of the most pronounced performance differentials. The Vanguard Real Estate ETF (VNQ) versus iShares US Real Estate ETF (IYR) pair yielded over 100% returns under the Bayesian strategy, compared to significantly lower gains in the reference strategy. Notably, the Bayesian approach reduced maximum drawdown from 0.304 to 0.039 and improved the Sortino ratio from -0.282 to 2.334, demonstrating its resilience in capturing short-term pricing dislocations while mitigating downside risk.

Similarly, in the commodities sector, the VanEck Rare Earth/Strategic Metals ETF (REMX) versus Global X Lithium & Battery Tech ETF (LIT) pair showed even greater outperformance, with returns exceeding 150% under the Bayesian strategy. Volatility was substantially reduced (0.138 vs. 0.22), reinforcing the Bayesian method's ability to optimize alpha extraction while maintaining risk efficiency.

2.3.5 Brazil Performance

Equity Pairs: Stock Performance Analysis

The Bayesian strategy demonstrated remarkable performance in Brazil's equity market, affirming the superiority of the proposed method across diverse sectors. The performance displayed for PETR4-PETR3 (Petróleo Brasileiro S.A. preferred shares versus Petróleo Brasileiro S.A. common shares) in the oil and gas sector shows that the Bayesian strategy effectively capitalizes on dislocations between Petrobras' share classes despite their fundamental linkage, achieving a cumulative return of 1.144 versus 0.972 for the reference strategy while reducing maximum drawdown by nearly 80% (0.022 compared to 0.107). This performance differential underscores the strategy's ability to capture price distortions and optimize trading signals even in securities with closely aligned fundamental attributes.

In the utilities sector, the Eletrobras - Centrais Elétricas Brasileiras S.A. ordinary shares (ELET3) versus Eletrobras - Centrais Elétricas Brasileiras S.A. preferred shares (ELET6) pair confirms the proposed method's superiority over standard cointegration by displaying the divergent performance trajectory between the two strategies. The Bayesian approach

displays consistent upward momentum while the reference strategy deteriorates steadily throughout the observation period. This pair exhibited dramatic improvements in risk-adjusted metrics, with Sortino ratio increasing from -0.471 to 1.462 and volatility reduced from 0.046 to 0.026.

The steel industry pair of Gerdau S.A. (GGBR4) versus Metalúrgica Gerdau S.A. (GOAU4) further illustrates this pattern, with the Bayesian strategy generating substantial positive returns exceeding 30% in the analyzed period (1.312 versus 0.975) while reducing volatility by 45% (0.029 versus 0.053) and maximum drawdown by 82% (0.025 versus 0.142). This contrasts markedly with the reference approach, which exhibits considerable volatility without meaningful cumulative gains.

The financial sector pair Itaúsa S.A. (ITSA4) versus Itaú Unibanco Holding S.A. (ITUB4) exhibits stepwise performance improvements under our method, with cumulative returns of 1.181 versus 0.930 for the reference strategy and significantly improved Sharpe (0.635 versus -0.151) and Sortino (0.944 versus -0.213) ratios. This contrasts with the standard cointegration, which fails to maintain consistent performance despite periodic recoveries.

Index and Sector Pairs: Index ETF Performance Analysis

Particularly notable is the performance differential in the Brazilian S&P 500 ETF pair of SPDR S&P 500 Brazil (SPXI11) versus iShares S&P 500 Brazil (IVVB11), where the Bayesian strategy generates exceptionally consistent returns throughout the observation period, culminating in gains exceeding 30% (1.336 versus 0.988) and with number of trades almost identical to the reference strategy. This pair exhibited extraordinary improvement in risk metrics, with maximum drawdown reduced by 84% (0.012 versus 0.077) and Sortino ratio increasing dramatically from -0.045 to 2.2. This demonstrates the methodology's effectiveness even when applied to instruments tracking the same underlying index in an emerging market context.

These findings suggest that emerging markets like Brazil may offer particularly fertile ground for the Bayesian pairs trading methodology, potentially due to reduced algorithmic

trading activity, wider bid-ask spreads, and more frequent pricing dislocations between related securities. The strategy's consistent outperformance across diverse Brazilian sectors, with considerably higher trade frequency (e.g., 203 trades versus 32 for ELET3-ELET6) yet substantially lower volatility, underscores its robustness in varied market contexts beyond developed markets.

2.3.6 General Remarks

The results demonstrate that the Bayesian approach delivers more stable performance compared to the standard cointegration strategy across the analyzed markets and asset classes. Visual inspection of cumulative returns (Figures 2.2 and 2.3) reveals that the proposed algorithm effectively refines entry and exit points. This confirms our hypothesis that the method produces more timely trading signals, optimizing trade execution and yielding more stable and reliable returns. Furthermore, significant differences in risk metrics between the two methods highlight the clear advantages of the proposed algorithm.

Lastly, the results also demonstrate that the proposed Bayesian method can serve as a foundational algorithm in more sophisticated strategies. In practice, it is common to see cointegration strategies combined with other techniques to enhance pair selection or trading signal generation. In this regard, since our method has proven to deliver better performance and more effective risk management, it can significantly enhance the overall strategy by replacing traditional cointegration in a compound pairs trading strategy.

In the following section, we will present simulations demonstrating how the Bayesian method reduces the occurrence of false positives in cointegration tests. This indicates that part of our method has broader applications — not only for pairs trading but also for improving the power of cointegration tests in some scenarios.

2.4 Simulation Study for Pair Selection

Some parameter combinations were tested for window size $w = \{65, 130\}$, lower and upper standardized spread thresholds $L = \pm\{1, 2\}$ and $U = \pm\{1, 2\}$, and alpha level $\alpha = \{0.05, 0.15\}$ for the quantiles q_α and $q_{1-\alpha}$ from the full conditional of β_1. After a thorough assessment of the cumulative returns and risk-return metrics, we selected the optimal setup for each pair for our analysis.

In this section, we evaluate the proposed method's effectiveness in identifying false positives from cointegration tests through simulation studies. While cointegration tests can yield both false positives and false negatives, false positives present a greater risk to investors. Trading pairs incorrectly identified as cointegrated can lead to substantial losses, as their prices may not maintain the assumed long-term equilibrium relationship. This risk amplifies when conducting multiple tests. Our proposed Bayesian method addresses this challenge by enhancing the pair selection process, offering a heuristic mechanism to detect potential false positives.

The cointegration test methodology includes a unit root test on the residuals as its second step, where the null hypothesis assumes these residuals are non-stationary. When evaluating this assumption, a type I error occurs if we incorrectly reject a true null hypothesis, leading to a false identification of cointegration. In other words, we might conclude that two price series are cointegrated when they actually have no long-term equilibrium relationship. Although we can theoretically control the type I error rate by establishing a significance level α, there is a specific situation in which we can significantly inflate the probability of observing false positives when applying the cointegration test in pairs trading. Clegg (2014) states that the use of a data mining procedure to identify cointegrated pairs on a list of potential candidates tends to inflate the chances of false positives.

If we perform multiple cointegration tests sequentially to identify potentially cointegrated pairs in an exchange, for example, the probability of observing at least one false positive increases beyond the nominal significance level α, resulting in an undesirably high number of false positives. This phenomenon is known as the family-wise error rate (FWER). As a

consequence, mistakenly selecting non-cointegrated pairs poses significant risks for investors, as the long-term equilibrium between the two assets is wrongly assumed. This undermines the assumption of a long-term equilibrium and leads to unreliable trading strategies. Therefore, correction methods based on the False Discovery Rate (FDR), such as the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), are unsuitable in this context because even a small number of false positives can incur significant costs.

Conversely, techniques like the Bonferroni correction (Bonferroni, 1936; Haynes, 2013) or the Holm-Bonferroni method (Holm, 1979) that control the FWER are excessively conservative. While they effectively reduce false positives, they significantly inflate the type II error rate, leading to the rejection of many genuinely cointegrated pairs. This is particularly problematic when using a data-mining approach that involves a large number of tests. For instance, considering the S&P500 index, the number of possible pairs exceeds 500^2. In such a scenario, the type II error would become so severe that nearly all pairs would be rejected, potentially excluding viable investment opportunities.

With that in mind and acknowledging the benefits of using the Bayesian framework, we hypothesized in Section 2.2.1 that the proposed method is not only capable of generating more accurate trading signals but also executing a more precise pair selection. This is mainly due to the adoption of a truncated normal as the prior distribution of $\boldsymbol{\beta}$, which also causes the full conditional to be a truncated normal. As presented in Section 2.2.1, by truncating the distribution at $\beta_1 < 0$, we can identify pairs that do not move together most of the time - which is a basic premise of the pairs trading strategy. This adds an extra layer of confirmation to the cointegration test, and, as we hypothesize, can help reduce the number of false positives through prior beliefs and heuristics.

The empirical results have shown that the Bayesian method delivers a superior performance in terms of profitability and risk management when compared to standard cointegration. Nonetheless, in order to evaluate the specific hypothesis of enhancement of the pair selection, we ran a set of simulations with both cointegrated and non-cointegrated series to determine if our algorithm is capable of detecting false positives in the cointegration test.

2.4.1 Design and Results

A total of 60,000 simulations were performed, organized into six batches of 10,000 simulations each. For each batch, 1,000 samples were generated from each of six selected window lengths {60, 90, 120, 180, 252, 500}, and combined with five different β_1 values {0.5, 0.75, 1.0, 1.25, 1.5} for the ordinary least squares (OLS) regression between the series y_t and x_t. Lastly, each possible combination of data size and β_1 was simulated in two scenarios—one in which the two series are co-integrated and one in which they are not.

To simulate a cointegrated relationship, we generate two series, $y_t = \sum_{i=1}^{t} \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0,1)$ i.i.d., and $x_t = \beta_1 y_t + \eta_t$, where $\eta_t \sim \mathcal{N}(0,1)$ i.i.d. The two series are non-stationary because they are cumulative sums of white noise, i.e. they follow a random walk. However, since x_t is a linear combination of y_t with an additional noise component, the regression of y_t on x_t should produce stationary residuals, thus characterizing the cointegration.

For a non-cointegrated relationship, we start by also simulating two non-stationary series, $y_t = \sum_{i=1}^{t} \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0,1)$ i.i.d., and $x_t = \sum_{i=1}^{t} \nu_i$, where $\nu_i \sim \mathcal{N}(0,1)$ i.i.d. Nonetheless, x_t is not represented as a linear combination of y_t anymore. Thus, regressing one on the other should not reveal a significant long-term relationship since both series are generated independently (each as a separate random walk). Therefore the residuals from this regression should be non-stationary.

Once we have the original series and their true nature, each simulated run is tested for cointegration using the Phillips-Perron framework at $\alpha = 0.05$ significance level and the result is stored. Our method is applied immediately by drawing realizations from the full conditional of $\boldsymbol{\beta}$. If a sequence of $m = 30$ negative β_1 values is drawn, then an annotation indicating the pair should be discarded is stored in our database.

When there was a false positive result from the cointegration test and our method correctly suggested discarding the pair we counted this as a correct decision. We also examined potential errors in the proposed method related to erroneously discarding pairs. A discarding error was recorded whenever our method recommended rejecting a pair that was, in fact, cointegrated, regardless of the cointegration test outcome.

Table 2.2 below show the summary of the results obtained in the simulations. Our method correctly identified around 75% of the total false positives in the simulations. Although it is not shown in the table, it is also important to highlight that the algorithm did not erroneously discard any truly cointegrated pairs.

Table 2.2: Performance of the Bayesian method in correctly rejecting false positives across different time series sizes and β_1 values. The table summarizes the number of false positives correctly rejected, the total number of false positive indications in the cointegration method, and the proportion of correct rejections for each combination of size and β_1 tested.

	Size						
	500	252	180	120	90	60	Total
$\beta_1 = 0.50$							
Cointegration False Positives	73	36	70	71	52	0	302
Correctly Rejected	43	17	70	70	35	0	235
Proportion	0.59	0.47	1.00	0.99	0.67	-	0.78
$\beta_1 = 0.75$							
Cointegration False Positives	73	36	70	72	34	2	287
Correctly Rejected	44	19	70	72	17	1	223
Proportion	0.60	0.53	1.00	1.00	0.50	0.50	0.78
$\beta_1 = 1.00$							
Cointegration False Positives	76	33	70	72	68	0	319
Correctly Rejected	43	16	70	72	17	0	218
Proportion	0.57	0.48	1.00	1.00	0.25	-	0.68
$\beta_1 = 1.25$							
Cointegration False Positives	71	37	70	71	37	1	287
Correctly Rejected	41	20	70	70	18	0	219
Proportion	0.58	0.54	1.00	0.99	0.49	0.00	0.76
$\beta_1 = 1.50$							
Cointegration False Positives	75	35	72	71	17	0	270
Correctly Rejected	45	17	72	70	0	0	204
Proportion	0.60	0.49	1.00	0.99	0.00	-	0.76
Overall Totals							
Cointegration False Positives	368	177	352	357	208	3	1465
Correctly Rejected	216	89	352	354	87	1	1099
Proportion	0.587	0.503	1.00	0.99	0.418	0.33	0.75

In summary, the simulation results indicate that the Bayesian method can correctly identify a considerable amount of false positive results from cointegration tests across various data sizes and β_1 values without erroneously discarding truly cointegrated pairs. The results suggest that this approach can serve as a valuable tool for enhancing cointegration analysis in general applications and not only in pairs trading. Further research and refinement may

improve the method's precision in pair selection and extend its application to other contexts. Future work could also focus on deriving theoretical results that would allow a transition from its current heuristic state to a more theoretically grounded approach.

2.5 Concluding Remarks

We developed a Bayesian method for pairs trading based on cointegration that improves profitability and risk management by dynamically adjusting trading signals based on the full conditional distribution of the hedge ratio β_1. The method applies Bayesian hierarchical modeling, incorporating truncated normal priors to refine pair selection and optimize entry and exit signals.

The proposed Bayesian algorithm significantly outperformed the standard cointegration strategy in all the analyzed pairs. The method's ability to adjust trading signals dynamically, based on the distribution of the hedge ratio, has proven to be a significant enhancement over the traditional cointegration framework, especially in its capacity to generate more timely and accurate entry and exit signals. Our algorithm has consistently demonstrated superior risk metrics compared to standard cointegration, with substantial reductions in volatility, and remarkable improvements in risk-adjusted return measures, with Sharpe and Sortino ratios showing dramatic enhancements.

Additionally, the Bayesian approach has shown the potential to reduce false positive rates in cointegration tests, offering a more refined tool for pair selection. This finding also suggests that this method has applicability beyond pairs trading, offering potential benefits for areas other than quantitative finance that also rely on cointegration.

Future research could explore the performance of this strategy in a broader range of markets and asset classes, including emerging markets, fixed income, and more volatile sectors such as cryptocurrencies. Testing different parameter configurations for window size, full conditional distribution quantiles, and stop-loss/take-profit thresholds, could also provide new insight into optimizing strategy for different market conditions. Lastly, incorporating transaction costs into the analysis would give a more realistic assessment of the strategy's

net profitability.

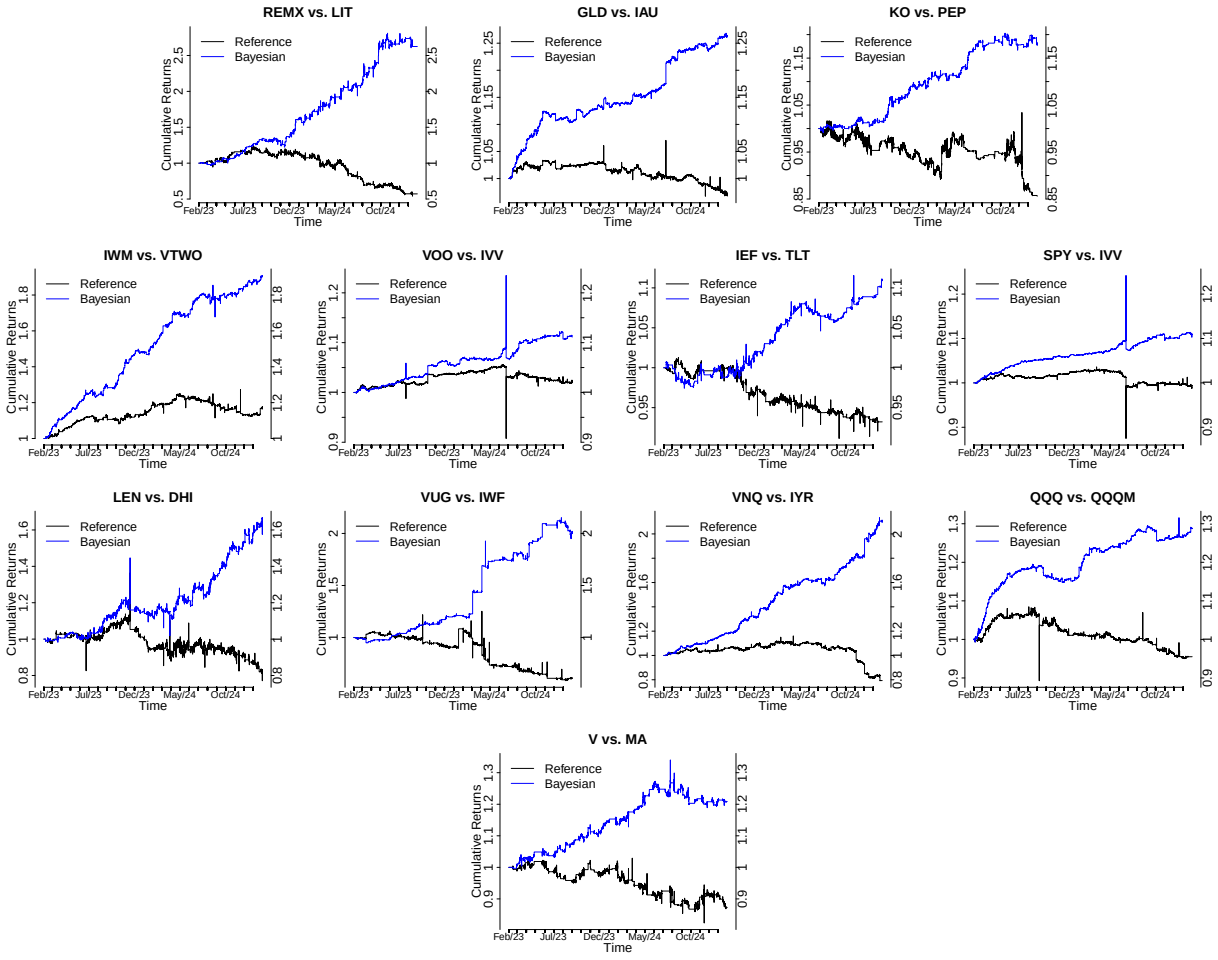

Figure 2.2: Cumulative returns (y-axis) as a multiple of the initial investment for selected pairs in the US market from 02/01/2023 to 01/31/2025 (x-axis). The blue curve indicates the cumulative returns of the proposed Bayesian method whereas the black curve shows the performance of the standard cointegration strategy. Each row was obtained using a different parameter combination for window length w and quantile parameter α. First row: $\alpha = 0.05$ and $w = 65$; second row: $\alpha = 0.05$ and $w = 130$; third row $\alpha = 0.15$ and $w = 65$; and fourth row $\alpha = 0.15$ and $w = 130$. The lower and upper standardized spread thresholds (L and U) used were ± 1; the stop-loss margin adopted was $\xi = 1$; and the take-profit standardized threshold spread was set as 0.

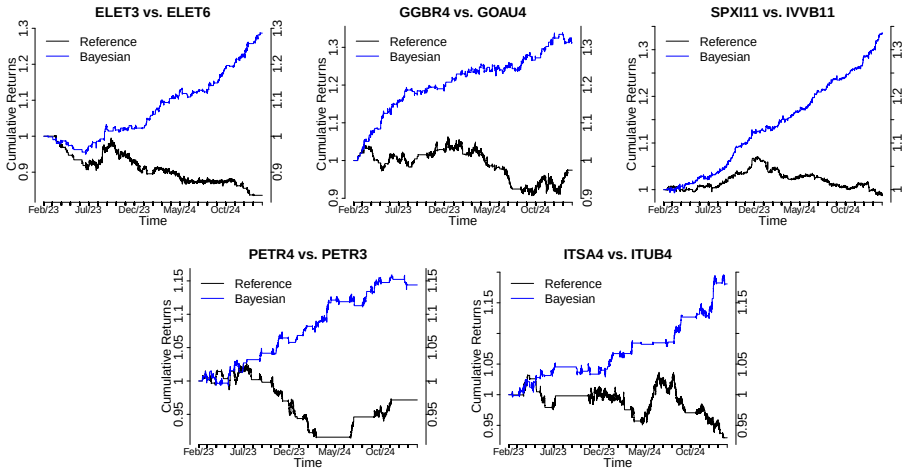

Figure 2.3: Cumulative returns (y-axis) as a multiple of the initial investment for selected pairs in the Brazilian market from 02/01/2023 to 01/31/2025 (x-axis). The blue curve indicates the cumulative returns of the proposed Bayesian method, whereas the black curve shows the performance of the standard cointegration strategy. Each row was obtained using a different parameter combination for window length w and quantile parameter α. First row: $\alpha = 0.05$ and $w = 65$; and second row: $\alpha = 0.05$ and $w = 130$. The lower and upper standardized spread thresholds (L and U) used were ± 1; the stop-loss margin adopted was $\xi = 1$; and the take-profit standardized threshold spread was set as 0.

Chapter 3

Non-Overlapping Block Bootstrap approach to distribution-based trading signals

3.1 Introduction

Pairs trading remains one of the most enduring and versatile statistical arbitrage strategies in quantitative finance, consistently generating interest among both practitioners and academics for its market-neutral characteristics and adaptability across various market conditions. As demonstrated in our previous work, the performance of traditional pairs trading strategies can be significantly enhanced by incorporating distribution-based approaches that capture the uncertainty in the hedge ratio estimation process.

Previously, we introduced a novel Bayesian approach to pairs trading that employed the full conditional distribution of the hedge ratio (β_1) to create a confirmatory layer for trading signals. This method successfully addressed key challenges in pairs trading: improving pair selection accuracy and generating more precise and timely trading signals. The empirical results across multiple markets and asset classes confirmed our hypothesis that incorporating distributional information about the hedge ratio leads to superior risk-adjusted performance

compared to traditional cointegration frameworks.

Building upon this distribution-based framework, we now introduce a nonparametric approach: the Non-Overlapping Block Bootstrap (NBB) method for pairs trading. While our Bayesian approach utilized hierarchical modeling with truncated normal priors to derive the distribution of the hedge ratio, the NBB method achieves a similar objective through resampling techniques that preserve the temporal dependence structure inherent in financial time series data.

3.1.1 Block Bootstrap Methods

Bootstrap methods have gained immense popularity in statistical inference for independent and identically distributed (i.i.d.) data, yet their application to dependent data structures such as time series has been comparatively limited. While Efron (1979) introduced the original bootstrap technique for i.i.d. data, extensions for dependent data became essential for applications in economics, finance, and other fields where temporal dependency is intrinsic to the data generation process. These adaptations for time-series data are commonly referred to as block bootstrap methods. They are specifically designed to preserve the dependence structure present in the original data. Several key approaches have emerged in the literature, each with distinct characteristics and applications.

The Moving Block Bootstrap (MBB), independently developed by Künsch (1989) and Liu and Singh (1992), constructs resamples by selecting blocks of consecutive observations from the original series. It operates by dividing the original time series into overlapping blocks of equal length and randomly sampling these blocks with replacement. This approach allows the dependency structure within each block to be preserved in the resampled data. The Non-Overlapping Block Bootstrap, conceptually based on Carlstein's (1986) idea of non-overlapping subseries for variance estimation, partitions the original series into disjoint blocks of equal length without any overlap. This method treats the data as naturally segmented into discrete chunks that are then resampled with replacement to form bootstrap samples. The NBB was further explored and systematically explained by Lahiri (2003) in

his comprehensive treatment of resampling methods for dependent data, where he analyzed its theoretical properties and performance relative to other block bootstrap methods. The Circular Block Bootstrap (CBB), proposed by Politis and Romano (1992), addresses the edge effects inherent in other block bootstrap methods by conceptualizing the time series as circular. This technique wraps the data around a circle, allowing blocks to include observations from both the end and the beginning of the series, ensuring each observation appears in exactly the same number of potential blocks.

Each approach offers specific advantages. The MBB utilizes overlapping blocks, providing more potential blocks (n-b+1 versus $\lfloor n/b \rfloor$ for NBB) and potentially reducing variance, but at the cost of increased computational complexity. Despite this theoretical advantage, our empirical testing showed that the MBB did not perform as well as the NBB for pairs trading applications. Besides that advantage, the NBB offers simpler implementation and clearer interpretation, with each observation belonging to exactly one block (except in cases with partial end blocks). This characteristic provides a direct correspondence between blocks and distinct periods in the time series, enhancing interpretability for financial applications like pairs trading. While the CBB provides theoretical advantages for certain statistics by addressing edge effects, these benefits may not outweigh the NBB's straightforward structure and implementation for many practical applications in finance.

Lahiri (2003) provides a comprehensive comparison of these methods, noting that while the MBB may achieve greater efficiency for some statistics, the NBB can offer computational advantages and more straightforward implementation. Importantly, for regression coefficients in particular, the performance differences between these methods tend to be less pronounced than for simpler statistics, such as sample means. This occurs because regression coefficients are influenced by relationships between variables rather than absolute values, making them more robust to the specific blocking scheme. Additionally, regression models inherently capture some of the dependency structure within the data, reducing the sensitivity to the exact block bootstrap method employed.

3.1.2 The non-overlapping block bootstrap approach in pairs trading

The NBB method represents an evolution in our distribution-based pairs trading framework. It generates a bootstrap distribution (also called empirical distribution) of the hedge ratio through resampling techniques that preserve the essential time-dependency characteristics of financial data. This approach is particularly valuable in pairs trading, where the relationship between assets may exhibit complex temporal dependencies that aren't fully captured by parametric models.

Like our Bayesian framework, the NBB method uses quantiles from the bootstrap distribution of the hedge ratio to create confirmatory thresholds for trading signals, maintaining the two-layer decision process that proved effective in enhancing strategy performance. However, the NBB approach offers distinct advantages in certain market conditions by accommodating nonparametric estimation of the hedge ratio distribution, potentially capturing market dynamics that might be challenging to model directly.

This chapter details the theoretical foundation, implementation framework, and performance characteristics of the NBB methodology. We demonstrate how this approach complements our earlier work, expanding the distribution-based pairs trading family with an alternative mechanism for generating robust trading signals. Our empirical analysis and simulation studies evaluate the strategy across diverse market conditions, demonstrating its effectiveness in both enhancing returns and managing risk.

By presenting this methodological extension, we continue to develop a comprehensive framework for distribution-based pairs trading strategies, offering practitioners a versatile toolkit that can be adapted to various market environments and investment objectives.

3.2 Methodology

3.2.1 Introduction

In this section, we introduce our Non-Overlapping Block Bootstrap approach to pairs trading, which builds upon the cointegration framework by dynamically enhancing the reliability of pair selection and refining the trading signals through probabilistic evaluation of the hedge ratio (denoted by β_1). The core of the strategy lies in generating the empirical distribution of the hedge ratio coefficient through a resampling technique that preserves the temporal dependence structure of the time series data while providing computational advantages through non-overlapping blocks.

With the bootstrap distribution derived, we identify two quantiles (e.g., 5^{th} and 95^{th}) from the distribution of β_1 and utilize them as confirmatory triggers for trading signals. This approach enables more precise and adaptive decision-making in pairs trading by incorporating the uncertainty and variability of the hedge ratio into the trading process.

We hypothesize that the proposed method is profit-wise superior to the regular cointegration framework because it may generate more timely trading signals through distribution-based confirmation thresholds.

Relying exclusively on the standardized spread for signal generation can result in entering trades just as the pair is about to lose its long-term equilibrium (cointegration) status or when the pair balance given by β_1 is excessively distorted. These two scenarios can lead to losses because the pair will not return to the long-term mean or because of an unbalanced share allocation. Furthermore, it is not uncommon for pairs trading practitioners to observe early stoppages that considerably affect the strategy's performance. We hypothesize that the adoption of the bootstrap distribution of β_1 as an extra layer of confirmation contributes to generating more trustworthy and timely trading signals, thus mitigating the aforementioned risks and improving the strategy's profitability.

We empirically evaluate the performance of our NBB method against the standard cointegration strategy in terms of profitability and risk metrics in Section 3.3. The patterns of

cumulative returns and drawdowns provide insights into the differences in trading signals between the two methods, aiding in the assessment of our hypothesis.

3.2.2 Theoretical background

Non-Overlapping Block Bootstrap Sampling Procedure

The NBB procedure generates bootstrap samples by randomly selecting k blocks with replacement from the collection $\{B_1, B_2, \ldots, B_k\}$ and concatenating them to form a new series. This resampling approach preserves the internal dependence structure within each block while treating different blocks as independent units.

Formally, let $I_1, I_2, \ldots, I_k$ be independent random variables, each uniformly distributed on $\{1, 2, \ldots, k\}$. These random indices determine which original blocks will be selected for inclusion in the bootstrap sample. Each I_j represents the index of the block from the original series that will become the j-th block in the bootstrap sample.

The bootstrap sample is then constructed as:

$$W_1^*, W_2^*, \ldots, W_m^*$$

where $m = kb$ is the total length of the bootstrap sample and

$$(W^*_{(j-1)b+1}, W^*_{(j-1)b+2}, \ldots, W^*_{jb}) = B_{I_j} = (W_{(I_j-1)b+1}, W_{(I_j-1)b+2}, \ldots, W_{I_jb}) \tag{3.1}$$

for $j = 1, 2, \ldots, k$.

This equation maps each block in the bootstrap sample to its corresponding block in the original sample. The left side of this equation represents the j-th block in the bootstrap sample, spanning from position $(j-1)b+1$ to position jb whereas the right side represents the I_j-th block from the original sample, which is selected to become the j-th block in the

bootstrap sample. The middle term B_{I_j} is a compact notation for the I_j-th original block.

The previous structure reveals that Non-Overlapping Block Bootstrap enables random reordering of the original data sequence, allowing blocks to appear in positions different from their original placement. Through resampling with replacement, some blocks may occur multiple times while others might not appear at all. This method preserves local dependence by maintaining the internal structure within each block, thus conserving temporal dependence among closely timed observations. Furthermore, NBB establishes independence between blocks by treating them as separate units, effectively capturing the concept that dependence naturally diminishes between observations distant in time. In summary, its resampling mechanism allows the NBB to capture the essential dependence structure of the time series while creating bootstrap samples that appropriately reflect the uncertainty in parameter estimation.

Adaptation for Partial End Blocks

In our implementation, we adopt a slight modification to the standard NBB approach to ensure complete coverage of the data series. While traditional NBB implementations often discard the remaining observations when n is not divisible by b, our approach incorporates a potential partial block at the end of the series. This is achieved by using $\lceil n/b \rceil$ instead of $\lfloor n/b \rfloor$ to determine the number of blocks, ensuring that all observations are included in the bootstrap process.

Mathematically, this adaptation modifies the block structure as follows:

$$B_j = \begin{cases} (W_{(j-1)b+1}, W_{(j-1)b+2}, \ldots, W_{jb}), & \text{if } j < k' \\ (W_{(j-1)b+1}, W_{(j-1)b+2}, \ldots, W_n), & \text{if } j = k' \end{cases} \tag{3.2}$$

where $k' = \lceil n/b \rceil$.

This adaptation has several theoretical implications. The inclusion of all data points potentially increases efficiency by utilizing the complete information available in the sample,

which may be particularly valuable in financial applications where recent data often contains important market information. Moreover, in pairs trading contexts, discarding the most recent observations could mean losing crucial signals for trading decisions.

However, this modification introduces blocks of unequal length, creating a deviation from the standard theoretical framework of the NBB. Lahiri (2003) discusses block bootstrap methods generally under the assumption of equal-sized blocks when establishing their asymptotic properties. The theoretical impact of this modification diminishes asymptotically as n increases, since the proportion of observations in the partial block becomes negligible relative to the entire sample.

The standard asymptotic theory for block bootstrap methods generally assumes that as $n \to \infty$, both the block length $b \to \infty$ and the number of blocks $n/b \to \infty$, while $b/n \to 0$. Under these conditions, the effect of a single partial block becomes asymptotically negligible. In finite samples, however, this modification may introduce additional variability in the bootstrap distribution, particularly if the partial block is substantially shorter than the full blocks and contains observations with distinct patterns.

3.2.3 Consistency of the Non-Overlapping Block Bootstrap for $\hat{\beta}_1$

To establish the consistency of the Non-Overlapping Block Bootstrap for the hedge ratio $\hat{\beta}_1$ in the linear regression model, we need to show that the bootstrap distribution of $\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1)$ correctly approximates the true sampling distribution of $\sqrt{n}(\hat{\beta}_1 - \beta_1)$. Formally, we must prove that:

$$\mathbb{P}^* \left(\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1) \leq x \right) \xrightarrow{p} \mathbb{P} \left(\sqrt{n}(\hat{\beta}_1 - \beta_1) \leq x \right) \quad \text{as } n \to \infty. \tag{3.3}$$

This requires demonstrating that the bootstrap estimator $\hat{\beta}_1^*$ satisfies the same asymptotic normality as the original OLS estimator $\hat{\beta}_1$, ensuring that the bootstrap variance estimator correctly converges to the true variance.

Convergence of the Bootstrap Variance to the True Variance

To ensure the validity of the bootstrap, we must show that the bootstrap variance estimator correctly approximates the true variance V, that is,

$$\text{Var}^*(\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1)) \xrightarrow{p} V. \tag{3.4}$$

We start by rewriting the OLS bootstrap estimator as:

$$\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1) = \frac{\sqrt{n}\sum_{t=1}^{n}(x_t^* - \bar{x}^*)u_t^*}{\sum_{t=1}^{n}(x_t^* - \bar{x}^*)^2}, \tag{3.5}$$

In Lahiri's original formulation for statistics based on sample means, the bootstrap variance is given by:

$$\text{Var}^*(T_n^*) = \frac{\ell}{n}\sum_{i=1}^{b}(U_i^{(2)} - \bar{U}^{(2)})^2 \tag{3.6}$$

where, $U_i^{(2)}$ represents the average of observations within each non-overlapping block; $\bar{U}^{(2)}$ is the overall mean of these block averages; ℓ is the block size; b is the number of blocks; and n is the sample size.

To adapt to the regression context, we replace: T_n^* with the regression estimator $\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1)$; and the block means $(U_i^{(2)})$ with block sums (S_i^*) of weighted residuals for each resampled block as:

$$S_i^* = \sum_{t=(i-1)l+1}^{il}(x_t^* - \bar{x}^*)u_t^*, \quad i = 1, \ldots, k \tag{3.7}$$

where, b is the block size and k is the number of blocks such that $n = kb$.

The numerator of the bootstrap estimator expression can be rewritten in terms of these blocks:

$$\sqrt{n}\sum_{t=1}^{n}(x_t^* - \bar{x}^*)u_t^* = \sqrt{n}\sum_{i=1}^{k} S_i^* \tag{3.8}$$

The conditional variance of the bootstrap estimator then becomes:

$$\mathrm{Var}^*(\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1)) = \mathrm{Var}^*\left(\frac{\sqrt{n}\sum_{i=1}^{k} S_i^*}{\sum_{t=1}^{n}(x_t^* - \bar{x}^*)^2}\right) \tag{3.9}$$

Since the denominator is considered non-random (conditioned on bootstrap observations), we have:

$$\mathrm{Var}^*(\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1)) = \frac{n}{\left(\sum_{t=1}^{n}(x_t^* - \bar{x}^*)^2\right)^2}\mathrm{Var}^*\left(\sum_{i=1}^{k} S_i^*\right) \tag{3.10}$$

This allows us to express the variance of the sum of blocks as the sum of covariances:

$$\mathrm{Var}^*\left(\sum_{i=1}^{k} S_i^*\right) = \sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{Cov}^*(S_i^*, S_j^*) \tag{3.11}$$

Using the fact that, by construction, blocks are independently resampled in the block bootstrap, we have:

$$\mathrm{Cov}^*(S_i^*, S_j^*) \approx 0, \quad \text{for } i \neq j \tag{3.12}$$

Therefore:

$$\sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{Cov}^*(S_i^*, S_j^*) = \sum_{i=1}^{k}\mathrm{Var}^*(S_i^*) \tag{3.13}$$

We thus arrive at the final expression:

$$\mathrm{Var}^*(\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1)) = \frac{n}{\left(\sum_{t=1}^{n}(x_t^* - \bar{x}^*)^2\right)^2}\sum_{i=1}^{k}\mathrm{Var}^*(S_i^*) \tag{3.14}$$

According to Lahiri (2003), when the regression error process $\{u_t\}$ satisfies certain standard moments and strong mixing conditions, we can conclude that equation (3.4) holds, i.e.:

$$\mathrm{Var}^*(\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1)) \xrightarrow{p} V$$

These assumptions hold so long as the block length b grows to infinity, but at a slower rate than the sample size n, i.e. $b^{-1} + n^{-1}b = o(1)$, when $n \to \infty$.

The aforementioned conditions are satisfied in our context due to the previously stated assumptions about the mixing properties of the time series and our appropriate choice of block length. The strong mixing conditions ensure that observations sufficiently far apart are nearly independent, which is crucial for the block bootstrap to correctly capture the dependence structure.

Asymptotic Normality of the Bootstrap Estimator

Having established the convergence of the bootstrap variance, we now need to prove that the bootstrap estimator satisfies the same asymptotic normality as the original OLS estimator. We begin by rewriting the numerator of the bootstrap estimator in equation 3.5 in terms of block sums:

$$\sqrt{n}\sum_{t=1}^{n}(x_t^* - \bar{x}^*)u_t^* = \sqrt{n}\sum_{i=1}^{k} S_i^* \tag{3.15}$$

where each S_i^* represents the sum of terms $(x_t^* - \bar{x}^*)u_t^*$ in the i-th block.

Since the blocks are independently resampled in the NBB, we can apply the Central Limit Theorem to the sum of independent block statistics:

$$\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\frac{S_i^* - E^*(S_i^*)}{\sqrt{Var^*(S_i^*)}} \xrightarrow{d^*} \mathcal{N}(0,1) \tag{3.16}$$

This shows that the standardized sum of blocks converges in distribution to a standard normal distribution.

We can combine this result with our previous finding on variance convergence. If we have:

$$\frac{\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1)}{\sqrt{\text{Var}^*(\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1))}} \xrightarrow{d^*} \mathcal{N}(0,1) \tag{3.17}$$

and

$$\sqrt{\text{Var}^*(\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1))} \xrightarrow{p} \sqrt{V} \tag{3.18}$$

Then, by the Slutsky Theorem:

$$\frac{\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1)}{\sqrt{\text{Var}^*(\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1))}} \cdot \sqrt{\text{Var}^*(\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1))} \xrightarrow{d^*} \mathcal{N}(0,1) \cdot \sqrt{V} \tag{3.19}$$

$$\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1) \xrightarrow{d^*} \mathcal{N}(0, V) \tag{3.20}$$

This establishes that the bootstrap estimator has the same asymptotic normal distribution as the original estimator.

Convergence in Probability of the CDF

To complete the evaluation of the NBB's consistency for $\hat{\beta}_1$, we can use Skorokhod's Representation Theorem (Skorokhod, 1965; Billingsley, 1999) to establish almost sure convergence. This theorem allows us to work with almost surely convergent random variables that have the same limiting distribution as the weakly convergent sequence, simplifying certain arguments through a stronger mode of convergence.

Let $\tilde{Z}_n^*$ be a redefined bootstrap estimator in a new probability space that preserves the same distribution as $\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1)$. The true limiting normal distribution of the original estimator is:

$$Z \sim \mathcal{N}(0, V),$$

which corresponds to the asymptotic distribution of $\sqrt{n}(\hat{\beta}_1 - \beta_1)$.

The almost sure convergence of the redefined bootstrap sequence,

$$\tilde{Z}_n^* \xrightarrow{a.s.} Z \quad \text{in an extended probability space,}$$

allows us to conclude pointwise convergence of the empirical cumulative distribution functions (CDFs), ensuring that the bootstrap correctly approximates the sampling distribution:

$$\mathbb{P}^*\left(\sqrt{n}(\hat{\beta}_1^* - \hat{\beta}_1) \leq x\right) \xrightarrow{p} \mathbb{P}\left(\sqrt{n}(\hat{\beta}_1 - \beta_1) \leq x\right) \quad \text{as } n \to \infty.$$

In summary, the Non-Overlapping Block Bootstrap is consistent for estimating the hedge ratio β_1 in regression models with dependent data. This consistency comes from three key properties: (1) the bootstrap variance estimator converges to the true variance V; (2) the bootstrap estimator satisfies the same asymptotic normality as the original estimator; and (3) the bootstrap distribution converges in probability to the real asymptotic distribution. These properties collectively ensure the reliability of the NBB method in dependent data settings.

3.2.4 On Bootstrap Quantiles and Coverage Probability

In our pairs trading application, we utilize direct quantiles from the empirical bootstrap distribution to establish trading decision thresholds. This approach differs from traditional statistical inference, where coverage probability of confidence intervals is a primary concern. We briefly address why coverage probability considerations are not critical in our context.

Trading Decisions vs. Statistical Inference

When using bootstrap methods for statistical inference, the coverage probability, i.e. the probability that a confidence interval contains the true parameter, is a fundamental concern. However, in our pairs trading application, our primary objective is not to make inferences about the true value of β_1, but rather to establish decision thresholds that effectively identify trading opportunities.

The quantiles q_α^* and $q_{1-\alpha}^*$ of the bootstrap distribution serve as statistical boundaries for trading decisions, where exceeding these thresholds indicates a potentially profitable deviation from the equilibrium relationship. In this decision-oriented framework, the traditional coverage probability is less relevant for several reasons:

1. Decision Optimality: Our thresholds are chosen to optimize trading performance rather than parameter estimation accuracy. The effectiveness of these thresholds is evaluated by trading metrics (e.g., returns, Sharpe ratio) rather than by their coverage of the true parameter.

2. Direct Probability Control: The quantiles directly control the probability of signal generation under the empirical distribution, providing a natural interpretation in terms of signal frequency rather than parameter coverage.

3. Empirical Performance Focus: In trading applications, the ultimate validation comes from out-of-sample performance rather than theoretical coverage properties.

Theoretical Justification

From a theoretical perspective, our approach can be justified within the framework of statistical decision theory rather than estimation theory. As noted by Efron (1987), bootstrap methods can be used to approximate the sampling distribution for making decisions, even in cases where exact interval coverage is not the primary concern.

When our objective is to establish decision rules for trading rather than to conduct formal hypothesis tests about model parameters, the bootstrap distribution provides a valid basis for defining statistically informed thresholds. The empirical quantiles directly characterize the uncertainty in the estimated relationship and translate this uncertainty into actionable trading boundaries.

In this context, we are less concerned with whether our intervals cover the true parameter at exactly the nominal rate, and more concerned with establishing consistent decision rules that effectively separate signal from noise in the trading context. Additionally, the capture of the time series dependence structure by the non-overlapping block bootstrap improves the robustness of our method.

Practical Implications

This distinction between inference-focused applications (where coverage probability is critical) and decision-focused applications (where performance metrics are paramount) allows us to use the empirical bootstrap distribution directly without implementing coverage-enhancing refinements such as the bias-corrected and accelerated (BCa) methods or the studentized bootstrap. Moreover, the use of the NBB improves the stability of decision thresholds in practical applications given that it accounts for the autocorrelation structure of the financial time series.

By focusing on the decision-theoretic aspects of our application rather than formal inference, we maintain the practical advantages of the bootstrap approach, i.e. robustness to certain forms of model misspecification, model-free estimation of uncertainty, and adaptability to changing market conditions, while ensuring that our method appropriately accounts for the time-dependent nature of financial data.

3.3 Results

This section presents the empirical analysis of the proposed block bootstrap approach. We compared its performance against the traditional cointegration method (referred to as the 'Reference' strategy) in terms of cumulative returns, volatility, maximum drawdown, Sharpe ratio (Sharpe, 1966), and Sortino ratio (Sortino and van der Meer, 1991). Our dataset consists of equity pairs from the U.S. and Brazilian markets, selected based on historical price co-movement and suitability for statistical arbitrage. We obtained intraday price data at a 30-minute frequency from two primary sources: Alpha Vantage Inc. for the U.S. market and the MetaTrader 5 database of XP Investiments for the Brazilian market. The dataset spans a two-year period from February 1, 2023, to January 31, 2025. In cases where data availability was constrained for one or both assets in a pair, we restricted our analysis to the overlapping period to ensure consistency in comparison.

For the trading strategy implementation, we utilized adjusted close prices to identify

entry and exit signals, while simulating trade executions at the midpoint between the open and adjusted close prices of each 30-minute interval. This approach provides a more realistic representation of achievable execution prices compared to alternative methods such as using daily close-to-close prices, as it better accounts for intraday price dynamics and execution slippage that traders would encounter in real-world scenarios.

3.3.1 Backtesting Methodology

To test the proposed method, we implemented an algorithm using a walk-forward framework that continuously adjusts strategy parameters based on new data. To achieve that, we used a sliding window that moves incrementally throughout the price time series. The sliding window comprises three stages: formation, decision, and trading periods. The formation period has length w and is used to estimate the trading parameters.

Assuming the decision period in which we opt to trade or stay out of the market occurs at time t, the formation period spans from $t - w + 1$ to $t - 1$, whereas the trading period is expressed as $t + 1$. If the pair is not cointegrated in the formation period, we slide the rolling window one discrete time unit forward and re-evaluate the cointegration status. Alternatively, if the pair is cointegrated in the formation window, we apply the NBB to estimate the empirical distribution $\hat{F}_n^*$ of the hedge ratio β_1. The procedure begins by dividing the original time series into non-overlapping blocks of length b, potentially retaining a partial block at the end. We then generate B bootstrap samples by randomly selecting $k = \lceil n/b \rceil$ blocks with replacement and concatenating them. For each bootstrap sample, we estimate the hedge ratio β_1^* using ordinary least squares regression. From these B bootstrap estimates, we construct the empirical distribution of β_1^*. The resulting distribution captures the uncertainty in the hedge ratio estimate, accounting for the temporal dependence structure in the financial time series.

Once we have generated the bootstrap distribution of β_1, we use quantiles from this distribution to confirm or reject trading signals from the standard cointegration approach. This creates a two-layer decision process: in the first layer, the standard cointegration method

generates initial signals based on the deviation of the spread ($y_t - \hat{\beta}_1 x_t$) from its historical mean, measured in terms of standard deviations (standardized spread); these signals are confirmed or rejected in the second layer based on whether the current estimate of β_1 falls within the quantiles (q_α and $q_{1-\alpha}$) derived from the bootstrap distribution.

For entry signals, if the current standardized spread crosses the predefined thresholds (e.g., $\pm 2\sigma$) and the current $\hat{\beta}_1$ is within the bootstrap-derived quantiles, the trade is executed. Otherwise, the signal is ignored.

For exit signals, we implement an asymmetric confirmation process. Take-profit signals are executed without additional confirmation when the standardized spread reverts to its equilibrium point 0. Conversely, stop-loss signals are activated when the standardized spread moves beyond a predetermined loss threshold and confirmed only if the current $\hat{\beta}_1$ falls outside the bootstrap-derived quantiles, indicating potential structural changes in the relationship between the assets.

All procedures are formally described in algorithms 2 and 3.

Algorithm 2 Backtesting the NBB approach for distribution-based trading signals

Require: Time series observations $\{(x_t, y_t)\}_{t=1}^{n}$, sliding window length w, block length b, bootstrap replications m, negative $\hat{\beta}_1^*$ tolerance proportion p, lower bound entry threshold L, upper bound entry threshold U, stop-loss margin ξ, quantile parameter α

flag $\leftarrow FALSE$ ▷ Flag indicating the existence of an active operation
for t in $[(w+1):n]$ **do**
  if flag is FALSE **then**
    Run cointegration test on data $\{(x_t, y_t) : t \in [t-w+1:t-1]\}$
    if pair is cointegrated **then**
      Run Non-Overlapping Block Bootstrap on data $\{(x_t, y_t) : t \in [t-w+1:t-1]\}$ and obtain the empirical distribution $\hat{F}_w^*$ for β_1
      if $\mathbb{P}(\hat{F}_w^* < p)$ **then** ▷ Evaluating the proportion of $-\beta_1$'s in the empirical distribution
        Discard the cointegration status for the period ▷ Using the algorithm to perform pair selection
      else
        Compute $\hat{\beta}^*_{1,[t-w+1:t-1]} = \mathbb{E}[\hat{F}_w^*]$
        Compute q_α and $q_{1-\alpha}$
        Compute spread $\delta_{[t-w+1:t-1]} = y_{[t-w+1:t-1]} - \hat{\beta}^*_{1,[t-w+1:t-1]}\ x_{[t-w+1:t-1]}$ on data from $t-w+1$ to $t-1$
        Compute standardized spread at time t:
        $z[t] = \frac{\delta[t] - \bar{\delta}_{[t-w+1:t-1]}}{s_{\delta_{[t-w+1:t-1]}}}$
        Fit $\hat{y}_{[t-w+1:t]} = \hat{\beta}_{0,[t-w+1:t]} + \hat{\beta}_{1,[t-w+1:t]}\ x_{[t-w+1:t]}$ to data $\{(x_t, y_t) : t \in [t-w+1:t]\}$
        Store $\hat{\beta}_{1,[t-w+1:t]}$
        if $(z[t] \geq U$ or $z[t] \leq L)$ and $q_\alpha < \hat{\beta}_{1,[t-w+1:t]} < q_{1-\alpha}$ **then** ▷ Look for entry triggers
          ENTRY Signal on $t+1$
          flag $\leftarrow$ TRUE
        end if
      end if
    end if

Algorithm 2 Backtesting the NBB approach for distribution-based trading signals (continued)

else ▷ Look for exit triggers

Fit $\hat{y}_{[t-w+1:t+1+k]} = \hat{\beta}_{0,[t-w+1:t+1+k]} + \hat{\beta}_{1,[t-w+1:t+1+k]}\ x_{[t-w+1:t+1+k]}$ to data $\{(x_t, y_t) : t \in [(t-w+1) : (t+1+k)]\}$, where $t+1+k$ represents current time

Store new $\hat{\beta}_{1,[t-w+1:t+1+k]}$

Compute spread $\delta_{[t-w+1:t+1+k]} = y_{[t-w+1:t+1+k]} - \hat{\beta}_{1,[t-w+1:t+1+k]}\ x_{[t-w+1:t+1+k]}$ on data $\{(x_t, y_t) : t \in [(t-w+1) : (t+1+k)]\}$

Compute standardized spread at current time $t+1+k$:

$z[t+1+k] = \frac{\delta[t+1+k] - \bar{\delta}_{[t-w+1:t+1+k]}}{s_{\delta_{[t-w+1:t+1+k]}}}$

if $(z[t+1+k] \geq U+\xi$ or $z[t+1+k] \leq L-\xi)$ and $(\hat{\beta}_{1,[t-w+1:t+1+k]} \leq q_\alpha$ or $\hat{\beta}_{1,[t-w+1:t+1+k]} \geq q_{1-\alpha})$ **then**

EXIT Signal ▷ Stop-loss signal

flag $\leftarrow$ FALSE

else if $(z[t+1+k] < 0$ or $z[t+1+k] > 0)$ **then**

EXIT Signal ▷ Take-profit signal

flag $\leftarrow$ FALSE

end if

end if

end for

Algorithm 3 Non-Overlapping Block Bootstrap for hedge ratio (β_1) distribution estimation

Require: Time series observations $\{(x_t, y_t)\}_{t=1}^{n}$, block length b, bootstrap replications m, negative β_1 tolerance proportion p

$\mathcal{B} \leftarrow \emptyset$ ▷ Set of bootstrap coefficients

$K \leftarrow \lceil T/b \rceil$ ▷ Number of blocks

for $i = 1$ to m **do**

  Select K blocks with replacement from the set $\{1, 2, \ldots, K\}$, denoted as $\{k_1, k_2, \ldots, k_K\}$

  For each block k_j, let $\mathcal{I}_{k_j} = \{(k_j - 1)b + 1, \ldots, \min(k_j b, T)\}$

  Let $\mathcal{I}_i = \bigcup_{j=1}^{K} \mathcal{I}_{k_j}$ ▷ Indices of observations in the i-th bootstrap sample

  Compute $\hat{\beta}_1^{*(i)}$ from the regression $\hat{y}_t^* = \hat{\beta}_0^* + \hat{\beta}_1^* x_t^*$ using observations $\{(X_t, Y_t) : t \in \mathcal{I}_i\}$

  $\mathcal{B} \leftarrow \mathcal{B} \cup \{\hat{\beta}_1^{*(i)}\}$

end for

$p^* \leftarrow \frac{|\{\hat{\beta}_1^{*(i)} \in \mathcal{B} : \hat{\beta}_1^{*(i)} < 0\}|}{|\mathcal{B}|}$ ▷ Proportion of negative coefficients

if $p^* > p$ **then**

  return $\emptyset$ ▷ Discard distribution if too many negative values

else

  return $\mathcal{B}$ ▷ Empirical bootstrap distribution of $\hat{\beta}_1$

end if

In the following section, we show that the proposed algorithm effectively filters out signals that might arise from temporary market noise while remaining responsive to genuine trading opportunities and structural changes in the asset relationship. By utilizing the bootstrap distribution, we introduce a probabilistic dimension to the trading decision process, making it more adaptive to market conditions and more robust against false signals compared to the standard cointegration strategy.

3.3.2 Parameter specification and sensitivity analysis

Our model requires the specification of several parameters that can be categorized into two groups. The first group includes parameters common to both our Bayesian approach and the standard cointegration strategy, specifically the formation window length (w), the standardized spread levels L and U, and the stop-loss margin ϵ. Note that the take-profit

margin is defined at the equilibrium point 0. The second group consists of parameters unique to our Bayesian framework, notably the quantile parameter (α), which is used to obtain the confirmation thresholds q_α and $q_{1-\alpha}$ under the full conditional distribution of β_1.

Formation window (w)

For intraday data in the 30-minute timeframe, our analysis indicates that the appropriate length w for the formation period ranges between 65 and 130 observations, representing approximately two and four trading weeks of data, respectively. These values were selected for testing as they provide sufficient data points for robust statistical analysis while remaining responsive to market conditions.

The sensitivity analysis revealed no consistent advantage between using $w = 65$ or $w = 130$ across all trading pairs. Some pairs performed better with the shorter window ($w = 65$), while others benefited from the extended formation period ($w = 130$). This variation likely reflects differences in individual pairs' underlying price dynamics and cointegration characteristics.

The absence of a clear optimal window length aligns with theoretical expectations. Excessively small windows can increase the chances of an incorrect rejection of the cointegration hypothesis (Type II errors) due to low statistical power, higher estimator variance, and insufficient capture of long-run relationships (Cheung and Lai, 1993; Otero and Smith, 2000). In theory, larger samples should improve cointegration detection by increasing statistical power; however, in practice, too large time series tend to result in a rejection of cointegration status due to the inclusion of structural breaks or regime changes that disrupt long-term equilibrium relationships (Gregory and Hansen, 1996; Hakkio and Rush, 1991). This dual challenge in window selection has been well documented in the econometric literature (Engle and Granger, 1987; Johansen, 1988, 1991), highlighting the importance of finding an appropriate balance in window sizing: large enough to ensure statistical robustness but constrained enough to avoid capturing multiple structural changes that could mask existing cointegrating relationships.

Standardized spread thresholds (L and U)

Our analysis reveals that thresholds of ± 1 standard deviation yield optimal results for he proposed strategies. This configuration strikes an effective balance between identifying trading opportunities and filtering out market noise. Notably, the method maintains robust performance and consistently outperforms the reference cointegration approach across all analyzed pairs, even when using the more conservative ± 2 standard deviation thresholds.

This resilience to threshold adjustments demonstrates the stability of the proposed method across varying entry levels. The consistent performance across different threshold settings suggests that the Bayesian framework effectively captures the underlying dynamics of pair relationships regardless of the specific entry criteria employed, highlighting its adaptive capability in diverse market conditions for the analyzed pairs.

Stop-loss margin (ξ)

The stop-loss margin ξ in the cointegration method defines the level of deviation in the standardized spread that, when exceeded, triggers the automatic closure of the position to limit potential losses. Typically, this parameter requires different specifications depending on the chosen levels for the standardized thresholds and the take-profit margin, allowing investors to implement proper risk management.

In our study, we tested ϵ in two configurations: half of the take-profit margin or equal to it. This approach was applied to both algorithms. Although the best results were obtained using a 1/1 ratio for the take-profit/stop-loss margin, the proposed Bayesian method performed well across all evaluated scenarios among the analyzed pairs. From a practical standpoint, it is well established that adopting smaller entry thresholds, such as ± 1, increases the number of operations but may also introduce considerable noise. We hypothesize that the Bayesian method's superior performance in high-risk scenarios stems from its enhanced capability to discriminate between noise and genuine trading signals. This characteristic, combined with an increased operational frequency, enables the algorithm to capitalize on more opportunities while maintaining precision, ultimately generating amplified returns in contexts where the

traditional method is less effective.

Quantile parameter (α)

The quantile parameter α exerts significant control over the responsiveness of the proposed model. This parameter effectively regulates the occurrence of entry and stop-loss signals in the following manner:

A smaller α value (e.g., 0.05) expands the area in which β_1 is assumed to be well-behaved, and make the Bayesian method less selective for entry signals and more selective for stop-loss signals. A larger α value (e.g., 0.15) creates a narrower middle band, which filters more potential entries compared to the reference method, but makes the algorithm less selective and thus more responsive to exit signals from the standard cointegration layer (Figure 3.1).

Block length (b)

The block length parameter b plays a crucial role in the Non-Overlapping Block Bootstrap implementation, as it directly impacts the preservation of the temporal dependence structure within financial time series data. Our empirical analysis indicates that the optimal block length varies based on the characteristic mean-reversion horizon of each specific pair.

For the pairs analyzed in our study, we tested block lengths of $\{13, 26, 39\}$ for $w = 65$ and $\{13, 26, 29, 65\}$ for $w = 130$, representing approximately 1 to 5 days of market activity in our 30-minute timeframe. While Lahiri (2003) establishes that the consistency of block bootstrap methods generally requires the block length to grow with the sample size, but at a slower rate (i.e., $b^{-1} + n^{-1}b = o(1)$ as $n \to \infty$), our approach prioritized blocks with practical financial meaning over strictly theoretical calculations. This balanced approach allows us to benefit from statistical theory while accounting for the unique characteristics of our application context.

Specifically, we selected block sizes that correspond to meaningful market periods (e.g., trading days) rather than adhering rigidly to the theoretical formula. This choice was motivated by the observation that financial time series often exhibit cyclical patterns aligned

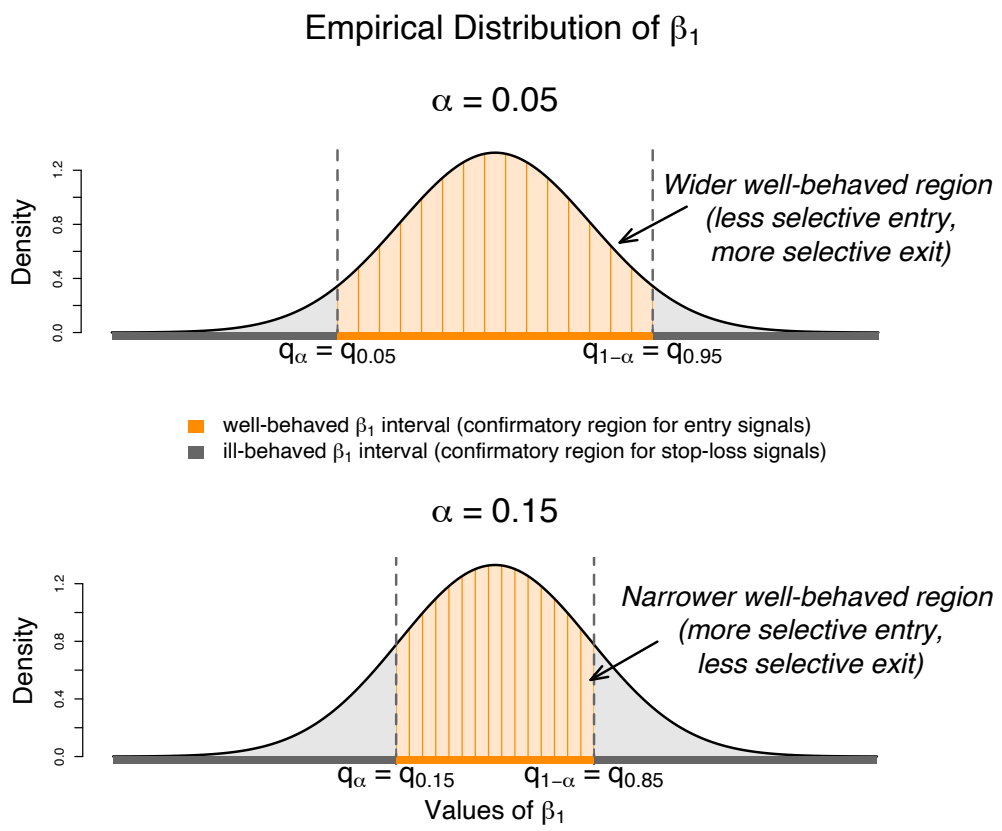

Figure 3.1: Quantile parameter α and 'well-behaved' intervals for β_1 under the empirical distribution $\hat{F}_w^*$. The upper and lower panels illustrate how different values of α affect the confirmatory regions for entry and exit signals. For $\alpha = 0.05$ (upper panel), the well-behaved region is wider, resulting in less selectivity for entry signals and more selectivity for exit signals. For $\alpha = 0.15$ (lower panel), the well-behaved region is narrower, resulting in more selectivity for entry signals and less selectivity for exit signals. The blue regions represent intervals where β_1 is considered well-behaved, while the gray regions indicate intervals where β_1 falls outside the quantiles defined in the full conditional distribution.

with natural trading intervals, and practitioners typically find greater intuitive value in parameters that correspond to recognizable market periods. Although this approach deviates somewhat from the pure statistical framework, it maintains theoretical consistency while enhancing interpretability in the trading context.

The sensitivity analysis revealed that moderate block lengths, i.e. $\{13, 26, 39\}$, generally provided the best balance between preserving the essential autocorrelation structure and ensuring sufficient variability in bootstrap resamples for the tested formation periods $\{65, 130\}$. Notably, the optimal block length exhibited some correlation with the forma-

tion window length (w). Pairs that performed optimally with shorter formation windows ($w = 65$) typically benefited from slightly shorter block lengths, while pairs optimized with extended formation periods ($w = 130$) showed better performance with moderately larger block sizes. This relationship aligns with theoretical expectations, as larger windows capture longer-term relationships that may require larger blocks to preserve their dependency structure.

From a practical standpoint, the block length directly impacts the balance between preserving local dependencies (within blocks) and establishing independence between blocks. This parameter proves particularly important in financial time series that exhibit complex dependency structures with both short-term momentum and mean-reversion characteristics occurring at different time scales. Our empirical results suggest that the proposed algorithm maintains robust performance across reasonable variations in block length, indicating stability with respect to this parameter.

3.3.3 General remarks

The empirical findings presented in Table 3.1 demonstrate that the NBB strategy consistently outperforms the traditional cointegration approach across the pairs analyzed. The NBB method displays superior risk-adjusted performance characterized by significantly reduced drawdowns while generating enhanced returns over time. This pattern of outperformance is evident across diverse market conditions, asset classes, and geographic regions, suggesting the method's robustness and adaptability.

One notable characteristic of the NBB approach is its effectiveness in both high and low volatility environments. In highly volatile sectors such as airlines (AAL-DAL, UAL-DAL) and commodities (REMX-LIT), the method demonstrated extraordinary capabilities in reducing downside risk while capturing upside potential. Simultaneously, in traditionally low-volatility segments like treasury bonds (SHY-IEF, IEF-TLT) and S&P 500 tracking ETFs (SPY-IVV, VOO-IVV), the NBB strategy maintained its edge, extracting meaningful alpha from even subtle pricing dislocations. This dual capability suggests the method's

Table 3.1: Comparison between the Non-Overlapping Block Bootstrap ("NBB") and the standard cointegration methods ("Ref.") for cumulative returns as a multiple of the initial investment and risk-return metrics from 02/01/2023 to 01/31/2025 on selected pairs in the U.S. and Brazilian markets. The lower and upper standardized spread thresholds (L and U) used were ± 1; the stop-loss margin adopted was $\xi = 1$; and the take-profit standardized threshold spread was set as 0.

Parameters			Cum. Ret.		Count		Vol.		Max. Draw.		Sharpe		Sortino	
Quantile-Window	Block Length	Pair	Ref.	NBB	Ref.	NBB	Ref.	NBB	Ref.	NBB	Ref.	NBB	Ref.	NBB
US														
$\alpha = .05$, $w = 65$	$b = 13$	KO-PEP	0.877	1.371	244	222	0.085	0.033	0.153	0.027	-0.102	0.902	-0.145	1.346
	$b = 26$	GLD-IAU	0.973	1.196	720	1002	0.042	0.017	0.094	0.014	-0.039	0.987	-0.054	1.593
		AAL-DAL	1.369	2.288	186	215	0.196	0.14	0.474	0.168	0.248	0.617	0.36	1.167
$\alpha = .05$, $w = 130$	$b = 26$	IWM-VTWO	1.17	2.006	303	592	0.08	0.05	0.113	0.051	0.223	1.333	0.316	1.994
		MSFT-AAPL	0.833	1.281	57	60	0.09	0.025	0.291	0.032	-0.146	0.957	-0.211	1.502
	$b = 39$	SPY-IVV	0.991	1.119	578	856	0.062	0.01	0.147	0.008	0.017	1.091	0.024	1.745
		IWN-MOO	0.564	1.447	93	105	0.16	0.101	0.39	0.079	-0.253	0.392	-0.355	0.589
		QQQ-QQQM	1.066	1.409	159	602	0.12	0.09	0.287	0.19	0.111	0.402	0.157	0.573
$\alpha = .15$, $w = 65$	$b = 13$	SHY-IEF	0.942	1.066	74	247	0.014	0.009	0.07	0.008	-0.4	0.679	-0.56	1.029
		IEF-TLT	0.912	1.198	117	312	0.053	0.024	0.11	0.02	-0.136	0.72	-0.191	1.05
		UAL-DAL	0.632	2.043	66	276	0.207	0.057	0.41	0.032	-0.101	1.193	-0.148	1.81
		HD-LOW	0.71	1.66	132	299	0.137	0.084	0.288	0.11	-0.164	0.608	-0.23	0.865
		LEN-DHI	0.812	1.947	123	344	0.25	0.135	0.343	0.23	0.048	0.529	0.068	0.765
		REMX-LIT	0.572	2.961	178	268	0.22	0.14	0.45	0.114	-0.127	0.796	-0.179	1.188
	$b = 26$	VOO-IVV	0.993	1.205	496	778	0.062	0.032	0.139	0.019	0.021	0.558	0.029	1.961
		VUG-IWF	0.611	1.505	107	215	0.316	0.062	0.356	0.076	0.017	0.647	0.024	1.028
		COP-EOG	1.225	2.242	241	305	0.134	0.081	0.114	0.052	0.208	0.976	0.301	1.471
$\alpha = .15$, $w = 130$	$b = 39$	SHY-TLT	0.948	1.027	27	81	0.013	0.006	0.062	0.007	-0.379	0.425	-0.526	0.675
		VNQ-IYR	0.803	2.623	554	604	0.092	0.053	0.322	0.028	-0.176	1.723	-0.232	2.837
	$b = 65$	V-MA	0.873	1.61	208	197	0.105	0.06	0.17	0.048	-0.068	0.772	-0.098	1.143
Brazil														
$\alpha = .05$, $w = 65$	$b = 13$	PETR4-PETR3	0.843	1.181	42	30	0.025	0.023	0.206	0.037	-0.626	0.678	-0.82	1.055
		ELET3-ELET6	0.836	1.258	32	65	0.046	0.034	0.158	0.055	-0.34	0.644	-0.471	0.95
	$b = 26$	SPXI11-IVVB11	0.988	1.265	478	335	0.025	0.018	0.077	0.013	-0.032	1.195	-0.045	1.927
		BOVA11-SMAL11	0.856	1.056	32	15	0.032	0.008	0.185	0.013	-0.444	0.681	-0.614	1.063
$\alpha = .05$, $w = 130$	$b = 26$	GGBR4-GOAU4	1	1.216	98	37	0.049	0.037	0.151	0.045	0.024	0.516	0.033	0.768
		BBDC3-BBDC4	1.106	1.16	42	29	0.031	0.031	0.039	0.037	0.318	0.461	0.455	0.662
	$b = 39$	ITSA4-ITUB4	0.93	1.164	30	27	0.04	0.04	0.101	0.083	-0.151	0.373	-0.213	0.54

fundamental strength lies in its ability to more accurately identify genuine mean-reversion opportunities while filtering out market noise regardless of the overall volatility regime.

Visual inspection of cumulative returns (Figures 3.2 and 3.3) reveals that the proposed algorithm effectively refines entry and exit points. This confirms our hypothesis that the method produces more timely trading signals, optimizing trade execution and yielding more stable and reliable returns. Furthermore, significant differences in risk metrics between the two methods highlight the clear advantages of the proposed algorithm, with consistent

reductions in volatility and maximum drawdown across nearly all tested pairs.

The NBB method achieved this improved performance with varying trade frequency, showing adaptability across different market conditions. Notably, in some pairs like GLD-IAU, the NBB method executed a higher number of trades (1002 versus 720), while maintaining better risk management with volatility reduced by 59

Lastly, the results also demonstrate that the proposed NBB method can serve as a foundational algorithm in more sophisticated strategies. In practice, it is common to see cointegration strategies combined with other techniques to enhance pair selection or trading signal generation. In this regard, since our method has proven to deliver better performance and more effective risk management, it can significantly enhance the overall strategy by replacing traditional cointegration in a compound pairs trading strategy.

In the following section, we will present simulations demonstrating how the NBB method reduces the occurrence of false positives in cointegration tests. This indicates that part of our method has broader applications — not only for pairs trading but also for improving the power of cointegration tests in some scenarios.

3.3.4 US Performance

Equity Pairs Analysis

In the equities segment, the comparison between Lennar Corporation (LEN) and D.R. Horton, Inc. (DHI) highlights the NBB strategy's effectiveness in cyclical sectors. While the reference strategy deteriorated to approximately 20% below initial value, the NBB approach achieved a cumulative return of 1.947 versus 0.812, with notably reduced volatility (0.135 versus 0.25) and a maximum drawdown of 0.23 versus 0.343. This significant divergence demonstrates the NBB method's ability to effectively capture mean-reverting opportunities in cyclical industries.

The aviation sector pair of American Airlines Group Inc. (AAL) versus Delta Air Lines, Inc. (DAL) exhibited one of the most substantial outperformances, with the NBB strategy generating a cumulative return of 2.288 compared to 1.369 for the reference strategy. This

remarkable performance was accompanied by reduced volatility (0.14 versus 0.196) and significantly improved maximum drawdown (0.168 versus 0.474). The superior performance in this cyclical and often volatile sector underscores the NBB method's robustness in industries characterized by complex market dynamics.

Similarly, the airline carrier pair United Airlines Holdings, Inc. (UAL) versus Delta Air Lines, Inc. (DAL) further validates this pattern, with the NBB approach achieving an exceptional cumulative return of 2.043 versus a disappointing 0.632 for the reference strategy. This substantial outperformance was complemented by dramatic improvements in risk metrics, with volatility reduced by over 72% (0.057 versus 0.207) and maximum drawdown by over 92% (0.032 versus 0.41). The NBB strategy's extraordinary performance in this pair demonstrates its effectiveness in extracting value from related securities within volatile industries.

In the retail home improvement sector, The Home Depot, Inc. (HD) versus Lowe's Companies, Inc. (LOW) pair showed consistent NBB strategy outperformance. While the reference strategy declined to approximately 29% below initial value, the NBB approach generated positive returns of 66% (1.66 versus 0.71). This was accompanied by improved risk metrics, with volatility reduced by nearly 39% (0.084 versus 0.137) and maximum drawdown by approximately 62% (0.11 versus 0.288). This performance differential highlights the NBB method's ability to identify profitable trading opportunities even during periods when the traditional approach struggles to maintain value.

In the financial sector, the Visa Inc. (V) versus Mastercard Incorporated (MA) pair further validates the NBB approach for highly correlated financial equities. The NBB strategy generated steady positive returns, reaching approximately 61% cumulative gain (1.61 versus 0.873), with reduced volatility (0.06 versus 0.105), reinforcing its ability to stabilize returns while maintaining positive performance. This pair exemplifies how the NBB filtering effectively captures short-term pricing inefficiencies in highly correlated stocks.

Similarly, in consumer staples, the Coca-Cola Company (KO) versus PepsiCo, Inc. (PEP) pair demonstrated consistent positive performance under the NBB approach, while the reference strategy exhibited steady decline. The NBB method reduced volatility by 61% (0.033

versus 0.085) and maximum drawdown by 82% (0.027 versus 0.153), confirming its effectiveness in defensive equity pairs.

Index and Sector Pairs Analysis

In the technology sector, the Invesco QQQ Trust (QQQ) versus Invesco NASDAQ 100 ETF (QQQM) pair showed the NBB method's ability to extract meaningful alpha despite the high correlation between these instruments. The NBB strategy achieved 40.9% cumulative returns (1.409 versus 1.066), with volatility reduced by 25% (0.09 versus 0.12) and drawdown reduced by 34% (0.19 versus 0.287). This highlights the NBB model's superior entry and exit timing capabilities for related instruments.

The analysis of bond market ETFs revealed significant advantages for the NBB approach across various duration profiles. The iShares 1-3 Year Treasury Bond ETF (SHY) versus iShares 7-10 Year Treasury Bond ETF (IEF) pair demonstrated the method's effectiveness in fixed income markets, with the NBB strategy generating a cumulative return of 1.066 versus 0.942 for the reference strategy, while reducing volatility by 36% (0.009 versus 0.014) and maximum drawdown by nearly 89% (0.008 versus 0.07). This dramatically improved the risk-adjusted performance metrics, with Sharpe ratio increasing from -0.4 to 0.679.

A similar pattern emerged in the medium-to-long duration fixed income pair iShares 7-10 Year Treasury Bond ETF (IEF) versus iShares 20+ Year Treasury Bond ETF (TLT), where the NBB strategy achieved a return of 1.198 versus 0.912, with volatility reduced by 55% (0.024 versus 0.053) and maximum drawdown by 82% (0.02 versus 0.11). These results in the fixed income segment demonstrate the NBB method's versatility across different asset classes beyond equities.

For the short-to-long duration treasury pair iShares 1-3 Year Treasury Bond ETF (SHY) versus iShares 20+ Year Treasury Bond ETF (TLT), the NBB approach again demonstrated superior performance with returns of 1.027 versus 0.948 for the reference strategy, while cutting volatility by 54% (0.006 versus 0.013) and maximum drawdown by 89% (0.007 versus 0.062). This consistent pattern across different segments of the yield curve reinforces the

robustness of the NBB methodology in fixed income markets.

The U.S. small-cap analysis examined the iShares Russell 2000 ETF (IWM) versus Vanguard Russell 2000 ETF (VTWO) pair, which exhibited strong performance under the NBB strategy with cumulative returns of 2.006 compared to 1.17 for the reference method. This was accompanied by substantial improvements in risk metrics, with volatility reduced by 38% (0.05 versus 0.08) and maximum drawdown by 55% (0.051 versus 0.113). The Sharpe ratio showed dramatic enhancement from 0.223 to 1.333, and the Sortino ratio from 0.316 to 1.994, demonstrating the NBB method's effectiveness in the small-cap segment.

An interesting comparison emerged in the natural resources sector through the iShares Russell 2000 Value ETF (IWN) versus VanEck Agribusiness ETF (MOO) pair, where the NBB strategy reversed the substantial decline seen in the reference approach (1.447 versus 0.564). This dramatic outperformance was achieved while reducing volatility by 37% (0.101 versus 0.16) and maximum drawdown by 80% (0.079 versus 0.39), showcasing the NBB method's particular efficacy in commodity-related sectors.

In the S&P 500 index tracking space, several ETF pairs demonstrated the NBB method's ability to extract value even from highly efficient instruments. The SPDR S&P 500 ETF Trust (SPY) versus iShares Core S&P 500 ETF (IVV) pair showed modest but consistent outperformance (1.119 versus 0.991) with dramatically reduced volatility (0.01 versus 0.062) and maximum drawdown (0.008 versus 0.147). Similarly, the Vanguard S&P 500 ETF (VOO) versus iShares Core S&P 500 ETF (IVV) pair demonstrated the NBB approach's effectiveness with returns of 1.205 versus 0.993, volatility reduction of 48% (0.032 versus 0.062), and maximum drawdown improvement of 86% (0.019 versus 0.139).

The analysis of growth ETFs through the Vanguard Growth ETF (VUG) versus iShares Russell 1000 Growth ETF (IWF) pair revealed substantial outperformance under the NBB strategy (1.505 versus 0.611), with volatility reduction of 80% (0.062 versus 0.316) and maximum drawdown improvement of 79% (0.076 versus 0.356). This dramatic performance differential suggests the NBB method's particular effectiveness in capturing pricing inefficiencies in growth-oriented instruments.

The energy sector was represented by ConocoPhillips (COP) versus EOG Resources, Inc.

(EOG), where the NBB strategy generated exceptional returns of 2.242 compared to 1.225 for the reference approach. The substantial outperformance was accompanied by significant risk reduction, with volatility decreased by 40% (0.081 versus 0.134) and maximum drawdown by 54% (0.052 versus 0.114). This resulted in dramatically improved risk-adjusted metrics, with Sharpe ratio increasing from 0.208 to 0.976 and Sortino ratio from 0.301 to 1.471.

The real estate and commodity sectors exhibited some of the most dramatic performance differentials. The Vanguard Real Estate ETF (VNQ) versus iShares US Real Estate ETF (IYR) pair generated returns exceeding 162% under the NBB strategy (2.623 versus 0.803), with significantly improved risk metrics. The maximum drawdown was reduced by 91% (0.028 versus 0.322) and volatility by 42% (0.053 versus 0.092), resulting in a remarkable improvement in Sortino ratio from -0.232 to 2.837.

Similarly, the VanEck Rare Earth/Strategic Metals ETF (REMX) versus Global X Lithium & Battery Tech ETF (LIT) pair showed exceptional outperformance, with returns exceeding 296% under the NBB strategy (2.961 versus 0.572), while reducing volatility by 36% (0.14 versus 0.22) and maximum drawdown by 75% (0.114 versus 0.45). This substantial performance differential suggests the NBB method's particular efficacy in capturing pricing dislocations in commodity-related instruments.

3.3.5 Brazil Performance

Equity Pairs Analysis

The NBB strategy demonstrated remarkable performance in Brazil's equity market, affirming the superiority of the proposed method across diverse sectors. For PETR4-PETR3 (Petróleo Brasileiro S.A. preferred shares versus Petróleo Brasileiro S.A. common shares) in the oil and gas sector, the NBB strategy effectively capitalized on dislocations between Petrobras' share classes, achieving a cumulative return of 1.181 versus 0.843 for the reference strategy while reducing maximum drawdown by 82% (0.037 versus 0.206).

In the utilities sector, the Eletrobras - Centrais Elétricas Brasileiras S.A. ordinary shares (ELET3) versus preferred shares (ELET6) pair confirms the NBB method's superiority, dis-

playing consistent upward momentum while the reference strategy deteriorates steadily. This pair exhibited dramatic improvements in risk-adjusted metrics, with Sortino ratio increasing from -0.471 to 0.95 and volatility reduced from 0.046 to 0.034.

The steel industry pair of Gerdau S.A. (GGBR4) versus Metalúrgica Gerdau S.A. (GOAU4) further illustrates this pattern, with the NBB strategy generating substantial positive returns of 21.6% (1.216 versus 1.0) while reducing volatility by 25% (0.037 versus 0.049) and maximum drawdown by 70% (0.045 versus 0.151). This contrasts with the reference approach, which exhibits considerable volatility without meaningful cumulative gains.

In the financial sector, the Banco Bradesco S.A. ordinary shares (BBDC3) versus Banco Bradesco S.A. preferred shares (BBDC4) pair demonstrated the NBB strategy's effectiveness even in highly liquid banking stocks. The NBB approach achieved a modest but consistent outperformance with returns of 1.16 versus 1.106 for the reference strategy, while maintaining the same level of volatility (0.031) but with slightly lower maximum drawdown (0.037 versus 0.039). The improved risk-adjusted metrics were evident in the Sharpe ratio increase from 0.318 to 0.461 and Sortino ratio from 0.455 to 0.662.

The financial holding company pair Itaúsa S.A. (ITSA4) versus Itaú Unibanco Holding S.A. (ITUB4) further validates the NBB method's efficacy in Brazil's banking sector. While the reference approach suffered a decline of 7% (0.93 final value), the NBB strategy generated returns of 16.4% (1.164 versus 0.93), with no increase in volatility (both at 0.04) but with an 18% reduction in maximum drawdown (0.083 versus 0.101). The risk-adjusted performance metrics showed substantial improvement, with Sharpe ratio increasing from -0.151 to 0.373 and Sortino ratio from -0.213 to 0.54.

Index and Sector Pairs Analysis

Particularly notable is the performance differential in the Brazilian S&P 500 ETF pair of SPDR S&P 500 Brazil (SPXI11) versus iShares S&P 500 Brazil (IVVB11). The NBB strategy generated consistent returns throughout the observation period, culminating in gains exceeding 26.5% (1.265 versus 0.988), with extraordinary improvement in risk metrics. Max-

imum drawdown was reduced by 83% (0.013 versus 0.077) and Sortino ratio increased dramatically from -0.045 to 1.927, demonstrating the methodology's effectiveness even when applied to instruments tracking the same underlying index in an emerging market context.

The Brazilian small-cap versus large-cap index ETF pair of iShares Bovespa Small Cap Fundo de Índice (SMAL11) versus iShares Bovespa Index Fund (BOVA11) showed the NBB strategy's ability to generate positive returns in a challenging market environment. While the reference approach declined by 14.4% (0.856 final value), the NBB method achieved a modest positive return of 5.6% (1.056 versus 0.856). This was accompanied by dramatic improvements in risk metrics, with volatility reduced by 75% (0.008 versus 0.032) and maximum drawdown by 93% (0.013 versus 0.185). The risk-adjusted performance metrics showed substantial enhancement, with Sharpe ratio improving from -0.444 to 0.681 and Sortino ratio from -0.614 to 1.063.

These findings suggest that emerging markets like Brazil may offer particularly fertile ground for the NBB pairs trading methodology, potentially due to reduced algorithmic trading activity, wider bid-ask spreads, and more frequent pricing dislocations between related securities. The strategy's consistent outperformance across diverse Brazilian sectors, with substantially lower volatility, underscores its robustness in varied market contexts beyond developed markets.

3.3.6 Comparison with Bayesian Method

When comparing the NBB approach with our Bayesian methodology described previously, both methods demonstrate significant improvements over the standard cointegration strategy, but with notable differences in performance characteristics and operational patterns.

The NBB method generally shows more consistent performance across a wider range of pairs, with fewer parameter sensitivities than observed in the Bayesian approach. For instance, the NBB method maintained robust performance across different window lengths (w = 65 or w = 130) and spread threshold levels, while the Bayesian method showed more variation in optimal parameters across different pairs.

In terms of trading frequency, the NBB method typically generates a number of trades closer to the reference cointegration strategy compared to the Bayesian approach. For example, in the GLD-IAU pair, the NBB method executed 1002 trades compared to 720 for the reference strategy, while the Bayesian method executed 1204 trades. This more moderate increase in trading activity could translate to lower transaction costs while still achieving superior performance, making the NBB method potentially more cost-efficient in real-world implementations.

Risk management appears particularly strong in the NBB method, which achieved some of the most dramatic reductions in volatility and maximum drawdowns. For the SPXI11-IVVB11 pair, the NBB approach reduced maximum drawdown by 83% (0.013 versus 0.077) compared to the Bayesian method's reduction of 79% (0.016 versus 0.077).

While both methods delivered impressive cumulative returns, the performance profiles differ slightly. The NBB method achieved the highest absolute returns for several pairs (such as REMX-LIT at 2.961 compared to the Bayesian method's 2.625), while also demonstrating more consistent improvements across the entire portfolio of tested pairs, with fewer instances of underperformance.

The empirical evidence suggests that the NBB method offers a more balanced combination of enhanced returns and risk reduction, with operational characteristics closer to the traditional cointegration approach, potentially making it more accessible for practitioners to implement while still capturing significant performance improvements.

In the fixed income sector, both methods showed substantial improvements over the reference strategy, but the NBB approach demonstrated particular strength in the SHY-IEF, IEF-TLT, and SHY-TLT pairs. The NBB method's efficacy across the yield curve spectrum suggests its potential utility for duration-based strategies in the fixed income space, an area where even modest improvements in risk-adjusted returns can translate to significant absolute performance given the scale of fixed income portfolios in institutional settings.

For equity sector pairs, particularly in cyclical industries like airlines (AAL-DAL, UAL-DAL) and homebuilders (LEN-DHI), the NBB method showed exceptional capability in managing downside risk while capitalizing on mean-reversion opportunities. This charac-

teristic makes it especially valuable for pairs trading in volatile sectors where the reference strategy's signals may lack sufficient precision to avoid adverse price movements.

The empirical evidence across both U.S. and Brazilian markets indicates that the NBB method provides a robust alternative to both traditional cointegration and the more complex Bayesian approach, offering a favorable compromise between implementation complexity and performance enhancement. Its ability to maintain performance consistency across different market environments and asset classes suggests particular utility as a foundational algorithm for systematic pairs trading strategies in diverse portfolio contexts.

In summary, despite their methodological differences, both NBB and Bayesian methods are innovative and inaugurate a family of strategies based on the hedge ratio distribution - the distribution-based methods. These approaches represent a significant evolution in pairs trading techniques, transcending the limitations of traditional methods by incorporating richer statistical information in the analysis. The ability of both methods to adapt to different asset classes and market conditions suggests broad possibilities for future research, such as modeling the distribution of β_1 with different models or development of hybrid frameworks that combine the advantages of each approach. The emergence of these distribution-based methods marks an important advancement in the evolution of quantitative statistical arbitrage strategies, offering portfolio managers more refined tools to navigate increasingly complex and interconnected markets.

3.4 Concluding Remarks

We introduced a Non-Overlapping Block Bootstrap approach to pairs trading that improves upon traditional cointegration methods. By deriving the empirical distribution of the hedge ratio (β_1) through resampling techniques that preserve temporal dependencies, our method enhances trading signal precision.

Our empirical analysis across U.S. and Brazilian markets demonstrates the NBB approach's consistent superiority over traditional cointegration. The method generated higher cumulative returns while significantly reducing volatility and maximum drawdown across

diverse asset classes and market conditions. Risk-adjusted metrics showed remarkable improvement, with Sortino ratios increasing from negative to strongly positive values in most pairs and maximum drawdowns reduced by up to 93%.

Compared to our previous Bayesian approach, the NBB method demonstrates more consistent performance with fewer parameter sensitivities and operational characteristics closer to traditional methods, balancing enhanced performance with implementation accessibility.

Together, NBB and Bayesian methods inaugurate a family of distribution-based strategies representing a significant advancement in pairs trading techniques. Future research could explore additional markets and asset classes, optimize parameter configurations across market regimes, and develop hybrid frameworks combining the advantages of both approaches to provide portfolio managers with increasingly refined tools for complex markets.

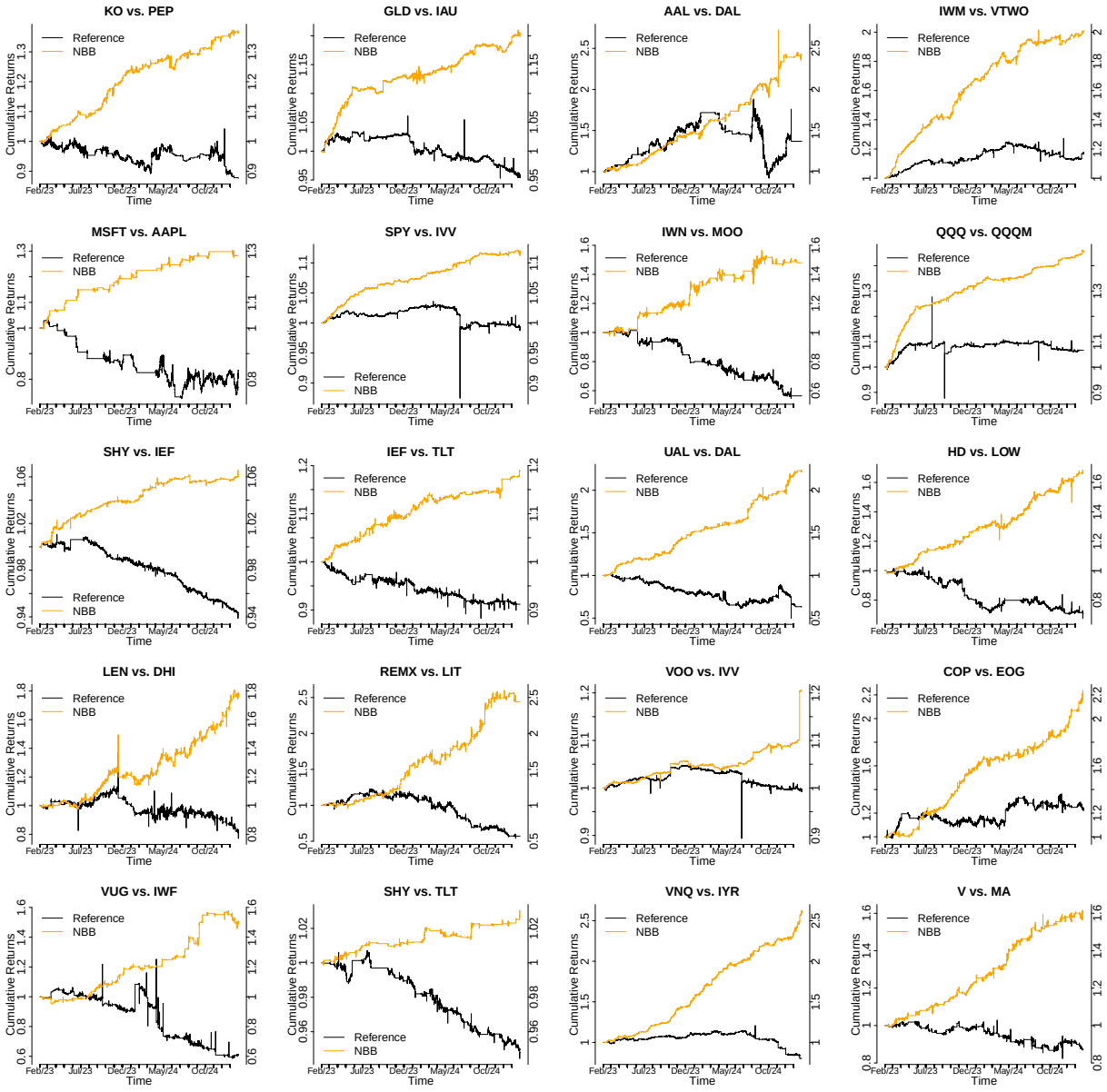

Figure 3.2: Cumulative returns (y-axis) as a multiple of the initial investment for selected pairs in the US market in the period comprised from 02/01/2023 to 01/31/2025 (x-axis).The orange curve indicates the cumulative returns of the proposed Non-Overlapping Block Bootstrap method whereas the black curve shows the performance of the standard cointegration strategy. The parameters used in each pair are described in Table 3.1.

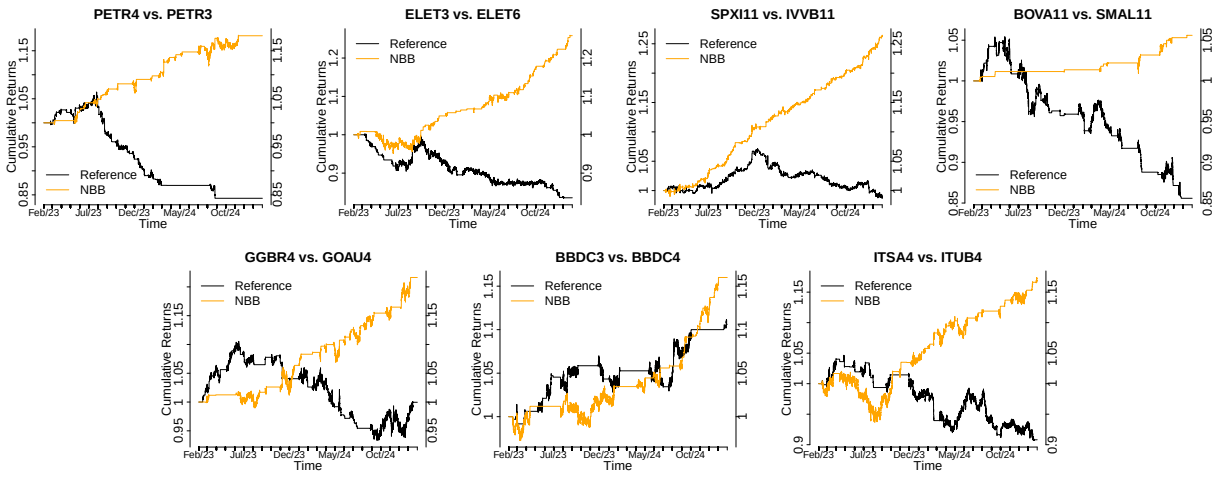

Figure 3.3: Cumulative returns (y-axis) as a multiple of the initial investment for selected pairs in the Brazilian market in the period comprised from 01/01/2014 to 05/10/2024 (x-axis) when applicable. The orange curve indicates the cumulative returns of the proposed Non-Overlapping Block Bootstrap method whereas the black curve shows the performance of the standard cointegration strategy. The parameters used in each pair are described in Table 3.1.

Chapter 4

Bayesian Hierarchical Modeling for Predicting Equipment Life-cycle in the Absence of Data

4.1 Introduction

Accurately predicting equipment failure is a crucial challenge in industrial maintenance and reliability engineering, with far-reaching implications for operational efficiency, safety, and economic sustainability. The evolution of predictive maintenance approaches has historically been marked by a progressive shift from reactive to preventive paradigms, and more recently, toward condition-based and predictive methodologies. These advanced approaches promise significant advantages, including reduced downtime, optimized maintenance scheduling, extended equipment life, and substantial cost savings (Mobley, 2002).

However, the implementation of sophisticated predictive methodologies has remained largely contingent on the availability of extensive historical failure data, a prerequisite that presents an insuperable barrier in many critical contexts. This dependence on historical data creates a paradox where the systems most in need of robust reliability predictions are often those for which historical failure data are least available. This is particularly evident

scenarios in which we have newly installed equipment or facilities; highly specialized or custom-engineered equipment with limited deployment; and safety-critical systems where failures are deliberately engineered to be extremely rare events.

The challenges of predicting equipment failure in data-scarce environments surface in many industries, such as highly specialized biological research facilities and manufacturing environments where equipment reliability directly impacts operational metrics and broader public safety and security considerations. The National Bio and Agro-Defense Facility (NBAF), a high-specialized biosafety laboratory designed to study diseases threatening livestock and public health, exemplifies this challenge. The facility houses highly specialized equipment operating in unique environmental conditions, with virtually no historical failure data available at commissioning and stringent requirements for operational reliability.

This chapter addresses this fundamental challenge by proposing an unorthodox Bayesian approach to estimate time-to-failure in scenarios lacking historical failure data. We explore the theoretical foundations of Bayesian methods for reliability prediction and develop a rigorous hierarchical framework that systematically incorporates expert knowledge to produce a simulated likelihood and informative priors to account for the scarcity of failure data. We then validate this methodology under the reliability theory through application to the NBAF case study. The analysis demonstrates how structural prior elicitation, likelihood simulation, hierarchical parameter sharing, and systematic uncertainty quantification can generate meaningful reliability predictions even in the complete absence of historical failure observations.

The significance of this work extends beyond methodological innovation to practical implementation. We propose establishing a framework for reliability prediction in data-scarce environments. Most fundamentally, the proposed Bayesian approach reframes the challenge of reliability prediction, moving from a paradigm that treats data scarcity as an insuperable barrier to one that acknowledges and systematically incorporates all available sources of knowledge within a coherent probabilistic framework. In doing so, it establishes an adaptive paradigm that evolves naturally with the incremental acquisition of data, bridging the gap between initial expert assessment and data-driven prediction.

4.2 Literature Review on Reliability and Failure Prediction Methods

4.2.1 Introduction

Reliability theory constitutes a cornerstone of the analysis and assurance of the performance and safety of complex technological systems. Initially formalized in the mid-twentieth century, its development was largely driven by military and industrial needs for systematic approaches to assess and improve system dependability. Reliability engineering has evolved from a narrow focus on component failure rates toward the integrative analysis of multi-component, multi-state systems that involve intricate interdependencies among hardware, software, human, and organizational elements (Zio, 2009).

This expansion led the field toward decision-based frameworks that explicitly quantify, evaluate and prioritize risk, reconceptualizing maintenance as an ongoing risk management activity rather than a purely technical function (Khan and Haddara, 2003). The reliability framework applies formal decision-theoretic principles to maintenance, evaluating alternatives through expected utility calculations that incorporate both probability and consequence dimensions (Clemen and Reilly, 2013). This foundation provides a rigorous basis for comparing maintenance strategies and resource allocation decisions under uncertainty, moving beyond simplistic heuristics toward quantifiable optimization that captures the nuanced relationship between intervention costs and risk reduction benefits.

The Probabilistic Foundation of Reliability Theory

At its core, reliability theory requires robust methods for quantifying and propagating uncertainty throughout the decision process. This requirement creates a natural alignment with probabilistic frameworks that can systematically represent both aleatory uncertainty (inherent randomness) and epistemic uncertainty (knowledge limitations) (Der Kiureghian and Ditlevsen, 2009).

For any alternative in the decision process, Clemen and Reilly (2013) argues that conse-

quence values should be weighted by their associated probabilities, building the risk profile for a given strategy. Thus, the risk metric typically takes the form:

$$\text{Risk} = \sum_i P(E_i) \times C(E_i) \tag{4.1}$$

where $P(E_i)$ represents the probability of event E_i, and $C(E_i)$ represents its consequence. This formulation, while conceptually straightforward, presents significant implementation challenges in data-scarce environments where reliably estimating $P(E_i)$ becomes problematic.

Traditional probabilistic methods become innocuous for rare events or novel systems without historical data. As Apostolakis (1990) emphasizes, this limitation has driven high-risk industries toward subjective probability interpretations that accommodate expert judgment, a shift that naturally aligns with Bayesian methods. This synergy between reliability engineering and Bayesian capabilities creates a powerful framework for decision-making under uncertainty, particularly valuable in environments with limited data (Rausand and Høyland, 2004).

Despite its straightforward concept, implementing risk-based maintenance in practice presents significant challenges. Data integration poses a significant hurdle, as reliability studies require the synthesis of heterogeneous data sources, including failure statistics, inspection results, operational variables, and expert judgments. This challenge grows exponentially with the complexity of the system (Rausand and Høyland, 2004). Model uncertainty further complicates implementation, allowing space for significant discrepancies between risk estimates and reality or benchmarks. The inconsistency in the Quantitative Risk Assessment (QRA) eventually undermines stakeholder confidence, putting the strategy under discredit (Pasman et al., 2011). The dynamic nature of risk profiles introduces additional complexity, as many systems exhibit time-varying risk characteristics due to aging, changing operational conditions, or evolving external environments, necessitating dynamic rather than static assessment approaches.

These challenges are particularly acute for newly commissioned facilities or specialized

equipment where historical failure data are unavailable, technical documentation may be limited, and operational experience is still developing. In such contexts, Bayesian methods offer a structured framework for representing initial knowledge states, incorporating multiple information sources, and systematically updating assessments as operational experience accumulates.

4.2.2 Probabilistic Survival Models

Traditional probabilistic survival models can serve as the quantitative foundation for estimating failure probabilities in reliability theory. These models provide the mathematical framework to estimate the "$P(E_i)$" component in risk equation 4.1. These probabilistic approaches can be broadly categorized into non-parametric, parametric, and semi-parametric methods, all of which present distinct challenges in data-scarce environments.

The Kaplan-Meier method (Kaplan and Meier, 1958), widely regarded as the cornerstone of non-parametric survival analysis, provides a step-function estimate of the survival curve directly from observed failure times. Similarly, the Nelson-Aalen estimator (Nelson, 1972; Aalen, 1978) offers a non-parametric approach to estimating the cumulative hazard function. These methods are particularly valuable in contexts where theoretical knowledge about the failure mechanism is limited or when it is desirable to avoid potentially distorting assumptions (Lawless, 2003).

While these methods make minimal distributional assumptions, they are fundamentally data-driven and require substantial empirical observations to produce meaningful estimates. As a consequence, they are effective for estimating survival functions within the range of observed data, but they are not suitable for extrapolations beyond that range as emphasized by Lawless (2003). Since these methods do not assume a specific functional form for the failure time distribution, this prevents reliable predictions outside the observed interval.

Parametric approaches based on specific probability distributions, such as Weibull, Log-normal, Exponential, and others, provide enhanced structure and extrapolation capability but with some limitations. They fundamentally depend on data sufficiency, as parameter

estimation via maximum likelihood techniques demands numerous failure observations to ensure adequate precision (Lawless, 2003). Additionally, standard implementations often impose parameter homogeneity by assuming uniform parameter values throughout the population, thereby disregarding the unit-to-unit variability inherent in many complex systems.

The model selection can also pose some problems in parametric methods. They typically treat model selection as deterministic rather than probabilistic (Raftery, 1995), which can result in excessively confident predictions when observations are limited. Furthermore, basic parametric models generally only accommodate time-independent covariates, struggling to incorporate dynamic factors that influence failure mechanisms, such as changing environmental conditions or maintenance activities. Finally, classical parametric approaches lack formal frameworks for integrating prior or subjective knowledge, rendering them unsuitable for scenarios where expert judgment represents a critical information source.

On a middle ground between parametric and non-parametric approaches, we have the Cox proportional hazards model (Cox, 1972). These methods allow covariate effects without specifying the baseline hazard distribution. This characteristic makes the model semiparametric, enabling estimation of hazard ratios and adjusted survival curves using a minimum of assumptions. This characteristic makes the model widely used and regarded as robust across a range of survival data scenarios (Kleinbaum and Klein, 2012).

The primary disadvantage of the Cox model relies on the proportional hazards assumption, which posits that covariate effects shift the hazard function multiplicatively and remain constant over time. This may not always hold in real-world data, leading to biased estimates and inaccurate conclusions, because the effect of a risk factor might change over time instead of being constant Kleinbaum and Klein (2012); Kumar and Klefsjö (2000).

4.2.3 Fundamental Limitations in Data-Scarce Contexts

Across all these traditional approaches, several common limitations emerge when applied to scenarios lacking historical data.

First, classical survival analysis operates within a frequentist statistical paradigm that

conceptualizes probability in terms of long-run frequencies. This framework becomes philosophically tenuous when applied to rare events or unique equipment where the "repeated trial" concept has limited meaning (Singpurwalla, 2006). The absence of a formal mechanism for incorporating subjective probability assessments restricts the integration of valuable expert knowledge.

Second, traditional methods typically assume homogeneity in model parameters (Gelman et al., 2013), which often does not correspond to the reality of complex facilities where equipment operates under variable conditions and possesses distinct characteristics. This limitation becomes particularly problematic when attempting to transfer knowledge between similar but non-identical systems, which is a common necessity when historical data are scarce.

Third, classical approaches typically treat uncertainty assessment as secondary to point estimation, focusing on confidence intervals that capture sampling variability rather than holistic uncertainty quantification (Droguett and Mosleh, 2012). This emphasis becomes increasingly problematic as data become scarce and epistemic uncertainty dominates.

Finally, traditional methods lack mechanisms for knowledge accumulation and updating, treating each analysis as a standalone exercise rather than part of a continuous learning process. This static approach fails to capitalize on the incremental nature of knowledge acquisition that characterizes the commissioning and early operation of new facilities (Martz and Waller, 1996).

Consequently, there is a pressing need for predictive approaches capable of overcoming these fundamental limitations by effectively integrating expert knowledge, quantifying multiple sources of uncertainty, facilitating knowledge transfer between similar systems, and establishing a framework for continuous model updating as operational experience accumulates. These requirements point naturally toward a Bayesian paradigm, with its explicit incorporation of prior knowledge and systematic uncertainty quantification.

4.3 Bayesian Survival Analysis

In contrast to classical survival methods, Bayesian survival analysis explicitly incorporates prior knowledge into predictive models. Bayesian methods involve updating prior beliefs about model parameters with observed data, thus providing posterior distributions that reflect updated knowledge .

The Bayesian paradigm interprets probability as a measure of subjective belief or degree of certainty about an event, rather than as a long-run frequency of occurrence This interpretation, sometimes called "epistemic", "subjective", or "personal" probability, provides a coherent framework for reasoning about unique events or parameters that cannot meaningfully be conceptualized as random variables with frequency distributions (Lindley, 2000). As Singpurwalla (2006) and Martz and Waller (1996) emphasize, subjective probability interpretation provides a more natural foundation for reliability engineering, particularly when dealing with complex one-of-a-kind systems or rare failure events.

4.3.1 Key Distinctions from Classical Methods

Several fundamental differences distinguish Bayesian from classical methods, with particular relevance for reliability analysis in data-scarce contexts:

1. Parameter Treatment Classical methods treat parameters as fixed quantities and focus on estimator properties (unbiasedness, efficiency, consistency) across hypothetical repeated sampling. Bayesian approaches treat parameters as random variables with distributions reflecting uncertainty about their values. This distinction becomes crucial for reliability prediction, where parameters may directly influence maintenance decisions and system design.

2. Incorporation of Prior Knowledge Perhaps the most distinctive feature of Bayesian inference is its formal incorporation of prior knowledge through the prior distribution. Classical methods typically incorporate external information only informally or through con-

straints on the parameter space, whereas Bayesian methods provide a structured mechanism for representing initial knowledge states (Gelman et al., 2013).

This capability is particularly valuable in engineering contexts where substantial theoretical knowledge, previous experience with similar systems, or expert judgment are also considered as "available information" and may have an even more significant impact on reliability inferential results when test data is scarce (Hamada et al., 2008).

3. Small Sample Performance Classical methods often rely on large-sample asymptotic properties that become questionable with limited data. Maximum likelihood estimators, while theoretically optimal under certain conditions, can produce biased or highly variable estimates with small samples. In contrast, Bayesian methods perform well even with minimal data by tempering likelihood information with prior knowledge (Hamada et al., 2008). This advantage is particularly relevant for specialized equipment with limited operational history, where maintenance decisions must often be made before substantial failure data accumulate (Johnson et al., 2005).

4. Decision-Making Framework Bayesian methods integrate seamlessly with decision theory through the notion of expected utility maximization under parameter uncertainty (Raiffa and Schlaifer, 1968). By averaging over the posterior distribution of model parameters, Bayesian decision analysis naturally accounts for both aleatory and epistemic uncertainties when evaluating alternative actions.

This integration proves particularly valuable for maintenance optimization, where decisions must balance multiple objectives (availability, cost, safety) under substantial uncertainty. As Martz and Waller (1996) demonstrates, Bayesian decision analysis can optimize maintenance intervals even with limited failure data by formally incorporating consequence structures, cost models, and imperfect inspection information.

4.3.2 Bayesian Hierarchical Modeling

Bayesian hierarchical modeling extends the basic Bayesian framework to accommodate multi-level data structures and partial pooling of information across related units. This approach proves particularly valuable for reliability analysis of complex systems comprising multiple components or for facilities with various equipment categories that share certain characteristics but differ in others (Hamada et al., 2008).

The hierarchical structure typically takes the form:

$$y_i \sim p(y_i|\theta_i)$$

$$\theta_i \sim p(\theta_i|\phi)$$

$$\phi \sim p(\phi)$$

where:

- y_i represents data for unit i
- θ_i represents unit-specific parameters
- ϕ represents hyperparameters that govern the distribution of unit-specific parameters

This structure creates a natural mechanism for sharing information across units through partial pooling, striking a balance between completely separate analyses that ignore similarities and completely pooled analyses that ignore differences. The degree of pooling emerges naturally from the data rather than requiring pre-specified assumptions, with stronger pooling when between-unit variation is small relative to within-unit uncertainty (Hamada et al., 2008).

In reliability contexts, hierarchical modeling enables more accurate estimation for units with limited data by utilizing information from similar units with more extensive histories. As demonstrated by Johnson et al. (2005), hierarchical Bayesian approaches can significantly

improve reliability predictions for low-failure-count components by appropriately sharing information across a system while still preserving important unit-specific characteristics.

4.4 Research Problem

All the previously discussed probabilistic methodologies, including the Bayesian approach, presuppose the availability of at least some historical failure data. This fundamental assumption creates a critical gap in reliability engineering: the inability to generate meaningful predictions in contexts of complete data absence. This gap is not merely a methodological inconvenience but represents a significant barrier to implementing risk-based management in precisely those environments where it would be most valuable: newly commissioned facilities, specialized equipment with limited deployment, or safety-critical systems engineered for extreme reliability.

The challenge of complete data absence is qualitatively different from the more commonly addressed problem of small sample sizes or censored data in survival analysis. With small samples, traditional techniques face limitations but can still produce estimates, albeit with higher uncertainty. In contrast, complete data absence renders traditional maximum likelihood estimation mathematically impossible and prevents even the updating of prior distributions in standard Bayesian frameworks. The absence of likelihood information creates a fundamental barrier that cannot be overcome through frequentist or common Bayesian approaches, regardless of their sophistication.

Despite the fact several approaches have been proposed to address data scarcity in reliability prediction, each of them faces substantial limitations when confronting complete data absence. Accelerated life testing, for example, while valuable for generating failure data more rapidly, requires physical specimens for testing and may introduce additional uncertainties when extrapolating to normal operating conditions. Similarly, data pooling, which aggregates data across similar equipment to expand the available information base, requires the existence of sufficiently similar units with documented failure histories. This condition is often not met for specialized or novel systems (Hamada et al., 2008; Meeker and Escobar,

1998).

Other common approaches face similar constraints. Using surrogate data from related components or systems as proxies introduces substantial model uncertainty and requires strong assumptions about the transferability of reliability characteristics across different contexts (Zio, 2009; Hamada et al., 2008; Johnson et al., 2005; Meeker and Escobar, 1998). Purely physics-based models, while theoretically capable of predicting failure without empirical data, require a comprehensive understanding of all potential failure modes and their underlying mechanisms, a level of knowledge rarely available for complex systems.

When facing complete data absence, these approaches either require resources unavailable during initial deployment (accelerated testing), rely on data that simply does not exist (pooling and surrogates), or demand theoretical knowledge beyond current capabilities (physics-based models). Consequently, reliability engineers often resort to heuristic approaches, conservative over-design, or arbitrary safety factors. These methods lack formal uncertainty quantification and may lead to suboptimal resource allocation.

4.4.1 The Bayesian Paradigm: An Incomplete Solution

The Bayesian framework, with its explicit incorporation of prior knowledge, initially appears to offer a natural solution to the data-scarcity problem. By encoding expert judgment, theoretical understanding, or manufacturer specifications into prior distributions, Bayesian methods theoretically provide a balancing mechanism between expert judgment and evidence coming from data.

However, standard Bayesian approaches still require a likelihood function based on observed data to update prior beliefs into posterior distributions. Bayesian inference fundamentally involves the synthesis of prior knowledge with empirical evidence through Bayes' theorem. In the complete absence of data, standard Bayesian updating reduces to a tautology where the posterior simply equals the prior, providing no formal mechanism for refined prediction or uncertainty quantification beyond initial assumptions.

Despite this limitation, the Bayesian paradigm offers two essential capabilities that make

it a promising foundation for addressing the data absence challenge:

- Explicit Knowledge Representation: The prior distribution provides a formal mechanism for encoding diverse knowledge sources, including theoretical understanding, physical constraints, expert judgment, and analogical reasoning.
- Uncertainty Propagation: The Bayesian framework offers established methods for propagating uncertainties through complex models and decision processes, enabling robust risk assessment even with limited information.

These capabilities suggest that while standard Bayesian approaches may be insufficient for the complete data absence problem, extensions of the Bayesian paradigm that address the likelihood challenge could potentially bridge this critical methodological gap.

4.5 Proposed Method

4.5.1 An Alternative Bayesian Approach to the Data Absence Problem

This research proposes an unorthodox approach to the Bayesian framework specifically designed to generate meaningful reliability predictions in the complete absence of historical failure data. Our methodology introduces two interconnected innovations that collectively address the fundamental challenge of prediction in data-scarce environments.

Previous work by Johnson et al. (2005) proposed a hierarchical Bayesian model for estimating the early reliability of complex systems when little or no system-level failure data are available. Their approach relied on borrowing data from comparable systems developed by different categories of manufacturers, incorporating a "quality" index to distinguish between systems produced by experienced and inexperienced manufacturers. While innovative, this approach has limitations when applied to unique or highly specialized equipment. Borrowing data from similar equipment can be problematic because failure mechanisms are often highly context-dependent, influenced by environmental conditions, maintenance practices,

and other operational characteristics that may differ substantially from those of the target equipment.

Our proposal goes beyond Johnson et al.'s 2005 approach by eliminating the need for borrowed failure data altogether. The foundation of our method consists of a hierarchical Bayesian structure in which we simulate the likelihood function directly from operational information of the equipment under study so we can depart from conventional reliance on empirical failure observations. We construct this function by synthesizing theoretical considerations, physical constraints, and domain knowledge about failure mechanisms. This innovative adaptation enables Bayesian updating to proceed even when conventional failure data are entirely absent, effectively circumventing the traditional prerequisite for historical observations.

Second, building upon this simulated likelihood, we conduct meticulous prior elicitation to allow for systematic knowledge integration across multiple sources. This hierarchical framework incorporates manufacturer specifications, engineering judgment, physical understanding of failure mechanisms, and analogical reasoning from related systems, with explicit quantification of the uncertainty associated with each source. Our approach carefully structures prior distributions based on readily available operational variables such as manufacturer-specified expected lifetime, equipment runtime, maintenance quality, and system criticality. All these types of information are typically accessible even in newly commissioned facilities. This structured knowledge representation creates a formal mechanism for transforming qualitative expert assessments into quantifiable probabilistic statements.

Another direct consequence of systematic knowledge integration through the priors is that we establish explicit relationships between observable operational variables and reliability parameters. By linking operational characteristics to the model's probabilistic structure, we create a mechanism for continuously refining predictions based on evolving operational experience, even before actual failures occur. This capability enables the model to learn from operational data that precedes failures, making prediction refinement possible during the critical early lifecycle phases where traditional methods remain uninformative.

This approach fundamentally reconceptualizes the reliability prediction problem in data-

scarce environments, moving from a paradigm that treats failure data as a prerequisite for meaningful analysis to one that utilizes all available knowledge sources within a coherent probabilistic framework. Most importantly, it provides a systematic pathway from initial expert assessment to data-driven prediction by establishing a model structure that can seamlessly incorporate empirical observations as they gradually accumulate, creating a natural bridge between the data-scarce commissioning phase and the data-rich operational phase without requiring methodological discontinuity.

4.5.2 Implementation Through Case Study: The NBAF Application

To validate this methodology, we apply it to the challenging case of the National Bio and Agro-Defense Facility (NBAF). This facility represents an ideal test environment for our approach due to several characteristics that embody the data absence challenge in high-consequence environments.

The NBAF exemplifies the complexity-criticality paradox common in specialized facilities. It houses highly sophisticated equipment operating under unique environmental conditions, including high containment pressurization systems, specialized decontamination protocols, and stringent operational requirements that create unprecedented reliability demands. These specialized operating conditions mean that even equipment with established reliability histories in conventional settings may exhibit different failure patterns in this unique context.

Compounding this complexity is the facility's status as a recently commissioned installation, which leaves it with virtually no historical failure data. Yet paradoxically, this same facility requires exceptionally robust reliability predictions to support comprehensive risk management and budget planning, precisely when such predictions are most difficult to generate using conventional methods. This timing mismatch between prediction needs and data availability creates the perfect test case for our methodology.

The stakes of prediction accuracy at NBAF are particularly high due to the facility's critical mission and safety implications. The facility's work with high-consequence pathogens

creates an extremely low tolerance for equipment failure, particularly for containment-critical systems where failures could potentially impact not only research continuity but also biosafety and biosecurity protocols.

Further complicating the validation challenge, many of the facility's specialized systems have no direct analogs elsewhere in the research infrastructure landscape. This uniqueness prevents straightforward data transfer or reliability inference from similar installations, eliminating another traditional approach to addressing data scarcity. Systems such as specialized air handling units, decontamination equipment, and containment barrier technologies have been custom-engineered for the specific requirements of NBAF, creating a prediction context where expert judgment must be formalized rather than simply supplemented with historical data.

Through this case study, we demonstrate how our proposed Bayesian approach can generate meaningful reliability predictions for critical NBAF equipment by systematically incorporating manufacturer specifications, engineering judgment, theoretical understanding of failure mechanisms, and facility-specific contextual factors. We further conduct validation sessions with stakeholders who systematically assess the prediction consistency. The empirical study suggests that this method offers a robust probabilistic framework for evaluating prediction accuracy even in the absence of failure data, addressing a fundamental challenge in reliability methodology for critical systems designed to operate without failures.

4.5.3 Hierarchical Bayesian Weibull Model

While our methodology employs the Weibull distribution for the NBAF case study, it is important to emphasize that the fundamental innovations proposed in this research, namely the simulated likelihood construction and meticulous prior development, are model-agnostic. The proposed approach using the hierarchical Bayesian framework could theoretically accommodate different distributions, hierarchical levels, and settings appropriate for other specific failure mechanisms. The critical elements of our approach lie not in the specific distributional form selected, but rather in the systematic handling of complete data absence through

principled prior construction and the development of simulated likelihood functions that do not require historical failure observations.

The Weibull distribution forms the foundation of our approach to the NBAF problem due to its exceptional versatility in capturing diverse failure patterns across equipment lifecycles. In its classical form, the Weibull hazard function is characterized by:

$$h(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1}, \quad t > 0, k > 0, \lambda > 0 \tag{4.2}$$

where k represents the shape parameter (also known as the failure mode index or Weibull slope) and λ the scale parameter (commonly referred to as the characteristic life parameter). This formulation elegantly accommodates various failure behaviors through the shape parameter k:

- $k < 1$: Decreasing failure rate (characteristic of early-life failures or "infant mortality")
- $k = 1$: Constant failure rate (simplifying to the exponential distribution)
- $k > 1$: Increasing failure rate (modeling wear-out and aging related failures)

This adaptability makes the Weibull model particularly suitable for reliability engineering applications, especially when modeling mechanical components subject to various degradation mechanisms (Rausand and Høyland, 2004; ReliaSoft, 2020; Omari and Ibrahim, 2011).

We employ a two-level hierarchical structure for each equipment $i \in \{1, 2, \ldots, g\}$ in a group:

$$t_i \sim Wei(k_i, \lambda_i) \tag{4.3}$$

$$k_i \sim \mathcal{LN}(\mu_i, \sigma_i^2) \tag{4.4}$$

$$\lambda_i \sim \Gamma(\alpha_i, \beta_i) \tag{4.5}$$

This formulation creates a coherent probabilistic framework where the Weibull likelihood for time-to-failure is paired with carefully selected prior distributions for the shape and scale parameters. Our choices of the Log-normal and Gamma distributions are not arbitrary but guided by both theoretical considerations and practical requirements. The selection process reflects a careful balance between mathematical properties, physical constraints, and interpretability needs in the data-absent reliability context.

For the shape parameter k_i, the Log-normal distribution is selected for several compelling reasons that combine mathematical rigor with practical utility. The distribution's strictly positive domain satisfies the natural constraint that the Weibull shape parameter must be positive, eliminating the need for artificial truncation or transformation that might complicate interpretation. Beyond this basic constraint, the Log-normal offers remarkable flexibility in representing uncertainty through its dispersion parameter σ_i^2, allowing for representation from highly confident narrow distributions to diffuse representations of limited knowledge.

This flexibility aligns with empirical observations about failure mechanisms in complex systems. Studies by Abernethy (2006) suggest that shape parameters for many engineering systems naturally exhibit right-skewed distributions, a pattern that the Log-normal effectively captures through its inherent asymmetry. Perhaps equally important in the context of communicating with domain experts, the Log-normal parameters μ and σ^2 have clear interpretations related to the central tendency and spread of the shape parameter, facilitating meaningful dialogue between reliability analysts and engineers who may lack statistical expertise but possess critical domain knowledge.

For the scale parameter λ_i, the Gamma distribution offers a complementary set of advantages that address both theoretical and practical concerns. Like the Log-normal, the Gamma distribution preserves the positivity of the scale parameter, maintaining the physical validity of the model without requiring constraints that might complicate posterior sampling or interpretation.

The practical benefits of the Gamma prior extend to its natural relationship with expected equipment lifetime. The parameters of the Gamma distribution (α and β) can be directly related to the expected value and variance of equipment lifetime, creating a straightforward

pathway for incorporating manufacturer specifications into the prior formulation. This parameter mapping facilitates the translation between engineering specifications (mean time to failure and associated uncertainty) and statistical parameters, bridging the gap between reliability engineering and statistical modeling. Furthermore, the shape and rate parameters of the Gamma distribution can be adjusted to represent varying degrees of certainty, from highly informative priors derived from substantial domain knowledge to appropriately diffuse priors when information is limited.

To simulate the Weibull likelihood, we rely, one more time, on prior knowledge and equipment operational characteristics. We specifically use this information to assign values to k_i and λ_i. For example, with recently installed equipment, we tend to simulate the Weibull distribution with $k_i < 1$, reflecting the higher probability of early-stage failure due to installation errors or design defects. Otherwise, we tend to simulate aging and wear-out failures by using $k > 1$. Although there seems to be a wide range of possibilities for the last case, according to Abernethy (2006), the common choice is usually a number between 1^+ and 4. According to the same author, extremely high k values such as 10, for example, *"are uncommon and represent a very rapid failure rate not typically seen in practical scenarios"*. To the parameter λ_i, which is strongly linked to the expected lifetime, we assign the average or median of the lifetimes provided by the manufacturer for a specific group of similar equipment.

The combination of Log-normal and Gamma priors with the Weibull likelihood creates a mathematically coherent framework that balances theoretical considerations with practical interpretability. This is a desired feature when communicating with facility engineers and maintenance personnel who must translate model outputs into operational decisions. The hierarchical structure enables systematic integration of available knowledge through the hyperparameters μ_i, σ_i, α_i, and β_i. By carefully parameterizing these priors based on available operational knowledge, our approach provides meaningful reliability predictions even in the complete absence of historical failure data, addressing the fundamental challenge that traditional methods cannot overcome.

4.5.4 Informative Prior Specification

One of the main contributions of this research is the development of systematic methods for specifying informative priors based on expert knowledge and equipment characteristics. Our prior specification approach is carefully designed to translate qualitative assessments and equipment metrics into probabilistic distributions that can function effectively without historical data. Operational variables such as the expected lifetime specified by the manufacturer, the service life of the equipment, the quality of maintenance, and the criticality of the system - all information typically accessible even in newly commissioned facilities - can be incorporated through structured relationships with these hyperparameters.

Furthermore, all the prior input variables are treated as proportions of the expected life provided by the manufacturer for the specific equipment, which consists of accessible information in most cases. This strategy provides significant advantages for reliability modeling in data-scarce environments:

- Dimensionless Scaling: By using proportions rather than absolute values, the model achieves scale-invariance across equipment with different lifespans. This enables consistent parameter interpretation regardless of whether the equipment has an expected lifetime of 5 years or 50 years.

- Transferability Across Equipment Types: The proportional approach facilitates knowledge transfer between different equipment categories, allowing the modeling framework to apply equally well to short-lived components and long-lived infrastructure without requiring recalibration of the fundamental mapping functions.

- Intuitive Expert Elicitation: It is usually more natural to express assessments in relative terms (e.g.,"this equipment has operated for approximately 40% of its expected lifetime") rather than making absolute judgments about failure distributions. The proportional framework aligns naturally with this intuitive reasoning process.

- Coherent Uncertainty Propagation: When equipment lifespan itself contains uncertainty, the proportional approach ensures that this uncertainty propagates consistently

through the model. For example, uncertainty about whether a component will last 8 or 10 years automatically translates into appropriate uncertainty about the current proportional age if the component is 4 years old.

- Integration with Manufacturer Specifications: Manufacturer-supplied lifetime estimates typically represent ideal conditions. By treating operational factors as proportions of these idealized estimates, the model naturally incorporates deviations from optimal conditions through the mapping functions, creating a principled framework for adjusting ideal-case estimates to reflect actual operating environments.

- Simplified Sensitivity Analysis: Proportional parameterization simplifies sensitivity analysis by providing a standardized domain (typically $[0, 1]$) for key variables, facilitating systematic exploration of parameter spaces and making results more comparable across different equipment types and failure modes.

The proportional approach further enables consistent integration across different information sources. For instance, when runtime hours are expressed as a proportion of expected lifetime runtime, and equipment age is expressed as a proportion of expected calendar life, these disparate operational metrics become directly comparable and can be integrated through the mapping functions without requiring complex conversions or arbitrary weighting schemes.

Gamma Parameters Mapping

The expected life of the equipment is mapped to the Gamma distribution's parameters, α and β, to model λ, the scale parameter of the Weibull distribution, using the following functions that incorporate variability:

$$\alpha_i = \frac{1}{\Delta_i^2} \tag{4.6}$$

$$\beta_i = \frac{\alpha_i}{\xi_i} \tag{4.7}$$

where ξ_i is the expected life of a specific equipment given by the manufacturer, and dispersion Δ_i is a measure of uncertainty derived from the difference between manufacturer specifications and engineering assessments for the expected life. This dispersion factor is calculated as:

$$\Delta_i = \max\left(\frac{|\xi_i - \hat{\mu}_{\text{eng},i}|}{\xi_i}, c\right) \tag{4.8}$$

The parameter c ($c > 0$) ensures numerical stability when the engineer's estimate closely matches the manufacturer's specification. $\hat{\mu}_{\text{eng},i}$ is the midpoint of the estimated minimum and maximum life values of the engineer:

$$\hat{\mu}_{\text{eng},i} = \frac{a_i + b_i}{2} \tag{4.9}$$

with a_i being the engineer's minimum estimated life and b_i the engineer's maximum estimated life for the equipment.

This mapping approach directly incorporates the discrepancy between manufacturer specifications and engineering assessments as a measure of uncertainty. The resulting Gamma distribution has a mean exactly equal to the manufacturer's expected life ($\alpha_i/\beta_i = \xi_i$), while its variance is controlled by the relative difference between expectations.

When engineering assessments differ significantly from manufacturer specifications, the resulting distribution exhibits greater dispersion, reflecting increased uncertainty (Figure 4.1). Conversely, when assessments closely align with specifications, the distribution becomes more concentrated, indicating higher confidence in the expected life estimate.

Finally, the mapping function ensures that, as the manufacturer's expected life increases, the survival curve shifts rightward while maintaining appropriate dispersion, representing a longer time until failure. Furthermore, equipment with higher dispersion (larger difference between manufacturer and engineering assessments) will exhibit survival curves with distinct

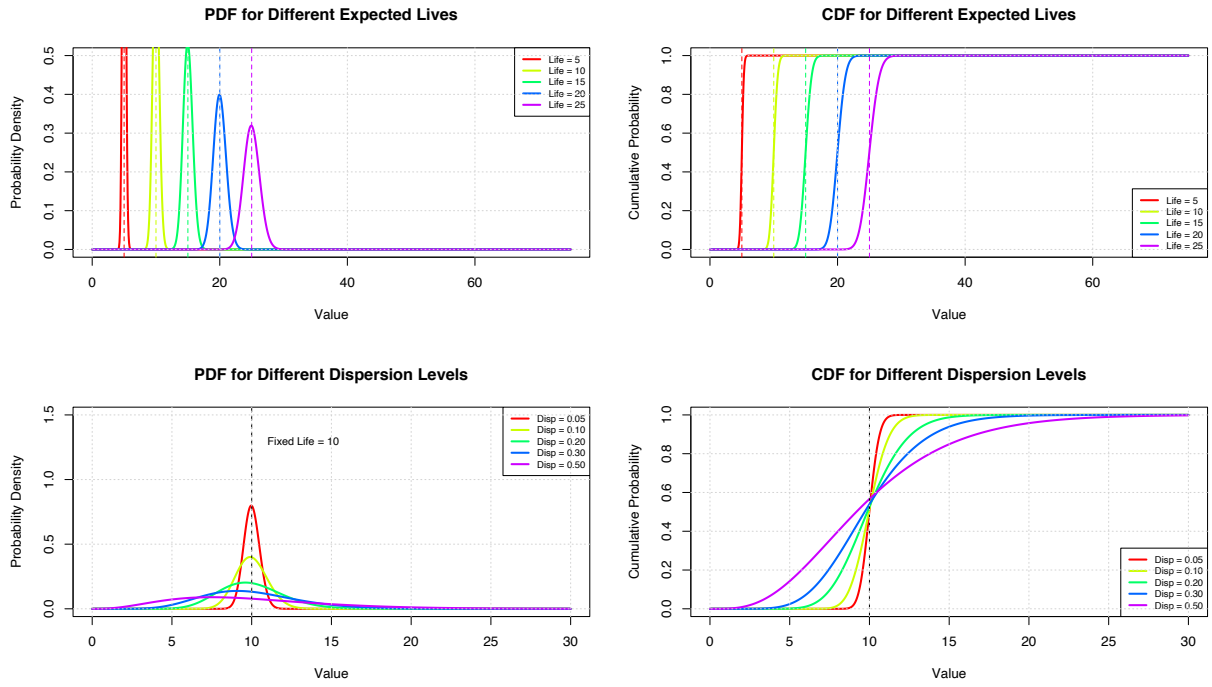

Figure 4.1: Sensitivity of the Gamma PDF and CDF to different expected-life (top row) and dispersion values (bottom row).

shapes, reflecting the increased uncertainty in their failure times (Figure 4.2). This mapping structure ensures that the relationship between expected lifetime, engineering assessment variability, and survival probabilities behaves in accordance with practical reliability engineering intuition while maintaining appropriate statistical properties and offering greater flexibility for different equipment types with varying levels of specification confidence.

Log-normal Parameters Mapping

The failure rate, represented by the shape parameter k, is mapped from several operational characteristics of the equipment-such as age, runtime, maintenance level, and criticality- as a proportion of the expected life. Each variable contributes to the prior for k through the following mapping function:

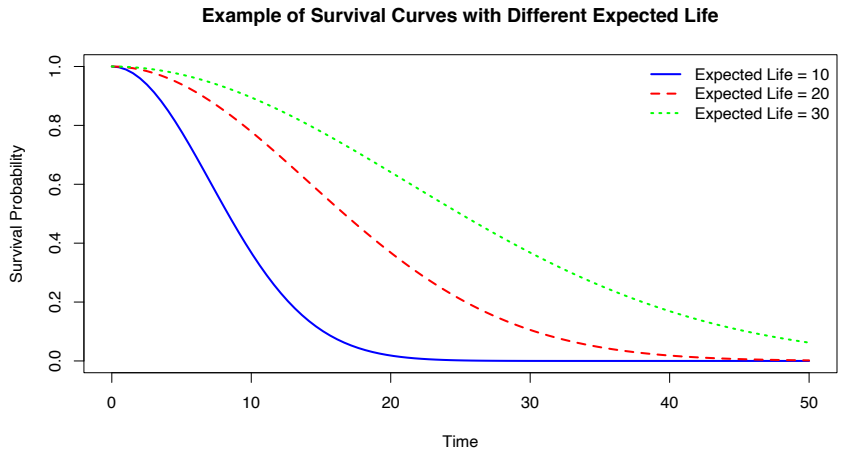

Figure 4.2: Survival curves for equipment with different expected lifetimes showing horizontal shifts and changes in curve shape.

$$\sigma_{k_i}^2 = x^2 \tag{4.10}$$

$$\mu_{k_i} = \log(\hat{k}_i) \times C - \frac{\sigma_{k_i}^2}{2} \tag{4.11}$$

where $\hat{k}_i$ represents the baseline failure rate for the analyzed equipment, x is the proportion of the expected life for the given input variable, and C is a scaling factor to make the survival curve more or less responsive. Note that equation 4.11 is derived from the formula relating the μ parameter of the Log-normal distribution to its expected value $\mathbb{E}[X] = e^{\mu + \frac{\sigma^2}{2}}$. It adjusts the μ parameter of the Log-normal distribution so that, when combined with the $\sigma^2/2$ term, the distribution has a behavior centered around the value k^C. The parameter C should be calibrated through sensitivity analysis to ensure appropriate responsiveness of the model to changes in input variables while maintaining plausible survival curves across the range of expected equipment lifetimes. This calibration has to be performed through extensive simulation testing with subject matter experts to validate that the resulting predictions are aligned with engineering expectations.

The value k can be empirically chosen depending on the life-cycle and type of equipment.

As previously mentioned, an appropriate choice for k allows us to mimic increased early life failure ($k < 1$); constant failure risk across the whole equipment life-cycle ($k = 1$); or increasing failure rate characteristic of mechanical systems that degrade over time ($k > 1$).

The mapping function adjusts both the position (μ) and spread (σ^2) of the Log-normal distribution, allowing the failure rate to vary according to the input characteristics in a way that reflects increasing uncertainty as equipment ages, is exposed to longer runtimes, or did not receive appropriate maintenance for example. Figure 4.3 illustrates the expected behavior of parameters μ and σ^2 for given values of k and C.

For categorical variables such as maintenance level and criticality, we assign different proportions to each level to integrate expert knowledge into the equipment's life-cycle model. In our NBAF case study, maintenance conditions are classified into three categories:"green" (good),"yellow" (moderate), and "red" (poor). We quantify these classifications with values of 0.1, 0.5, and 0.9 respectively, following the principle that higher proportions reflect greater failure probabilities. Similarly, we apply this approach to criticality ratings, which range from "1" (critical equipment) to "5" (less critical components). This methodology can be adjusted to capture asymmetric relationships between categories, enabling the model to reflect non-uniform impact of different classification levels on equipment reliability.

Independently of the values of $\hat{k}_i$ and C, as the proportion increases, both μ and σ^2 reflect greater uncertainty in the failure rate, which is typical as equipment ages or is exposed to worse conditions of the input variables. This mapping affects directly the dispersion of the survival curve as illustrated in Figure 4.4.

4.5.5 Single Equipment vs. Group Modeling

The hierarchical Bayesian structure allows us to model equipment-specific variability and shared characteristics, providing appropriate failure probability calculations for each case.

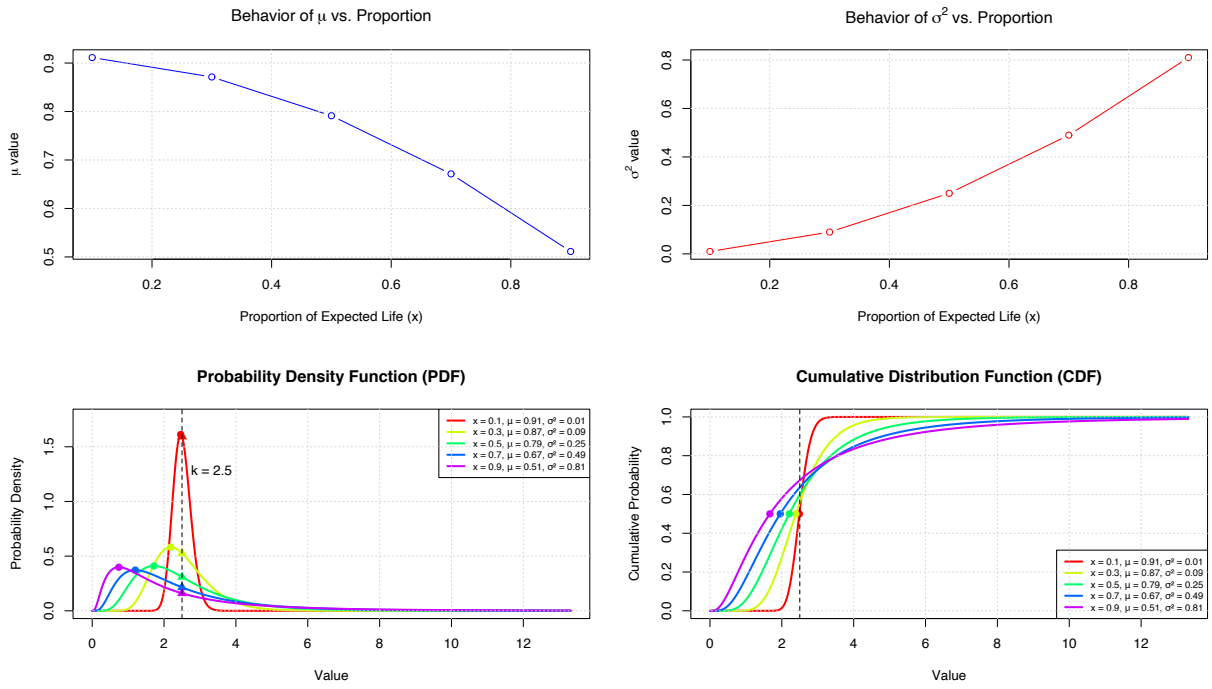

Figure 4.3: Sensitivity of the Log-normal parameters (top row) and PDF (bottom row) to different choices of x and considering $k = 2.5$.

Failure Probability for a Single Equipment

For individual equipment analysis, we employ Metropolis-Hastings Markov Chain Monte Carlo (MCMC) to generate the posterior distribution for each equipment. While the prior distributions for k and λ are specified analytically, the resulting posterior distributions generally do not have closed-form expressions.

Our model assumes that, conditional on specific values of k and λ, the time to failure follows a Weibull distribution with the corresponding survival function for equipment i:

$$S_i(t|k,\lambda) = \exp\left(-\left(\frac{t}{\lambda}\right)^k\right) \tag{4.12}$$

To account for parameter uncertainty, we use the posterior samples from MCMC to

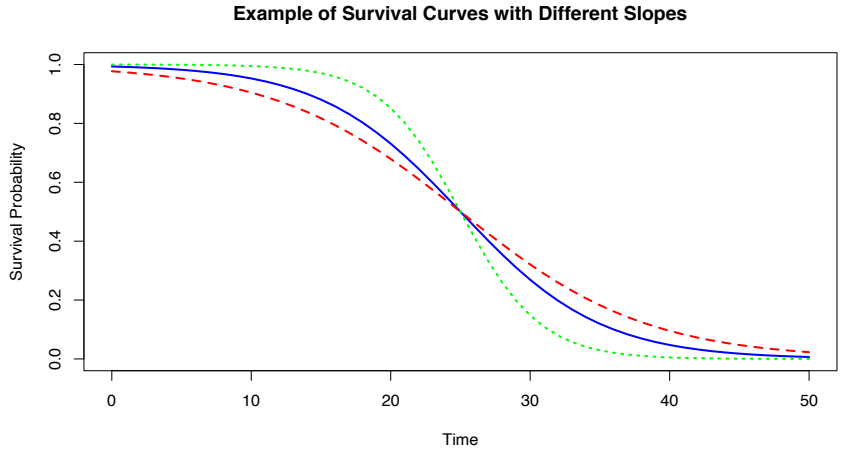

Figure 4.4: Survival curves for equipment with different operational characteristics that affect failure rate k showing changes in curve dispersion.

compute the marginal survival function by numerical integration:

$$S_i(t) = \mathbb{E}_{k,\lambda|t}[S_i(t|k,\lambda)] \approx \frac{1}{N}\sum_{j=1}^{N} S(t|k_j,\lambda_j) \tag{4.13}$$

where N is the number of posterior samples, and (k_j, λ_j) represent the j-th sample from the joint posterior distribution.

The probability of failure at time t for equipment i (p_{t_i}) is then computed directly from this marginal survival function:

$$p_{t_i} = 1 - S_i(t) \tag{4.14}$$

Failure Probability for Groups

For groups of equipment, our concern is concentrated on group cumulative failure probabilities. In a Bayesian framework, these probabilities incorporate the posterior distributions of each equipment's parameters.

At a high level, the probability that at least m out of n equipment fail by time t can be expressed as:

$$P(N(t) \geq m) = \int_{\theta_1} \cdots \int_{\theta_n} P(N(t) \geq m|\theta_1, \ldots, \theta_n) \cdot p(\theta_1|t_1) \cdots p(\theta_n|t_n) d\theta_1 \cdots d\theta_n, \quad (4.15)$$

where $\theta_i = (k_i, \lambda_i)$ represents the Weibull parameters for equipment i, and $p(\theta_i|t_i)$ is the posterior distribution of these parameters obtained through Metropolis-Hastings sampling.

Expanding the conditional probability $P(N(t) \geq m|\theta_1, \ldots, \theta_n)$, we have:

$$P(N(t) \geq m|\theta_1, \ldots, \theta_n) = \sum_{x=m}^{n} \sum_{S \in \mathcal{C}(n,x)} \int_0^t \cdots \int_0^t \prod_{i \in S} f_i(t_i|\theta_i) \prod_{j \notin S} [1 - F_j(t|\theta_j)] dt_1 \cdots dt_x, \quad (4.16)$$

where $N(t)$ is the number of equipment that fail until time t; $\mathcal{C}(n, x)$ represents all possible combinations of x equipment from the group of n; S denotes a specific combination of x equipment that fail; $f_i(t_i|\theta_i)$ is the Weibull likelihod for equipment i given its parameters; and $F_j(t|\theta_j)$ is the Weibull cumulative distribution function for equipment j given its parameters.

This high-dimensional nested integral is typically intractable analytically since we need to integrate each possible failing and non-failing equipment combination over all the failure time possibilities from the current time to time t. Monte Carlo simulation is particularly well-suited for this task (Gilks et al., 1996). We then implement a two-stage approach combining Metropolis-Hastings MCMC with Monte Carlo simulation as follows:

1. For each piece of equipment in the group, parameters such as expected life, age, maintenance level, runtime, and criticality are used to generate appropriate prior distributions: Gamma distributions for λ and Log-normal distributions for k.

2. For each equipment i, we perform Metropolis-Hastings MCMC to obtain posterior samples of k_i and λ_i that account for both the prior distributions and simulated data from a Weibull likelihood:

$$p(k_i, \lambda_i|t_i) \propto p(t_i|\hat{k}_i, \hat{\lambda}_i) \cdot p(k_i) \cdot p(\lambda_i) \quad (4.17)$$

3. In each Monte Carlo simulation iteration j (of J total iterations), for each equipment i, we randomly select one posterior sample $(k_i^{(j)}, \lambda_i^{(j)})$ from its respective posterior distribution.

4. Using these sampled parameters, we generate a remaining time to failure $R_i^{(j)}$ for each equipment using the Weibull distribution and adjusting for the current age:

$$T_i^{(j)} \sim \text{Weibull}(k_i^{(j)}, \lambda_i^{(j)}) \tag{4.18}$$

$$\text{and} \quad R_i^{(j)} = T_i^{(j)} - \text{Age}_i \tag{4.19}$$

5. For each simulation j, we count the number of failures occurring within the specified time horizon t:

$$N_j(t) = \sum_{i=1}^{g} \mathbf{1}(R_i^{(j)} \leq t) \tag{4.20}$$

where $\mathbf{1}(\cdot)$ is the indicator function and g is the total number of equipment in the group.

6. After J simulation iterations, the probability of at least m failures out of total g equipment within time t is computed as:

$$P(N(t) \geq m) = \frac{1}{J} \sum_{j=1}^{J} \mathbf{1}(N_j(t) \geq m) \tag{4.21}$$

These steps are summarized in Algorithm 4.

For group-level analysis, only the probability of failure and the corresponding histogram showing the distribution of the number of failures across simulations (Figure 4.5) can be provided. The appropriate visualization for group failure dynamics through a survival curve would require more complex multi-state representations or multivariate survival functions that are beyond the scope of standard visualization techniques.

Algorithm 4 Bayesian Hierarchical Monte Carlo for Equipment Failure Probability

Require: Equipment group $\mathcal{G}$ with g items, time horizon t, MCMC iterations M, MC simulations J, failure threshold m

Ensure: $P(N(t) \geq m)$

▷ Phase 1: Posterior sampling via MCMC for individual equipment in the group

for $i = 1$ to g **do**
  Set $\mu_{k_i}, \sigma^2_{k_i}, \alpha_{\lambda_i}, \beta_{\lambda_i} \leftarrow f(\text{params}_i)$ ▷ Map equipment parameters to priors
  $\mathcal{D}_i \sim \text{Weibull}(k_{\text{base}}, \lambda_{\text{base}})$ ▷ Generate simulated data
  $k_i^{(0)}, \lambda_i^{(0)} \leftarrow$ initial values
  for $s = 1$ to M **do**
    $k_i^* \sim q_k(k_i^* | k_i^{(s-1)})$ ▷ Propose k from Log-normal
    $\lambda_i^* \sim q_\lambda(\lambda_i^* | \lambda_i^{(s-1)})$ ▷ Propose λ from Gamma
    $\alpha \leftarrow \min\left\{1, \frac{p(\mathcal{D}_i | k_i^*, \lambda_i^*) p(k_i^*) p(\lambda_i^*)}{p(\mathcal{D}_i | k_i^{(s-1)}, \lambda_i^{(s-1)}) p(k_i^{(s-1)}) p(\lambda_i^{(s-1)})}\right\}$
    if $U(0,1) < \alpha$ **then**
      $k_i^{(s)} \leftarrow k_i^*$, $\lambda_i^{(s)} \leftarrow \lambda_i^*$
    else
      $k_i^{(s)} \leftarrow k_i^{(s-1)}$, $\lambda_i^{(s)} \leftarrow \lambda_i^{(s-1)}$
    end if
  end for
  $\mathcal{P}_i \leftarrow \{(k_i^{(s)}, \lambda_i^{(s)})\}_{s=b}^{M}$ with thinning τ ▷ Store posterior after burn-in b
end for

▷ Phase 2: Monte Carlo simulation using posterior samples

Initialize $N \leftarrow [0]^J$
for $j = 1$ to J **do**
  for $i = 1$ to g **do**
    $(k_i^{(j)}, \lambda_i^{(j)}) \sim \mathcal{P}_i$ ▷ Random sample from posterior
    $A_i \leftarrow \text{AgeFactor}_i \cdot \text{ExpectedLife}_i$ ▷ Current age
    $T_i^{(j)} \sim \text{Weibull}(k_i^{(j)}, \lambda_i^{(j)})$ ▷ Time to failure
    $R_i^{(j)} \leftarrow T_i^{(j)} - A_i$ ▷ Remaining life
    if $R_i^{(j)} \leq t$ **then**
      $N[j] \leftarrow N[j] + 1$
    end if
  end for
end for
$S \leftarrow \sum_{j=1}^{J} \mathbb{1}(N[j] \geq m)$ ▷ Count simulations with at least m failures
return $P(N(t) \geq m) = S/J$

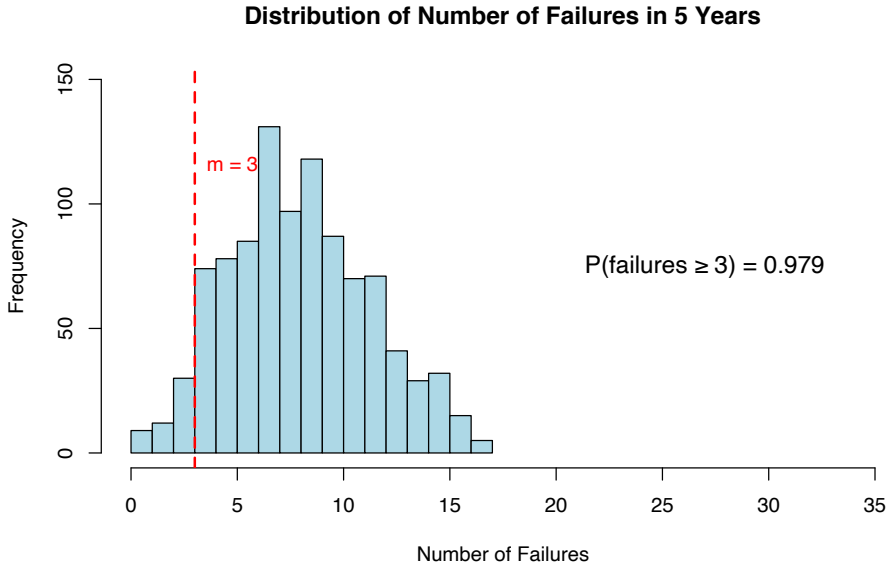

Figure 4.5: Distribution of the number of failures obtained through Monte Carlo simulations for group-level analysis. The histogram illustrates the frequency of occurrence of different failure quantities, enabling visualization of the uncertainty associated with the failure prediction for the equipment group. This representation complements the point probability of failure, offering a more comprehensive perspective on the probabilistic behavior of the system.

4.5.6 Model Validation

Application to NBAF's specialized equipment demonstrated the Bayesian model's capability to provide reliable survival estimates without initial historical data. The model parameters and prior distributions were calibrated through a comprehensive, multi-stage process combining computational testing with expert validation.

Initial parameter values were established based on theoretical considerations provided by experts and stakeholders. These parameters were then refined through multiple structured review sessions, where stakeholders assessed the plausibility of generated survival curves. To facilitate this validation process, an interactive application (Figure 4.6) was developed using the Shiny framework (Chang et al., 2023), allowing stakeholders to test different input variable combinations and immediately evaluate the resulting survival curves and probability estimates. This interactive approach enabled more intuitive assessment and fine-tuning of

the model. To ensure appropriate model responsiveness, systematic iterative sensitivity analyses were performed until the model produced survival curves and failure probabilities that aligned with stakeholder expectations across diverse equipment types and conditions.

The appropriateness of the mapping functions used in the priors was critical in this process. These functions enabled the posterior density to successfully reflect differences in the input parameters. Expected lifetime values incorporated into the Gamma distribution maintained the appropriate shape of the survival curve while causing a horizontal shift and dispersion changes proportional to changes in the expected lifetime. Complementarily, the other operational variables incorporated into the Log-normal distribution produced appropriate changes in the dispersion of the survival function, effectively increasing or decreasing the area under the curve (survival probabilities).

4.6 Conclusion and Future Research

The hierarchical Bayesian framework developed in this work effectively overcomes the problem of scarce historical data by incorporating structured expert knowledge through informative priors and simulating the likelihood function. It integrates seamlessly with reliability theory, providing a quantitative foundation for prioritizing maintenance activities based on assessed failure risks.

The proposed approach represents not just a methodological contribution but also a practical solution to a critical problem faced by many new or highly specialized facilities. By providing a statistical framework that evolves with data acquisition, this model fills an important gap between initial qualitative judgment and traditional data-driven quantitative approaches.

Among the method's restrictions, it's important to note that the likelihood function is also "contaminated" by prior beliefs, as the data generation process itself incorporates expert judgments. The quality of predictions remains highly dependent on the accuracy of expert knowledge incorporated via priors, requiring well-structured processes for expert knowledge elicitation. More specifically, in the case of single equipment estimates, there is

a disproportionate weight given to the expected lifetime parameter, since it is used both to simulate from the Weibull distribution and in the Gamma prior. This dual use of the same parameter may amplify its influence on the posterior distributions, potentially leading to overconfidence in estimates when this particular input has high uncertainty.

Several promising directions exist for future research in this area. Development of more systematic processes for expert knowledge elicitation and validation would strengthen the model's foundation. The framework could be extended to incorporate other time-dependent covariates and environmental factors, allowing for more dynamic reliability assessments. Integration with condition monitoring systems would enable real-time prediction updates as equipment performance data is collected. Additionally, expanding the hierarchical levels to better capture interdependencies among equipment in complex systems would provide a more holistic view of facility reliability. Perhaps most importantly, continuous empirical validation through comparison of model predictions with operational outcomes as data becomes available will be essential for refining the model and establishing its long-term validity in practical applications.

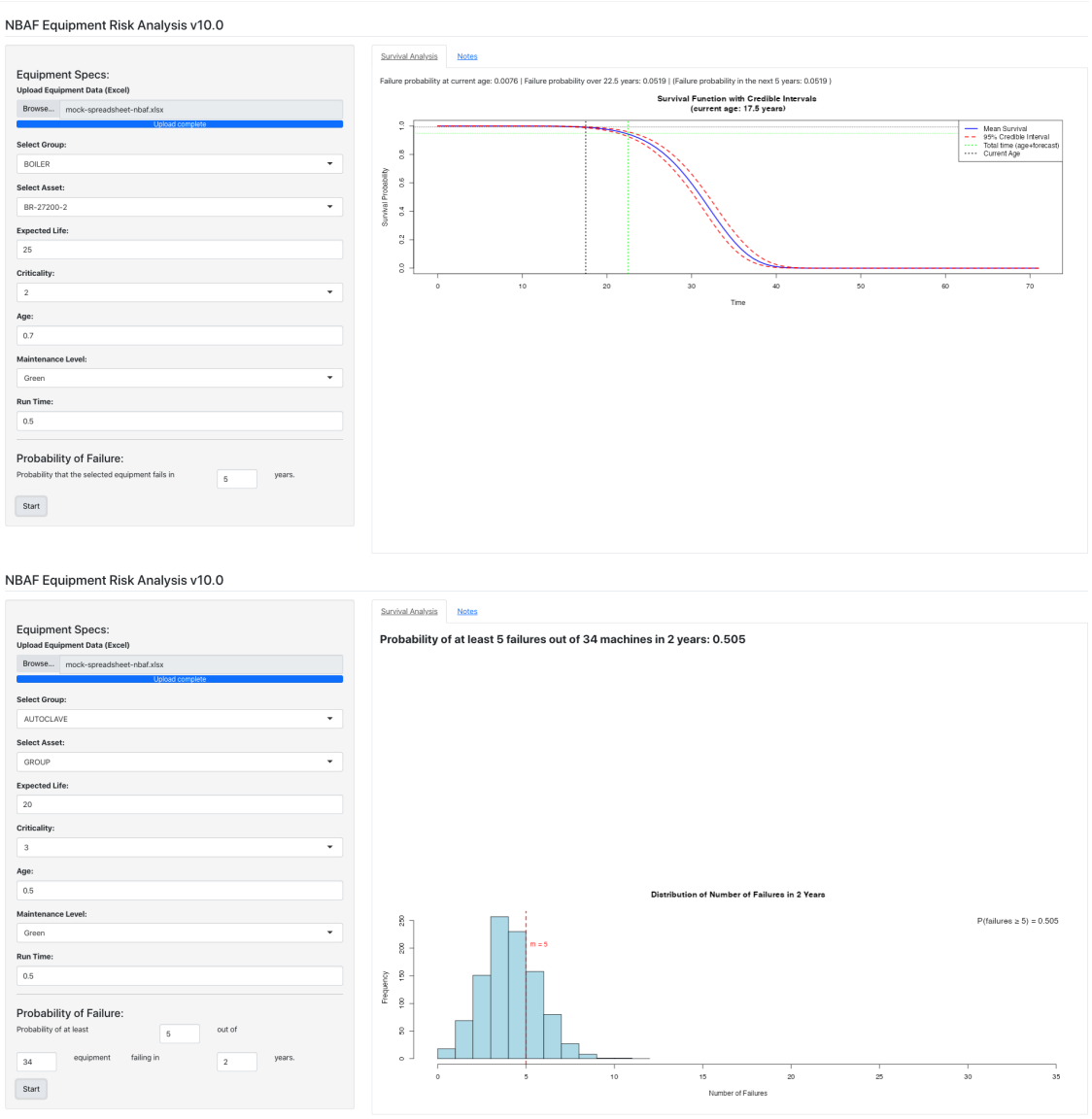

Figure 4.6: Interactive Shiny application developed for stakeholder validation of the Bayesian model. The application allows testing of different parameter combinations and immediate visualization of resulting survival curves (top panel) alongside probability estimates (bottom panel). This interactive tool facilitated intuitive assessment and fine-tuning of the model parameters through structured review sessions with domain experts, ensuring the model produced results aligned with initial expectations across diverse equipment types and operational conditions.

Bibliography

Evan Gatev, William N Goetzmann, and K Geert Rouwenhorst. Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies*, 19(3):797–827, 2006.

Marco Avellaneda and Jeong-Hyun Lee. Statistical arbitrage in the us equities market. *Quantitative Finance*, 10(7):761–782, 2010. doi: 10.1080/14697680903124632. URL https://doi.org/10.1080/14697680903124632.

Dominique M. Guillaume, Michel M. Dacorogna, Rakhal R. Davé, Ulrich A. Müller, Richard B. Olsen, and Olivier V. Pictet. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, 1(2):95–129, 1997. doi: 10.1007/s007800050018. URL https://doi.org/10.1007/s007800050018.

Thomas Lux and Michele Marchesi. Volatility clustering in financial markets: A microsimulation of interacting agents. *International Journal of Theoretical and Applied Finance (IJTAF)*, 03(04):675–702, 2000. URL https://EconPapers.repec.org/RePEc:wsi:ijtafx:v:03:y:2000:i:04:n:s0219024900000826.

Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001. doi: 10.1080/713665670. URL https://doi.org/10.1080/713665670.

Carol Alexander and Anca Dimitriu. The cointegration alpha: Enhanced index tracking and long-short equity market neutral strategies. *ISMA Finance Discussion Paper*, (08), 2002. doi: 10.2139/ssrn.315619. URL http://dx.doi.org/10.2139/ssrn.315619.

Joseph Engelberg, Pengjie Gao, and Ravi Jagannathan. An anatomy of pairs trading: the

role of idiosyncratic news, common information and liquidity. In *Third Singapore International Conference on Finance*, 2009.

Binh Do, Robert Faff, and Kais Hamza. A new approach to modeling and estimation for pairs trading. In *Proceedings of 2006 financial management association European conference*, volume 1, pages 87–99, 2006.

Binh Do and Robert Faff. Does naïve pairs trading still work. In *Working Paper*. Citeseer, 2008.

Heiko Jacobs and Martin Weber. On the determinants of pairs trading profitability. *Journal of Financial Markets*, 23:75–97, 2015. ISSN 1386-4181. doi: https://doi.org/10.1016/j.finmar.2014.12.001. URL https://www.sciencedirect.com/science/article/pii/S1386418114000809.

Sandro Andrade, Vadim Di Pietro, and M Seasholes. Understanding the profitability of pairs trading. *Unpublished working paper, UC Berkeley, Northwestern University*, 2005.

George Papadakis and Peter Wysocki. Pairs trading and accounting information. *Boston university and mit working paper*, 2007.

John Paul Broussard and Mika Vaihekoski. Profitability of pairs trading strategy in an illiquid market with multiple share classes. *Journal of International Financial Markets, Institutions and Money*, 22(5):1188–1201, 2012. doi: 10.1016/j.intfin.2012.06. URL https://ideas.repec.org/a/eee/intfin/v22y2012i5p1188-1201.html.

Rong Qi Liew and Yuan Wu. Pairs trading: A copula approach. *Journal of Derivatives & Hedge Funds*, 19:12–30, 2013. doi: 10.1057/jdhf.2013.1. URL https://doi.org/10.1057/jdhf.2013.1.

Yolanda Stander, Daniel Marais, and Ilse Botha. Trading strategies with copulas. *Journal of Economic and Financial Sciences*, 6(1):83–107, 2013. doi: 10.10520/EJC135921. URL https://journals.co.za/doi/abs/10.10520/EJC135921.

Hossein Rad, Rand Kwong Yew Low, and Robert W. Faff. The profitability of pairs trading strategies: Distance, cointegration, and copula methods. *Quantitative Finance*, 2016. doi: 10.1080/14697688.2016.1164337. URL https://ssrn.com/abstract=2614233.

Jing Wu. Pairs trading-copula vs cointegration, August 2018. URL https://www.quantconnect.com/learning/articles/investment-strategy-library/pairs-trading-copula-vs-cointegration. [Online; posted 2018-08-18].

Thomas S Ferguson. *Optimal Stopping and Applications*, 2007. URL https://www.math.ucla.edu/~tom/Stopping/Contents.html.

Stefan Rampertshammer. An ornstein-uhlenbeck framework for pairs trading. 2007. URL https://api.semanticscholar.org/CorpusID:117033128.

Li Qiang. Pair trading in optimal stopping theory. Master's thesis, Uppsala University, Uppsala, Sweden, 2009. URL https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-119421.

Tim Leung and Xin Li. Optimal mean reversion trading with transaction costs and stop-loss exit. *International Journal of Theoretical and Applied Finance*, 18(3), April 26 2015. URL http://dx.doi.org/10.2139/ssrn.2222196.

Sylvia Endres and Johannes Stübinger. Optimal trading strategies for lévy-driven ornstein-uhlenbeck processes. FAU Discussion Papers in Economics 17/2017, Friedrich-Alexander-Universität Erlangen-Nürnberg, Institute for Economics, Nürnberg, 2017. URL https://hdl.handle.net/10419/169117.

Da-Ming Zhu, Li Lin, Lei Li, and Gaoxiu Wang. Optimal pairs trading with dynamic mean-variance objective. *Mathematical Methods of Operations Research*, 94(1): 145–168, 2021. doi: 10.1007/s00186-021-00751-z. URL https://doi.org/10.1007/s00186-021-00751-z.

Jakub W. Jurek and Halla Yang. Dynamic portfolio selection in arbitrage. *EFA 2006 Meetings*, 2007. URL https://ssrn.com/abstract=882536.

Supakorn Mudchanatongsuk, James A Primbs, and Wilfred Wong. Optimal pairs trading: A stochastic control approach. In *2008 American control conference*, pages 1035–1039. IEEE, 2008.

Haipeng Xing. A singular stochastic control approach for optimal pairs trading with proportional transaction costs. *Journal of Risk and Financial Management*, 15(4), 2022. ISSN 1911-8074. doi: 10.3390/jrfm15040147. URL https://www.mdpi.com/1911-8074/15/4/147.

Ying Han and Wei Wei. Unsupervised learning for pairs trading: A clustering approach. *Quantitative Finance*, 23(1):45–62, 2023. doi: 10.1080/14697688.2022.2109876.

Christopher Krauss and Johannes Stübinger. Non-linear dependence modelling with bivariate copulas: statistical arbitrage pairs trading on the s&p 100. *Applied Economics*, 49(52): 5352–5369, 2017. doi: 10.1080/00036846.2017.1305097. URL https://doi.org/10.1080/00036846.2017.1305097.

Thomas Fischer and Christopher Krauss. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2): 654–669, 2018. doi: 10.1016/j.ejor.2017.11.054.

Saeid Fallahpour, Hasan Hakimian, Khalil Taheri, and Ehsan Ramezanifar. Pairs trading strategy optimization using the reinforcement learning method: a cointegration approach. *Soft Computing*, 20:5051–5066, 2016.

Andrew Brim. Deep reinforcement learning pairs trading. *All Graduate Plan B and other Reports, 1425*, 2019. URL https://digitalcommons.usu.edu/gradreports/1425.

Taewook Kim and Ha Young Kim. Optimizing the pairs-trading strategy using deep reinforcement learning with trading and stop-loss boundaries. *Complexity*, 2019(52):20, 2019. doi: 10.1155/2019/3582516. URL https://doi.org/10.1155/2019/3582516.

Cheng Wang, Patrik Sandås, and Peter Beling. Improving pairs trading strategies via reinforcement learning. In *2021 International Conference on Applied Artificial Intelligence (ICAPAI)*, pages 1–7, 2021. doi: 10.1109/ICAPAI49758.2021.9462067.

Yuyang Zhang, Stefan Zohren, and Stephen Roberts. Deep learning for multiple asset trading. *arXiv preprint arXiv:1907.06527*, 2019.

Marcos Lopez de Prado David H. Bailey, Jonathan M. Borwein and Qiji Jim Zhu. The probability of backtest overfitting. *Journal of Computational Finance*, 18(2):37–55, 2014. doi: 10.21314/JCF.2014.304.

Bryan Kelly Shihao Gu and Dacheng Xiu. Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273, 2020. doi: 10.1093/rfs/hhaa009.

Andrés L. Suárez-Cetrulo, David Quintana, and Alejandro Cervantes. Machine learning for financial prediction under regime change using technical analysis: A systematic review. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(2):67–79, 2023. doi: 10.9781/ijimai.2023.06.003. URL https://www.ijimai.org/journal/sites/default/files/2024-03/ip2023_06_003.pdf.

Carlos Eduardo de Moura, Adrian Pizzinga, and Jorge Zubelli and. A pairs trading strategy based on linear state space models and the kalman filter. *Quantitative Finance*, 16(10): 1559–1573, 2016. doi: 10.1080/14697688.2016.1164886. URL https://doi.org/10.1080/14697688.2016.1164886.

Burak Alparslan Eroğlu, Haluk Yener, and Taner Yiğit. Pairs trading with wavelet transform. *Quantitative Finance*, 23(7-8):1129–1154, 2023. doi: 10.1080/14697688.2023.2230249. URL https://doi.org/10.1080/14697688.2023.2230249.

Jing-You Lu, Hsu-Chao Lai, Wen-Yueh Shih, Yi-Feng Chen, Shen-Hang Huang, Hao-Han Chang, Jun-Zhe Wang, Jiun-Long Huang, and Tian-Shyr Dai. Structural break-aware pairs trading strategy using deep reinforcement learning. *The Journal of Supercomputing*, 78:3843–3882, 2022. doi: 10.1007/s11227-021-04013-x.

Ganapathy Vidyamurthy. *Pairs Trading: quantitative methods and analysis*, volume 217. John Wiley & Sons, 2004.

Robert F. Engle and C. W. J. Granger. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276, 1987.

Søren Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2):231–254, 1988. ISSN 0165-1889. doi: https://doi.org/10.1016/0165-1889(88)90041-3. URL https://www.sciencedirect.com/science/article/pii/0165188988900413.

Hiro Y. Toda and Taku Yamamoto. Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics*, 66(1-2):225–250, 1995.

Allan W Gregory and Bruce E Hansen. Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics*, 70(1):99–126, 1996.

Pentti Saikkonen and Helmut Lütkepohl. Testing for the cointegrating rank of a var process with structural shifts. *Journal of Business & Economic Statistics*, 18(4):451–464, 2000.

Anindya Banerjee, Juan J. Dolado, and Ricardo Mestre. Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis*, 19(3):267–283, 1998.

M. Hashem Pesaran, Yongcheol Shin, and Richard J. Smith. Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, 16(3):289–326, 2001.

Joakim Westerlund. Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics*, 69(6):709–748, 2007.

Peter Pedroni. Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, 61(S1):653–670, 1999.

Peter Pedroni. Panel cointegration: Asymptotic and finite sample properties of pooled time series tests with an application to the ppp hypothesis. *Econometric Theory*, 20(3):597–625, 2004.

C. Bracegirdle and David Barber. Bayesian conditional cointegration. In *29th International Conference on Machine Learning (ICML 2012)*, volume 29, 2012. URL http://web4.cs.ucl.ac.uk/staff/C.Bracegirdle/BayesianConditionalCointegration.php.

Peter C. B. Phillips and Pierre Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988. ISSN 00063444. URL http://www.jstor.org/stable/2336182.

Peter J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1:221–233, 1967.

Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980. doi: 10.2307/1912934.

Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987. doi: 10.2307/1913610.

Alvin C. Rencher and G. Bruce Schaalje. *Linear Models in Statistics*. John Wiley & Sons, Inc, USA, 2nd edition, 2007. ISBN 9780471754985.

Siddhartha Chib, Federico Nardari, and Neil Shephard. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341–371, 2006. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2005.06.026. URL https://www.sciencedirect.com/science/article/pii/S0304407605001478.

Omar Aguilar and Mike West. Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, 18(3):338–357, 2000. ISSN 07350015. URL http://www.jstor.org/stable/1392266.

M.B. Hooten and T. Hefley. *Bringing Bayesian Models to Life*. CRC Press, USA, 1st edition, 2019. ISBN 9780429243653.

Alan E. Gelfand, Adrian F. M. Smith, and Tai-Ming Lee. Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association*, 87(418):523–532, 1992. ISSN 01621459. URL http://www.jstor.org/stable/2290286.

William F. Sharpe. Mutual fund performance. *Journal of Business*, 39(1):119–138, 1966. doi: 10.1086/294846.

Frank A. Sortino and Robert van der Meer. Downside risk. *Journal of Portfolio Management*, 17(4):27–31, 1991. doi: 10.3905/jpm.1991.409343.

Yin-Wong Cheung and Kon S Lai. Finite-sample sizes of johansen's likelihood ratio tests for cointegration. *Oxford Bulletin of Economics and Statistics*, 55(3):313–328, 1993.

Jesús Otero and Jeremy Smith. Testing for cointegration: power versus frequency of observation—further monte carlo results. *Economics Letters*, 67(1):5–9, 2000.

Craig S Hakkio and Mark Rush. Cointegration: how short is the long run? *Journal of International Money and Finance*, 10(4):571–581, 1991.

Søren Johansen. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: Journal of the Econometric Society*, 59(6): 1551–1580, 1991.

Matthew Clegg. On the persistence of cointegration in pairs trading. *ERN: North America (Developed Markets) (Topic)*, 2014. URL https://api.semanticscholar.org/CorpusID:154111565.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series*

B (Methodological), 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x. URL https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Carlo Emilio Bonferroni. *Teoria statistica delle classi e calcolo delle probabilità*. Libreria internazionale Seeber, Firenze, Italy, 1936.

Winston Haynes. Bonferroni correction. In Werner Dubitzky, Olaf Wolkenhauer, Kwang-Hyun Cho, and Hiroki Yokota, editors, *Encyclopedia of Systems Biology*, pages 154–154. Springer New York, New York, NY, 2013. ISBN 978-1-4419-9863-7. doi: 10.1007/978-1-4419-9863-7_1213. URL https://doi.org/10.1007/978-1-4419-9863-7_1213.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

Bradley Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7 (1):1–26, 1979.

Hans R. Künsch. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17(3):1217–1241, 1989.

Regina Y. Liu and Kesar Singh. Moving blocks jackknife and bootstrap capture weak dependence. In Raoul LePage and Lynne Billard, editors, *Exploring the Limits of Bootstrap*, pages 225–248. John Wiley & Sons, New York, 1992.

Edward Carlstein. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 14(3):1171–1179, 1986.

Soumendra Nath Lahiri. *Resampling methods for dependent data*. Springer, New York, 2003. ISBN 978-0-387-00928-0. doi: 10.1007/978-0-387-21700-9.

Dimitris N. Politis and Joseph P. Romano. The circular block bootstrap. *Journal of the American Statistical Association*, 87(418):369–375, 1992.

A. V. Skorokhod. *Studies in the Theory of Random Processes*. Addison-Wesley, Reading, MA, 1965.

Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 2nd edition, 1999.

Bradley Efron. Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397):171–185, 1987. doi: 10.1080/01621459.1987.10478410.

R. Keith Mobley. *An Introduction to Predictive Maintenance*. Butterworth-Heinemann, Amsterdam, 2002.

Enrico Zio. Reliability engineering: Old problems and new challenges. *Reliability Engineering & System Safety*, 94(2):125–141, 2009.

Faisal I. Khan and Mahmoud M. Haddara. Risk-based maintenance (rbm): a quantitative approach for maintenance/inspection scheduling and planning. *Journal of Loss Prevention in the Process Industries*, 16(6):561–573, 2003.

Robert T. Clemen and Terence Reilly. *Making Hard Decisions with Decision Tools Suite*. Duxbury, 3rd edition, 2013. ISBN 978-0-538-79757-3.

Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? *Structural Safety*, 31(2):105–112, 2009.

George Apostolakis. The concept of probability in safety assessments of technological systems. *Science*, 250(4986):1359–1364, 1990.

Marvin Rausand and Arnljot Høyland. *System Reliability Theory: Models, Statistical Methods, and Applications*. John Wiley & Sons, Hoboken, NJ, 2004.

Hans J. Pasman, William J. Rogers, and M. Sam Mannan. Risk assessment: What is it worth? shall we just do away with it, or can it do a better job? *Safety Science*, 49(10): 1425–1432, 2011.

Edward L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

Wayne Nelson. Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4):945–966, 1972.

Odd Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4):701–726, 1978.

Jerald F. Lawless. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, Hoboken, NJ, 2nd edition, 2003.

Adrian E. Raftery. Bayesian model selection in social research. *Sociological Methodology*, 25: 111–163, 1995.

David R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–220, 1972.

David G. Kleinbaum and Mitchel Klein. *Survival Analysis: A Self-Learning Text*. Springer, New York, 2012.

D. Kumar and B. Klefsjö. Proportional hazards model: a review. *Reliability Engineering & System Safety*, 44(2):177–188, 2000.

Nozer D. Singpurwalla. *Reliability and Risk: A Bayesian Perspective*. John Wiley & Sons, Chichester, 2006.

Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL, 3 edition, 2013.

Enrique L. Droguett and Ali Mosleh. Integrated treatment of model and parameter uncertainties through a bayesian approach. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 227(1):41–54, 2012.

Harry F. Martz and Ray A. Waller. *Bayesian Reliability Analysis*. John Wiley & Sons, New York, 1996.

Dennis V. Lindley. The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):293–337, 2000.

Michael S. Hamada, Alyson G. Wilson, C. Shane Reese, and Harry F. Martz. *Bayesian Reliability*. Springer, New York, 2008.

Valen E. Johnson, Albert Moosman, and Peter Cotter. A hierarchical model for estimating the reliability of complex systems. *Bayesian Statistics*, 7:199–213, 2005.

Howard Raiffa and Robert Schlaifer. *Applied Statistical Decision Theory*. Harvard University Press, Cambridge, MA, 1968.

William Q. Meeker and Luis A. Escobar. *Statistical Methods for Reliability Data*. John Wiley & Sons, New York, 1998.

ReliaSoft. Life data analysis reference. *HBM Prenscia Inc.*, 2020.

Mohammed Ahmed Al Omari and Noor Akma Ibrahim. Bayesian survival estimator for weibull distribution with censored data. *Journal of Applied Sciences*, 11(5):393–396, 2011. doi: 10.3923/jas.2011.393.396. URL https://scialert.net/abstract/?doi=jas.2011.393.396.

Robert B. Abernethy. *The New Weibull Handbook*. Independent, 5 edition, 2006.

Walter R. Gilks, Sylvia Richardson, and David J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, London, 1996. ISBN 9780412055515.

Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. shiny: Web application framework for R. 2023. URL https://CRAN.R-project.org/package=shiny. R package version 1.7.5.