LUISS

Department of Economics and Finance
Chair of Empirical Finance

PAIR TRADING STRATEGY

IMPLEMENTATION OF PARTIAL COINTEGRATION MODEL ON ITALIAN STOCK MARKET

Candidate
Andrea Mangiucca

Supervisor
Prof. Giacomo Morelli

Co-Supervisor
Prof. Paolo Santucci de Magistris

ACADEMIC YEAR 2022-2023

Abstract

This thesis investigates the use of the Partial Cointegration Model (PCI) for pair trading strategies, focusing on the Italian stock market. It explores the theoretical basis of the PCI model and its practical application through simulation and empirical analysis. This study aims to demonstrate the PCI model's effectiveness in identifying statistical arbitrage opportunities, contributing to the quantitative finance field by offering insights into optimizing trading strategies in the context of a quantitatively evolving financial scenario.

Contents

Introduction

In the contemporary financial landscape, characterized by an increasing shift towards quantitative analysis, traditional trading methodologies are facing challenges in their efficacy, giving rise to the adoption of advanced econometric models. These models, particularly in the realm of pair trading, have shown promise in exploiting the relative mispricing between closely linked assets through statistical arbitrage. Our thesis aims to explore the application of the Partial Cointegration Model (PCI), as initially conceptualized by Clegg & Krauss, within the Italian market – a market potentially ripe with inefficiencies compared to its larger counterparts like the American. The intent is to investigate the PCI model's viability as a trading strategy, evaluating its profitability through practical application.

Our discussion begins with a comprehensive overview of pair trading concepts and the spectrum of approaches that can be adopted. Subsequently, we delve into the theoretical underpinnings of the PCI model, dissecting its logic and methodology for parameter estimation. The narrative progresses to outline the construction of the trading strategy, detailing the entry and exit rules for both the partial cointegration and the classical cointegration approach. Through rigorous simulation, the robustness of the PCI model is scrutinized, ensuring the reliability of estimates and the strategy's profitability. A comparative analysis of these methodologies' performance is conducted using real data from the most capitalized stocks in the Italian market across a defined period of three years, from 2021 to 2024.

In the concluding sections, the thesis will highlight the superior performance and profitability of the PCI model over traditional cointegration approach. Discussions will extend to the model's limitations and potential avenues for enhancing its efficacy. This comprehensive exploration aims not only to contribute to the academic discourse on quantitative finance but also to offer actionable insights for practitioners in the field.

1. Pair Trading Literature Review

This chapter is dedicated to an exhaustive review of the literature on pair trading, also referred to as statistical arbitrage. It commences with an examination of the fundamental principles and rationale underpinning the strategy before progressing to a detailed discussion of various methodologies. These range from the relatively straightforward, such as the well-known distance method discussed in Section 1.1, to the more complex, such as the Partial Cointegration model (PAR) and other approaches in the following sections.

1.1. *Pair Trading Fundamentals*

Pair trading is usually a market-neutral trading strategy that capitalizes on the relationship between two historically correlated securities. The fundamental idea behind is that two assets that exhibit a historical tendency to move together will continue to maintain this relationship over time. Practitioners using this scheme seek to identify pairs whose prices have diverged from their expected behaviour.

When such a divergence occurs, a trader would simultaneously buy one asset and sell the other, with the expectation that the prices will eventually revert to their historical relation. This convergence allows the dealer to profit from the narrowing of the spread between the two, independent of the market direction.

The rationale for this strategy is rooted in the belief that the prices of the pairs are mean-reverting; that is, when they deviate from their historical norm, they will tend to return to the mean, thanks to the forces of supply and demand, arbitrage, or other market mechanisms.

Investors rely on quantitative methods to identify couples and to determine the timing of the operations. Statistical measures such as correlation and standard deviation are key tools in assessing the strength and stability of the relationship between the couples. A successful arbitrage[1] does not require both the long and short positions to make a profit; instead, the strategy is deemed successful if for example the long position outpaces the short. This makes pair trading a popular strategy among hedge funds[2] and retail investors who seek to

1. Buying and selling an asset simultaneously in different markets to exploit price differences.
2. A private investment fund using varied and complex strategies, including leverage, to achieve

reduce market risk and generate returns that are not correlated with the broader market trends.

1.2. *Distance Method*

An examination is conducted of the method utilized in Gatev et al. (2006), a seminal work in the domain of this ambit. This study employs one of the most rudimentary quantitative techniques, which is predicated upon a distance metric. This section is devoted to analysing the tracking variance, which is defined as the average of the sum of squared deviations of normalized prices over a series of time intervals. To illustrate, considering two stocks, A and B, the tracking variance over a duration extending from time 1 to T is delineated by the following expression:

$$TV = \frac{1}{T}\sum_{t=1}^{T}(Q_{A_t} - Q_{B_t})^2.$$

Herein, Q_A and Q_B symbolize the normalized prices, where $Q_t = P_t/P_1$, with P denoting the price of the two securities at time t. The selection of stock is informed by this metric, opting for those with the minimal tracking variance. Upon the identification of pairs, the ensuing strategy is articulated:

a. monitor the divergence Δ midst the normalized prices of the pair, which will yield a temporal series of the spread;

b. establish a threshold, designated as 2σ in this context, with σ representing the standard deviation of the spread;

c. activate the trading mechanism upon the breach of this threshold: execute a short position on the stock with the elevated price and a long position on the stock with the diminished price;

d. liquidate the positions when the stock prices converge (the spread reverts to zero) or in the event of a divergence that prompts the activation of a stop-loss order.[3]

In the subsequent discourse, a detailed examination of the naive distance method will not be pursued. This decision is informed by findings from various studies, including those by Krauss et al. (2015), which have empirically demonstrated a decline in the profitability of such simplistic approaches over the years. The attenuation in efficacy is largely attributed to the increasing sophistication of market participants and the widespread adoption of more advanced statistical

high returns.

3. An order to close a trade position when it reaches a certain price, to limit potential losses.

methods in trading strategies, leading to a diminution of the opportunities that such naive approaches once exploited. Thus, our focus will shift towards more intricate methodologies that may offer sustainable profitability in the contemporary trading landscape.

1.3. *Classical Cointegration Approach*

A time series is classified as stationary when the underlying stochastic process is defined by constant parameters that are invariant through time. Stationary series are amenable to prediction and exhibit mean-reverting behaviour. Theoretical arbitrage opportunities arise when one can exploit the tendency of asset prices to revert to their mean by purchasing securities when their prices are below this mean and selling them when prices exceed it, thereby purportedly securing a risk-free[4] profit predicated on the premise of mean reversion.

Empirically, one may model a generic stock price time series as a random walk, which constitutes the simplest archetype of an integrated process of the first order, or $I(1)$. The inherent unpredictability of random walk[5] sequences underscores the milestone achievement of Engle & Granger (1987), whose seminal contributions have had profound implications for the development of statistical arbitrage strategies rooted in cointegration.

The linchpin of such strategies is the recognition of the non-stationary nature of assets prices. To be pertinent for cointegration analysis, a time series must comport with the characteristics of an integrated process of order $I(1)$. The aim is to discern couples that are cointegrated, indicating a persistent, long-term equilibrium relationship. This is formally expressed by the equation

$$\alpha x_t + \beta y_t = z_t,$$

where z_t is stationary, $I(0)$, ad x_t, y_t are the stock price time series that are individually non-stationary but together form a stationary linear combination. The investor's goal is to leverage short-term deviations from this equilibrium, with the cointegration coefficient serving as a guide to identify and capitalize on transient mispricing.

The identification of stationary pairs within this framework necessitates the implementation of statistical tests. These tests critically assess the cointegration level to ensure that any identified relationship is statistically robust. Successful testing leads to the selection of pair candidates suitable for a cointegration-based trading strategy.

4. The rate of return of an investment with zero risk, typically represented by the yield on government bonds like U.S. Treasury bills.
5. The Random Walk Model is a theory suggesting that stock market prices move unpredictably and are not influenced by past movements.

1.4. *Partial Cointegration Approach*

As previously mentioned in the introduction, each cointegrated or partially-cointegrated series is comprised of two distinct elements: a mean-reverting component, which is permanent, and a random walk component, which is temporary. A fundamental issue with classical cointegration is the assumption that any shock causing a deviation from the mean spread is temporary. However, this deviation can become permanent for various reasons, such as changes in interest rates or fundamental shifts in a company's stock, as discussed in the seminal works of Clegg & Krauss (2016). In such cases, a pair that appeared cointegrated in historical analysis may not exhibit the same behaviour in future scenarios.

The Partial Cointegration (PCI) model addresses this challenge by incorporating a stochastic element into the process, distinguishing it from the mean-reverting component to provide more reliable signals.

Although the aggregate spread is directly observable, the individual components, the mean-reverting and the random walk, are not, and must be estimated using specific procedures. The nuances of partial cointegration, along with its practical application, will be thoroughly examined in subsequent chapters. This section aims to provide the reader with the essential theoretical framework necessary for a comprehensive understanding of the entire thesis.

1.5. *Other Approaches*

In the realm of statistical arbitrage, a plethora of methodologies exists beyond the common cointegration approach, catering to a range of complexities and statistical foundations. Notable among these are the Copula procedure, which captures the dependence structure between asset pairs; the cointegrated Logistic Mixture Autoregressive (LMAR) model, which introduces logistic transition functions[6] to model the mean-reverting behaviour; the quasi-multivariate method that extends the univariate case to multiple assets; and the Support Vector Machines, which employs machine learning algorithms[7] for pattern recognition and predictions. While some of these techniques are intricate and others more straightforward, the intent of this paragraph is to acknowledge the diversity of strategies within pair trading. However, these alternative methods will not be dissected in detail within this thesis, as our focus is dedicated to an in-depth analysis of a stat arb strategy through the lens of the Partial Cointegration (PCI) model.

6. Functions in logistic regression modelling the probability of an outcome based on input variables.
7. Machine learning is a branch of artificial intelligence where computers learn from data to improve their performance on tasks without explicit programming.

2. Methodology

In this chapter a comprehensive dissection of the trading model will be undertaken from a theoretical standpoint. Section 2.1 will delve into the conceptual underpinnings of the framework, laying the groundwork for the subsequent empirical analysis.

Section 2.2 is designed to elucidate the methodologies employed in estimating the parameters that are critical to the model's functionality. Moving forward, Section 2.3 will be dedicated to validating the reliability of our estimates. This will be accomplished through a series of simulations designed to test their robustness under various market conditions.

Following this we will explore an advanced procedure known as the Kalman Filter. This powerful algorithm enables the separation of the mean-reverting from the random walk component of the residual series. Concluding the Methodology section base on the renewed paper *Pairs trading with partial cointegration* by Clegg & Krauss (2016), we will discuss the Akaike Information Criterion (AIC) test. The AIC test, along with other criteria, plays a pivotal role by aiding in the selection of the most statistically valid pairs.

2.1. *Representation*

The Partial Cointegration model represents a nuanced variation of traditional cointegration, which accommodates for the presence of both mean-reverting and random walk elements within a residual series. This framework is set forth in line with the principles established by Engle & Granger (1987).

Definition: The components of the vector X_t are said to exhibit partial cointegration of order d, b, denoted as $X_t \sim \mathrm{PCI}\,(d, b)$, if (i) each constituent of X_t is integrated of order d; and (ii) there exist a non-zero vector α such that $Z_t = a' X_t$ decomposed into a sum $Z_t = R_t + M_t$, where $R_t \sim I(d)$ and $M_t \sim I(d-b)$.

In this section, the focus is placed on the most elementary form of partial cointegration. This involves the examination of two price time series, denoted as $X_1 = (X_{1,t})_{t\in T}$ and $X_2 = (X_{2,t})_{t\in T}$, which are considered to be partially cointegrated if there exist parameters β, ρ, σ_M, σ_R and initial conditions m_0, r_0 that satisfy the specified model

$$X_{2,t} = \beta X_{1,t} + W_t.$$

Here, W_t represents a composite error term composed of

$$W_t = M_t + R_t,$$

where M_t and R_t are defined as

$$\begin{aligned} M_t &= \rho M_{t-1} + \epsilon_{M,t}, \quad \epsilon_{M,t} \sim N(0, \sigma_M^2), \\ R_t &= R_{t-1} + \epsilon_{R,t}, \qquad \epsilon_{R,t} \sim N(0, \sigma_R^2), \end{aligned}$$

with β being the hedge ratio parameter, $\rho \in (-1, 1)$ representing the $AR(1)$ coefficient, and $\epsilon_{M,t}$, $\epsilon_{R,t}$ being mutually independent Gaussian white noise processes with zero mean and respective variances $\sigma_M^2, \sigma_R^2 \in \mathbb{R}^+$.

To simplify the estimation of the model, we set the initial values $m_0 = 0$ and r_0 being the initial residuals defined as

$$r_0 = X_{2,0} - \beta X_{1,0}.$$

The time series X_2 and X_1 are thus interconnected by a partially autoregressive (PAR) model $W = (W_t)_{t \in T}$ first discussed by Summers (1986) and Poterba & Summers (1988), and subsequently detailed by Clegg (2015).

A central metric within the PAR model is the ratio of variance attributable to mean reversion, articulated as

$$R_{MR}^2 = \frac{\mathrm{Var}\big[(1-B)M_t\big]}{\mathrm{Var}\big[(1-B)W_t\big]} = \frac{2\sigma_M^2}{2\sigma_M^2 + (1+\rho^2)\sigma_R^2}, \quad R_{MR}^2 \in [0, 1], \tag{1}$$

where B signifies the backshift operator. If $R_{MR}^2 = 0$, the autoregressive component is nonexistent, thus rendering the series a pure random walk. Conversely, if $R_{MR}^2 = 1$, the random walk component is absent, and the series becomes fully autoregressive.

Given that the error term W_t is not directly observable, a state space representation becomes necessary for analysis. Brockwell & Davis (2010) lay the groundwork by providing an introductory overview of state space models. For a more extensive examination, Durbin & Koopman (2012) present a detailed discussion. The state space framework comprises two fundamentals: the observation equation and the state equation. These are conventionally expressed as

$$X_t = H_t Z_t + V_t \tag{2}$$

$$Z_t = F_t Z_{t-1} + G_t U_t + W_t. \tag{3}$$

The system under consideration is characterized by a state variable Z_t, as indicated in equation 3, which may elude direct observation. This state is presumed to evolve according to a linear dynamic, potentially subject to an exogenous input U_t. Accompanying the state is a stochastic term W_t, endowed with a covariance matrix Q_t, encapsulating the system's inherent noise. The part of the system that can be observed is signified by X_t, detailed in equation 2, which is reliant on the hidden state Z_t through a linear relationship captured by the matrix H_t, and is also affected by its own noise component V_t, with an associated covariance matrix R_t. It is posited that V_t is a zero-mean noise term and that the system is devoid of any external control input, denoted by U_t. Furthermore, the matrices H_t that describe the dependence on the hidden state and F_t that account for the state transition are considered static over time. Under these assumptions, the system's equations are simplified, leading to a more tractable form:

$$\begin{aligned} X_t &= HZ_t, \\ Z_t &= FZ_{t-1} + W_t. \end{aligned}$$

Within the Partial Cointegration model framework, we have two observable variables X_1 and X_2, as well as two hidden state variables M and R. For simplicity in representation, X_1 is treated as a third hidden state variable, meaning it is included in both the observation and state equations. Thus, the observation equation for the PCI system is articulated as

$$X_t = \begin{bmatrix} X_{2,t} \\ X_{1,t} \end{bmatrix} = HZ_t = \begin{bmatrix} \beta & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_{1,t} \\ M_t \\ R_t \end{bmatrix},$$

and the corresponding hidden state equation is

$$Z_t = \begin{bmatrix} X_{1,t} \\ M_t \\ R_t \end{bmatrix} = FZ_t + W_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ M_{t-1} \\ R_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{X,t} \\ \epsilon_{M,t} \\ \epsilon_{R,t} \end{bmatrix}.$$

In this model, $\epsilon_{X,t}$ signifies the stochastic variance component of the X_1 process in its first differences and is assumed to follow a normal distribution[1] with zero mean and variance σ_X^2, denoted as $\epsilon_{X,t} \sim N(0, \sigma_X^2)$. This term is considered to be statistically independent from the noise components $\epsilon_{M,t}$ and $\epsilon_{R,t}$.

2.2. *Parameters Estimations*

Referenced in Appendix A is a demonstration affirming the identifiability of the Partial Cointegration (PCI) model. This implies that for any given realization of a PCI system – potentially infinite in its manifestations – there exists a distinct set of parameter that are responsible for that specific occurrence.

Parameter estimation within the model is conducted via the maximum likelihood method [2], leveraging the capabilities of the associated Kalman filter. Given that the system's parameters are established and the innovations $\epsilon_{X,t}$, $\epsilon_{M,t}$ and

1. A symmetric, bell-shaped frequency distribution common in statistics.
2. A method for estimating parameters in a statistical model, maximizing the likelihood of observed data.

$\epsilon_{R,t}$ are independent, identically distributed Gaussian variables with zero mean, the Kalman filter operates to minimize the mean-squared error of the parameter estimates. It remains the optimal linear estimator even when the innovations are white, uncorrelated, and possess a non-Gaussian distribution, as asserted by Simon (2006).

Within this framework, Θ_t encapsulates all available information up to and including the time t, while Φ symbolizes the parameters vector β, ρ, σ_X, σ_R and σ_R. The Kalman filter defines the one-step ahead prediction error as

$$e_t = X_t - E[X_t|\Theta_{t-1}, \Phi].$$

For the PCI model, this prediction error can be expressed as

$$e_t = \begin{bmatrix} \beta e_{X,t} + \epsilon_{M,t} + \epsilon_{R,t} \\ \epsilon_{X,t} \end{bmatrix}.$$

Assuming that the joint probability $p(\beta e_{X,t} + \epsilon_{M,t} + \epsilon_{R,t}, \epsilon_{X,t})$ is equal to $p(\epsilon_{M,t} + \epsilon_{R,t}, \epsilon_{X,t})$, since the errors are uncorrelated, the likelihood function for the Kalman filter of the PCI model is delineated as

$$\mathcal{L}(\Phi) = p(X_1|\Phi) \prod_{k=2}^{n} p(\epsilon_{M,k} + \epsilon_{R,k}; 0, \sigma_M^2 + \sigma_R^2) \prod_{k=2}^{n} p(\epsilon_{X,k}; 0, \sigma_X^2),$$

where p denotes the probability density function of the normal distribution, and $p(X_1|\Phi)$ is a constant term corresponding to the initial observation. The optimization process focuses solely on the parameters β, ρ, σ_M and σ_R thereby disregarding the constant term. Consequently, the maximum likelihood estimates for β, ρ, σ_M and σ_R are obtained by maximizing the likelihood function

$$\mathcal{L}_{MR}(\beta, \rho, \sigma_M, \sigma_R) = \prod_{k=2}^{n} p(\epsilon_{M,k} + \epsilon_{R,k}; 0, \sigma_M^2 + \sigma_R^2).$$

This likelihood function is used as the objective function in the optimization algorithm. A full derivation of it is shown in Appendix B.

2.3. *Simulation of* PAR *Model*

TAB. 2.1

Parameter	Value
ρ	1
σ_M	0.5
σ_R	0.5
Sample size	300

In order to visualize the behavior of a Partially Autoregressive (PAR) model, we shall embark on a simulation exercise. The model parameters are set as follows: a value of 1is assigned to the autoregressive coefficient (ρ), while the standard deviations of the mean-reverting component (σ_M) and the random walk component (σ_R) are both set to 0.5. The sample size for this simulation is determined to be 300 data points as shown in table 2.1.

The generation process for the mean-reverting and random walk components will be conducted separately, adhering to the PAR model's stipulated structure. This bifurcated generation ensures that each component accurately reflects its theoretical properties – mean-reversion and random walk, respectively.

For the synthesis of these time series elements, the Statsmodels library in Python offers a suite of functions aptly suited for such stochastic simulations. Upon the creation of these individual series, they will be aggregated to yield a composite sequence. Below in figure 2.1, the plot of the simulation is presented to illustrate the outcome of this synthesis visually.

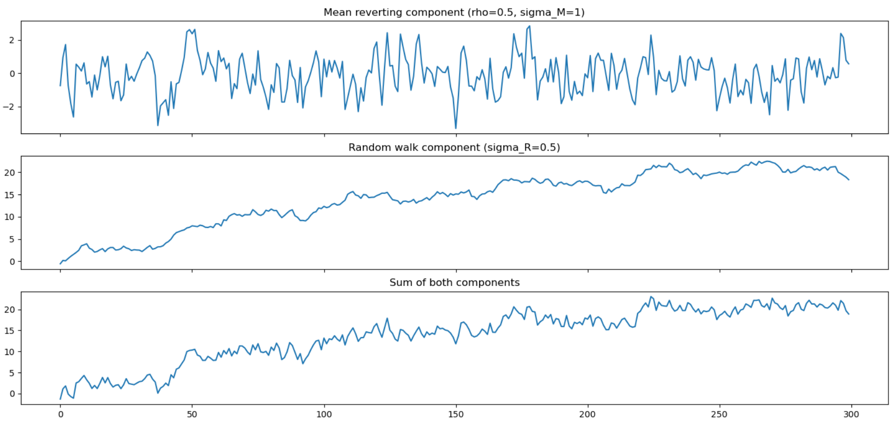


FIG. 2.1

2.3.1 Consistency of the Estimates

Utilizing a maximum likelihood estimation approach, our estimators should exhibit three key asymptotic properties – efficiency, lack of bias, and consistency in terms of mean squared error (MSE) – provided that the necessary regularity conditions are met. To validate these theoretical properties, we conduct analyses on an ensemble of synthetic data sets generated under the framework of Partial Cointegration (PCI) models.

In the reference work by Clegg and Krauss, a series of simulations are executed to explore the behavior of estimators across various levels of the mean-reverting variance σ_M and the autoregressive coefficient ρ. Their findings indicate that for every positive value of σ_M, the mean squared error (MSE) tends to zero, affirming the consistency of the MSE of the estimates. The scenario where σ_M is set to zero is an exception; here, the model reduces to a pure random walk, and consequently, the estimates become inconsistent due to the absence of variance in the mean-reverting component.

Similarly, for all values within the range $[0.6, 1)$ for ρ, the estimates remain consistent. It is only at the upper bound of this interval, where ρ =1, that the model's estimations falter. At this point, the aggregate residual of the series is effectively a combination of two random walk processes, leading to inconsistent estimates.

For the purpose of this thesis, rather than conducting multiple simulations, we will generate a single one adhering to the parameter values specified in table 2.2. This aims to verify whether the estimated parameters align with their true values. The primary focus of this study is not on the exhaustive verification of the model's statistical properties but rather on the practical application of the trading strategy and the analysis of its performance. By limiting the scope of simulation, we allocate more resources to comprehensively assess the efficacy of the trading strategy derived from the PCI model in real-world scenarios.

TAB. 2.2

β	σ_X	σ_M	σ_R	ρ	m_0	r_0
1.0000	0.0236	1.0000	1.0000	0.9000	0.0000	0.0000

To construct and plot a sample of partially cointegrated time series, first we produce X_1 as a geometric random walk[3] starting at the value of 100. Then we generate PAR residuals and calculate X_2 according to the model definition of the paper (table 2.2 and figure 2.2).

FIG. 2.2

3. A model where logarithms of successive price steps are independent and identically distributed, typically used in finance.

The parameter β[4] is set to unity to represent a standard equilibrium relationship between the paired series. The time series X_t, signifying price, is constructed as a cumulative return series commencing at a value of 100, with a mean of zero, and a standard deviation $\sigma_X = 0.0236$. This value mirrors the median volatility of all stocks that have been part of the S&P 500 index from 1990 until 2015. The standard deviations for the mean-reverting component σ_M and the random walk component σ_R are both set for simplicity to one. In this context, the absolute values of these deviations are less significant than their ratio, as this affects the variance portion attributed to mean-reversion as specified in equation 1. The autoregressive coefficient ρ is determined at 0.90 in the baseline case, equating to a mean-reversion half-life of approximately 6.60 days.

Our forthcoming step encompasses the deployment of a procedure for the estimation of the model's parameters. Initially, alpha and beta are estimated by employing an Ordinary Least Squares (OLS)[5] regression of X_2 on X_1. These preliminary estimations are then leveraged to ascertain initial values for the parameters of the PAR model using the lagged-variance method. These are refined by fitting to the residuals W_t. With these preliminary estimates established, the ensuing phase involves their concurrent optimization via a maximum likelihood approach.

Tab. 2.3

β	1.017 393 379 309 3202
σ_X	0.897 453 189 861 2403
σ_M	1.067 106 928 991 5014
σ_R	0.929 718 060 458 8074
$\ln L$	−1985.304 095 002 2935

In table 2.3 we observe the estimated values of the model's parameters. Notably, these estimates align closely with the true values, confirming the robustness of the model.

2.4. *Kalman Filter Application*

The Kalman filter is a recursive statistical algorithm that estimates the state of a linear dynamic system from a series of incomplete and noisy measurements. Theoretically, it operates under the premise that both the system and the measurement processes are governed by Gaussian noise, enabling the filter to produce estimates that are optimal in the sense of minimizing the mean of the squared error. In the context of our model, we employ the Kalman filter to disentangle the mean-reverting component from the stochastic trend, or random walk, within a time series. The extracted mean-reverting is then harnessed to generate trading signals, while the actual trading is executed on the overall

4. A measure in finance showing how an asset's price movements are related to the market or to other asset class. Here it is used to identify the weights of the two components of the couple.
5. A method for estimating unknown parameters in a linear regression model.

spread, which encompasses the total residual series.

In our case, upon obtaining the fitted values for parameters ρ, σ_M and σ_R, the Kalman gain, denoted by κ, is calculated. This gain can be precisely derived using a closed-form expression as indicated in Clegg (2015), or estimated via the iterative process.

Once κ is determined, the estimated values of M_t and R_t are computed utilizing the following recursive relationships

$$M_t = \rho M_{t-1} + \kappa E_t, \qquad R_t = R_{t+1} + (1-\kappa)E_t,$$
$$E_t = W_t - \rho M_{t-1} - R_{t-1}, \qquad W_t = P_t - \beta Q_t,$$

where E_t denotes the one-step prediction error from the Kalman filter. These equations are utilized within the formation period to generate an in-sample estimate of M_t and its standard deviation σ. In the subsequent trading period, the same equations are employed daily to update estimates of M_t.

2.5. *Properties of* AIC *test*

The Akaike Information Criterion (AIC), developed by Hirotsugu Akaike in 1974, serves as a pivotal tool for selecting statistical models by striking a delicate balance between effectively fitting data and minimizing model complexity. It quantifies the trade-off between these two critical aspects by estimating the information loss incurred when a model is utilized to represent the underlying data-generation process.

Calculation of the AIC involves considering the number of independent variables and the maximum likelihood estimate of its fit. The preferred model is one that attains the lowest AIC value, signifying its ability to elucidate a substantial portion of data variability using the fewest inputs.

The test is computed using the formula

$$\text{AIC} = 2K - 2\ln \mathcal{L}$$

Here, K denotes the count of independent variables, and $\mathcal{L}$, represents the likelihood function. The iterative progression of K begins at 2 and increases by 2 for each additional input incorporated.

When one model demonstrates an AIC value surpassing another one by more than two units, it is conventionally regarded as significantly superior. This discrepancy in AIC values implies a balance between goodness of fit and simplicity.

Our specific aim is to leverage this test to discern the most fitting model among PAR, AR, and RW for our dataset. Our goal involves pinpointing pairs exhibiting reduced AIC values relative to the PAR model. This approach will facilitate the selection of tradable pairs that manifest a superior fit concerning the PCI, potentially enhancing the identification of profitable trading opportunities.

3. Trading Strategy

The primary objective of this chapter is to examine the steps required to implement the trading strategy based on the methodologies described in the previous modules. Section 3.1 will delineate the division of data into training and testing phases. Additionally, this part will expound upon the methodology used to compute returns for individual paired trades as well as for the overall portfolio. Section 3.2 will discuss the rationale for selecting the tradable couples, finally Section 3.3 will address the generation of trading signals, including the establishment of trigger thresholds for both traditional and PCI-based strategies.

3.1. *Study Periods*

In this part, we aim to elucidate the rationale behind data partitioning and delineate the accurate methodology for computing Profit and Loss (P&L) within this specific arbitrage approach.

3.1.1 Splitting Data

Our analysis bifurcates the trading timeline into two distinct periods: the in-sample, or training period, and the out-of-sample, or testing period. During the in-sample period, we focus on calibrating our strategy by estimating the model's parameters, applying the selection criteria for tradable pairs, and conducting a preliminary back-test[1] to gauge profitability. The testing period then serves to validate the strategy's effectiveness in a real-world scenario, providing a robust measure of its potential profitability outside the confines of the training dataset. We use 2/3 of the dataset as train and the remain part of 1/3 as test.

3.1.2 Asset Return

The daily returns of the long and short positions within the pair are computed in the following manner, respectively, for a long and a short position:

$$r_{i,t} = sr_{i,t} \cdot w_{i,t}, \quad r_{i,t} = -sr_{i,t} \cdot w_{i,t},$$

1. That is, testing a trading strategy on past data to see how it would have performed.

where $r_{i,t}$ denotes the return of asset i at time t, factoring in reinvested payoffs, and $sr_{i,t}$ signifies the simple return of asset i, derived from the percentage change in prices between time t and $t-1$. Here, $w_{i,t}$ represents the cumulative wealth level of asset i at time t, calculated as

$$w_{i,t} = w_{i,t-1}(1 + r_{i,t-1}) = \prod_{t=1}^{t-1}(1 + r_{i,t}),$$

which is the product of wealth accumulation over time, considering the returns at each time step t.

3.1.3 Portfolio Returns

In our trading framework, which incorporates constraints on quantity, we cannot equally distribute the weight of assets y and x within a pair as was possible in Gatev et al. (2006), since the cointegration coefficient β is not always equal to one. Consequently, the allocation of weights to each asset in the pair at any given time t is determined by their respective market prices. Moreover, it's crucial to adjust the computed return by the number of pairs to ensure proportionality. Specifically, the weights of assets y and x in the pair are governed by the following equations

$$q_{y,t} = \frac{p_{y,t}}{p_{y,t} + \beta\, p_{x,t}} \frac{1}{n},$$

$$q_{x,t} = \frac{\beta\, p_{x,t}}{p_{y,t} + \beta\, p_{x,t}} \frac{1}{n} = (1 - q_y)\frac{1}{n}.$$

Here, $p_{y,t}$ and $p_{x,t}$ signify the market prices of y and x assets at time t, and β represents the cointegration coefficient, and n is the count of pairs. These weights remain static for the duration that the position is open but are recalculated to reflect market price shifts whenever the position is closed and subsequently reopened.

To calculate the intrapair returns, we apply the weights q_t to the individual asset return at each time t:

$$rp_{j,t} = \frac{q_{j,t}\, sr_{j,t}\, w_{j,t}}{\sum_{j \in P} q_{j,t}\, w_{j,t}}.$$

The total return of the portfolio at time t is thus formulated as

$$rp_t = \sum_{j \in P} rp_{j,t}.$$

Each long position on a spread involves purchasing a quantity (x) of the second asset (x_2) and selling a calculated quantity (βx) of the first asset (x_1). Conversely, initiating a short position on the spread entails selling a quantity (x) of the second asset and buying a calculated quantity (βx) of the first asset.

3.2. *Pairs Selection*

This section delineates the selection criteria utilized for pinpointing the most advantageous stock pairs on which to apply our trading strategy.

We initiate our selection by aggregating a set of potential pairs from the universe of available stocks, constraining our choices to those within the same Global Industry Classification Standard (GICS) sector.[2] This ensures computational efficiency and curtails the incidence of identifying couples with insignificant correlations, in alignment with the methodologies established by Gatev et al. (2006) and Do & Faff (2010, 2012).

Within the Cointegration strategy, a pair's eligibility is contingent upon it not substantiating the null hypothesis of no cointegration at a 5% significance level, as evidenced by both the Augmented Dickey-Fuller (ADF) and Johansen cointegration tests. This approach is corroborated by Huck (2015) and Rad et al. (2015), and operationalized by Caldeira & Moura (2013). The selection for the Partial Cointegration (PCI) strategy incorporates distinct criteria: a pair is considered suitable if the AIC for the PAR is lower than that of the competing AR and random walk models, if the mean reversion coefficient ρ exceeds 0.5, and if the proportion of variance due to mean reversion R^2 is greater than 0.5. These conditions exclude pairs that mean revert too quickly, thereby reducing the influence of the bid-ask spread on trading gains and ensuring robustness in parameter estimation.

An additional criterion for both strategies is a β value for the couples that is close to 1, specifically between 0.75 and 1.25, to maintain market-neutral exposure.

In conclusion, we conduct an in-sample back-test. Each pair is assigned an in-sample Sharpe Ratio following the methodologies of Dunis et al. (2010), Bertram (2010), and Caldeira & Moura (2013), with pairs organised based on this metric. The top-ranked are then selected as tradable couples. This process culminates in the creation of a portfolio comprising the top twenty stocks and their corresponding partners.

3.3. *Trading Signals*

This segment revisits the comprehensive steps undertaken for parameter estimation and the construction of a trading strategy underpinned by econometric modelling. Additionally, it will elaborate on the methodology for generating entry and exit signals or trigger levels.

2. A classification system defining sectors and industries for stocks.

3.3.1 Cointegration-based Pairs Trading

We embark on a methodical approach by initially generating all potential pairs from a chosen universe of stocks. This extensive collection forms the basis of our analysis. We then proceed by dividing the dataset into two distinct segments, allocating two-thirds for in-sample analysis and one-third for out-of-sample testing. The in-sample data will undergo the steps detailed below, while the out-of-sample data is reserved for the final back-testing phase.

The next critical step involves parameter estimation. Here, we employ the Ordinary Least Squares (OLS) regression to determine the key parameters: α, which represents the constant term, and β, denoting the hedge ratio. This is achieved by regressing the second price of the pair against the first.

Then, we apply two stationarity tests, namely the Augmented Dickey-Fuller (ADF) test on the regression residuals and the Johansen test on the price series. Only the pairs that successfully reject the null hypothesis of non-stationarity at a 95% confidence level are considered further.

Upon identifying the viable pairs, we proceed to spread construction. This involves combining the prices of the two stocks and the zero mean residuals W_t, creating what are known as synthetic spreads. These spreads form the basis of our trading signals.

$$W_t = X_{2,t} - \alpha - \beta X_{1,t}$$

Signal generation is a crucial step where we normalize the spread of the residual series. We calculate the mean and standard deviation of Wt and then determine the Z-score[3] using the formula

$$Z_t = \frac{W_t - \text{mean}}{\sigma}.$$

Finally, the strategy undergoes a test in the form of out-of-sample back-testing.

For the Cointegration strategy, the thresholds for initiating and closing trades are set ($\tau_0 = \pm 2$ and $\tau_0 = 0$), respectively, following Huck (2015) and Rad et al. (2015). This strategy does not incorporate a stop-loss mechanism, which aligns with the recommendations of Nath (2003) and Caldeira & Moura (2013) and reflects the evolution of pairs trading profitability as analyzed by Jacobs & Weber (2015).

3.3.2 PCI-based Pairs Trading

In the methodology for PCI-based pairs trading, the initial stages mirror those of the standard cointegration strategy, involving the generation of stock pairs and the division of the dataset into segments for in-sample estimation and

3. Z-score is a statistical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of standard deviations from the mean.

out-of-sample validation.

The estimation of parameters is a multi-step process where we determine the constituents of the PAR model, including the vector $[\rho, \sigma_M, \sigma_R]$, as well as the cointegration parameters α and β. Given the potential complexity of parameter estimation, it is prudent to synthesize the process into discrete stages, commencing with an initial guess of the parameters and culminating in an optimization via maximum likelihood. The model is articulated as follows:

$$\begin{aligned} X_{2,t} &= \alpha + \beta X_{1,t} + W_t, & W_t &= M_t + R_t, \\ M_t &= \rho M_{t-1} + \epsilon_{M,t}, & \epsilon_{M,t} &\sim N(0, \sigma_M^2), \\ R_t &= R_{t-1} + \epsilon_{R,t}, & \epsilon_{R,t} &\sim N(0, \sigma_R^2), \end{aligned}$$

Initially, β is estimated by regressing the price of stock X_2 on that of X_1. The intercept α is then obtained by adjusting the first observation of X_2 with the product of the first observation of X_1 and the estimated β:

$$\alpha = X_{2,0} - \beta X_{1,0}.$$

With α and β established, we construct the residual series W_t, which represents the synthetic spread, by inverting the cointegration equation:

$$X_{2,t} = \alpha + \beta X_{1,t} + W_t.$$

Having computed the synthetic spread, we proceed to estimate the parameters $[\rho, \sigma_M, \sigma_R]$ using an initial estimate derived from the Lagged Variance method. This method is encapsulated by the following formulas:

$$\begin{aligned} \omega_k &= \mathrm{Var}\big[(1-B)^k X_t\big], & \rho &= \frac{\omega_1 - 2\omega_2 + \omega_3}{2(\omega_1 - \omega_2)}, \\ \sigma_M^2 &= \frac{1}{2}\frac{\rho+1}{\rho-1}(\omega_2 - 2\omega 1), & \sigma_R^2 &= \frac{1}{2}(\omega_2 - \sigma_M^2), \end{aligned}$$

where B denotes the backshift operator.

Following the preliminary estimation, we engage in the optimization of $[\alpha, \beta, \rho, \sigma_M, \sigma_R]$ through maximization of the log-likelihood function. This optimization is iteratively performed ten times to ensure robustness – initially using the parameters extrapolated via Lagged Variance and subsequently employing random values.

Regarding the mean-reverting component M_t and the random walk component R_t, both critical yet unobservable elements that must be inferred through estimation. The Kalman filter serves as the inferential tool, applying the previously optimized parameters: the mean reversion coefficient ρ, the standard deviation of the mean-reverting process σ_M, and the standard deviation of the random walk σ_R. The Kalman gain κ, a pivotal factor in updating the estimations, can be ascertained either by a closed-form solution or through iterative approximations provided by the Kalman filter methodology as discussed in

Section 2.4.

Upon entering the trading phase, identical computational procedures are executed on a daily basis to refresh M_t estimates. A normalized Z-score, given by $Z_t = M_t/\sigma$, is then crafted to formulate trading signals. The thresholds for initiating ($\tau_0 = \pm 1$) and terminating ($\tau_0 = \pm 0.5$) trades have been systematically optimized by conducting multiple simulations on synthetic data as explicated by Clegg & Kraus. A salient aspect of the PCI strategy, distinguishing it from its cointegration-only counterpart, is the inherent risk of loss stemming from the random walk component – a risk absent when trades are concluded before the end of the trading interval in a strictly cointegration-based model, where profits are typically assured.

4. Trading on Simulated Data

Within this section, the focus shifts to the execution of the previously delineated steps, aiming to back-test the PCI strategy across various simulated datasets. This approach is undertaken to conduct a thorough validity analysis, leveraging simulated data to subject the method to a broad spectrum of hypothetical market scenarios. Such testing is instrumental in assessing the strategy's durability and its ability to withstand market fluctuations that some historical data may not capture.

Additionally, the empirical trials are designed to critically evaluate the performance differentials exhibited by the PCI in comparison to the traditional cointegration model. The intention is to determine the relative effectiveness of the PCI approach when confronted with data characterized by partial cointegration. By back-testing against simulated scenarios, we can gather empirical evidence on the strategy's operational soundness and its capacity to generate favourable outcomes.

4.1. *Parameters Calibration*

In the referenced study by Clegg and Krauss, a comprehensive analysis is conducted through a series of simulations to evaluate the impact of adjusting certain parameters, such as the mean reversion coefficient (ρ) and the proportion of variance due to mean reversion (R^2_{MR}), on the profitability and effectiveness of the strategies. This critical examination serves to underpin the selection criteria for pairs as articulated in Section 3.2. The findings lay the groundwork for the subsequent examinations, which are not aimed at exhaustive parameter optimization but rather at affirming the strategy's performance across different simulated market scenarios. The ultimate goal is to apply these well-founded thresholds to real-world market data in the forthcoming procedures.

Both cointegration and Partial Cointegration (PCI) strategies exhibit limitations in generating substantial returns when the random walk component dominates the variance of the pair ($R^2_{MR} < 0.5$). However, as the mean-reverting component gains prominence, returns escalate across all values of ρ, suggesting that a strong mean-reverting dynamic is conducive to enhanced trading outcomes.

The simulations further reveal that the PCI strategy outperforms classical cointegration when $\rho < 0.9$, losing its edge as ρ escalates towards one, where the $AR(1)$ process verges on becoming a random walk. As the R^2_{MR} metric approaches unity, the PCI and traditional cointegration methodologies converge, diminishing the differential benefits and aligning the strategies more closely.

The study emphasizes that variations in the first differences of the spread process's variance only affect the absolute magnitude of returns without altering the relative strategic advantage between the methodologies. Similarly, the spread W_t remains unaffected by changes in the standard deviation of the series σ_x, reaffirming the strategies' robustness against different levels of underlying volatility.

The investigation concludes that only pairs with a pronounced mean-reverting trait are suitable for trading. The PCI strategy is particularly effective for intermediate values of ρ and tends to converge with the performance of the cointegration-based strategy as either ρ or R^2_{MR} increase.

Clegg and Kraus show that as the opening threshold becomes grater, the monthly mean return tends to decline. This inverse relationship between opening thresholds and mean returns can be attributed to the trade-off between the frequency and the profitability per trade. A higher threshold may result in fewer trades but potentially higher returns, and vice versa. Notably, the difference in average profit between a moderate threshold ($\tau_0 = \pm 1.0$) and a high threshold ($\tau_0 = \pm 1.5$) is exacerbated when accounting for the amplified transaction costs associated with fewer number of trades. For pragmatic trading implementations, an opening threshold (τ_0) of ± 1.0 and a closing threshold (τ_c) respectively of ± 0.5 strikes a balance between the frequency of trading opportunities and the impact of transaction costs.

4.2. *Cointegration Results*

The commencement of the back-test and parameters calibration phase involves generating a simulated instance of two partially cointegrated time series, X_1 and X_2, replicating the method outlined in Section 2.3. This simulation spans 1000 days (figure 4.1), designated as the training period, followed by a 125 observations trading period, as shown in figure 4.2. The ordinary least squares (OLS) regressor is applied during the training period to ascertain the parameters alpha and beta, which are used in the computation of the spread.

TAB. 4.1

	Total return	APR	Sharpe	Max DD	Max DD duration
CI algo	−0.001613	−0.003248	−0.010944	−0.059508	49.0

FIG. 4.1

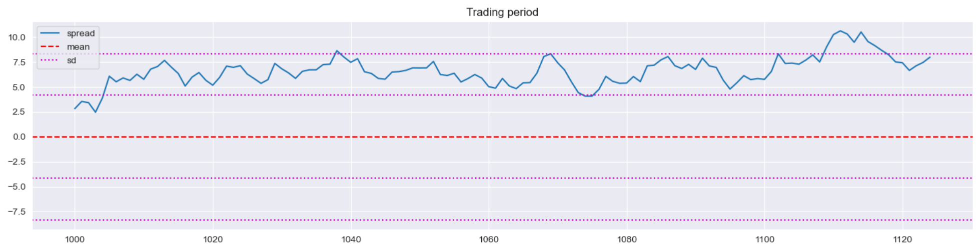


FIG. 4.2

Subsequent to the generation of the time series, the Johansen and ADF tests are applied to both series to verify stationarity, thus confirming the viability of a cointegration strategy. With the stationary relationship validated, the spread is then computed and charted for the subsequent trading period.

From the graphical representation, it becomes evident that the conventional cointegration-based trading strategy does not yield profitable outcomes within this simulation. The spread fails to revert to its historical mean, indicating a divergence rather than the anticipated mean reversion. Nevertheless, for the sake of thoroughness, the strategy is back-tested. The thresholds for trade initiation and termination are established following the methodology of Huck (2015) and Rad et al. (2015), with open trades triggered at a Z-score of ± 2 and closed at a Z-score of 0.

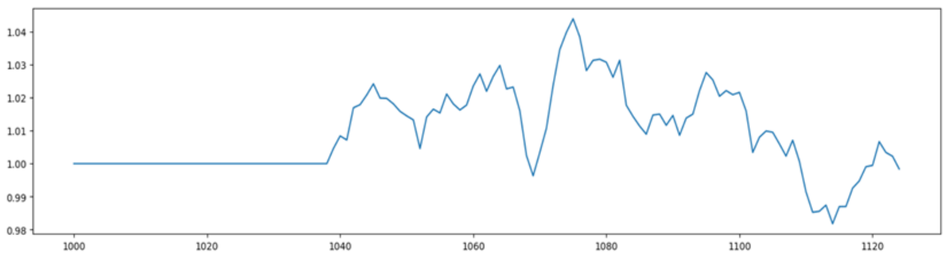


FIG. 4.3

The performance analysis, illustrated in the graph plotting the historical cumulative returns (figure 4.3), corroborates the initial assessment. It reveals that the cointegration-based strategy does not exhibit profitability in this simulated environment, as the spread does not demonstrate the expected mean-reverting behaviour within the trading period.

4.3. PCI *Results*

The approach to back-testing the Partial Cointegration (PCI) based trading strategy begins with parameter estimation for the model on historical data to establish the spread. The Kalman filter is then applied to tease out the mean-reverting component, which is crucial for generating trading signals. The trading rules are constructed around the Z-score of the mean-reverting part, a methodology that was detailed in the previous section on parameter calibration.

Figure 4.4 illustrates the residual series W_t over the training period (shown in blue) and the trading period (depicted in yellow), providing a visual representation of the model's behavior over time.

Following the application of the Kalman filter, the mean-reverting component along with its standard deviation (σ_M) is plotted (figure 4.5), as indicated by the dotted line in the second graph. This visualization is particularly informative as it highlights the thresholds at which the trading signals for entering trades will be triggered, based on the standardized deviation of the mean-reverting component.

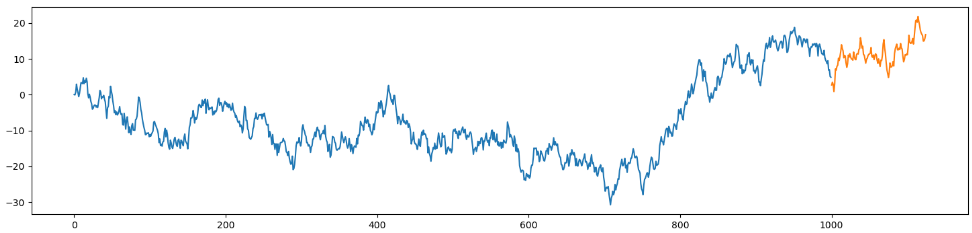


FIG. 4.4

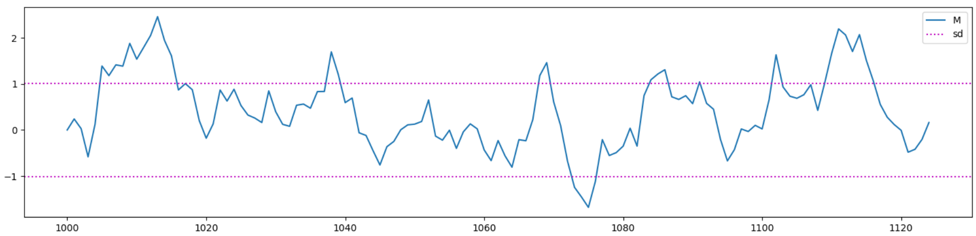


FIG. 4.5

The backtesting of the strategy is conducted under the following rules:

- Long position is opened when it is below $-\sigma_M$;
- Long position is closed when it is above $0.5\,\sigma_M$;

- Short position is opened when it is above σ_M
- Short position is closed when it is below $-0.5\,\sigma_M$.

The displayed results (table 4.2) illustrate the comparative performance of a trading strategy predicated on partial method against a traditional cointegration, while figure 4.6 represents the P&L chart of the PCI approach. The analysis reveals a significant outperformance by the partial, as evidenced by the positive returns over a span of 120 observations. The strategy not only yielded positive returns but also exhibited a Sharpe Ratio[1] exceeding 1, indicative of a robust risk-adjusted despite the relatively modest absolute returns.

TAB. 4.2

	Total return	APR	Sharpe	Max DD	Max DD duration
CI algo	−0.001 613	−0.003 248	−0.010 944	−0.059 508	49.0
PCI algo	0.037 885	0.077 486	1.099 020	−0.031 999	65.0

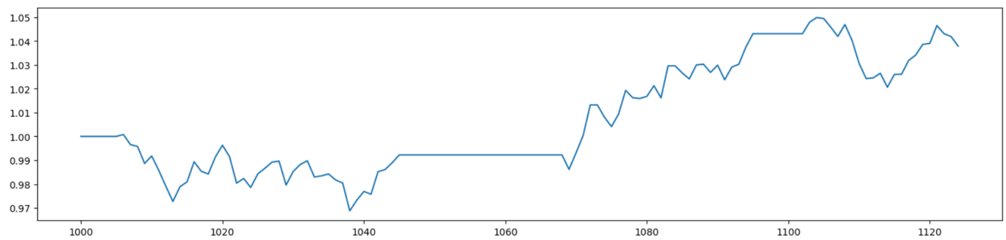


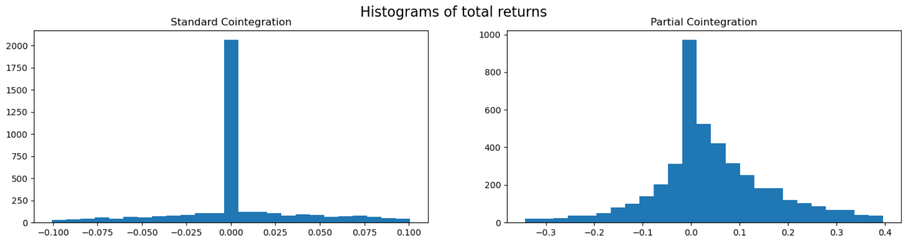
FIG. 4.6

4.4. *Multiple Simulations*

While the results are promising, they should be interpreted with caution, as a single data sample may not fully capture the variability and potential risks inherent in live market conditions. To ensure a comprehensive evaluation of the trading strategy, we will extend our analysis beyond a single data set and conduct 5000 simulations. Each simulation will generate a pair of partially cointegrated time series, reflecting the same underlying parameters. Upon these simulated data sets, we will back-test both the classical cointegration and the partial cointegration strategies, recording their respective performance metrics.

The histograms presented (figure 4.7a and figure 4.7b) offer a visual com-

1. A measure of risk-adjusted return, comparing the excess return of an investment to its volatility. Sharpe Ratio = $(R_p - R_f)/\sigma_p$, where R_p is the return of the portfolio, R_f is the risk-free rate, and σ_p is the standard deviation (volatility) of the portfolio's excess return.



a b

Fig. 4.7

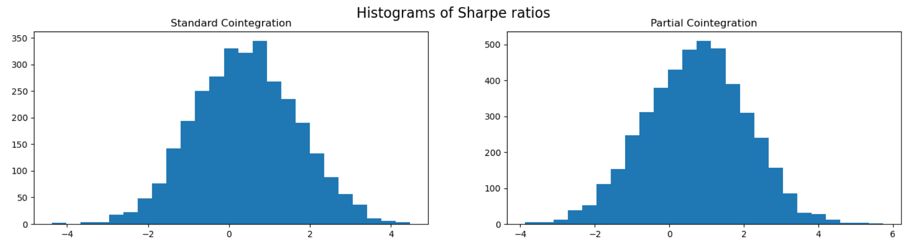


Fig. 4.8

parison of the distributions of total returns and Sharpe ratios for standard and partial cointegration. The returns distribution for the first one is centred around zero, indicating an equal likelihood of positive and negative outcomes. In contrast, the second one demonstrates a positive skewness, suggesting a higher prevalence of profitable trades compared to losses.

Furthermore, the distribution of Sharpe ratios, showcases a marked disparity between the two strategies. The PCI not only achieves a higher average Sharpe, which approximates 1, but also shows a distribution that is decisively more favourable than that of the classical method, whose average is closer to zero. This implies that the PC provides a more consistent and superior risk-adjusted performance relative to its counterpart.

5. Empirical Application on Real Data

In this final part of our study, we will put theory into practice by applying our methods in a live market environment.

Section 5.1 will outline the dataset we have selected for applying our strategies, detailing the reasons for focusing on that particular market segment.

Section 5.2 will evaluate the effectiveness of cointegration against the PCI methodology. We will demonstrate that, for the chosen dataset and specific historical timeframe, our Partial Cointegrated approach can yield superior risk-adjusted returns when compared to standard benchmarks. The essence of our work is captured in this segment, as it will confirm whether our sophisticated econometric model can actually be utilized to generate profits in real-world.

In Section 5.3, we will examine and calculate two risk management measures, specifically historical and parametric Value-at-Risk (VaR). These metrics will help us understand the necessary capital requirements to manage and absorb potential daily losses.

Finally, Section 5.4 will address additional considerations for refining and enhancing the entire process to maximize performance outcomes.

5.1. *Data Preparation*

Our pairs trading strategy is designed to test the effectiveness of statistical arbitrage on the Italian Stock Exchange from October 27, 2015, to December 20, 2023, with data provided by Bloomberg.[1] We have chosen the Italian market for its potential to unveil inefficiencies and opportunities that might not be as evident in larger markets like the American one, and which might be overlooked by simpler models.

Using Python and Microsoft Excel, we have deployed two strategies across the top 40 stocks of the FTSE MIB Index. The leading stock index of Borsa Italiana, undergoes a quarterly review. It reflecting the most actively traded stock classes, accounts for about 80% of the total Italian market capitalization.

With each review, it may be updated to include or exclude stocks based on criteria such as market capitalization and liquidity. For our study, we have used the stock composition as per the most recent update from Borsa Italiana's official

1. A global company providing financial software tools, data, and other financial services.

website[2] (21 December 2023).

Two-thirds of the period is dedicated to in-sample analysis for estimating parameters and extrapolating time series for the mean-reverting portion, which spans from October 27, 2015, to March 21, 2021. The remaining one-third of the period, from March 22, 2021, to December 20, 2023, is allocated for out-of-sample back-testing to evaluate the strategies' performance.

5.2. *Standard vs* PCI *Performance Analysis*

5.2.1 Strategy 1

Our analysis begins with the traditional method of cointegration. After running a linear regressor across all potential stock pair combinations to estimate alpha and beta parameters for spread construction, we involve a two-step verification process that incorporates both the Augmented Dickey-Fuller (ADF) test and the Johansen test on the residual series of each pairs. A stock couple is deemed suitable for trading only if it successfully passes the ADF test by rejecting the hypothesis of non-stationarity at a 95% confidence level and similarly surpasses the Johansen test's threshold by dismissing the null hypothesis of non-cointegration at a 90% level. This adjustment to a lower confidence level for the Johansen test is a pragmatic response to the limited number of pairs typically rejecting the null hypothesis at the more conventional 95% threshold.

An additional layer of selection is based on the beta values obtained from the regression. To ensure a market-neutral position only pairs with a β within the range of 0.75 to 1.25 are chosen. This range ensures that the capital allocated to each stock in a pair is balanced, maintaining the desired market neutrality. After applying this additional condition we are left with the following pairs (table 5.1).

TAB. 5.1

Stock Pairs	α	β	ADF	Johansen
G-UCG	1.173 650	0.809 321	1.0	1.0
SRG-BMED	2.529 798	0.979 232	1.0	1.0
MB-CNHI	0.181 598	0.901 064	1.0	1.0
AZM-BGN	8.440 925	1.024 048	1.0	1.0
LDO-UCG	2.660 296	1.034 159	1.0	1.0
ENI-SPM	−1.787 521	0.880 781	1.0	1.0
ENI-UCG	0.825 725	0.936 846	1.0	1.0

2. https://www.borsaitaliana.it/homepage/homepage.htm.

The next steps involve running an in-sample back-test on the Z-score of each synthetic spread series, then ranking the pairs in descending order based on their Sharpe Ratio (SR). Finally, only pairs with an SR greater than zero are selected, under the premise that pairs unprofitable in-sample are unlikely to be profitable out-sample. As a result, we have the following seven pairs (table 5.2).

TAB. 5.2

Stock Pairs	ENI-UCG	ENI-SPM	G-UCG	SRG-BMED	LDO-UCG	MB-CNHI	AZM-BGN
Sharpe	0.804804	0.791131	0.718494	0.710586	0.668573	0.629183	0.580099

TAB. 5.3

Stock Pairs	ENI-UCG	ENI-SPM	G-UCG	SRG-BMED	LDO-UCG	MB-CNHI	AZM-BGN
Returns	0.955632	0.364960	1.884446	1.185734	1.118688	1.120573	1.533610

In our concluding analysis, we back-tested our strategy on a test dataset to assess its performance.

Of the seven pairs tested, two recorded losses, with one pair experiencing a substantial 70% decline. However, the losses from these underperforming pairs were fully offset and surpassed by gains from the remaining five, particularly a pair yielding over 80% profit, as shown in figure 5.1 and table 5.3.

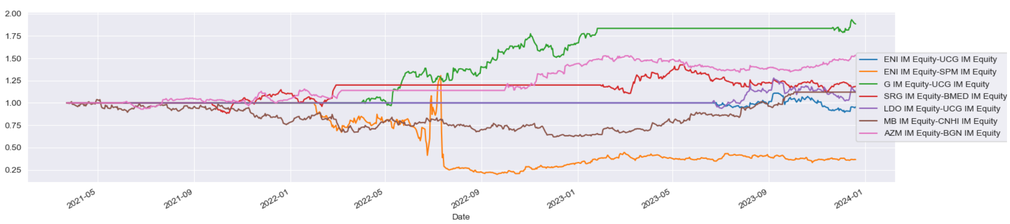


FIG. 5.1

Despite the mixed individual pair performances, the overall portfolio, with equal capital allocation to each pair, concluded with a total cumulative return of 23.25%. This translates to a respectable expected yearly rate of 7.22%, suggesting potential attractiveness. However, the strategy was largely dormant for the initial part of the testing period, with significant activity and gains materializing only

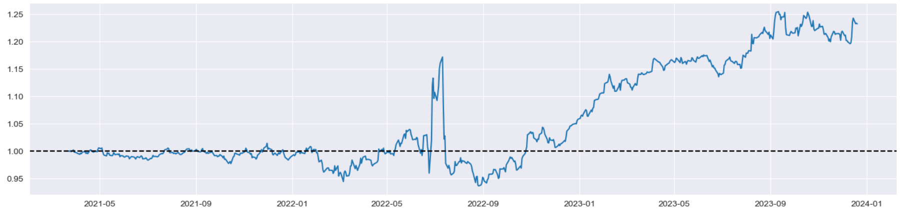


FIG. 5.2

in the later stages (figure 5.2).

This situation significantly affected our primary risk-adjusted metric, the Sharpe Ratio, which registered a relatively low 0.62 (table 5.4). A critical contributor to this subdued ratio was the pronounced volatility during June and July 2022, particularly marked by the significant losses from the ENI *vs* SPM pair. This period coincided with central banks' decisions to hike interest rates, ending a long era of low rates. Such a shift likely induced alterations in company fundamentals or the statistical patterns of price series. The next section will explore how the PCI approach improves this problem, solving problems related to idiosyncratic shocks, that assumed to be temporally, when instead they are permanent (Jondeau et al. 2015).

Delving into additional performance metrics, the Probability Sharpe Ratio[3] stands at 85.26%, indicating a strong likelihood that the portfolio's Sharpe Ratio reflects real performance and not just random fluctuation. The Sortino Ratio[4], which only considers downside risk, is recorded at 0.92, surpassing the Sharpe Ratio due to its exclusive focus on negative returns, thereby not penalizing positive performance. Lastly, the Ω Ratio[5] is observed at 1.16, signifying that the investment's returns are more favorably skewed above the minimum accept-

3. A probabilistic version of the Sharpe Ratio, assessing the likelihood of achieving a certain risk-adjusted return: PSR $= \Phi\big(\sqrt{T-k}\,(\hat{SR} - SR^*)/\hat{\sigma}_{\hat{SR}}\big)$, where Φ is the cumulative distribution function of the standard normal distribution, T is the track record, k is the number of estimated parameters, $\hat{SR}$ is the estimated Sharpe ration of the strategy, SR^* is the benchmark Sharpe ration and $\hat{\sigma}_{\hat{SR}}$ is the standard error of the Sharpe Ratio estimator.
4. A modification of the Sharpe Ratio that differentiates harmful volatility from total overall volatility by using the standard deviation of negative asset returns, called downside deviation: Sortino Ratio $= (R_p - R_f)/\sigma_d$, where R_p is the return of the portfolio, R_f is the risk-free rate and σ_d is the donwside deviation, which is the standard deviation of the negative portfolio returns.
5. A measure of the performance of an investment relative to a minimum acceptable return, considering both upside and downside volatility, given by

$$\Omega \text{ Ratio} = \int_{\text{MAR}}^{\infty} \big(1 - F(r)\big)\mathrm{d}r \Big/ \int_{-\infty}^{\text{MAR}} F(r)\mathrm{d}r,$$

where $F(r)$ is the cumulative distribution function of the returns, MAR is the minimum acceptable return, the numerator calculates the area under the return distribution curve above the MAR, representing gains and the denominator calculates the area under the return distribution curve below the MAR, representing losses.

TAB. 5.4. Cointegration Performance Analysis

Start Period	2021 03 23	Expected Daily %	0.03%
End Period	2023 12 20	Expected Monthly %	0.62%
Risk-Free Rate	0.0%	Expected Yearly %	7.22%
Time in Market	100.0%	Best Day	9.83%
Cumulative Return	23.25%	Worst Day	−7.46%
CAGR %	7.91%	Best Month	6.21%
Sharpe	0.62	Worst Month	−11.17%
Prob. Sharpe	85.26%	Best Year	16.5%
Smart Sharpe	0.52	Worst Year	−1.07%
Sortino	0.92	Avg. Drawdown	−2.7%
Smart Sortino	0.77	Avg. Drawdown Days	46
Sortino/$\sqrt{2}$	0.65	Recovery Factor	1.16
Smart Sortino/$\sqrt{2}$	0.55	Ulcer Index	0.07
Ω	1.16	Serenity Index	0.22
Max Drawdown	−20.1%	Avg. Up Month	2.17%
Longest DD Days	323	Avg. Down Month	−2.48%
Volatility (ann.)	13.52%	Win Days %	54.55%
Calmar	0.39	Win Month %	67.65%
Skew	0.96	Win Quarter %	50.0%
Kurtosis	41.6	Win Year %	66.67%
		N. of trades	10

able return than below, offering a balanced view of the potential for both gain and loss.

In the out-sample period, the strategy executed a total of 10 trades, averaging 1 or 2 trades per pair, ensuring continuous market engagement with all couples held in position by the end of the back-test. This reflects a 100% market participation rate,[6] typical of a strategy always maintaining at least one open position. The trading frequency is relatively low, characteristic of cointegration approaches, with just a few trades annually. This observation suggests potential benefits in exploring the strategy over different sub-periods or employing a more frequent trading interval to possibly enhance trading activity and risk-adjusted performance. The upcoming discussion on adopting a partial cointegration approach is anticipated to further address and potentially increase trade frequency. For the comprehensive performance assessment, the Quantstats Python library was utilized, offering a robust set of performance metrics commonly employed in hedge fund strategies.

Table 5.5 showcases the most significant drawdowns experienced during the back-testing period, with the most severe reaching approximately 20% and

6. Amount of time in which you have opened at least one position of the strategy.

lasting 323 days until full recovery. Given these substantial drawdowns and the strategy's year-long period of underperformance or losses, the fluctuating and inconsistent returns cast doubt on the method's reliability in the examined market over the specified period.

TAB. 5.5

Start	Valley	End	Days	Max Drawdown	99% Max Drawdown
2022 07 12	2022 08 26	2023 05 31	323	−20.095866	−19.652739
2022 05 31	2022 06 24	2022 06 28	28	−7.684061	−4.585677
2021 12 01	2022 03 07	2022 05 16	166	−6.894592	−5.888763
2023 09 12	2023 12 11	2023 12 20	99	−4.674510	−4.579572
2022 06 40	2022 07 04	2022 07 07	7	−3.617322	−3.279390

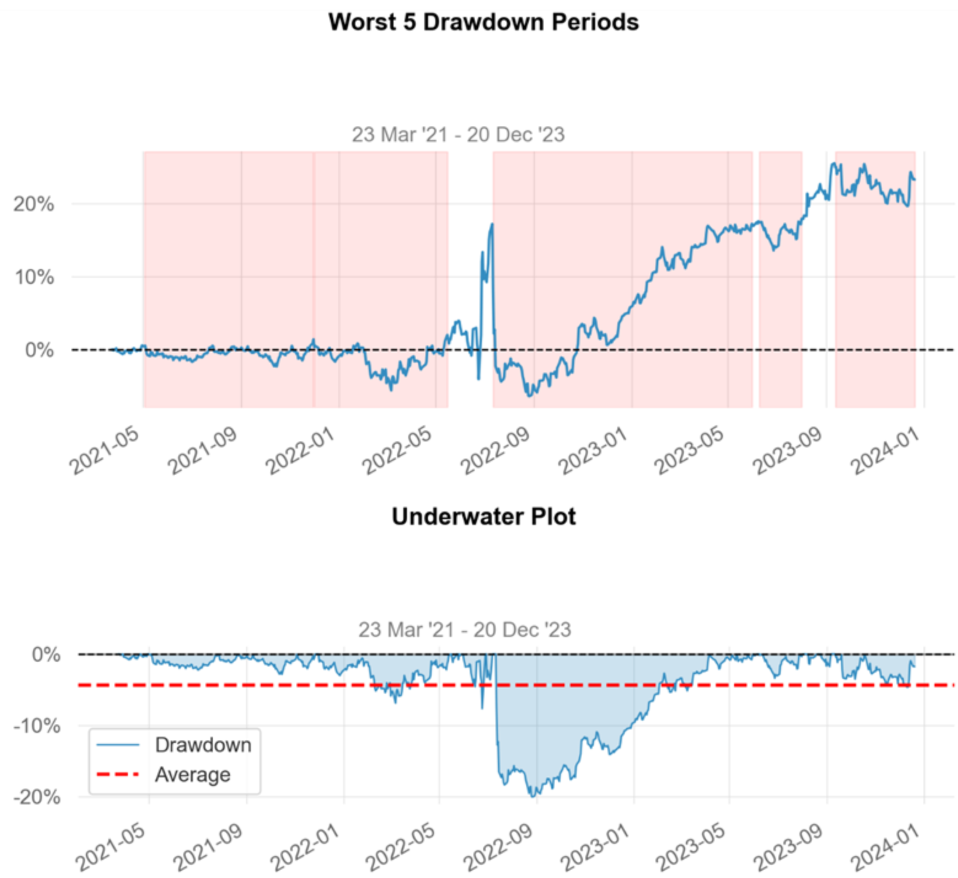


FIG. 5.3

The distribution is notably positively skewed, as displayed in figure 5.4, with a skewness measure of 0.96. This metric indicates a longer or heavier right tail, suggesting more frequent occurrences of returns significantly above the mean than significantly below. This particularity is typically favoured in finance as it implies the potential for larger gains over losses.

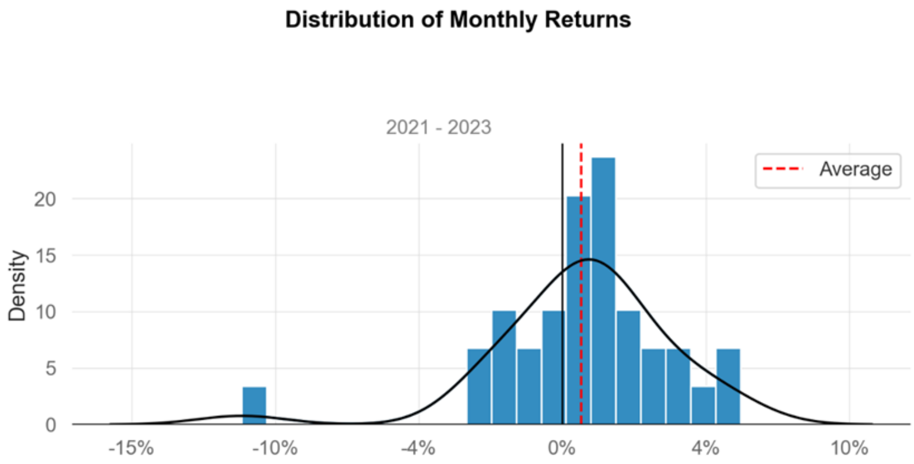


FIG. 5.4

However, the strategy also exhibits a high kurtosis[7] of 41.69. This high number means the returns are more prone to extreme values, both positive and negative, than a normal distribution would suggest. While the pronounced peak implies frequent modest incomes, the fat tails indicate a higher probability of significant fluctuations, introducing elevated levels of investment risk and potential for unexpected extreme events.

Figure 5.5 reveals the monthly return percentages, pinpointing July as the most challenging month, coinciding with central bank policy shifts that introduced significant market volatility. Interestingly, post-policy enactment, the strategy appeared to stabilize and even perform well, suggesting potential market correction or fortuitous timing.

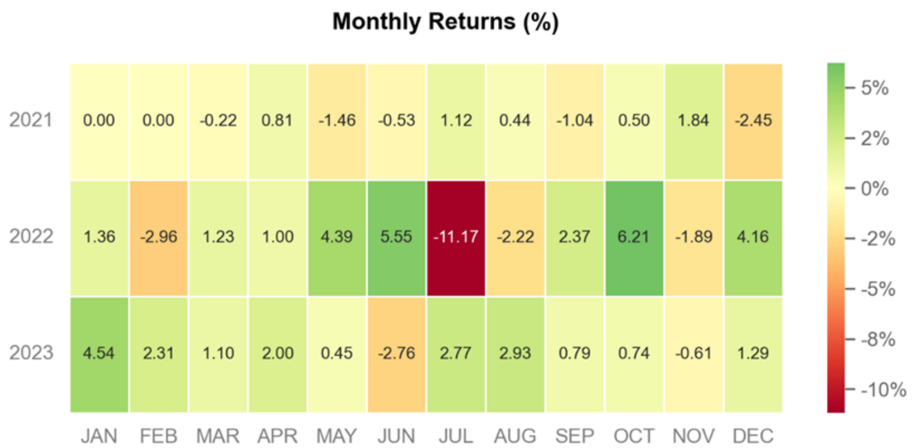


FIG. 5.5

7. Statistical measure that describes the shape of a distribution's tails in relation to its central peak, indicating the presence of extreme values (outliers) in comparison to a normal distribution.

Given these observations, it seems prudent to reevaluate the approach under different training and testing periods. For instance, recalibrating the model with data from 2021 and 2022 and then back-testing it for just the most recent year could provide insights into how the strategy adapts to the long-term effects of the interest rate shift. This approach would allow for the integration of this significant market event into the strategy's parameters. Additionally, using a two-year window for parameter estimation might offer a more accurate or relevant basis for predicting the thresholds that trigger trade decisions in the context of these synthetic spread series.

In our final analysis, we compare our strategy's performance against the FTSE MIB, the benchmark Italian stock index. A cursory review of the performance graph reveals that the market, like our strategy, struggled during the same period (figure 5.6).

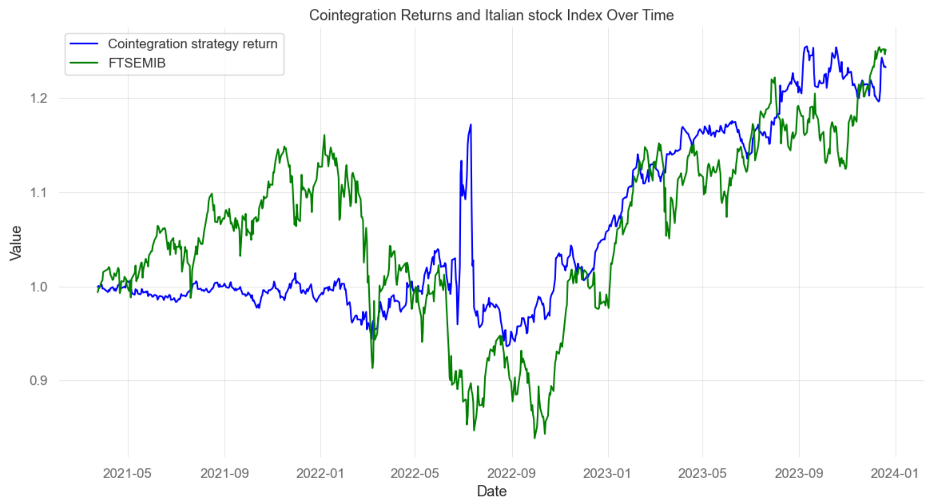


FIG. 5.6

TAB. 5.6

Cumulative Return	25.15 %
CAGR%	8.52 %
Sharpe	0.51
Max Drawdown	−27.73 %
Longest DD Days	540
Volatility (ann.)	19.47 %
Calmar	0.31
Skew	−0.48
Kurtosis	3.65

Specifically, the market exhibited a slightly lower Sharpe Ratio compared to our strategy (table 5.6). Furthermore, the maximum drawdown was more severe, increasing from 20% to 27.7%, with a longer recovery period extending beyond a year. The benchmark's additional performance metrics align with the empirically demonstrated stylize facts: a generally negative skewness and a kurtosis approaching the normative value around 3, reflecting typical market behavior.

In summary, while the classic cointegration strategy may not have delivered robust or particularly favorable results during this specific historical period, it's important to contextualize its performance by comparing it with its benchmark, over the same timeframe. This comparison reveals that despite the strategy's less-than-ideal results, it still managed to outperform the market.

5.2.2 Strategy 2

After confirming the theoretical soundness and simulated effectiveness of the model, we proceed to apply the PCI approach in a practical context to assess its real-world profitability. We begin by estimating the vector of parameters $[\rho, \sigma_M, \sigma_R]$. The selection of the most promising asset pairs is based on a set of criteria. The AIC for the PAR must be lower than that for the AR and RW models, ensuring we choose the model that best fits the data with the least complexity. We seek couples with an AR coefficient greater than 0.5 to ensure a substantial level of mean reversion. Moreover, we require that the proportion of variance due to mean reversion is over 0.5.

In addition to the metrics expressed in Clegg and Krauss's research, we incorporate two auxiliary conditions. We focus on pairs where the AR coefficient is not only above 0.5 but also below 0.8, as lower values have been associated with better performance. Finally, to maintain dollar neutrality, we select pairs where the β is between 0.75 and 1.25.

Upon applying the established conditions, we identify only two asset pairs that fulfil all the criteria for potential profitability, as shown in table 5.7

TAB. 5.7

Stock Pairs	α	β	ρ	σ_M	σ_R
SRG-TRN	0.742 387	0.993 45	0.685 573	0.034 416	0.023 692
HER-TRN	2.112 866	1.039 41	0.560 373	0.044 284	0.036 306

Fortunately, both selected pairs exhibit strong performance in the in-sample back-test, with Sharpe ratios of approximately 1.4 and 1.3, indicating their viability as tradable pairs (table 5.8).

Before proceeding to the out-of-sample back-test, we construct the residual series W_t and isolate the mean-reverting from the random walk part. Figures 5.7, 5.8, display the synthetic spreads, the AR component and its Z-score, with horizontal lines marking the thresholds for initiating and closing trades.

TAB. 5.8

Stock Pairs	Sharpe
SRG-TRN	1.394 371
HER-TRN	1.283 294

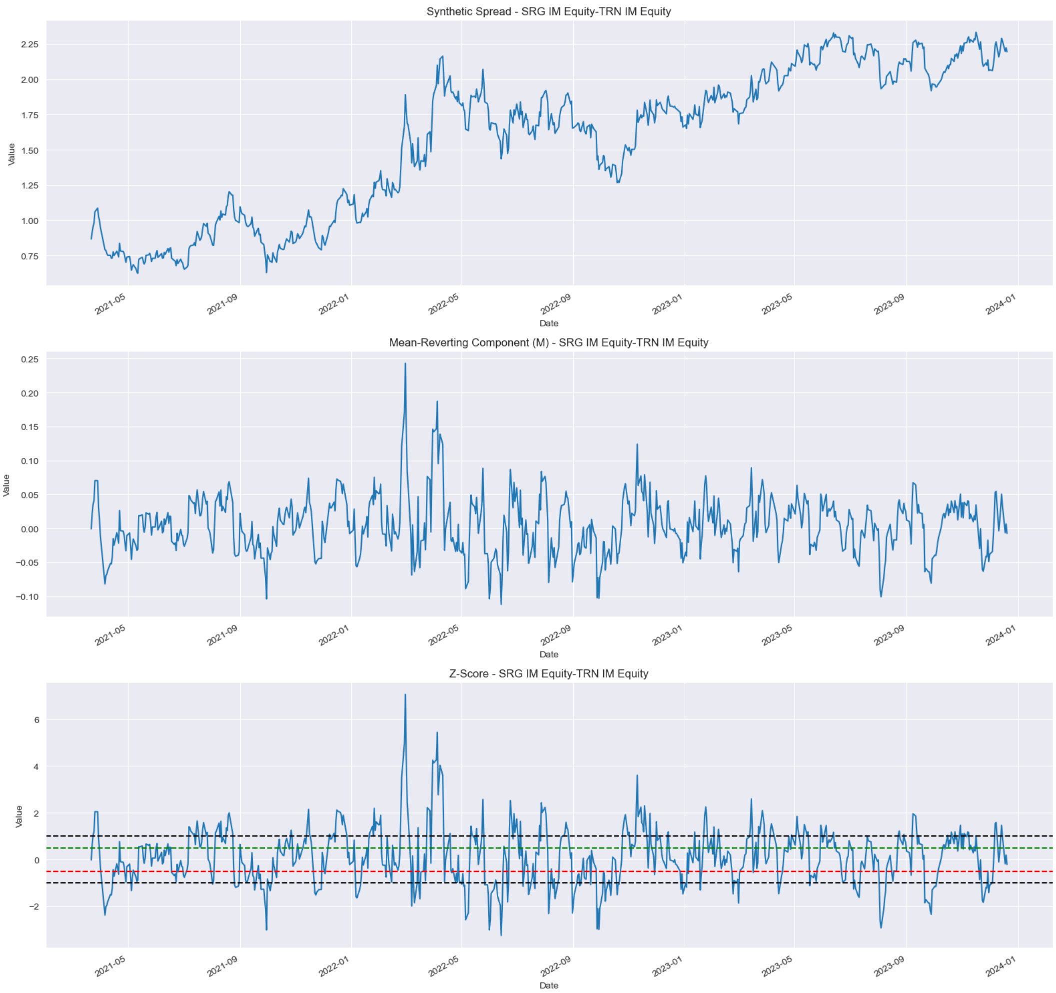


Fig. 5.7

Tab. 5.9

SRG-TRN	HER-TRN
1.236 072	1.856 616

When the back-test is carried out to the test dataset, both pairs demonstrate robust returns, with SRG *vs* TRN yielding 23.6% and HER *vs* TRN yielding 85.66% as we can see in figure 5.9 and table 5.9. Despite a lacklustre performance in the first year, the strategy's returns significantly improved in the second and third years.

We end up with a cumulative return of 53.14%, assuming equal investment in both stock pairs (figure 5.10). This strategy appears less impacted by the July 2022 interest rate shifts, showing steadier performance without significant volatility spikes. Post the interest rate hike, the strategy's performance improved, though it's unclear if this uptick is directly attributable to central bank policy changes.

The performance metrics (table 5.10) reveal that market engagement has decreased to 86% from the previous strategy's 100%, primarily due to more frequent but shorter-duration trades. The number of trades has surged from 10

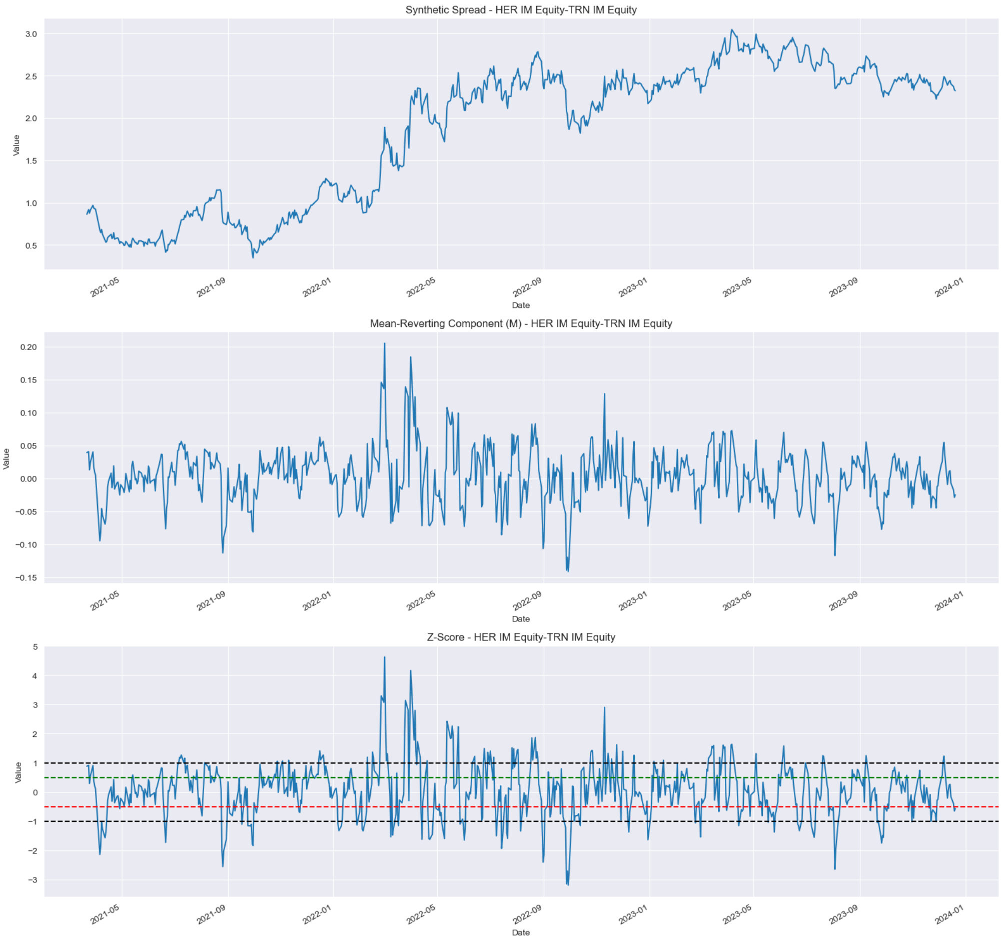


Fig. 5.8

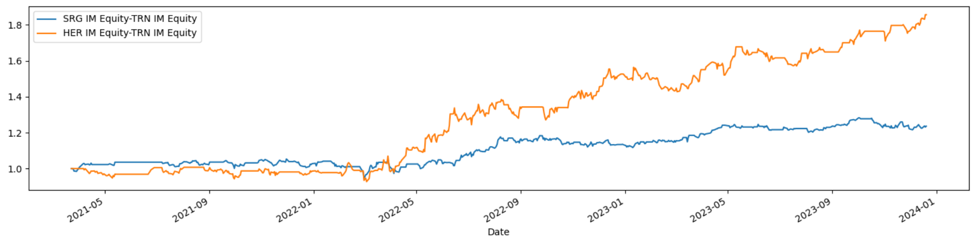


Fig. 5.9

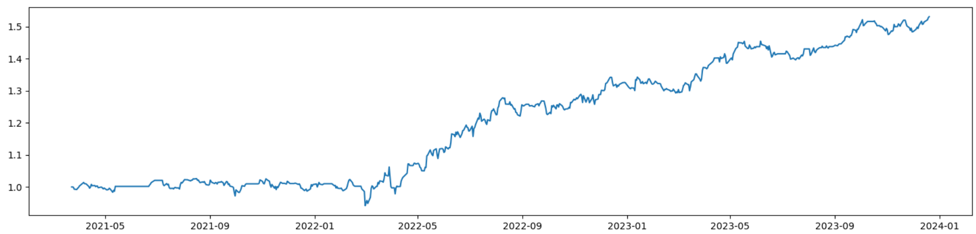


Fig. 5.10

to a total of 57, averaging about 10 trades per year. Concurrently, the duration of trades has reduced from around a month to just a few days.

Tab. 5.10. Partial Cointegration Performance Analysis

Start Period	2021 03 23	Expected Daily %	0.06%
End Period	2023 12 20	Expected Monthly %	1.26%
Risk-Free Rate	0.0%	Expected Yearly %	15.27%
Time in Market	86.0%	Best Day	3.42%
Cumulative Return	53.15%	Worst Day	−4.5%
CAGR %	16.8%	Best Month	7.02%
Sharpe	1.38	Worst Month	−3.54%
Prob. Sharpe	98.94%	Best Year	31.52%
Smart Sharpe	1.34	Worst Year	0.76%
Sortino	2.14	Avg. Drawdown	−1.85%
Smart Sortino	2.07	Avg. Drawdown Days	22
Sortino/$\sqrt{2}$	1.51	Recovery Factor	6.47
Smart Sortino/$\sqrt{2}$	1.46	Ulcer Index	0.02
Ω	1.3	Serenity Index	5.48
Max Drawdown	−8.22%	Avg. Up Month	3.14%
Longest DD Days	219	Avg. Down Month	−1.04%
Volatility (ann.)	11.5%	Win Days %	52.81%
Calmar	2.04	Win Month %	55.88%
Skew	0.0	Win Quarter %	75.0%
Kurtosis	4.67	Win Year %	100.0%
		N. of trades	57

Despite the potential impact on transaction costs, 10 trades per year are considered manageable. The notable cumulative return of 53.15% far exceeds the previous strategy's 23.25%. A Sharpe ratio of 1.38 indicates a respectable risk-adjusted performance, especially compared to the benchmark's lower ratio of 0.51. The probability-adjusted Sharpe Ratio's increase to approximately 99% instills further confidence in the strategy's robustness. Enhanced performance is also evident in other metrics like the Sortino Ratio and Ω, all of which are superior to those of the previous approach.

The strategy exhibits manageable drawdown levels (table 5.11), with the highest being only 6% and a recovery period of no more than three months for losses, aside from the initial unprofitable year.

Tab. 5.11

Start	Valley	End	Days	Max Drawdown	99% Max Drawdown
2021 08 17	2022 03 21	2022 03 24	219	−8.219499	−7.370208
2022 03 30	2022 04 05	2022 04 20	21	−7.874333	−6.128989
2022 08 10	2022 08 29	2022 11 07	89	−4.463634	−4.154324
2023 06 07	2023 07 17	2023 09 11	96	−4.001707	−4.848091
2023 01 16	2023 02 27	2023 03 23	64	−3.691508	−3.611328

The higher number of trades in this strategy yields a symmetric distribution of returns (figure 5.11), approximately zero skew, and a kurtosis of 4.67, aligning closely with a normal one. Although the positive skew of classical cointegration is generally favourable, its lower trade frequency can cast doubt on its effectiveness. Conversely, the PCI model's increased market executions lend more credibility to its results.

FIG. 5.11

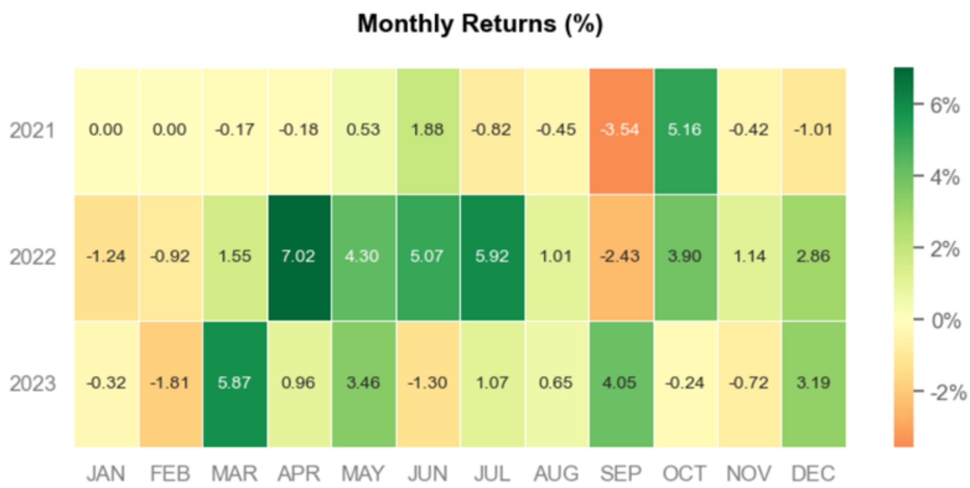


FIG. 5.12

Figure 5.12 detailing monthly results suggests that while it would have been illustrative to see the cointegration strategy falter post-July 2022 – thereby highlighting the PAR model's capability to mitigate the effects of permanent shocks not addressed by traditional cointegration – the reality is different. Contrary to expectations, the cointegration strategy actually performed better after the central banks' decision. Notably, this strategy began to excel even before the rate change, presenting a contrast to the anticipated narrative.

In conclusion, by comparing our strategy's performance with the FTSE MIB in figure 5.13, it seems that shortly after the market began to decline, our method started to generate profits. The fact that it earns during market downturns, while the benchmark index declines, indicates that the strategy is market-neutral and not correlated with general market trends.

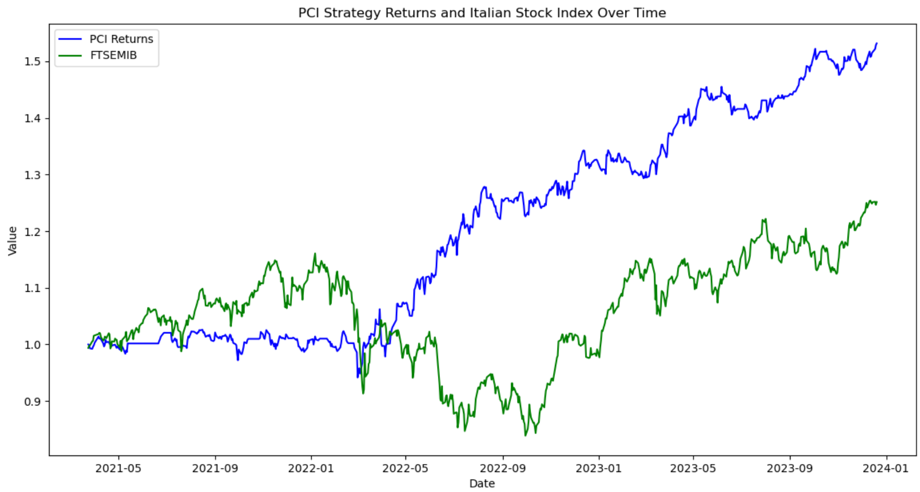


FIG. 5.13

The strategy significantly outperforms the Buy and Hold approach on the Italian Index, evidenced by its 53.15% return compared to the benchmark's 25.15% and a superior Sharpe Ratio of 1.38 against the benchmark's 0.51 (FTSE MIB performance metrics are shown in previous Subsection 5.2.1). It excels across various performance metrics, including trade frequency and controlled drawdowns. However, the reliance on just two pairs raises questions about the breadth and luck factor in its success. Therefore, expanding the study to include a larger sample beyond 40 stocks would provide a more comprehensive and credible evaluation of the strategy's efficacy.

5.3. *Value-at-Risk Computation*

Value at Risk (VaR) is a widely used risk management tool that quantifies the potential loss of a risky asset or portfolio over a defined period, for a given confidence level. Essentially, VaR provides a probabilistic estimate of the minimum loss expected from an investment over a specified time frame due to market risks. It's particularly useful in financial institutions for determining capital reserves, risk management, and regulatory compliance.

Instead, Conditional Value at Risk (CVaR), also known as Expected Shortfall (ES), is a risk assessment measure that provides an estimate of the expected losses that will occur beyond the VaR threshold within a given confidence level. Unlike the first one, which gives a threshold value for losses, CVaR gives the average of the losses that exceed the VaR, effectively capturing the tail risk or the extreme loss scenarios.

We will explore and calculate the potential future losses using two different approaches: a historical one and parametric one.

5.3.1 Historical and Parametric VaR

The Historical VaR method involves computing the potential loss by directly examining historical price movements or returns of the asset or portfolio. By ranking these historical changes from worst to best and selecting a percentile that corresponds to the desired confidence level, Historical VaR directly reflects the observed variability and distribution of returns over the chosen historical period.

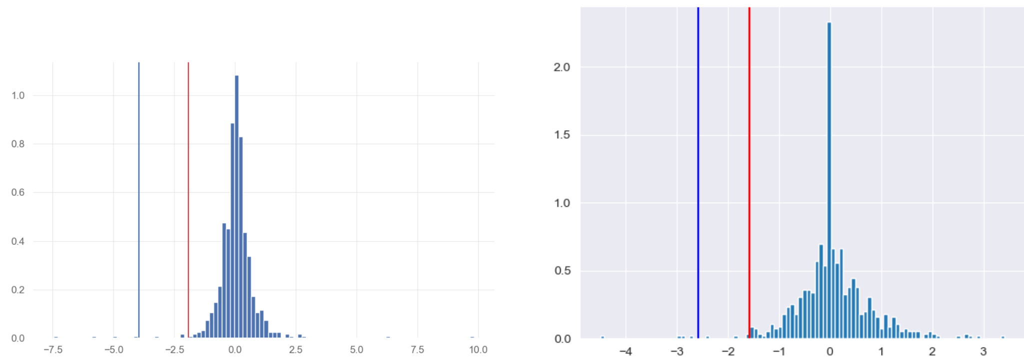


FIG. 5.14

When comparing the historical outcomes at 99% confidence level between cointegration and partial-cointegration (figure 5.14), the second demonstrates a lower VaR, indicating a reduced risk level. Specifically, the partial-cointegration's VaR is −1.58 compared to the cointegration's −1.91, both suggesting a moderate risk of loss. A more pronounced difference emerges when considering Conditional VaR. Here, the partial-cointegration strategy shows a CVaR of approximately −2.57, significantly lower than the cointegration strategy's −3.95, as shown in table 5.12. This indicates that the partial-cointegration method not only has a lower average loss beyond the VaR threshold but also a diminished likelihood of incurring large, extreme losses, providing a more favorable risk profile especially in the tail ends of the loss distribution.

Unlike the Historical VaR, Parametric VaR assumes that asset returns are

TAB. 5.12

	Cointegration	PCI
Historical VaR (99)	−1.910 044 552 074 3254 %	−1.582 153 720 815 9167 %
Historical CVaR (99)	−3.955 276 809 954 533 %	−2.568 303 842 317 08 %

normally distributed and utilizes the mean and standard deviation of these returns to estimate risk. It calculates VaR using the formula VaR $= Z\sigma\sqrt{t}$, where Z is the value related to the desired confidence level (e.g., 1.65 for 95% confidence), σ is the standard deviation of the returns, and t is the time horizon. This method is computationally straightforward and allows for easy aggregation of risks across different assets. However, its reliance on the normal distribution assumption and a constant volatility may not capture extreme events or changes in market conditions effectively.

Even with the alternative method, the results affirm that the partial cointegration (PCI) strategy presents lower VaR and CVaR values compared to the traditional model, with an increased difference particularly in the CVaR. The historical approach, which yields slightly better results than the parametric one, emphasizes its effectiveness due to its reliance on actual market data, capturing anomalies and extreme market conditions more accurately (table 5.13).

TAB. 5.13

	Cointegration	PCI
Parametric VaR (99)	−1.947 288 394 326 874 %	−1.620 563 611 998 130 %
Parametric CVaR (99)	−4.246 453 854 000 351 %	−2.708 941 879 226 996 %

5.4. *Other Considerations*

Upon comprehensive analysis of the model both theoretically and practically, we are prepared to offer final observations and suggest avenues for future research. Notably, neither of the strategies employed a stop-loss mechanism. In cointegration, it might be prudent to implement a stop at levels of ±4 standard deviations, while for partial cointegration, following Krauss's academic suggestions, closing trades at a 10% loss could be considered. Additionally, time-based stop-loss mechanisms, such as automatically closing trades after a set duration, could significantly impact strategy performance and should be explored.

Transaction costs, including commissions, the impact of bid-ask spreads and slippage,[8] were not accounted for in our analysis. Even though the trade fre-

8. The difference between the expected price of a trade and the price at which the trade is executed.

quency is moderate, incorporating these costs is crucial for a realistic assessment of strategy profitability.

Finally, integrating machine learning could offer a dynamic approach to both classical cointegration and partial strategies. Parameters could be estimated over a specific in-sample period (e.g., a month) and then applied in trading for the subsequent 15 days (out-sample), followed by a rolling window adjustment. This method, although computationally intensive, would adapt more fluidly to changing market conditions, and it would be interesting to analyze the outcomes of such approach.

6. Conclusion

In conclusion, our research demonstrates how, in the current financial landscape, not only have traditional macro strategies and technical analysis seen a decline in their potential, but even quantitative approaches of a relatively simple nature are experiencing diminished effectiveness. This can be attributed to the fact that the widespread adoption of simplistic strategies leads to market efficiency against these methods, consequently reducing the profitability of such simplistic strategies. In contrast, more complex quantitative approaches, which are less commonly employed, still present opportunities to exploit market inefficiencies for profit. The complexity and competitiveness of the ever-evolving financial market dictate that profitability for traders today necessitates a continual pursuit and study of innovative models to seize opportunities not yet recognized by competing traders.

Appendices

A. Identifiability

This appendix details the conditions for the identifiability of a system from its data. A system is identifiable if two conditions hold true for a pair of series X_1 and X_2: to begin, the first differences of X_1, therefore $(1 - B)X_{1,t}$, are stationary; and second, these differences are independent from the first differences of a noise and market impact term, $(1 - B)(M_t + R_t)$.

To confirm the identifiability of a system, we analyze its state space representation. Identifiability is assured if the first difference of $X_{1,t}$, denoted as D_t and having variance σ_x^2, is independent from the noise $\epsilon_{M,t}$ and $\epsilon_{R,t}$. The first difference of $X_{1,t}$ is then independent of the combined market and noise impact term $(1 - B)(M_t + R_t)$, leading to a zero covariance between D_t and $(1 - B)(M_t + R_t)$, so

$$\mathrm{Cov}\left[D_t,\, (1 - B)(M_t + R_t)\right] = 0.$$

Considering this, the first difference of $X_{2,t}$ is expressed as $(1 - B)(\beta X_{1,t} + M_t + R_t)$. The covariance between the first differences of $X_{1,t}$ and $X_{2,t}$ can be broken down into two parts: the covariance of D_t with itself and the covariance of D_t with the impact term, the latter of which is zero. Thus, the covariance is simplified to β times the variance of D_t. In other words:

$$(1 - B)X_{2,t} = (1 - B)(\beta X_{1,t} + M_t + R_t) = \beta D_t + (1 - B)(M_t + R_t).$$

Consequently,

$$\begin{aligned}\mathrm{Cov}\left[(1 - B)X_{1,t},\, (1 - B)X_{2,t}\right] &= \mathrm{Cov}\left[D_t,\, \beta D_t + (1 - B)(M_t + R_t)\right] = \\ &= \mathrm{Cov}[D_t,\, \beta D_t] + \mathrm{Cov}\left[D_t,\, (1 - B)(M_t + R_t)\right] = \\ &= \beta\,\mathrm{Var}[D_t].\end{aligned}$$

The parameter β is recoverable via the ratio of the covariance of the first differences of $X_{1,t}$ and $X_{2,t}$ to the variance of the first difference of $X_{1,t}$:

$$\beta = \frac{\mathrm{Cov}\left[(1 - B)X_{1,t},\, (1 - B)X_{2,t}\right]}{\mathrm{Var}\left[(1 - B)X_{1,t}\right]}.$$

Once β is obtained, the serie W_t can be computed:

$$W_t = X_{2,t} - \beta X_{1,t} = M_t + R_t.$$

the sequence W_t is recognized to be partially autoregressive and identifiable (Clegg, 2015). Knowing that

$$v_k = \mathrm{Var}\left[(1 - B^k)W_t\right],$$

we can estimate the subsequent parameters:

$$\rho = -\frac{v_1 - 2v_2 + v_3}{2v_1 - v_2}, \quad \sigma_M^2 = \frac{1}{2}\frac{\rho + 1}{\rho - 1}(v_2 - 2v_1), \quad \sigma_R^2 = \frac{1}{2}(v_2 - 2\sigma_M^2).$$

B. Likelihood Function Proof

Let Θ_t represent the information set available up to and including a specific point in time, t, and let Φ symbolize the set of parameters $[\beta, \rho, \sigma_X, \sigma_M, \sigma_R]$. With $X_1, X_2, \ldots, X_n$ denoting the sequence of observations, and Φ the corresponding parameter values, the likelihood function can be expressed accordingly as

$$\mathcal{L}(\phi) = p(X_1|\Phi)\prod_{k=2}^{n} p(X_k|\Theta_{k-1}, \Phi).$$

Through the Markov property the expression becomes

$$\mathcal{L}(\phi) = p(X_1|\Phi)\prod_{k=2}^{n} p(X_k|X_{k-1}, \Phi).$$

We can extend this expression for the element X_t obtaining

$$\mathcal{L}(\phi) = p(X_{1,1}, X_{2,1}|\Phi)\prod_{k=2}^{n} p(X_{1,k}, X_{2,k}|X_{1,k-1}, X_{2,k-1}, \Phi).$$

We remember from the rules of probability that $p(A, B|C) = p(A|B, C)\,p(B|C)$. Concentrating only on this term of the product $p(X_{1,k}, X_{2,k}|X_{1,k-1}, X_{2,k-1}, \Phi)$ we have

$$\begin{aligned} p(X_{1,k}, X_{2,k}|X_{1,k-1}, X_{2,k-1}, \Phi) &= \\ &= p(X_{2,k}|X_{1,k}X_{1,k-1}, X_{2,k-1}, \Phi)\,p(X_{1,k}|X_{1,k-1}, X_{2,k-1}, \Phi) \end{aligned}$$

We advance by individually assessing the two terms on the right side of the preceding equation, beginning with the first term for analysis

$$\begin{aligned} p(X_{1,k}|X_{1,k-1}, X_{2,k-1}, \Phi) &= p(X_{1,k} - X_{1,k-1}|X_{1,k-1}, X_{2,k-1}, \Phi) = \\ &= p(\epsilon_{X,k}|X_{1,k-1}, X_{2,k-1}, \Phi) = \\ &= \phi(\epsilon_{X,k}; 0, \sigma_X^2). \end{aligned}$$

Here, ϕ symbolizes the probability density function of the normal distribution characterized by a mean of zero and a variance of σ_X^2. To assess the probability $p(X_{2,k}|X_{1,k}X_{1,k-1}, X_{2,k-1}, \Phi)$ we observe that

$$\begin{aligned}
X_{2,k} - E(X_{2,k}|X_{1,k}m\Theta_{k-1}, \Phi) &= \beta X_{1,k} + M_k + R_k + \\
&\quad - E(\beta X_{1,k} + M_k + R_k|X_{1,k}\Theta_{k-1}, \Phi) = \\
&= M_k + R_k - E(M_k + R_k|X_{1,k}, \Theta_{k-1}, \Phi) = \\
&= M_k - E(M_k|X_{1,k}, \Theta_{k-1}, \Phi) + \\
&\quad + R_k - E(R_k|X_{1,k}, \Theta_{k-1}, \Phi) = \\
&= \rho M_{k-1} + \epsilon_{M,k} - E(\rho M_{k-1} + \epsilon_{M,k}|X_{1,k}, \Theta_{k-1}, \Phi) + \\
&\quad + R_{k-1} + \epsilon_{R,k} - E(R_{k-1} + \epsilon_{R,k}|X_{1,k}, \Theta_{k-1}, \Phi) = \\
&= \epsilon_{M,k} - E(\epsilon_{M,k}|X_{1,k}, \Theta_{k-1}, \Phi) + \\
&\quad + \epsilon_{R,k} - E(\epsilon_{R,k}|X_{1,k}, \Theta_{k-1}, \Phi) = \\
&= \epsilon_{M,k} + \epsilon_{R,k}.
\end{aligned}$$

And so

$$\begin{aligned}
p(X_{2,k}|X_{1,k}X_{1,k-1}, X_{2,k-1}, \Phi) &= p(\epsilon_{M,k} + \epsilon_{R,k}|X_{1,k}X_{1,k-1}, X_{2,k-1}, \Phi) \\
&= \phi(\epsilon_{M,k} + \epsilon_{R,k}; 0, \sigma_M^2 + \sigma_R^2).
\end{aligned}$$

Integrating the previously discussed components and revisiting the likelihood function, we arrive at a comprehensive formula that encapsulates all the analysed elements:

$$\begin{aligned}
\mathcal{L}(\phi) &= p(X_{1,1}, X_{2,1}, \Phi) \prod_{k=2}^{n} p(X_{1,k}, X_{2,k}|X_{1,k-1}, X_{2,k-1}, \Phi) = \\
&= p(X_{1,1}, X_{2,1}, \Phi) \prod_{k=2}^{n} \phi(\epsilon_{X,k}; 0, \sigma_X^2)\phi(\epsilon_{M,k} + \epsilon_{R,k}; 0, \sigma_M^2 + \sigma_R^2) \\
&= p(X_{1,1}, X_{2,1}, \Phi) \left(\prod_{k=2}^{n} \phi(\epsilon_{X,k}; 0, \sigma_X^2) \right) \left(\prod_{k=2}^{n} \phi(\epsilon_{M,k} + \epsilon_{R,k}; 0, \sigma_M^2 + \sigma_R^2) \right) \\
&= p(X_{1,1}, X_{2,1}, \Phi)\, \mathcal{L}_X(\sigma_X)\, \mathcal{L}_{MR}(\beta, \rho, \sigma_M, \sigma_R).
\end{aligned}$$

Bibliography

Bertram, W. K., 2010a. *A non-stationary model for statistical arbitrage trading*. SSRN Electronic Journal.

Bertram, W. K., 2010b. *Analytic solutions for optimal statistical arbitrage trading*. Physica A: Statistical Mechanics and its Applications 389 (11), 2234–2243.

Bozdogan, H., 1987. *Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions*. Psychometrika 52, 345–370.

Brockwell, P. J., Davis, R. A., 2010. *Introduction to time series and forecasting*, 2nd edition. Springer texts in statistics. Springer, New York.

Caldeira, J. F., Moura, G. V., 2013. *Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy*. Brazilian Review of Finance 11 (1), 49–80.

Clegg, M., 2015. *Modeling time series with both permanent and transient components using the partially autoregressive model*. SSRN Electronic Journal.

Clegg, M., Krauss, C., 2016. *Pairs trading with partial cointegration*. Independent (Clegg), University of Erlangen-Nürnberg (Krauss).

Clegg, M., Krauss, C., Rende, J., 2016. *partialCI: An R package for the analysis of partially cointegrated time series*. Independent (Clegg), University of Erlangen-Nürnberg (Krauss and Rende).

Do, B., Faff, R., 2010. *Does simple pairs trading still work?* Financial Analysts Journal 66 (4), 83–95.

Do, B., Faff, R., 2012. *Are pairs trading profits robust to trading costs?* Journal of Financial Research 35 (2), 261–287.

Dunis, C. L., Giorgioni, G., Laws, J., Rudy, J., 2010. *Statistical arbitrage and high-frequency data with an application to Eurostoxx 50 equities*. Working paper, Liverpool Business School.

Durbin, J., Koopman, S. J., 2012. *Time series analysis by state space methods*. Vol. 38 of Oxford Statistical Science Series. Oxford University Press, Oxford.

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

Gatev, E., Goetzmann, W. N., Rouwenhorst, K. G., 1999. *Pairs trading: Performance of a relative value arbitrage rule*. Working paper, Yale School of Management's International Center for Finance.

Gatev, E., Goetzmann, W. N., Rouwenhorst, K. G., 2006. *Pairs trading: Performance of a relative-value arbitrage rule*. Review of Financial Studies 19 (3), 797–827.

Huck, N., 2015. Pairs trading: *Does volatility timing matter?* Applied Economics 47 (57), 6239–6256.

Huck, N., Afawubo, K., 2015. *Pairs trading and selection methods: is cointegration superior?* Applied Economics 47 (6), 599–613.

Jacobs, H., Weber, M., 2015. *On the determinants of pairs trading profitability*. Journal of Financial Markets 23, 75–97

Johansen, S., 1988. *Statistical analysis of cointegration vectors*. Journal of Economic Dynamics and Control 12 (2-3), 231–254.

Johansen, S., Juselius, K., 1990. *Maximum likelihood estimation and inference on cointegration–with applications to the demand for money*. Oxford Bulletin of Economics and Statistics 52 (2), 169–210.

Jondeau, E., Lahaye, J., Rockinger, M., 2015. *Asymmetry in the price impact of trades in an high-frequency microstructure model with jumps*. Journal of Banking and Finance, forthcoming.

Krauss, C., 2015. *Statistical arbitrage pairs trading strategies: Review and outlook*. IWQW Discussion Paper Series, University of Erlangen-Nürnberg.

Krauss, C., Herrmann, K., Teis, S., 2015. *On the power and size properties of cointegration tests in the light of high-frequency stylized facts*. IWQW Discussion Paper Series, University of Erlangen-Nürnberg.

Nath, P., 2003. *High frequency pairs trading with U.S. treasury securities: Risks and rewards for hedge funds*. Working paper, London Business School.

Poterba, J. M., Summers, L. H., 1988. *Mean reversion in stock prices*. Journal of Financial Economics 22 (1), 27–59.

Rad, H., Low, Rand Kwong Yew, Faff, R. W., 2015. *The profitability of pairs trading strategies: Distance, cointegration, and copula Methods*. Working paper, University of Queensland.

Simon, D., 2006. *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, Hoboken, N.J.

Summers, L. H., 1986. *Does the stock market rationally reflect fundamental values?* The Journal of Finance 41 (3), 591–601.