



Evaluation of individual stock portfolio performance on the Indonesia stock exchange using the Ortiz, Jorion, and Markowitz approaches

Enggal Dwi Mulyaningtyas^{1,*}, Rofikoh Rokhim¹

¹ Master of Management, Faculty of Economics and Business, Universitas Indonesia, Central Jakarta, Jakarta 10430, Indonesia.

*Correspondence: enggal.dwi@ui.ac.id

Received Date: May 21, 2026

Revised Date: July 10, 2026

Accepted Date: July 21, 2026

ABSTRACT

Background: Indonesia's capital market has experienced a sustained increase in investor participation, creating a stronger need for systematic and implementable portfolio construction methods. This study evaluates estimation risk in mean-variance optimization by comparing traditional Markowitz optimization with two mean-shrinkage approaches: the parameter-focused Bayes-Stein estimator of Jorion and the decision-focused optimal shrinkage of means proposed by Ortiz et al. **Methods:** The study uses monthly individual stock data from the Indonesia Stock Exchange over January 2006-December 2025. Excess returns are calculated relative to a monthly risk-free proxy, and portfolios are evaluated using a 120-month rolling-window out-of-sample backtest under long-only constraints. The sample is a balanced panel of 74 stocks with complete monthly data, and the monthly deposit insurance rate of the Indonesia Deposit Insurance Corporation (LPS) is used as the risk-free proxy. Portfolio performance is assessed using monthly and annualized Sharpe ratios, while weight stability is assessed using average weight volatility and turnover. Newey-West tests are used to evaluate whether differences across methods are statistically significant. **Findings:** The Ortiz approach consistently selects an optimal shrinkage intensity of zero, making its weights and performance effectively identical to the traditional Markowitz portfolio. Markowitz and Ortiz record an annualized out-of-sample Sharpe ratio of 0.0713, while Jorion records 0.0608. The statistical tests indicate that differences in out-of-sample performance and stability are not significant across the three approaches. In economic terms, the annualized Sharpe ratios of all three methods are very low (below 0.08), indicating that long-only optimization of individual Indonesian stocks delivered only marginal risk-adjusted excess returns over the sample period. **Conclusion:** In the Indonesian stock market setting, more complex mean-shrinkage methods do not automatically produce superior portfolio outcomes. **Novelty/Originality of this article:** This article provides Indonesian market evidence on the comparison between parameter-focused and decision-focused shrinkage approaches within a consistent rolling-window portfolio backtesting framework.

KEYWORDS: optimal shrinkage of means; out-of-sample; portfolio optimization; Sharpe ratio; turnover.

1. Introduction

Indonesia's capital market has become increasingly relevant for empirical portfolio research because the investor base has expanded substantially and market activity has become more dynamic. Data from the Indonesian Central Securities Depository show that the number of capital market Single Investor Identification (SID) increased from 3.88

Cite This Article:

Mulyaningtyas, E. D. & Rokhim, R. (2026). Performance evaluation of individual stock portfolios using the Ortiz, Jorion, and Markowitz approaches on the Indonesia stock exchange. *Economic Military and Geographically Business Review*, 4(1), 125-140 <https://doi.org/10.61511/emagrap.v4i1.2026.3789>

Copyright: © 2026 by the authors. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



million in 2020 to 10.31 million in 2022 and continued to increase to 14.83 million in 2024 (*Kustodian Sentral Efek Indonesia*, 2023, 2024). This expansion implies that a larger number of investors now face decisions about asset allocation, risk control, and portfolio implementation under market uncertainty.

As more investors enter the capital market, investment decisions are no longer limited to stock selection, but also involve the construction of portfolios that can balance return expectations and risk exposure. This issue is especially important because individual investors may be attracted to stocks with high historical returns without fully considering whether such returns are persistent, whether the associated risk is acceptable, or whether the allocation remains stable when market conditions change.

In an emerging market such as Indonesia, several market characteristics amplify the challenge of portfolio construction. These conditions can make historical average returns less reliable than estimates of future expected returns. Therefore, a portfolio that appears optimal within the historical estimation period may become less effective when applied to a future period. One of the central issues addressed in this study is the gap between in-sample attractiveness and out-of-sample performance.

The Indonesian stock market also experiences episodes of severe market stress. During the first quarter of 2020, the Indonesia Composite Index experienced a sharp decline during capital outflow and exchange-rate pressure, illustrating the vulnerability of emerging markets to sudden shocks (*Otoritas Jasa Keuangan/OJK*, 2020). In such an environment, portfolio construction cannot rely only on historical average returns. Parameter uncertainty, instability in expected returns, and the risk that an apparently optimal portfolio performs poorly outside the estimation sample must also be addressed.

Modern Portfolio Theory provides a classical foundation for portfolio selection. However, in practice traditional mean-variance optimization is highly sensitive to estimation error, especially error in the sample mean. This sensitivity often produces extreme and unstable portfolio weights, a problem emphasized in portfolio literature (Michaud, 1989; DeMiguel et al., 2009). The issue is particularly relevant for individual stocks in emerging markets, where historical returns can be noisy, heterogeneous, and regime-dependent.

The sensitivity of mean-variance optimization to estimation error is important because it is difficult to estimate the expected return vector with precision. Expected-return estimates are particularly susceptible to sampling error relative to covariance estimates, especially when the number of assets is large relative to the available historical observations (Chopra & Ziemba, 1993). A small difference in the estimated mean return can result in a large change in the portfolio weights because the optimization process attempts to allocate more weight to assets that appear to offer superior risk-adjusted opportunities. When the noise drives the estimated superiority rather than the true expected return, the resulting portfolio can become unstable and perform poorly in the next period.

This estimation problem creates a practical trade-off between using expected returns to differentiate assets and limiting the optimizer's sensitivity to noisy historical estimates. This behavior underlies Michaud's (1989) characterization of mean-variance optimizers as estimation-error maximizers. Consequently, the performance of an optimization method should be judged not only by how well it fits the historical sample, but also by how well the resulting portfolio performs in an out-of-sample setting.

Shrinkage estimators are commonly used to stabilize noisy estimates. In portfolio analysis, Jorion (1986) proposed a Bayes-Stein estimator that shrinks the sample means toward a more stable target. This approach is parameter-focused because the estimator is designed to improve the estimated mean before optimizing the portfolio. More recently, Ortiz et al. (2025) proposed an optimal shrinkage of means in the Markowitz model from a decision-focused perspective. Instead of prioritizing the accuracy of the mean estimate, the approach selects shrinkage intensity based on the quality of the resulting portfolio decision.

Despite the extensive literature on estimation risk, an important gap remains. Most empirical evidence on mean-shrinkage estimators has been produced using developed-market data, whereas evidence for emerging markets, where return noise, liquidity frictions, and structural breaks are more pronounced, is still limited. In addition, the decision-focused optimal shrinkage of means proposed by Ortiz et al. (2025) has not yet been tested on individual stocks in the Indonesian market under realistic long-only constraints. Therefore, it is unclear whether the theoretical benefits of shrinkage documented in developed markets carry over to an emerging-market setting in which the estimation environment is arguably more hostile.

This study compares three approaches: traditional Markowitz optimization, Jorion's parameter-focused shrinkage estimator, and Ortiz's decision-focused optimal shrinkage of means. The research asks whether these approaches generate different out-of-sample portfolio performances and whether they differ in terms of portfolio weight stability. The contribution of this study lies in providing evidence of the Indonesian market using a consistent rolling-window out-of-sample backtesting framework for individual stocks.

The originality of this article lies in its application of the decision-focused shrinkage framework to individual stocks in the Indonesian market. Therefore, the results of this study contribute to the literature by showing whether the theoretical benefits of shrinkage are realized when applied to a long-horizon Indonesian stock sample under realistic long-only portfolio constraints.

1.1 Portfolio optimization and risk estimation

Modern portfolio theory was developed by Markowitz (1952), who formalized portfolio selection as a trade-off between expected return and variance. In its plug-in form, mean-variance optimization replaces the unknown population parameters with their sample counterparts and treats the estimates as if they were the true values.

A central limitation of this plug-in approach is its sensitivity to estimation error in the input parameters. Jobson & Korkie (1980) demonstrated that estimated efficient frontiers can substantially deviate from their population counterparts. Best & Grauer (1991) showed that optimal weights are extremely sensitive to small changes in expected returns. Chopra & Ziemba (1993) quantified the relative importance of estimation errors and found that means errors are far more damaging than errors in variance or covariance errors. Green & Hollifield (1992) explained why mean-variance efficient portfolios often take extreme positions. Michaud (1989) summarized the problem by describing optimizers as estimation-error maximizers. Taken together, these findings indicate that uncertainty in expected-return estimates is a major source of instability in practical mean-variance portfolio construction.

1.2 Shrinkage estimators: parameter-focused and decision-focused approaches

Shrinkage estimation is one widely used approach for reducing the influence of estimation error in expected-return inputs. Jorion (1986) introduced the Bayes-Stein estimator, which shrinks the sample mean vector toward a stable target derived from the minimum-variance portfolio. Jorion (1991) later showed that Bayesian and equilibrium-based estimators can improve portfolio choice. Beyond Bayes-Stein shrinkage, other parameter-focused approaches have addressed uncertainty in portfolio inputs. Black & Litterman (1992) combined market-equilibrium returns with investor views, while Kan & Zhou (2007) derived portfolio rules that explicitly account for parameter uncertainty. A parallel literature, most prominently Ledoit & Wolf (2003, 2004), applies shrinkage to the covariance matrix, whose estimators improve the conditioning of the risk input. These contributions are parameter-focused in the sense that they aim to improve the statistical quality of the inputs before optimization takes place.

More recent research adopts a decision-focused perspective in which the quality of the final portfolio decision, is the criterion for selecting the estimator, rather than the

accuracy of the parameter estimate. Ortiz et al. (2025) proposed an optimal shrinkage of means within the Markowitz model, in which the shrinkage intensity is chosen to optimize the resulting portfolio decision. Importantly, the framework permits a shrinkage intensity of zero when introducing shrinkage does not improve the portfolio decision. Related developments include enhanced portfolio optimization that corrects noisy input directions (Pedersen et al., 2021), robust Bayesian portfolio optimization under ambiguity (Zapata Quimbayo et al., 2025), and forecast-based Markowitz implementations using time-series and machine learning models (Ślusarczyk & Ślepaczuk, 2025).

1.3 Constraints, comparative evidence, and the context of emerging markets

Empirical comparisons temper optimization expectations. DeMiguel et al. (2009) showed that, across the portfolio strategies and datasets they examined, optimized approaches did not consistently outperform the simple 1/N allocation rule out-of-sample. Brandt (2010) also highlighted the practical challenges involved in translating theoretically optimal portfolio rules into implemented investment strategies. Constraints themselves act as implicit shrinkage: Jagannathan & Ma (2003) showed that imposing no-short-sale constraints can improve performance even when the constraints are wrong in a statistical sense, and Clarke et al. (2006) documented the strong empirical performance of constrained minimum-variance portfolios.

Emerging-market portfolio optimization may face additional challenges arising from heterogeneous liquidity, trading costs, and price formation that differ substantially across stocks and over time (O'Hara, 1995; Hasbrouck, 2007; Harris, 2023), which affects both the quality of return estimates and the implementability of optimized portfolios. Taken together, the literature establishes strong theoretical motivations for shrinkage, but predominantly developed-market evidence on its realized benefits. This study contributes to filling this gap by comparing a parameter-focused estimator (Jorion) and a decision-focused estimator (Ortiz) with the traditional Markowitz benchmark on Indonesian individual stocks within a single, consistent out-of-sample design.

2. Methods

This study uses a quantitative design based on monthly stock price data for individual stocks listed on the Indonesia Stock Exchange. The observation period spans from January 2006 to December 2025. The long horizon is intended to capture different market conditions, including crisis, recovery, and normal market periods. The study also uses a monthly risk-free proxy to calculate excess returns, because portfolio optimization is evaluated relative to the compensation obtained above a relatively risk-free alternative.

The sample consists of 74 individual stocks selected using a balanced-panel approach: a stock is included only if it has complete and continuous month-end closing prices from January 2006 to December 2025. Stocks listed after the beginning of the observation period, delisted before its end, subject to prolonged trading suspensions, or containing missing price observations are excluded. This criterion keeps the asset dimension constant across all rolling windows, so that the estimators reflect the differences between methods rather than the changes in data availability. This study explicitly acknowledges that this design choice induces potential survivorship bias, because the sample is tilted toward firms that survived the entire period. Therefore, the interpretation of the results is restricted to the universe of continuously listed stocks. Because the same universe is used for all three methods, the bias affects the level of performance rather than the cross-method comparison, which is the object of the research questions.

The monthly risk-free rate is proxied by the deposit insurance rate (*Tingkat Bunga Penjaminan*) published by the Indonesia Deposit Insurance Corporation (*Lembaga Penjamin Simpanan*, LPS). First, deposits below the LPS ceiling rate are guaranteed by law, so the rate is default-free in the domestic legal sense, consistent with Sharpe (1964) and Bodie et al. (2021) definitions of a risk-free asset. Second, the LPS series is available

continuously at a monthly frequency since January 2006, whereas structured government bond yield data only became consistently available around late 2009 and remained sparse before 2012. Third, the LPS rate functions as a domestic lower bound for the riskless return available to ordinary investors. Accordingly, the reported Sharpe ratios should be read as performance relative to this domestic proxy rather than to a theoretical government-bill rate.

The main analytical variable is the excess return of individual stocks. Monthly stock returns are calculated from monthly price changes and then reduced by the monthly risk-free rate. This transformation is consistent with the use of the Sharpe ratio as the primary risk-adjusted performance measure (Sharpe, 1966). The monthly simple and excess returns are expressed as follows.

$$r_{i,t} = \frac{(P_{i,t} - P_{i,t-1})}{P_{i,t-1}} \quad (\text{Eq. 1})$$

$$x_{i,t} = r_{i,t} - r_{f,t} \quad (\text{Eq. 2})$$

Where $r_{i,t}$ is the simple return of stock in month t , $P_{i,t}$ and $P_{i,t-1}$ denote the corresponding month-end prices, $r_{f,t}$ is the monthly risk-free proxy, and $x_{i,t}$ is the monthly excess return used in portfolio estimation. Excess returns are used throughout the analysis to maintain consistency with the Sharpe-ratio objective, which evaluates portfolio return relative to the risk-free proxy per unit of total portfolio volatility (Sharpe, 1966). The use of excess return rather than raw return is intended to align the empirical analysis with the concept of risk compensation. In portfolio evaluation, an investor is not only interested in whether a portfolio produces a positive return, but also whether the return is sufficient relative to a low-risk alternative. By subtracting the monthly risk-free rate, the analysis focuses on the additional return generated by bearing equity risk. This treatment is also consistent with the Sharpe ratio, which measures the excess return per unit of total portfolio volatility.

Before the rolling-window procedure is implemented, the data are organized into a monthly return panel. Stocks with incomplete data over the required estimation horizon are carefully treated to ensure that a consistent set of observations within each rolling window is used in the optimization process. This is important because missing values can affect the mean returns and covariance matrix estimation. In portfolio optimization, inconsistent data treatment may produce weight differences caused by data availability rather than the optimization method itself. Therefore, maintaining consistency in the input data is necessary to make meaningful comparisons across Markowitz, Jorion, and Ortiz.

The portfolio formation procedure applies a 120-month rolling estimation window. The relevant expected return vector and covariance matrix are estimated for each window, portfolio weights are optimized under long-only constraints, and the resulting weights are applied to the next month to obtain out-of-sample performance. This process is repeated throughout the sample period to create a time series of realized out-of-sample portfolio returns. The Markowitz portfolio directly uses the historical sample mean as the expected return input. The Jorion portfolio applies a Bayes-Stein shrinkage mechanism that pulls the sample mean toward a more stable target before optimization (Jorion, 1986). The Ortiz portfolio applies a decision-focused shrinkage framework, where the optimal shrinkage intensity is chosen based on the portfolio decision rather than only the estimation accuracy of the mean parameter (Ortiz et al., 2025). To ensure comparability, all three approaches are implemented under the same return data, covariance estimation, and long-only optimization constraint.

Several implementation details are specified for reproducibility. The covariance matrix in every window is the sample covariance matrix of the in-window excess returns, and the same covariance estimate is used for all three methods so that performance differences can be attributed to the mean estimator alone; a small ridge term (10^{-8} on the diagonal) is added for numerical stability. Portfolio weights are obtained by maximizing

the in-sample Sharpe ratio subject to full-investment and long-only constraints, numerically solved with the Sequential Least Squares Programming (SLSQP) algorithm in SciPy (maximum 2,000 iterations, convergence tolerance 10^{-9}); if the solver fails to converge, an equal-weight allocation is used as a fallback. Portfolios are rebalanced monthly: the weights estimated at the end of each window are held for one month and applied to the following month's returns. Any month containing a missing observation is removed from the training data within each estimation window. If a stock has a missing return in the out-of-sample month, its weight is set to zero and the remaining weights are renormalized. Newey–West-adjusted standard errors are estimated using a lag length of three months to accommodate short-horizon serial dependence in the monthly pairwise-difference series.

The 120-month rolling estimation window was selected to provide a balance between statistical reliability and adaptability. A longer window provides more observations for estimating the expected returns and covariances, which can reduce the sampling error. However, using an excessively long window may incorporate outdated market conditions and reduce the model's ability to adapt to structural changes. A shorter window may be more responsive to recent market developments, but it may also produce unstable estimates due to limited observations. Therefore, the 120-month window represents a compromise by using ten years of historical monthly data to estimate parameters while still allowing the portfolio to be updated over time.

The long-only constraint is also important from an implementation perspective. Although unconstrained mean-variance optimization can generate mathematically optimal portfolios, it may produce negative weights that imply short-selling. Short-selling may be restricted, costly, or inconsistent with the mandate of ordinary investors in many practical investment settings. The long-only constraint makes the analysis more realistic for investors who allocate wealth across selected stocks without taking short positions. At the same time, the constraint may reduce the extremity of the optimized weights, which means that the empirical benefit of shrinkage must be evaluated under a more practical portfolio setting.

The shrinkage-based expected return input can be summarized as follows: Equation 3 represents the parameter-focused Bayes–Stein logic in Jorion, while Equations 4 and 5 represent the decision-focused logic in Ortiz, where the shrinkage intensity is selected based on the quality of the resulting portfolio decision rather than the parameter estimate alone.

$$\tilde{\mu}^{\text{Jorion}} = (1 - \delta_j) \hat{\mu} + \delta_j \mu_0 \quad (\text{Eq. 3})$$

$$\tilde{\mu}^{\text{Ortiz}}(\delta) = (1 - \delta) \hat{\mu} + \delta \bar{\mu} \quad (\text{Eq. 4})$$

$$\delta^* = \arg \min_{\delta} L(w(\tilde{\mu}(\delta)), w_{\text{ref}}) \quad (\text{Eq. 5})$$

Performance is evaluated using the mean monthly out-of-sample return, monthly standard deviation, monthly Sharpe ratio, and annualized Sharpe ratio. The annualized Sharpe ratio is calculated as the monthly Sharpe ratio multiplied by . The weight stability is evaluated using average standard deviation of the weights and the average turnover. Pairwise statistical comparisons are conducted on the monthly out-of-sample return differences and monthly turnover differences between portfolio methods. Because these difference series may exhibit heteroskedasticity and serial dependence, particularly when portfolios are formed from overlapping rolling estimation windows, Newey–West-adjusted standard errors are used with a lag length of three months, following Newey & West (1987). The resulting tests distinguish descriptive numerical differences from differences supported by statistical inference.

The choice of evaluation metrics directly follows the two research questions. The Sharpe ratio is used because the optimization objective itself is the maximization of risk-adjusted excess return. Therefore, the out-of-sample Sharpe ratio measures the quantity the methods are designed to deliver (Sharpe, 1966). Average weight volatility captures how much the allocation to each stock fluctuates across rebalancing dates, while average turnover, which is defined as the sum of absolute weight changes between consecutive rebalancing dates, approximates trading intensity and implicit implementation cost of each method. Together, these metrics evaluate both the quality and the implementability of the portfolio decision, which is the appropriate standard for comparing parameter-focused and decision-focused estimators.

The statistical testing procedure is included because numerical differences in performance metrics may not necessarily indicate meaningful superiority. Portfolio returns are time-series data, and the differences between methods may exhibit autocorrelation or heteroskedasticity, particularly when overlapping rolling windows are used to form portfolios. The Newey-West adjustment is therefore used to obtain standard errors that are more robust to these time-series properties. This allows the study to distinguish between merely descriptive and statistically supported differences.

The analysis also separates performance and stability because a portfolio method may look attractive from a return perspective but still be difficult to implement. Turnover is included alongside return-based metrics to capture the implementation intensity associated with each portfolio strategy. Therefore, from an investor's perspective, a method that improves the Sharpe ratio but produces excessive turnover may not necessarily be superior. This study adopts a more comprehensive view of portfolio performance by evaluating both Sharpe ratio and turnover.

Table 1. Research design and operationalization

Component	Operationalization
Data	Monthly individual stock prices listed on the Indonesia Stock Exchange from January 2006 to December 2025.
Sample	Balanced panel of 74 individual stocks with complete monthly closing prices over the full period; stocks with later listing, delisting, prolonged suspension, or missing data are excluded.
Risk-free proxy	The monthly deposit insurance rate (<i>Tingkat Bunga Penjaminan</i>) of the Indonesia Deposit Insurance Corporation (LPS).
Covariance estimator	The sample covariance of in-window excess returns, identical across methods, with a 10^{-8} diagonal ridge for numerical stability.
Optimizer and rebalancing	Maximization of the in-sample Sharpe ratio under fully invested, long-only, no-leverage constraints, solved with SLSQP (SciPy); monthly rebalancing; equal-weight fallback if the solver fails.
Main variable	Monthly excess return, calculated as stock return minus monthly risk-free return.
Estimation window	Rolling the 120-month in-sample window.
Portfolio constraints	Long-only weights, fully invested, and no leverage.
Compared methods	Traditional Markowitz, Jorion Bayes-Stein shrinkage, and Ortiz optimal shrinkage of means.
Performance metrics	Mean monthly return, standard deviation, monthly Sharpe ratio, and annualized Sharpe ratio.
Stability metrics	The average standard deviation of weights and average turnover.
Statistical test	Newey-West adjusted test of the mean differences.

3. Results and Discussion

3.1 Descriptive pattern of the excess return data

The descriptive analysis indicates substantial cross-sectional variation in the excess returns of individual Indonesian stocks. Average excess returns and volatility vary across

the sampled stocks, indicating considerable cross-sectional differences in their return-risk profiles. This cross-sectional dispersion is relevant to portfolio optimization because differences in estimated returns and risks can materially influence the allocation of portfolio weights across stocks.

The time-series dimension of the data also shows strong market instability. The lowest cross-sectional mean excess return occurred in September 2008, while February 2020 also recorded a pronounced negative cross-sectional mean during the onset of the COVID-19 market shock. These findings imply that the optimization input is not stable over time. Therefore, the study context is appropriate for evaluating whether shrinkage methods can improve portfolio performance and weight stability. Table 2 reports the months with the lowest cross-sectional mean excess return, documenting the episodes of market stress referred to above. The most negative cross-sectional mean occurred in September 2008 during the global financial crisis (-0.2542), followed in magnitude by February 2020 at the onset of the COVID-19 market shock (-0.1370). The concentration of large negative cross-sectional returns in distinct stress periods illustrates the presence of substantial downside variation within the historical sample used for portfolio estimation.

Table 2. Months with the lowest cross-sectional mean excess return

Month	Cross-sectional Mean	Cross-sectional Median	Cross-sectional SD
Sep 2008	-0.2542	-0.1500	0.2758
Feb 2020	-0.1370	-0.0900	0.2327
Aug 2008	-0.1087	-0.0700	0.1706
Aug 2011	-0.1021	-0.0800	0.1182
Mar 2024	-0.0960	-0.0300	0.2751

Note: Values are monthly excess returns in decimal form over the LPS risk-free proxy, computed each month across the 74 sample stocks.

The heterogeneity of excess returns provides an important explanation for the difficulty of mean estimation in this sample. Stocks with very high average returns may appear attractive during the estimation period, but such performance may be driven by a few extreme monthly observations rather than persistent return characteristics. Conversely, stocks with low or negative average returns may not always remain unattractive in subsequent periods. This condition creates a challenge for any optimization method that relies on historical means, because the optimizer must use limited historical information to distinguish between signal and noise.

The descriptive results further indicate that the distribution of individual-stock excess returns varies over time. The substantial differences observed across stress periods indicate that the return-generating environment is unlikely to be constant throughout the 2006–2025 sample. Accordingly, the rolling-window design allows expected returns and covariances to be updated as the historical information set changes, rather than imposing one parameter estimate over the full sample. This design enables the subsequent analysis to compare how the three portfolio methods perform and how their allocations evolve as the underlying information set changes.

3.2 Out-of-sample portfolio performance

The out-of-sample backtest produces 119 monthly portfolio return observations. The comparison shows that the traditional Markowitz and Ortiz methods generate the same mean monthly return, standard deviation, and Sharpe ratio. Consequently, under the decision criterion applied in this sample, introducing positive shrinkage did not improve the portfolio solution relative to the no-shrinkage case.

Jorion's parameter-focused shrinkage approach produced numerically lower out-of-sample performance metrics than Markowitz and Ortiz. Markowitz and Ortiz recorded a mean monthly return of 0.0010, a monthly standard deviation of 0.0497, a monthly Sharpe ratio of 0.0206, and an annualized Sharpe ratio of 0.0713. Jorion records a mean monthly

return of 0.0009, a monthly standard deviation of 0.0511, a monthly Sharpe ratio of 0.0176, and an annualized Sharpe ratio of 0.0608.

The identical results between Markowitz and Ortiz are particularly informative because they show that the decision-focused shrinkage mechanism does not mechanically impose shrinkage. Instead, the method evaluates whether shrinkage improves the resulting portfolio decision. In this sample, the optimal shrinkage intensity is zero, indicating that the portfolio selected by the Ortiz framework coincides with the traditional Markowitz portfolio. This result is important because it prevents an overly simplistic interpretation that shrinkage should always be positive or that a more advanced method must always generate different weights.

The numerically lower Jorion performance can be interpreted in terms of the conceptual distinction between parameter regularization and realized portfolio decision quality. A parameter-focused shrinkage estimator may reduce the variance of the expected return estimate, however, this does not guarantee that the final optimized portfolio will produce better realized returns. One possible explanation is that shrinkage toward the selected target may alter portfolio allocations without improving their alignment with subsequent realized returns. Therefore, the empirical results support the view that portfolio models should be evaluated based on realized portfolio outcomes rather than only on the theoretical attractiveness of the estimator.

Table 3. Out-of-sample portfolio performance

Method	Mean Monthly Return	SD Monthly Return	Sharpe OOS Monthly	Sharpe OOS Annual
Markowitz	0.0010	0.0497	0.0206	0.0713
Jorion	0.0009	0.0511	0.0176	0.0608
Ortiz	0.0010	0.0497	0.0206	0.0713

Note: Returns are monthly excess returns in decimal form over the LPS risk-free proxy; OOS = out-of-sample; Sharpe OOS Annual = monthly Sharpe ratio multiplied by $\sqrt{12}$; portfolios are formed from 74 stocks; 119 monthly observations.

The Newey–West tests detected no statistically significant differences in mean monthly out-of-sample returns among the three pairwise method comparisons. The comparison between Markowitz and Jorion has a p-value of 0.7798, while Markowitz versus Ortiz and Jorion versus Ortiz have p-values of 0.7243 and 0.7798, respectively. These tests provide no evidence of statistically significant differences in mean monthly out-of-sample returns across the three methods.

From an economic perspective, the differences in the Sharpe ratio among the three methods are small. The annualized Sharpe ratio of Markowitz and Ortiz is 0.0713, while Jorion records 0.0608. Although Markowitz and Ortiz appear numerically higher, the difference is insufficient to establish practical dominance. This implies that the choice among these methods should not be based only on the reported Sharpe ratio for investors. Other considerations may also be relevant, such as model transparency, ease of implementation, computational complexity, and robustness across alternative samples. All three methods produced annualized Sharpe ratios below 0.08, indicating that realized excess returns were small relative to portfolio volatility over the out-of-sample evaluation period. Before transaction costs, all three long-only portfolio approaches generated low risk-adjusted excess returns regardless of the mean estimator used.

The absence of statistically significant pairwise differences in mean monthly out-of-sample returns suggests that the observed numerical ranking should be interpreted cautiously. Small differences in portfolio backtesting can arise from particular market episodes, selected stocks, or the timing of rebalancing. Without statistical support, it would be risky to conclude that one approach consistently outperforms the others. The Newey–West test therefore acts as a safeguard against overinterpreting small numerical differences.

Several design features improve the internal comparability of the three portfolio methods. The inference relies on Newey-West heteroskedasticity- and autocorrelation-consistent standard errors, which protect the conclusions against the serial dependence induced by overlapping estimation windows. The use of an identical covariance estimator, identical constraints, and an identical asset universe across the three methods isolates the effect of the mean estimator, thus the null result cannot be attributed to uncontrolled implementation differences. The sensitivity of the conclusions to the estimation window length, sub-period analysis, and alternative shrinkage targets remains to be examined; these extensions are identified as robustness checks for future work in the concluding section.

Table 4. Newey-West test for out-of-sample performance differences

Comparison	Observations	Mean Diff.	NW SE	<i>t</i> Stat.	<i>p</i> -value
Markowitz vs Jorion	119	0.0001	0.0005	0.2796	0.7798
Markowitz vs Ortiz	119	0.0000	0.0000	-0.3527	0.7243
Jorion vs Ortiz	119	-0.0001	0.0005	-0.2796	0.7798

Note: Mean Diff. = mean of the monthly out-of-sample return difference between the two methods (decimal); NW SE = Newey-West heteroskedasticity- and autocorrelation-consistent standard error (lag = 3); OOS = out-of-sample.

3.3 Portfolio weight stability

The stability analysis provides a similar conclusion. Markowitz and Ortiz produced identical average standard deviations of weights and average turnover. Jorion produces slightly higher variability and turnover. Thus, in this sample, Jorion's parameter-focused shrinkage did not produce lower observed weight variability or turnover than the Markowitz benchmark. Theoretically, shrinkage should reduce sensitivity to noisy sample means, but its practical effect depends on the relationship between the shrinkage target, the covariance structure, and the optimization constraint.

The average turnover is economically important because it approximates the portfolio rebalancing intensity. Although transaction costs are not included in the return calculation, higher turnover may reduce net performance in real-world implementation. Therefore, the stability result reinforces the interpretation that a more complex estimator should be evaluated not only from pre-cost return metrics but also from implementability. This implementation consideration may be particularly relevant for individual-stock portfolios because heterogeneous liquidity can cause the trading cost associated with a given level of turnover to vary across securities. The stability results are relevant because optimized portfolios are often criticized for producing weights that change too frequently. Even when transaction costs are excluded from the formal calculation, turnover remains a useful proxy for implementation difficulty. A portfolio with high turnover may require frequent trading, which can reduce net return once brokerage fees, bid-ask spreads, taxes, and market impact are considered. For individual stocks in an emerging market, this issue can be more important because liquidity conditions differ across stocks and trading costs may not be uniform.

This suggests that the expected return input shrinkage does not necessarily reduce the rebalancing intensity. The expected return vector determines the final portfolio weights, as well as the covariance matrix and optimization constraints. Consequently, even if shrinkage stabilizes the estimated mean vector, portfolio weights may still vary across rolling windows as expected-return estimates and covariance structures evolve under the imposed constraints. This reinforces the importance of empirically evaluating stability rather than assuming that shrinkage automatically leads to more implementable portfolios.

Table 5. Portfolio weight stability and rebalancing intensity

Method	Average SD of Weights	Average Turnover
Markowitz	0.0154	0.1824
Jorion	0.0158	0.1882
Ortiz	0.0154	0.1824

Note: Average SD of Weights = standard deviation of each stock's weight across rebalancing dates, averaged over the 74 stocks; Average Turnover = average of the sum of absolute weight changes between consecutive rebalancing dates (decimal per month).

The Newey–West tests detected no statistically significant pairwise differences in mean monthly turnover among the three methods. The p-values were 0.1231 for Markowitz versus Jorion, 0.7164 for Markowitz versus Ortiz, and 0.1231 for Jorion versus Ortiz. Therefore, the evidence does not support a statistically significant difference in portfolio stability among the three approaches.

Table 6. Newey–West test of pairwise mean monthly turnover differences

Comparison	Observations	Mean Diff.	NW SE	<i>t</i> Stat.	p-value
Markowitz vs Jorion	119	-0.0058	0.0038	-1.5419	0.1231
Markowitz vs Ortiz	119	0.0000	0.0000	0.3633	0.7164
Jorion vs Ortiz	119	0.0058	0.0038	1.5419	0.1231

Note: Differences are computed on the monthly turnover series of each pair of methods; NW SE = Newey–West standard error (lag = 3).

3.4 Discussion

The main empirical finding is that the Ortiz approach selects a zero shrinkage intensity throughout the tested rolling windows. This is an important result because it shows that a decision-focused method may rationally choose not to shrink when shrinkage does not improve the portfolio decision. Accordingly, the zero-shrinkage outcome should be interpreted as an admissible solution of the decision-focused framework rather than as a failure of the method. Rather, it suggests that the optimal decision under the specified constraints for this Indonesian stock sample coincides with the traditional Markowitz allocation.

This result also illustrates that the statistical rationale for shrinkage does not necessarily translate into superior realized portfolio outcomes. In parameter estimation, shrinkage can reduce variance by moving noisy estimates toward a stable target. However, portfolio optimization is a downstream decision-making problem. Even if shrinkage improves the statistical properties of the estimated mean vector, it may not improve realized portfolio performance when the resulting allocations do not better align with subsequent return-risk outcomes under the imposed constraints. This explains why Jorion's estimator can be theoretically attractive but still generate slightly weaker out-of-sample performance in this application.

These findings are consistent with the broader literature emphasizing that optimized portfolios are highly sensitive to inputs and that empirical superiority must be tested out-of-sample rather than assumed from theoretical elegance alone (Michaud, 1989; DeMiguel et al., 2009). The implication for practitioners is that model complexity should be justified by measurable improvements in performance, stability, or implementation cost. In this study, the more complex shrinkage approaches do not significantly improve the traditional Markowitz benchmark.

The findings have several implications for Indonesian portfolio management practice. First, portfolio managers should be cautious when assuming that more sophisticated estimation methods will automatically improve investment outcomes. Model sophistication must be supported by empirical validation, particularly through out-of-sample testing. Second, long-only constraints may themselves limit extreme allocations

generated from noisy estimates, which could reduce the incremental effect of shrinkage on portfolio weights. Third, the results suggest that model selection should consider not only theoretical elegance but also whether the method produces measurable improvements under realistic investment constraints.

The results also have implications for future portfolio optimization research in emerging markets. One possible explanation for the limited improvement from mean shrinkage is that the main source of instability may not be only expected return estimation. Covariance estimation, liquidity constraints, transaction costs, and regime shifts may also affect realized portfolio performance. Therefore, future research can extend this study by combining mean shrinkage with covariance shrinkage, robust optimization, or transaction-cost-aware optimization. Such extensions may provide a more complete assessment of whether advanced portfolio methods can improve Indonesian market performance.

In addition, the use of individual stocks creates both strength and limitation. It provides a direct test of portfolio optimization at the asset level and exposes the analysis to stock-specific noise and survivorship considerations. The results may be affected by survivorship bias if the sample only includes stocks with complete historical data because firms with incomplete data or delisting events are excluded. Future studies may address this issue by using an unbalanced panel or by explicitly incorporating delisted stocks where data are available. This would help determine whether the conclusions remain robust when the investable universe changes over time. The implications for practitioners differ by investor type. For institutional applications, the constrained Markowitz solution may serve as a transparent benchmark against which additional modeling complexity can be evaluated. Whether covariance shrinkage, transaction-cost-aware rebalancing, or other extensions provide greater incremental benefits should be tested directly in future analyses. For retail applications, the low reported Sharpe ratios indicate that the optimized portfolios generated limited risk-adjusted excess returns before implementation costs. However, because the study does not include a simple diversified benchmark or explicit transaction-cost model, no direct conclusion can be drawn regarding their relative advantage over lower-complexity allocation strategies.

Finally, the three limitations qualify the findings. First, the balanced-panel requirement introduces survivorship bias, which may affect estimated performance levels and limits the generalizability of the results to a broader, time-varying investable universe. Using the same stock universe across methods improves internal comparability, but does not eliminate this limitation. Second, transaction costs, bid-ask spreads, taxes, and market impact are excluded from the analysis. Because turnover differs descriptively across methods, incorporating explicit implementation costs could alter their net performance rankings and should be examined directly. Third, the results are conditional on a single estimation window length and a single shrinkage target; therefore sensitivity analyses across window lengths, sub-periods, and alternative targets are a natural robustness agenda for subsequent work.

4. Conclusions

This study evaluates the performance and stability of individual stock portfolios on the Indonesia Stock Exchange using Markowitz, Jorion, and Ortiz approaches over January 2006-December 2025. The results show that Ortiz's decision-focused optimal shrinkage of means consistently chooses zero shrinkage, making its portfolio allocation and performance identical to traditional Markowitz's. Jorion's parameter-focused shrinkage approach records slightly lower out-of-sample performance and slightly higher turnover, but these differences are not statistically significant based on Newey-West tests. The balanced-panel design introduces survivorship bias and therefore limits the generalizability of the findings beyond continuously listed stocks. In addition, all three methods produced annualized Sharpe ratios below 0.08 before transaction costs, indicating low realized risk-adjusted excess returns over the evaluated sample. The main

contribution of this study is to show that, in the context of the Indonesian stock market, more advanced mean-shrinkage methods do not automatically outperform a constrained Markowitz benchmark.

The findings highlight the need for disciplined out-of-sample testing, transparent model validation, and careful consideration of implementation costs before adopting more complex portfolio optimization methods. In practical terms, the constrained Markowitz solution can serve as a transparent benchmark against which additional estimator complexity is evaluated. However, because the study does not include simpler diversified benchmark portfolios or explicit transaction-cost modeling, the relative attractiveness of the optimized strategies for different investor types cannot be determined from the present results. Future research could extend the analysis by incorporating transaction costs, testing alternative risk-free proxies and estimation-window lengths, applying covariance shrinkage, and using a time-varying or unbalanced investable universe to address potential survivorship bias.

Overall, the study emphasizes that the value of a portfolio optimization model should be assessed through its decision outcome. A method that is theoretically more advanced is not necessarily superior if it does not improve out-of-sample performance or implementation stability. In the context of the Indonesian stock market, the evidence suggests that constrained Markowitz optimization remains a relevant benchmark and that shrinkage-based improvements require careful validation before being adopted in practice. This conclusion does not reject the usefulness of shrinkage methods; however, it highlights that their effectiveness depends on the empirical setting, the selected target, the optimization constraint, and the market environment in which the portfolio is implemented.

Acknowledgement

The author expresses sincere appreciation to the Faculty of Economics and Business, Universitas Indonesia, for the academic environment, resources, and support that made the completion of this research possible. The author also gratefully acknowledges PT Reasuransi Indonesia Utama (Indonesia Re) for its support during the preparation of this manuscript. The Indonesia Stock Exchange (IDX) and the Indonesia Deposit Insurance Corporation (*Lembaga Penjamin Simpanan*, LPS) are gratefully acknowledged for making the underlying market and interest rate data publicly available for this study. On a personal note, this research was carried out alongside full-time work and pregnancy, and the author wishes to thank her husband, Rafif Hibatullah, for his unwavering support and encouragement throughout this journey, and her unborn child, who accompanied every stage of this work.

Author Contribution

Conceptualization, E.D.M. and R.R.; Methodology, E.D.M.; Software, E.D.M.; Validation, E.D.M. and R.R.; Formal Analysis, E.D.M.; Investigation, E.D.M.; Resources, E.D.M.; Data Curation, E.D.M.; Writing - Original Draft Preparation, E.D.M.; Writing - Review and Editing, E.D.M. and R.R.; Visualization, E.D.M.; Supervision, R.R.; Project Administration, E.D.M.

Funding

This research received no external funding.

Ethical Review Board Statement

Not available. This study does not involve human participants, animals, public health intervention, or personally identifiable data.

Informed Consent Statement

Not available.

Data Availability Statement

The study uses publicly available market data and processed research datasets. The processed data and code may be made available by the author upon reasonable request, subject to data provider terms and academic requirements.

Conflicts of Interest

The author declares no conflict of interest.

Declaration of Generative AI Use

During the preparation of this work, the authors used Gemini to assist in improving syntax code. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Open Access

©2026. The authors. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit: <http://creativecommons.org/licenses/by/4.0/>

References

- Best, M. J., & Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies*, 4(2), 315–342. <https://doi.org/10.1093/rfs/4.2.315>
- Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5), 28–43. <https://doi.org/10.2469/faj.v48.n5.28>
- Bodie, Z., Kane, A., & Marcus, A. J. (2021). *Investments* (12th ed.). McGraw Hill.
- Brandt, M. W. (2010). Portfolio choice problems. In Y. Aït-Sahalia & L. P. Hansen (Eds.), *Handbook of financial econometrics: Tools and techniques* (Vol. 1, pp. 269–336). North-Holland.
- Chopra, V. K., & Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2), 6–11. <https://doi.org/10.3905/jpm.1993.409440>
- Clarke, R., de Silva, H., & Thorley, S. (2006). Minimum-variance portfolios in the U.S. equity market. *The Journal of Portfolio Management*, 33(1), 10–24. <https://doi.org/10.3905/jpm.2006.661366>
- DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, 22(5), 1915–1953. <https://doi.org/10.1093/rfs/hhm075>
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417. <https://doi.org/10.2307/2325486>
- Green, R. C., & Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? *The Journal of Finance*, 47(5), 1785–1809. <https://doi.org/10.1111/j.1540-6261.1992.tb04683.x>
- Harris, L. (2023). *Trading and exchanges: Market microstructure for practitioners*. Oxford University Press. <https://doi.org/10.1093/oso/9780195144703.001.0001>
- Hasbrouck, J. (2007). *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*. Oxford University Press.

- Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4), 1651–1683. <https://doi.org/10.1111/1540-6261.00580>
- Jobson, J. D., & Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371), 544–554. <https://doi.org/10.1080/01621459.1980.10477507>
- Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3), 279–292. <https://doi.org/10.2307/2331042>
- Jorion, P. (1991). Bayesian and CAPM estimators of the means: Implications for portfolio selection. *Journal of Banking & Finance*, 15(3), 717–727. [https://doi.org/10.1016/0378-4266\(91\)90094-3](https://doi.org/10.1016/0378-4266(91)90094-3)
- Kan, R., & Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3), 621–656. <https://doi.org/10.1017/S0022109000004129>
- Kustodian Sentral Efek Indonesia. (2023). *Statistik pasar modal Indonesia Februari 2023*. Kustodian Sentral Efek Indonesia.
- Kustodian Sentral Efek Indonesia. (2024). *Statistik pasar modal Indonesia 2024*. Kustodian Sentral Efek Indonesia.
- Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5), 603–621. [https://doi.org/10.1016/S0927-5398\(03\)00007-0](https://doi.org/10.1016/S0927-5398(03)00007-0)
- Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4), 110–119. <https://doi.org/10.3905/jpm.2004.110>
- Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15(5), 850–859. <https://doi.org/10.1016/j.jempfin.2008.03.002>
- Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1), 13–37. <https://doi.org/10.2307/1924119>
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- Michaud, R. O. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45(1), 31–42. <https://doi.org/10.2469/faj.v45.n1.31>
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708. <https://doi.org/10.2307/1913610>
- O'Hara, M. (1995). *Market microstructure theory*. Blackwell.
- Ortiz, R., Contreras, M., & Mellado, C. (2025). Optimal shrinkage of means in the Markowitz model. *International Review of Financial Analysis*, 104, Article 104136. <https://doi.org/10.1016/j.irfa.2025.104136>
- Otoritas Jasa Keuangan. (2020). *Laporan triwulan I-2020*. Otoritas Jasa Keuangan.
- Pedersen, L. H., Babu, A., & Levine, A. (2021). Enhanced portfolio optimization. *Financial Analysts Journal*, 77(2), 124–151. <https://doi.org/10.1080/0015198X.2020.1854543>
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425–442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>
- Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(1), 119–138. <https://doi.org/10.1086/294846>
- Ślusarczyk, D., & Ślepaczuk, R. (2025). Optimal Markowitz portfolio using returns forecasted with time series and machine learning models. *Journal of Big Data*, 12, Article 127. <https://doi.org/10.1186/s40537-025-01164-z>
- Zapata Quimbayo, C. A., Carmona Espejo, D. F., & Gamboa Hidalgo, J. (2025). Robust Bayesian portfolio optimization. *International Review of Financial Analysis*, 103, Article 104215. <https://doi.org/10.1016/j.irfa.2025.104215>

Biographies of Authors

Enggal Dwi Mulyaningtyas, graduate student in Financial Management at Universitas Indonesia and an actuary at PT Reasuransi Indonesia Utama (Persero). Her work focuses on actuarial modeling, insurance reserving, and IFRS 17 implementation. She has experience in data analytics, financial projections, and risk assessment for insurance and reinsurance portfolios.

- Email: enggal.dwi@ui.ac.id
- ORCID: N/A
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: N/A

Rofikoh Rokhim, Professor of Management, Banking, and Finance at Universitas Indonesia. She earned her PhD in Economics from Université de Paris I Panthéon-Sorbonne, focusing on macro and microeconomic studies applied to finance and development. Her research covers banking, finance, sustainability, microfinance, and economic policy. She has published extensively in national and international journals and serves as a consultant for financial institutions.

- Email: rofikoh.rokhim@ui.ac.id
- ORCID: 0000-0002-2847-0552
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: <https://www.ui.ac.id/staff/rofikoh-rokhim>