International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 05 May 2025
DOI: 10.47191/ijcsrr/V8-i5-69, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

Evaluating Validity of Expected Shortfall Estimation Method in Measuring Market Risk for Bank's Trading Book Position

Kresna Nuswantoro[1], Rofikoh Rokhim[2]
[1,2] Master of Management, Faculty of Economics and Business, Universitas Indonesia

ABSTRACT: Expected Shortfall (ES) has become a pivotal risk measure in financial regulation, particularly under the Basel III Fundamental Review of the Trading Book (FRTB), which replaces Value-at-Risk (VaR) due to ES's ability to capture tail risk better. This study investigates the performance of two ES estimation methods—GARCH-t and Historical Simulation (HS)—in the context of foreign exchange (FX) against the IDR currency. Unlike VaR, ES incorporates the magnitude of extreme losses and is thus more sensitive to volatility dynamics. However, each estimation method presents trade-offs between responsiveness, robustness, and capital allocation. Using exchange rate data from 2007 to 2024, the ES values were computed under each method and validated through Acerbi–Szekely Z-tests on both tails. The backtesting results reveal that GARCH-t performs best during stress periods but demonstrates instability in calm markets. In contrast, Basic HS demonstrates more consistent backtest in the overall performance score across years. These findings offer practical insights into ES model implementation, emphasizing the importance of model selection, dual-tail backtesting, and supervisory alignment with FRTB.

KEYWORDS: Backtesting, Expected Shortfall, FRTB, Foreign Exchange, Market Risk.

1. INTRODUCTION

Market risk is a primary concern for banks and financial institutions, directly affecting their stability, profitability, and regulatory capital requirements (Jorion, 2007). It arises from fluctuations in financial asset prices, including foreign exchange (FX) rates, interest rates, equity prices, and commodity prices (Basel Committee on Banking Supervision, 2019). Among these, FX risk is particularly significant due to its high volatility and sensitivity to global macroeconomic shifts, monetary policies, and geopolitical uncertainty (Bekaert & Hodrick, 2011).

Traditionally, banks have used Value-at-Risk (VaR) as the standard metric for quantifying market risk. VaR estimates the potential loss in a portfolio over a given time horizon at a specified confidence level (Greuning & Bratanovic, 2020). However, it has notable limitations, particularly its inability to capture tail risk—extreme losses that occur beyond the VaR threshold (Yamai & Yoshiba, 2005). These limitations became evident during the 2008 Global Financial Crisis, where many institutions underestimated their actual exposures (Crouhy et al., 2014).

To address these shortcomings, the Basel Committee introduced the Fundamental Review of the Trading Book (FRTB), which replaces VaR with Expected Shortfall (ES) as the preferred measure for market risk under Basel III. Unlike VaR, which only estimates a cut-off loss threshold, ES measures the average expected loss beyond that threshold, making it more effective in capturing extreme market downturns (Basel Committee on Banking Supervision, 2019). While the shift is expected to improve financial resilience, implementing ES poses practical challenges—particularly in selecting accurate estimation models that prevent both risk underestimation and excessive capital conservatism (Wessels & Vuuren, 2023).

Multiple models have been developed to estimate ES. Parametric approaches like GARCH (Bollerslev,1986), aim to capture volatility dynamics. Yet, no consensus exists on the most reliable method, especially in the context of FX portfolios in emerging markets with elevated volatility (Righi & Ceretta, 2015). ES estimation is further complicated by model risk, where specification and estimation errors affect capital accuracy (Lazar & Zhang, 2019). Despite improved backtesting methods, forecast results still vary widely depending on model structure and assumptions (Taylor, 2020)

The transition from VaR to ES also impacts capital planning and compliance. A well-calibrated ES model can balance risk coverage and capital efficiency (Basel Committee on Banking Supervision, 2019). A well-calibrated ES model ensures that banks maintain adequate reserves while optimizing capital efficiency (Wessels & Vuuren, 2023). FRTB impact assessments suggest capital requirements for FX risk may increase by over 100% on average (PWC, 2016). As such, selecting an effective ES estimation model is vital for aligning regulatory compliance with operational efficiency requirements (Grajales & Medina Hurtado, 2023).

Recent studies provide mixed findings on model performance. GARCH models with heavy tails outperform simpler methods in capturing FX risk (Afuecheta et al., 2024). Historical Simulation tends to underestimate ES, especially with limited data (García-Risueño, 2025). (Lyócsa et al. (2024) emphasize that incorporating implied volatility enhances ES accuracy. Grajales & Medina Hurtado (2023) note that ES models under FRTB significantly elevate capital for FX options and advocate harmonizing internal and standardized approaches.

Although several studies have explored FX volatility using GARCH-type models in emerging markets (Epaphra, 2017; Nugroho & Susanto, 2017), few have addressed the direct estimation of Expected Shortfall (ES). Existing ES research often emphasizes developed markets or equity portfolios. Moreover, while Z2 backtesting for ES has been used (Lazar & Zhang, 2019; Righi & Ceretta, 2015), its application to FX portfolios remains limited. This research fills that gap by empirically evaluating and backtesting two ES models—GARCH based model and Historical Simulation—specifically for Indonesian FX portfolios. The inclusion of the Acerbi & Székely (2014) Z2 framework enhances model validation by directly testing ES forecasts, offering stronger regulatory alignment than traditional VaR-based backtests.

The objective of this research is to identify the most effective method for estimating Expected Shortfall (ES) in measuring market risk within bank trading books, specifically for foreign exchange (FX) portfolios in emerging markets. The study evaluates the accuracy and reliability of three ES models using robust backtesting techniques. It aims to prevent both underestimation—which can lead to regulatory non-compliance due to insufficient capital coverage—and overestimation, which may result in inefficient capital allocation. The findings are intended to guide academics, practitioners, and regulators—supporting Basel III FRTB and OJK requirements and enhancing risk management in Indonesia's banking sector.

2. LITERATURE REVIEW

Building on the shift from VaR to ES established under FRTB (Basel Committee on Banking Supervision, 2019), recent research has focused on evaluating various estimation models suitable for regulatory implementation. Expected Shortfall (ES) is recognized for its coherence, subadditivity, and superior tail risk sensitivity compared to VaR. Despite its theoretical appeal, ES is non-elicitable, which complicates direct backtesting. Regulatory frameworks like Basel III acknowledge this by allowing indirect testing via VaR or using specialized frameworks such as Acerbi and Szekely's (2014) backtesting approach. The shift toward ES emphasizes the need for models that are both accurate and practically implementable.

Parametric models, particularly the GARCH family introduced by Bollerslev (1986), remain widely used for capturing time-varying volatility in financial markets. Afuecheta et al. (2024) found that GARCH models with Student's t distribution and dynamic correlations provide accurate ES forecasts for major African currencies. Lyócsa et al. (2024) also found enhanced ES performance when implied volatility was integrated into EGARCH models. These findings support GARCH's strength in modeling FX volatility in emerging markets.

Historical Simulation (HS) is a non-parametric approach that estimates ES directly from past returns without assuming distributional properties. While easy to implement, HS often underestimates tail risk during volatile periods due to its backward-looking nature. García-Risueño (2025) and Righi & Ceretta (2015) demonstrated that HS typically performs poorly under stress, requiring adjustment or augmentation to meet Basel expectations.

Backtesting ES models remains challenging due to its non-elicitability. While Basel primarily prescribes backtesting VaR at the 99% level, researchers now advocate direct ES evaluation using the Magnitude Test (Z2) and the Acerbi–Szekely framework (2014). Johansson & Ludolphy (2020); Lazar & Zhang (2019) and Schmutz & Schneider (2023) applied these techniques to GARCH and VWHS models, finding them more compliant under regulatory traffic-light systems than HS or normal distribution assumptions.

Empirical applications in FX and emerging markets are growing but still limited in scope. Nugroho & Susanto (2017) used APARCH with Student's t distribution to model IDR/USD and JPY/IDR volatility, while Epaphra (2017) applied GARCH and EGARCH to African currencies. These studies emphasized volatility modelling but did not evaluate ES estimation. Meanwhile, Grajales & Medina Hurtado (2023) revealed that adopting FRTB-compliant ES and the Sensitivities-Based Method significantly increases capital requirements for FX options under stress scenarios.

Although many studies have explored risk modelling techniques for FX markets, few have directly compared GARCH-t and HS for ES estimation within an emerging market framework. Even fewer applied rigorous backtesting such as the Z2 test to FX portfolios. This study addresses these gaps by evaluating the comparative performance of three ES models on Indonesian FX data, contributing to more robust risk assessment practices under Basel III and OJK standards.

3. RESEARCH METHOD

3.1 Overview of Downside Risk Estimation and Approaches

This study evaluates three commonly used methods for estimating Expected Shortfall (ES)—a downside risk measure adopted under Basel III's Fundamental Review of the Trading Book (FRTB)—to assess market risk in foreign exchange (FX) portfolios. The models include: (i) the GARCH model, a parametric method that captures time-varying volatility and fat tails; and (ii) the Historical Simulation (HS) method, which relies on empirical return distributions. These approaches are examined for their effectiveness in capturing tail risk, in line with regulatory expectations under FRTB.

3.2 GARCH Model

The parametric Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, introduced by Bollerslev (1986), is applied with a student's t-distribution to accommodate heavy tails in FX return data. The model accounts for volatility clustering observed in financial markets. The conditional variance equation is given by:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \quad (1)$$

The returns are assumed to follow:

$$R_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim t(0, \sigma_t^2, v) \quad (2)$$

Where v represents the degree of freedom. The ES at confidence level α is computed using the properties of the t-distribution and the estimated conditional volatility as follows:

$$ES_\alpha = \frac{v + {t_{\alpha,v}}^2 t}{v - 1} . \frac{dt(t_{\alpha,v}, v)}{1 - \alpha} . \sigma_t . \sqrt{\frac{v-2}{v}} \quad (3)$$

3.3 Historical Simulation

The Historical Simulation (HS) method estimates Value-at-Risk (VaR) and Expected Shortfall (ES) based entirely on empirical return distributions. This non-parametric technique does not rely on any distributional assumptions and instead uses historical return data to construct the loss distribution. It is widely used by financial institutions due to its simplicity and ease of implementation.In this study, a 250-day rolling window is employed for each currency in the FX portfolio. Daily returns are sorted from worst to best, and the VaR at the 97.5% confidence level is identified as the 6th worst return. The ES is calculated as the average of the five worst returns, representing the average loss in the tail beyond the VaR threshold:

$$VaR_\alpha = quantile\ (R_t, \alpha) \quad (4)$$

$$ES_\alpha = \frac{1}{1 - \alpha} \sum_{i=1}^{N} R_t . L_i(R_t > VaR_\alpha) \quad (5)$$

This approach assumes that the historical return distribution adequately represents future risk. However, it suffers from a key limitation: it does not account for changes in market volatility over time, which can lead to risk underestimation during periods of financial stress. As such, the HS method is commonly used as a benchmark but may not be responsive enough for dynamic market conditions under the FRTB framework.

3.4 Backtesting and Evaluation Framework

This study adopts the backtesting methodology introduced by Acerbi & Székely (2014), which allows for direct validation of Expected Shortfall (ES) forecasts. Unlike traditional VaR-based backtesting approaches, this method evaluates both underestimation and overestimation of tail risk, making it particularly relevant for regulatory applications under the FRTB.

The test involves evaluating whether the ES forecast at each time t accurately reflects the conditional expectation of losses beyond the VaR threshold. The hypotheses are defined as:
H₀: $ES_{\alpha,t}^{P} = ES_{\alpha,t}^{F}$- The ES forecast is unbiased (actual equals forecasted).
H₁: $ES_{\alpha,t}^{P} < ES_{\alpha,t}^{F}$- the model underestimate ES.
H₁: $ES_{\alpha,t}^{P} > ES_{\alpha,t}^{F}$- the model overestimate ES.

Where:

- $ES_{\alpha,t}^{P}$is the actual loss beyond VaR (empirical shortfall)
- $ES_{\alpha,t}^{F}$is the forecasted ES by the model.

The test involves evaluating whether the ES forecast at each time *t* accurately reflects the conditional expectation of losses beyond the VaR threshold. The hypotheses are defined as:

$$ES_{\alpha,t}(r_t) = \mathrm{E}\left[r_t | r_t > VaR_{\alpha,t}\right] = \mathrm{E}\left[\frac{[r_t \,.\, I_t]}{(1 - \alpha)}\right] \qquad (6)$$

where:
r_t is the fx return at time *t*,
I_t is an indicator function defined as:

$$I_t = \begin{cases} 1 \text{ if } r_t > VaR_\alpha \\ 0 \text{ if } r_t \leq VaR_\alpha \end{cases}$$

α is the confidence level (97,5%).

The Z-statistic aggregates deviations between realized losses and forecasted ES over T observations, scaled by the tail probability:

$$Z_t = -\frac{1}{T(1-\alpha)} \sum_{t=1}^{T} \frac{r_t . I_t}{ES_{\alpha,t}(r_t)} + 1 \qquad (7)$$

where:
T is the number of observations,
r_t is the return at time *t*,
$ES_{\alpha,t}(r_t)$ is the forecasted ES at confidence level α,
Under the null hypothesis of correct ES estimation, the expected value of the Z-statistic is:

$$\mathrm{E}\,[Z_t] = 0 \qquad (8)$$

Following Acerbi & Székely (2014), rejection thresholds are set based on the significance level of the Z-statistic:

- For underestimation:
  Reject H₀ if $Z_t < -0{,}7$ for the significance level 5%
  Reject H₀ if $Z_t < -1{,}8$ for the significance level 1%
- For overestimation:
  Reject H₀ if $Z_t > 0{,}53$ for the significance level 5%
  Reject H₀ if $Z_t > 0{,}93$ for the significance level 1%

To align with the Basel framework, results of the Z-statistic are interpreted using a traffic-light system, mirroring the Basel traffic-light approach, which classifying model performance as follows:

Table 1. Summary Rejection and Traffic Light Result Procedure based on Z2 Test Statistic

Zone	*Z Statistic Range*	*Interpretation*
Green	-0,7 ≤ Zt ≤ 0,59	ES model is valid
Amber	-1,8 ≤ Zt < -0,7 or 0,59 < Zt ≤ 0,93	ES model shows mild bias, revision advised
Red	Zt < -1,8 or Zt > 0,93	Severe underestimation or overestimation, model is rejected

This structured framework helps determine whether a bank's internal model satisfies Basel III requirements or requires recalibration.

3.5 Data

This study utilizes daily exchange rate data for eleven major currency pairs against the Indonesian Rupiah (IDR), collected from Refinitiv for the period 2 January 2007 to 31 December 2024. The selected currencies include USD, EUR, JPY, GBP, CNH, AUD, CAD, CHF, HKD, SGD, and SAR. Ten of these are the most actively traded currencies globally, based on Bank for International Settlements (2022) rankings, while SAR is included due to its high demand in Indonesia. The exchange rates used are mid-market closing prices published by Bank Indonesia, calculated as the average of daily bid and ask rates. Daily log-returns are computed using the continuous compounding method:

$$R_t = ln \frac{P_t}{P_{t-1}} \tag{9}$$

where $R_t$is the return at time *t*, and P_t are the exchange rates at time *t* and *t−1*, respectively (Ballotta & Fusai, 2017). This method is preferred for modeling continuous risk and volatility in financial time series.

4. RESULT AND DISCUSSION

4.1 Descriptive Statistics of Return Series

Table 2 presents the descriptive statistics for the daily log-returns of eleven major foreign exchange (FX) pairs against the Indonesian Rupiah (IDR). Across all currency pairs, the mean and median returns are close to zero, indicating the absence of consistent long-term appreciation or depreciation trends. This aligns with the expectations of efficient market behaviour in FX markets. In terms of volatility, measured by standard deviation, JPY/IDR (0.0076), AUD/IDR (0.0071), and CHF/IDR (0.0071) exhibit the highest fluctuations. Conversely, SGD/IDR (0.0041) and USD/IDR (0.0044) show more stable return patterns, reflecting their lower exposure to emerging market shocks. The skewness values indicate that most return distributions are right-skewed (positively skewed), suggesting a higher likelihood of extreme positive returns. However, AUD/IDR is negatively skewed (-0.37463), revealing its tendency toward negative shocks. Meanwhile, the kurtosis values highlight the presence of fat tails in all currencies—particularly CHF/IDR (80.06), SGD/IDR (52.73), and USD/IDR (41.84)—suggesting higher probabilities of extreme outcomes. These findings affirm the importance of tail risk measures such as Expected Shortfall (ES) in FX portfolio risk modelling.

Table 2. Descriptive Statistic of FX Returns

	USD	*EUR*	*JPY*	*GBP*	*CNY*	*AUD*	*CAD*	*CHF*	*HKD*	*SGD*	*SAR*
Mean	0,0001	0,0001	0,0001	0,0000	0,0001	0,0001	0,0001	0,0002	0,0001	0,0002	0,0001
Median	0,0000	0,0000	(0,0001)	0,0001	0,0001	0,0002	0,0002	0,0001	0,0001	0,0002	0,0000
Maximum	0,0762	0,0644	0,0942	0,0874	0,0750	0,0566	0,0719	0,1657	0,0766	0,0821	0,0759
Minimum	(0,0647)	(0,0582)	(0,0670)	(0,0597)	(0,0639)	(0,0695)	(0,0487)	(0,0743)	(0,0647)	(0,0574)	(0,0640)
Std. Dev.	0,0044	0,0060	0,0076	0,0065	0,0044	0,0071	0,0058	0,0071	0,0044	0,0041	0,0044
Skewness	0,81033	0,10077	0,66861	0,00715	0,63248	(0,37463)	0,21603	2,79964	0,87245	1,39519	0,83013
Kurtosis	41,84191	8,32769	10,56023	14,03438	36,78850	9,51645	9,97049	80,05598	42,70301	52,72749	39,85353

4.2 Log Return Volatility

Figure 1 displays the daily log-returns for each FX pair throughout the full sample period from January 2007 to December 2024. The time series plots clearly reveal volatility clustering, particularly during known periods of global financial turmoil, such as the 2008–2009 Global Financial Crisis and the 2020 COVID-19 pandemic. Currencies such as JPY, AUD, and CHF experienced heightened volatility and frequent spikes during these periods. These patterns validate the need for dynamic risk modelling approaches such as GARCH, which capture time-varying volatility more effectively than static estimators. Overall, the charts reinforce the limitations of assuming normality or constant variance in FX return distributions.

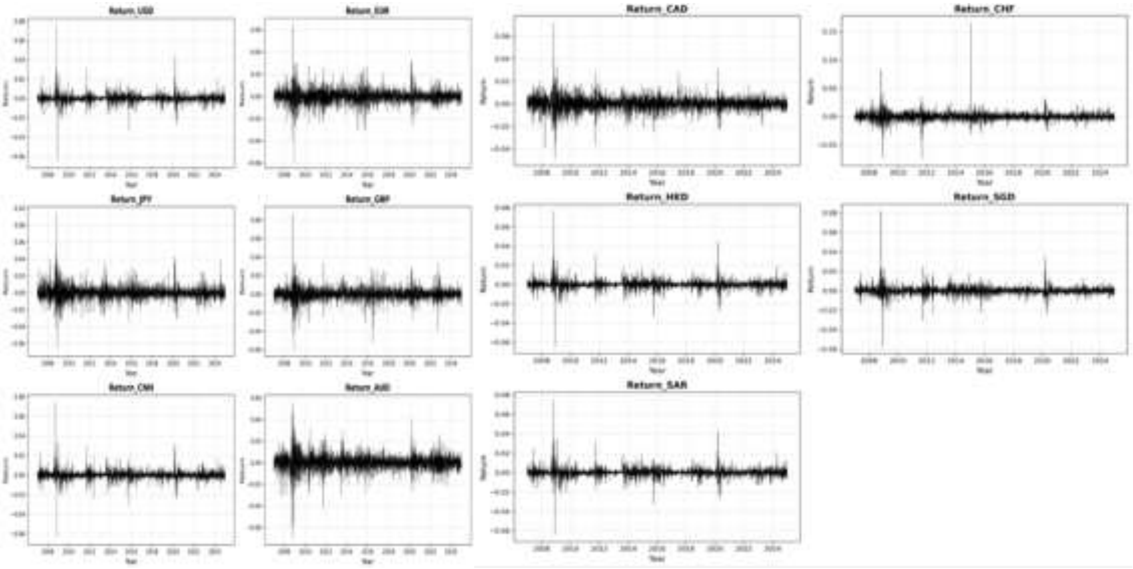

Figure 1. Daily Log-Returns of Major Currency Pairs

4.3 Annual Volatility (Standard Deviation)

Table 3 reports the annualized standard deviation of daily returns for each FX pair. The highest levels of volatility occurred during 2008, 2009, and 2020, coinciding with major global crisis periods. For instance, in 2008, AUD/IDR reached a volatility of 1.46%, while SGD/IDR peaked at 0.845%—both values significantly above their long-run averages.

The annual volatility levels declined after each crisis period, indicating a transition from turbulent to more stable market conditions. The inclusion of both high-stress and stable periods in the sample enhances the robustness of ES model evaluation, as it allows performance to be assessed across diverse market regimes.

By spanning from 2007 to 2024, the dataset captures two major systemic shocks, allowing the ES estimation framework to account for regime changes and tail dependencies. This broad temporal coverage ensures a more accurate and representative assessment of FX market risk under the Basel III FRTB framework.

Table 3. Yearly Standard Deviation Analysis

	USD	*EUR*	*JPY*	*GBP*	*CNY*	*AUD*	*CAD*	*CHF*	*HKD*	*SGD*	*SAR*
2007	0,00408	0,00438	0,00755	0,00482	0,00443	0,00664	0,00595	0,00559	0,00410	0,00351	0,00443
2008	0,00835	0,01054	0,01507	0,01165	0,00832	0,01460	0,01085	0,01252	0,00845	0,00846	0,00838
2009	0,00572	0,00777	0,01134	0,01018	0,00557	0,00983	0,00806	0,00849	0,00571	0,00496	0,00601
2010	0,00299	0,00692	0,00750	0,00647	0,00320	0,00781	0,00628	0,00619	0,00290	0,00308	0,00314
2011	0,00376	0,00719	0,00662	0,00553	0,00405	0,00829	0,00635	0,00990	0,00364	0,00438	0,00376
2012	0,00282	0,00539	0,00555	0,00460	0,00388	0,00556	0,00447	0,00536	0,00289	0,00417	0,00284
2013	0,00384	0,00566	0,00892	0,00592	0,00409	0,00694	0,00523	0,00632	0,00382	0,00409	0,00385

2014	0,00441	0,00528	0,00612	0,00503	0,00441	0,00570	0,00533	0,00547	0,00438	0,00411	0,00441
2015	0,00525	0,00839	0,00668	0,00644	0,00537	0,00687	0,00637	0,01327	0,00525	0,00462	0,00522
2016	0,00388	0,00533	0,00813	0,00851	0,00349	0,00578	0,00524	0,00535	0,00380	0,00319	0,00388
2017	0,00167	0,00425	0,00528	0,00568	0,00246	0,00451	0,00468	0,00428	0,00167	0,00212	0,00167
2018	0,00377	0,00428	0,00510	0,00469	0,00342	0,00426	0,00459	0,00459	0,00372	0,00286	0,00377
2019	0,00288	0,00339	0,00470	0,00504	0,00279	0,00377	0,00369	0,00406	0,00286	0,00228	0,00289
2020	0,00686	0,00702	0,00836	0,00743	0,00601	0,00727	0,00602	0,00736	0,00685	0,00574	0,00682
2021	0,00257	0,00300	0,00366	0,00379	0,00213	0,00450	0,00389	0,00362	0,00252	0,00206	0,00260
2022	0,00286	0,00541	0,00638	0,00698	0,00374	0,00698	0,00432	0,00523	0,00280	0,00271	0,00284
2023	0,00362	0,00398	0,00610	0,00451	0,00338	0,00599	0,00423	0,00490	0,00365	0,00263	0,00362
2024	0,00382	0,00339	0,00630	0,00384	0,00356	0,00427	0,00344	0,00462	0,00381	0,00275	0,00382

4.4 Statistical Test Result

Before proceeding with volatility and risk modelling, we conducted three standard diagnostic tests—ADF for stationarity, Jarque-Berra for normality, and White test for heteroskedasticity. Table 4 shows that all FX return series satisfy the assumption of stationarity and exhibit significant non-normality and heteroskedasticity. These results support the use of GARCH-family models with heavy-tailed innovations for risk estimation.

Table 4. Summary of Preliminary Statistical Tests for FX Return Series (Normality, Stationarity, and Heteroskedasticity)

Currency	*JB Statistic*	*JB p-value*	*ADF Statistic*	*ADF p-value*	*White Statistic*	*White p-value*
USD	320.701,50	0,000	-13,348	0,01	9,6769	0,00792
EUR	12.688,64	0,000	-15,121	0,01	67,4087	0,00000
JPY	20.720,97	0,000	-14,635	0,01	91,1759	0,00000
GBP	36.021,06	0,000	-15,120	0,01	31,4977	0,00000
CNH	247.833,53	0,000	-13,269	0,01	16,0428	0,00033
AUD	16.663,59	0,000	-17,242	0,01	112,7894	0,00000
CAD	18.213,13	0,000	-17,536	0,01	85,6652	0,00000
CHF	1.178.011,99	0,000	-16,709	0,01	9,1376	0,01037
HKD	334.094,34	0,000	-13,257	0,01	10,1793	0,00616
SGD	509.943,95	0,000	-13,982	0,01	16,5197	0,00026
SAR	291.012,98	0,000	-13,399	0,01	12,4751	0,00195

Note: ADF = Augmented Dickey-Fuller test for stationarity; JB = Jarque-Berra test for normality; White = ARCH LM test for heteroskedasticity. Results are based on daily log-returns from January 2007 to December 2024.

4.5 Estimated Parameters of GARCH (1,1)-*t*

To model time-varying volatility in the FX return series, we estimated univariate GARCH (1,1) models with Student's t-distributed errors for each currency. This specification was selected due to its ability to capture both volatility clustering and excess kurtosis, which were evident in the preliminary diagnostic tests. The estimation was conducted on eleven currency pairs against the Indonesian Rupiah (IDR) using software R Studio, covering daily data from January 2007 to December 2024.

Table 5. Estimated Parameters of GARCH (1,1)-*t* Model for Exchange Rate Return Series

Currency	*Omega*	*p-value*	*Alpha*	*p-value*	*Beta*	*p-value*	*Degrees of Freedom*	*p-value*	*Log Likelihood*	*AIC*	*BIC*
USD	0,00000025	0,571	0,155	0,000	0,844	0,000	3,595	0,000	19.179,79	-8,7158	-8,7086
EUR	0,00000025	0,121	0,050	0,000	0,943	0,000	6,445	0,000	16.933,07	-7,6946	-7,6873
JPY	0,00000079	0,023	0,078	0,000	0,910	0,000	5,505	0,000	15.972,24	-7,2578	-7,2506
GBP	0,00000060	0,009	0,047	0,000	0,936	0,000	5,812	0,000	16.596,36	-7,5415	-7,5343
CNY	0,00000067	0,171	0,189	0,000	0,809	0,000	3,398	0,000	18.885,22	-8,5819	-8,5747
AUD	0,00000032	0,206	0,042	0,000	0,951	0,000	7,593	0,000	16.238,39	-7,3788	-7,3716
CAD	0,00000029	0,116	0,041	0,000	0,949	0,000	6,122	0,000	17.023,30	-7,7356	-7,7283
CHF	0,00000081	0,006	0,058	0,000	0,922	0,000	5,138	0,000	16.517,98	-7,5059	-7,4986
HKD	0,00000026	0,572	0,162	0,000	0,837	0,000	3,687	0,000	19.196,55	-8,7234	-8,7162
SGD	0,00000030	0,313	0,085	0,000	0,896	0,000	4,195	0,000	19.138,51	-8,6970	-8,6898
SAR	0,00000029	0,517	0,158	0,000	0,841	0,000	3,549	0,000	19.098,96	-8,6791	-8,6718

Note: ω = constant, α = ARCH coefficient, β = GARCH coefficient, df = degrees of freedom for Student's t-distribution.

The estimated parameters—presented in **Table 5**. —include the constant term (ω), the ARCH coefficient (α), the GARCH coefficient (β), the degrees of freedom (df) for the t-distribution, and the log-likelihood value indicating model fit. The GARCH (1,1)-*t* model estimates reveal meaningful differences in volatility dynamics across currencies. A very small omega (ω) value across all series indicates that exchange rate volatility lacks a substantial constant component and is primarily driven by previous shocks or past volatility. The alpha (α) coefficient captures the immediate impact of past shocks on current volatility, while the beta (β) coefficient reflects the persistence or clustering of volatility over time. The degrees of freedom estimates lie between approximately 4 and 7, confirming the presence of heavy-tailed return distributions. This justifies the use of the GARCH-t model as it provides more accurate risk estimates in the presence of extreme market movements, which is critical for Expected Shortfall (ES) measurement under the Basel III FRTB framework.

4.6 Expected Shortfall (ES) Estimation Results

This section presents the estimation results of Expected Shortfall (ES) for each currency using two different approaches: the GARCH (1,1)-*t* model and Basic Historical Simulation (HS). The ES was calculated at the 97.5% confidence level over the full sample period from January 2007 to December 2024. Figures 2 to 3 shows the daily return series alongside the corresponding upper and lower ES bounds estimated under each method for all 11 currencies. Each figure provides a visual comparison between observed returns (black line) and the conditional ES bands (blue for upper, red for lower), enabling assessment of model responsiveness to market fluctuations and extreme events.

Figure 2 displays the ES estimation using the GARCH (1,1)-*t* model. This method captures time-varying volatility by modelling conditional variance and allows for heavy-tailed innovations. The resulting ES bands exhibit high sensitivity to market stress periods, such as the 2008 global financial crisis and the 2020 COVID-19 shock. Notably, currencies such as JPY, CHF, and AUD show pronounced ES expansions during these events, reflecting the model's effectiveness in capturing tail risk and volatility clustering. The wider ES range during turbulence suggests that the GARCH-t model provides a more conservative and responsive measure of downside risk.

Figure 3 presents the ES estimates based on the Historical Simulation method. Unlike the GARCH-t model, HS relies solely on empirical return distributions without assuming any specific volatility structure. As a result, the estimated ES bands appear smoother and less reactive to sudden shifts in return volatility. During stress events, such as in CHF and GBP markets, the HS model tends to underestimate ES, failing to reflect recent volatility increases. This limitation is particularly concerning under Basel III FRTB, where underestimation may lead to inadequate capital reserves.

Overall, the comparative analysis highlights key differences in ES behavior across methods. While GARCH-t offers the highest responsiveness to market shocks. The HS method, though simple, risks systematic underestimation during high-volatility regimes.

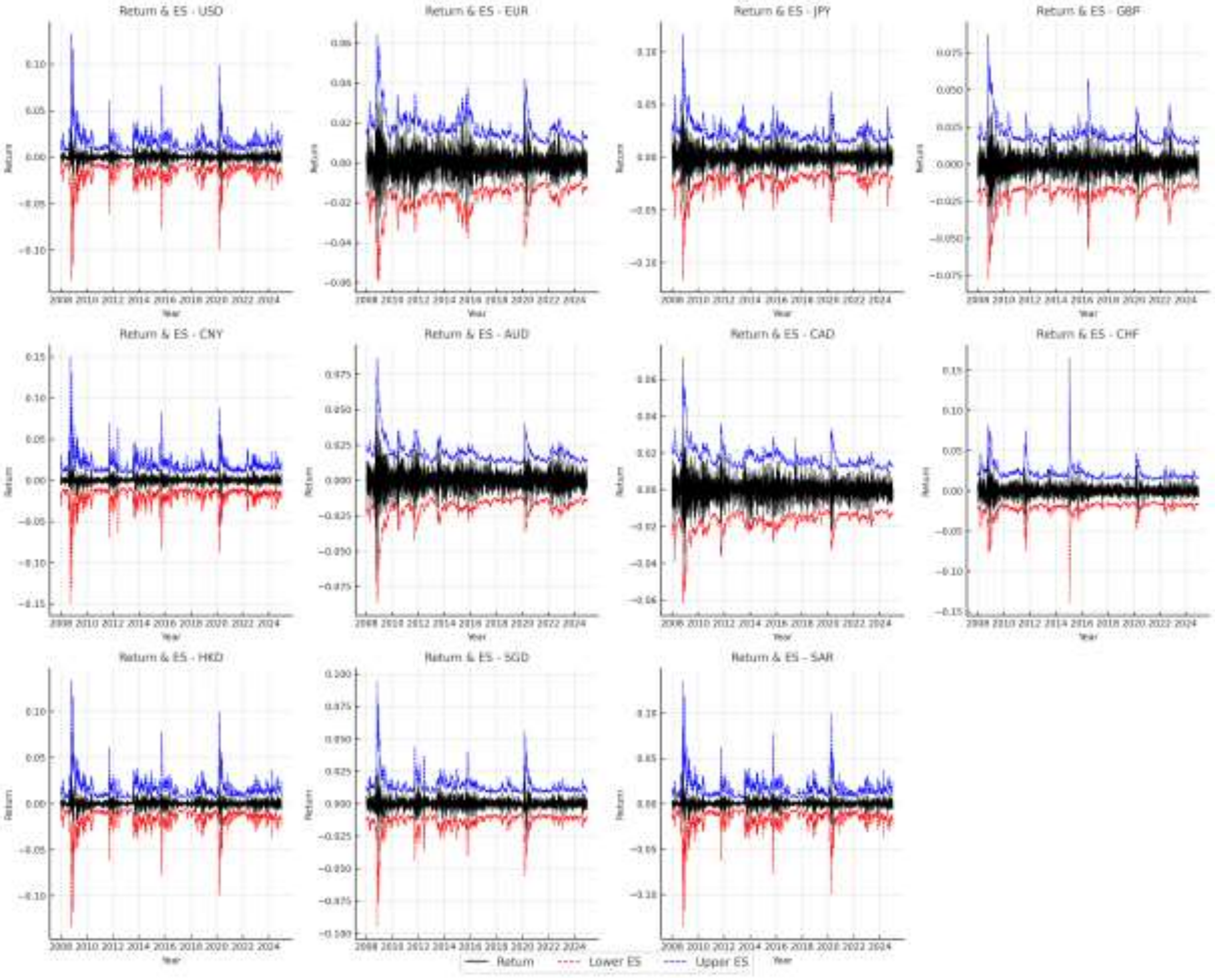

Figure 2. Estimated Expected Shortfall (ES) Using GARCH(1,1)-*t* Model for 11 Currencies

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 05 May 2025
DOI: 10.47191/ijcsrr/V8-i5-69, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

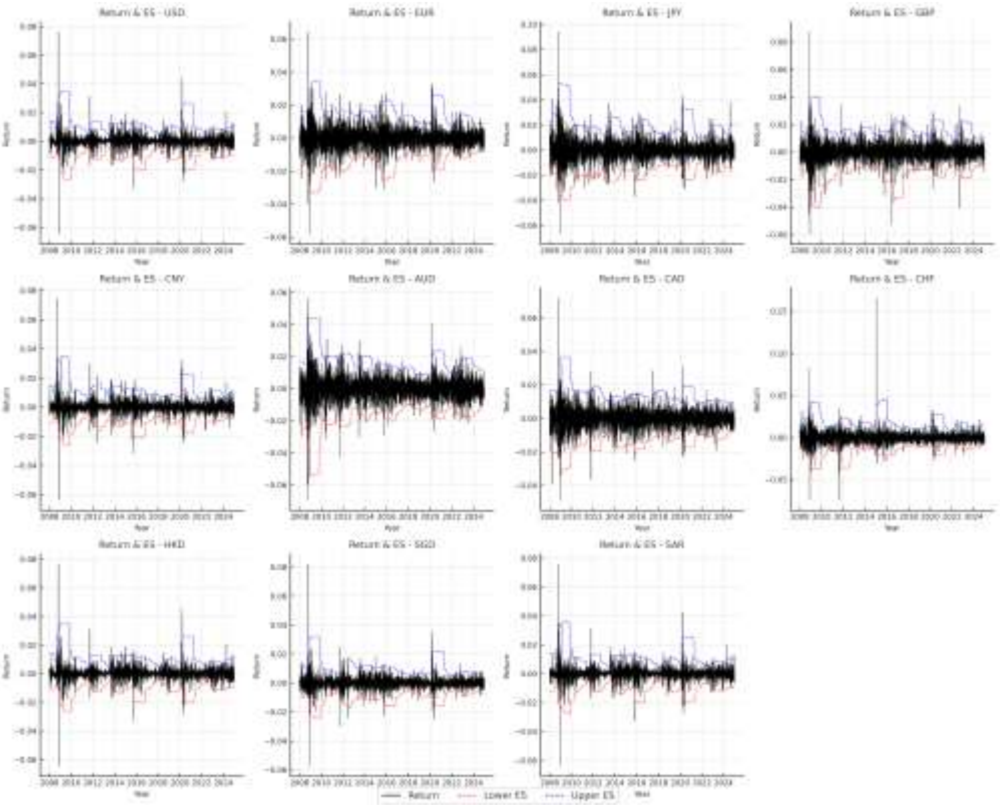

Figure 3. Estimated Expected Shortfall (ES) Using Historical Simulation (HS) for 11 Currencies

4.7 Backtesting Analysis of ES Models

The effectiveness of Expected Shortfall (ES) estimation methods in accurately measuring market risk is assessed using the Acerbi & Szekely (2014) Z2 Test. This test enables the detection of both underestimation and overestimation biases in ES forecasts. To support a robust evaluation, this study applies backtesting on both lower-tail ES (representing depreciation risks) and upper-tail ES (representing appreciation risks), acknowledging that FX exposure in trading portfolios can result in losses in either direction depending on the position (long or short). The traffic-light framework proposed by Basel is then used to classify the model performance based on the Z-statistic outcomes (BCBS, 2019).

The Z2 test evaluates whether actual returns falling beyond the Value-at-Risk (VaR) threshold are consistent with ES estimates based on the **Table 1.** This subsection summarizes the Z-statistic values from the backtest for all two estimation methods—Basic Historical Simulation (Basic HS) and GARCH (1,1)-*t*—applied across 11 major currencies over the period from 2008 to 2024.

Table 6. Summary of Backtesting Z2 Test Result

Methods/ Traffic Zone	*Green*		*Amber*		*Red*	
	Lower ES	*Upper ES*	*Lower ES*	*Upper ES*	*Lower ES*	*Upper ES*
GARCH-t	75	85	87	79	25	23
HS	106	95	73	82	8	10

From the **Table 6** above, Historical Simulation (HS) performs well by lower frequency of Red Zone but shows a slightly higher frequency in Amber Zones. In contrast, GARCH-*t* models exhibit the highest frequency of Red Zone results, indicating a tendency to severely over- or underestimate ES. While GARCH-*t* still performs reasonably well, its tail risk estimation may require further refinement or adjustment to cope with sharp regime changes or structural breaks in return dynamics. An interesting insight emerges when observing asymmetry between Lower and Upper ES performances: GARCH-t suffers from high Amber counts across both tails, signaling imbalanced tail behaviors, which can lead to inconsistent capital allocation if not addressed. While HS performs better for Lower ES but shows more Amber and Red Zones in Upper ES, suggesting more frequent mild overestimation in appreciating currencies.

To align Expected Shortfall (ES) model evaluation with regulatory standards, this section interprets the Z-test results under the Basel III traffic-light backtesting framework. The framework categorizes each Z-test outcome into one of three zones based on pre-defined thresholds, indicating the degree of model accuracy and reliability. These zones are crucial for determining whether a risk model is considered valid and compliant with capital adequacy standards under the Fundamental Review of the Trading Book (FRTB).

The regulatory traffic-light system does not isolate backtesting outcomes by tail direction. In this study, however, both Lower and Upper ES tests are conducted independently to capture full risk exposure in the bank's foreign exchange portfolio. The traffic-light classification is applied separately to each tail, and a model is flagged as non-compliant if either tail falls into the Red Zone.

4.8 Model Performance Comparison

To evaluate the robustness and consistency of each Expected Shortfall (ES) estimation model across varying market conditions, we construct a Model Performance Score. This metric combines the backtesting results derived from the Z-test traffic-light classification system (as per Basel standards), using the formula:

$$Performance\ Score = \ Green - Red - 0{,}5 \times \ Amber \qquad (10)$$

This formulation penalizes models for frequent Amber and Red outcomes, rewarding those with more valid (Green zone) Z-test results. To facilitate model comparison across the entire 2008–2024 period, Table 7 summarizes the total performance score per method. Notably, HS leads with the highest overall score (105.5), while GARCH-*t*, despite strong results in crisis periods, accumulates a lower net score (29), affected by higher volatility in stable years.

Table 7. Summary of Model Performance Score Across Years

Year	*GARCH-t*	*HS*
Total Model Performance Score	29	105,5

To capture year-over-year performance dynamics, **Figure 4** below plots the annual model score trends, highlighting consistency across different ES estimation methods. It reveals that GARCH-t performs strongly in stress periods (e.g., 2008, 2020), aligning with its volatility-adaptive nature. HS show higher stability across non-crisis periods, consistently outperforming after 2014. GARCH-t exhibits fluctuations, with notable underperformance in 2012 and 2021, where in that year there are one of the lowest average standard deviation as mentioned in Table 5, indicating potential over estimation in capital allocation as the Z2 result are more than 0,59 and 0,93, which may suggest occasional over-sensitivity during calm period or high-volatility transitions.

These results suggest that while GARCH-*t* demonstrates strength during turbulent market phases, while HS offer greater consistency and reliability across all phases of the market cycle. The use of performance scores derived from traffic-light backtesting zones proves to be a transparent and practical tool for comparative evaluation of ES model robustness.

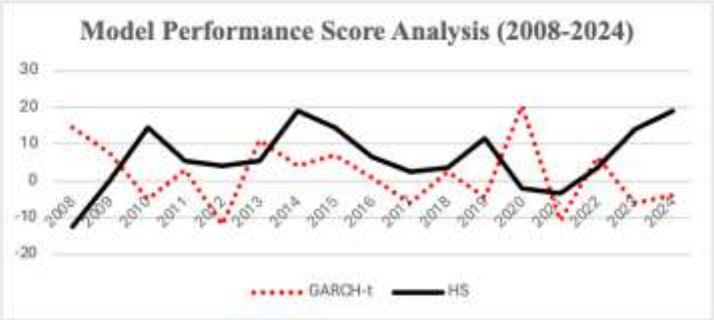

Figure 5. Model Performance Score Trend Across Years

4.9 Discussion of Model Suitability

The comparative evaluation of Expected Shortfall (ES) models in this study highlights important distinctions in their suitability across different market regimes and regulatory expectations. While all models—GARCH-*t* and Historical Simulation (HS)—provide valid approaches to ES estimation, their performance under the Acerbi–Székely backtesting framework reveals divergent strengths and weaknesses depending on the nature of the market environment.

The GARCH-t model demonstrates notable strength during periods of market stress. In years such as 2008 and 2020, it produced the highest performance scores among all models, reflecting its capacity to capture tail risk dynamics and respond to volatility clustering. This aligns with its theoretical underpinnings in modelling conditional heteroscedasticity and fat-tailed return distributions. However, despite its crisis-resilient characteristics, GARCH-t performs inconsistently in stable market conditions. Over the full sample period (2008–2024), it exhibits the highest number of Amber and Red zone violations, indicating substantial backtesting rejections under Basel III criteria. This outcome diverges from previous studies such as Lazar and Zhang (2019), who observed that GARCH-type models generally required fewer corrections and performed reliably in tail-risk estimation. The inconsistency in GARCH-t performance highlights the importance of careful model calibration and cautions against unqualified reliance on heavy-tailed parametric models for daily risk management.

In contrast, the Historical Simulation (HS) method, while frequently critiqued in the literature for underestimating ES—particularly by García-Risueño (2025)—exhibited a more under performance. In this study, HS performed surprisingly well in normal market conditions, consistently maintaining a high number of Green Zone results and relatively few backtesting violations. Its underperformance was concentrated primarily in stress periods, where its static distributional assumptions were inadequate to reflect sudden spikes in market volatility. These findings suggest that HS may function well as a low-complexity, regulatory-compliant model in normal condition where operational simplicity and capital efficiency are crucial, despite its limited tail flexibility.

This study also extends previous work by explicitly validating ES models using the Acerbi–Székely Z-test for both tails, following the approach of Righi and Ceretta (2015) and Lazar and Zhang (2019). However, unlike those studies—which found that parametric models such as GARCH generally passed ES backtests—our findings reveal that GARCH-*t* models in fact produce more rejections than either historical or semi-parametric alternatives. This divergence highlights the importance of market context and data characteristics when applying model validation tools.

Furthermore, as emphasized by Grajales and Medina Hurtado (2023), the adoption of ES under the FRTB framework has substantial capital implications, particularly when models are not well-calibrated to capture tail risk. In such cases, backtesting failures can directly translate into stringent capital charges. These findings underscore the importance of aligning model selection not only with statistical accuracy, but also with capital optimization objectives. GARCH-t models may serve effectively in stress-testing frameworks, while HS is more suited to daily internal risk measurement in stable conditions. Ultimately, institutions may benefit from adopting a hybrid model governance framework, combining ES estimation method strategies to enhance both compliance reliability and capital efficiency.

5. CONCLUSION

This study evaluates the performance of Expected Shortfall (ES) estimation models—GARCH-*t* and Historical Simulation (HS)—in measuring market risk within Indonesian bank trading book portfolios under the Basel III Fundamental Review of the Trading Book (FRTB) framework. Through rigorous statistical validation, including Acerbi–Székely's Z-test on both lower and upper ES forecasts, the analysis offers new empirical evidence on model reliability across varying economic regimes.

The findings reveal a fundamental trade-off between tail risk sensitivity and consistency with regulatory compliance. The GARCH-*t* model demonstrates strong performance during high-volatility stress periods—such as those in 2008 and 2020—aligning with its theoretical ability to capture volatility clustering and fat-tailed return distributions. However, its elevated rate of backtesting rejections in stable markets raises concerns about over-sensitivity and potential capital inefficiencies. By contrast, Historical Simulation models deliver more stable results over time, with achieving the greatest overall consistency across both tails, particularly in normal market conditions.

From a regulatory perspective, the results suggest that Indonesian authorities and financial institutions should strengthen supervisory frameworks by incorporating dual-tail ES validation and ensuring that calibration practices balance capital efficiency with stress responsiveness. Furthermore, as the current OJK framework on market risk is still in its early stages, this study offers timely insights that can support regulatory refinement and supervisory development from OJK's perspective, while also guiding banks in strengthening their internal model governance in line with evolving international standards.

In conclusion, this research contributes to the literature on ES estimation in emerging markets and FX-exposed portfolios by validating the performance of alternative models and drawing attention to their practical implications for capital adequacy and regulatory compliance. Future research could explore the use of ensemble models, machine learning-based calibration techniques, or regime-switching frameworks to improve robustness and adaptability across different market environments.

REFERENCES

1. Acerbi, C., & Székely, B. (2014). BACKTESTING EXPECTED SHORTFALL Introducing three model-independent, non-parametric back-test methodologies for Expected Shortfall. https://api.semanticscholar.org/CorpusID:56135689
2. Afuecheta, E., Okorie, I. E., Nadarajah, S., & Nzeribe, G. E. (2024). Forecasting Value at Risk and Expected Shortfall of Foreign Exchange Rate Volatility of Major African Currencies via GARCH and Dynamic Conditional Correlation Analysis. In Computational Economics (Vol. 63, Issue 1). Springer US. https://doi.org/10.1007/s10614-022-10340-9
3. Ballotta, L., & Fusai, G. (2017). A Gentle Introduction to Value at Risk. https://doi.org/10.13140/RG.2.2.23435.49448
4. Basel Committee On Banking Supervision. (2019). Basel Committee on Banking Supervision Consultative Document Revisions to the minimum capital requirements for market risk #408 (Vol. 2019, Issue June). https://www.bis.org/bcbs/publ/d436.pdf
5. Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 32(3), 407–327. https://econpapers.repec.org/RePEc:eee:econom:v:31:y:1986:i:3:p:307-327
6. Epaphra, M. (2017). Modeling Exchange Rate Volatility: Application of the GARCH and EGARCH Models. Journal of Mathematical Finance, 07(01), 121–143. https://doi.org/10.4236/jmf.2017.71007
7. García-Risueño, P. (2025). Historical Simulation Systematically Underestimates the Expected Shortfall. Journal of Risk and Financial Management, 18(1). https://doi.org/10.3390/jrfm1801003
8. Grajales, C. A., & Medina Hurtado, S. (2023). Sensitivities-based method and expected shortfall for market risk under FRTB: its impact on options risk capital. Journal of Economics, Finance and Administrative Science, 28(55), 96–115. https://doi.org/10.1108/JEFAS-12-2021-0268
9. Johansson, E., & Ludolphy, L. E. (2020). An Empirical Study : Expected Shortfall Estimation Methods for a Bank ' s Trading Book.
10. Lazar, E., & Zhang, N. (2019). Model Risk of Expected Shortfall. Journal of Banking and Finance, 105, 74–93. https://doi.org/10.1016/j.jbankfin.2019.05.017
11. Lyócsa, Š., Plíhal, T., & Výrost, T. (2024). Forecasting day-ahead expected shortfall on the EUR/USD exchange rate: The (I)relevance of implied volatility. International Journal of Forecasting, 40(4), 1275–1301. https://doi.org/10.1016/j.ijforecast.2023.11.003

12. Nugroho, D. B., & Susanto, B. (2017). Volatility modeling for IDR exchange rate through APARCH model with student- t distribution. AIP Conference Proceedings, 1868(November 2022). https://doi.org/10.1063/1.4995120
13. Otoritas Jasa Keuangan. (2015). POJK No. 46/POJK.03/2015 tentang Penetapan Systemically Important Bank dan Capital Surcharge. Jakarta: Otoritas Jasa Keuangan.
14. Otoritas Jasa Keuangan. (2016). POJK No. 18/POJK.03/2016 tentang Penerapan Manajemen Risiko bagi Bank Umum. Jakarta: Otoritas Jasa Keuangan.
15. Otoritas Jasa Keuangan. (2022). SEOJK No. 23/SEOJK.03/2022 tentang Perhitungan Aset Tertimbang Menurut Risiko Untuk Risiko Pasar Bagi Bank Umum. Jakarta: Otoritas Jasa Keuangan.
16. PWC. (2016). Basel IV: Revised Standardised Approach for Market Risk. 1–48.
17. Righi, M. B., & Ceretta, P. S. (2015). A comparison of Expected Shortfall estimation models Marcelo Brutti Righi, Paulo Sergio Ceretta. Journal of Economics and Business, 78(2015), 14–47.
18. Schmutz, R. E., & Schneider, L. (2023). Into the Trading Book : Estimating Expected Shortfall. June.
19. Taylor, J. W. (2020). Forecast combinations for value at risk and expected shortfall. International Journal of Forecasting, 36(2), 428–441. https://doi.org/10.1016/j.ijforecast.2019.05.014
20. Wessels, C., & Vuuren, G. Van. (2023). Basel ' s Fundamental Review of the Trading Book : Implementation Principles for the Internal Models Approach. Review of Economics and Finance, 21, 1878–1892.
21. Yamai, Y., & Yoshiba, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. Journal of Banking and Finance, 29(4), 997–1015. https://doi.org/10.1016/j.jbankfin.2004.08.010

Cite this Article: Nuswantoro, K., Rokhim, R. (2025). Evaluating Validity of Expected Shortfall Estimation Method in Measuring Market Risk for Bank's Trading Book Position. International Journal of Current Science Research and Review, 8(5), pp. 2579-2592. DOI: https://doi.org/10.47191/ijcsrr/V8-i5-69