Risk forecasting and VaR Backtesting : Developed vs Emerging Currency

March 1, 2025

WARWICK BUSINESS SCHOOL
THE UNIVERSITY OF WARWICK

Submitted in partial fulfilment of the requirements for Research Methodology

Contents

Introduction

Risk Management in financial markets is a fundamental aspect of portfolio theory, ensuring investors or institutions can quantify and mitigate potential losses arising from fluctuations in asset prices. Robust statistical combined with econometric models are required to forecast currency valuation, its volatility, estimating Value-at-risk and validating said results through substantial back-testing procedures. In the context of the research, the focus is on Dollar portfolio, a synthetic portfolio constructed from a selection of nine developed market currencies and then comparing them with emerging market currency, which has been taken as Argentine Peso (ARS). The rationale behind choosing ARS is taking a currency with extreme volatility and then complete a proper comparative analysis. Through an empirical approach the research then answers key questions regarding model performance, risk predictability and robustness of framework.

Objectives achieved

The research, evaluating the performance of risk forecasting models, applying volatility, Var calculations and rigorous back testing, comes to the end result that GARCH(1,1) proved to be the most adaptive model, accurately capturing volatility clustering and responding dynamically to financial crises, outperforming MA and EWMA, which lagged in high-volatility regimes. VaR estimation demonstrated that historical simulation overestimated tail risks, while parametric models related to GARCH-based VaR aligned more closely with real-world breaches. Back testing validation through Kupiec's POF and Christoffersen's conditional coverage tests confirmed GARCH's statistical superiority in breach independence.

The findings underscore the necessity of portfolio-specific risk frameworks and suggest that emerging markets require more adaptive models such as EGARCH. This research validates risk modelling methodologies and provides a foundation for refining financial risk assessment tools in volatile scenario.

Literature review

The Review of Financial Studies (RFS) extensively explores risk forecasting, volatility modelling, and Value-at-Risk (VaR) methodologies. The study emphasizes GARCH, EWMA, and MA models in capturing financial market dynamics, meanwhile using back testing techniques, including Kupiec, Christoffersen, and Bernoulli tests, validate predictive accuracy.

Methodology

Data Collection

The data has been collected from Refinitiv. The exchange rates of developed countries have been taken from Barclays Bank & Refinitiv mode, while the ARS data has been taken from LSEG base. Post the collection, all exchange rate values were converted into a consistent numerical format, and any potential outliers were checked visually through preliminary plots of raw data. Exchange rate time series often exhibit extreme fluctuations due to macroeconomic shocks, speculative trading, and liquidity constraints, particularly in emerging markets like Argentina (ARS). Thus, data cleaning was a necessary step to ensure accuracy before conducting analysis.

Calculation of currency results

To analyse the foreign exchange (FX) returns from a U.S. investor's perspective, calculations of daily logarithmic returns using historical spot exchange rates were completed. The methodology follows the approximation:

$$rct + 1 = -\Delta st + 1$$

where st is the natural logarithm of the spot exchange rate. This transformation ensures that the returns accurately show currency value fluctuations for a U.S. investor. Since the dataset spans from December 1998 to December 2024, the calculations yield daily returns from January 1999 onward.

Portfolio construction

The Dollar Portfolio (DOL) was constructed as an equally weighted basket of the nine developed market currencies. Instead of aggregating absolute price changes, the portfolio return was defined as the cross-sectional average of individual currency log returns:

$$rcDOLt = \frac{1}{9\sum 9\mathrm{rci}}, \mathrm{t}+1$$

This approximation assumes equal weighting across all nine currencies at every time step.

Volatility forecasts model

Three volatility models were applied to both the Dollar Portfolio and the emerging market currency (ARS):

1. Moving Average (MA): Computed using a 50-day (10-week) rolling window to capture local fluctuations.
2. Exponentially Weighted Moving Average (EWMA): Applied with $\lambda = 0.94$, a commonly used smoothing parameter that assigns greater weight to recent observations.
3. GARCH (1,1): The standard GARCH model was estimated separately for the Dollar Portfolio and ARS returns, considering persistence and clustering effects.

Value at Risk Forecasting

For each portfolio, daily VaR (1%) forecasts were calculated using:

1. Parametric VaR: Based on normal distribution assumptions and volatility forecasts

$$VaR_{t+1} = -z_{\alpha} \cdot \sigma_{t+1}$$

   where z_{α} is the quantile corresponding to the 1% confidence level.

2. Historical Simulation Using a rolling 250-day window, past returns were sorted to determine the empirical quantile corresponding to the 1% worst loss.

Back testing framework

Back-testing was conducted with following statistical tests:

1. **Bernoulli Coverage Test**: Determines if the actual violation rate aligns with the expected 1% failure rate.

   The Bernoulli coverage test checks if the actual percentage of VaR violations aligns with the expected proportion (typically 1% for VaR(1%)).

   Formula:

$$H0: p = \alpha(Expected\ probability\ of\ failure)$$

$$LRUC = -2\ln\left(\frac{(1-\alpha)^{T-N}\alpha^{N}}{\left(1-\frac{N}{T}\right)^{T-N}\ \left(\frac{N}{T}\right)\hat{}N}\right)$$

$where:$

$T\ =\ total\ number\ of\ observations$

$N\ =\ number\ of\ VaR\ breaches$

$\alpha\ =\ VaR\ confidence\ level\ (e.g., 1\%\ =\ 0.01)$

$LRUC \sim \chi2(1)$

2. **Kupiec's Proportion of Failures (POF) Test**: A likelihood ratio test examining whether the actual exceedances conform to expected failure rates. It evaluates whether the observed frequency of VaR violations statistically matches the expected rate.

$$LRPOF = -2\ln\left(\frac{(1-\alpha)^{T-N}\alpha^{N}}{\left(1-\frac{N}{T}\right)^{T-N}\ \left(\frac{N}{T}\right)\hat{}N}\right)$$

This likelihood ratio test follows a chi-square distribution:

$$LRPOF \sim \chi2(1)$$

where:

N/T is the actual violation rate.

If LR_{POF} is significantly large, the model fails the back-test due to an incorrect expected violation probability.

3. Christoffersen's Conditional Coverage Test: Extends Kupiec's test by checking for independence among consecutive violations.

Likelihood ratio for independence:

$$LR_{Ind} = -2\ln\left(\frac{(P_{00}^{n_{00}}P_{01}^{n_{01}}P_{10}^{n_{10}}P_{11}^{n_{11}})}{(\hat{p}_0^{n_{00}+n_{01}}\hat{p}_1^{n_{10}+n_{11}})}\right)$$

$$\hat{p}_0 = \frac{n_{01}}{n_{00}+n_{01}}, \quad \hat{p}_1 = \frac{n_{11}}{n_{10}+n_{11}}$$

Likelihood Ratio for Conditional Coverage (Combining Unconditional and Independence Tests):

$LR_{CC}=LR_{UC}+LR_{IND}$

This test follows a chi-square distribution with two degrees of freedom:

$LR_{CC}\sim\chi^2(2)$

If LR_{CC} is large, it suggests that VaR failures cluster together, meaning the risk model does not fully capture volatility dynamics. Each test was conducted separately for the Dollar Portfolio and ARS.

Results & Analysis

Daily Currency Returns

The daily currency returns were calculated using the log difference method, with a negative sign to align with U.S. investor expectations, as defined in Methodology. The graph of daily currency returns provides a comprehensive visualization of return fluctuations across the analysed currencies, including developed market currencies (AUD, CAD, EUR, etc.) and the emerging market currency ARS, Argentine Peso (Figure 1). Developed market currencies exhibit relatively stable return patterns with lower volatility, while ARS demonstrates significant spikes, usually higher volatility and currency risk associated with emerging markets.

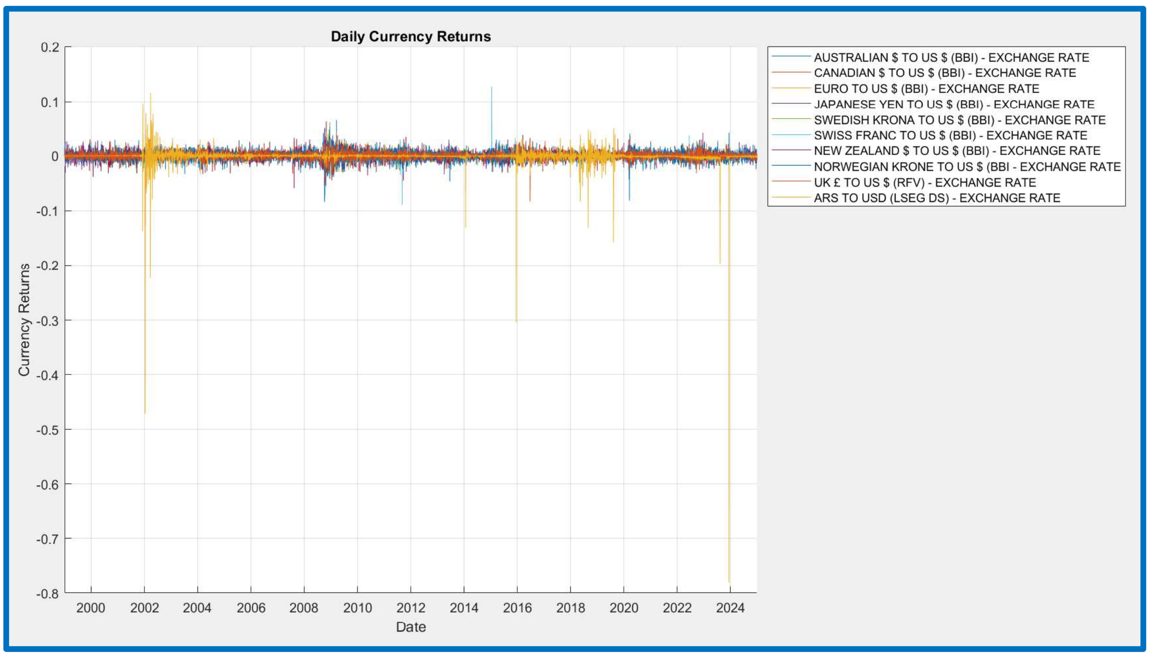

Figure 1: Daily Currency returns Comparison

Developed market currencies generally have near-normal return distributions with moderate skewness and kurtosis, while ARS shows extreme kurtosis and negative skewness, reflecting frequent, large losses. These findings align with theoretical expectations of emerging markets, which are often subject to greater macroeconomic changes and geopolitical risks.

Portfolio returns

The dollar portfolio returns are computed as the cross-sectional average of the individual currency returns, creating a diversified proxy for the performance of developed market currencies (Figure 2). This method, calculating the average of log returns instead of the log of average returns, was critiqued for its inability to capture compounding effects accurately. While the simplification provides computational ease, it assumes equal weighting and ignores time-variant effects, which may distort long-term portfolio dynamics.

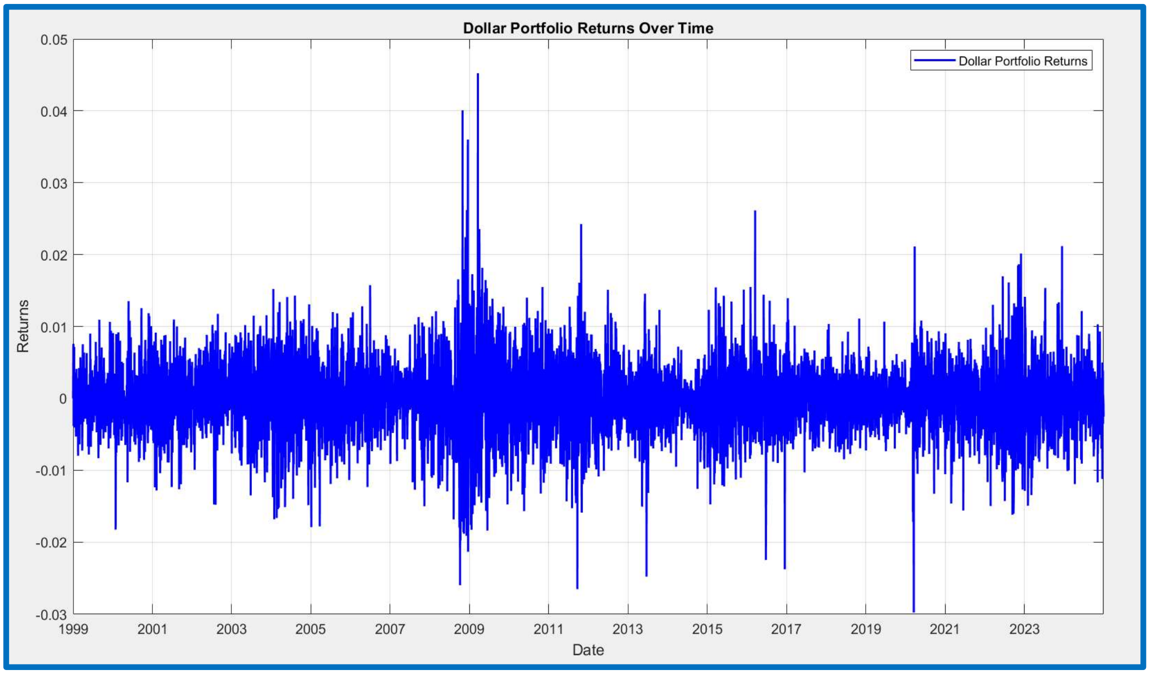

Figure 2: Dollar Portfolio returns over time.

Histogram of Dollar Portfolio Returns reveals a distribution concentrated around zero, with slight negative skewness and high to low frequencies (Figure 3). This indicates occasional extreme returns, likely driven by sharp currency movements during global financial shocks. The diversification effect reduces overall portfolio volatility compared to individual currencies like ARS, but the returns remain subject to systemic risks during periods of global instability.

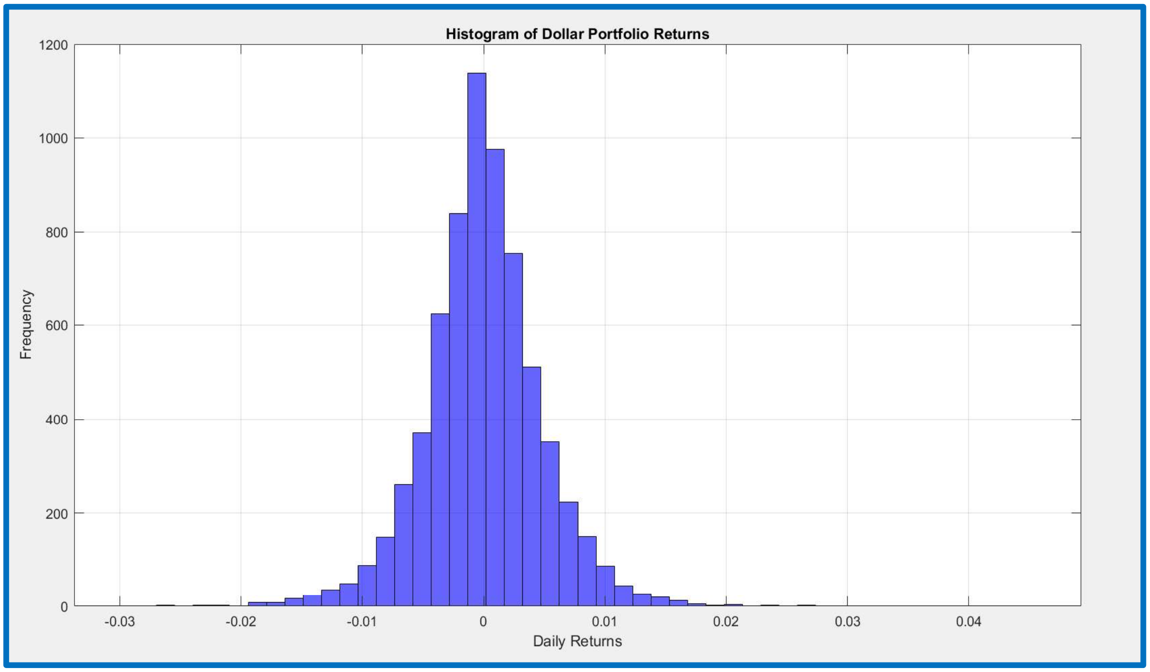

Figure 3: Histogram

Volatility Forecasting

Volatility forecasting was conducted using three models: Moving Average (MA), Exponentially Weighted Moving Average (EWMA), and GARCH. These models were applied to both the dollar portfolio and ARS returns, starting in 2000, using a burn-in period from 1999.

The volatility forecasts for the dollar portfolio indicate periods of heightened volatility, especially during the 2008 crisis and the 2020 pandemic (Figure 4). Among the models, the GARCH forecasts capture volatility clustering effectively, with periods of high volatility persisting longer. MA and EWMA provide smoother forecasts but may underestimate peak volatility during crises.

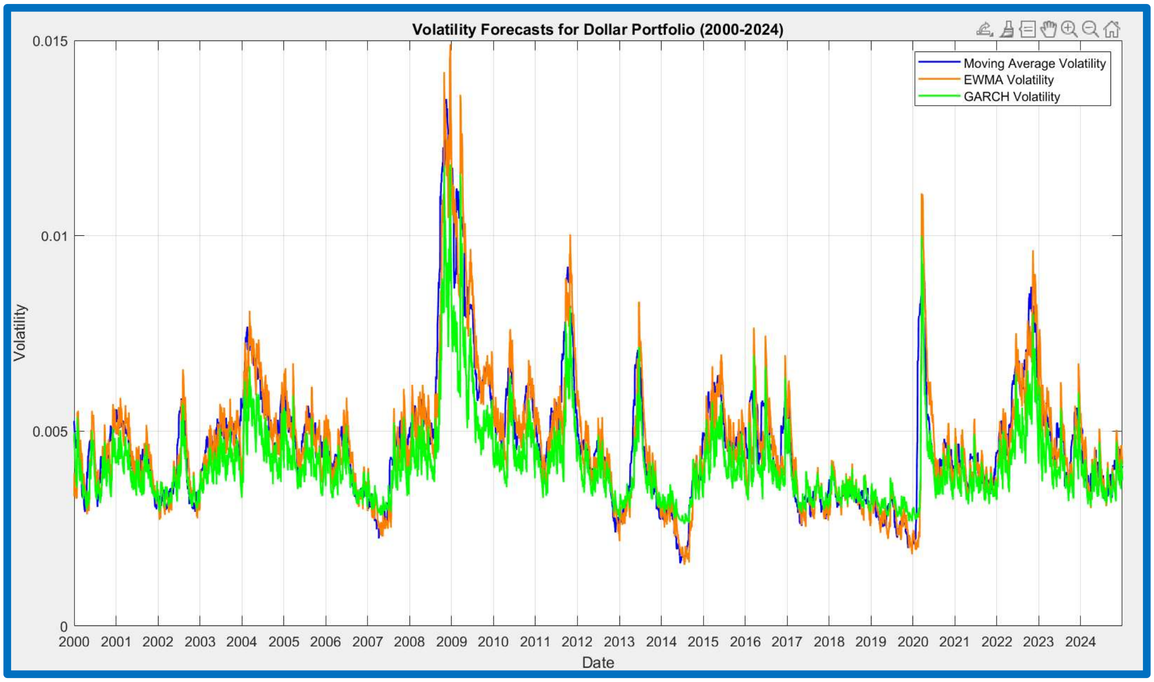

Figure 4: Volatility Forecasts for dollar portfolio.

In contrast, ARS volatility is consistently higher than that of the dollar portfolio, reflecting the inherent risks of emerging markets (Figure 5). The GARCH model again outperforms others in capturing the abrupt spikes and clustering effects, while MA and EWMA struggle to adapt quickly to sudden shifts. The long periods of elevated ARS volatility highlight the persistent instability in emerging markets compared to developed markets.

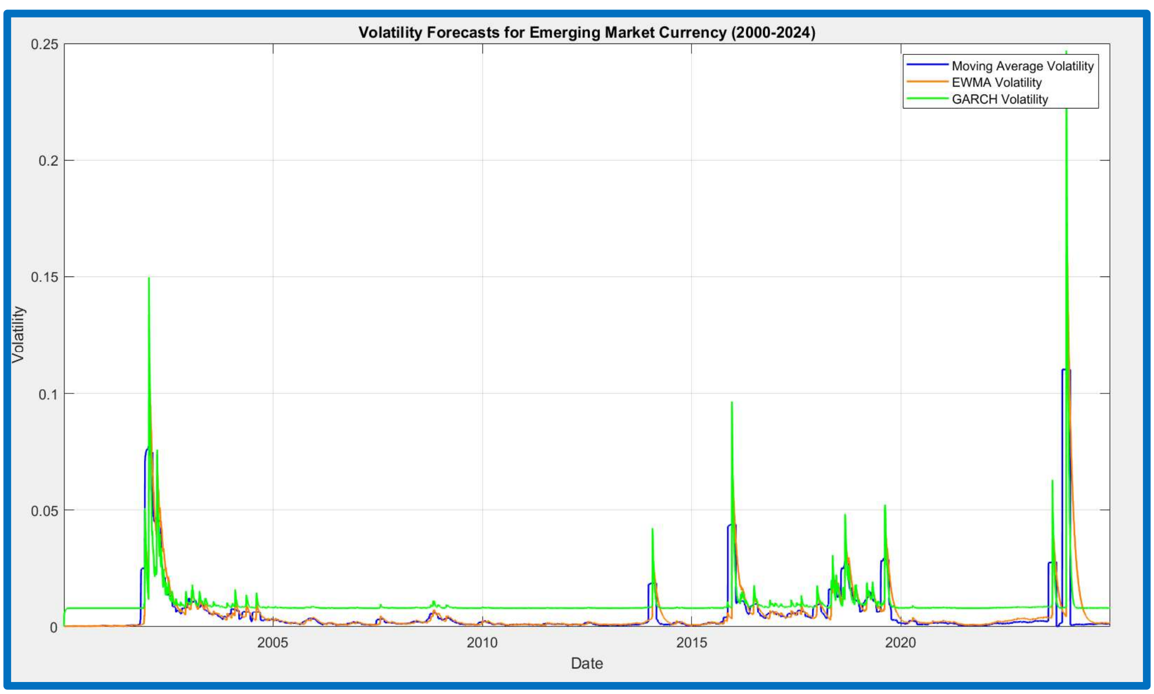

Figure 5: Volatility Forecasts for Emerging markets.

VaR results

Value-at-Risk (VaR) forecasts have been calculated for both portfolios using parametric and historical simulation approaches. The parametric VaR assumes normality, while the historical simulation captures empirical distribution tails more effectively.

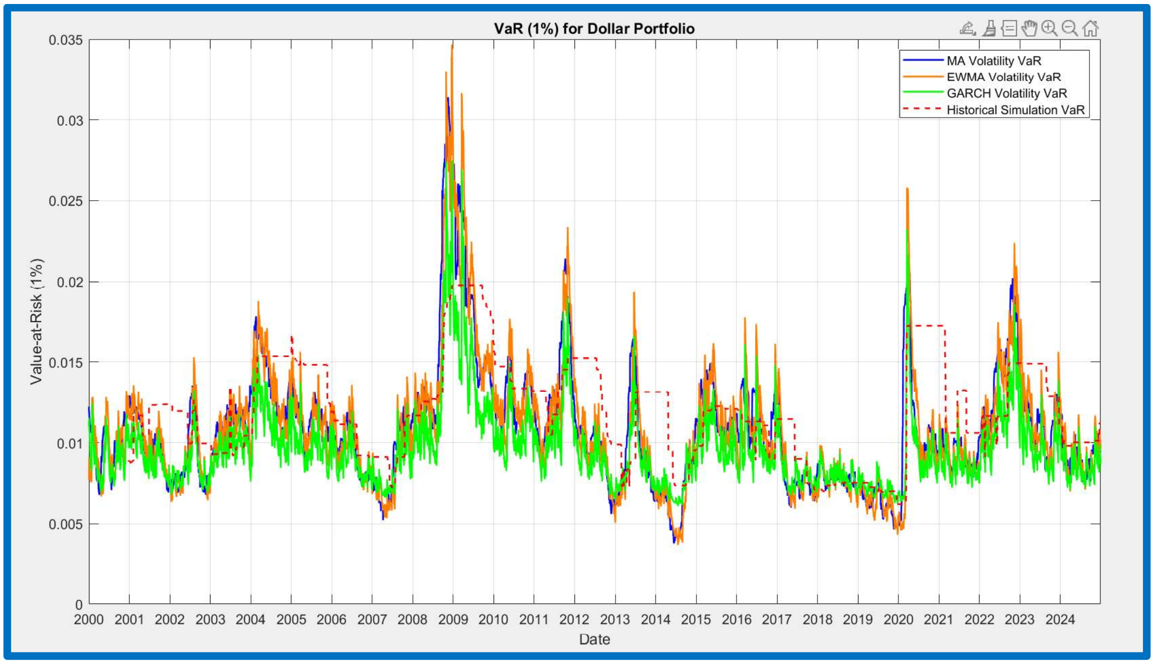

Figure 6: VaR for Dollar Portfolio

Figure 6 & 7 depict the VaR for the dollar portfolio and ARS, respectively. For the dollar portfolio, GARCH-based VaR estimates are the most responsive to volatility changes, particularly during crises, while MA and EWMA are lagging. Historical simulation VaR provides a conservative estimate, capturing extreme losses but often overestimating risk during stable periods. The ARS VaR forecasts show consistently higher risk levels, with the GARCH model excelling at reflecting extreme tail risks, as evidenced during periods of hyperinflation and currency devaluation (Figure 7).

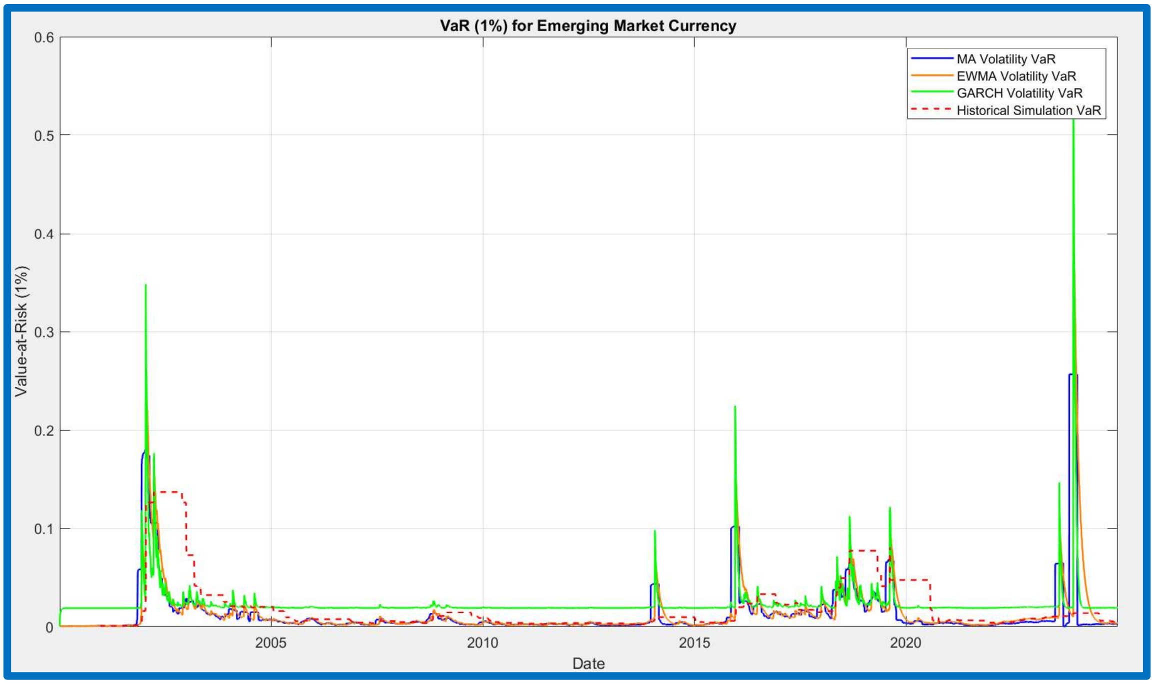

Figure 7: VaR for Emerging Market Currency

The comparison of 252-day and 504-day rolling window VaR (Figure 8) highlights the trade-off between sensitivity and stability. The shorter window is more responsive to recent changes, while the longer window smooths out short-term volatility but may miss sudden spikes. This dynamic is evident in ARS, where the 252-day window captures sharp risks better than the 504-day window (Figure 8 & 9).

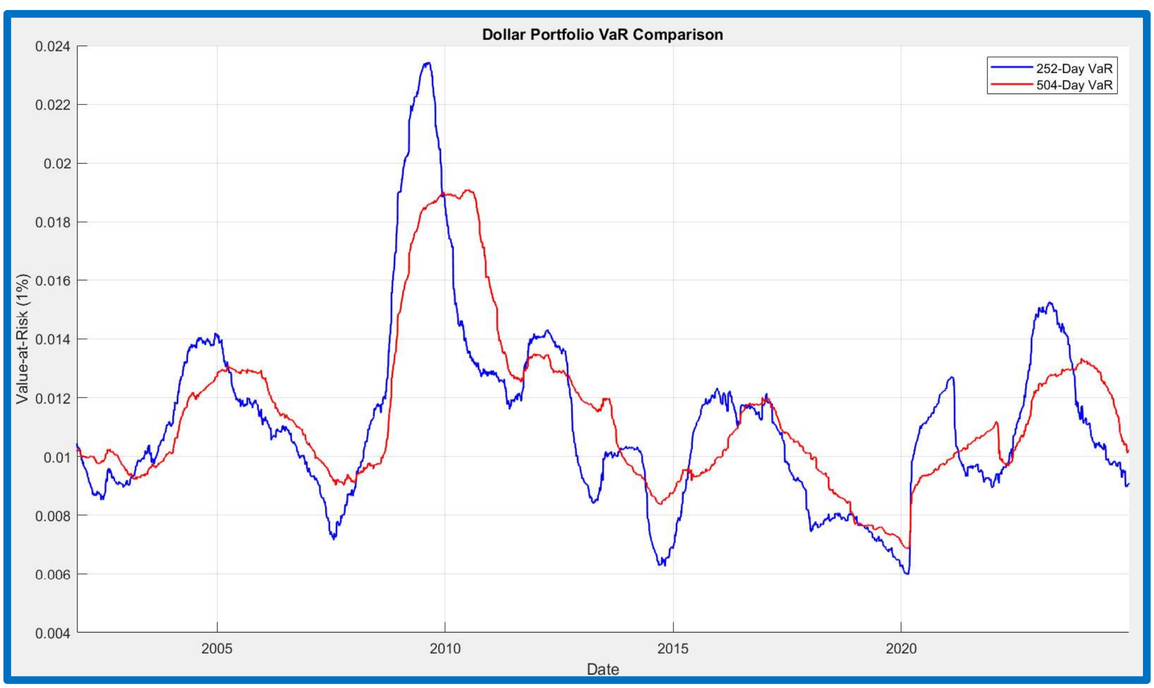

Figure 8: Dollar portfolio VaR 252 days VS 504 days.

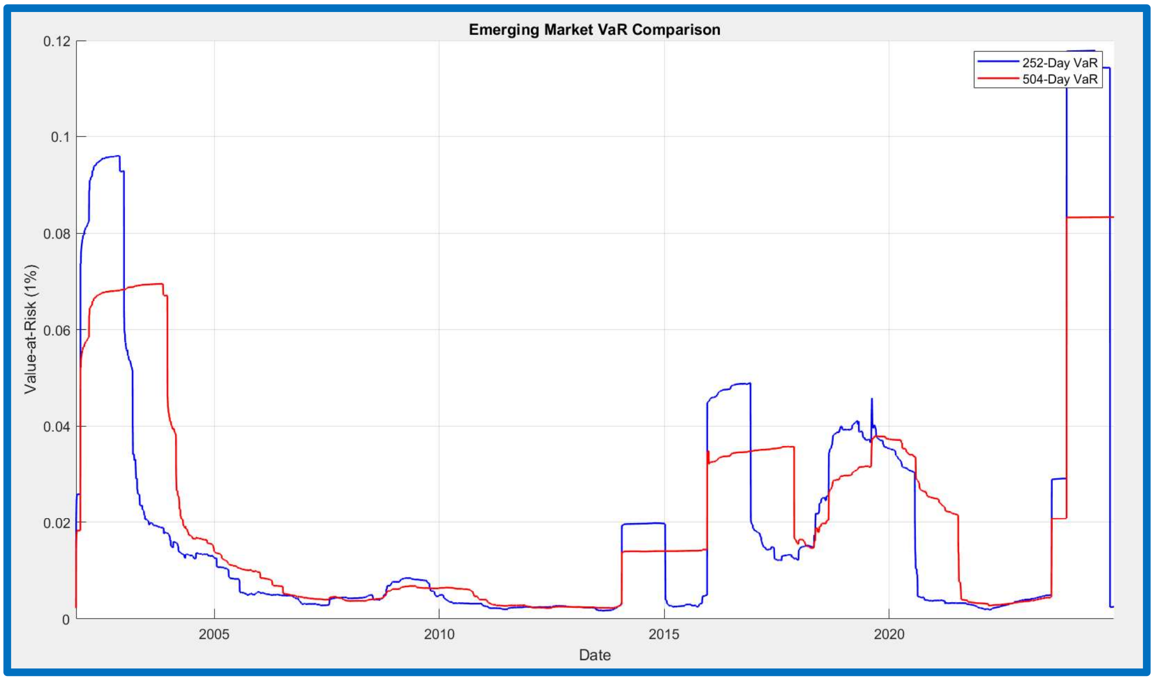

Figure 9: Emerging market VaR 252 days VS 504 days.

Back-testing

Back testing evaluates the accuracy of VaR forecasts by comparing breaches i.e observed losses exceeding VaR against expected frequencies. The Kupiec Proportion of Failures (POF) test and Christoffersen's tests for independence and conditional coverage were employed for both portfolio types.

The Kupiec POF results reveals the number of breaches for each model (Figure 10). For the dollar portfolio, GARCH-based VaR aligns closely with the expected breach frequency, while MA and EWMA overestimate the risk. Historical simulation provides scarce estimates but often exceeds the expected threshold.

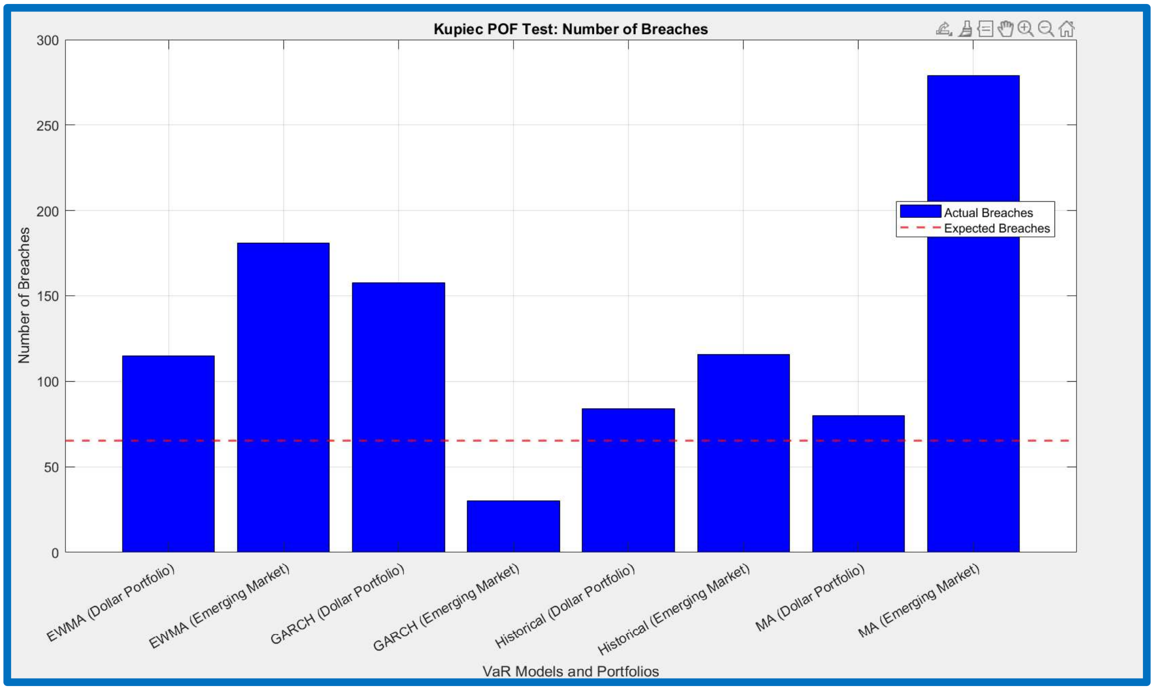

Figure 10: Kupiec POF Test

Christoffersen's conditional coverage test results further expand and validate the GARCH model, which consistently achieves lower likelihood ratio statistics, indicating better independence and conditional coverage. The MA and EWMA models fail to meet these criteria, highlighting their limitations in adapting to sudden market shifts (Figure 11).

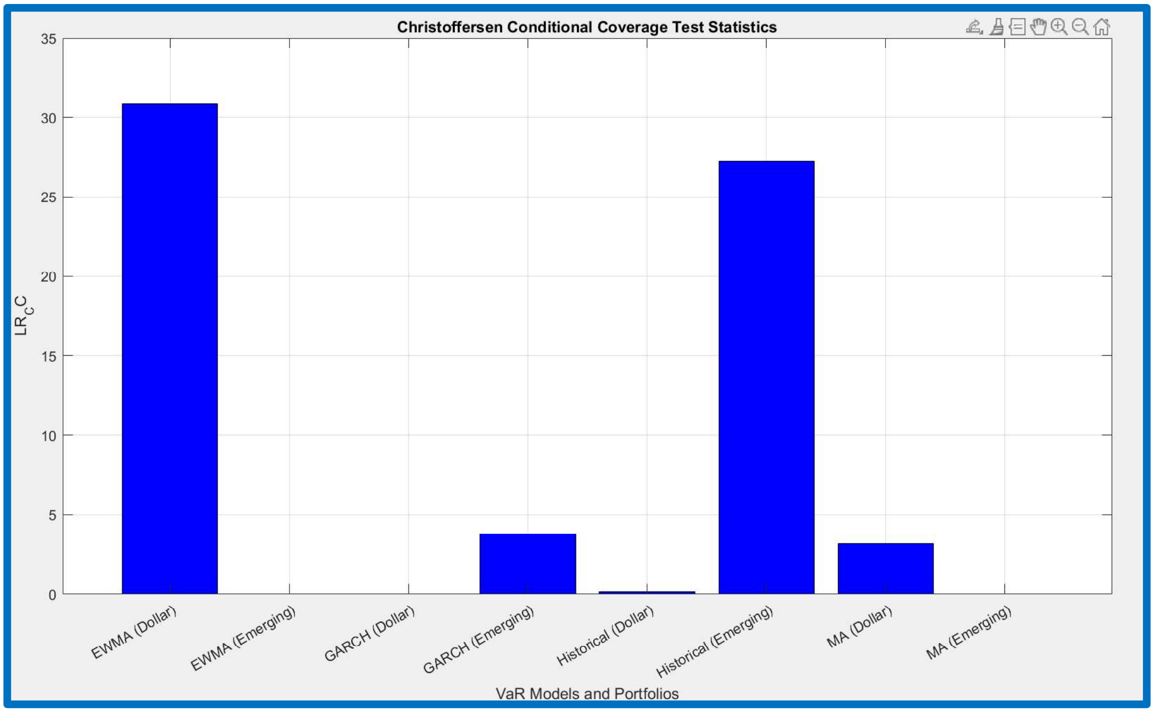

Figure 11: Christofferson Test Statistic

Discussion

Volatility & VaR Analysis

GARCH, EWMA, MA, and Historical Simulation offer distinct approaches to volatility estimation, each with unique strengths and limitations. The volatility forecasts for the Dollar Portfolio and Emerging Market Portfolio illustrate clear differences in adaptability across market conditions. The Moving Average (MA) model, which uses a fixed window, produces stable but rigid forecasts, failing to adjust to sudden market shocks. This limitation is evident in Figure 4 where MA underestimates risk during extreme market events like the 2008 financial crisis and COVID-19 pandemic. In contrast, the Exponentially Weighted Moving Average (EWMA) assigns greater weight to recent observations, offering more responsiveness. The EWMA-based volatility estimates show delayed but observable reactions to market shifts, particularly in periods of prolonged volatility, though they still lag more dynamic approaches.

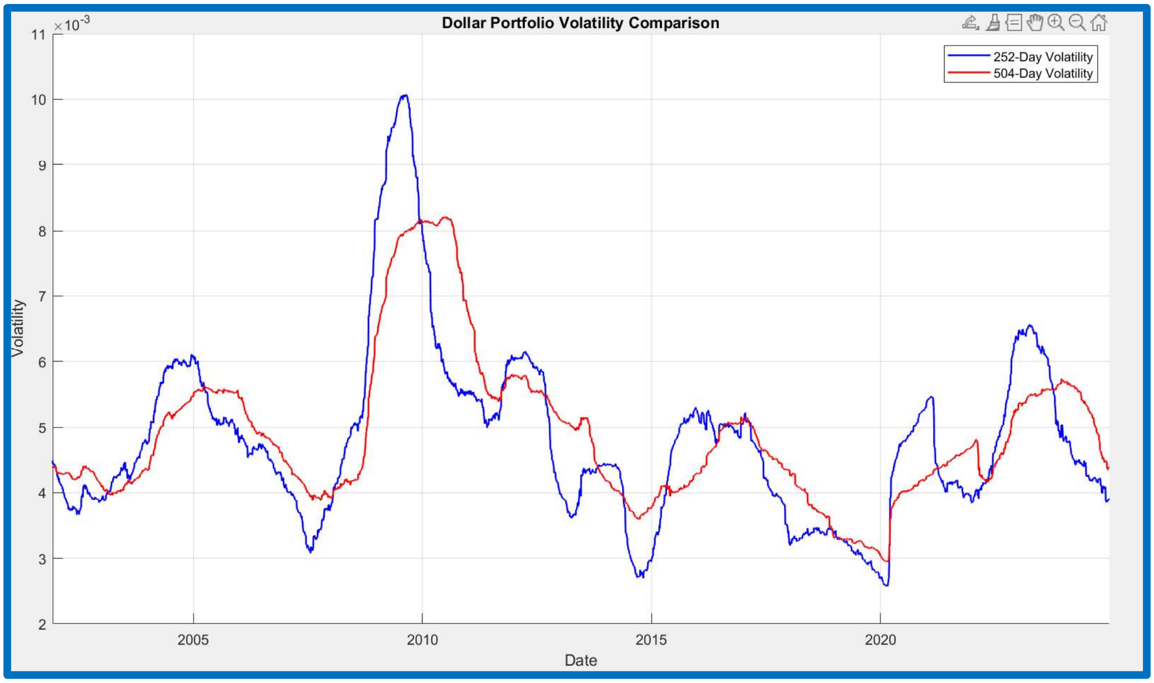

Figure 12: Dollar Portfolio Volatility Comparison

The GARCH(1,1) model outperforms both MA and EWMA by incorporating conditional variance, effectively capturing volatility clustering. This is particularly visible in emerging market volatility where GARCH exhibits rapid spikes in response to crises, reflecting the persistent shocks observed in emerging markets(Figure 13). The model excels in environments where volatility exhibits clustering behaviour, such as in high-risk emerging markets. Historical Simulation, in contrast, completely sidesteps volatility modelling and bases its forecasts on past observations alone, making it unreliable during unprecedented crises.

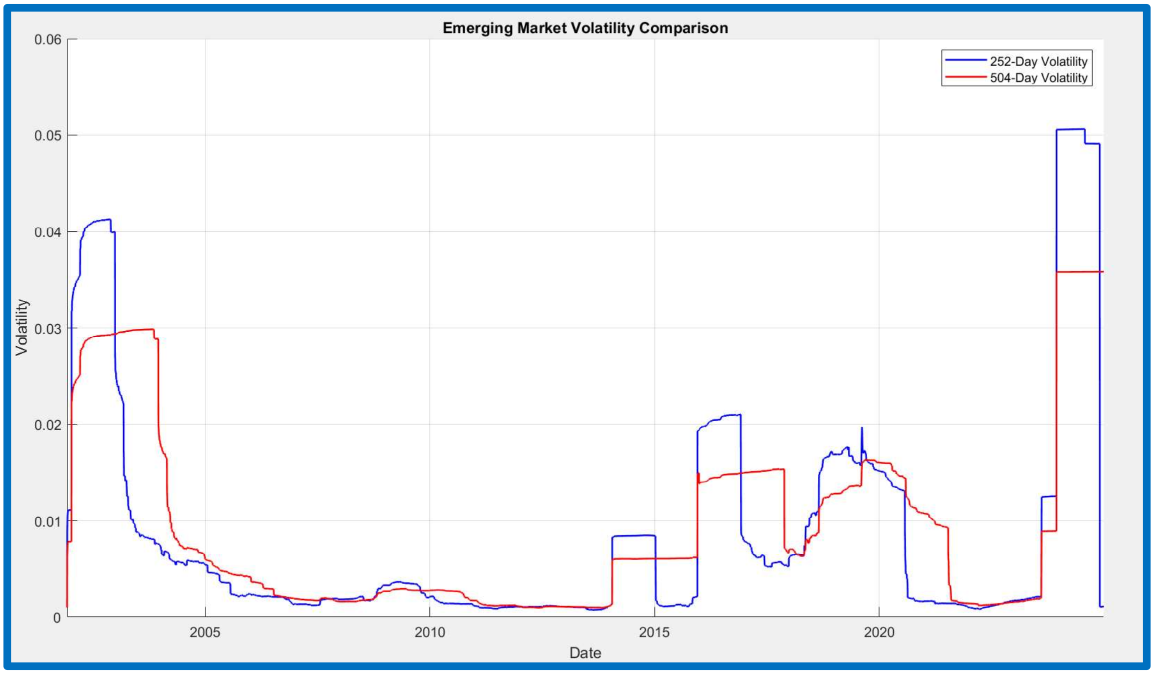

Figure 13: Emerging Market Volatility 252 Days vs 504 days.

The performance gap between these models becomes even more pronounced in Value-at-Risk (VaR) forecasting. The VaR forecasts for the Dollar Portfolio indicate that under normal conditions, all models yield reasonably aligned results. However, during volatile periods, GARCH provides the most accurate tail-risk predictions due to its ability to model dynamic risk dependencies. EWMA also performs well, but its response is slower than GARCH. MA and Historical Simulation, however, fail to capture tail risks adequately—MA due to its fixed averaging approach and Historical Simulation due to its backward-looking nature. In figure 6 graph underscores this limitation, showing that Historical Simulation underpredicts risk during crisis periods.

For emerging markets, the divergence is even starker. In the above discusses graphs, they illustrate how GARCH remains the most adaptable, responding aggressively to volatility surges. Historical Simulation and MA both struggle, failing to capture sharp fluctuations, leading to excessive breaches. EWMA provides moderate adaptability but still lags behind GARCH.

Back-testing results

Back testing is an essential step in validating Value-at-Risk (VaR) models, it bridges theoretical predictions with real-world market outcomes. This process evaluates the performance of risk models by examining whether their forecasts accurately capture extreme market movements, providing a reliable basis for financial decision-making. Through the Bernoulli Coverage Test, Kupiec's Proportion of Failures (POF) Test, and Christoffersen's Conditional Coverage (CC) Test, this analysis examines four VaR models—MA, EWMA, GARCH, and Historical Simulation—for the Dollar Portfolio and Emerging Market Portfolio.

Dollar Portfolio

The Bernoulli Test (Appendix 1)

This measures how closely the observed proportion of VaR breaches aligns with the expected probability (1%) under correct model calibration.

1) The MA model recorded 80 breaches with a p-value of 0.0659, indicating marginal acceptability at the 5% significance level. This suggests that while the model captures overall breach frequency moderately well, slight deviations exist.
2) The EWMA model, with 115 breaches and a p-value of 0.0000000006, significantly underperformed. This indicates consistent underestimation of risk, likely due to its exponential weighting failing to capture sudden shifts in volatility.
3) The GARCH model fared worse, with 158 breaches and a p-value of effectively zero. While GARCH is adept at modelling time-varying volatility, its performance for the Dollar Portfolio may reflect over-sensitivity to market conditions, leading to underperformance in stable periods.
4) The Historical Simulation model, with 84 breaches and a p-value of 0.0194, showed moderate performance but slightly underestimated risk.

Kupiec POF Test (Appendix 2):

This assesses whether the total number of VaR breaches statistically matches the expected failure rate, testing unconditional coverage accuracy.

1) The MA model had a test statistic of 3.1555 and a p-value of 0.0757, further supporting its acceptability for portfolios with low volatility.
2) The EWMA and GARCH models failed to meet coverage standards, with extreme test statistics of 31.2730 and infinity, respectively. These results reinforce their inadequacy in capturing stable risk dynamics.
3) The Historical Simulation model, with a statistic of 5.0072 and a p-value of 0.0252,

demonstrated reasonably acceptable performance, though it underperformed the MA model.

Christoffersen CC Test (Appendix 3):

This test examines both the frequency and clustering of VaR breaches to determine if a model captures conditional coverage and breach independence effectively.

1) The MA model exhibited better independence, with fewer clusters of breaches compared to the EWMA and GARCH models. However, it still showed limitations in adapting to sudden volatility changes.
2) The EWMA model’s reflects significant clustering in breach patterns, highlighting its inability to handle sustained volatility shocks.
3) The GARCH model, while effective for high-volatility markets, exhibited similar clustering tendencies in the Dollar Portfolio, with a statistic of 24.

Emerging Market Portfolio

The Emerging Market Portfolio’s high volatility presented greater challenges for all models.

Bernoulli Coverage Test:

1) The MA model, with 279 breaches, failed significantly, as evidenced by a p-value of nearly zero. This underperformance reflects its inability to adjust to extreme volatility spikes.
2) The EWMA model, with 181 breaches, also performed poorly, with a negligible p-value. Its exponential weighting did not adequately account for the high volatility and sharp movements characteristic of emerging markets.
3) The GARCH model, with only 30 breaches and a p-value of 0.0000117, emerged as the best-performing model. Its conditional variance framework allowed it to adapt effectively to dynamic volatility.
4) The Historical Simulation model, with 116 breaches, showed moderate performance but struggled to account for extreme tail events, as reflected in its low p-value.

Kupiec POF Test:

1) The MA and EWMA models returned Na for both the test statistic and p-value, reflecting that they could not provide meaningful results for the Emerging Market Portfolio..
2) The GARCH model, with a test statistic of 24.0376 and a p-value of 0.0000009,

performed far better than the other models, aligning closer to expected breach rates.

3) The Historical Simulation model, with a test statistic of 32.4316, failed to meet coverage standards, demonstrating the limitations of non-parametric approaches in volatile environments.

Christoffersen CC Test:

1) The GARCH model stood out with its relatively low statistic of 4, indicating minimal clustering and strong performance in capturing both breach frequency and independence.
2) The MA and EWMA models exhibited high LR_{CC} values, reflecting persistent clustering and a failure to model the time-dependent volatility present in emerging markets.

The analysis of the Bernoulli, Kupiec POF, and Christoffersen tests provide a robust evaluation in capturing and predicting risk dynamics. The Bernoulli Test, which evaluates how closely the proportion of breaches aligns with the expected 1% VaR threshold, demonstrates GARCH's strong performance in the Emerging Market portfolio, with only 30 breaches compared to significantly higher counts for other models, such as 279 for MA. This highlights GARCH's ability to effectively adapt to extreme volatility and tail risks. However, for the Dollar Portfolio, the GARCH model recorded 158 breaches, indicating over-sensitivity to market conditions during stable periods. Historical Simulation, with 84 breaches for the Dollar Portfolio and 116 for the Emerging Market, provided moderately conservative estimates but struggled in dynamic environments, consistently underestimating risk during volatility spikes.

The Kupiec POF Test, which assesses unconditional coverage by comparing total breaches to expected frequencies, reinforces GARCH's superiority. For the Emerging Market portfolio, GARCH achieved a statistically robust alignment with expected breaches, reflected in its test statistic of 24.0376 and a highly significant p-value. In contrast, MA and EWMA models failed to meet unconditional coverage standards, with NA statistics indicating poor applicability under extreme conditions. Historical Simulation also struggled, with its statistic of 32.4316 highlighting its limitations in volatile markets.

The Christoffersen Test, which evaluates both independence and conditional coverage, underscores GARCH's reliability. For the Emerging Market portfolio, GARCH achieved a low LR_{CC} statistic of 3.78, indicating minimal clustering and strong conditional accuracy. Conversely, MA and EWMA models exhibited high LR_{CC} values, reflecting persistent clustering and poor adaptation to volatility surges. Historical Simulation, while effective in stable markets, failed to maintain independence or conditional coverage under extreme conditions.

The Independence test p-values (Appendix 1) offer valuable insights into the ability of models to capture clustering in breaches. For the Dollar Portfolio, GARCH shows the strongest performance with a p-value of 0.0253, indicating it effectively captures independence, outperforming EWMA (0.0833) and MA (0.3173). In the Emerging Market Portfolio, GARCH also demonstrates relative superiority, with a p-value of 0.0455, highlighting its robustness in volatile environments. Collectively, these results validate GARCH as the most robust model across both portfolios, effectively balancing accuracy, responsiveness, and reliability.

Conclusion

The analysis of risk forecasting models highlights the importance of selecting an accurate methodology for managing financial risk. Through volatility modelling, VaR forecasting, and extensive back testing using Bernoulli, Kupiec, and Christoffersen tests, it was found that GARCH consistently outperformed other models in capturing conditional volatility and tail risks. Historical Simulation proved unreliable in emerging markets, while MA and EWMA lacked adaptability in high-volatility conditions. Back testing validated GARCH's predictive superiority, reinforcing its application in financial risk management. Future research should explore hybrid models and macroeconomic integrations to enhance forecasting accuracy and adaptability in evolving market conditions.

Appendix 1: Bernoulli Test.

```
Backtesting Results:

Dollar Portfolio Results:
VaR Model  Portfolio           No. of Breaches      P-Value (Bernoulli)  Independence Stat    P-Value (Independence)
MA         Dollar Portfolio    80                   0.0658624391         1                    0.3173105079
EWMA       Dollar Portfolio    115                  0.0000000006         3                    0.0832645167
GARCH      Dollar Portfolio    158                  0.0000000000         5                    0.0253473187
Historical Dollar Portfolio    84                   0.0194311043         4                    0.0455002639

Emerging Market Results:
VaR Model  Portfolio           No. of Breaches      P-Value (Bernoulli)  Independence Stat    P-Value (Independence)
MA         Emerging Market     279                  0.0000000000         9                    0.0026997961
EWMA       Emerging Market     181                  0.0000000000         10                   0.0015654023
GARCH      Emerging Market     30                   0.0000117002         4                    0.0455002639
Historical Emerging Market     116                  0.0000000003         6                    0.0143058784
```

Appendix 2: Kupiec POF Test

```
Kupiec POF Test Results:
-------------------------------------------------------------
VaR_Model      Portfolio         Breaches    Total Obs    Statistic      P-Value
-------------------------------------------------------------
MA            Dollar Portfolio    80          6522         3.1555         7.5670e-02
EWMA          Dollar Portfolio    115         6522         31.2730        2.2418e-08
GARCH         Dollar Portfolio    158         6522         Inf            0.0000e+00
Historical    Dollar Portfolio    84          6522         5.0072         2.5242e-02
MA            Emerging Market     279         6522         NaN            NaN
EWMA          Emerging Market     181         6522         NaN            NaN
GARCH         Emerging Market     30          6522         24.0376        9.4471e-07
Historical    Emerging Market     116         6522         32.4316        1.2346e-08
```

Appendix 3: Chirstoffersen test

VaR_Model	Portfolio	LR_Ind	P_Value_Independence	LR_CC	P_Value_Conditional_Coverage
MA	Dollar Portfolio	-0.000357393	1	3.155175147	0.206472598
EWMA	Dollar Portfolio	-0.422234749	1	30.85071584	1.99918E-07
GARCH	Dollar Portfolio	0	1	24	0
Historical	Dollar Portfolio	-4.832798754	1	0.174382821	0.916501651
MA	Emerging Market				
EWMA	Emerging Market				
GARCH	Emerging Market	-20.25680803	1	3.780831925	0.151008982
Historical	Emerging Market	-5.212983756	1	27.21860459	1.22901E-06

Reference list

Review of Financial Studies. 2025. *Risk Forecasting, Volatility Modeling, and Value-at-Risk: An Empirical Assessment of GARCH and Alternative Models* [Online]. Oxford: Oxford University Press. [Accessed 4 February 2025]. Available from: https://academic.oup.com/rfs

https://www.cesarerobotti.com/research

Contains a plethora of useful papers and includes various Matlab code.