Dynamic Volatility Arbitrage Strategy: Harnessing EGARCH Forecasting and VIX Discrepancies in Market Trading

Xianyang (Eric) Hu, Xinran (Kevin) Li, Chang (Rex) Liu

December 2023

Contents

1 Introduction and Assumptions

We aim to implement a trading strategy using volatility arbitrage. Volatility arbitrage involves exploiting differences between the implied volatility of options and the expected future volatility. We use statistical models to predict future volatility and then compare these predictions with the implied volatility (VIX) reflected in current option prices. The forecasted volatility is annualized and compared with the VIX index, a popular measure of the stock market's expectation of volatility based on S&P 500 index options. A dynamic trading rule is applied: The strategy takes a 'Long' position if the forecasted error (the difference between the EGARCH forecast and the VIX) is greater than a certain threshold, and a 'Short' position otherwise.s The threshold is determined by the standard deviation of the forecast errors multiplied by a threshold factor. This strategy hinges on the belief that the options market might misprice the true level of futurep volatility.

The following foundational assumptions are posited:

- *Market Efficiency.* Markets are presumed efficient, incorporating all available information into asset prices. This assumption is pivotal as it predicates the identification of volatility arbitrage opportunities on anomalies rather than systemic mispricing.
- *Liquidity.* Assets and options targeted are assumed to exhibit high liquidity, ensuring minimal market impact on entry and exit of positions, and maintaining narrow bid-ask spreads.
- *No Arbitrage.* It is posited that under normal market conditions, arbitrage opportunities do not present risk-free profits. This aligns with the theoretical fair pricing of options and derivatives.
- *Volatility Predictability.* The strategy presupposes the statistical predictability of future market volatility, essential for comparing forecasted volatility against current implied volatility.
- *Rational Market Participants.* Market participants are assumed rational, basing decisions on available data and logical assessments. This underpins the predictability of market responses to stimuli.
- *Stable Regulatory Environment.* The strategy assumes a consistent regulatory framework, essential for the stability and predictability of market operations and trading conditions.

These assumptions form the theoretical bedrock for our strategy's design, implementation, and anticipated performance within the financial market context.

2 Data Collection and Preprocessing

Utilizing the 'yfinance' library, we download the S&P 500 Index data, denoted by 'ĜSPC', for a specified time period. This library facilitates direct access to Yahoo Finance's historical data, focusing on the daily closing prices. Post data acquisition, we preprocess it to calculate log returns, an essential step for financial analyses involving volatility. The log returns are computed as the natural logarithm of the ratio of consecutive day's closing prices. This computation not only aligns with models that assume normally distributed returns, like GARCH, but also is crucial for meaningful comparisons with volatility measures such as the VIX, ensuring a continuous and stabilized variance of returns over time.

Specifically, we have VIX designed to be an estimate of the realized volatility (RV) over the next 30 trading days and is approximately given by,

$$\text{VIX} \approx 100 \times \hat{\sigma}, \quad \text{where} \quad \hat{\sigma}^2 \approx \mathbb{E}_{\mathbb{Q}}\left[RV^2\right].$$

The definition of realized volatility RV over the next n trading steps is given by,

$$RV := \text{ annualized standard deviation of } \{R_0, \ldots, R_{n-1}\}$$

where $R_i = \ln\left(S_{t_{i+1}}/S_{t_i}\right)$ is the log-return of asset from day t_i to day t_{i+1}. Then, $\Delta t := t_{i+1} - t_i$ corresponds to the length of one step, and the annualized variance of each R_i is equal to $\mathbb{E}_{\mathbb{Q}}\left[\text{var}\left(R_i\right)\right]/\Delta t$, where the expected value is taken under the risk-neutral probability. Accordingly, we define $\hat{\sigma}^2$ by,

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=0}^{n-1}\frac{\mathbb{E}_{\mathbb{Q}}\left[\text{var}\left(R_i\right)\right]}{\Delta t}.$$

Above formulae for VIX and $\hat{\sigma}^2$ are formal and motivates the precise definition given by CBOE.

We divide the S&P 500 log returns data into distinct sets for robust model training and evaluation: 60% is used as the training set to develop and calibrate the model, the subsequent 20% forms the validation set for model tuning and intermediate evaluation, and the final 20% serves as the test set. This partition ensures a comprehensive approach to model training, allows for effective hyperparameter tuning during validation, and provides a reliable assessment of the model's predictive performance on unseen data, thus ensuring both accuracy and generalizability in real-world scenarios.

3 Model Selection

3.1 GARCH-type models

In the development of our strategy, Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, encompassing GARCH, EGARCH, T-GARCH, and IGARCH, stand out as our initial choice for volatility forecasting. These models are chosen for their advanced capability to capture the dynamic nature of financial market volatility, including the persistence of volatility over time and its asymmetric response to market news. This aligns well with our strategy's need for precise volatility forecasting and exploiting discrepancies between forecasted and implied volatilities in option pricing.

- The typical *GARCH model* captures volatility clustering. Its variance equation is typically expressed as

  $$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{1}$$

  where p, q are the order of the model, and α_i, β_j are parameters.
  GARCH is the baseline model for volatility forecasting and is effective in capturing the clustering of volatility that is common in stock market returns.

- *Exponential GARCH models* (EGARCH) allow for asymmetric responses to shocks, crucial for modeling leverage effects.

  $$\log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{p} \alpha_i g(\epsilon_{t-i}) + \sum_{j=1}^{q} \beta_j \log(\sigma_{t-j}^2) \tag{2}$$

  where $g(\epsilon)$ captures the asymmetry in volatility.
  It accounts for the asymmetric impact of positive and negative shocks (leverage effect), which is often observed in equity markets.

- *Threshold GARCH models* (TGARCH) explicitly model the asymmetric impact of positive and negative returns.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p}(\alpha_i \epsilon_{t-i}^2 + \gamma_i \epsilon_{t-i}^2 I_{[\epsilon_{t-i}<0]}) + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{3}$$

where $I_{[\cdot]}$ denotes the indicator function for the threshold effect.
Like EGARCH, TGARCH models the asymmetric volatility response to positive and negative returns, which can be important for stock indices.

- *Integrated GARCH models* (IGARCH) model persistent long-term volatility, indicating enduring shocks.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + (1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j)\sigma_{t-1}^2 \tag{4}$$

It is suitable for modeling long-term persistence in volatility, which can be a feature of major stock indices.

3.2 Maximum Likelihood Estimation in Model Training

In the training of GARCH family models, Maximum Likelihood Estimation (MLE) is employed to determine the optimal parameters of the model, which include α_0, $\alpha_1, \ldots, \alpha_p$, and $\beta_1, \ldots, \beta_q$. MLE is a statistical method used to estimate the model parameters by maximizing the likelihood function, which represents the probability of the observed data given the parameters.

For GARCH models, the likelihood function can be expressed in terms of the conditional distribution of the returns given the past information. Assuming the standardized residuals are normally distributed, the log-likelihood function for a GARCH(p, q) model can be formulated as:

$$\ln L(\alpha, \beta) = -\frac{n}{2}\ln(2\pi) - \sum_{t=1}^{n} \ln(\sigma_t^2) - \sum_{t=1}^{n} \frac{\epsilon_t^2}{\sigma_t^2}, \tag{5}$$

where σ_t^2 is the conditional variance defined as in equation (1), and ϵ_t are the model residuals.

The MLE process involves finding the set of parameters that maximize this log-likelihood function. This optimization is typically performed using numerical techniques since an analytical solution is not feasible for most GARCH-type models.

3.3 Hyperparameter Tuning with MLE

Hyperparameter tuning leverages MLE to determine the order of the GARCH model, denoted by p and q, and the asymmetric response parameter o for TGARCH models. The grid search performed over potential values of p, q, and o involves fitting the GARCH model to the training data using MLE for each combination and evaluating the goodness of fit using the Akaike Information Criterion (AIC). The AIC incorporates the maximized log-likelihood and penalizes the number of parameters, balancing fit and complexity:

$$AIC = 2k - 2\ln(L(\hat{\alpha}, \hat{\beta})), \tag{6}$$

where k is the number of parameters, and $L(\hat{\alpha}, \hat{\beta})$ is the likelihood function evaluated at the estimated parameters.

3.4 Model Validation and Testing

After selecting the best model based on AIC, the chosen model undergoes validation and testing. The rolling forecast methodology is applied to the validation and test sets to predict future volatility. The predictive performance of the model is quantitatively assessed using the RMSE on the out-of-sample test data:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{\sigma}_t^2 - \sigma_{t,\text{actual}}^2)^2}, \tag{7}$$

where $\hat{\sigma}_t^2$ is the forecasted conditional variance, and $\sigma_{t,\text{actual}}^2$ is the actual realized variance. Models with lower RMSE values are considered better at capturing the dynamics of volatility. Thus, the model training, selection, and validation process is grounded on rigorous statistical estimation techniques, ensuring the reliability of the volatility forecasts produced by the GARCH family models.

4 Strategy Formulation

4.1 Analysis of SPX/VIX Correlation

We first conducted a detailed analysis of the correlation between the S&P 500 Index (SPX) and the Volatility Index (VIX) over various time frames, including 1 month, 3 months, 6 months, and 1 year. Using yfinance for data retrieval, we observed a typically negative correlation between SPX movements and expected market volatility as reflected by the VIX. This inverse relationship (see Figure 1) is particularly evident in the short term and aligns with the market logic where rising volatility expectations coincide with market downturns. Since VIX is a measure of future volatility, so when the market falls, there's greater future volatility. If a market rises, usually it's assumed that future volatility is less.

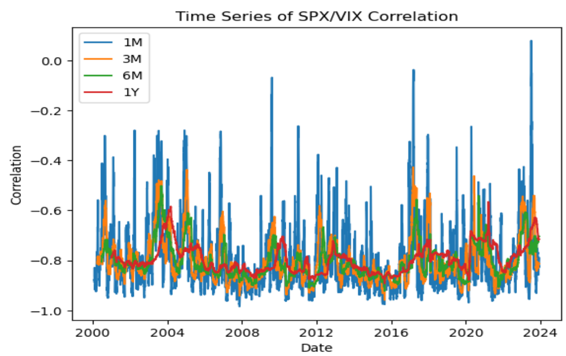


Figure 1: SPX/VIX Correlation with respect to time

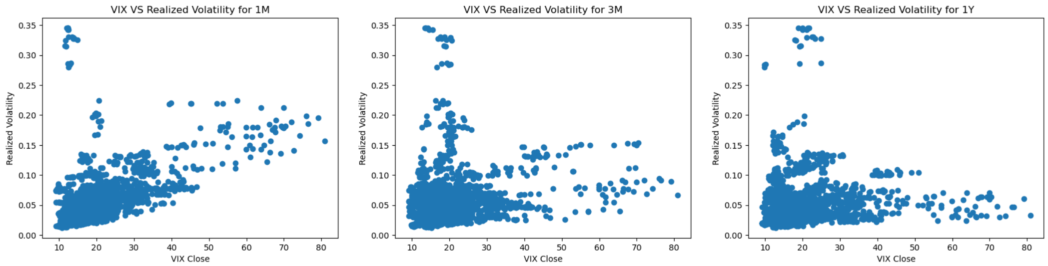
4.2 Realized Volatility of S&P 500

Further, we compared the VIX with the S&P 500's realized volatility over different periods. At a 1-month horizon, the VIX showed a consistent but not perfect prediction of short-term volatility, with occasional significant deviations. Over 3 months, the data suggested two distinct predictive trends, indicating the VIX's less straightforward predictability in medium-term scenarios. Extending to a 1-year period, the relationship deviated from linearity, suggesting the VIX's reduced efficacy in forecasting long-term volatility. These findings highlight the complexity of volatility forecasting and the importance of adapting strategies to different market conditions and time horizons. See Figure 2.

4.3 Optimizing and Evaluating GARCH-Type Models

In this subsection, we focus on selecting the best parameters for various GARCH-type models (GARCH, EGARCH, TGARCH, IGARCH) and analyzing their performance using historical S&P 500 data.

Recall from §3 that the dataset is partitioned into training (60%), validation (20%), and testing (20%) sets. This split allows us to rigorously tune the parameters



(a) (b) (c)

Figure 2: VIX vs. Realized Volatility of S&P 500 for different time horizons

for each model on the validation set and then evaluate their performance on the testing set. Using the 'arch' Python library, we employ a systematic approach to determine the best parameters for each model type. This process involves iterating over a range of parameter values and selecting the combination that yields the lowest Akaike Information Criterion (AIC) score, indicating the best fit to the data.

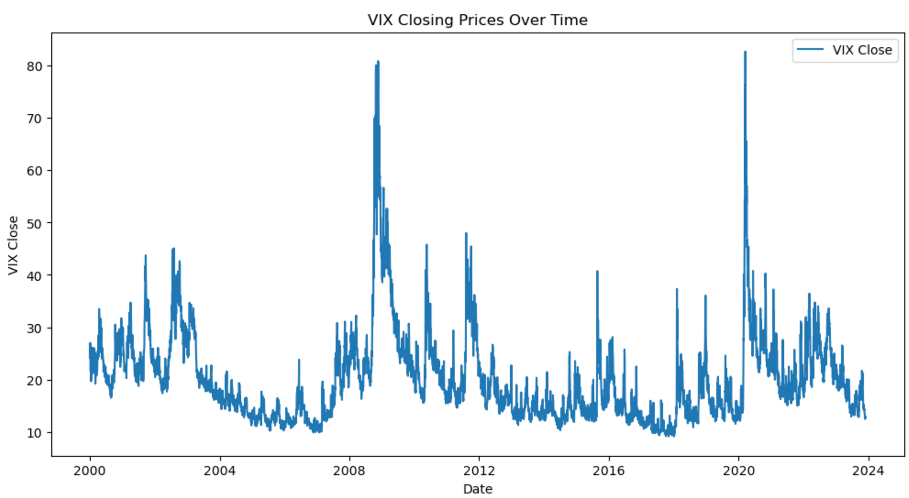
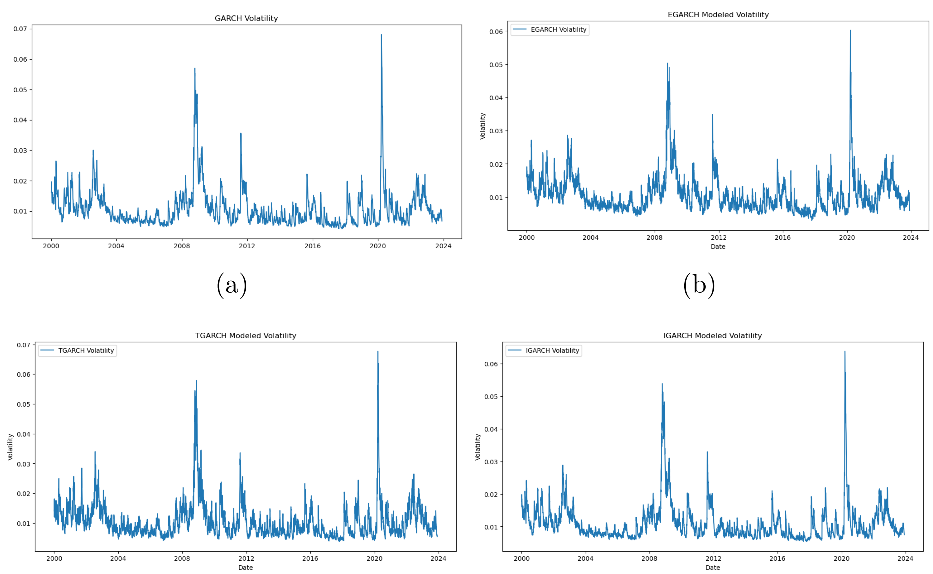


Figure 3: VIX Closing Prices over Time

For the GARCH and EGARCH models, we explore various combinations of the order of ARCH (p) and GARCH (q) terms. For the TGARCH model, which accounts for the asymmetric effects of shocks on volatility, we additionally iterate over the order of asymmetry terms (o). The IGARCH model, a special case of the GARCH model with the constraint that the sum of ARCH and GARCH coefficients equals one, is also included in our analysis. The results are shown as follows:

```
GARCH Best Params: (4, 0, 2)
EGARCH Best Params: (8, 0, 4)
TGARCH Best Params: (1, 4, 3)
IGARCH Best Params: 10
```



(a) (b) (c) (d)

Figure 4: Prediction by Different GARCH-Type Models with Best Parameters

Once the best parameters are identified for each model, we fit these models to the training data and perform rolling forecasts on the testing set. The Root Mean Square Error (RMSE) is computed for each model's forecasts against the actual data, offering a quantitative measure of their predictive accuracy. Additionally, we report the AIC and Bayesian Information Criterion (BIC) scores for each model, providing further insights into their relative performance.

```
GARCH - RMSE: 0.01370315643627133 AIC: 14341.8662586546 BIC: 14380.739467117168
EGARCH - RMSE: 0.013672341973333101 AIC: 14469.60378647515 BIC: 14534.392467246096
TGARCH - RMSE: 0.013677621956685647 AIC: 13820.151611805783 BIC: 13884.940292576728
IGARCH - RMSE: 0.013671882180070151 AIC: 14028.242046537469 BIC: 14105.988463462603
```

The results demonstrate varying levels of performance across different models, with each model capturing different aspects of the volatility dynamics in the SPX. This comparative analysis not only sheds light on the effectiveness of these models in volatility forecasting but also provides valuable insights into the complex behavior of financial market volatility. One should note that the EGARCH model has the superior RMLE, AIC, and BIC performance.

4.4 Strategy Simulation with EGARCH Forecasts

In this section, we detail a dynamic trading strategy using EGARCH model forecasts to exploit volatility differences between the model's predictions and the actual market volatility indicated by the VIX index. The strategy hinges on a rolling forecast method, continuously updating the EGARCH model with the latest data and aligning these forecasts with the VIX for direct comparison.

A key feature of our strategy is the implementation of a threshold factor within our dynamic trading rule. This factor determines when to switch between 'Long' and 'Short' positions based on the standard deviation of the forecast errors (the difference between EGARCH forecasts and the VIX). This threshold-based approach allows for adaptive positioning in response to market volatility fluctuations.

Position sizes are dynamically adjusted according to the level of forecasted volatility, balancing potential returns with associated risks. The strategy's effectiveness is evaluated through both the total simulated PnL and visual analysis, where annualized EGARCH forecasts are plotted against the VIX (Figure 5), offering insights into the strategy's performance in navigating market volatility.

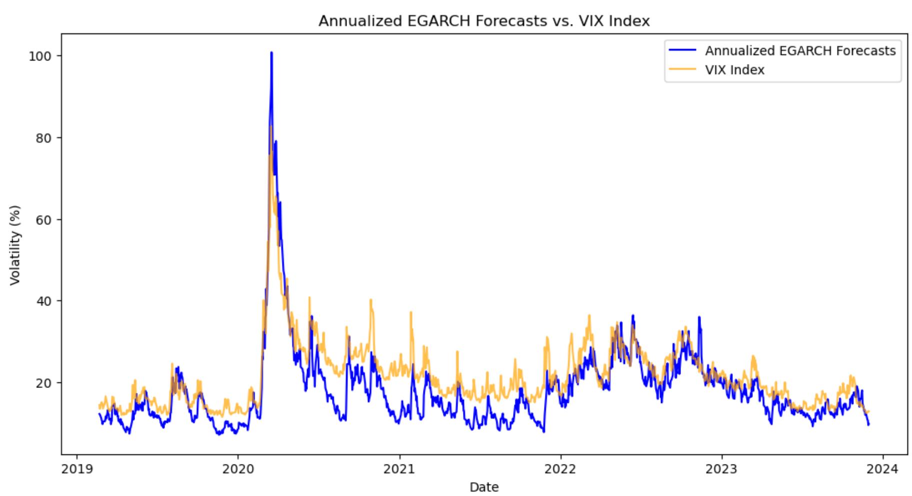


Figure 5: Annualized EGARCH Forecasts compared to VIX Index

5 Backtesting

Our backtesting process involves applying the strategy on a rolling window basis to out-of-sample data, allowing us to evaluate its performance over time and under varying market conditions. We also adjust for transaction costs and slippage to ensure a more realistic assessment of the strategy's profitability. Before start, we introduce some performance metrics.

5.1 Performance Metrics

- *Sharpe Ratio* is essential for assessing the risk-adjusted return of the strategy. It compares the strategy's excess returns (over a risk-free rate) to its volatility, offering a measure of return per unit of risk. A higher Sharpe Ratio indicates a more favorable risk-reward balance.
- *Alpha* measures the strategy's excess return over a benchmark (typically the market index or a similar asset). It represents the value added by the strategy's unique approach, after accounting for market movements. Positive alpha indicates outperformance relative to the benchmark.
- *Beta Neutrality.* This concept is crucial for strategies aiming to be market-neutral. A beta-neutral strategy seeks to have a beta (a measure of sensitivity to market movements) close to zero, implying that its performance is largely independent of market swings. This is often a desired attribute in hedging strategies or those aiming to minimize systemic risk.
- *Transaction Costs.* Including transaction costs in the analysis is vital for a realistic assessment of a trading strategy. Costs such as brokerage fees, bid-ask spreads, and slippage can significantly impact net returns, especially in strategies with frequent trading.

5.2 Results and Strength/weakness Analysis

We backtested the above strategy using the last 20% of the data as the test set.

- The strategy yielded a beta of $\boxed{-3.43}$, indicating an inverse relationship with the market, and an alpha of $\boxed{0.00\%}$, suggesting returns in line with the risk-free rate after adjusting for market exposure.

- The strategy achieved an annualized return rate of $\boxed{14.56\%}$, showing strong performance over the long term. The cumulative PnL is visualized in Figure 6.
- With a Sharpe Ratio of $\boxed{0.4747}$, the strategy's risk-adjusted return is moderately positive, indicating a decent return per unit of risk taken.
- The strategy's performance was further evaluated considering transaction costs of 0.05 basis points per trade. This adjustment resulted in a total PnL with transaction costs being *significantly lower*, highlighting the impact of trading expenses on net profitability.

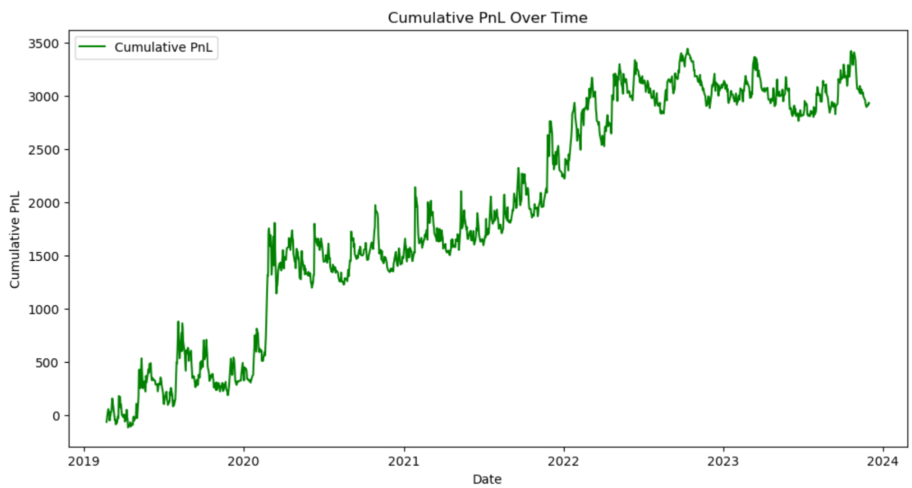


Figure 6: Cumulative PnL over Time

Based on the results, we see that the strategy demonstrated a high annualized return rate and the ability to capitalize on volatility discrepancies. Additionally, the use of a dynamic trading rule based on EGARCH forecasts allowed for adaptive positioning in varying market conditions.

However, the negative beta value indicates a significant inverse market correlation, suggesting potential underperformance in bullish market conditions. The Sharpe Ratio, while positive, suggests moderate efficiency in risk-adjusted terms, indicating room for improvement in managing volatility and drawdowns. The inclusion of transaction costs noticeably reduced the strategy's profitability, highlighting the impact of trading expenses.

Bibliography

[BD21] Pedro Barroso and Andrew Detzel. Do limits to arbitrage explain the benefits of volatility-managed portfolios? *Journal of Financial Economics*, 140(3):744–767, 2021. doi:10.1016/j.jfineco.2021.02.009.

[PS16] Ioannis Psaradellis and Georgios Sermpinis. Modelling and trading the u.s. implied volatility indices. evidence from the vix, vxn and vxd indices. *International Journal of Forecasting*, 32:1268–1283, 10 2016. doi:10.1016/j.ijforecast.2016.05.004.

[SYY15] Robert F. Stambaugh, Jianfeng Yu, and Yu Yuan. Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, 70(5):1903–1948, 2015. URL: http://www.jstor.org/stable/43611124.