Proceedings of Machine Learning Research 204:1–11, 2023 Conformal and Probabilistic Prediction with Applications

Market Implied Conformal Volatility Intervals

Alejandro Canete ACANETE@UCHICAGO.EDU
Data Science Institute, University of Chicago, Chicago, IL, USA

Editor: Harris Papadopoulos, Khuong An Nguyen, Henrik Boström and Lars Carlsson

Abstract

Volatility is a fundamental input for pricing and risk management of financial instruments. In the following work we propose an algorithm to estimate the market implied uncertainty of future realized volatility. Our method interprets the market implied volatility as a point prediction of future realized volatility and applies online conformal prediction to estimate the uncertainty of this prediction. We analyze rolling coverage and width of several nonconformity scores over 15 years of daily data. The results suggest that conformal prediction can be used to infer market implied prediction intervals for realized volatility.
Keywords: Volatility, Conformal Prediction, Time Series

1. Introduction

Volatility as a measure of uncertainty and variability of the returns of financial assets is an active area of resarch in financial mathematics and econometrics (McAleer and Medeiros, 2008; Poon and Granger, 2003; Bhowmik and Wang, 2020), and a foundational concept for practitioners (Bennett and Gil, 2012; Sebastian and Taylor, 2022). One common way to model financial returns has the form:

$$\frac{dP_t}{P_t} = \mu_t + \sigma_t dB_t, \tag{1}$$

where P_t is the price of the asset at time t, μ_t represents the drift, B_t is a standard Brownian motion and σ_t is the instantaneous volatility, which is unobservable. The historical volatility (Ederington and Guan, 2006) can be estimated given an historical sample of returns.

The realized volatility is a class of estimators that use intraday data (McAleer and Medeiros, 2008; Gatheral and Oomen, 2010) or a small sample of daily data (Sebastian and Taylor, 2022) to infer the instantaneous volatility of a particular trading day.

Forecasting future historical volatility is fundamental for practitioners since the pricing of expected risk exposure is done in a forward looking way (Bennett and Gil, 2012). The forecast of one day ahead realized volatility is particularly relevant now that the volume share of options that expire on the same day has grown from less than 5% in 2016 to more than 40% by the end of 2022 (Brogaard et al., 2023; Beckmeyer et al., 2023).

The implied volatility, is the input volatility that matches the Black-Scholes theoretical price of a European financial option with its market counterpart (Hagan et al., 2002; Stefanica and Radoičić, 2017; Gatheral, 2011). The relationship between implied volatility and realized volatility has been studied from different angles and approaches (Christensen and Prabhala, 1998; Liang et al., 2020; Gatheral, 2008). In this study we use conformal prediction (Vovk et al., 2005; Fontana et al., 2023; Angelopoulos and Bates, 2021) as another approach to study its relationship.

2. Methodology

The proposed approach builds online one day ahead realized volatility prediction intervals based on the previous day's implied volatility. Proxies for realized and implied volatility were used. Normalizers based on historical statistics and implied by the market were evaluated within a conformal forecasting setting.

2.1. Volatility Proxies

Following the approach in (Aït-Sahalia and Kimmel, 2007) we use the VIX volatility index (Whaley, 2009) as a proxy for the implied volatility σ:

$$\sigma_t \sim VIX_t, \tag{2}$$

we use the SPDR S&P 500 Exchange-Traded Fund (SPY ETF) as a proxy for the S&P 500 index (Guo et al., 2021), and we use the Garman-Klass Yang-Zhang extension formula (Fałdziński and Osińska, 2016), with $N = 1$, as a proxy of the realized volatility for a single day:

$$h_t = \sqrt{252}\sqrt{\ln^2\left(\frac{O_t}{C_{t-1}}\right) + \frac{1}{2}\ln^2\left(\frac{H_t}{L_t}\right) - (2\ln(2) - 1)\ln^2\left(\frac{C_t}{O_t}\right)}, \tag{3}$$

where O_t, H_t, L_t, C_t are the open, highest, lowest and close market prices for a given trading day t.

2.2. Base Prediction Model

The VIX is considered a biased (and inefficient) estimator of future volatility (Bennett and Gil, 2012) so as in (Christensen and Prabhala, 1998) we build a simple point forecast based on VIX:

$$\hat{h}_{t+1} = f_t(\text{VIX}_t) = a_{t-1}\text{VIX}_t + b_{t-1}, \tag{4}$$

where a_{t-1} and b_{t-1} are updated on-line using exponential moving average regression as in (Loveless et al., 2013).

2.3. Data

We will be working with daily end-of-day data from `https://finance.yahoo.com/` for the VIX values and SPY prices and with daily end-of-day data from `https://www.cboe.com/` for values of the SKEW and VVIX indices (Hung et al., 2022). The full sample consist of 4223 days starting on 2006/03/15 and ending on 2023/03/17.

2.4. Online Conformal Prediction Approach

We will be using the ACI framework of (Gibbs and Candes, 2021) with some ideas from (Barber et al., 2022) to build a prediction interval $C_t(\alpha)$ for trading day t. We will assume the trader has a target level of coverage $\alpha \in (0, 1)$, such that h_t belongs to $\hat{C}_t^\alpha$ at least $100(1 - \alpha)\%$ of the time. The ACI framework continosuly adjusts α_t (and thus how the empirical quantile is computed) to account for distributional drift so that in the long run the algorithm can attain the target coverage. The proposed framework entails making the following decisions:

Decision 1 Choose a one day ahead realized volatility prediction $\hat{h}_{t-1}$ and a measure of uncertainty $u_{t-1} \in \mathbb{R}^+$ that can be known before the market opens, then define a normalized nonconformity score based on this measure of uncertainty, $s_t = S(\hat{h}_{t-1}, h_t, u_{t-1})$. This is motivated by the fact that volatility clusters in time and thus non-conformity score normalized by an uncertainty measure (coming directly from the market) should make the conditional coverage level closer to the target one. This was corroborated by our experiments.

Decision 2 Choose a weighting scheme to define an historical sample of scores $\mathcal{S}_t = \{(w_\tau, s_\tau)\}_{\tau=1}^t$, then choose a method to build a prediction interval based on the weighted empirical quantiles of the historical normalized nonconformity scores. Weighted samples allow the trader to encode prior beliefs of the relative frequency of certain market conditions.

Decision 3 Choose a coverage loss $l_t = L(h_t, \hat{C}_t(\alpha_{t-1}), \alpha)$ and a learning rate γ. Once the realized volatility is known (after the market closes), update the target confidence rate as: $\alpha_t := \alpha_{t-1} + \gamma l_t$. This is the ACI update step. This provides the coverage guarantees of the proposed approach.

Since volatility is non negative, we used a normalized nonconformity score in log space

$$s_t = \frac{|\log(h_t) - \log(\hat{h}_{t-1})|}{u_{t-1}}. \tag{5}$$

Similarly as proposed in (Barber et al., 2022) we use weighted quantiles. We use an exponentially decayinig weighting scheme by making:

$$w_i := (1-k)^{t-i},\ \forall i \leq t, \tag{6}$$

This way we give more importance to recent history while still retaining (in the sample and thus in the quantiles) tail events that happened a relatively long time ago (e.g. the pasts financial crisis). We define the prediction intervals as:

$$\hat{C}_t(\alpha_{t-1}) = \exp\left\{\log f_{t-1}\left(\mathrm{VIX}_{t-1}\right) \pm Q_{1-\alpha_{t-1}/2}\left(\frac{\sum_{i=1}^{t-1} w_i \delta_{s_i}}{\sum_{i=1}^{t-1} w_i}\right) u_t\right\}, \tag{7}$$

Where $Q_\tau(\cdot)$ denotes the τ-quantile of its argument, and δ_a denotes the point mass at a.

Once the realized volatility h_t is revealed, (once the market closes), we update α_t using the ACI update step by making $l_t := \alpha - \mathrm{err}_t$, where $\mathrm{err}_t := \mathcal{I}\left\{h_t \in \hat{C}_t(\alpha_{t-1})\right\}$ and $\mathcal{I}\{X\} \in \{0, 1\}$ is the indicator function.

3. Experiments

We evaluate the following non conformity scores defined by the uncertainty measure u_t used to normalize the base non conformity score.

BASE No normalization, i.e. $u_t = 1$.

VIX The VIX index measures the market's expectations of future volatility and was used in a similar way in (Gibbs and Candes, 2021).

VVIX The VVIX index measures of the volatility of the VIX (Park, 2015).

SKEW The SKEW index measures the potential downside risk in financial markets (Bevilacqua and Tunaru, 2021).

STD-VIX The empirical exponentially weighted standard deviation of VIX.

STD-RV The empirical exponentially weighted standard deviation of the realized volatility (h_t).

For details on how VIX, VVIX and SKEW are defined and use in practice please refer to https://www.cboe.com/indices/.

Similarly, as in (Bhatnagar et al., 2023), for all experiments we report 100 times the average of:

Width $\{\mathrm{width}_t = \max C_t(\alpha_{t-1}) - \min C_t(\alpha_{t-1})\}$,

Coverage $\left\{1 - \frac{1}{252}\sum_{\tau=t-251}^{t} \mathrm{err}_\tau \quad : \quad t > 252\right\}$,

Error $\left\{\max_{[t-251,t]} \left|\alpha - \frac{1}{252}\sum_{\tau=t-251}^{t} \mathrm{err}_\tau\right| \quad : \quad t > 252\right\}$,

All exponential memory parameters (for weights, the exponetial regression and exponential standard deviation of VIX) were set at 125 days (i.e., $k = 2.0/(125 + 1)$)) the learning parameter γ was set equal to 0.2.

Table 1 contains the metrics of the experiments with $\alpha = 0.1$, while Table 2 contains the metrics of the experiments with $\alpha = 0.2$. The mean absolute error (MAE) across all experiments was 5.107. The coverage is not materially different across different scores. The BASE, SKEW and VVIX scores have the lowest error around 1 point, while the VIX, STD-VIX and STD-H have materially higher errors (1.5X to 2.0X). Similarly the width of BASE, SKEW and VVIX is at most half the width of the other evaluated scores. The score that performs the best across width, coverage and error is the one based on SKEW, an example of how the prediction intervals looked like during the first half of 2020 are presented in Figure 1. All of the experiments were also conducted with a simple base forecasting model based on the previous day's realized volatility ($\hat{h}_t = f_{t-1}(h_{t-1})$). All of the metrics were virtually the same, with the exception of MAE, wich was slightly higher at 5.68.

We believe that SKEW outperformed the other evaluated normalizers because it captures downside risk premimum (Bevilacqua and Tunaru, 2021), and thus the tails of the normalized non-conformity scores population would be thinner and thus the finite sample empirical estimates of the quantiles would be more accurate. This was partly corroborated by the fact that the excess kurtosis of the normalized non-conformity scores was the lowest for the ones normalized via SKEW. This needs further investigation.

Table 1: Results using $\hat{h}_t = f_{t-1}(\text{VIX}_{t-1})$ and $\alpha = 0.1$

Score	Width	Coverage	Error
BASE	29.993	89.853	1.224
SKEW	**29.570**	**89.870**	**1.095**
VVIX	31.077	89.824	1.285
VIX	77.093	89.643	2.073
STD-VIX	2224.840	89.338	2.329
STD-H	338.144	89.604	2.040

Table 2: Results using $\hat{h}_t = f_{t-1}(\text{VIX}_{t-1})$ and $\alpha = 0.2$

Score	Width	Coverage	Error
BASE	22.187	79.811	1.340
SKEW	**22.024**	**79.813**	**1.314**
VVIX	23.168	79.775	1.484
VIX	28.538	79.586	2.035
STD-VIX	50.139	79.554	2.093
STD-H	73.859	79.643	1.992

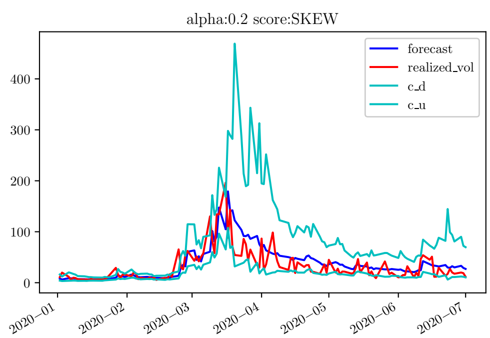

Figure 1: Prediction intervals with $\alpha = 0.2$ during the first half of 2020

3.1. Competitive Base Prediction Models

We rerun the experiments using the nonconformity score based on SKEW, but with two competitive base forecasters (both of them based on historical returns), NoVaS (Chen and Politis, 2020) and HAR (Clements and Preve, 2021). Note that for both of these approaches outperformance over GARCH(1,1) has been reported in several works (Chen and Politis, 2020; Vortelinos, 2017). We also evaluate a HAR model with VIX added as a factor, we will refer to this approach as HAR-VIX. For completeness, we also evaluated a NoVaS approach that uses VIX (for the studentization), we will refer to this approach as NoVaS-VIX. Details of these methods are found in appendix A. The results for $\alpha = 0.1$ and $\alpha = 0.2$ are presented in Tables 3 and 4, respectively. The MAE for HAR was 5.32, while the MAE for NoVaS was 6.04, and the MAE for HAR-VIX was 5.224. Across all of these models the coverage and error are fairly consistent and not materially different. HAR-VIX is the model with the smallest width among all of the evaluated approaches.

Table 3: Competitive base prediction models using SKEW based nonconformity score and $\alpha = 0.1$

Model	**Width**	**Coverage**	**Error**
HAR	28.248	89.708	1.405
HAR-VIX	**26.053**	**89.849**	**1.281**
NoVaS	32.012	89.797	1.489
NoVaS-VIX	45.813	89.718	1.600

Table 4: Competitive base prediction models using SKEW based nonconformity score and $\alpha = 0.2$

Model	**Width**	**Coverage**	**Error**
HAR	21.626	79.740	1.354
HAR-VIX	**20.181**	**79.759**	**1.316**
NoVaS	23.814	79.619	1.591
NoVaS-VIX	35.345	79.766	1.569

4. Related Work

In (Chen and Politis, 2020) a fully non-parametric realized volatility prediction interval method is proposed. The base forecaster is NoVaS (Politis and Thomakos, 2013) which has a fully online setting version and the prediction intervals are built using sequential bootstrap. The coverage of the intervals or the effect of other market measures of uncertainty (like SKEW and VVIX) are not covered.

The conditional realized volatility forecasting capability of VVIX and SKEW were studied in (Hung et al., 2022) in a parametric setting concluding that VVIX was informative while SKEW was not. Coverage of prediction intervals were also analyzed in the context of parametric VaR models.

5. Conclusions and Future Work

In the present work we used the ACI framework (Gibbs and Candes, 2021) with an exponentially decaying weighting scheme (Barber et al., 2022) to build market implied realized volatility predictive intervals. Several simple (and competitive) base prediction models were evaluated and several empirical and market based measures of uncertainty were used to define different nonconformity scores. The ACI framework performs consistently in terms of coverage regardless of the based predictor used. The discriminant factor across all evaluated approaches became the average width of the generated intervals. SKEW as an uncertainty measure within normalized nonconformity scores outperformed (in terms of coverage and width) the other evaluated measures. For the competitive base predictors used (NoVaS and HAR) the addition of VIX was always beneficial. Across all of the evaluated approaches HAR-VIX with SKEW nonconformity score was the highest performer.

5.1. Future Work

A direct comparison with coverage and width in mind of (Politis and Thomakos, 2013) and (Hung et al., 2022) is outstanding. Rerunning this study with FACI (Gibbs and Candès, 2022) and SAOCP (Bhatnagar et al., 2023) is also outstanding. Including Analysis of the potential uses of width_t as a factor in certain modeling tasks like in Hung et al. (2022) for SKEW could open interesting research directions. Replicating the study using a structural simulation approach like in (Cont and Vuletić, 2022) could help understand the limitations and scope of the presented approach. Using option prices directly, like in (Taylor et al., 2010), could potentially allow us to define more targeted nonconformity measures. Finally using intraday data would allow us to use more sophisticated base predictors like (Moreno-Pino and Zohren, 2022) and better volatility proxies (Christensen et al., 2010).

Acknowledgments

We extend our thanks to Presidential Professor Jim Gatheral and Professor Andrew Lesniewski, both of the Department of Mathematics at Baruch College, City University of New York, for their helpful insights and suggestions. We also appreciate the valuable comments provided by the anonymous reviewers.

References

Yacine Aït-Sahalia and Robert Kimmel. Maximum likelihood estimation of stochastic volatility models. *Journal of financial economics*, 83(2):413–452, 2007.

Anastasios N Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.

Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. Conformal prediction beyond exchangeability. *arXiv preprint arXiv:2202.13415*, 2022.

Heiner Beckmeyer, Nicole Branger, and Leander Gayda. Retail traders love 0dte options... but should they? *Available at SSRN 4404704*, 2023.

Colin Bennett and Miguel A Gil. Volatility trading. *Trading volatility, Correlation, Term Structure and Skew, Santander Equity Derivatives*, 2012.

Mattia Bevilacqua and Radu Tunaru. The skew index: Extracting what has been left. *Journal of Financial Stability*, 53:100816, 2021.

Aadyot Bhatnagar, Huan Wang, Caiming Xiong, and Yu Bai. Improved online conformal prediction via strongly adaptive online learning. *arXiv preprint arXiv:2302.07869*, 2023.

Roni Bhowmik and Shouyang Wang. Stock market volatility and return analysis: A systematic literature review. *Entropy*, 22(5):522, 2020.

Jonathan Brogaard, Jaehee Han, and Peter Y Won. How does zero-day-to-expiry options trading affect the volatility of underlying assets? *Available at SSRN 4426358*, 2023.

Jie Chen and Dimitris N Politis. Time-varying novas versus garch: point prediction, volatility estimation and prediction intervals. *Journal of Time Series Econometrics*, 12(2), 2020.

Bent J Christensen and Nagpurnanand R Prabhala. The relation between implied and realized volatility. *Journal of financial economics*, 50(2):125–150, 1998.

Kim Christensen, Roel Oomen, and Mark Podolskij. Realised quantile-based estimation of the integrated variance. *Journal of Econometrics*, 159(1):74–98, 2010.

Adam Clements and Daniel PA Preve. A practical guide to harnessing the har volatility model. *Journal of Banking & Finance*, 133:106285, 2021.

Rama Cont and Milena Vuletić. Simulation of arbitrage-free implied volatility surfaces. *Available at SSRN 4299363*, 2022.

Louis H Ederington and Wei Guan. Measuring historical volatility. *Journal of Applied Finance*, 16(1), 2006.

Marcin Fałdziński and Magdalena Osińska. Volatility estimators in econometric analysis of risk transfer on capital markets. *Dynamic Econometric Models*, 16:21–35, 2016.

Matteo Fontana, Gianluca Zeni, and Simone Vantini. Conformal prediction: a unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.

Jim Gatheral. Consistent modeling of spx and vix options. In *Bachelier congress*, volume 37, pages 39–51, 2008.

Jim Gatheral. *The volatility surface: a practitioner's guide*. John Wiley & Sons, 2011.

Jim Gatheral and Roel CA Oomen. Zero-intelligence realized variance estimation. *Finance and Stochastics*, 14(2):249–283, 2010.

Isaac Gibbs and Emmanuel Candes. Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, 34:1660–1672, 2021.

Isaac Gibbs and Emmanuel Candès. Conformal inference for online prediction with arbitrary distribution shifts. *arXiv preprint arXiv:2208.08401*, 2022.

Wei Guo, Sebastian A Gehricke, Xinfeng Ruan, and Jin E Zhang. The implied volatility smirk in spy options. *Applied Economics*, 53(23):2671–2692, 2021.

Patrick S Hagan, Deep Kumar, Andrew S Lesniewski, and Diana E Woodward. Managing smile risk. *The Best of Wilmott*, 1:249–296, 2002.

Jui-Cheng Hung, Hung-Chun Liu, and J Jimmy Yang. Does the tail risk index matter in forecasting downside risk? *International Journal of Finance & Economics*, 2022.

Chao Liang, Yu Wei, and Yaojie Zhang. Is implied volatility more informative for forecasting realized volatility: An international perspective. *Journal of Forecasting*, 39(8):1253–1276, 2020.

Jacob Loveless, Sasha Stoikov, and Rolf Waeber. Online algorithms in high-frequency trading. *Communications of the ACM*, 56(10):50–56, 2013.

Michael McAleer and Marcelo C Medeiros. Realized volatility: A review. *Econometric reviews*, 27(1-3):10–45, 2008.

Fernando Moreno-Pino and Stefan Zohren. Deepvol: Volatility forecasting from high-frequency data with dilated causal convolutions. *arXiv preprint arXiv:2210.04797*, 2022.

Yang-Ho Park. Volatility-of-volatility and tail risk hedging returns. *Journal of Financial Markets*, 26:38–63, 2015.

Dimitris N Politis and Dimitrios D Thomakos. *NoVaS transformations: flexible inference for volatility forecasting.* Springer, 2013.

Ser-Huang Poon and Clive W J Granger. Forecasting volatility in financial markets: A review. *Journal of economic literature*, 41(2):478–539, 2003.

Mark Sebastian and L Celeste Taylor. *Trading Options for Edge: A Professional Guide to Volatility Trading.* Walter de Gruyter GmbH & Co KG, 2022.

Dan Stefanica and Radoš Radoičić. An explicit implied volatility formula. *International Journal of Theoretical and Applied Finance*, 20(07):1750048, 2017.

Stephen J Taylor, Pradeep K Yadav, and Yuanyuan Zhang. The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks. *Journal of Banking & Finance*, 34(4):871–881, 2010.

Dimitrios I Vortelinos. Forecasting realized volatility: Har against principal components combining, neural networks and garch. *Research in international business and finance*, 39:824–839, 2017.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.

Robert E Whaley. Understanding the vix. *The Journal of Portfolio Management*, 35(3): 98–105, 2009.

Appendix A. Base Competitive Prediction Models

A.1. NoVaS Methodology

The NoVaS methodology (Politis and Thomakos, 2013) attempts to map a financial return time series to a sample from a Gaussian stationary series, to achieve this at each step the method finds the memory m that minimizes the excess kurtosis of studentized returns z_t:

$$z_t(m) = \frac{r_t}{F_t(m)}, \quad \text{for} \quad t > m, \tag{8}$$

where $F_t(m) = \sqrt{\sum_{i=0}^{m} a_i r_{t-i}^2}$ and $a_i = 1/(m+1)$. The empirical kurtosis to minimize is defined as:

$$\text{KURT}_t(z) = \frac{n^{-1}\sum_{i=t-n+1}^{t}(z_t - \bar{z})^4}{\left(n^{-1}\sum_{i=t-n+1}^{t}(z_t - \bar{z})^2\right)^2}, \tag{9}$$

each day m^* is chosen as:

$$m_t^* = \text{argmin}_m |\text{KURT}_t(z_t(m)) - 3|, \tag{10}$$

we use $F_t(m_t^*)$ as the volatility forecast for $t+1$. For all experiments we used $n = 252$. For NoVaS-VIX we used the following studentization factor:

$$F_t^{\text{VIX}}(m) = \sqrt{\sum_{i=0}^{m} a_i f_{t-i}(\text{VIX}_{t-i})^2}, \tag{11}$$

where f_t is defined as in Equation 4.

A.2. HAR Methodology

The Heterogeneous Autoregressive Model (HAR) is a competitive and popular model for Realized Volatitlity forecasting (Clements and Preve, 2021) defined as:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1}^{(5)} + \beta_3 RV_{t-1}^{(22)} + \epsilon_t, \tag{12}$$

where RV_t is the realized variance of day t, $\epsilon_t \sim \mathcal{N}(0, \eta)$ and,

$$RV_t^{(j)} = \frac{1}{j}\sum_{n=0}^{j-1} RV_{t-n}, \tag{13}$$

in our experiments we replaced RV_t with h_t^2 and the β parameters were estimated by a rolling one year regression via ordinary least squares (OLS). The HAR-VIX model was specified by adding the previous day's VIX as a factor to the original HAR definition as:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1}^{(5)} + \beta_3 RV_{t-1}^{(22)} + \beta_4 \mathrm{VIX}_{t-1}^2 + \epsilon_t. \tag{14}$$