

Article

Information Transmission Across Markets: Tail Risk Spillovers and Cross-Market Volatility Forecasting

Shaocong Peng  and Yun Shi 

School of Statistics, East China Normal University, Shanghai 200062, China; 52264404007@stu.ecnu.edu.cn

* Correspondence: yshi@fem.ecnu.edu.cn

Abstract

This paper examines tail risk spillovers and cross-market volatility forecasting between the U.S. equity market and the crude oil market. Using realized and implied volatility within a heterogeneous autoregressive (HAR) framework, we document asymmetric and time-varying tail risk transmission across the two markets. Motivated by these findings, we propose several cross-market volatility forecasting strategies, including direct information augmentation, threshold-based designs, forecast averaging, and transfer learning. The results show that incorporating cross-market information improves volatility forecasts primarily at medium and longer horizons, consistent with the forward-looking nature of implied volatility. Moreover, the relative effectiveness of different transmission mechanisms varies across markets, with transfer learning performing particularly well in the crude oil market. Overall, the findings highlight the importance of linking tail risk spillovers to volatility forecasting and demonstrate that flexible cross-market information transmission can enhance predictive performance across markets and horizons.

Keywords: tail risk spillovers; volatility forecasting; cross-market information; HAR models; crude oil; transfer learning

MSC: 91G70; 91B05; 62M10; 91B84; 91B76

1. Introduction

Financial markets are increasingly interconnected, with shocks originating in one market rapidly transmitting to others. Among these linkages, the interaction between equity markets and commodity markets—particularly crude oil—has attracted sustained attention from both academics and policymakers [1–5]. Crude oil plays a dual role as a key production input and a financial asset, implying that oil price fluctuations may affect stock markets through macroeconomic channels, corporate cash flows, and investor risk perceptions [6–8]. Conversely, equity market conditions may convey information about aggregate demand and financial stress that is relevant for oil price dynamics [2,9]. Understanding how risk propagates between these two markets is therefore central to the assessment of financial stability and risk management.

A growing body of work studies stock–oil linkages through tail- and quantile-based methods, highlighting strong nonlinearities, asymmetries, and state dependence. Ref. [10] provided early evidence by applying Granger causality in risk based on VaR exceedance indicators and kernel tests. They document statistically significant and asymmetric extreme-risk spillovers between WTI crude oil futures and the S&P 500, with spillovers strengthening markedly after the global financial crisis and occurring both contemporaneously and with



Academic Editors: Kuang-Hsun Shih,
Shu-Ping Lin and Yi-Hsien Wang

Received: 16 January 2026

Revised: 11 February 2026

Accepted: 12 February 2026

Published: 15 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

delayed effects. Building on quantile-oriented spillover designs, ref. [11] adopted a quantile-on-quantile framework for emerging and frontier markets and showed that oil–stock return exposures are highly heterogeneous across market states, supporting the interpretation of oil price risk as a proxy for global growth concerns and motivating the use of oil market signals in financial risk modeling. Using a multivariate quantile approach, ref. [12] proposed a VAR-for-VaR framework and pseudo impulse responses to study spillovers across quantiles, finding pronounced asymmetries in cross-market risk transmission and stronger spillover effects in the post-2008 period, with similar patterns also present for major international equity indices. More recently, ref. [13] investigated crude oil and BRICS stock markets under extreme shocks by decomposing returns into extreme positives, extreme negatives, and normal components and showed that oil–stock causal links are substantially stronger under extreme states than in normal times, with asymmetric effects across the sign of tail shocks. Extreme market movements, rather than average fluctuations, are often responsible for systemic events and large economic losses. However, most existing studies remain primarily concerned with detecting the presence of tail risk spillovers, rather than assessing their economic relevance for forecasting and decision-making.

The heterogeneous autoregressive (HAR) model has become a benchmark for volatility forecasting because it parsimoniously captures persistent realized-volatility dynamics by aggregating components across multiple horizons. Ref. [14] provides the foundational motivation by proposing the HAR-RV model as an additive volatility cascade across time scales, showing that this simple AR-type structure can reproduce key empirical regularities while delivering strong forecasting performance. Subsequent research refines how the HAR model is implemented in practice. Ref. [15] showed that predictive accuracy can depend materially on practitioner choices such as estimators, transformations of realized volatility, and forecast combination schemes, with simple adjustments often improving out-of-sample performance relative to the standard HAR benchmark. Beyond point forecasts, related work studies volatility and returns jointly to improve predictive distributions. Ref. [16] developed a bivariate return–realized-volatility framework and documented that incorporating intraday realized-volatility information and multi-component volatility dynamics improves multi-horizon density forecasts relative to models based only on daily returns. A further extension augments HAR with forward-looking information from option markets. Ref. [17] showed that implied volatility delivers sizeable gains over HAR in forecasting realized volatility and that additional predictors (e.g., leverage, overnight returns, and higher-order RV measures) can further enhance accuracy with economic significance. Finally, ref. [18] demonstrated that implied volatility contains incremental information for forecasting both continuous and jump components of realized volatility across asset classes and proposed a vector HAR system to address simultaneity concerns. Despite these advances, HAR and HAR-IV models remain largely single-market in nature, relying primarily on own-market realized and implied measures and typically abstracting from cross-market information transmission.

The disconnect between strong evidence of tail risk spillovers and the reliance on market-specific volatility models is notable. While tail risk spillovers document that extreme risks are transmitted across markets, it remains unclear whether—and how—this information can be systematically incorporated into volatility forecasting frameworks. In particular, spillover effects may be episodic rather than constant, and their relevance for forecasting may depend on market conditions, forecast horizons, and the nature of the transmission mechanism. Recent studies begin to move in this direction by explicitly embedding cross-market signals into volatility prediction models: ref. [19] showed that cross-market information flow contains both linear and nonlinear predictive content for stock market volatility and that its contribution is horizon-dependent, with stronger gains

at medium and longer horizons than at very short horizons; ref. [20] documented that speculative sentiment carries incremental forecasting power for gold-market volatility, with cross-market sentiment from equities becoming more informative at longer horizons and delivering economically meaningful gains in portfolio performance; and ref. [21] provided evidence that equity market jump measures can improve forecasts of crude oil implied volatility (OVX), especially when combined with hybrid data-driven and parametric modeling strategies. Taken together, these contributions suggest that cross-market predictors can enhance volatility forecasts, but the evidence also points to heterogeneous mechanisms and horizon-specific effects, reinforcing the need for a unified framework that links tail risk spillovers to volatility forecasting and systematically compares alternative transmission designs.

This paper aims to fill this gap by linking tail risk spillovers to cross-market volatility forecasting in a unified framework. Focusing on the U.S. equity market and the crude oil market, we first document asymmetric and time-varying tail risk spillovers using tail-based Granger causality tests and connectedness measures. These results provide direct empirical motivation for incorporating cross-market information into volatility forecasting models. Building on this evidence, we propose several cross-market extensions of the HAR framework, including direct information augmentation with realized and implied volatility from the external market, threshold-dependent designs that activate cross-market information only during extreme states, forecast averaging to address model uncertainty, and a transfer learning approach that regularizes volatility dynamics across markets.

Our empirical results show that incorporating cross-market information leads to economically meaningful improvements in volatility forecasting, particularly at medium and longer horizons. The effectiveness of different transmission mechanisms varies across markets: while forecast averaging performs robustly in the equity market, transfer learning proves especially effective in the crude oil market, where volatility dynamics are more volatile and noisy. Taken together, the findings highlight the importance of integrating tail risk spillovers into volatility forecasting models and demonstrate that flexible cross-market information transmission can enhance predictive performance across markets and horizons.

The remainder of this paper is organized as follows. Section 2 investigates tail risk spillovers between the U.S. equity market and the crude oil market. Using tail-based Granger causality tests and a VAR-based connectedness framework, this section documents the direction, asymmetry, and time variation of extreme risk transmission across markets and provides direct empirical motivation for incorporating cross-market information into volatility forecasting models. Section 3 introduces the volatility forecasting framework. Building on the heterogeneous autoregressive (HAR) paradigm, this section first presents benchmark market-specific models and then develops several cross-market extensions, including direct information augmentation, threshold-dependent designs, forecast averaging, and a transfer learning approach that allows volatility dynamics in one market to borrow information from another. Section 4 evaluates the out-of-sample forecasting performance of the competing models across markets, forecast horizons, and rolling-window specifications. This section assesses both statistical and economic forecasting gains and examines how the effectiveness of different cross-market transmission mechanisms varies across markets and time horizons.

2. Tail Risk Spillovers

This section investigates tail risk spillovers between the U.S. equity market and the crude oil market. Building on the tail risk measures constructed in the previous section, we examine whether extreme risks in one market contain predictive information for tail events in the other and how such transmission evolves over time.

We proceed in three steps: First, we employ tail Granger causality tests to assess directional predictability between downside and upside tail risks across markets in Section 2.1. Second, in Section 2.2, we complement the causality analysis with a connectedness framework based on tail risk indicators to capture the time-varying intensity and direction of spillovers. Finally, we present empirical evidence using rolling-window estimates to document the dynamics of tail risk transmission and to motivate the cross-market volatility forecasting strategies developed Section 2.3.

2.1. Tail Granger Causality in Risk

This subsection follows the framework of [10,22] to examine extreme risk spillovers across markets by testing Granger causality in the tail of the return distribution.

Let Y_{lt} denote the daily return of market $l \in \{1, 2\}$. Returns are assumed to follow a general location–scale representation,

$$Y_{lt} = \mu_{lt} + \sigma_{lt}\varepsilon_{lt}, \quad (1)$$

where μ_{lt} is the conditional mean; σ_{lt}^2 is the conditional variance; and the standardized innovation ε_{lt} follows a martingale difference sequence with unit variance, denoted by $\varepsilon_{lt} \sim \text{m.d.s}(0, 1)$.

The conditional variance is modeled using the heterogeneous autoregressive (HAR) framework. In the standard HAR model, realized volatility evolves as

$$RV_{lt} = \beta_{l,0} + \beta_{l,d}RV_{l,t-1}^{(d)} + \beta_{l,w}RV_{l,t-1}^{(w)} + \beta_{l,m}RV_{l,t-1}^{(m)} + u_{lt}, \quad (2)$$

where $RV^{(d)}$, $RV^{(w)}$, and $RV^{(m)}$ denote daily, weekly, and monthly realized volatility components, respectively.

To incorporate forward-looking information, we extend the benchmark HAR model to a HAR^{IV} specification by including implied volatility at corresponding horizons:

$$RV_{lt}^{IV} = \beta_{l,0} + \sum_{h \in \{d,w,m\}} \beta_{l,h}RV_{l,t-1}^{(h)} + \sum_{h \in \{d,w,m\}} \gamma_{l,h}IV_{l,t-1}^{(h)} + u_{lt}. \quad (3)$$

To clarify the construction of the implied-volatility regressors in Equations (3), we note that the implied volatility indices used in the empirical analysis (*VIX* for equities and *OVX* for crude oil) are quoted as approximately 30-calendar-day option-implied volatility measures. Following the standard practice in the HAR-IV literature [17], we extract heterogeneous implied-volatility components by temporal aggregation of the daily index levels. Specifically, let $IV_{l,t}$ denote the daily closing level of the implied volatility index for market $l \in \{1, 2\}$. The daily, weekly, and monthly implied-volatility components are constructed as

$$IV_{l,t}^{(d)} = IV_{l,t}, \quad IV_{l,t}^{(w)} = \frac{1}{5} \sum_{i=0}^4 IV_{l,t-i}, \quad IV_{l,t}^{(m)} = \frac{1}{22} \sum_{i=0}^{21} IV_{l,t-i}.$$

All implied-volatility components enter the forecasting regressions with a one-period lag, ensuring that the information set at time t contains only data observable up to that date. This construction mirrors the aggregation of realized-volatility components in the HAR framework and is adopted purely for consistency in temporal structure.

The conditional standard deviation is then given by $\sigma_{lt} = \sqrt{RV_{lt}^{IV}}$. It is worth emphasizing that the HAR^{IV} model is used exclusively to proxy conditional volatility and does not model return dynamics, consistent with the framework in [16]. The conditional mean is specified as the one-period lagged return, $\mu_{lt} = Y_{l,t-1}$, which provides a parsimonious benchmark and avoids imposing additional structure on return dynamics. Our results are

robust to alternative specifications of the conditional mean, including rolling historical averages and $AR(1)$ prediction.

Given the location–scale representation in Equation (1), the Value at Risk (VaR) at confidence level $\alpha \in (0, 1)$ is defined on both sides of the conditional return distribution as

$$\begin{aligned} VaR_{lt}^D &= -\mu_{lt} + \sigma_{lt} z_{lt}(\alpha), \\ VaR_{lt}^U &= \mu_{lt} + \sigma_{lt} z_{lt}(1 - \alpha), \end{aligned} \quad (4)$$

where $z_{lt}(\cdot)$ denotes the corresponding quantile of the standardized innovation ε_{lt} . The superscripts D and U denote downside and upside tail events, respectively. Distinguishing between D and U allows us to examine potential asymmetries in tail risk transmission across markets. In particular, extreme losses and extreme gains may be driven by different economic mechanisms and may propagate differently across markets. As a result, separating downside and upside tail events makes it possible to identify not only same-tail spillovers (DD and UU) but also cross-tail spillovers (DU and UD), where extreme movements of one sign in one market are associated with extreme movements of the opposite sign in another market.

Rather than analyzing VaR directly, the tail Granger causality framework focuses on the occurrence of extreme tail events relative to these thresholds. Accordingly, we define the downside and upside tail risk indicators as

$$Z_{lt}^D = \mathbf{1}(Y_{lt} < -VaR_{lt}^D), \quad Z_{lt}^U = \mathbf{1}(Y_{lt} > VaR_{lt}^U), \quad l = 1, 2.$$

Tail Granger causality from market 2 to market 1 is defined as the incremental predictive content of past tail risk events in market 2 for future tail risk events in market 1. To formalize this notion, let $\mathcal{I}_{t-1}$ denote the global information set available at time $t - 1$, defined as the union of the market-specific information sets,

$$\mathcal{I}_{t-1} = \mathcal{I}_{1,t-1} \cup \mathcal{I}_{2,t-1}.$$

For each market $l \in \{1, 2\}$, the information set $\mathcal{I}_{l,t-1}$ consists of backward-looking information derived from historical price dynamics, including past returns and realized volatility at daily, weekly, and monthly horizons, as well as forward-looking information extracted from option markets and summarized by implied volatility measures at corresponding horizons.

The null hypothesis of no tail risk Granger causality from market 2 to market 1 is then given by

$$H_0: E(Z_{1t}^k | \mathcal{I}_{1,t-1}) = E(Z_{1t}^k | \mathcal{I}_{t-1}), \quad k \in \{D, U\}. \quad (5)$$

which states that, for each tail type $k \in \{D, U\}$, conditioning on the information from market 2 does not improve the prediction of the corresponding downside or upside tail risk events in market 1. In practice, the distinction between DD, DU, UU, and UD cases is implemented through the composition of the information set $\mathcal{I}_{t-1}$, which includes the relevant downside or upside tail indicator from market 2 depending on the configuration under consideration, thereby giving rise to four separate hypothesis tests.

Following [22], the null hypothesis is tested using kernel-weighted cross-correlations of the risk indicators. Let

$$\hat{C}^{k_1, k_2}(j) \equiv \begin{cases} T^{-1} \sum_{t=1+j}^T (\hat{Z}_{1t}^{k_1} - \hat{\alpha}_1) (\hat{Z}_{2,t-j}^{k_2} - \hat{\alpha}_2), & 0 \leq j \leq T-1, \\ T^{-1} \sum_{t=1-j}^T (\hat{Z}_{1,t-j}^{k_1} - \hat{\alpha}_1) (\hat{Z}_{2t}^{k_2} - \hat{\alpha}_2), & 1-T \leq j < 0, \end{cases} \quad (6)$$

where $\hat{\alpha}_l = T^{-1} \sum_{t=1}^T \hat{Z}_{lt}$, $l = 1, 2$ denotes the sample mean of the tail event indicators, and $k_1, k_2 \in \{D, U\}$ denotes the type of tail risk indicators. For notational simplicity, the superscript $k \in \{D, U\}$ is suppressed in the subsequent expressions for the tail event indicators and their correlation $\hat{C}^{k_1 k_2}(j)$. Throughout the remainder of this subsection, Z_{lt} should be understood as a generic tail indicator, which can represent any of the four configurations (Z_{1t}^D, Z_{2t}^D) , (Z_{1t}^D, Z_{2t}^U) , (Z_{1t}^U, Z_{2t}^U) , or (Z_{1t}^U, Z_{2t}^D) , corresponding to DD, DU, UU, and UD cases, respectively.

The corresponding sample cross-correlation at lag j is defined as

$$\hat{\rho}(j) = \frac{\hat{C}(j)}{\sqrt{\hat{\alpha}_1 \hat{\alpha}_2 (1 - \hat{\alpha}_1)(1 - \hat{\alpha}_2)}}.$$

The one-sided test statistic aggregates the squared cross-correlations across positive lags using kernel weights,

$$Q_1(M) = \frac{T \sum_{j=1}^{T-1} k^2(j/M) \hat{\rho}^2(j) - C_{1T}(M)}{\sqrt{D_{1T}(M)}}, \quad (7)$$

where $k(\cdot)$ is a kernel function and M is a bandwidth parameter, with centering and scaling terms

$$C_{1T}(M) = \sum_{j=1}^M (1 - j/T) k^2\left(\frac{j}{M}\right),$$

$$D_{1T}(M) = 2 \sum_{j=1}^M (1 - j/T)(1 - (j+1)/T) k^4\left(\frac{j}{M}\right).$$

The kernel weighting scheme assigns greater weight to lower-order lags while progressively downweighting higher-order lags, thereby capturing economically relevant short- to medium-term tail dependence and reducing the influence of sampling noise, where $k(\cdot)$ is a kernel function and M is a bandwidth parameter. In the empirical implementation, we employ the Daniell kernel, which assigns equal weights within a given bandwidth and zero weight outside. The finite-sample correction terms $C_{1T}(M)$ and $D_{1T}(M)$ ensure proper centering and scaling of the statistic. Under standard regularity conditions, $Q_1(M)$ converges in distribution to a standard normal distribution under the null hypothesis and large positive values reject H_0 in favor of Granger causality in risk. Since the downside and upside VaR thresholds generally differ across the two markets, the tail Granger causality analysis yields four distinct test statistics and corresponding hypothesis test results, associated with the DD, DU, UU, and UD configurations. Technical details regarding the asymptotic theory and implementation are provided in [22].

2.2. Connectedness Analysis Based on Tail Risk Measures

Building on the tail risk measures constructed in the previous section, this subsection employs a VAR-based connectedness framework, rather than Granger causality tests, to quantify the dynamic transmission of extreme risk between two markets, following [23,24].

Let $\mathbf{Y}_t = (VaR_{1t}, VaR_{2t})'$ denote a two-dimensional vector of tail risk measures, where Y_{lt} represents the estimated tail risk (e.g., VaR-based measures) for market l at time t . We assume that $\mathbf{Y}_t$ follows a VAR(p) process,

$$\mathbf{Y}_t = \sum_{k=1}^p \mathbf{A}_k \mathbf{Y}_{t-k} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim (0, \boldsymbol{\Sigma}), \quad (8)$$

where ε_t is a vector of innovations with covariance matrix Σ .

Under standard regularity conditions, the vector autoregressive models admit the vector moving-average (VMA) representation

$$\mathbf{Y}_t = \sum_{h=0}^{\infty} \mathbf{\Psi}_h \varepsilon_{t-h}, \quad (9)$$

where $\mathbf{\Psi}_0$ is a 2×2 identity matrix and $\mathbf{\Psi}_h$ captures the dynamic propagation of shocks between the two markets.

We employ the generalized forecast error variance decomposition (GFEVD), which is invariant to variable ordering. The contribution of shocks in market j to the H_C -step-ahead forecast error variance of market i is given by

$$\theta_{ij}(H_C) = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H_C-1} (\mathbf{e}_i' \mathbf{\Psi}_h \Sigma \mathbf{e}_j)^2}{\sum_{h=0}^{H_C-1} \mathbf{e}_i' \mathbf{\Psi}_h \Sigma \mathbf{\Psi}_h' \mathbf{e}_i}, \quad i, j \in \{1, 2\}, \quad (10)$$

where $\mathbf{e}_i$ denotes a selection vector with unity in the i th position.

Since $\theta_{i1}(H_C) + \theta_{i2}(H_C) \neq 1$ in general, the variance decompositions are normalized as

$$\tilde{\theta}_{ij}(H_C) = \frac{\theta_{ij}(H_C)}{\theta_{i1}(H_C) + \theta_{i2}(H_C)}, \quad i, j \in \{1, 2\}. \quad (11)$$

Directional connectedness is defined as the spillover of tail risk between the two markets. The connectedness *from* market j to market i is given by

$$C_{i \leftarrow j}(H_C) = \tilde{\theta}_{ij}(H_C) \times 100, \quad i \neq j, \quad (12)$$

while the connectedness *to* market j from market i is defined as

$$C_{i \rightarrow j}(H_C) = \tilde{\theta}_{ji}(H_C) \times 100, \quad i \neq j. \quad (13)$$

2.3. Empirical Analysis of Tail Risk Spillovers

2.3.1. Dataset

Our empirical analysis of both the Granger causality tests and the connectedness measures is based on the interaction between the U.S. equity market and the crude oil market.

The equity market is represented by the S&P 500 index. Daily price data for the S&P 500 are obtained from the Center for Research in Security Prices (CRSP). Realized volatility (RV) for the equity market is obtained from the Risk Lab at the University of Chicago Booth School of Business, maintained by Professor Dacheng Xiu. These RV measures have been widely used in the volatility forecasting literature [25,26]. The underlying estimation is based on quasi-maximum likelihood estimation (QMLE) of volatility from moving average models $MA(q)$, constructed using nonzero transaction price returns sampled at the highest available frequency for days with at least 12 observations (see [27,28] for methodological details). Implied volatility for the equity market is measured by the CBOE *VIX* index (dividing the index values by 100, same for crude oil *IV*), which reflects option-implied expectations of volatility for the S&P 500.

The crude oil market is proxied by WTI light sweet crude oil futures. Daily futures price data are obtained from the CBOE. WTI crude oil prices are constructed from NYMEX WTI futures contracts using a liquidity-based selection rule. Specifically, for each trading day, we select the futures contract with the highest trading volume and use its settlement price as the daily oil price observation. This approach allows the price series to reflect the most actively traded contract at each point in time and avoids imposing an *ex ante*

rolling schedule. Daily crude oil returns are computed from this volume-based series, rather than from a pre-constructed continuous futures contract. As a result, the series implicitly incorporates contract transitions when trading activity shifts across maturities, while always relying on the most liquid available contract. No additional price adjustments are applied at contract switches.

To address the potential issue of roll-induced discontinuities in WTI futures data, we conduct a robustness check following the approach of [29]. Specifically, we construct a time series with constant time-to-maturity (e.g., 1-month maturities) by reorganizing the daily prices of WTI futures contracts such that each column corresponds to a fixed time-to-maturity instead of the delivery month. This approach avoids jumps in prices due to contract rolls. A robustness check using constant-maturity series yields qualitatively similar results (not reported), confirming that our findings are robust to roll effects. All results that follow are based on the volume-based series. To ensure consistency across markets, the oil and equity datasets are aligned by common trading days. Observations corresponding to non-overlapping trading days, such as U.S. equity holidays or oil-market-specific non-trading days, are removed so that all empirical analyses are conducted on a synchronized sample.

Realized volatility for the crude oil market is also obtained from the Risk Lab, ensuring consistency in RV construction across markets. Implied volatility for the crude oil market is measured by the CBOE Crude Oil Volatility Index (OVX). Although OVX is constructed from options written on the United States Oil Fund (USO) rather than directly on crude oil futures, USO closely tracks WTI crude oil futures through a rolling futures strategy, making OVX an accepted proxy for forward-looking uncertainty in the crude oil market.

The sample period spans from 10 May 2007 to 26 May 2023. The starting date is determined by the availability of the OVX index, while the ending date is constrained by the availability of realized volatility measures for crude oil futures. In the empirical analysis, all regressions are conducted using the logarithm of realized volatility.

Table 1 presents the descriptive statistics of the main variables employed in the empirical analysis. The sample covers the period from 10 May 2007 to 26 May 2023, capturing several episodes of heightened volatility in both equity and crude oil markets. Both implied and realized volatility measures exhibit substantial dispersion over the sample period. The standard deviation of OVX (17.44) is notably higher than that of VIX (9.21), indicating greater variability and more pronounced uncertainty in the crude oil market. Similarly, realized volatility in the crude oil market (RV^{CLF}) shows higher variability than its equity market counterpart (RV^{SPX}), consistent with the more volatile nature of energy markets.

Table 2 reports pairwise correlations among the main variables. Implied and realized volatility measures within each market are strongly correlated, while cross-market volatility correlations are also substantial, suggesting the presence of potential volatility spillovers between equity and crude oil markets. While the tail Granger causality tests focus on identifying directional predictability between extreme risk events across markets, the connectedness analysis complements this perspective by quantifying the time-varying intensity and persistence of tail risk spillovers. Together, these two approaches offer a comprehensive view of how tail risks propagate across markets and evolve over time. Importantly, establishing empirical evidence of tail risk spillovers is a prerequisite for incorporating cross-market information into volatility forecasting models. The results documented in this section therefore provide direct motivation for the cross-market HAR-type volatility models examined in the subsequent sections.

Table 1. Descriptive statistics of equity and crude oil market variables.

	<i>VIX</i>	<i>OVX</i>	<i>RV^{SPX}</i>	<i>RV^{CLF}</i>	<i>SPX</i>	<i>CLF</i>
Mean	20.62	39.04	12.67	34.64	2211.20	73.74
Median	18.34	35.77	10.13	29.95	1987.00	73.36
Std. Dev.	9.21	17.44	9.26	20.76	1050.38	23.25
Min	9.15	14.50	2.38	8.14	676.50	10.01
1st Quartile	14.46	29.55	7.15	23.49	1341.50	53.62
3rd Quartile	23.85	44.89	15.13	39.37	2858.70	93.03
Max	82.69	236.80	132.65	300.00	4793.50	145.29

This table reports descriptive statistics for the variables used in the empirical analysis, including the mean, median, standard deviation, minimum, first quartile, third quartile, and maximum. *VIX* denotes the CBOE implied volatility index for the S&P 500, while *OVX* represents the CBOE crude oil volatility index. *RV^{SPX}* and *RV^{CLF}* denote realized volatility for the equity market and the crude oil market, respectively. *SPX* denotes the level of the S&P 500 index, and *CLF* denotes the front-month WTI crude oil futures price. The sample spans from 10 May 2007 to 26 May 2023.

Table 2. Pairwise correlations among equity and crude oil market variables.

	<i>VIX</i>	<i>OVX</i>	<i>RV^{SPX}</i>	<i>RV^{CLF}</i>	<i>SPX</i>	<i>CLF</i>
<i>VIX</i>	1.00					
<i>OVX</i>	0.72	1.00				
<i>RV^{SPX}</i>	0.88	0.59	1.00			
<i>RV^{CLF}</i>	0.65	0.92	0.58	1.00		
<i>SPX</i>	−0.17	0.05	−0.16	0.06	1.00	
<i>CLF</i>	−0.12	−0.40	−0.06	−0.40	−0.21	1.00

This table reports pairwise Pearson correlation coefficients among the main variables used in the empirical analysis. *VIX* and *OVX* denote implied volatility indices for the equity and crude oil markets, respectively. *RV^{SPX}* and *RV^{CLF}* denote realized volatility measures. *SPX* denotes the S&P 500 index level, and *CLF* denotes the front-month WTI crude oil futures price.

2.3.2. Empirical Evidence on Tail Risk Spillovers

This subsection provides empirical evidence on tail risk spillovers between the U.S. equity market and the crude oil market. Building on the tail Granger causality in risk framework and the tail-based connectedness measures introduced in the previous section, we assess both the existence and the dynamic nature of cross-market tail risk transmission.

Figure 1 presents rolling-window Value-at-Risk (VaR) estimates for the U.S. equity market and the crude oil market at the 95% and 99% confidence levels. As described in Equation (4), the tail risk measures are constructed from a location–scale representation in which the conditional mean is specified as the one-period lagged return, the conditional variance is proxied by volatility forecasts from the HAR^{IV} model, and the standardized innovation is assumed to follow a standard normal distribution. The rolling estimation framework captures the time-varying nature of tail risk in both markets, allowing the conditional risk bounds to adjust dynamically to changing market conditions.

Several notable differences emerge across markets. While both markets experience episodes of heightened volatility, the crude oil market exhibits more frequent and more severe VaR violations, particularly at the 99% confidence level. This pattern reflects the oil market's greater exposure to supply-side shocks, geopolitical tensions, and abrupt demand disruptions [30–32]. VaR violations cluster during periods of systemic stress, most prominently around the global financial crisis and the COVID-19 pandemic. These episodes are characterized by sharp increases in both the magnitude and frequency of exceedances, suggesting heightened tail dependence and creating conditions under which cross-market tail risk spillovers are more likely to arise.

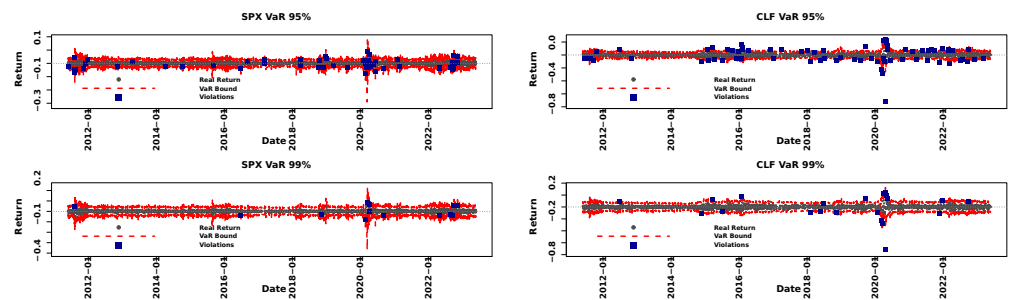


Figure 1. Rolling-window downside and upside Value-at-Risk estimates for equity and crude oil markets: this figure plots rolling-window Value-at-Risk (VaR) estimates for the S&P 500 (SPX) and WTI crude oil futures (CLF) at the 95% and 99% confidence levels. Black dots denote realized daily returns; red dashed lines represent the estimated VaR bounds; and blue squares indicate VaR violations, defined as realized returns exceeding the corresponding VaR thresholds. The VaR estimates are constructed using the normal quantile, with the tail Granger causality statistic is computed based on the Daniell kernel and bandwidth parameter $M = 20$.

Table 3 reports the tail Granger causality test statistics from Equation (7), which assess directional predictability between extreme risk events in the crude oil market and the U.S. equity market. The results are presented for both downside and upside tail configurations and are computed under alternative bandwidth parameters M and kernel functions used to estimate the long-run covariance matrix. To evaluate robustness, we report results across commonly used kernel specifications (Daniell, Bartlett, and Parzen) and a range of bandwidth choices (see [10,22] for detailed discussions of kernel and bandwidth selection). Following [22] and common practice, we set baseline $M = 20$ to capture short-to-medium lag dependence.

Overall, the results provide strong evidence of statistically significant tail risk spillovers between the two markets. Importantly, the strength and significance of these spillovers vary across tail states, directions, and quantile levels, indicating a pronounced state-dependent asymmetry rather than a single dominant transmission pattern. While the magnitude of the test statistics changes with the choice of M and kernel, the qualitative presence of tail risk spillovers remains robust across specifications.

Focusing first on spillovers from the crude oil market to the equity market, statistically significant Granger causality is most frequently observed in downside-related configurations. At the 1% tail level, significant spillovers emerge in cross-tail states, indicating that extreme oil price movements—whether upward or downward—contain predictive information for subsequent tail risk in equity markets. At the 5% quantile, spillovers become more concentrated in downside-to-downside configurations, suggesting that moderately adverse oil market conditions also transmit risk to equities, albeit with reduced intensity.

In contrast, spillovers from the equity market to the crude oil market display a broader pattern at the most extreme tail level. At the 1% quantile, significant Granger causality is detected across multiple tail configurations, including downside-to-downside, downside-to-upside, and upside-to-upside states, pointing to a pervasive influence of extreme equity market conditions on oil market tail risk. At the 5% quantile, however, significant spillovers are largely confined to downside-related states, while upside and cross-tail effects weaken substantially. These directional features are stable across alternative kernel functions and bandwidth choices.

Table 3. Granger causality spillover test results.

M	Kernel	α	Crude Oil $\rightarrow$ S&P 500				S&P 500 $\rightarrow$ Crude Oil			
			DD	DU	UU	UD	DD	DU	UU	UD
20	Daniell	1%	0.57	5.66 ***	11.78 ***	27.63 ***	10.22 ***	3.75 ***	11.09 ***	2.23 **
		5%	3.19 ***	3.78 ***	−1.04	13.33 ***	9.43 ***	18.16 ***	−0.10	0.39
30	Daniell	1%	11.04 ***	4.04 ***	8.23 ***	31.85 ***	13.14 ***	8.06 ***	16.94 ***	8.81 ***
		5%	3.70 ***	4.06 ***	−1.28	12.37 ***	11.76 ***	18.35 ***	1.57 *	1.20
60	Daniell	1%	18.42 ***	1.80 **	7.45 ***	32.00 ***	12.02 ***	9.79 ***	31.17 ***	18.57 ***
		5%	5.40 ***	3.11 ***	−0.31	10.41 ***	11.47 ***	16.78 ***	5.49 ***	4.54 ***
20	Bartlett	1%	0.47	6.03 ***	12.94 ***	25.60 ***	8.17 ***	2.95 ***	8.14 ***	1.24
		5%	2.72 ***	4.57 ***	−1.00	12.46 ***	8.70 ***	17.64 ***	−0.56	0.05
30	Bartlett	1%	8.32 ***	4.65 ***	10.43 ***	29.93 ***	11.29 ***	6.78 ***	14.73 ***	6.60 ***
		5%	3.17 ***	4.49 ***	−1.16	12.25 ***	10.87 ***	18.20 ***	1.01	1.01
60	Bartlett	1%	15.88 ***	1.81 **	8.90 ***	32.53 ***	12.24 ***	9.36 ***	28.11 ***	16.11 ***
		5%	5.04 ***	3.58 ***	−0.43	11.33 ***	11.61 ***	17.83 ***	4.61 ***	3.83 ***
20	Parzen	1%	4.45 ***	5.12 ***	10.52 ***	29.21 ***	11.43 ***	5.58 ***	12.37 ***	4.09 ***
		5%	3.04 ***	4.05 ***	−1.18	12.93 ***	10.38 ***	18.28 ***	0.23	0.50
30	Parzen	1%	13.09 ***	3.29 ***	8.50 ***	31.89 ***	13.18 ***	9.23 ***	20.74 ***	10.96 ***
		5%	3.53 ***	4.00 ***	−1.08	11.78 ***	11.95 ***	17.98 ***	2.45 ***	2.10 **
60	Parzen	1%	17.96 ***	0.24	7.31 ***	32.05 ***	11.80 ***	9.46 ***	31.94 ***	19.61 ***
		5%	6.17 ***	2.83 ***	−0.17	10.55 ***	11.19 ***	16.77 ***	6.12 ***	4.75 ***

This table reports Granger causality spillover test results based on tail risk measures for different bandwidth parameters (M) and kernel functions. Reported values are Q statistics, with statistically significant statistics shown in bold. Superscripts ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. DD and UU denote downside-to-downside and upside-to-upside spillovers, while DU (UD) represents downside-to-upside (upside-to-downside) spillovers. Statistical significance is assessed using two-sided standard normal critical values.

Taken together, the results highlight clear directional and state-dependent asymmetries in tail risk transmission between the two markets. These patterns can be interpreted through the economic transmission channels articulated in the framework of [10]. Specifically, the significant *DD* spillover from the equity market to the crude oil market is consistent with the “aggregate demand” channel. As noted by [10], an extreme crash in the stock market often signals deteriorating expectations for future business activity and a contraction in the global economic system. This negative outlook triggers a sharp reduction in the derived demand for energy products, thereby precipitating a collapse in oil prices. Conversely, the strong spillover from crude oil upside risk to stock market downside risk (UD) points to the prevalence of “supply-side” shocks or increases in precautionary demand driven by supply shortfall concerns. Structurally, such extreme oil price surges transmit risk by elevating production costs and inducing inflationary pressures, which squeeze consumer discretionary spending and corporate profit margins. Consequently, investors revise their valuations downward based on expected changes in real cash flows and dividends. Thus, the observed asymmetry is not merely statistical but reflects the distinct structural shocks—demand contraction versus cost-push inflation—that drive tail risk transmission in each direction.

Beyond macroeconomic shocks, the financialization of commodities [33,34] also provides a crucial structural explanation for the observed asymmetric spillovers. As crude oil has evolved into a key asset class for institutional portfolio allocation, its price dynamics are increasingly driven by financial factors rather than purely physical supply and demand. During extreme equity market downturns, systemic risk propagates to the oil market through cross-market rebalancing and deleveraging, explaining why equity market spillovers tend to be more pervasive. Furthermore, *liquidity constraints* during deep tail events can amplify these transmissions. According to the “liquidity spiral” framework [35], investors facing margin calls or funding shortages in the equity market may be forced

to liquidate their liquid commodity positions [36]. This liquidity-driven contagion creates a potent downside-to-downside transmission channel that reinforces the asymmetric influence of the equity market on oil volatility.

To assess the sensitivity of the results to the construction of tail risk measures, Appendix A Table A1 reports robustness results based on empirical quantiles computed exclusively from rolling-window in-sample observations. The resulting spillover patterns are broadly consistent with those obtained under the normal-quantile specification, both in terms of the direction of tail risk transmission and the relative strength of significant configurations. Unless otherwise stated, subsequent VaR-based analyses rely on the baseline specification using normal quantiles, the Daniell kernel, and a bandwidth parameter of $M = 20$.

Although the magnitude and significance of spillovers differ between the 1% and 5% quantile levels, both sets of results consistently point to economically and statistically meaningful tail risk interactions between the equity and crude oil markets. The variation across quantiles reflects differences in the severity of extreme events rather than the absence of interdependence. These findings establish a clear empirical foundation for the subsequent analyses. We next examine the time-varying evolution of tail risk transmission using connectedness measures and then assess whether incorporating cross-market tail information improves volatility forecasting performance within a HAR-type modeling framework.

Before turning to the dynamic connectedness analysis, it is useful to briefly clarify the role of the U.S. dollar in the interaction between the equity and crude oil markets. As both markets are denominated in U.S. dollars, exchange rate movements constitute an important channel through which global financial conditions can simultaneously affect risk dynamics in both markets. More generally, similar to other macro-financial variables, fluctuations in the U.S. dollar can act as a common transmission mechanism that links commodity and equity markets by shaping global liquidity conditions, risk appetite, and valuation effects. In this study, the U.S. dollar is therefore treated as part of the broader macro-financial environment within which tail risk transmission is examined.

While the tail Granger causality tests provide evidence on directional predictability between extreme risk events across markets, they do not describe how the strength of tail risk transmission evolves over time. To complement this perspective, we next examine rolling-window connectedness measures constructed from the tail risk indicators. Unlike the Granger causality results, which focus on statistical significance, the connectedness measures are primarily used here as descriptive tools to illustrate the time variation and heterogeneity in tail risk spillover intensity across different tail-state configurations.

Figure 2 plots the rolling-window directional connectedness measures $C_{i \rightarrow j}(H_C)$ defined in Equation (13) under four tail-state configurations. Across all panels, the connectedness measures exhibit substantial time variation, indicating that the intensity of tail risk spillovers is not constant but fluctuates over time. Moreover, spillover strength differs systematically across tail-state configurations, suggesting that tail risk transmission depends on the prevailing joint tail states of the two markets.

This pronounced time variation is closely aligned with the regime-dependent nature of oil–equity spillovers documented in [37]. In their framework, spillover intensity and apparent asymmetry tend to strengthen during periods of financial stress, heightened uncertainty, and binding liquidity constraints, while remaining comparatively muted in normal times. Changes in investors’ interpretation of oil price movements—shifting from supply-side considerations to aggregate-demand and macroeconomic signals—play a central role in shaping these regime shifts.

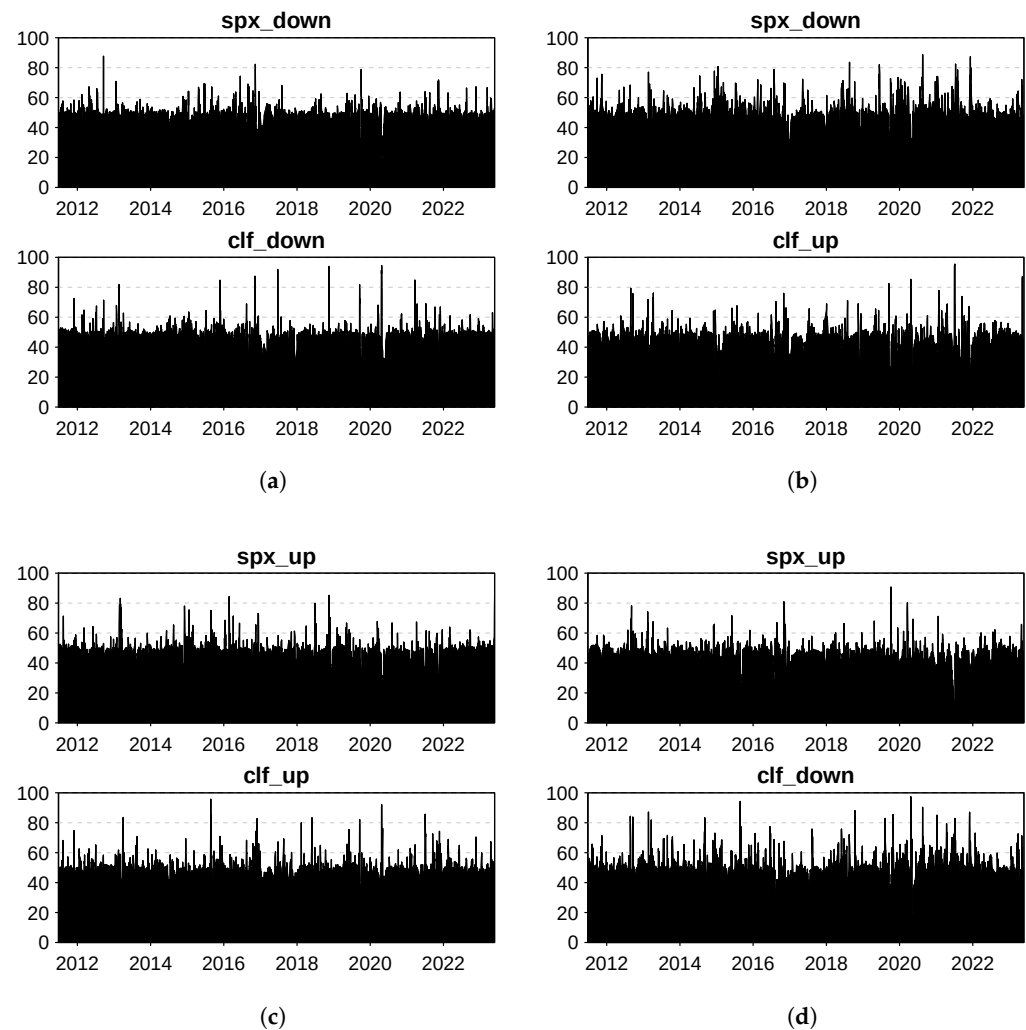


Figure 2. Time-varying directional connectedness (*TO*) based on tail risk indicators: This figure plots rolling-window VAR-based directional connectedness measures computed from the tail risk *Var*. Each panel corresponds to one tail-state configuration: (a) downside and downside; (b) downside and upside; (c) upside and upside; and (d) upside and downside. The plotted series is the *TO* spillover measure $C_{i \rightarrow j}(H_C)$ for the chosen direction. In a two-market system, the *TO* spillover from market 1 to market 2 is numerically equivalent to the *FROM* spillover received by market 2 from market 1, so reporting *TO* is sufficient to characterize directional transmission.

From this perspective, the episodic behavior observed in Figure 2 suggests that cross-market tail risk spillovers are not a permanent feature of the oil–equity relationship but instead emerge more forcefully during specific high-risk periods. This observation has direct implications for volatility modeling: the relevance of cross-market information is time-varying and state-dependent, rather than constant over the sample.

Taken together, the tail Granger causality and connectedness analyses jointly indicate that tail risk spillovers are both statistically significant and time-varying, providing a coherent empirical foundation for incorporating cross-market information in a flexible and state-dependent manner.

3. Volatility Forecasting Models

This section introduces the volatility forecasting framework employed in this study. Building on the volatility dynamics and risk measures constructed in the previous sections, we first establish benchmark HAR-type models and then propose several extensions that

incorporate cross-market information motivated by tail risk spillovers. Section 3.1 presents the benchmark HAR and HAR^{IV} models, which serve as market-specific baselines for volatility forecasting. Section 3.2 introduces cross-market HAR-type specifications that allow volatility dynamics in one market to depend on information from the other market. Section 3.3 describes the multi-step-ahead forecasting strategy used for out-of-sample evaluation.

3.1. Benchmark HAR-Type Models

The heterogeneous autoregressive (HAR) model [14–16] and HAR^{IV} model extension with implied volatility [17,18] have already been introduced in the context of modeling the price process and constructing tail risk measures in Equation (2) and Equation (3). Here, we briefly recall their role and economic interpretation. The HAR framework provides a parsimonious and transparent reduced-form representation of realized volatility dynamics by aggregating information across multiple time horizons. Motivated by the heterogeneous market hypothesis, the model captures the idea that volatility persistence arises from the interaction of market participants operating at different investment horizons, ranging from short-term traders to long-term institutional investors. In the present study, the HAR model plays a dual role. First, it serves as the baseline specification for forecasting realized volatility in each market. Second, it provides the volatility dynamics underlying the construction of conditional risk measures used in the tail Granger causality and connectedness analyses. This dual usage ensures internal consistency between the risk spillover analysis and the subsequent forecasting exercise.

To incorporate forward-looking information from option markets, we adopt the HAR^{IV} specification, which augments the HAR structure with implied volatility components aggregated over matching horizons. By aligning the temporal aggregation of realized and implied volatility, the HAR^{IV} model preserves the heterogeneous-agent interpretation while allowing volatility forecasts to reflect both historical volatility dynamics and market expectations embedded in option prices.

Importantly, both the HAR and HAR^{IV} models are market-specific in nature: all realized and implied volatility components are drawn from the same underlying market. As such, these benchmark models do not account for cross-market information transmission, even though the empirical evidence in Section 2.3 demonstrates the presence of statistically significant and time-varying tail risk spillovers between markets. This limitation motivates the cross-market extensions introduced in the next subsection, where volatility forecasts for one market are augmented with information extracted from the other market in a manner that is both economically interpretable and consistent with the documented tail risk spillover patterns.

3.2. Cross-Market HAR Models

Motivated by the empirical evidence of tail risk spillovers between stock and oil markets, we extend the HAR framework by incorporating volatility information from an external market. For expositional clarity, we denote the target market as market 1 and the external market as market 2. In the empirical application, market 1 and market 2 correspond to the S&P 500 index and crude oil futures, respectively, or vice versa.

Let $RV_{l,t}$ and $IV_{l,t}$ denote the realized and implied volatility of market l , for $l \in \{1, 2\}$. Both realized and implied volatility measures are constructed over heterogeneous horizons, consistent with the HAR framework.

3.2.1. Direct Cross-Market Augmentation

Empirical evidence presented in Section 2.3 documents significant and time-varying tail risk spillovers between stock and oil markets. In particular, both the VAR-based tail Granger causality tests and the dynamic connectedness analysis indicate that extreme risk

conditions in one market contain predictive information about future risk dynamics in the other market. These findings motivate an extension of the benchmark HAR^{IV} framework that allows volatility information to flow across markets in a systematic manner.

To operationalize this idea, we augment the HAR^{CRV} model of the target market (market 1) by incorporating volatility information from an external market (market 2). We consider two direct cross-market HAR specifications that differ in the information set drawn from market 2.

The first specification introduces only realized volatility measures from market 2, capturing the transmission of ex post volatility shocks across markets. This *RV-only* cross-market HAR model is given by

$$RV_{1,t+1}^{HARCRV} = Z'_{1,t}\beta_1 + \sum_{h \in \{d,w,m\}} \beta_{2,h} RV_{2,t}^{(h)} + \varepsilon_{t+1}, \quad (14)$$

where under-scripts 1 and 2 denote market 1 and market 2, respectively, and $h \in \{d, w, m\}$ indexes daily, weekly, and monthly horizons.

The second specification further enriches the information set by including implied volatility from market 2, allowing forward-looking risk expectations in the external market to directly affect volatility dynamics in the target market. The resulting *RV-IV* cross-market HAR model can be written in compact form as

$$RV_{1,t+1}^{HARCRVIV} = Z'_{1,t}\beta_1 + \sum_{h \in \{d,w,m\}} \beta_{2,h} RV_t^{(2,h)} + \sum_{h \in \{d,w,m\}} \gamma_{2,h} IV_t^{(2,h)} + \varepsilon_{t+1}, \quad (15)$$

In both specifications, $Z'_{1,t}\beta_1$ denotes the standard own-market HAR^{IV} components, defined as

$$Z'_{1,t}\beta_1 = \beta_0 + \sum_{h \in \{d,w,m\}} \beta_{1,h} RV_{1,t}^{(h)} + \sum_{h \in \{d,w,m\}} \gamma_{1,h} IV_{1,t}^{(h)}.$$

Both realized and implied volatility components are constructed consistently with the heterogeneous autoregressive framework across all horizons.

Equations (14) and (15) extend the information set of the benchmark HAR^{IV} model along two dimensions. First, it incorporates backward-looking realized volatility and forward-looking implied volatility from market 2, allowing cross-market information to enter the forecasting equation directly. Second, by including volatility measures at heterogeneous horizons, the model preserves the economic interpretation of the HAR framework in which agents operating at different investment horizons contribute differently to volatility dynamics.

From an economic perspective, tail risk spillovers indicate that extreme risk conditions in one market may be associated with future risk dynamics in another market through broad economic linkages rather than a single dominant channel. In particular, stock and commodity markets are interconnected on both the demand side and the supply side; shocks to one market may therefore be reflected in the risk environment of the other [1,2,4,10].

The direct cross-market HAR^{CRVIV} specification in Equation (15) is designed to accommodate such interdependencies in a parsimonious and agnostic manner. By augmenting the information set of market 1 with realized and implied volatility measures from market 2 across heterogeneous horizons, the model allows cross-market information to enter the volatility forecasting process without committing to a specific economic propagation channel.

Importantly, this formulation remains purely reduced-form. It does not presuppose how volatility is transmitted across markets, nor does it require explicit assumptions about the underlying supply-side or demand-side mechanisms. Instead, it treats volatility measures from market 2 as potential predictors and evaluates their relevance based on

whether they contain incremental forecasting information beyond own-market realized and option-implied volatility. The presence and strength of cross-market spillovers are thus determined empirically rather than imposed ex ante.

In addition to individual model specifications, we consider a simple forecast combination strategy to account for model uncertainty regarding the strength and nature of cross-market spillovers. Specifically, we construct an averaged volatility forecast by taking the arithmetic mean of the fitted values obtained from the HAR^{IV} model, the HAR^{CRV} model, and the HAR^{CRVIV} model:

$$\widehat{RV}_{t+1}^{AVG} = \frac{1}{3} \left(\widehat{RV}_{t+1}^{HAR^{IV}} + \widehat{RV}_{t+1}^{HAR^{CRV}} + \widehat{RV}_{t+1}^{HAR^{CRVIV}} \right). \quad (16)$$

This combination approach exploits the complementary information embedded in different cross-market specifications while mitigating the risk of model misspecification. From an economic perspective, the averaged forecast reflects a diversified view of volatility dynamics, integrating own-market persistence, realized cross-market shocks, and forward-looking risk expectations from the external market.

3.2.2. Threshold-Dependent Cross-Market HAR Model

Empirical evidence presented earlier suggests that cross-market risk transmission is concentrated in extreme market conditions. To reflect this feature, we augment the cross-market HAR framework by allowing information from the external market to enter the volatility forecasting equation only when a spillover-related state variable lies outside a predefined range [38–40].

Specifically, we consider the following threshold-activated HAR specification:

$$RV_{t+1}^{Thresh} = \mathbf{Z}'_{1,t} \boldsymbol{\beta}_1 + \mathbb{I}(S_t \notin [c_L, c_U]) \left(\beta_{2,d} RV_{2,t}^{(d)} + \gamma_{2,d} IV_{2,t}^{(d)} \right) + \varepsilon_{t+1}, \quad (17)$$

where $\mathbf{Z}'_{1,t} \boldsymbol{\beta}_1 = \beta_0 + \sum_{h \in \{d,w,m\}} \beta_{1,h} RV_{1,t}^{(h)} + \sum_{h \in \{d,w,m\}} \beta_{1,h} IV_{1,t}^{(h)}$ denotes the standard own-market HAR^{IV} components, including daily, weekly, and monthly realized volatility as well as implied volatility. $RV_{t,d}^{(2)}$ and $IV_{t,d}^{(2)}$ are the daily realized and implied volatility from the external market. $\mathbb{I}(\cdot)$ is an indicator function that takes the value of one when the state variable S_t falls outside the interval $[c_L, c_U]$, and zero otherwise.

This formulation implies that cross-market information is activated only during extreme states of the spillover variable, corresponding to periods of unusually high stress or abnormal market conditions. For most observations, the indicator equals zero, reflecting the absence of economically meaningful spillovers in normal times.

The lower and upper bounds (c_L, c_U) are determined endogenously from the data. We adopt a grid-search procedure over a set of admissible threshold values and select the interval that minimizes the residual sum of squares (RSS) of the forecasting regression. Formally, the estimated thresholds are obtained as

$$(\hat{c}_L, \hat{c}_U) = \arg \min_{(c_L, c_U) \in \mathcal{C}} \sum_{t=1}^T \hat{\varepsilon}_{t+1}^2(c_L, c_U), \quad (18)$$

where $\mathcal{C}$ denotes the set of candidate threshold pairs, and $\hat{\varepsilon}_{t+1}(c_L, c_U)$ is the regression residual associated with a given threshold choice.

We consider alternative definitions of the state variable S_t to capture different dimensions of extreme market conditions in the external market. First, S_t is defined as the realized or implied volatility of the external market. In this case, the indicator $\mathbb{I}(S_t \notin [c_L, c_U])$ primarily captures periods of abnormally high volatility, with the upper bound c_U playing a dominant role and $c_L = 0$ ($\mathbb{I}(S_t \geq c_U)$). This specification reflects the notion that volatility

spillovers are most relevant when the external market experiences exceptionally high uncertainty. Second, we consider S_t defined as the return of the external market. Under this specification, the indicator function activates cross-market predictors when returns fall outside a symmetric or asymmetric interval around zero, capturing extreme positive or negative price movements. This setting allows for spillovers driven by sharp market rallies or crashes, consistent with the idea that large return shocks transmit information and risk across markets.

Across these alternative specifications, the common modeling principle is that cross-market volatility transmission is triggered only when the external market exhibits extreme behavior, either in terms of volatility or returns. This approach provides a flexible yet economically intuitive framework for incorporating state-dependent spillovers into volatility forecasting.

3.2.3. Transfer Learning-Based HAR Model

While the direct cross-market HAR specifications incorporate information from both markets symmetrically, empirical evidence on tail risk spillovers suggests that information transmission is often directional. To exploit this feature, we adopt a transfer learning framework in which volatility dynamics estimated from an external source market (market 2) are used to inform parameter estimation in the target market (market 1).

The transfer learning procedure proceeds in two steps. In the first step, we estimate a baseline HAR^{IV} model using data from market 2 and obtain the corresponding parameter vector, denoted by $\hat{\beta}^{(2)}$. This source model captures the intrinsic volatility dynamics of market 2 and serves as a reference structure for the target market.

In the second step, we estimate the HAR^{IV} model for market 1 while shrinking its parameter vector toward $\hat{\theta}^{(2)}$ through a ridge-type penalty. Formally, the transfer learning HAR model is obtained by solving

$$\hat{\beta}_{1,transfer} = \arg \min_{\beta} \sum_{t=1}^T (RV_{1,t+1} - Z'_{1,t}\beta)^2 + \lambda \|\beta - \hat{\beta}_2\|_2^2, \quad (19)$$

where $Z_{1,t}$ denotes the HAR^{IV} regressors constructed from market 1, and $\lambda \geq 0$ is a tuning parameter that controls the strength of information transfer from the source market. $\hat{\beta}_{1,transfer}$ represents the transfer learning parameter for market 1, while $\hat{\beta}_2$ denotes the parameter estimated solely from the $HAR - IV$ model based on market 2 data.

To ensure comparability between parameters across markets and to avoid scale-driven distortions in the shrinkage term, all regressors are standardized to have zero mean and unit variance prior to estimation. This normalization step ensures that the penalty reflects genuine similarities in volatility dynamics rather than differences in measurement units or volatility levels.

From an economic perspective, the transfer learning framework imposes a form of parameter similarity across markets, reflecting the idea that volatility formation mechanisms may share common features when markets are strongly connected, particularly during periods of elevated tail risk. The ridge penalty allows the target market to borrow information from the source market while retaining flexibility to deviate when local market conditions differ. As λ increases, the model places greater weight on the volatility structure learned from market 2, whereas setting $\lambda = 0$ reduces the specification to a standard HAR^{IV} model estimated solely on market 1 data.

Overall, this approach provides a disciplined way to incorporate cross-market information at the parameter level, complementing the direct predictor-based spillover models and offering an alternative channel through which tail risk transmission can enhance volatility forecasting.

3.3. Multi-Step Ahead Forecasting Strategy

Short-horizon volatility forecasts are often dominated by strong serial dependence in realized volatility, which may limit the ability to disentangle the incremental predictive content of alternative information sets. From a methodological perspective, focusing exclusively on one-step-ahead forecasts can therefore place disproportionate weight on short-run persistence, potentially masking more gradual effects associated with cross-market information. To complement the one-step-ahead analysis, we consider multi-step ahead volatility forecasting over longer horizons.

Formally, for a given forecast horizon $H \geq 1$, the H -step-ahead forecasting equation is specified as

$$RV_{t+H} = \mathbf{Z}'_{1,t} \boldsymbol{\beta}_{1,H} + \varepsilon_{t+H}, \quad (20)$$

where $\mathbf{Z}'_{1,t} \boldsymbol{\beta}_{1,H} = \beta_0 + \sum_{h \in \{d,w,m\}} \beta_{1,h}^H RV_{1,t}^{(h)} + \sum_{h \in \{d,w,m\}} \gamma_{1,h}^H IV_{1,t}^{(h)}$ denotes the standard own-market HAR^{IV} components, and $\boldsymbol{\beta}_H$ is a horizon-specific parameter vector. This specification allows the relationship between predictors and future volatility to vary flexibly with the forecast horizon.

For expositional clarity, Equation (20) presents the baseline formulation. All alternative models considered in the paper—including the direct cross-market HAR specifications, threshold-activated models, and transfer learning-based models—are extended to the multi-step setting in an analogous manner by replacing the one-step-ahead target variable with RV_{t+H} and allowing model parameters to depend on H . This direct multi-step framework ensures that forecasts at each horizon are estimated independently and rely solely on information available at time t .

3.4. Statistical Loss Functions

To evaluate the out-of-sample forecasting performance of the competing volatility models, we employ several commonly used statistical loss functions that capture different aspects of forecast accuracy. These loss measures are reported in the empirical performance section and provide complementary perspectives on model performance.

First, we consider the quasi-likelihood (QLIKE) loss function, defined as

$$L_{t+1}^{QLIKE} = \frac{RV_{t+1}}{\widehat{RV}_{t+1|t}} - \log\left(\frac{RV_{t+1}}{\widehat{RV}_{t+1|t}}\right) - 1. \quad (21)$$

QLIKE is robust to measurement error in realized volatility and assigns relatively higher penalties to forecast errors during periods of elevated volatility. As a result, smaller QLIKE values indicate more accurate volatility forecasts, particularly in high-volatility regimes.

As a benchmark accuracy measure, we also report the root mean squared error (RMSE), given by

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(RV_{t+1} - \widehat{RV}_{t+1|t} \right)^2}. \quad (22)$$

RMSE measures the average magnitude of forecast errors in volatility levels, with lower RMSE values corresponding to smaller deviations between realized and predicted volatility.

Together, these loss functions provide complementary statistical and economic criteria for evaluating volatility forecasts, ensuring that improvements in predictive accuracy are reflected in both overall fit and tail risk measurement.

4. Volatility Forecasting Performance

This section evaluates the out-of-sample volatility forecasting performance of the proposed models under a rolling-window framework. Building on the tail risk spillover

evidence documented in the previous sections, the analysis assesses whether incorporating cross-market information improves predictive accuracy and economic relevance relative to standard market-specific benchmarks.

Section 4.1 describes the rolling-window forecasting design and details the estimation of all model parameters, including threshold values and regularization parameters, using only information available at each forecast origin. Section 4.2 compares the statistical forecasting accuracy of competing models based on standard loss functions, with particular emphasis on QLIKE. Finally, Section 4.3 examines the economic significance of the volatility forecasts by implementing a simple volatility timing strategy, illustrating how improvements in statistical accuracy translate into portfolio performance gains.

4.1. Rolling Window and Parameter

The forecasting analysis is implemented using the same dataset described in Section 2.3.1, which contains realized and implied volatility measures for the S&P 500 and the crude oil market over the full sample period. Out-of-sample forecasts are generated under a rolling-window scheme in which model parameters are re-estimated at each forecast origin using a fixed-length estimation window and then used to produce multi-horizon volatility forecasts. We consider two rolling-window lengths, $W = 500$ and $W = 1000$ trading days, in order to assess the sensitivity of forecasting performance to the amount of historical information used in estimation. Shorter windows allow the models to adapt more quickly to recent structural changes, while longer windows provide greater estimation stability.

Within each rolling window, threshold parameters in the conditional forecasting model are estimated endogenously using only in-sample observations. Specifically, a sufficiently wide search range is specified for the conditioning variable so as to cover both lower and upper tail regions. Candidate threshold pairs are constructed on a finite grid based on empirical quantiles of the conditioning variable computed from the rolling estimation window. For each candidate pair, the corresponding threshold-based HAR forecasting regression is re-estimated, and the threshold combination that yields the smallest in-sample residual sum of squares in Equation (18) is selected. Both the threshold values and the associated regression coefficients are therefore determined entirely within the rolling window and are subsequently used to generate out-of-sample forecasts. Table A2 provides descriptive summary statistics of the resulting threshold estimates across the out-of-sample period, showing that the selected threshold values exhibit limited dispersion and remain broadly comparable across rolling window lengths $W = 500$ and $W = 1000$. Figure A1 further illustrates the time-varying nature of the threshold estimates, showing that the high threshold values, used to trigger market entries during high volatility periods, remain consistent across both window sizes. This stability indicates that the threshold mechanism captures persistent changes in the volatility environment rather than transitory noise.

For the transfer learning specifications, the regularization parameter λ is determined in a data-driven manner within each rolling window. At each forecast origin, λ is selected by cross-validation using only the in-sample observations in the corresponding estimation window. Standard 5-fold cross-validation is primarily employed, where data is partitioned into 5 equally sized folds, and λ is selected to minimize the mean squared error (MSE) on these Holdout sets. The selected value of λ is then fixed when producing the out-of-sample forecast for that window, ensuring that regularization is tuned without recourse to future information. As a robustness check, we additionally conducted a time-series-aware cross-validation (CV) scheme. This approach considers temporal dependencies by using sequentially ordered folds, aligning with the characteristics of time-series data. Time-series CV was performed using a rolling split design, where training sets incrementally

"roll" forward in time to reflect temporal evolution. Results from the time-series CV are largely consistent with those obtained from standard CV, with minor variations. While the time-series CV performs slightly worse in terms of the forecasting error, the similarity in outcomes supports the robustness of our transfer learning specifications.

Figure A2 displays the empirical distribution of the selected λ values for a rolling window of $W = 500$. We find that the results for $W = 1000$ exhibit a nearly identical pattern; thus, they are omitted for brevity. Crucially, the resulting values of λ remain in a moderate range for the majority of the period in both markets. This moderate magnitude indicates that the transfer learning framework does not simply force the target market to "mimic" the source market's structure (which would require an excessively large λ) but rather selectively incorporates cross-market information to regularize the forecasting process. Furthermore, there is a distinct asymmetry in the regularization strength: the equity market (SPX) often exhibits higher penalty coefficients and greater dispersion compared to the crude oil market (WTI), suggesting that the degree of information integration is more volatile and aggressive when transferring oil market signals to equities.

In unreported results, we also consider the alternative choice λ_{1se} , defined as the largest regularization parameter within one standard error of the minimum cross-validation error, which corresponds to a more conservative specification. The main forecasting conclusions remain qualitatively unchanged under this alternative choice.

4.2. Statistical Forecasting Accuracy

Tables 4 and 5 summarize the out-of-sample forecasting performance for WTI crude oil futures under rolling estimation windows of $W = 500$ and $W = 1000$, respectively. A first striking feature of the crude oil results is that cross-market models tend to deliver improvements that are not confined to a single horizon or a single window length. Relative to the benchmark HAR^{IV} model, adding information from the equity market often improves forecasting accuracy at the 30-day and 60-day horizons and, in several cases, also produces non-negligible gains at the one-day horizon. This broader distribution of gains across horizons contrasts with the equity market case discussed below and suggests that oil volatility is more responsive to external information, consistent with the view that oil market uncertainty is tightly linked to macro-financial conditions and to shifts in aggregate risk sentiment captured by equity markets.

The horizon patterns in the crude oil market are also informative about the role of forward-looking versus backward-looking cross-market information. Models that include both realized and implied volatility from the external market (i.e., HAR^{CRVIV}) frequently outperform the benchmark and the RV-only augmentation (i.e., HAR^{CRV}), particularly at longer horizons. This evidence is consistent with the interpretation that option-implied volatility conveys forward-looking information about future uncertainty that is not fully captured by realized volatility persistence alone. In the crude oil setting, this forward-looking channel is especially relevant because oil markets are exposed to abrupt repricing driven by global demand expectations, policy uncertainty, and risk-off episodes, all of which may be reflected earlier and more sharply in equity-market implied volatility. Thus, when equity-market implied volatility is incorporated into the oil volatility forecasting equation, it can help forecast not only short-run movements but also more persistent volatility components that matter for multi-week horizons.

Table 4. Performance comparison of volatility models for crude oil futures (500-day rolling window).

Model	1 DAY HORIZON			30 DAY HORIZON			60 DAY HORIZON		
	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE
No_threshold									
HAR^{IV}	0.01562	0	9.203	0.06469	0	17.85	0.08497	0	19.34
HAR^{CRV}	0.01556	0.3727	9.234	0.05903	8.759	17.45	0.07895	7.094	18.92
HAR^{CRVIV}	0.01556	0.3545	9.272	0.05544	14.3	18.31	0.06647	21.78	17.07
$HAR^{TRANSFER}$	0.01574	−0.7545	9.327	0.06328	2.182	18.18	0.0786	7.504	19.02
$MODEL^{AVG}$	0.01545	1.113	9.179	0.0569	12.04	17.49	0.07042	17.13	18.02
RV_threshold									
HAR^{IV}	0.01562	0	9.203	0.06469	0	17.85	0.08497	0	19.34
HAR^{CRV}	0.01569	−0.4621	9.356	0.06121	5.379	20.03	0.08459	0.4486	19.3
HAR^{CRVIV}	0.01597	−2.223	9.678	0.06245	3.473	20.75	0.08501	−0.04621	19.36
$MODEL^{AVG}$	0.01562	−0.01516	9.289	0.06155	4.856	18.72	0.08424	0.8682	19.29
RETURN_threshold									
HAR^{IV}	0.01562	0	9.203	0.06469	0	17.85	0.08497	0	19.34
HAR^{CRV}	0.01571	−0.588	9.357	0.06492	−0.3523	19.71	0.08546	−0.5671	19.35
HAR^{CRVIV}	0.01566	−0.2653	9.256	0.06166	4.69	19.9	0.08533	−0.4229	19.34
$MODEL^{AVG}$	0.01559	0.2164	9.233	0.06277	2.979	18.62	0.08475	0.2629	19.31

This table reports out-of-sample volatility forecasting performance for the WTI crude oil futures using a rolling estimation window of $W = 500$ trading days. Forecast horizons are $h \in \{1, 30, 60\}$ days. All models are estimated on log realized volatility, and volatility forecasts are converted back to levels using the lognormal transformation before computing loss functions. Reported statistics are average losses over the out-of-sample period; smaller values of QLIKE and RMSE indicate better forecasting performance. Q_{desc} (%) reports the percentage improvement in QLIKE relative to the benchmark HAR^{IV} model within the same window length and forecast horizon, computed as $100 \times (QLIKE_{HAR} - QLIKE_{model}) / QLIKE_{HAR}$, so positive (negative) values indicate improvement (deterioration). The blocks No_threshold, RV_threshold, and RETURN_threshold correspond to alternative rules regulating the inclusion of cross-market information, as described in Section 3.2. $MODEL^{AVG}$ denotes forecast averaging across the selected cross-market specifications.

Table 5. Performance comparison of volatility models for crude oil futures (1000-day rolling window).

Model	1 DAY HORIZON			30 DAY HORIZON			60 DAY HORIZON		
	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE
No_threshold									
HAR^{IV}	0.01606	0	9.774	0.07045	0	19.02	0.09761	0	20.91
HAR^{CRV}	0.01587	1.214	9.784	0.06731	4.461	18.60	0.09762	−0.008903	20.99
HAR^{CRVIV}	0.01578	1.741	9.736	0.06593	6.413	18.67	0.09039	7.398	20.70
$HAR^{TRANSFER}$	0.01604	0.1053	9.903	0.06698	4.935	18.91	0.08769	10.16	20.40
$MODEL^{AVG}$	0.01581	1.565	9.738	0.06669	5.345	18.65	0.09169	6.059	20.63
RV_threshold									
HAR^{IV}	0.01606	0	9.774	0.07045	0	19.02	0.09761	0	20.91
HAR^{CRV}	0.01606	0.01149	9.782	0.06848	2.800	20.21	0.09749	0.124	20.94
HAR^{CRVIV}	0.01619	−0.8054	10.21	0.06875	2.416	20.25	0.09748	0.1304	20.93
$MODEL^{AVG}$	0.01600	0.3467	9.834	0.06827	3.102	19.32	0.09718	0.4363	20.90
RETURN_threshold									
HAR^{IV}	0.01606	0	9.774	0.07045	0	19.02	0.09761	0	20.91
HAR^{CRV}	0.01608	−0.09931	9.786	0.07100	−0.7716	19.11	0.09794	−0.3415	20.92
HAR^{CRVIV}	0.01609	−0.1876	9.802	0.06842	2.890	18.67	0.09738	0.2348	20.91
$MODEL^{AVG}$	0.01601	0.3171	9.763	0.06939	1.514	18.86	0.09727	0.3421	20.90

This table reports out-of-sample volatility forecasting performance for the WTI crude oil futures using a rolling estimation window of $W = 1000$ trading days. Forecast horizons are $h \in \{1, 30, 60\}$ days. All models are estimated on log realized volatility, and volatility forecasts are converted back to levels using the lognormal transformation before computing loss functions. Reported statistics are average losses over the out-of-sample period; smaller values of QLIKE and RMSE indicate better forecasting performance. Q_{desc} (%) reports the percentage improvement in QLIKE relative to the benchmark HAR^{IV} model within the same window length and forecast horizon, computed as $100 \times (QLIKE_{HAR} - QLIKE_{model}) / QLIKE_{HAR}$, so positive (negative) values indicate improvement (deterioration). The blocks No_threshold, RV_threshold, and RETURN_threshold correspond to alternative rules regulating the inclusion of cross-market information, as described in Section 3.2. $MODEL^{AVG}$ denotes forecast averaging across the selected cross-market specifications.

Model rankings in the crude oil market further highlight an important role for transfer learning and forecast combination. In terms of model-specific performance, the transfer learning specification often plays a prominent role in the crude oil market. $HAR^{TRANSFER}$

frequently delivers competitive performance and, in several configurations, provides improvements that are comparable to or stronger than those from direct cross-market augmentation. This pattern suggests that “how” cross-market information is imported matters: disciplining oil volatility dynamics by shrinking parameters toward structures learned from the equity market can be particularly effective when oil volatility is noisy and subject to episodic structural changes. In such an environment, directly injecting equity volatility regressors may add signal, but it may also add noise or collinearity; transfer learning can mitigate these issues by borrowing information at the parameter level while still allowing the oil market to deviate when local conditions differ. Forecast averaging, $MODEL^{AVG}$, also performs robustly in the crude oil market. Its gains are consistent with the idea that there is model uncertainty regarding the dominant spillover channel at any given time—sometimes realized spillovers dominate, sometimes implied spillovers dominate, and sometimes regularization-based transfer is more stable. Averaging can therefore provide a reliable improvement by smoothing over these regime shifts.

Turning to the threshold designs, the crude oil results underscore that spillover regulation is not uniformly beneficial. While $RV_threshold$ specifications can improve forecasting accuracy in certain cases, they also lead to clear deteriorations in others, especially when the threshold restricts the model from exploiting cross-market information during periods that are not classified as “extreme” by the chosen state variable. This observation is economically plausible: oil volatility can react to macro-financial news and risk repricing even outside the most extreme volatility regimes, so a strict volatility threshold may inadvertently shut down a useful transmission channel. The $RETURN_threshold$ designs tend to be less reliable, which is also intuitive in this context: extreme returns are relatively rare and may not align well with the persistence and multi-horizon structure of realized volatility. Taken together, the crude oil evidence suggests that unconditional cross-market information inclusion (or flexible mechanisms such as averaging and transfer learning) is often the most robust strategy, whereas hard threshold rules require careful calibration to avoid discarding informative signals.

Comparing results across rolling-window lengths reveals additional insights specific to the crude oil market. The 500-day window can be more adaptive to recent structural breaks and regime changes in oil markets, potentially capturing rapid shifts in volatility formation after major events. However, this adaptiveness comes with higher estimation variability, which may amplify noise in certain model variants. The 1000-day window generally provides smoother performance and more stable rankings across specifications, reflecting the benefits of estimation stability when volatility persistence is strong. Importantly, the fact that cross-market improvements persist under both window lengths reinforces the conclusion that equity market information contains incremental predictive content for oil volatility beyond what is captured by the oil market’s own realized and implied volatility dynamics.

Overall, the crude oil market results indicate that cross-market information transmission is not only statistically relevant (as documented in the spillover analysis) but also economically useful for forecasting. The improvements are more consistent than in the equity market, are present across multiple horizons, and are supported by both direct predictor augmentation and parameter-transfer mechanisms. This market-specific robustness strengthens the central message of the paper: linking tail risk spillovers to forecasting is particularly valuable when the target market is strongly exposed to external macro-financial information, as is the case for crude oil.

We next turn to the S&P 500, reported in Tables 6 and 7 for rolling windows of $W = 500$ and $W = 1000$, respectively. While cross-market information from the crude oil market can improve equity volatility forecasts, the pattern of gains is more horizon-dependent and,

on average, less uniform than in the crude oil case. At the one-day horizon, improvements from cross-market augmentation are generally modest and mixed across specifications. In several cases, unconditional cross-market augmentation leads to small deteriorations relative to the benchmark HAR^{IV} model, indicating that very short-horizon equity volatility dynamics are dominated by strong own-market persistence and microstructure-driven fluctuations that may not benefit substantially from external signals. This result is consistent with the broader volatility forecasting literature, where one-step-ahead realized volatility is often well explained by its own recent history, leaving limited incremental explanatory power for additional regressors.

In contrast, at the 30-day and 60-day horizons, cross-market models exhibit more consistent improvements in QLIKE. Under the No_threshold design, HAR^{CRV} , HAR^{CRVIV} , and $MODEL^{AVG}$ frequently outperform the benchmark, suggesting that information from the crude oil market becomes more relevant for forecasting medium- to long-horizon equity volatility. This horizon-dependent pattern implies that cross-market spillovers primarily affect persistent volatility components rather than short-lived fluctuations. A complementary explanation relates to the construction and information horizon of implied volatility. The VIX index (also for OVX) is designed to measure market expectations of volatility over approximately the next 30 calendar days, implying that its information content is naturally aligned with medium-term uncertainty rather than very short-lived daily movements. Consequently, cross-market information transmitted through implied volatility is expected to matter more for volatility forecasts at horizons beyond the daily frequency. The stronger performance of cross-market models at the 30-day and 60-day horizons is therefore economically intuitive, as these horizons better match the forward-looking window embedded in option-implied volatility measures.

The RMSE results for the equity market broadly align with the QLIKE-based findings, although the magnitude of improvements is typically smaller. This difference is expected because QLIKE is more sensitive to forecast errors during high-volatility periods and is more robust to measurement error in realized volatility. The fact that QLIKE improvements exceed RMSE improvements suggests that the main benefit of cross-market information for equity volatility forecasting arises precisely in high-volatility or stressed regimes, when correctly predicting volatility is most economically valuable.

Rolling-window comparisons also reveal a trade-off between adaptability and stability in the equity market. The 500-day window occasionally delivers larger improvements at particular horizons but exhibits greater variability across model specifications, reflecting sensitivity to short-lived episodes and estimation noise. The 1000-day window produces smoother and more stable performance patterns, indicating that longer windows may better capture the persistent components of volatility dynamics that drive medium- and long-horizon forecasts.

Finally, the threshold designs yield a similar qualitative message in the equity market as in the crude oil market, but with some notable differences. Volatility-based thresholds (RV_threshold) deliver mixed performance: they can improve forecasts in some configurations but also underperform the no-threshold counterparts, especially at longer horizons under the 1000-day window. Return-based thresholds (RETURN_threshold) are generally less effective and tend to produce limited or unstable gains. Taken together, these results suggest that, within our framework and for the equity market, unconditional inclusion of cross-market information is often (though not always) the most robust approach, with forecast averaging providing a particularly stable improvement by diversifying across alternative cross-market specifications.

Across all dimensions, forecast averaging performs robustly in the equity market, especially at longer horizons. Although no single cross-market specification dominates

uniformly, $MODEL^{AVG}$ remains competitive by balancing the strengths of individual models and mitigating specification-specific weaknesses. The transfer learning specification also exhibits noteworthy performance in certain settings: while it does not always dominate at longer horizons in the equity market, it can perform comparably to the benchmark at the one-day horizon and deliver stable improvements at medium horizons, suggesting that regularization-based information transfer may be a useful complement when direct cross-market augmentation is less effective.

Taken together with the spillover evidence and the empirical behavior of the regularization parameter, the differences in forecasting performance across markets suggest a plausible, albeit inferential, explanation for the underlying information dynamics. While these observations do not constitute a formal proof of the transmission mechanism, they provide a consistent basis for reasoning why transfer learning is more effective in the oil market than in the equity market. Based on the tail Granger causality analysis (Table 3), the equity market exerts more pervasive and robust spillovers on the crude oil market. This suggests that the equity market serves as a stable source of information that helps regularize the more “noisy” oil market. The relatively stable and lower λ values observed in Figure A2 for the equity-to-oil direction support this inference, indicating that the model can consistently leverage equity signals to improve oil volatility forecasts. In contrast, the equity market models derive less benefit from oil market information. The higher and more volatile λ values required when transferring oil signals to equities imply that the oil market’s information is more difficult to distill, requiring more aggressive filtering to manage its inherent noise. Consequently, the observed discrepancy in model performance is likely rooted in this fundamental asymmetry in the quality and stability of information flow between the two markets.

This discrepancy in forecasting performance can be linked to the market-specific roles within the broader financial system. The equity market, acting as the information hub and a source of funding liquidity, provides a more stable and “clean” structural signal. This is empirically reflected in the relatively lower and more stable λ values for the equity-to-oil specification, indicating that the oil market benefits from learning this robust financial structure. In contrast, the “noisiness” of the crude oil market—partly driven by speculative flows and idiosyncratic supply shocks—requires the equity market model to adopt a more aggressive and volatile filtering mechanism (as evidenced by the higher and more dispersed λ in Figure A2). Consequently, the financialized nature of the oil market makes it a responsive recipient of equity signals, while its inherent volatility limits the reverse information flow, leading to the observed gains in predictive accuracy being concentrated in the oil market.

Additional forecasting results for alternative horizons are reported in Appendix A Tables A3 and A4 for the S&P 500, and Tables A5 and A6 for the crude oil market. Several horizon-dependent patterns are worth highlighting. Overall, forecasts at intermediate and longer horizons (e.g., 15-day and 45-day) tend to be more favorable than the 1-day horizon, although the relative ranking between 15-day and 45-day horizons is not uniform and varies across model specifications and markets. In terms of model class, specifications without thresholding generally deliver stronger and more stable performance across horizons. In particular, for the S&P 500 with $W = 500$, the CRVIV specification exhibits notably strong performance at the 15-day and 45-day horizons. These patterns are consistent with the main findings reported in this section, with the appendix tables providing the detailed comparisons.

Table 6. Performance comparison of volatility models for the S&P 500 (500-day rolling window.)

Model	1 DAY HORIZON			30 DAY HORIZON			60 DAY HORIZON		
	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE
No_threshold									
HAR^{IV}	0.0457	0	4.086	0.1216	0	6.642	0.121	0	6.511
HAR^{CRV}	0.04606	−0.7811	4.117	0.1152	5.243	6.593	0.1157	4.337	6.354
HAR^{CRVIV}	0.04652	−1.789	4.271	0.1169	3.806	6.617	0.1172	3.105	6.357
$HAR^{TRANSFER}$	0.04468	2.221	4.085	0.1189	2.154	6.752	0.1256	−3.862	6.878
$MODEL^{AVG}$	0.04581	−0.253	4.142	0.1139	6.298	6.529	0.1118	7.57	6.319
RV_threshold									
HAR^{IV}	0.0457	0	4.086	0.1216	0	6.642	0.121	0	6.511
HAR^{CRV}	0.04641	−1.55	4.179	0.1188	2.268	6.554	0.117	3.263	6.371
HAR^{CRVIV}	0.04736	−3.629	4.306	0.1186	2.392	6.582	0.1182	2.266	6.4
$MODEL^{AVG}$	0.04598	−0.6099	4.137	0.1184	2.614	6.552	0.1173	3.016	6.402
RETURN_threshold									
HAR^{IV}	0.0457	0	4.086	0.1216	0	6.642	0.121	0	6.511
HAR^{CRV}	0.04598	−0.6095	4.141	0.1224	−0.6563	6.664	0.1215	−0.4235	6.515
HAR^{CRVIV}	0.04614	−0.9743	4.211	0.1219	−0.2684	6.672	0.1209	0.09606	6.502
$MODEL^{AVG}$	0.04574	−0.08629	4.125	0.1212	0.283	6.643	0.1206	0.2846	6.498

This table reports out-of-sample volatility forecasting performance for the S&P 500 using a rolling estimation window of $W = 500$ trading days. Forecast horizons are $h \in \{1, 30, 60\}$ days. All models are estimated on log realized volatility, and volatility forecasts are converted back to levels using the lognormal transformation before computing loss functions. Reported statistics are average losses over the out-of-sample period; smaller values of QLIKE and RMSE indicate better forecasting performance. Q_{desc} (%) reports the percentage improvement in QLIKE relative to the benchmark HAR^{IV} model within the same window length and forecast horizon, computed as $100 \times (QLIKE_{HAR} - QLIKE_{model}) / QLIKE_{HAR}$, so positive (negative) values indicate improvement (deterioration). The blocks No_threshold, RV_threshold, and RETURN_threshold correspond to alternative rules regulating the inclusion of cross-market information, as described in Section 3.2. $MODEL^{AVG}$ denotes forecast averaging across the selected cross-market specifications.

Table 7. Performance comparison of volatility models for the S&P 500 (1000-day rolling window.)

Model	1 DAY HORIZON			30 DAY HORIZON			60 DAY HORIZON		
	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE
No_threshold									
HAR^{IV}	0.04709	0	4.009	0.1378	0	6.864	0.1501	0	7.025
HAR^{CRV}	0.04705	0.08298	4.025	0.1319	4.297	6.813	0.1481	1.291	7.063
HAR^{CRVIV}	0.04712	−0.05388	4.067	0.1326	3.772	6.836	0.1472	1.946	7.059
$HAR^{TRANSFER}$	0.04606	2.187	3.975	0.1269	7.916	6.766	0.1453	3.19	7.105
$MODEL^{AVG}$	0.04692	0.372	4.018	0.1317	4.408	6.798	0.1446	3.641	6.977
RV_threshold									
HAR^{IV}	0.04709	0	4.009	0.1378	0	6.864	0.1501	0	7.025
HAR^{CRV}	0.04728	−0.4007	4.031	0.1348	2.169	6.819	0.1513	−0.8453	7.034
HAR^{CRVIV}	0.04756	−0.9837	4.02	0.1354	1.732	6.831	0.1511	−0.6924	7.025
$MODEL^{AVG}$	0.04703	0.1325	3.978	0.1349	2.099	6.816	0.1495	0.4179	7.004
RETURN_threshold									
HAR^{IV}	0.04709	0	4.009	0.1378	0	6.864	0.1501	0	7.025
HAR^{CRV}	0.0474	−0.6536	4.058	0.1383	−0.3382	6.87	0.1508	−0.507	7.034
HAR^{CRVIV}	0.04749	−0.8394	4.099	0.1383	−0.3383	6.86	0.1506	−0.326	7.037
$MODEL^{AVG}$	0.04723	−0.2888	4.048	0.1376	0.1492	6.855	0.15	0.0601	7.022

This table reports out-of-sample volatility forecasting performance for the S&P 500 using a rolling estimation window of $W = 1000$ trading days. Forecast horizons are $h \in \{1, 30, 60\}$ days. All models are estimated on log realized volatility, and volatility forecasts are converted back to levels using the lognormal transformation before computing loss functions. Reported statistics are average losses over the out-of-sample period; smaller values of QLIKE and RMSE indicate better forecasting performance. Q_{desc} (%) reports the percentage improvement in QLIKE relative to the benchmark HAR^{IV} model within the same window length and forecast horizon, computed as $100 \times (QLIKE_{HAR} - QLIKE_{model}) / QLIKE_{HAR}$, so positive (negative) values indicate improvement (deterioration). The blocks No_threshold, RV_threshold, and RETURN_threshold correspond to alternative rules regulating the inclusion of cross-market information, as described in Section 3.2. $MODEL^{AVG}$ denotes forecast averaging across the selected cross-market specifications.

To formally assess whether the differences in average forecasting performance are statistically significant, we conduct Diebold–Mariano (DM) tests based on out-of-sample loss differentials. The DM tests are implemented using the QLIKE loss function, with the null hypothesis that the competing model and the benchmark deliver equal predictive accuracy. We adopt a one-sided alternative that the competing model outperforms the benchmark and compute test statistics using HAC-robust variance estimators with a Bartlett kernel to account for serial correlation induced by overlapping forecasts. The benchmark model in all comparisons is HAR^{IV} , and negative DM statistics indicate superior forecasting performance of the competing specification.

Tables 8 and 9 report the resulting DM test statistics for the crude oil market and the S&P 500, respectively, across different forecast horizons and rolling-window lengths. Several clear patterns emerge. First, the 30-day horizon consistently exhibits the strongest and most robust evidence of predictive gains, with statistically significant improvements observed for multiple competing specifications across both markets. This finding provides formal inferential support for the superior average QLIKE performance documented earlier at this intermediate horizon. Second, at the 1-day and 60-day horizons, evidence in favor of the transfer learning and model-averaging specifications is more selective but still present in a number of configurations, indicating that these models can deliver meaningful gains outside the intermediate horizon as well. In contrast, threshold-based specifications generally perform less favorably in the DM comparisons, with fewer statistically significant improvements across horizons. This pattern closely mirrors the corresponding QLIKE results and suggests that, in the present setting, thresholding does not systematically enhance out-of-sample volatility forecasting accuracy. Overall, the DM test results reinforce the main conclusions drawn from the average loss comparisons, highlighting the particular relevance of the 30-day horizon and the relative advantage of non-threshold-based specifications.

Table 8. Diebold–Mariano test results based on QLIKE loss for crude oil: $W = 500$ vs. $W = 1000$.

Model	WINDOW $W = 500$			WINDOW $W = 1000$		
	1 DAY	30 DAY	60 DAY	1 DAY	30 DAY	60 DAY
No_threshold						
HAR^{IV} (Bench)	–	–	–	–	–	–
HAR^{CRV}	–3.946 ***	–2.133 **	–1.440 *	–4.262 ***	–1.928 **	–1.623 *
HAR^{CRVIV}	–3.888 ***	–2.078 **	–1.572 *	–4.695 ***	–1.416 *	–1.299 *
$HAR^{TRANSFER}$	–4.833 ***	–2.461 ***	–2.515 ***	–2.390 ***	–1.908 **	–1.778 **
$MODEL^{AVG}$	–5.739 ***	–2.434 ***	–1.887 **	–5.948 ***	–1.889 **	–1.530 *
RV_threshold						
HAR^{IV} (Bench)	–	–	–	–	–	–
HAR^{CRV}	–2.357 ***	–2.064 **	0.989	–2.056 **	–1.437 *	0.867
HAR^{CRVIV}	0.224	–1.728 **	1.405	–0.584	–1.656 **	1.152
$MODEL^{AVG}$	–3.684 ***	–2.053 **	0.804	–2.861 ***	–1.893 **	0.874
RETURN_threshold						
HAR^{IV} (Bench)	–	–	–	–	–	–
HAR^{CRV}	–1.704 **	1.761	1.886	0.706	1.741	–0.799
HAR^{CRVIV}	–2.982 ***	–2.098 **	2.117	0.275	–1.890 **	–1.113
$MODEL^{AVG}$	–3.885 ***	–2.666 ***	1.614	–0.156	–1.607 *	0.535

This table reports one-sided Diebold–Mariano test statistics based on the QLIKE loss function. The benchmark model is HAR^{IV} . Negative test statistics indicate superior forecasting performance of the listed model relative to the benchmark. HAC-robust variance estimates with a Bartlett kernel are used to account for serial correlation in the loss differentials. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 9. Diebold–Mariano test results based on QLIKE loss for the S&P 500: $W = 500$ vs. $W = 1000$.

Model	WINDOW $W = 500$			WINDOW $W = 1000$		
	1 DAY	30 DAY	60 DAY	1 DAY	30 DAY	60 DAY
No_threshold						
HAR^{IV} (Bench)	–	–	–	–	–	–
HAR^{CRV}	3.281	−1.301 *	−1.321 *	−0.815	−2.071 **	−0.351
HAR^{CRVIV}	2.798	−0.930	−1.268	−1.932 **	−1.471 *	−1.175
$HAR^{TRANSFER}$	−7.205 ***	−2.063 **	−0.988	−5.951 ***	−2.603 ***	−2.144 **
$MODEL^{AVG}$	1.827	−1.944 **	−2.218 **	−5.341 ***	−2.182 **	−1.798 **
RV_threshold						
HAR^{IV} (Bench)	–	–	–	–	–	–
HAR^{CRV}	4.068	0.229	−1.628 *	1.236	−2.216 **	1.109
HAR^{CRVIV}	4.794	−1.026	−1.068	1.384	−1.735 **	1.462
$MODEL^{AVG}$	2.933	−1.530 *	−1.911 **	−1.120	−2.557 ***	0.706
RETURN_threshold						
HAR^{IV} (Bench)	–	–	–	–	–	–
HAR^{CRV}	2.355	1.791	−0.876	2.735	−0.285	0.781
HAR^{CRVIV}	2.607	−1.029	−1.616 *	2.141	0.226	−1.621 *
$MODEL^{AVG}$	2.467	−1.254	−2.467 ***	1.671	−1.578 *	−1.865 **

This table reports one-sided Diebold–Mariano test statistics based on the QLIKE loss function. The benchmark model is HAR^{IV} . Negative test statistics indicate superior forecasting performance of the listed model relative to the benchmark. HAC-robust variance estimates with a Bartlett kernel are used to account for serial correlation in the loss differentials. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

4.3. Economic Significance: A Volatility Timing Application

To assess the economic relevance of the forecasting improvements documented above, we complement the statistical evaluation based on QLIKE and RMSE with a simple volatility-timing (VT) exercise. Following the univariate volatility timing framework in [41,42], we construct portfolio weights using only volatility forecasts, without requiring explicit return predictability.

Specifically, let $\widehat{RV}_{i,t}$ denote the h -step-ahead forecast of realized volatility for asset i at time t . The portfolio weight assigned to asset i is defined as

$$w_{i,t} = \frac{\left(1/\widehat{RV}_{i,t}\right)^\eta}{\sum_{j=1}^N \left(1/\widehat{RV}_{j,t}\right)^\eta}, \quad (23)$$

where $\eta > 0$ controls the aggressiveness of the volatility timing strategy. Larger values of η place relatively more weight on assets with lower forecasted volatility. This strategy requires only univariate volatility forecasts and avoids portfolio optimization or matrix inversion, thereby reducing estimation error and transaction costs.

Given the portfolio weights, realized portfolio returns are computed as $r_{p,t} = \sum_{i=1}^N w_{i,t} r_{i,t}$. Portfolio turnover is measured following [42] as the sum of absolute changes in portfolio weights before and after rebalancing.

Table 10 reports the economic performance of the volatility timing strategy under alternative volatility forecasting models. Several patterns emerge. First, models that incorporate cross-market information—in particular HAR^{CRVIV} —consistently deliver lower portfolio volatility and higher Sharpe ratios relative to the benchmark HAR^{IV} model across different values of η . Second, these gains are achieved without materially increasing portfolio turnover, indicating that the improvements are not driven by excessive trading. We also experiment with additional combinations of model specifications and tuning parameters; the corresponding portfolio results, which are not reported for brevity, display patterns that are qualitatively consistent with the QLIKE- and Diebold–Mariano-based comparisons in Section 4.2. Several patterns emerge.

Table 10. Economic performance of volatility timing strategies based on volatility forecasts.

Model	Volatility of Portfolio Returns (%)		
	$\eta = 1$	$\eta = 2$	$\eta = 4$
HAR^{IV} (Benchmark)	18.50	17.36	17.45
HAR^{CRV}	18.47	17.33	17.45
HAR^{CRVIV}	18.42	17.30	17.43
$HAR^{TRANSFER}$	18.54	17.30	17.40
$MODEL^{AVG}$	18.46	17.33	17.44
Model	Portfolio Turnover (TO)		
	$\eta = 1$	$\eta = 2$	$\eta = 4$
HAR^{IV} (Benchmark)	0.0336	0.0338	0.0163
HAR^{CRV}	0.0316	0.0313	0.0145
HAR^{CRVIV}	0.0299	0.0294	0.0133
$HAR^{TRANSFER}$	0.0338	0.0345	0.0170
$MODEL^{AVG}$	0.0314	0.0311	0.0144
Model	Sharpe Ratio		
	$\eta = 1$	$\eta = 2$	$\eta = 4$
HAR^{IV} (Benchmark)	0.509	0.578	0.652
HAR^{CRV}	0.520	0.592	0.663
HAR^{CRVIV}	0.528	0.602	0.674
$HAR^{TRANSFER}$	0.512	0.583	0.656
$MODEL^{AVG}$	0.519	0.591	0.664
Model	Annualized Return (%)		
	$\eta = 1$	$\eta = 2$	$\eta = 4$
HAR^{IV} (Benchmark)	9.42	10.03	11.38
HAR^{CRV}	9.61	10.26	11.56
HAR^{CRVIV}	9.73	10.41	11.74
$HAR^{TRANSFER}$	9.49	10.09	11.40
$MODEL^{AVG}$	9.59	10.24	11.58

This table reports economic performance measures from a volatility timing (VT) strategy based on the volatility forecasts generated by different models. The VT strategy follows [41] and allocates portfolio weights inversely proportional to forecasted variances, with tuning parameter $\eta \in \{1, 2, 4\}$ controlling the aggressiveness of rebalancing. The portfolio consists of the S&P 500 and WTI crude oil assets. All volatility forecasts are obtained from the No_threshold specifications using a one-day forecast horizon and a rolling window of $W = 500$. Portfolio return volatility is annualized. Turnover (TO) measures rebalancing intensity. Sharpe ratios are computed assuming a zero risk-free rate. Annualized returns are calculated as the mean daily return multiplied by 252.

Overall, the results suggest that the statistical gains documented in the forecast evaluation translate into economically meaningful benefits in a simple asset allocation setting. Even in a parsimonious volatility timing framework that relies solely on volatility forecasts, incorporating cross-market information and transfer-based regularization improves risk-adjusted portfolio performance.

5. Conclusions

This paper examines tail risk spillovers and volatility forecasting between the U.S. equity market and the crude oil market. Using realized and implied volatility within a HAR-type framework, we document pronounced, asymmetric, and time-varying tail risk transmission across the two markets and show that such spillovers contain economically meaningful information for volatility forecasting.

Building on these findings, we propose several cross-market volatility forecasting strategies, including direct information augmentation, threshold-based designs, forecast averaging, and transfer learning. The empirical results indicate that incorporating cross-

market information improves forecasting performance primarily at medium and longer horizons, consistent with the forward-looking nature of implied volatility. Moreover, the relative effectiveness of different transmission mechanisms varies across markets: while forecast averaging performs robustly in the equity market, transfer learning proves particularly effective in the crude oil market, where volatility dynamics are noisier.

Overall, this paper highlights the importance of linking tail risk spillovers to volatility forecasting and demonstrates that flexible cross-market information transmission can enhance predictive performance across markets and horizons.

Several avenues for future research naturally emerge. First, more refined state-selection mechanisms could be developed to regulate cross-market information transmission, potentially integrating volatility, liquidity, and broader macro-financial conditions. Second, the framework could be extended to allow for richer cross-market and cross-asset information integration, enabling volatility forecasts to draw simultaneously on signals from multiple related markets. Finally, the horizon-dependent effectiveness of cross-market information points to the existence of intertemporal spillover mechanisms whose economic foundations remain unclear and warrant further investigation.

Author Contributions: Conceptualization, S.P. and Y.S.; methodology, S.P.; software, S.P.; validation, S.P. and Y.S.; formal analysis, S.P. and Y.S.; investigation, S.P. and Y.S.; resources, S.P. and Y.S.; data curation, S.P.; writing—original draft preparation, S.P.; writing—review and editing, S.P. and Y.S.; visualization, S.P.; supervision, S.P. and Y.S.; project administration, S.P.; funding acquisition, Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China [grant 71931004].

Data Availability Statement: The data that support the findings will be available in Risk Lab at the University of Chicago Booth School of Business at <https://dachxiu.chicagobooth.edu/> (accessed on 14 November 2025), and CME group at <https://www.cmegroup.com/> (accessed on 14 November 2025) following an embargo from the date of publication to allow for commercialization of research findings.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Table A1. Granger causality spillover test results: robustness using empirical quantiles.

M	Kernel	α	Crude Oil $\rightarrow$ S&P 500				S&P 500 $\rightarrow$ Crude Oil			
			DD	DU	UU	UD	DD	DU	UU	UD
20	Daniell	1%	−2.64	−2.57	22.50 ***	44.31 ***	39.93 ***	7.45 ***	19.91 ***	−1.84
		5%	5.52 ***	12.65 ***	4.19 ***	22.32 ***	11.34 ***	16.75 ***	8.36 ***	18.99 ***
30	Daniell	1%	−3.25	−2.61	16.74 ***	51.25 ***	52.20 ***	14.61 ***	31.11 ***	4.03 ***
		5%	6.34 ***	10.79 ***	4.00 ***	21.74 ***	11.88 ***	14.71 ***	12.79 ***	19.57 ***
60	Daniell	1%	−4.55	−3.41	16.35 ***	46.16 ***	55.32 ***	18.11 ***	46.62 ***	23.56 ***
		5%	8.43 ***	7.65 ***	4.50 ***	20.03 ***	11.20 ***	12.39 ***	17.66 ***	23.21 ***
20	Bartlett	1%	−2.31	−2.31	24.26 ***	41.80 ***	32.16 ***	5.98 ***	15.90 ***	−2.31
		5%	4.36 ***	14.36 ***	4.21 ***	23.81 ***	11.03 ***	18.21 ***	7.43 ***	18.51 ***
30	Bartlett	1%	−2.84	−2.84	20.27 ***	48.95 ***	44.73 ***	12.34 ***	27.31 ***	0.92
		5%	5.56 ***	12.45 ***	4.17 ***	23.11 ***	11.61 ***	16.25 ***	11.15 ***	19.71 ***
60	Bartlett	1%	−4.04	−4.04	18.46 ***	47.97 ***	53.97 ***	17.09 ***	43.00 ***	19.40 ***
		5%	7.87 ***	9.11 ***	4.51 ***	21.50 ***	11.33 ***	13.64 ***	16.45 ***	23.15 ***

Table A1. Cont.

M	Kernel	α	Crude Oil $\rightarrow$ S&P 500				S&P 500 $\rightarrow$ Crude Oil			
			DD	DU	UU	UD	DD	DU	UU	UD
20	Parzen	1%	−2.64	−2.64	20.29 ***	47.70 ***	44.07 ***	10.33 ***	23.41 ***	−1.53
		5%	5.50 ***	12.20 ***	4.06 ***	22.14 ***	11.43 ***	16.04 ***	9.59 ***	18.79 ***
30	Parzen	1%	−3.25	−3.25	17.20 ***	51.68 ***	52.71 ***	16.44 ***	36.28 ***	5.08 ***
		5%	6.54 ***	10.20 ***	4.11 ***	21.40 ***	11.53 ***	14.00 ***	13.62 ***	20.48 ***
60	Parzen	1%	−4.61	−4.61	16.11 ***	44.02 ***	55.43 ***	17.58 ***	46.39 ***	28.99 ***
		5%	8.62 ***	6.95 ***	4.36 ***	19.83 ***	10.81 ***	11.92 ***	18.34 ***	23.54 ***

This table provides robustness results using empirical quantiles to construct tail-event indicators and reports Granger causality spillover test results based on tail risk measures for different bandwidth parameters (M) and kernel functions. Reported values are Q statistics, with statistically significant statistics shown in bold. Superscripts ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. DD and UU denote downside-to-downside and upside-to-upside spillovers, while DU (UD) represents downside-to-upside (upside-to-downside) spillovers. Statistical significance is assessed using two-sided standard normal critical values.

Table A2. Summary statistics of optimal RV thresholds for 30-day volatility forecasts.

Statistic	HAR^{CRV}				HAR^{CRVIV}			
	W = 500		W = 1000		W = 500		W = 1000	
	Low	High	Low	High	Low	High	Low	High
Panel A: S&P 500 Models								
Mean	2.557	3.789	2.332	3.793	2.557	3.751	2.332	3.757
SD	0.287	0.398	0.195	0.323	0.287	0.384	0.195	0.325
Min	2.097	2.959	2.097	3.323	2.097	2.959	2.097	3.277
Q1	2.449	3.556	2.097	3.572	2.449	3.463	2.097	3.559
Median	2.483	3.762	2.449	3.753	2.483	3.740	2.449	3.645
Q3	2.799	4.053	2.456	4.022	2.799	3.955	2.456	4.018
Max	3.108	4.846	2.682	4.534	3.108	4.534	2.682	4.534
Panel B: Crude oil Models								
Mean	1.149	2.869	0.993	2.967	1.149	2.836	0.993	2.883
SD	0.262	0.459	0.116	0.554	0.262	0.442	0.116	0.501
Min	0.865	2.162	0.865	2.221	0.865	2.162	0.865	2.221
Q1	0.897	2.520	0.865	2.347	0.897	2.511	0.865	2.349
Median	1.103	2.782	0.966	3.232	1.103	2.777	0.966	2.928
Q3	1.308	3.237	1.103	3.453	1.308	3.233	1.103	3.281
Max	1.973	3.798	1.117	3.745	1.973	3.833	1.117	3.929

This table reports summary statistics of the optimal lower and upper threshold parameters (c_L, c_U) estimated for the RV-threshold cross-market HAR models at the 30-day forecast horizon. Thresholds are applied to the logarithm of realized volatility and are estimated endogenously within each rolling window by minimizing the in-sample residual sum of squares of the forecasting regression. Results are reported separately for rolling window lengths $W = 500$ and $W = 1000$. Panel A presents threshold estimates for S&P 500 volatility models incorporating crude oil information, while Panel B reports results for crude oil volatility models incorporating S&P 500 information. All threshold values are expressed in log scale and summarize the distribution of rolling-window estimates over the out-of-sample period.

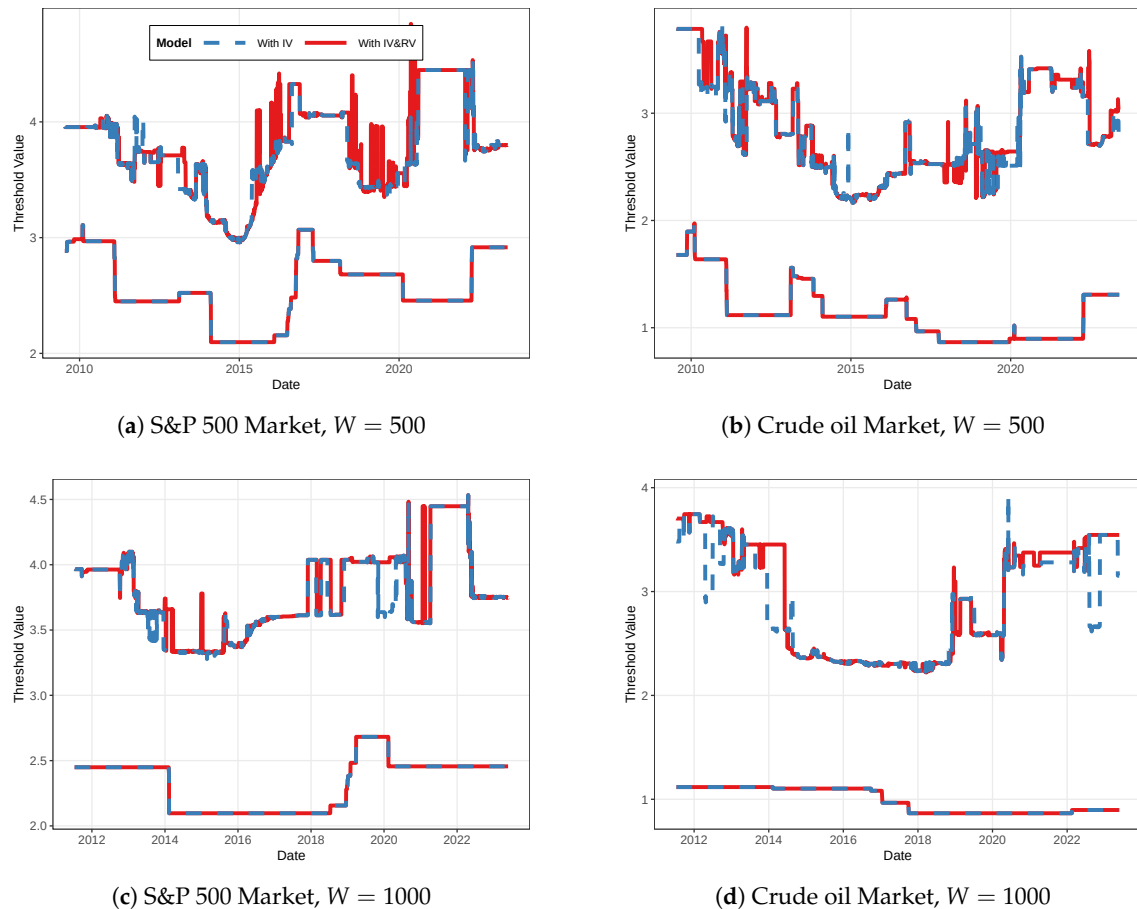


Figure A1. Time-varying threshold estimates for tail risk identification across markets and window sizes. This figure displays the estimated threshold values for tail risk identification over time. Each panel shows results for a specific market (S&P 500 or crude oil) and rolling window size ($W = 500$ or $W = 1000$). The solid lines represent the model without implied volatility (without IV), while the dashed lines represent the model with implied volatility (With IV). The lower threshold is not considered in this model, as only high volatility periods trigger market entries, which is reflected by the overlap of lower threshold values across both models.

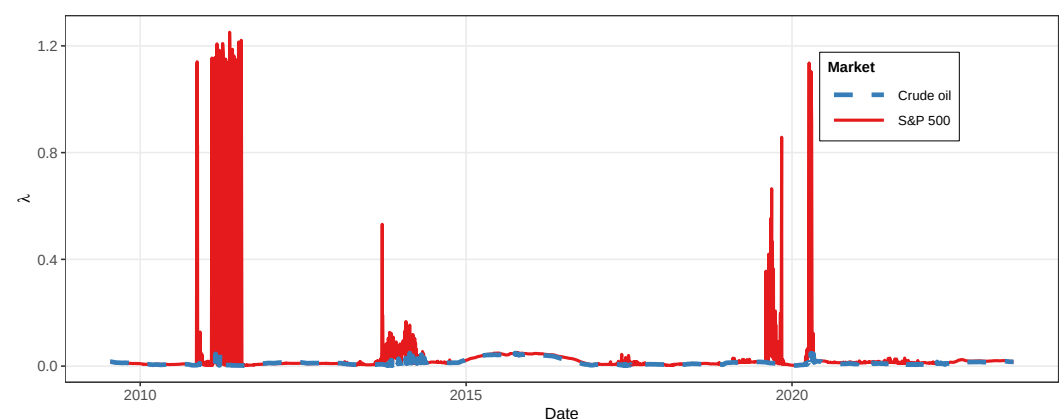


Figure A2. Time-varying lambda estimates ($W = 500$): This figure displays the transfer learning tuning parameter λ values over time for S&P 500 and crude oil markets under rolling window size $W = 500$. Solid lines represent the S&P 500 market and dashed lines represent the crude oil market.

Table A3. Volatility models for S&P 500 across additional forecast horizons (500-day rolling window).

Model	15 DAY HORIZON			45 DAY HORIZON		
	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE
No_threshold						
HAR^{IV}	0.1082	0	6.388	0.1232	0	6.656
HAR^{CRV}	0.1052	2.811	6.654	0.1182	4.113	6.552
HAR^{CRVIV}	0.1075	0.6997	6.688	0.1220	0.9769	6.617
$HAR^{TRANSFER}$	0.1051	2.891	6.426	0.1254	−1.727	6.904
$MODEL^{AVG}$	0.1037	4.155	6.485	0.1163	5.611	6.510
RV_threshold						
HAR^{IV}	0.1082	0	6.388	0.1232	0	6.656
HAR^{CRV}	0.1072	0.9386	6.375	0.1211	1.760	6.561
HAR^{CRVIV}	0.1073	0.8337	6.407	0.1215	1.430	6.525
$MODEL^{AVG}$	0.1066	1.472	6.363	0.1199	2.674	6.543
RETURN_threshold						
HAR^{IV}	0.1082	0	6.388	0.1232	0	6.656
HAR^{CRV}	0.1085	−0.271	6.436	0.1250	−1.447	6.683
HAR^{CRVIV}	0.1093	−1.030	7.444	0.1237	−0.4061	6.668
$MODEL^{AVG}$	0.1082	−0.004942	6.571	0.1233	−0.07606	6.657

This table reports out-of-sample volatility forecasting performance for the S&P 500 using a rolling estimation window of $W = 500$ trading days. Forecast horizons are $h \in \{15, 45\}$ days. All models are estimated on log realized volatility, and volatility forecasts are converted back to levels using the lognormal transformation before computing loss functions. Reported statistics are average losses over the out-of-sample period; smaller values of QLIKE and RMSE indicate better forecasting performance. Q_{desc} (%) reports the percentage improvement in QLIKE relative to the benchmark HAR^{IV} model within the same window length and forecast horizon, computed as $100 \times (QLIKE_{HAR} - QLIKE_{model}) / QLIKE_{HAR}$, so positive (negative) values indicate improvement (deterioration). The blocks No_threshold, RV_threshold, and RETURN_threshold correspond to alternative rules regulating the inclusion of cross-market information, as described in Section 3.2. $MODEL^{AVG}$ denotes forecast averaging across the selected cross-market specifications.

Table A4. Volatility models for S&P 500 across additional forecast horizons (1000-day rolling window).

Model	15 DAY HORIZON			45 DAY HORIZON		
	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE
No_threshold						
HAR^{IV}	0.1186	0	6.530	0.1452	0	7.063
HAR^{CRV}	0.1163	2.013	6.551	0.1384	4.662	6.971
HAR^{CRVIV}	0.1171	1.329	6.550	0.1403	3.369	7.038
$HAR^{TRANSFER}$	0.1096	7.622	6.433	0.1379	5.047	7.064
$MODEL^{AVG}$	0.1158	2.434	6.516	0.1383	4.790	6.961
RV_threshold						
HAR^{IV}	0.1186	0	6.530	0.1452	0	7.063
HAR^{CRV}	0.1177	0.7908	6.560	0.1437	1.023	7.039
HAR^{CRVIV}	0.1173	1.131	6.558	0.1433	1.330	7.042
$MODEL^{AVG}$	0.1172	1.245	6.531	0.1428	1.634	7.028
RETURN_threshold						
HAR^{IV}	0.1186	0	6.530	0.1452	0	7.063
HAR^{CRV}	0.1193	−0.555	6.583	0.1457	−0.3353	7.069
HAR^{CRVIV}	0.1192	−0.4368	6.635	0.1448	0.2931	7.065
$MODEL^{AVG}$	0.1187	−0.05668	6.565	0.1447	0.3222	7.057

This table reports out-of-sample volatility forecasting performance for the S&P 500 using a rolling estimation window of $W = 1000$ trading days. Forecast horizons are $h \in \{15, 45\}$ days. All models are estimated on log realized volatility, and volatility forecasts are converted back to levels using the lognormal transformation before computing loss functions. Reported statistics are average losses over the out-of-sample period; smaller values of QLIKE and RMSE indicate better forecasting performance. Q_{desc} (%) reports the percentage improvement in QLIKE relative to the benchmark HAR^{IV} model within the same window length and forecast horizon, computed as $100 \times (QLIKE_{HAR} - QLIKE_{model}) / QLIKE_{HAR}$, so positive (negative) values indicate improvement (deterioration). The blocks No_threshold, RV_threshold, and RETURN_threshold correspond to alternative rules regulating the inclusion of cross-market information, as described in Section 3.2. $MODEL^{AVG}$ denotes forecast averaging across the selected cross-market specifications.

Table A5. Volatility models for crude oil futures across additional forecast horizons (500-day rolling window).

Model	15 DAY HORIZON			45 DAY HORIZON		
	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE
No_threshold						
HAR^{IV}	0.04712	0	15.83	0.08093	0	19.22
HAR^{CRV}	0.04468	5.162	15.43	0.07545	6.771	18.91
HAR^{CRVIV}	0.04231	10.21	15.21	0.06795	16.04	18.54
$HAR^{TRANSFER}$	0.04598	2.423	15.55	0.07718	4.631	19.15
$MODEL^{AVG}$	0.04311	8.506	15.24	0.07081	12.5	18.57
RV_threshold						
HAR^{IV}	0.04712	0	15.83	0.08093	0	19.22
HAR^{CRV}	0.04531	3.830	15.14	0.07987	1.301	19.13
HAR^{CRVIV}	0.04657	1.163	15.61	0.08040	0.6545	19.67
$MODEL^{AVG}$	0.04540	3.647	15.24	0.07973	1.474	19.15
RETURN_threshold						
HAR^{IV}	0.04712	0	15.83	0.08093	0	19.22
HAR^{CRV}	0.04694	0.369	15.40	0.08135	−0.5224	19.24
HAR^{CRVIV}	0.04656	1.192	15.62	0.08060	0.3996	19.28
$MODEL^{AVG}$	0.04641	1.491	15.48	0.08057	0.4399	19.21

This table reports out-of-sample volatility forecasting performance for the WTI crude oil futures using a rolling estimation window of $W = 500$ trading days. Forecast horizons are $h \in \{15, 45\}$ days. All models are estimated on log realized volatility, and volatility forecasts are converted back to levels using the lognormal transformation before computing loss functions. Reported statistics are average losses over the out-of-sample period; smaller values of QLIKE and RMSE indicate better forecasting performance. Q_{desc} (%) reports the percentage improvement in QLIKE relative to the benchmark HAR^{IV} model within the same window length and forecast horizon, computed as $100 \times (QLIKE_{HAR} - QLIKE_{model}) / QLIKE_{HAR}$, so positive (negative) values indicate improvement (deterioration). The blocks No_threshold, RV_threshold, and RETURN_threshold correspond to alternative rules regulating the inclusion of cross-market information, as described in Section 3.2. $MODEL^{AVG}$ denotes forecast averaging across the selected cross-market specifications.

Table A6. Volatility models for crude oil futures across additional forecast horizons (1000-day rolling window).

Model	15 DAY HORIZON			45 DAY HORIZON		
	QLIKE	Q_{desc} (%)	RMSE	QLIKE	Q_{desc} (%)	RMSE
No_threshold						
HAR^{IV}	0.04878	0	16.30	0.09279	0	20.75
HAR^{CRV}	0.04680	4.049	15.84	0.09243	0.3955	20.76
HAR^{CRVIV}	0.04518	7.364	15.64	0.08697	6.275	20.76
$HAR^{TRANSFER}$	0.04718	3.275	16.05	0.08491	8.500	20.36
$MODEL^{AVG}$	0.04616	5.374	15.82	0.08870	4.411	20.61
RV_threshold						
HAR^{IV}	0.04878	0	16.30	0.09279	0	20.75
HAR^{CRV}	0.04834	0.9016	15.95	0.09249	0.3334	20.71
HAR^{CRVIV}	0.04861	0.3449	16.01	0.09267	0.1289	20.72
$MODEL^{AVG}$	0.04805	1.495	15.90	0.09236	0.4682	20.71
RETURN_threshold						
HAR^{IV}	0.04878	0	16.30	0.09279	0	20.75
HAR^{CRV}	0.04902	−0.5084	16.08	0.09322	−0.4606	20.77
HAR^{CRVIV}	0.04819	1.208	15.86	0.09264	0.1713	20.75
$MODEL^{AVG}$	0.04834	0.8954	16.03	0.09257	0.2414	20.74

This table reports out-of-sample volatility forecasting performance for the WTI crude oil futures using a rolling estimation window of $W = 1000$ trading days. Forecast horizons are $h \in \{15, 45\}$ days. All models are estimated on log realized volatility, and volatility forecasts are converted back to levels using the lognormal transformation before computing loss functions. Reported statistics are average losses over the out-of-sample period; smaller values of QLIKE and RMSE indicate better forecasting performance. Q_{desc} (%) reports the percentage improvement in QLIKE relative to the benchmark HAR^{IV} model within the same window length and forecast horizon, computed as $100 \times (QLIKE_{HAR} - QLIKE_{model}) / QLIKE_{HAR}$, so positive (negative) values indicate improvement (deterioration). The blocks No_threshold, RV_threshold, and RETURN_threshold correspond to alternative rules regulating the inclusion of cross-market information, as described in Section 3.2. $MODEL^{AVG}$ denotes forecast averaging across the selected cross-market specifications.

References

1. Park, J.; Ratti, R.A. Oil price shocks and stock markets in the US and 13 European countries. *Energy Econ.* **2008**, *30*, 2587–2608.
2. Apergis, N.; Miller, S.M. Do structural oil-market shocks affect stock prices? *Energy Econ.* **2009**, *31*, 569–575.
3. Maghyereh, A.I.; Awartani, B.; Bouri, E. The directional volatility connectedness between crude oil and equity markets: New evidence from implied volatility indexes. *Energy Econ.* **2016**, *57*, 78–93.
4. Aromi, D.; Clements, A. Spillovers between the oil sector and the S&P500: The impact of information flow about crude oil. *Energy Econ.* **2019**, *81*, 187–196.
5. Ding, H.; Kim, H.G.; Park, S.Y. Crude oil and stock markets: Causal relationships in tails? *Energy Econ.* **2016**, *59*, 58–69.
6. Basher, S.A.; Haug, A.A.; Sadorsky, P. Oil prices, exchange rates and emerging stock markets. *Energy Econ.* **2012**, *34*, 227–240.
7. Gozgor, G.; Lau, C.K.M.; Bilgin, M.H. Commodity markets volatility transmission: Roles of risk perceptions and uncertainty in financial markets. *J. Int. Financ. Mark. Institutions Money* **2016**, *44*, 35–45.
8. Degiannakis, S.; Filis, G.; Arora, V. Oil prices and stock markets: A review of the theory and empirical evidence. *Energy J.* **2018**, *39*, 85–130.
9. Liu, Z.; Ding, Z.; Lv, T.; Wu, J.S.; Qiang, W. Financial factors affecting oil price change and oil-stock interactions: A review and future perspectives. *Nat. Hazards* **2019**, *95*, 207–225.
10. Du, L.; He, Y. Extreme risk spillovers between crude oil and stock markets. *Energy Econ.* **2015**, *51*, 455–465.
11. Balçilar, M.; Demirer, R.; Hammoudeh, S. Quantile relationship between oil and stock returns: Evidence from emerging and frontier stock markets. *Energy Policy* **2019**, *134*, 110931.
12. Wen, D.; Wang, G.J.; Ma, C.; Wang, Y. Risk spillovers between oil and stock markets: A VAR for VaR analysis. *Energy Econ.* **2019**, *80*, 524–535.
13. Wang, L.; Ma, F.; Niu, T.; He, C. Crude oil and BRICS stock markets under extreme shocks: New evidence. *Econ. Model.* **2020**, *86*, 54–68.
14. Corsi, F. A simple approximate long-memory model of realized volatility. *J. Financ. Econom.* **2009**, *7*, 174–196.
15. Clements, A.; Preve, D.P. A practical guide to harnessing the har volatility model. *J. Bank. Financ.* **2021**, *133*, 106285.
16. Maheu, J.M.; McCurdy, T.H. Do high-frequency measures of volatility improve forecasts of return distributions? *J. Econom.* **2011**, *160*, 69–76.
17. Kambouroudis, D.S.; McMillan, D.G.; Tsakou, K. Forecasting realized volatility: The role of implied volatility, leverage effect, overnight returns, and volatility of realized volatility. *J. Futur. Mark.* **2021**, *41*, 1618–1639.
18. Busch, T.; Christensen, B.J.; Nielsen, M.Ø. The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *J. Econom.* **2011**, *160*, 48–57.
19. Wang, Y.; Chen, Z.; Ji, X. Cross-market information transmission and stock market volatility prediction. *N. Am. J. Econ. Financ.* **2023**, *68*, 101977.
20. Niu, Z.; Demirer, R.; Suleman, M.T.; Zhang, H. Speculation, cross-market sentiment and the predictability of gold market volatility. *J. Behav. Financ.* **2024**, *25*, 278–295.
21. Jiang, G.; Qiao, G.; Wang, L.; Ma, F. Hybrid forecasting of crude oil volatility index: The cross-market effects of stock market jumps. *J. Forecast.* **2024**, *43*, 2378–2398.
22. Hong, Y.; Liu, Y.; Wang, S. Granger causality in risk and detection of extreme risk spillover between financial markets. *J. Econom.* **2009**, *150*, 271–287.
23. Diebold, F.X.; Yilmaz, K. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *J. Econom.* **2014**, *182*, 119–134.
24. Antonakakis, N.; Chatziantoniou, I.; Gabauer, D. Refined measures of dynamic connectedness based on time-varying parameter vector autoregressions. *J. Risk Financ. Manag.* **2020**, *13*, 84.
25. Shiba, S.; Aye, G.C.; Gupta, R.; Goswami, S. Forecastability of agricultural commodity futures realised volatility with daily infectious disease-related uncertainty. *J. Risk Financ. Manag.* **2022**, *15*, 525.
26. Shi, S.; Yu, J. Volatility puzzle: Long memory or antipersistence. *Manag. Sci.* **2023**, *69*, 3861–3883.
27. Xiu, D. Quasi-maximum likelihood estimation of volatility with high frequency data. *J. Econom.* **2010**, *159*, 235–250.
28. Da, R.; Xiu, D. When moving-average models meet high-frequency data: Uniform inference on volatility. *Econometrica* **2021**, *89*, 2787–2825.
29. Geman, H.; Kharoubi, C. WTI crude oil futures in portfolio diversification: The time-to-maturity effect. *J. Bank. Financ.* **2008**, *32*, 2553–2559.
30. Gil-Alana, L.A.; Monge, M. Crude oil prices and COVID-19: Persistence of the shock. *Energy Res. Lett.* **2020**, *1*, 1–4.
31. Zhang, W.; Hamori, S. Crude oil market and stock markets during the COVID-19 pandemic: Evidence from the US, Japan, and Germany. *Int. Rev. Financ. Anal.* **2021**, *74*, 101702.
32. Huang, M.; Shao, W.; Wang, J. Correlations between the crude oil market and capital markets under the Russia–Ukraine conflict: A perspective of crude oil importing and exporting countries. *Resour. Policy* **2023**, *80*, 103233.

33. Tang, K.; Xiong, W. Index investment and the financialization of commodities. *Financ. Anal. J.* **2012**, *68*, 54–74.
34. Cheng, I.H.; Xiong, W. Financialization of commodity markets. *Annu. Rev. Financ. Econ.* **2014**, *6*, 419–441.
35. Brunnermeier, M.K.; Pedersen, L.H. Market liquidity and funding liquidity. *Rev. Financ. Stud.* **2009**, *22*, 2201–2238.
36. Kyle, A.S.; Xiong, W. Contagion as a wealth effect. *J. Financ.* **2001**, *56*, 1401–1440.
37. Xu, W.; Ma, F.; Chen, W.; Zhang, B. Asymmetric volatility spillovers between oil and stock markets: Evidence from China and the United States. *Energy Econ.* **2019**, *80*, 310–320. <https://doi.org/10.1016/j.eneco.2019.01.014>.
38. Tsay, R.S. Testing and modeling threshold autoregressive processes. *J. Am. Stat. Assoc.* **1989**, *84*, 231–240.
39. Hansen, B.E. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **1996**, *64*, 413–430.
40. Jawadi, F.; Ftiti, Z.; Louhichi, W. Forecasting energy futures volatility with threshold augmented heterogeneous autoregressive jump models. *Econom. Rev.* **2020**, *39*, 54–70.
41. Kirby, C.; Ostdiek, B. It's all in the timing: Simple active portfolio strategies that outperform naive diversification. *J. Financ. Quant. Anal.* **2012**, *47*, 437–467.
42. Clements, A.; Vasnev, A.L. Forecast combination puzzle in the HAR model. *J. Forecast.* **2024**, *43*, 118–137.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.