Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# A Sparse GARCH-MIDAS Approach with Minimax Concave Penalty for Crude Oil Volatility

**Xiaogang Dong**
Changchun University of Technology
**Hongzhen Jin**
Changchun University of Technology
**Xiwen Qin**
qinxiwen@ccut.edu.cn

Changchun University of Technology
**Liping Yuan**
Changchun University of Technology
**Shuo Li**
Jilin Tonglian Credit Service Co., Ltd

**Research Article**

**Keywords:**

**Posted Date:** November 13th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10212852/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# A Sparse GARCH-MIDAS Approach with Minimax Concave Penalty for Crude Oil Volatility

Xiaogang Dong* Hongzhen Jin† Xiwen Qin‡ Liping Yuan§ Shuo Li¶

July 1, 2026

## Abstract

Sparse GARCH-MIDAS specifications used in the volatility forecasting literature typically combine a Lasso-type penalty with a single information criterion, which leaves two issues unaddressed: uniform shrinkage biases large coefficients, and the retained factor set can be sensitive to the particular data split. We replace the Lasso penalty with the Minimax Concave Penalty (MCP), solved by ADMM, and add a path-level inclusion-frequency check from 5-fold expanding-window cross-validation, keeping only factors that are both HQIC-optimal in sample and stable across folds. Applied to WTI crude oil between February 2010 and November 2025, with 53 candidate factors covering uncertainty and geopolitics, oil-market fundamentals, macro-finance, and market sentiment and stress, the rule retains six core drivers spanning all four categories: the Commodity Markets EMV Tracker, drilling of oil and gas wells (IPN213111S), narrowly defined climate policy uncertainty, fiscal policy uncertainty, the World Sentiment Index, and the real broad dollar index. Our results show that the six-factor model attains the lowest in-sample information criteria, the lowest out-of-sample losses at the daily, weekly, and monthly horizons, and a Mincer–Zarnowitz slope close to the theoretical value of one; a joint Wald test on the six loadings rejects the zero-loading null at the 1% level, and the Diebold–Mariano and Clark–West tests are consistent with this ranking. A dense benchmark chosen by an information criterion alone produces substantially higher out-of-sample losses and a negative out-of-sample $R^2$, which shows that the cross-sample stability constraint, not just the penalty, accounts for much of the gain.

**Keywords:** GARCH-MIDAS; Minimax Concave Penalty (MCP); Crude oil volatility; High-dimensional variable selection; Dual-threshold selection; Out-of-sample forecasting.

**JEL Classification:** C22, C52, C58, G17, Q47.

*School of Mathematics and Statistics, Changchun University of Technology, 2055 Yan'an Street, Chaoyang District, Changchun 130012, Jilin Province, China. Email: dongxiaogang@ccut.edu.cn.

†School of Mathematics and Statistics, Changchun University of Technology. Email: 2697106897@qq.com.

‡Corresponding author. School of Mathematics and Statistics, Changchun University of Technology. Email: qinxiwen@ccut.edu.cn.

§School of Mathematics and Statistics, Changchun University of Technology. Email: 1202312001@stu.ccut.edu.cn.

¶Jilin Tonglian Credit Service Co., Ltd., Changchun, Jilin Province, China. Email: 2697106897@qq.com.

# 1 Introduction

Crude oil is a foundational commodity underpinning global energy, industry, transportation, and a wide range of downstream industries, and its price volatility therefore has broad economic consequences. Movements in crude oil prices feed directly into production costs and, in turn, into inflation expectations and wage–price dynamics; sharp fluctuations also complicate the conduct of monetary policy and amplify macroeconomic uncertainty, potentially dampening corporate investment (Gao et al., 2022). As an important risk factor, crude oil price volatility also influences asset pricing in financial markets (Diaz et al., 2016; Salisu et al., 2022). Given the central role of crude oil in the global economy, understanding and forecasting its price volatility is of substantial importance to policymakers, institutional investors, and commodity-market participants.

Identifying these drivers from a large candidate set, without sacrificing interpretability or predictive accuracy, is not straightforward. Traditional GARCH-family models capture the short-run dynamics of crude oil volatility well but become difficult to estimate once a large set of macroeconomic signals enters the specification (Herrera et al., 2018). Crude oil volatility responds to several interacting forces—macroeconomic fundamentals, geopolitical events, and financial-market uncertainty—so the modeling framework needs to accommodate multivariate high-dimensional information (Bakas and Triantafyllou, 2019; Wei et al., 2017). The GARCH-MIDAS framework of Engle et al. (2013) offers a unified specification for this purpose, decomposing the daily conditional variance into a short-run GARCH component and a low-frequency MIDAS component that aggregates macroeconomic signals. When the number of candidate predictors is large, however, direct maximization of the log-likelihood becomes prone to over-parameterization and degraded out-of-sample performance. Lasso-type penalties have been brought into the GARCH-MIDAS volatility-forecasting literature for factor selection (Chuang and Yang, 2022; Fang et al., 2020), but existing studies typically rely on a single information criterion to choose the regularization parameter, do not offer cross-sample robustness guarantees, and inherit the estimation bias that uniform Lasso shrinkage induces on large coefficients.

This paper combines two ingredients to address these limitations. The first is the Minimax Concave Penalty (MCP) of Zhang (2010), which delivers near-unbiasedness for large coefficients together with selection consistency and so corrects a known weakness of Lasso-type penalties; we solve the resulting nonconvex penalized problem with the Alternating Direction Method of Multipliers (ADMM) of Boyd et al. (2011). The second is a dual-threshold selection rule—"HQIC $\cap$ path-level cross-validation frequency"—that pairs the information-criterion optimum with a cross-sample robustness constraint. Using WTI crude oil data from February 2010 to November 2025, we identify six core drivers from a set of 53 candidate macroeconomic factors spanning four categories: uncertainty and geopolitics, oil-market fundamentals, macro-finance, and market sentiment and stress, and then evaluate their out-of-sample forecasting performance and economic value.

Our analysis yields three main findings. First, the six core factors identified by the dual-threshold rule—Commodity Markets EMV Tracker, Industrial Production of Drilling Oil and Gas Wells (IPN213111S), narrowly defined Climate Policy Uncertainty (CPU_narrow), Fiscal

Policy Uncertainty (EPU_Fiscal), World Sentiment Index (WSI), and Real Broad Dollar Index (FX_BG)—attain the lowest in-sample HQIC and BIC, and a joint Wald test on the six loadings rejects at the 1% level ($W = 25.25$, $p < 0.001$), so the four classes of macroeconomic signals each carry independent predictive content after the others are controlled for. Second, across 1,434 daily out-of-sample observations from 5-fold expanding-window cross-validation, the six-factor model has the lowest MAE at the daily ($h = 1$), weekly ($h = 5$), and monthly ($h = 22$) horizons; the Diebold–Mariano, Clark–West, and Mincer–Zarnowitz tests are consistent with this ranking, the model attains an out-of-sample $R^2$ of 0.449 that exceeds all alternatives, and its unbiasedness slope of 1.099 is close to the theoretical value of one. Third, the dense HQIC benchmark (M5) with 21 factors that does not impose the path-frequency robustness constraint produces an out-of-sample RMSE that is 53% higher and an out-of-sample $R^2$ of $-0.287$, which points to severe overfitting when the regularization parameter is selected by an information criterion alone.

The paper contributes to the existing literature in three respects. On the methodological side, it brings the MCP penalty into the GARCH-MIDAS framework for crude oil volatility forecasting and provides a folded-concave alternative to the Lasso-based sparse GARCH-MIDAS studies of Fang et al. (2020), Chuang and Yang (2022), and Zhao (2022), attenuating the estimation bias that uniform shrinkage tends to induce on large coefficients. On the selection side, it pairs the information-criterion optimum with a cross-sample path-level stability check; this lowers the influence of any single sample split on the retained factor set, and a fully out-of-sample factor re-selection experiment offers further evidence on this point. On the empirical side, it assembles a candidate pool of 53 factors covering four categories of macroeconomic drivers—uncertainty and geopolitics, oil-market fundamentals, macro-finance, and market sentiment and stress—and evaluates the resulting model with Diebold–Mariano, Clark–West, and Mincer–Zarnowitz tests, Kupiec (1995) VaR backtesting, volatility-targeted portfolios, and a HAR-RV benchmark (Corsi, 2009).

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the dataset and the construction of candidate factors. Section 4 details the GARCH-MIDAS-MCP modeling framework, the ADMM solution algorithm, the dual-threshold selection rule, and the multiple testing procedures. Section 5 presents full-sample estimation results, out-of-sample forecast evaluation, the economic value analysis, and robustness checks. Section 6 concludes and discusses policy implications.

## 2 Related Literature

This section reviews related literature along three streams: the GARCH-MIDAS framework and its extensions; high-dimensional sparse estimation methods applied to volatility forecasting; and the empirical identification of macroeconomic drivers of crude oil volatility.

### 2.1 The GARCH-MIDAS Framework and Its Extensions

The GARCH-MIDAS model of Engle et al. (2013) decomposes the daily conditional variance into a short-run GARCH component and a low-frequency MIDAS component, providing a unified

specification for jointly modeling high-frequency price information and low-frequency macroeconomic signals. The framework has been widely applied in financial and commodity markets and extended along several dimensions. Pan et al. (2017) introduce regime switching to capture structural breaks; Asgharian et al. (2013) compare the predictive content of multiple macroeconomic factors within the GARCH-MIDAS framework; Conrad and Kleen (2020) propose a multiplicative-component extension that decomposes conditional variance more flexibly; Salisu et al. (2022) document the predictive ability of the global financial cycle for crude oil volatility within a GARCH-MIDAS specification; and Amendola et al. (2021) propose a Double Asymmetric GARCH-MIDAS-X model in which short- and long-run components respond asymmetrically to positive and negative shocks. Recent studies have continued to extend the framework in multiple directions: in *Computational Economics*, Akgun and Gulay (2025) compare 11 GARCH variants and deep-learning models in cryptocurrency markets, and Bera et al. (2026) integrate signal-decomposition techniques with the Realized GARCH model to build a hybrid forecasting framework; Nguyen and Javed (2023), in *Journal of Empirical Finance*, study the dynamic dependence between stock and bond returns within a mixed-frequency GAS-MIDAS copula framework. Taken together, this body of work shows that the GARCH-MIDAS framework is flexible enough to incorporate multiple macroeconomic drivers of crude oil volatility. Many of these contributions work with a pre-specified set of predictors, which is well suited to the structural questions they investigate. The angle pursued in this paper is different: when the candidate pool itself becomes large, how should the most informative predictors be identified in a way that is robust across sample splits?

### 2.2 High-Dimensional Sparse Estimation in Volatility Forecasting

An influential early work introducing high-dimensional sparse estimation into the GARCH-MIDAS framework is Fang et al. (2020), who apply the Adaptive Lasso penalty (Zou, 2006) to select core factors from a set of macroeconomic variables in equity markets. Subsequent studies extend this idea to commodity markets: Chuang and Yang (2022) and Zhao (2022) apply Lasso-type methods to identify core drivers of crude oil volatility, while Qiao et al. (2024) combine regularization with principal components analysis to handle high-dimensional features of Chinese crude oil futures. In a related direction, Bonnier (2022) uses a general-to-specific (GETS) algorithm to select exogenous predictors of crude oil volatility from a large candidate set, providing an alternative non-penalty-based perspective on high-dimensional predictor selection. This stream of work makes a case for regularization in high-dimensional GARCH-MIDAS estimation and leaves room for two refinements that the present paper takes up. One concerns the penalty itself: Lasso-type penalties impose uniform shrinkage on all nonzero coefficients, which biases large effects downward (Zhao and Yu, 2006). Folded-concave penalties such as SCAD (Fan and Li, 2001) and MCP (Zhang, 2010) are designed to attenuate this bias and have been studied extensively in linear and generalized linear models, but their behavior in the GARCH-MIDAS context has received less attention. The other concerns the choice of the regularization parameter, which in this literature is typically tied to a single information criterion; pairing the information-criterion optimum with a cross-sample stability check can lower the sensitivity to any one data split. The dual-threshold rule in this paper combines these two refinements within

the GARCH-MIDAS framework.

### 2.3 Macroeconomic Drivers of Crude Oil Volatility

The literature on macroeconomic drivers of crude oil volatility falls into four streams. In the economic-policy-uncertainty stream, Wei et al. (2017) jointly examine speculative, fundamental, and uncertainty indicators and find that EPU-type indices serve as dominant drivers of crude oil volatility in their sample; Liang et al. (2020) compare the predictive content of multiple uncertainty indicators for crude oil volatility. In the geopolitical-risk stream, the GPR index of Caldara and Iacoviello (2022) has been widely applied in crude oil market analyses, and Luo et al. (2022) assess the predictive ability of sentiment and geopolitical indicators in the presence of structural breaks. In the climate-policy and energy-related uncertainty stream, the CPU index of Gavriilidis (2021) provides a quantitative measure for assessing the impact of climate policy on energy markets, and Dang et al. (2023) construct an energy-related uncertainty index. In the broader macro-finance and uncertainty stream, Sheng et al. (2020) evaluate the spillover effects of structural oil shocks on macroeconomic uncertainty, and Wen et al. (2024) examine the predictive content of multiple uncertainty variables for crude oil volatility. Recent contributions in *Computational Economics* further extend the frontier: Kang et al. (2026) adopt an interpretable machine-learning framework with 140 macroeconomic and uncertainty predictors; Gupta et al. (2024) quantify the predictive content of geopolitical risk for safe-haven assets using more than a century of data; Léber and Egyed (2025) integrate media and social media sentiment with a GARCH-LSTM specification for Value-at-Risk forecasting; and García-Medina and Aguayo-Moreno (2024) propose an LSTM-GARCH hybrid model for cryptocurrency portfolio volatility. Many of these studies focus on a specific class of candidate factors, which is a sensible choice given the structural questions they address. The angle we take here is different: when representative indicators from these streams are placed together in one high-dimensional candidate pool, which of them survive? We assemble such a pool of 53 factors spanning the four categories reviewed above and let the MCP path together with the path-level stability check decide which combinations carry the strongest and most robust predictive content for crude oil volatility in our sample.

## 3 Data

### 3.1 WTI Crude Oil Prices and Realized Volatility

We use WTI crude oil as the proxy for the global crude oil market. Daily WTI spot prices are obtained from the DCOILWTICO series compiled by FRED from EIA data, covering January 4, 2010 to December 31, 2025—a total of 4,011 daily observations. We drop the extreme observation on April 20, 2020, when the WTI front-month futures contract settled at a historically negative price, and apply $\pm 25\%$ winsorization to the remaining series of daily log returns (affecting seven extreme observations) in order to mitigate the disproportionate influence of a small number of extreme values on parameter estimation. The cleaned sample contains 3,971 daily observations spanning 190 complete months from February 2010 to November 2025 (the partial first and last

months are excluded to ensure each month contains a full set of trading days). The monthly log realized variance is defined as $\log RV_t = \log \sum_{i=1}^{N_t} r_{i,t}^2$ (we follow the GARCH-MIDAS literature in loosely referring to $\log RV_t$ as "realized volatility" throughout the paper for consistency with the established terminology), where $N_t$ is the number of valid trading days in month $t$, yielding 190 monthly observations. Figure 1 displays the time series of WTI prices and daily returns; dashed vertical lines mark four major events within the sample: the November 2014 OPEC inaction, the July 2018 escalation of U.S.–China trade tensions, the March 2020 COVID-19 global pandemic (together with the April 2020 negative WTI futures price), and the February 2022 outbreak of the Russia–Ukraine conflict.

### 3.2 Construction of 53 Candidate Macroeconomic Factors

The 53 monthly candidate factors come from four categories. The uncertainty and geopolitics category (14 factors) includes the Economic Policy Uncertainty (EPU) index of Baker et al. (2016) with its Fiscal and Trade subcomponents and the Monetary Policy Uncertainty (BBD_MPU) index of the same authors, a quant-based U.S. economic policy uncertainty variant (EURQ_US), the Geopolitical Risk (GPR) index of Caldara and Iacoviello (2022) together with its Threats and Acts subcomponents, the narrow and broad versions of the Climate Policy Uncertainty (CPU) index of Gavriilidis (2021), the Energy-related Uncertainty Index (ETU) of Dang et al. (2023), and the World Uncertainty, World Pandemic Uncertainty, and World Trade Uncertainty indices (WUI, WPUI, WTUI) of Ahir et al. (2022). The oil-market fundamentals category (15 factors) covers U.S. crude oil production, imports, exports, refinery capacity utilization, commercial and Strategic Petroleum Reserve inventories, industrial production indices for oil and gas extraction (IPG211S), drilling of oil and gas wells (IPN213111S), and coal mining (IPN2121S), and prices of related energy commodities (Brent crude, gasoline, Henry Hub natural gas, European natural gas, Australian coal, and copper). The macro-finance category (15 factors) covers industrial production, total industrial capacity utilization (TCU), new orders, nonfarm payrolls, the unemployment rate, PPI, the federal funds rate, the 10-year Treasury yield, the term spread, the credit spread, M2, broad dollar indices (BIS-weighted nominal, FRED real broad goods-and-services, and FRED real dollar vs. advanced foreign economies), and the house price index. The market sentiment and stress category (9 factors) includes the Equity Market Volatility (EMV) trackers of Baker et al. (2026) (overall, policy, commodity, and disease subcomponents), the St. Louis Fed and Kansas City Fed financial stress indices, the VIX, the consumer confidence index, and the World Sentiment Index (WSI) of Ahir et al. (2022).

Raw factors are processed according to their statistical properties: for the 28 factors with clear time trends or growth dynamics (e.g., prices, production levels, monetary aggregates, currency indices, economic policy uncertainty indices, and other macroeconomic indicators), we use 100 times the consecutive monthly log difference as the transformed series, capturing the percentage growth rate; for the 6 factors that are I(1) level variables stationary after first-differencing (two interest rates—the federal funds rate and the 10-year Treasury yield—the unemployment rate, refinery capacity utilization, the VIX, and the consumer confidence index), we take the consecutive monthly first difference as the transformed series; for the 19 factors that exhibit mean-reverting behavior in their original level form (uncertainty indices, sentiment

and stress indices, the term and credit spreads, and TCU), we retain the original level without further differencing. All transformed series are then subjected to two preprocessing steps: first, values outside the 1st and 99th percentiles are truncated at the corresponding percentile cutoffs, in order to mitigate the impact of a small number of outliers on parameter estimation; second, each series is standardized using its full-sample mean and standard deviation, yielding zero-mean unit-variance series and removing scale heterogeneity across factors. The complete factor list and transformation method are reported in Table 7 in the Appendix.

Table 1 reports descriptive statistics for the WTI daily return, the monthly $\log RV_t$, and the six core factors selected by the dual-threshold rule (formally defined as the BALANCED_CORE set in Section 5.1). WTI daily returns exhibit pronounced negative skewness ($-0.19$) and high excess kurtosis (21.70); the Jarque–Bera normality test rejects the null at the 1% level. These features are consistent with the asymmetric return distribution and tail-fatness commonly observed in crude oil markets. The monthly $\log RV_t$ exhibits skewness and excess kurtosis close to those of a normal distribution, and the ADF test rejects the unit-root null at the 1% level. After standardization, the six core factors have means and standard deviations close to zero and one, respectively; their skewness ranges from $-2.03$ to 2.67 and excess kurtosis from $-0.51$ to 9.60, indicating substantial heterogeneity in their cross-sectional distributions. Descriptive statistics and stationarity tests for all 53 candidate factors—together with the WTI daily return, its absolute value, its square, and the monthly $\log RV$ (57 variables in total), with Benjamini–Hochberg FDR-adjusted $p$-values—are provided in Table 8 of the Appendix.

Table 1: Descriptive statistics.

| Variable | N | Mean | Std. Dev. | Skew. | Kurt. | JB | ADF | Q(10) |
|---|---|---|---|---|---|---|---|---|
| WTI daily return | 3,971 | 0.0009 | 2.6384 | −0.19 | 21.70 | 0.000*** | 0.000*** | 0.000*** |
| Monthly $\log RV$ | 190 | 0.0056 | 0.9996 | 0.40 | 0.21 | 0.087* | 0.002*** | 0.000*** |
| EMV_Commodity | 190 | −0.0012 | 1.0025 | 2.67 | 9.60 | 0.000*** | 0.000*** | 0.000*** |
| IPN213111S | 190 | 0.0000 | 1.0000 | −2.03 | 6.39 | 0.000*** | 0.000*** | 0.000*** |
| CPU_narrow | 190 | 0.0000 | 1.0000 | 0.10 | 1.52 | 0.000*** | 0.000*** | 0.000*** |
| EPU_Fiscal | 190 | −0.0009 | 1.0026 | 2.12 | 5.07 | 0.000*** | 0.029** | 0.000*** |
| WSI | 190 | 0.0011 | 1.0025 | −1.44 | 3.85 | 0.000*** | 0.001*** | 0.000*** |
| FX_BG | 190 | 0.0000 | 1.0000 | 0.13 | −0.51 | 0.342 | 0.000*** | 0.000*** |

*Notes:* WTI returns are expressed in percent. Kurtosis is the excess kurtosis. The JB, ADF, and Q(10) columns report $p$-values for the Jarque–Bera normality test, the Augmented Dickey–Fuller unit-root test, and the Ljung–Box test with 10 lags, all adjusted using the Benjamini–Hochberg FDR procedure. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

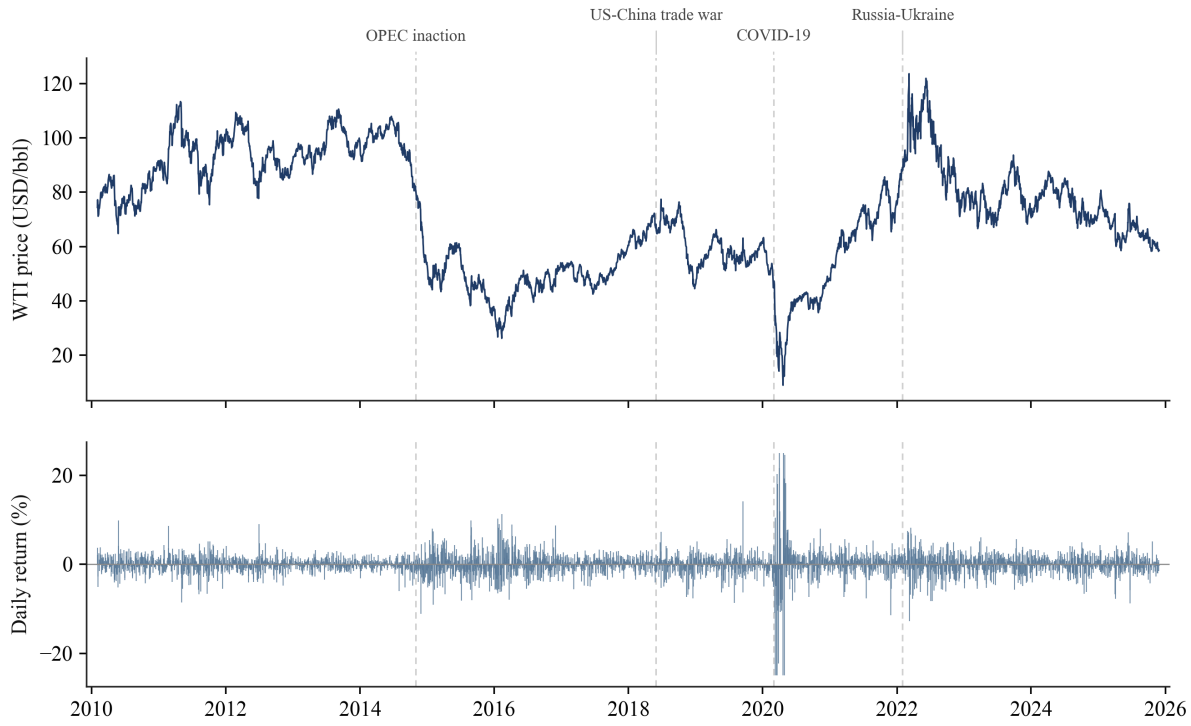



Figure 1: WTI crude oil daily prices and daily returns.

*Notes:* The figure displays daily WTI spot prices (top panel, USD per barrel) and daily log returns (bottom panel, in percent) from January 4, 2010 to December 31, 2025. Dashed vertical lines mark four key events: the November 2014 OPEC inaction, the July 2018 escalation of U.S.–China trade tensions, the March 2020 COVID-19 pandemic (together with the April 2020 negative WTI futures price), and the February 2022 outbreak of the Russia–Ukraine conflict.

# 4 Methodology

## 4.1 The GARCH-MIDAS-X Model

Let $r_{i,t}$ denote the log return on the $i$-th trading day of month $t$. Following the GARCH-MIDAS framework of Engle et al. (2013), we decompose the daily conditional variance into the product of a high-frequency GARCH component and a low-frequency MIDAS component:

$$r_{i,t} = \mu + \sqrt{\tau_t \cdot g_{i,t}} \cdot \varepsilon_{i,t}, \quad i = 1, \ldots, N_t \tag{1}$$

$$\varepsilon_{i,t} \mid \Phi_{i-1,t} \sim N(0,1) \tag{2}$$

where $\mu$ is the constant mean and $N_t$ is the number of trading days in month $t$. The daily conditional variance $\sigma^2_{i,t} = \tau_t \cdot g_{i,t}$ is decomposed into a short-run component $g_{i,t}$ and a long-run component $\tau_t$. To account for the asymmetric response of crude oil returns to positive and negative shocks, the short-run component follows the GJR-GARCH(1,1) specification of Glosten et al. (1993):

$$g_{i,t} = (1 - \alpha - \beta - \gamma/2) + [\alpha + \gamma \cdot \mathbf{1}\{r_{i-1,t} - \mu < 0\}] \cdot \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta \cdot g_{i-1,t} \tag{3}$$

with $\alpha > 0$, $\beta > 0$, $\gamma$ measuring the leverage effect, and the covariance stationarity condition $\alpha + \beta + \gamma/2 < 1$. The long-run component $\tau_t$ aggregates the past $K$ periods of low-frequency information via a MIDAS weighting scheme. We jointly include the realized volatility and a set of macroeconomic factors in the long-run component:

$$\log \tau_t = m + \beta_{RV} \cdot \sum_{k=1}^{K} \phi_k(\omega_{RV}) \log RV_{t-k} + \sum_{j=1}^{J} \theta_j \cdot \sum_{k=1}^{K} \phi_k(\omega) \cdot X_{j,t-k} \tag{4}$$

where the realized variance is defined as the monthly sum of squared daily returns:

$$RV_t = \sum_{i=1}^{N_t} r_{i,t}^2 \tag{5}$$

$X_{j,t}$ is the value of the $j$-th standardized macroeconomic factor in month $t$, and $J$ is the total number of candidate factors. The MIDAS weights follow the single-parameter Beta form (Engle et al., 2013); we adopt the $k/(K+1)$ normalization, which is also used in subsequent GARCH-MIDAS studies (Conrad and Kleen, 2020) and avoids the boundary case $0^{\omega-1}$ at $k = K$:

$$\phi_k(\omega) = \frac{(1 - k/(K+1))^{\omega-1}}{\sum_{l=1}^{K}(1 - l/(K+1))^{\omega-1}} \tag{6}$$

satisfying $\phi_k \geq 0$ and $\sum_{k=1}^{K} \phi_k = 1$. When $\omega > 1$, the weights decline with the lag order, assigning higher weights to more recent information. We use distinct shape parameters $\omega_{RV}$ and $\omega$ for realized volatility and the exogenous factors, respectively.

### 4.2 MCP Penalty and Sparse Variable Selection

Model parameters are estimated by Quasi-Maximum Likelihood (QMLE). Under the conditional normality assumption, the sample log-likelihood function is:

$$\ell(\Theta) = -\frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \left[ \log(2\pi \tau_t g_{i,t}) + \frac{(r_{i,t} - \mu)^2}{\tau_t g_{i,t}} \right] \tag{7}$$

where $T$ denotes the total number of months in the sample and $n = \sum_{t=1}^{T} N_t$ is the total number of daily observations. When a large number of exogenous factors enter the long-run component, directly maximizing $\ell(\Theta)$ tends to produce overfitting and impedes the identification of factors that genuinely carry predictive information. We therefore apply the Minimax Concave Penalty (MCP) of Zhang (2010) to regularize the exogenous-factor coefficients:

$$P_\lambda(\theta_j; a) = \begin{cases} \lambda|\theta_j| - \dfrac{\theta_j^2}{2a}, & \text{if } |\theta_j| \leq a\lambda \\ \dfrac{a\lambda^2}{2}, & \text{if } |\theta_j| > a\lambda \end{cases} \tag{8}$$

where $\lambda > 0$ controls the degree of sparsity and $a > 1$ is the concavity parameter, set to $a = 3$ in this paper. MCP imposes near-Lasso shrinkage on small coefficients but its penalty diminishes as $|\theta_j|$ grows and becomes constant when $|\theta_j| > a\lambda$, so that large nonzero coefficients are not

biased and the estimator satisfies the oracle property in large samples (Fan and Li, 2001; Zhang, 2010). Parameter estimation is carried out by minimizing the penalized negative log-likelihood:

$$\hat{\Theta}(\lambda) = \arg\min_{\Theta} \left\{ -\ell(\Theta) + n \cdot \sum_{j=1}^{J} P_{\lambda}(\theta_j; a) \right\} \tag{9}$$

where $n$ is the number of daily observations. The penalty is applied only to the exogenous-factor coefficients $\theta_j$; the short-run GARCH parameters, the MIDAS shape parameters, and the realized-volatility coefficient $\beta_{RV}$ are not penalized.

### 4.3 The Dual-Threshold Selection Rule

This paper introduces a dual-threshold selection rule combining the Hannan–Quinn information criterion with a path-level cross-validation frequency criterion. The first threshold uses the Hannan–Quinn information criterion (HQIC; Hannan and Quinn, 1979) to select the optimal $\lambda$ along the full-sample MCP path:

$$\mathrm{HQIC}(\lambda) = -2 \cdot \ell(\hat{\Theta}(\lambda)) + 2 \cdot k(\lambda) \cdot \log(\log n) \tag{10}$$

where $k(\lambda)$ is the total number of free parameters. HQIC strikes a balance between the over-dense tendency of AIC and the corner-solution tendency of BIC in small monthly samples, and remains relatively flat around the optimum, which helps reduce sensitivity to numerical perturbations. The second threshold computes a path-level inclusion frequency for each factor via 5-fold expanding-window cross-validation: for each candidate factor $j$, we count its inclusion frequency $f_j^{\mathrm{path}}$ across the $30 \times 5 = 150$ path-fold nodes and retain only those factors that are selected at the full-sample HQIC optimum and satisfy $f_j^{\mathrm{path}} \geq 90\%$. Relative to the conventional approach of relying solely on an information criterion to choose the regularization parameter, the path-level CV frequency threshold acts as a complementary tool that effectively reduces the out-of-sample overfitting risk—the global HQIC optimum may be influenced by idiosyncratic features of a specific data split, while factors that are robustly selected across sample periods are more likely to reflect true predictive content.

### 4.4 The ADMM Solution Algorithm

Because of the nonconvexity of MCP, the optimization problem does not admit a closed-form solution. We employ the Alternating Direction Method of Multipliers (ADMM; Boyd et al., 2011). Introducing the auxiliary variable $z = \theta$ and the dual variable $u$, the problem is decomposed into a smooth subproblem in $\Theta$ and a proximal-operator step in $\theta_j$ that are updated alternately: the former is solved by L-BFGS-B on the GARCH-MIDAS negative log-likelihood, and the latter has a closed-form firm-thresholding solution. The solution is computed along the $\lambda$ path from large to small values, using the solution at the previous node as the warm-start initial value, which substantially accelerates the overall path. We solve the full-sample MCP path on a grid of 30 logarithmically equispaced $\lambda$ nodes, where the maximum regularization parameter $\lambda_{\max}$ is determined by the KKT condition and the minimum is $\lambda_{\min} = 0.005 \cdot \lambda_{\max}$.

We also conduct a convergence sensitivity test under `max_outer` $\in \{20, 30, 50\}$ configurations of the ADMM outer-loop iteration count; the sparsity at the HQIC-optimal $\lambda$ is $k \in \{21, 21, 23\}$, so the factor identification does not change materially with the algorithm configuration.

### 4.5 Significance Tests

Crude oil daily returns typically exhibit heavy tails, so the conditional normality assumption may fail. In that case, the conventional Hessian-based standard errors are inconsistent (White, 1982). We therefore use the robust standard errors of Bollerslev and Wooldridge (1992) for inference, which remain asymptotically valid under conditional distribution misspecification. In our sample, the median ratio of robust to MLE standard errors across models lies between 1.40 and 1.45, with a maximum of 1.81, indicating a moderate degree of conditional misspecification.

To test the joint significance of the selected exogenous factors as a group, we use the Wald test. Under the null $H_0 : \theta_{\text{sub}} = \mathbf{0}$, the Wald statistic

$$W = \hat{\theta}'_{\text{sub}} \cdot \hat{\Sigma}^{-1}_{\text{sub}} \cdot \hat{\theta}_{\text{sub}} \tag{11}$$

is asymptotically $\chi^2(p)$, where $\hat{\theta}_{\text{sub}}$ is the subvector of factor coefficients to be tested and $\hat{\Sigma}_{\text{sub}}$ is the corresponding robust covariance submatrix. As a complement to the Wald test, we also perform the likelihood-ratio (LR) test:

$$\text{LR} = 2 \cdot [\ell(\hat{\Theta}_U) - \ell(\hat{\Theta}_R)] \tag{12}$$

where $\hat{\Theta}_U$ and $\hat{\Theta}_R$ are the QMLE estimators under the unrestricted and restricted ($\theta_{\text{sub}} = \mathbf{0}$) models, respectively. Under the null, LR is asymptotically $\chi^2(p)$ and can be cross-checked against the Wald test.

## 5 Empirical Findings

### 5.1 Full-Sample Estimation and Factor Identification

We solve the full-sample MCP path on a grid of 30 logarithmically equispaced $\lambda$ nodes, with maximum regularization parameter $\lambda_{\max} = 0.0120$ determined by the KKT condition. HQIC attains its minimum at $\lambda^* = 0.0068$, corresponding to 21 nonzero factors; AIC favors a denser solution ($k = 34$); and BIC, given the small monthly sample, prefers a corner solution ($k = 0$). We use HQIC as the primary criterion for selecting the regularization parameter, because its value lies between that of AIC and BIC and the criterion is relatively flat around the optimum. Figure 2 displays the sparsity along the $\lambda$ grid together with the evolution of the three information criteria.

To further filter out factors that are selected only incidentally at the global optimum, we combine the HQIC-optimal solution with a path-level inclusion-frequency threshold from 5-fold expanding-window cross-validation. This yields six core factors, which we hereafter refer to as the BALANCED_CORE set (so named because the selected factors balance across the four

candidate categories): Commodity Markets EMV Tracker ($f^{\text{path}} = 91\%$), Industrial Production of Drilling Oil and Gas Wells (IPN213111S; 98%), narrowly defined Climate Policy Uncertainty (CPU_narrow; 93%), Fiscal Policy Uncertainty (EPU_Fiscal; 95%), World Sentiment Index (WSI; 97%), and Real Broad Dollar Index (FX_BG; 92%). These six factors balance across the four categories: two from uncertainty and geopolitics, one from oil-market fundamentals, one from macro-finance, and two from market sentiment and stress. As a robustness check, we replace the expanding-window CV with a rolling-window CV (window length $W = 120$ months) and recompute the path frequencies. Of the six BALANCED_CORE factors, four still satisfy $f^{\text{path}} \geq 90\%$ under the rolling design. The remaining two—EPU_Fiscal (80%) and the World Sentiment Index (68%)—fall below the threshold, with WSI showing the greater sensitivity. The reduction in WSI's path frequency suggests that part of its contribution depends on the cumulative information available in expanding windows; we retain WSI in M4 because it satisfies the dual-threshold rule under the full sample, and we acknowledge this CV-design sensitivity as a limitation in Section 6. Figure 3 plots the standardized time series of the six BALANCED_CORE factors.

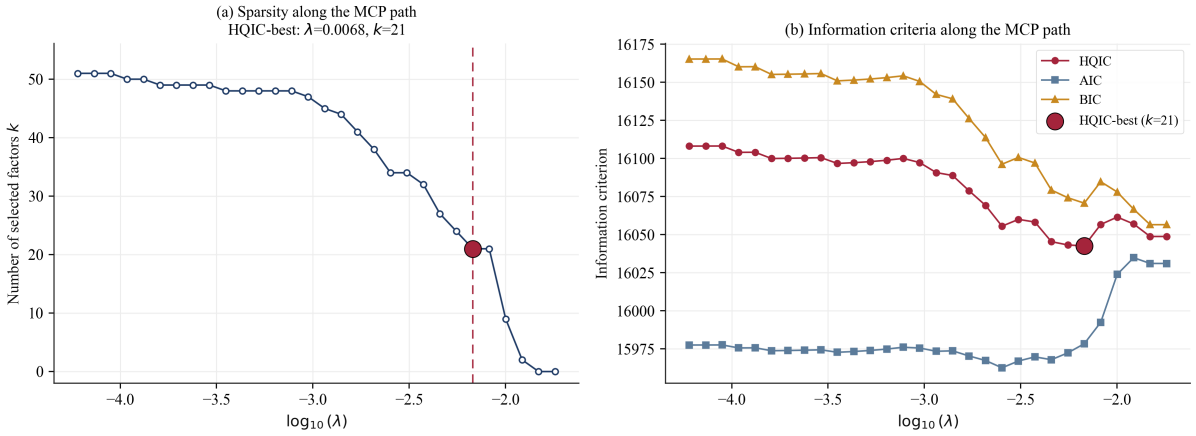



Figure 2: Sparsity and information criteria along the MCP path.

*Notes:* Panel (a) plots the number of selected factors (sparsity) along the 30 logarithmically equispaced $\lambda$ nodes of the MCP path. The red dashed line and red circle mark the HQIC optimum ($\lambda^* = 0.0068$, $k = 21$). Panel (b) overlays three information criteria—AIC, HQIC, and BIC—along the same $\lambda$ grid; the red circle marks the HQIC minimum.

To evaluate the predictive contribution of the core factor set, we specify five nested models for comparison. M1 is the GJR-GARCH(1,1) benchmark without a long-run component; M2 (GM-RV) augments M1 with $\log RV_t$ as the only long-run driver; M3 (GM-EMV) adds to M2 the Commodity Markets EMV Tracker—the factor with the largest absolute loading on the HQIC-optimal MCP path; M4 (GM-6Core) is our main model, with the six core factors identified by the dual-threshold rule; and M5 (GM-21HQIC) is a benchmark that contains all 21 nonzero factors at the HQIC optimum, used to assess the out-of-sample robustness of relying solely on an information criterion for regularization. Table 2 reports the full-sample QMLE estimates and information criteria. M4 attains the lowest BIC (16,007.81) and HQIC (15,994.26), the best performance under the two parsimony-leaning criteria; M5 attains the lowest AIC (15,938.46), but its BIC and HQIC are both higher than those of M4. This is consistent with the finding in Fang et al. (2020) that high-dimensional GARCH-MIDAS models tend toward over-parameterization

under the AIC criterion, foreshadowing M5's out-of-sample overfitting risk.

Table 2: Full-sample GARCH-MIDAS model estimation results.

| Parameter | M2 | M3 | M4 | M5 |
|---|---|---|---|---|
| $\mu$ | −0.0042 | −0.0021 | −0.0048 | −0.0028 |
| | (0.0310) | (0.0305) | (0.0301) | (0.0298) |
| $\alpha$ | 0.0548*** | 0.0453*** | 0.0386** | 0.0213 |
| | (0.0162) | (0.0150) | (0.0153) | (0.0152) |
| $\beta$ | 0.8568*** | 0.8617*** | 0.8642*** | 0.8752*** |
| | (0.0226) | (0.0240) | (0.0292) | (0.0339) |
| $\gamma$ | 0.0898*** | 0.0900*** | 0.0833*** | 0.0905*** |
| | (0.0241) | (0.0237) | (0.0245) | (0.0223) |
| $m$ | 1.5910*** | 1.5447*** | 1.4918*** | 1.4498*** |
| | (0.1408) | (0.1156) | (0.0936) | (0.0814) |
| $\beta_{RV}$ | 0.3469*** | 0.2823** | 0.3244*** | 0.1863 |
| | (0.1175) | (0.1235) | (0.1190) | (0.1541) |
| $\theta_{\text{EMV_Commodity}}$ | — | 0.2527*** | 0.4768*** | 0.5882*** |
| | — | (0.0734) | (0.1224) | (0.1367) |
| $\theta_{\text{IPN213111S}}$ | — | — | 0.1556** | 0.1188* |
| | — | — | (0.0749) | (0.0689) |
| $\theta_{\text{CPU_narrow}}$ | — | — | −0.2388** | −0.2061 |
| | — | — | (0.1015) | (0.1982) |
| $\theta_{\text{EPU_Fiscal}}$ | — | — | −0.2301** | −0.2161** |
| | — | — | (0.0953) | (0.0874) |
| $\theta_{\text{WSI}}$ | — | — | 0.0617 | 0.1177 |
| | — | — | (0.0732) | (0.0814) |
| $\theta_{\text{FX_BG}}$ | — | — | 0.0892 | 0.0108 |
| | — | — | (0.0891) | (0.1945) |
| $\omega_{RV}$ | 2.4920*** | 2.4923*** | 2.4944*** | 2.4985*** |
| | (0.0056) | (0.0055) | (0.0050) | (0.0028) |
| $\omega_X$ | — | 25.0679 | 10.6341*** | 9.4659** |
| | — | (19.9122) | (3.7473) | (4.7582) |
| $k$ | 8 | 9 | 14 | 29 |
| NLL | 8,007.55 | 7,990.89 | 7,967.63 | 7,940.23 |
| AIC | 16,031.10 | 15,999.79 | 15,963.27 | **15,938.46** |
| BIC | 16,056.56 | 16,028.42 | **16,007.81** | 16,030.73 |
| HQIC | 16,048.81 | 16,019.71 | **15,994.26** | 16,002.64 |

*Notes:* QMLE estimates for the full sample (February 2010–November 2025; monthly $M = 190$, daily $N = 3{,}971$). Values in parentheses are Bollerslev–Wooldridge sandwich standard errors. M1 (GJR-GARCH) has NLL = 8,012.90, AIC = 16,035.80, BIC = 16,051.71, HQIC = 16,046.87. The joint Wald test of the six factor loadings in M4 yields $W = 25.25$ ($p < 0.001$); the LR test of M4 vs. M2 yields LR = 79.83 ($p < 0.001$). Bold values mark the minimum in each row. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

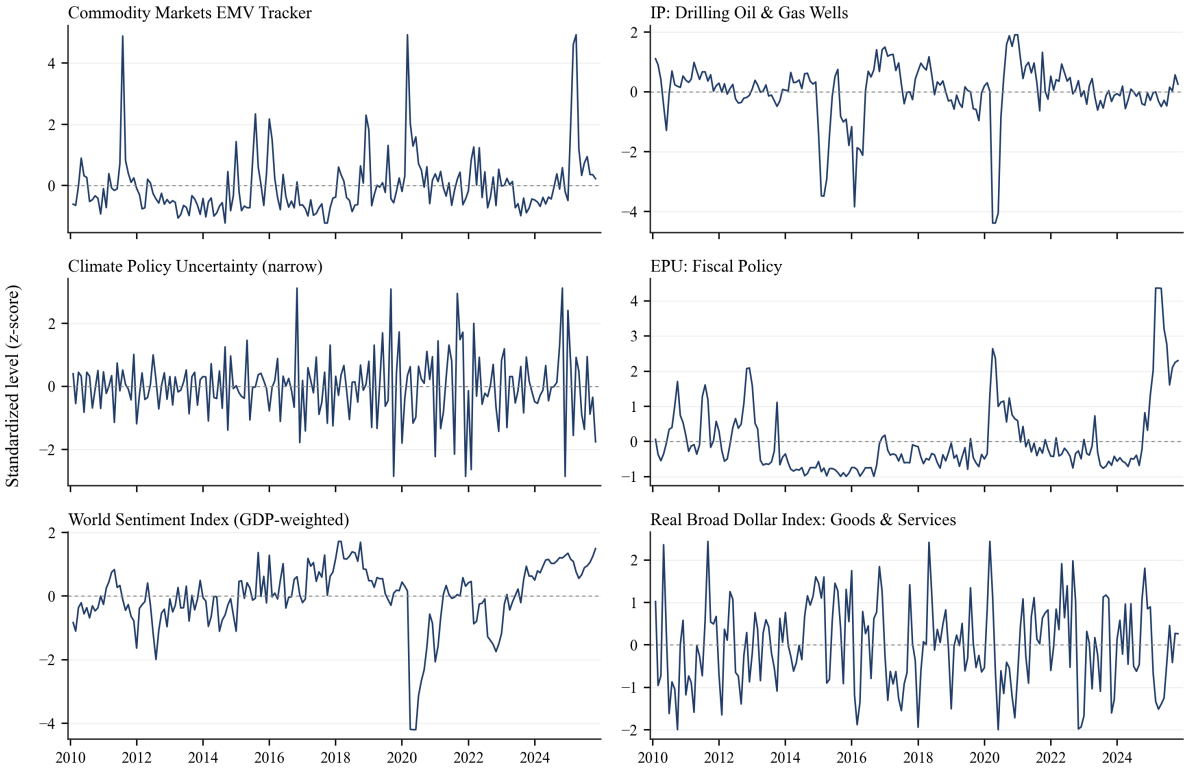



Figure 3: Standardized time series of the six BALANCED_CORE factors.

*Notes:* The figure displays the standardized ($z$-score) series of the six BALANCED_CORE factors identified by the dual-threshold rule over the sample period February 2010 to November 2025.

The parameter estimates for the M4 main model corroborate several stylized facts of crude oil volatility. The ARCH coefficient $\hat{\alpha} = 0.039$ is significant at the 5% level, and the GARCH coefficient $\hat{\beta} = 0.864$ is highly significant at the 1% level; their sum $\hat{\alpha} + \hat{\beta} + \hat{\gamma}/2 = 0.944$ is close to one, reflecting strong volatility persistence in the crude oil market. The asymmetry parameter $\hat{\gamma} = 0.083$ is significant at the 1% level, in line with the leverage effect documented for crude oil returns. The log-RV loading $\hat{\beta}_{RV} = 0.324$ is significant at the 1% level, indicating that past realized volatility itself remains an important driver of the long-run component even after including the six core factors. Among the six core factor loadings, the Commodity Markets EMV Tracker attains the strongest positive and significant loading ($\hat{\theta} = +0.477$, 1% level), the loading on IPN213111S is significantly positive (+0.156, 5% level), the loadings on Climate Policy Uncertainty and Fiscal Policy Uncertainty are significantly negative ($-0.239$ and $-0.230$, both at the 5% level), and the loadings on the World Sentiment Index and the Real Broad Dollar Index, while path-stable, are not individually significant at the 5% level. The joint Wald test of the six loadings yields $W = 25.25$ (df $= 6$, $p < 0.001$), rejecting the null that the six loadings are jointly zero; the likelihood-ratio test of M4 vs. M2 yields LR $= 79.83$ (df $= 6$, $p < 0.001$).

To measure the share of the conditional variance attributable to the long-run component, we compute the same-frequency variance ratio following Conrad and Kleen (2020):

$$\text{vr} = \frac{\text{Var}(\log \tau_t^{\text{daily}})}{\text{Var}(\log \sigma_t^{2,\text{daily}})} \times 100\% \tag{13}$$

where $\sigma_t^{2,\text{daily}} = \tau_t^{\text{daily}} \times g_t$ is the daily total conditional variance. For M4 the variance ratio is 39.1%, while for M2 it is only 11.1%; that is, the six core factors raise the long-run component's share of the total conditional variance by about 28 percentage points relative to the baseline. Figure 4 compares the annualized long-run component estimated by M4 with the annualized daily realized volatility; the four vertical dashed lines in Figure 4 mark notable points of the WTI price cycle within the sample period (the 2014 oil-price collapse and 2016 trough, the COVID-19/negative-price episode in 2020, and the 2022 Russia–Ukraine conflict), which differs slightly from the geopolitical-event annotations in Figure 1.

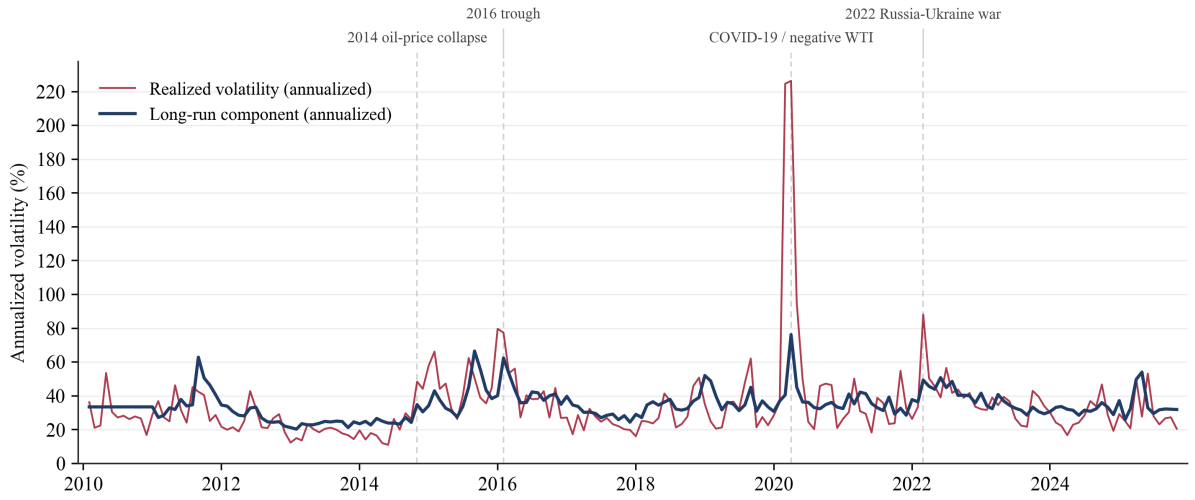



Figure 4: M4 long-run component vs WTI realized volatility.

*Notes:* The figure compares the annualized long-run component estimated from the M4 main model (dark blue line) with the annualized daily WTI realized volatility (red line), in percent. Four vertical dashed lines mark the 2014 oil-price collapse, the 2016 trough, the COVID-19/negative-price event in 2020, and the 2022 Russia–Ukraine conflict.

## 5.2 Out-of-Sample Forecast Performance

Out-of-sample forecast performance is evaluated under 5-fold expanding-window cross-validation. Each training set terminates at month 120, 134, 148, 162, and 176, respectively, and each testing window has a length of 14 months, producing a total of 1,434 daily out-of-sample forecasts. For each forecast horizon $h \in \{1, 5, 22\}$ trading days—corresponding to the daily, weekly, and monthly evaluation cycles—we use four loss functions: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), QLIKE loss, and the heteroskedasticity-robust Mean Absolute Error (HMAE) of Patton (2011). Table 3 reports the out-of-sample results, including the HAR-RV benchmark of Corsi (2009).

M4 has the lowest MAE at all three horizons; at $h = 5$ and $h = 22$, it also has the lowest RMSE, QLIKE, and HMAE. The improvement of M4 over the GARCH benchmark M1 in MAE rises monotonically from 9% at $h = 1$ to 20% at $h = 22$, consistent with the theoretical expectation that low-frequency macroeconomic factors primarily improve the long-run forecast component, which dominates multi-step-ahead conditional variances. M5 performs worst on nearly all horizon–loss combinations, and the gap to M4 is substantial: at $h = 22$, M5's MAE (226.48) is 53% higher than that of M4 (147.69), and its RMSE (795.70) is 53% higher than that

of M4 (520.39). This pattern is what one expects when the regularization parameter is chosen by HQIC alone, with no path-frequency robustness constraint: the model fits the training sample well but extrapolates poorly. Note also that the HAR-RV QLIKE is much higher than that of all GARCH-family models at all horizons, which suggests that high-frequency aggregation alone does not substitute for the high–low frequency information fusion of GARCH-MIDAS in a monthly framework for crude oil volatility.

Table 3: Out-of-sample forecast results (with HAR-RV benchmark).

| Model | MAE | RMSE | QLIKE | HMAE |
|---|---|---|---|---|
| $h = 1$ | | | | |
| HAR-RV | — | 122.05 | 2.830 | — |
| M1 (GJR) | 12.39 | 46.25 | 1.440 | 284.60 |
| M2 (GM-RV) | 12.09 | 46.64 | 1.494 | 292.90 |
| M3 (GM-EMV) | 12.43 | **46.15** | 1.467 | 307.52 |
| M4 (GM-6Core) | **11.28** | 46.27 | 1.465 | 307.42 |
| M5 (GM-21HQIC) | 13.82 | 56.02 | 1.527 | 307.00 |
| $h = 5$ | | | | |
| HAR-RV | — | 543.90 | 1.796 | — |
| M1 (GJR) | 43.08 | 140.97 | 0.484 | 1.398 |
| M2 (GM-RV) | 42.12 | 140.52 | 0.494 | 1.410 |
| M3 (GM-EMV) | 42.47 | 138.42 | 0.472 | 1.455 |
| M4 (GM-6Core) | **37.14** | **137.21** | **0.466** | **1.216** |
| M5 (GM-21HQIC) | 50.33 | 210.42 | 0.531 | 1.297 |
| $h = 22$ | | | | |
| HAR-RV | — | 1,785.99 | 1.578 | — |
| M1 (GJR) | 183.56 | 552.96 | 0.646 | 0.893 |
| M2 (GM-RV) | 179.32 | 550.06 | 0.659 | 0.878 |
| M3 (GM-EMV) | 174.11 | 552.94 | 0.639 | 0.872 |
| M4 (GM-6Core) | **147.69** | **520.39** | **0.540** | **0.658** |
| M5 (GM-21HQIC) | 226.48 | 795.70 | 0.622 | 0.753 |

*Notes:* Forecast horizon $h$ in trading days. Forecasts are generated under 5-fold expanding-window cross-validation, producing 1,434 daily out-of-sample observations. HAR-RV (Corsi, 2009) generates monthly RV forecasts under the same OOS window and then converts them to a daily-comparable scale; because HAR-RV does not provide daily-level MAE/HMAE under the monthly framework, the corresponding cells are marked "—". Lower loss values are better. Bold values mark the column minimum (excluding the HAR-RV benchmark).

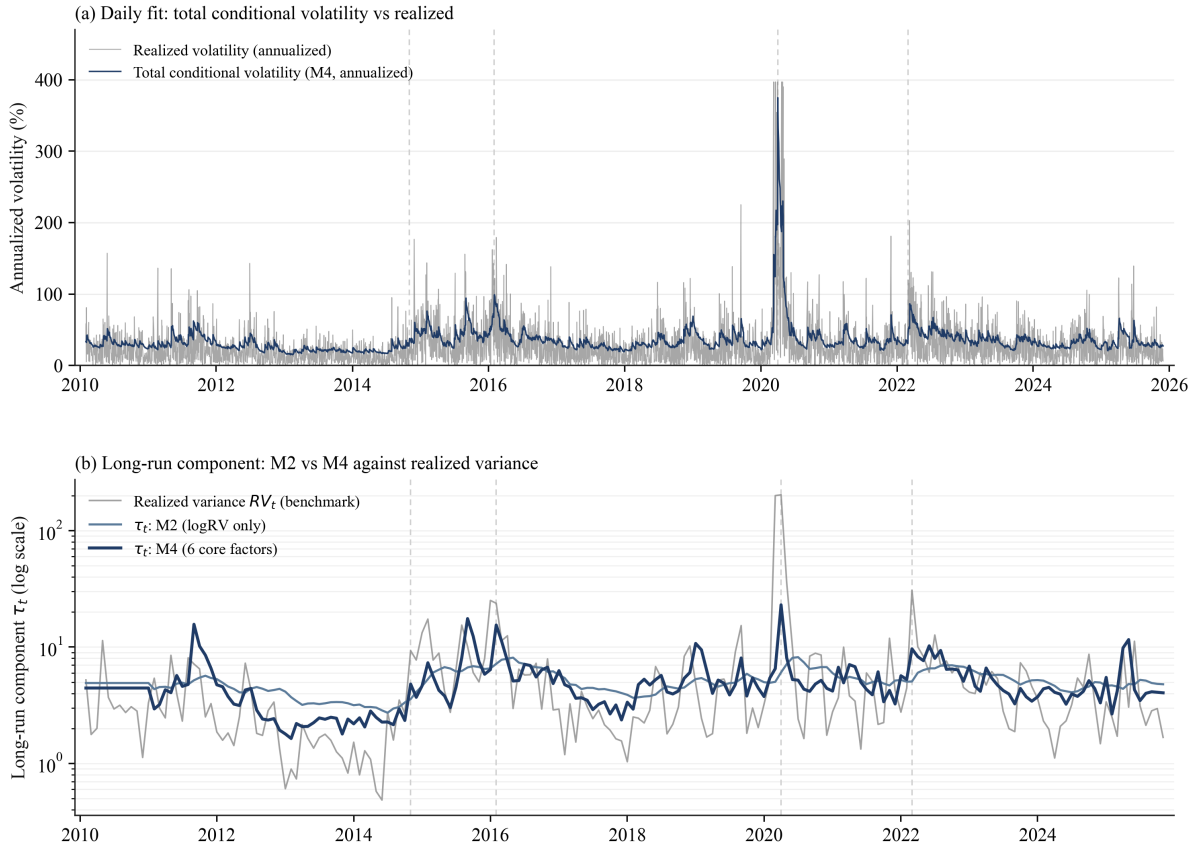



Figure 5: M4 total conditional volatility fit and M2/M4 long-run component comparison.
*Notes:* Panel (a) compares the annualized daily total conditional volatility from M4 (dark line) with the annualized daily realized volatility (light line), in percent. Panel (b) compares, on the log scale, the fits of M2 ($\log RV$ only) and M4 (six core factors) to the monthly realized variance $RV_t$.

Table 4 further reports the results of statistical tests. Panel A reports Diebold and Mariano (1995) tests with M4 as the reference model; positive DM statistics indicate that the comparison model has higher losses than M4:

$$\mathrm{DM} = \frac{\bar{d}}{\sqrt{\widehat{\mathrm{Var}}(\bar{d})/T_{\mathrm{oos}}}} \tag{14}$$

Under the QLIKE loss, M4 significantly outperforms M5 at both $h = 1$ and $h = 5$ at the 5% level. Panel B reports the Clark and West (2007) nested forecast test with M2 as the restricted benchmark model:

$$\mathrm{CW} = \frac{\bar{f}}{\sqrt{\widehat{\mathrm{Var}}(\bar{f})/T_{\mathrm{oos}}}} \tag{15}$$

where $f_t = (\sigma_t^2 - \hat{\sigma}^2_{\mathrm{M2},t})^2 - (\sigma_t^2 - \hat{\sigma}^2_{\mathrm{M4},t})^2 + (\hat{\sigma}^2_{\mathrm{M2},t} - \hat{\sigma}^2_{\mathrm{M4},t})^2$ is the Clark–West adjusted loss differential and $\bar{f}$ denotes its sample mean. M4 produces positive and significant CW statistics at all three horizons, so the six core factors bring out-of-sample predictive gains over M2. Panel C reports the Mincer and Zarnowitz (1969) regression for 22-day cumulative variance:

$$\sigma_t^{2,*} = a + b \cdot \hat{\sigma}_t^2 + u_t \tag{16}$$

where $\sigma_t^{2,*}$ is the realized cumulative variance over the next 22 trading days and $\hat{\sigma}_t^2$ is the model's 22-day-ahead cumulative variance forecast. Unbiased forecasts require $b = 1$. M4 attains the highest in-sample $R^2$ (0.460) and out-of-sample $R^2$ (0.449)—exceeding all alternatives—and its slope estimate $\hat{b} = 1.099$ is close to the theoretical value of one. M5's out-of-sample $R^2$ is $-0.287$, i.e., its predictive accuracy is worse than the naive "sample-mean predictor," and its slope $\hat{b} = 0.417$ severely underestimates the true variance level.

Table 4: Out-of-sample predictive ability tests.

| *Panel A: Diebold–Mariano test (QLIKE loss, M4 as reference)* | | | |
|---|---|---|---|
| Comparison | $h = 1$ | $h = 5$ | $h = 22$ |
| M4 vs M1 (GJR) | $-0.68$ | 0.63 | 1.73* |
| M4 vs M2 (GM-RV) | 1.19 | 0.97 | 1.51 |
| M4 vs M3 (GM-EMV) | 0.07 | 0.20 | 1.28 |
| M4 vs M5 (GM-21HQIC) | 2.47** | 2.26** | 1.76* |
| *Panel B: Clark–West nested forecast test (M2 as restricted)* | | | |
| Comparison | $h = 1$ | $h = 5$ | $h = 22$ |
| M3 vs M2 | 1.25 | 0.90 | 0.52 |
| M4 vs M2 | 1.68** | 1.56* | 1.32* |
| M5 vs M2 | 1.13 | 0.81 | 1.00 |
| *Panel C: Mincer–Zarnowitz regression ($h = 22$ cumulative)* | | | |
| Model | MZ $R^2$ | OOS $R^2$ | Slope $\hat{b}$ |
| M1 (GJR) | 0.386 | 0.378 | 0.879 |
| M2 (GM-RV) | 0.388 | 0.385 | 0.941 |
| M3 (GM-EMV) | 0.401 | 0.378 | 0.810 |
| M4 (GM-6Core) | **0.460** | **0.449** | 1.099 |
| M5 (GM-21HQIC) | 0.302 | $-0.287$ | 0.417 |

*Notes:* Panel A: DM statistics under QLIKE loss; positive values indicate the comparison model has higher losses than M4. Panel B: positive and significant values indicate out-of-sample predictive gains over M2. Panel C: MZ regressions for the 22-day cumulative variance. ***, **, * denote significance at the 1%, 5%, 10% levels.

### 5.3 Economic Value Evaluation: VaR Backtesting and Portfolio Analysis

To assess the practical value of each model in risk-management decisions, we apply four backtests to the 1-day-ahead VaR forecasts: the Kupiec (1995) unconditional coverage (UC) test, the Christoffersen (1998) independence (IND) and conditional coverage (CC) tests, and the Engle and Manganelli (2004) dynamic-quantile (DQ) test with 4 lags. Under conditional normality, the VaR is given by $\sqrt{\hat{\tau}_t \cdot \hat{g}_{i,t}} \times \Phi^{-1}(\alpha)$. The UC likelihood-ratio statistic is:

$$\mathrm{LR}_{\mathrm{UC}} = 2 \cdot \left[ N_v \cdot \log \frac{N_v}{T_{\mathrm{oos}} \cdot \alpha} + (T_{\mathrm{oos}} - N_v) \cdot \log \frac{T_{\mathrm{oos}} - N_v}{T_{\mathrm{oos}} \cdot (1 - \alpha)} \right] \tag{17}$$

Under $H_0 : E(N_v/T_{\mathrm{oos}}) = \alpha$, $\mathrm{LR}_{\mathrm{UC}}$ is asymptotically $\chi^2(1)$. Table 5, Panel A, reports the four backtest results for the 95% one-sided VaR. Under the UC test, all models except M5 (which is marginally rejected at the 10% level with a hit rate of 6.14% and $p = 0.056$) fail to reject the null at $p > 0.4$, indicating reasonable unconditional coverage. However, the independence (IND), conditional coverage (CC), and dynamic-quantile (DQ) tests reject the null at conventional sig-

nificance levels for M2–M5, suggesting that while the unconditional hit rates are well calibrated, the conditional normal VaR fails to capture the clustering of extreme returns. Appendix D reports Student-$t$ VaR and ES backtests that partially address this issue.

Panel B further evaluates the investment value of each model. We construct a volatility-targeted portfolio with an annualized target volatility of 15%, adjusting the leverage on a long crude oil position each day based on the 1-day-ahead conditional variance forecasts; the unleveraged portion earns the risk-free rate. Portfolio performance is summarized by the annualized Sharpe ratio and the Welch and Goyal (2008) Certainty-Equivalent Return (CER):

$$\mathrm{CER}(\gamma) = \bar{r} - 0.5\gamma \cdot \sigma_p^2, \quad \text{for risk-aversion coefficients } \gamma \in \{2, 5\} \tag{18}$$

Notably, the simpler GJR-GARCH benchmark (M1) attains the highest Sharpe ratio and CER in this portfolio exercise, despite its weaker statistical accuracy under loss-based metrics. This suggests that the economic value of volatility forecasts can be sensitive to the specific portfolio rule used, an issue we discuss further in Section 6.

Table 5: Economic value evaluation: VaR backtesting and volatility-targeted portfolio.

| *Panel A: 95% 1-day-ahead VaR backtesting ($T_{\text{oos}} = 1{,}434$)* | | | | | |
|---|---|---|---|---|---|
| Model | Hit rate (%) | UC $p$ | IND $p$ | CC $p$ | DQ $p$ |
| M1 (GJR) | 5.44 | 0.451 | 0.706 | 0.701 | 0.274 |
| M2 (GM-RV) | 5.02 | 0.971 | 0.011** | 0.038** | 0.053* |
| M3 (GM-EMV) | 5.02 | 0.971 | 0.011** | 0.038** | 0.048** |
| M4 (GM-6Core) | 5.44 | 0.451 | 0.010** | 0.028** | 0.036** |
| M5 (GM-21HQIC) | 6.14 | 0.056* | 0.056* | 0.026** | 0.005*** |
| *Panel B: Volatility-targeted portfolio (target annualized volatility 15%)* | | | | | |
| Model | Ann. ret. (%) | Ann. vol. (%) | SR | CER ($\gamma = 2$) | CER ($\gamma = 5$) |
| M1 (GJR) | +1.73 | 15.68 | +0.111 | −0.723 | −4.409 |
| M2 (GM-RV) | −0.03 | 15.88 | −0.002 | −2.549 | −6.330 |
| M3 (GM-EMV) | +0.92 | 15.53 | +0.059 | −1.496 | −5.115 |
| M4 (GM-6Core) | +0.52 | 16.01 | +0.033 | −2.041 | −5.887 |
| M5 (GM-21HQIC) | −0.04 | 16.34 | −0.003 | −2.712 | −6.717 |

*Notes:* Panel A: $p$-values for the four backtests of the 95% one-sided VaR. UC = Kupiec unconditional coverage; IND = Christoffersen independence; CC = joint conditional coverage; DQ = Engle–Manganelli dynamic-quantile test (with 4 lags). Panel B: Volatility-targeted strategy with a 15% annualized target. SR = annualized Sharpe ratio; CER = certainty-equivalent return. Bold values mark the column maximum.

### 5.4 Robustness Checks

To verify the robustness of our main conclusions, we conduct three robustness checks. First, we replace the default 5-fold expanding-window CV (with 14-month testing windows) with a 3-fold large-test-window split (25 months per fold, with 1,538 daily OOS predictions in total). At $h = 1$, M4 has QLIKE = 1.504 (default 1.465, deviation +2.7%); at $h = 5$, QLIKE = 0.480 (default 0.466, deviation +3.0%); and at $h = 22$, QLIKE = 0.522 (default 0.540, deviation −3.5%). The absolute deviations are all within the ±5% range, so M4's out-of-sample advantage holds up under this alternative split. In sharp contrast, M5's QLIKE deviations under the alternative split reach +5.8%, +13.5%, and +2.4%, respectively—at $h = 5$, the deviation exceeds the ±10%

range.

Second, we examine the fully out-of-sample factor re-selection robustness. Under the 5-fold expanding-window CV, we separately re-execute the "HQIC ∩ path-freq ≥ 85%" dual-threshold rule for each fold to obtain fold-specific BALANCED_CORE sets, and then generate OOS forecasts based on the fold-specific sets in the corresponding testing windows. Table 6 reports the results.

Table 6: Robustness checks.

| *Panel A: Alternative sample split (3 folds, $T_{\text{oos}} = 1{,}538$; QLIKE loss)* | | | |
|---|---|---|---|
| Model | $h = 1$ | $h = 5$ | $h = 22$ |
| M1 (GJR) | 1.474 (+2.4%) | 0.493 (+2.0%) | 0.614 (−5.1%) |
| M2 (GM-RV) | 1.532 (+2.6%) | 0.507 (+2.7%) | 0.626 (−5.0%) |
| M3 (GM-EMV) | 1.500 (+2.3%) | 0.480 (+1.7%) | 0.616 (−3.6%) |
| M4 (GM-6Core) | 1.504 (+2.7%) | 0.480 (+3.0%) | 0.522 (−3.5%) |
| M5 (GM-21HQIC) | 1.615 (+5.8%) | 0.603 (+13.5%) | 0.637 (+2.4%) |
| *Panel B: Fully OOS factor re-selection (Conditional vs Fully OOS, QLIKE loss, M4)* | | | |
| Horizon | Cond. OOS | Fully OOS | Δ (%) |
| $h = 1$ | 1.465 | 1.557 | +6.3% |
| $h = 5$ | 0.466 | 0.544 | +16.6% |
| $h = 22$ | 0.540 | 0.707 | +30.9% |

*Notes:* Panel A: QLIKE values under the alternative 3-fold large-test-window split (25 months per fold) together with deviations from the default 5-fold split (in parentheses). Panel B: Conditional OOS uses the six full-sample-selected BALANCED_CORE factors and re-estimates parameters in each fold's training sample; Fully OOS re-executes the dual-threshold selection rule on each fold to obtain fold-specific sets. Δ (%) = relative difference of Fully vs. Conditional; DM statistics with Conditional as reference are 2.28** ($h = 1$), 2.18** ($h = 5$), and 1.73* ($h = 22$). ***, **, * denote significance at the 1%, 5%, 10% levels.

Third, as a methodological robustness supplement, we conduct a sensitivity test on the ADMM outer-iteration count in the MCP algorithm. Under three configurations `max_outer` $\in \{20, 30, 50\}$, the number of nonzero factors at the HQIC-optimal $\lambda$ is $k = 21$, 21, and 23, respectively—the core factor identification result is essentially stable. Combining the evidence from Tables 3 through 6, the six BALANCED_CORE factors come out ahead of the alternative models on six fronts: full-sample information criteria, out-of-sample loss functions, pairwise predictive accuracy tests (Diebold–Mariano and Clark–West), Mincer–Zarnowitz unbiasedness, reasonable VaR coverage at the 95% one-sided level (comparable to the simpler benchmarks), and stability across alternative sample splits and factor re-selection schemes. These pieces of evidence point in the same direction and, taken together, support the dual-threshold factor identification rule.

# 6 Conclusion

This paper develops an "MCP penalty + path-level cross-validation frequency" dual-threshold factor identification rule within the GARCH-MIDAS framework, addressing two limitations of existing Lasso-based sparse GARCH-MIDAS studies: estimation bias from uniform shrinkage and cross-sample instability of single-criterion selection. The empirical analysis covers WTI

crude oil from February 2010 to November 2025. From a set of 53 candidate factors covering uncertainty and geopolitics, oil-market fundamentals, macro-finance, and market sentiment and stress, the dual-threshold rule retains six core drivers; the main model built on these factors outperforms the GARCH benchmark, the GM-RV baseline, and the dense HQIC benchmark on in-sample information criteria (BIC and HQIC), 5-fold CV factor robustness, and out-of-sample multi-loss evaluation, with the largest and most statistically robust gap relative to the dense HQIC benchmark. The Diebold–Mariano, Clark–West, Mincer–Zarnowitz, and Kupiec VaR backtests, the alternative-sample splits, and the fully out-of-sample factor re-selection check point in the same direction, and the six core factors raise the long-run component's share of the total conditional variance by about 28 percentage points relative to the baseline.

The findings carry several practical implications for risk management and policymaking in the crude oil market. The significantly negative effect of climate-policy and fiscal-policy uncertainty on long-run crude oil volatility suggests that market participants view these two categories of policies as structural signals about long-horizon fossil-fuel demand, providing a quantitative basis for assessing market responses to energy-transition policies. The strongest positive loading on the Commodity Markets EMV Tracker is consistent with a risk-contagion channel between commodity and energy markets, and monitoring volatility in the energy and commodity sectors of equity markets can serve as a forward-looking indicator of crude oil volatility. The set of core factors identified in this paper achieves reasonable coverage accuracy at the 95% one-sided VaR, and can serve as a low-frequency macroeconomic input for VaR models and stress-testing frameworks used by financial institutions.

This study has several limitations that can be addressed in future work. First, the analysis focuses only on WTI crude oil and does not cover other major crude benchmarks (Brent, Dubai) or related energy commodities (natural gas, coal); future work could extend the dual-threshold MCP framework to joint multi-commodity modeling. Second, while four of the six BALANCED_CORE factors maintain $f^{\text{path}} \geq 90\%$ under the rolling-window CV, EPU_Fiscal falls below the threshold (80%) and the World Sentiment Index shows considerably higher sensitivity (68%) to the choice of expanding-window versus rolling-window cross-validation; future work could explore alternative path-frequency thresholds or weighted-CV designs to mitigate this sensitivity. Third, the paper assumes conditional normality in the GARCH-MIDAS specification and does not consider time variation in factor loadings or nonlinear structures; in an energy market with frequent structural breaks, state-dependent parameters or machine-learning extensions (such as Random Forest and XGBoost-based interpretable machine learning, in the spirit of Kang et al. (2026) and Gupta et al. (2024)) may further enhance forecasting accuracy. Fourth, the portfolio analysis reveals that statistically optimal forecasts under M4 do not translate into optimal risk-adjusted returns; the underlying mechanism—e.g., how the leverage adjustment responds to extreme volatility—warrants further investigation using microstructure data.

## Statements and Declarations

**Funding.** This research is supported by the General Program of the National Social Science Fund of China (Grant No. 23BTJ047). The views expressed are those of the authors and do not necessarily reflect those of the funding agency.

**Competing Interests.** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Data Availability Statement.** Data can be made available on request.

**Ethics Approval.** This study is based on publicly available market and macroeconomic data and does not involve any human or animal subjects; no ethics approval was required.

**Consent for Publication.** All authors have read and approved the final version of the manuscript and consent to its submission and to its publication upon acceptance.

**Author Contributions.** Xiaogang Dong conceived the research idea. Hongzhen Jin performed experimentations and took the lead in writing the manuscript. Xiwen Qin supervised the findings and reviewed the manuscript. Liping Yuan and Shuo Li provided critical feedback and helped in framing the manuscript.

**Acknowledgments.** The authors thank the editor and the anonymous referees for their constructive comments; all remaining errors are the authors' own.

# A Complete List and Transformations of the 53 Candidate Macroeconomic Factors

Table 7 details the codes, descriptions, sources, and transformation methods of the 53 candidate macroeconomic factors, organized by the four categories. "$100\Delta\log$" indicates 100 times the consecutive monthly log difference; "$\Delta$" indicates first differencing; "Level" indicates the original level form. All transformed series are then winsorized at the 1st and 99th percentiles and standardized using the full-sample mean and standard deviation.

Table 7: Complete list of 53 candidate macroeconomic factors.

| Code | Description | Source | Trans. |
|---|---|---|---|
| *Category 1: Uncertainty and Geopolitics (14 factors)* | | | |
| BBD_MPU | Monetary Policy Uncertainty Index | Baker et al. (2016) | $100\Delta\log$ |
| EPU | Economic Policy Uncertainty (overall) | Baker et al. (2016) | $100\Delta\log$ |

*Continued on next page*

*Table 7 continued from previous page*

| Code | Description | Source | Trans. |
|---|---|---|---|
| EPU_Fiscal | EPU Fiscal Policy subcomponent | Baker et al. (2016) | 100Δ log |
| EPU_Trade | EPU Trade Policy subcomponent | Baker et al. (2016) | 100Δ log |
| CPU_narrow | Climate Policy Uncertainty (narrow) | Gavriilidis (2021) | Level |
| CPU_broad | Climate Policy Uncertainty (broad) | Gavriilidis (2021) | Level |
| ETU | Energy-related Uncertainty Index | Dang et al. (2023) | Level |
| WPUI | World Pandemic Uncertainty Index | Ahir et al. (2022) | Level |
| WTUI | World Trade Uncertainty Index | Ahir et al. (2022) | Level |
| WUI | World Uncertainty Index | Ahir et al. (2022) | Level |
| GPR | Geopolitical Risk Index | Caldara & Iacoviello (2022) | Level |
| GPRA | Geopolitical Acts subcomponent | Caldara & Iacoviello (2022) | Level |
| GPRT | Geopolitical Threats subcomponent | Caldara & Iacoviello (2022) | Level |
| EURQ_US | U.S. Economic Policy Uncertainty (quant.) | EIA / FRED | 100Δ log |
| *Category 2: Oil-market Fundamentals (15 factors)* | | | |
| PROD_US | U.S. monthly crude oil production | EIA | 100Δ log |
| IMP_US | U.S. monthly crude oil imports | EIA | 100Δ log |
| EXP_US | U.S. monthly crude oil exports | EIA | 100Δ log |
| REFINE_US | U.S. refinery capacity utilization | EIA | Δ |
| STOCK_COMM | U.S. commercial crude oil inventories | EIA | 100Δ log |
| STOCK_SPR | U.S. Strategic Petroleum Reserve | EIA | 100Δ log |
| IPG211S | IP of Oil and Gas Extraction | FRED IPG211S | 100Δ log |
| IPN213111S | IP of Drilling Oil and Gas Wells | FRED IPN213111S | 100Δ log |
| IPN2121S | IP of Coal Mining | FRED IPN2121S | 100Δ log |
| BRENT | Brent crude oil spot price | EIA / FRED | 100Δ log |
| GASREG | U.S. retail gasoline price | EIA | 100Δ log |
| HHNGAS | Henry Hub natural gas price | EIA / FRED | 100Δ log |
| NGAS_EU | European natural gas price | World Bank | 100Δ log |
| COAL_AU | Australian coal price | World Bank | 100Δ log |
| COPPER | International copper price (LME spot) | World Bank | 100Δ log |
| *Category 3: Macro-Finance (15 factors)* | | | |
| INDPRO | Industrial Production Index | FRED INDPRO | 100Δ log |
| TCU | Capacity Utilization (total industry) | FRED TCU | Level |
| NEWORDER | New Orders (non-defense capital goods) | FRED NEWORDER | 100Δ log |
| PAYEMS | Nonfarm Payrolls | FRED PAYEMS | 100Δ log |

*Continued on next page*

*Table 7 continued from previous page*

| Code | Description | Source | Trans. |
|---|---|---|---|
| UNRATE | Unemployment Rate | FRED UNRATE | $\Delta$ |
| PPIACO | PPI All Commodities | FRED PPIACO | $100\Delta \log$ |
| FF | Effective Federal Funds Rate | FRED FEDFUNDS | $\Delta$ |
| GS10 | 10-Year Treasury Yield | FRED GS10 | $\Delta$ |
| T10Y3M | 10Y minus 3M Term Spread | FRED T10Y3M | Level |
| BAA10Y | Baa minus 10Y Credit Spread | FRED BAA10YM | Level |
| M2 | M2 Money Supply | FRED M2SL | $100\Delta \log$ |
| FX_BIS | BIS Weighted Nominal Eff. Exch. Rate | BIS | $100\Delta \log$ |
| FX_BG | Real Broad Dollar Index (G&S) | FRED RTWEXBGS | $100\Delta \log$ |
| FX_AFE | Dollar vs Advanced Foreign Economies | FRED RTWEX-AFEGS | $100\Delta \log$ |
| HPI | U.S. House Price Index (Case-Shiller) | FRED CSUSH-PINSA | $100\Delta \log$ |
| *Category 4: Market Sentiment and Stress (9 factors)* | | | |
| EMV_Overall | Equity Market Volatility (overall) | Baker et al. (2026) | Level |
| EMV_Policy | EMV Policy subcomponent | Baker et al. (2026) | Level |
| EMV_Commodity | EMV Commodity Markets subcomponent | Baker et al. (2026) | Level |
| EMV_Disease | EMV Disease/Pandemic subcomponent | Baker et al. (2026) | Level |
| STLFSI | St. Louis Fed Financial Stress Index | FRED STLFSI4 | Level |
| KCFSI | Kansas City Fed Financial Stress Index | FRED KCFSI | Level |
| VIX | CBOE Volatility Index | FRED VIXCLS | $\Delta$ |
| CCI | Consumer Confidence Index (Michigan) | FRED UMCSENT | $\Delta$ |
| WSI | GDP-weighted World Sentiment Index | Ahir et al. (2022) | Level |

# B Descriptive Statistics and Stationarity Tests for All 57 Variables

Table 8 expands Table 1 of the main text, reporting descriptive statistics and stationarity tests for the WTI daily return, its absolute value, its square, the monthly $\log RV$, and all 53 standardized candidate factors. Monthly factors are standardized (mean zero, unit variance); WTI returns are in percent and kurtosis is the excess kurtosis. The $q$ columns report Benjamini–Hochberg-FDR-adjusted $p$-values for the Jarque–Bera (JB), Augmented Dickey–Fuller (ADF), and Ljung–Box (LB, 10 lags) tests, with the false discovery rate controlled at $\alpha = 0.05$. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 8: Descriptive statistics for all 57 variables (with BH–FDR adjustment).

| Variable | $N$ | Mean | Std. Dev. | Skew. | Kurt. | JB $q$ | ADF $q$ | LB $q$ |
|---|---|---|---|---|---|---|---|---|
| WTI daily return | 3,971 | 0.0009 | 2.6384 | −0.19 | 21.70 | 0.000*** | 0.000*** | 0.000*** |
| WTI \|return\| | 3,971 | 1.7025 | 2.0155 | 5.23 | 47.09 | 0.000*** | 0.000*** | 0.000*** |
| WTI return$^2$ | 3,971 | 6.9596 | 33.886 | 14.91 | 249.74 | 0.000*** | 0.000*** | 0.000*** |
| Monthly log $RV$ | 190 | 0.0056 | 0.9996 | 0.40 | 0.21 | 0.087* | 0.002*** | 0.000*** |
| EMV_Commodity | 190 | −0.0012 | 1.0025 | 2.67 | 9.60 | 0.000*** | 0.000*** | 0.000*** |
| IPN213111S | 190 | 0.0000 | 1.0000 | −2.03 | 6.39 | 0.000*** | 0.000*** | 0.000*** |
| CPU_narrow | 190 | 0.0000 | 1.0000 | 0.10 | 1.52 | 0.000*** | 0.000*** | 0.000*** |
| EPU_Fiscal | 190 | −0.0009 | 1.0026 | 2.12 | 5.07 | 0.000*** | 0.029** | 0.000*** |
| WSI | 190 | 0.0011 | 1.0025 | −1.44 | 3.85 | 0.000*** | 0.001*** | 0.000*** |
| FX_BG | 190 | 0.0000 | 1.0000 | 0.13 | −0.51 | 0.342 | 0.000*** | 0.000*** |
| BBD_MPU | 190 | 0.0000 | 1.0000 | 0.45 | 2.29 | 0.000*** | 0.000*** | 0.000*** |
| EPU | 190 | 0.0000 | 1.0000 | 0.90 | 3.93 | 0.000*** | 0.000*** | 0.538 |
| EPU_Trade | 190 | 0.0000 | 1.0000 | 0.57 | 9.90 | 0.000*** | 0.000*** | 0.040** |
| CPU_broad | 190 | 0.0000 | 1.0000 | 0.13 | 1.37 | 0.001*** | 0.000*** | 0.000*** |
| ETU | 190 | 0.0000 | 1.0000 | 0.40 | 2.85 | 0.000*** | 0.000*** | 0.000*** |
| WPUI | 190 | 0.0000 | 1.0000 | −0.46 | 2.91 | 0.000*** | 0.000*** | 0.001*** |
| WTUI | 190 | 0.0000 | 1.0000 | 0.25 | 5.25 | 0.000*** | 0.000*** | 0.018** |
| WUI | 190 | 0.0000 | 1.0000 | 0.07 | 0.51 | 0.393 | 0.000*** | 0.060* |
| GPR | 190 | 0.0000 | 1.0000 | 0.50 | 1.99 | 0.000*** | 0.000*** | 0.019** |
| GPRA | 190 | −0.0009 | 1.0026 | 1.16 | 1.38 | 0.000*** | 0.003*** | 0.000*** |
| GPRT | 190 | 0.0000 | 1.0000 | −0.05 | 2.35 | 0.000*** | 0.000*** | 0.038** |
| PROD_US | 190 | 0.0000 | 1.0000 | −0.28 | 2.44 | 0.000*** | 0.000*** | 0.110 |
| IMP_US | 190 | 0.0000 | 1.0000 | 0.02 | −0.13 | 0.929 | 0.000*** | 0.000*** |
| EXP_US | 190 | 0.0000 | 1.0000 | −0.11 | 2.32 | 0.000*** | 0.279 | 0.000*** |
| REFINE_US | 190 | 0.0000 | 1.0000 | −0.45 | −0.16 | 0.052* | 0.000*** | 0.000*** |
| STOCK_COMM | 190 | 0.0000 | 1.0000 | 0.07 | −0.40 | 0.526 | 0.137 | 0.000*** |
| STOCK_SPR | 190 | 0.0000 | 1.0000 | −2.84 | 8.89 | 0.000*** | 0.003*** | 0.000*** |
| IPG211S | 190 | 0.0000 | 1.0000 | −0.50 | 3.09 | 0.000*** | 0.000*** | 0.015** |
| IPN2121S | 190 | 0.0000 | 1.0000 | 0.12 | 0.37 | 0.513 | 0.000*** | 0.309 |
| BRENT | 190 | 0.0000 | 1.0000 | −0.53 | 0.88 | 0.001*** | 0.000*** | 0.006*** |
| GASREG | 190 | 0.0000 | 1.0000 | −0.23 | 0.50 | 0.202 | 0.000*** | 0.000*** |
| HHNGAS | 190 | 0.0000 | 1.0000 | −0.44 | 2.17 | 0.000*** | 0.000*** | 0.894 |
| NGAS_EU | 190 | 0.0000 | 1.0000 | 0.39 | 1.63 | 0.000*** | 0.078* | 0.002*** |
| COAL_AU | 190 | 0.0000 | 1.0000 | 0.19 | 1.75 | 0.000*** | 0.003*** | 0.000*** |
| COPPER | 190 | 0.0000 | 1.0000 | −0.26 | −0.09 | 0.395 | 0.002*** | 0.001*** |
| INDPRO | 190 | 0.0000 | 1.0000 | −0.71 | 7.57 | 0.000*** | 0.000*** | 0.093* |
| TCU | 190 | 0.0164 | 0.9767 | −1.31 | 3.25 | 0.000*** | 0.003*** | 0.000*** |
| NEWORDER | 190 | 0.0000 | 1.0000 | −0.14 | 0.72 | 0.121 | 0.004*** | 0.000*** |

*Continued on next page*

*Table 8 continued from previous page*

| Variable | $N$ | Mean | Std. Dev. | Skew. | Kurt. | JB $q$ | ADF $q$ | LB $q$ |
|---|---|---|---|---|---|---|---|---|
| PAYEMS | 190 | 0.0000 | 1.0000 | 3.12 | 13.43 | 0.000*** | 0.000*** | 0.000*** |
| UNRATE | 190 | 0.0000 | 1.0000 | −3.33 | 16.70 | 0.000*** | 0.000*** | 0.000*** |
| PPIACO | 190 | 0.0000 | 1.0000 | 0.10 | 0.47 | 0.406 | 0.000*** | 0.000*** |
| FF | 190 | 0.0000 | 1.0000 | 1.41 | 6.44 | 0.000*** | 0.029** | 0.000*** |
| GS10 | 190 | 0.0000 | 1.0000 | 0.10 | 0.45 | 0.423 | 0.000*** | 0.022** |
| T10Y3M | 190 | 0.0000 | 1.0000 | −0.01 | 0.53 | 0.393 | 0.000*** | 0.004*** |
| BAA10Y | 190 | 0.0000 | 1.0000 | 0.12 | 0.28 | 0.591 | 0.000*** | 0.153 |
| M2 | 190 | 0.0000 | 1.0000 | 2.68 | 12.93 | 0.000*** | 0.000*** | 0.000*** |
| FX_BIS | 190 | 0.0000 | 1.0000 | 0.10 | −0.31 | 0.591 | 0.000*** | 0.000*** |
| FX_AFE | 190 | 0.0000 | 1.0000 | −0.06 | −0.36 | 0.591 | 0.000*** | 0.004*** |
| HPI | 190 | 0.0000 | 1.0000 | 0.40 | 0.62 | 0.025** | 0.028** | 0.000*** |
| EMV_Overall | 190 | 0.0000 | 1.0000 | −0.22 | 4.06 | 0.000*** | 0.000*** | 0.002*** |
| EMV_Policy | 190 | −0.0007 | 1.0026 | 2.48 | 8.10 | 0.000*** | 0.000*** | 0.000*** |
| EMV_Disease | 190 | 0.0000 | 1.0000 | 0.86 | 8.37 | 0.000*** | 0.000*** | 0.001*** |
| STLFSI | 190 | −0.0005 | 1.0026 | 0.98 | 0.63 | 0.000*** | 0.001*** | 0.000*** |
| KCFSI | 190 | −0.0108 | 0.9914 | 1.04 | 0.84 | 0.000*** | 0.000*** | 0.000*** |
| VIX | 190 | −0.0022 | 1.0022 | 1.20 | 1.18 | 0.000*** | 0.000*** | 0.000*** |
| CCI | 190 | 0.0000 | 1.0000 | −0.29 | 0.93 | 0.013** | 0.000*** | 0.000*** |
| EURQ_US | 190 | 0.0000 | 1.0000 | 0.36 | 0.92 | 0.007*** | 0.000*** | 0.000*** |

# C Complete M5 (21-Factor) Estimation Results

Table 9 reports the full QMLE estimation results for the HQIC dense benchmark M5, including the short-run GARCH parameters, the MIDAS shape parameters, and the loadings together with Bollerslev–Wooldridge robust standard errors for all 21 factors selected at the HQIC optimum. The factor loadings are sorted in descending order of absolute MCP value at the HQIC-optimal $\lambda$. M5 has 29 free parameters, with NLL = 7,940.23 (the lowest among all models). However, its BIC = 16,030.73 and HQIC = 16,002.64 both exceed those of the main model M4 (16,007.81 and 15,994.26, respectively), reflecting that the in-sample fit improvement is insufficient to offset the model-complexity penalty. In M5, only EMV_Commodity is significant at the 1% level ($t = 4.30$) and EPU_Fiscal at the 5% level ($t = -2.47$); the other 19 factor loadings are not significant, consistent with the cross-sample instability of some HQIC-optimal factors documented in Section 5. Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 9: Complete M5 (GM-21HQIC) estimation results.

| Parameter | Estimate | Robust SE | $t$-stat | $p$-value | Sig. |
|---|---|---|---|---|---|
| $\mu$ | −0.0028 | 0.0298 | −0.09 | 0.9252 | |
| $\alpha$ | 0.0213 | 0.0152 | 1.40 | 0.1603 | |
| $\beta$ | 0.8752 | 0.0339 | 25.80 | 0.0000 | *** |
| $\gamma$ | 0.0905 | 0.0223 | 4.06 | 0.0000 | *** |
| $m$ | 1.4498 | 0.0814 | 17.82 | 0.0000 | *** |
| $\omega$ | 9.4659 | 4.7582 | 1.99 | 0.0467 | ** |
| $\omega_{RV}$ | 2.4985 | 0.0028 | 886.69 | 0.0000 | *** |
| $\beta_{RV}$ | 0.1863 | 0.1541 | 1.21 | 0.2266 | |
| $\theta_{\text{EMV_Commodity}}$ | 0.5882 | 0.1367 | 4.30 | 0.0000 | *** |
| $\theta_{\text{IPN213111S}}$ | 0.1188 | 0.0689 | 1.72 | 0.0847 | * |
| $\theta_{\text{CPU_narrow}}$ | −0.2061 | 0.1982 | −1.04 | 0.2985 | |
| $\theta_{\text{CPU_broad}}$ | −0.1746 | 0.2173 | −0.80 | 0.4218 | |
| $\theta_{\text{FX_AFE}}$ | 0.1159 | 0.1913 | 0.61 | 0.5445 | |
| $\theta_{\text{REFINE_US}}$ | −0.1049 | 0.0917 | −1.14 | 0.2524 | |
| $\theta_{\text{IPG211S}}$ | −0.1645 | 0.2293 | −0.72 | 0.4732 | |
| $\theta_{\text{EPU_Fiscal}}$ | −0.2161 | 0.0874 | −2.47 | 0.0134 | ** |
| $\theta_{\text{WSI}}$ | 0.1177 | 0.0814 | 1.45 | 0.1485 | |
| $\theta_{\text{FX_BG}}$ | 0.0108 | 0.1945 | 0.06 | 0.9558 | |
| $\theta_{\text{INDPRO}}$ | −0.1198 | 0.1364 | −0.88 | 0.3799 | |
| $\theta_{\text{M2}}$ | −0.1603 | 0.0953 | −1.68 | 0.0926 | * |
| $\theta_{\text{GPRT}}$ | −0.0882 | 0.1038 | −0.85 | 0.3957 | |
| $\theta_{\text{WPUI}}$ | −0.1693 | 0.1300 | −1.30 | 0.1931 | |
| $\theta_{\text{IMP_US}}$ | 0.1727 | 0.1033 | 1.67 | 0.0946 | * |
| $\theta_{\text{HHNGAS}}$ | 0.0769 | 0.0742 | 1.04 | 0.3004 | |
| $\theta_{\text{PROD_US}}$ | −0.0522 | 0.2448 | −0.21 | 0.8310 | |
| $\theta_{\text{COAL_AU}}$ | 0.1168 | 0.0926 | 1.26 | 0.2071 | |
| $\theta_{\text{PAYEMS}}$ | 0.1677 | 0.1164 | 1.44 | 0.1498 | |
| $\theta_{\text{BBD_MPU}}$ | 0.1551 | 0.1400 | 1.11 | 0.2678 | |
| $\theta_{\text{EMV_Disease}}$ | −0.0898 | 0.1241 | −0.72 | 0.4693 | |
| $k$ | 29 | — | — | — | |
| NLL | 7,940.23 | — | — | — | |
| AIC | 15,938.46 | — | — | — | |
| BIC | 16,030.73 | — | — | — | |
| HQIC | 16,002.64 | — | — | — | |

# D Supplementary 99% VaR and Student-$t$ VaR/ES Backtesting

Table 10 supplements two sets of results not displayed in Panel A of Table 5: (i) the Kupiec UC and the other three tests at the 99% level, and (ii) the $Z^2$ statistics for VaR and ES backtests at the 95% and 99% levels, replacing the conditional normal distribution with a Student-$t$ distribution (degrees of freedom df = 8.7, obtained from sample-residual moment estimation). The $Z^2$ statistic of McNeil and Frey (2000) is a joint unconditional-coverage plus tail-mean test that is asymptotically standard normal under the null. In Panel A, the null is $H_0 : E(N_v/T_{\text{oos}}) = 0.01$ for the 99% one-sided VaR; tests include the Kupiec UC, Christoffersen IND/CC, and Engle–Manganelli DQ. In Panels B–C, subscripts $N$ and $t$ refer to the conditional normal and Student-$t$ distributions, respectively, with $p$-values computed via stationary bootstrap ($B$ = 2,000). The results show that at the 99% level the actual hit rates of all six models significantly exceed the nominal 1% (1.88%–2.02%), and even with a Student-$t$ distribution replacing the conditional normal, the $Z^2$ $p$-values remain below 0.001 in all cases, reflecting that the tail behavior of crude oil daily returns is far heavier than what either the conditional normal or the Student-$t$ alone can capture—an observation that also appears in recent *Computational Economics* studies including Akgun and Gulay (2025). Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

Table 10: Supplementary 99% VaR and Student-$t$ VaR/ES backtesting.

| *Panel A: 99% 1-day-ahead VaR four-test results ($T_{\text{oos}}$ = 1,434)* | | | | | |
|---|---|---|---|---|---|
| Model | Hit rate (%) | UC $p$ | IND $p$ | CC $p$ | DQ $p$ |
| M1 (GJR) | 1.88 | 0.003*** | 0.103 | 0.003*** | 0.000*** |
| M2 (GM-RV) | 1.95 | 0.001*** | 0.002*** | 0.000*** | 0.000*** |
| M3 (GM-EMV) | 1.88 | 0.003*** | 0.001*** | 0.000*** | 0.000*** |
| M4 (GM-6Core) | 2.02 | 0.001*** | 0.000*** | 0.000*** | 0.000*** |
| M5 (GM-21HQIC) | 1.95 | 0.001*** | 0.000*** | 0.000*** | 0.000*** |
| *Panel B: Student-t (*df = 8.7*) VaR/ES $Z^2$ statistic (95% level)* | | | | | |
| Model | $\text{Hit}_N$ (%) | $\text{Hit}_t$ (%) | $Z_N^2$ | $Z_t^2$ | $p_t$ |
| M1 (GJR) | 5.44 | 5.65 | +2.18 | +2.11 | 0.000 |
| M2 (GM-RV) | 5.02 | 5.23 | +2.24 | +2.16 | 0.000 |
| M3 (GM-EMV) | 5.02 | 5.30 | +2.20 | +2.12 | 0.000 |
| M4 (GM-6Core) | 5.44 | 5.86 | +2.22 | +2.13 | 0.000 |
| M5 (GM-21HQIC) | 6.14 | 6.35 | +2.20 | +2.13 | 0.000 |
| *Panel C: Student-t (*df = 8.7*) VaR/ES $Z^2$ statistic (99% level)* | | | | | |
| Model | $\text{Hit}_N$ (%) | $\text{Hit}_t$ (%) | $Z_N^2$ | $Z_t^2$ | $p_t$ |
| M1 (GJR) | 1.88 | 1.26 | +2.25 | +2.23 | 0.000 |
| M2 (GM-RV) | 1.95 | 1.26 | +2.30 | +2.32 | 0.000 |
| M3 (GM-EMV) | 1.88 | 1.19 | +2.24 | +2.25 | 0.000 |
| M4 (GM-6Core) | 2.02 | 1.60 | +2.29 | +2.20 | 0.000 |
| M5 (GM-21HQIC) | 1.95 | 1.46 | +2.35 | +2.30 | 0.000 |

*Notes:* The Student-$t$ distribution is standardized to unit variance. For df = 8.7, the rescaled 5% critical value ($\approx -1.62$) is slightly less extreme than the standard normal one ($-1.645$), so $\text{Hit}_t$ slightly exceeds $\text{Hit}_N$ at the 95% level (Panel B); at the 99% level the rescaled critical value ($\approx -2.50$) is more extreme than the normal counterpart ($-2.326$), restoring $\text{Hit}_t < \text{Hit}_N$ (Panel C). ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

## References


Ahir, H., Bloom, N., and Furceri, D. (2022). The world uncertainty index. *NBER Working Paper*, (29763).

Akgun, O. B. and Gulay, E. (2025). Dynamics in realized volatility forecasting: Evaluating GARCH models and deep learning algorithms across parameter variations. *Computational Economics*, 65(6):3971–4013.

Amendola, A., Candila, V., and Gallo, G. M. (2021). Choosing the frequency of volatility components within the double asymmetric GARCH-MIDAS-X model. *Econometrics and Statistics*, 20:12–28.

Asgharian, H., Hou, A. J., and Javed, F. (2013). The importance of the macroeconomic variables in forecasting stock return variance: A GARCH-MIDAS approach. *Journal of Forecasting*, 32(7):600–612.

Bakas, D. and Triantafyllou, A. (2019). Volatility forecasting in commodity markets using macro uncertainty. *Energy Economics*, 81:79–94.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Baker, S. R., Bloom, N., Davis, S. J., and Kost, K. (2026). Policy news and stock market volatility. *Journal of Financial Economics*, 175:104187.

Bera, K., Raj, P., and Selvaraju, N. (2026). An integrated framework for volatility prediction: Leveraging decomposition techniques with realized GARCH models. *Computational Economics*, 67(4):3275–3309.

Bollerslev, T. and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2):143–172.

Bonnier, J.-B. (2022). Forecasting crude oil volatility with exogenous predictors: As good as it GETS? *Energy Economics*, 111:106059.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122.

Caldara, D. and Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review*, 112(4):1194–1225.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Chuang, O.-C. and Yang, C. W. S. (2022). Identifying the determinants of crude oil market volatility by the multivariate GARCH-MIDAS model. *Energies*, 15(8):2945.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

Conrad, C. and Kleen, O. (2020). Two are better than one: Volatility forecasting using multiplicative component GARCH-MIDAS models. *Journal of Applied Econometrics*, 35(1):19–45.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Dang, T. H.-N., Nguyen, C. P., Lee, G. S., Nguyen, B. Q., and Le, T. T. (2023). Measuring the energy-related uncertainty index. *Energy Economics*, 124:106817.

Diaz, E. M., Molero, J. C., and Perez de Gracia, F. (2016). Oil price volatility and stock returns in the G7 economies. *Energy Economics*, 54:417–430.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Engle, R. F., Ghysels, E., and Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *The Review of Economics and Statistics*, 95(3):776–797.

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Fang, T., Lee, T.-H., and Su, Z. (2020). Predicting the long-term stock market volatility: A GARCH-MIDAS model with variable selection. *Journal of Empirical Finance*, 58:36–49.

Gao, L., Hitzemann, S., Shaliastovich, I., and Xu, L. (2022). Oil volatility risk. *Journal of Financial Economics*, 144(2):456–491.

García-Medina, A. and Aguayo-Moreno, E. (2024). LSTM-GARCH hybrid model for the prediction of volatility in cryptocurrency portfolios. *Computational Economics*, 63(4):1511–1542.

Gavriilidis, K. (2021). Measuring climate policy uncertainty. *SSRN Working Paper*, (3847388).

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

Gupta, R., Karmakar, S., and Pierdzioch, C. (2024). Safe havens, machine learning, and the sources of geopolitical risk: A forecasting analysis using over a century of data. *Computational Economics*, 64(1):487–513.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B*, 41(2):190–195.

Herrera, A. M., Hu, L., and Pastor, D. (2018). Forecasting crude oil price volatility. *International Journal of Forecasting*, 34(4):622–635.

Kang, Y., Ryu, D., and Webb, R. I. (2026). Uncertainty indicators as key predictors of oil volatility: An interpretable machine learning approach. *Computational Economics*. Advance online publication.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2):73–84.

Léber, D. and Egyed, B. (2025). The sentiment augmented GARCH-LSTM hybrid model for value-at-risk forecasting. *Computational Economics*. Advance online publication.

Liang, C., Wei, Y., Li, X., Zhang, X., and Zhang, Y. (2020). Uncertainty and crude oil market volatility: New evidence. *Applied Economics*, 52(27):2945–2959.

Luo, J., Demirer, R., Gupta, R., and Ji, Q. (2022). Forecasting oil and gold volatilities with sentiment indicators under structural breaks. *Energy Economics*, 105:105751.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, 7(3-4):271–300.

Mincer, J. A. and Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, pages 3–46. NBER Books.

Nguyen, H. and Javed, F. (2023). Dynamic relationship between stock and bond returns: A GAS MIDAS copula approach. *Journal of Empirical Finance*, 73:272–292.

Pan, Z., Wang, Y., Wu, C., and Yin, L. (2017). Oil price volatility and macroeconomic fundamentals: A regime switching GARCH-MIDAS model. *Journal of Empirical Finance*, 43:130–142.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.

Qiao, G., Pan, Y., Liang, C., Wang, L., and Wang, J. (2024). Forecasting Chinese crude oil futures volatility: New evidence based on dual feature processing of large-scale variables. *Journal of Forecasting*, 43(7):2495–2521.

Salisu, A. A., Gupta, R., and Demirer, R. (2022). Global financial cycle and the predictability of oil market volatility: Evidence from a GARCH-MIDAS model. *Energy Economics*, 108:105934.

Sheng, X., Gupta, R., and Ji, Q. (2020). The impacts of structural oil shocks on macroeconomic uncertainty: Evidence from a large panel of 45 countries. *Energy Economics*, 91:104940.

Wei, Y., Liu, J., Lai, X., and Hu, Y. (2017). Which determinant is the most informative in forecasting crude oil market volatility: Fundamental, speculation, or uncertainty? *Energy Economics*, 68:141–150.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

Wen, D., He, M., Wang, Y., and Zhang, Y. (2024). Forecasting crude oil market volatility: A comprehensive look at uncertainty variables. *International Journal of Forecasting*, 40(3):1022–1041.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942.

Zhao, J. (2022). Exploring the influence of the main factors on the crude oil price volatility: An analysis based on GARCH-MIDAS model with Lasso approach. *Resources Policy*, 79:103031.

Zhao, P. and Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7:2541–2563.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.