# Middle East Geopolitical Risk and Monthly Systematic Tail Risk Identification for Time Series Momentum in Chinese Commodity Futures

Zihan Ye
Sino-German College of Intelligent Manufacturing, Shenzhen Technology University, Electronic Science and Technology, Shenzhen, Guangdong, 518118, China
zihanye687@gmail.com

**ABSTRACT**

Fluctuations in Middle East Geopolitical Risk (ME-GPR), a regional application of Geopolitical Risk (GPR), rarely remain confined to political news. They permeate commodity futures markets through oil supply expectations, shipping security, global risk-off sentiment, and dollar-based financial conditions. Leveraging 42 continuous contracts of Chinese commodity futures, this paper constructs two Time Series Momentum (TSMOM) strategies with $J=60$ and $J=250$, and derives ME-GPR, Middle East and North Africa Geopolitical Risk (MENA-GPR), change terms, and a positive 'Shock' variable from the Caldara-Iacoviello country-level GPR data. Here, 'Shock' denotes an upward month-to-month risk-change indicator rather than a separately identified event shock. The analysis does not target extreme single-day tail loss events. Instead, it identifies whether the strategy enters a systematic tail risk state over the following month; the model draws on risk information observable in the previous month to assess the next-month high-risk state. The empirical evaluation is conducted strictly at a monthly frequency.

Empirical diagnostics indicate that the monthly Logistic scoring architecture retains non-trivial predictive utility for isolating systematic tail states. The more stable predictive content, however, comes from market-risk benchmark variables, including the CBOE Volatility Index (VIX), the CBOE Crude Oil Volatility Index (OVX), Brent, the U.S. Dollar Index (DXY), commodity market volatility, and strategy-state variables. In the 2022-2025 main sample, the Market-risk benchmark Logistic model records an Area Under the Curve (AUC) of 0.701; in the 2021-2025 robustness sample, the AUC is 0.730. The inclusion of GPR variables does not yield a consistent statistical gain in out-of-sample predictive performance. Monthly dynamic risk control compresses risk exposure and maximum drawdown: in the main sample, the gross maximum drawdown narrows from -15.01% to -9.47%, and the 5 basis points (bp)-cost drawdown narrows from -17.17% to -10.16%. After transaction costs based on actual turnover are included, however, absolute-return gains remain weakly robust. Consequently, the architecture is better interpreted as a risk-budgeting tool than as an alpha-generating strategy.

**Keywords:** Middle East geopolitical risk; Time Series Momentum; commodity futures; systematic tail risk states; risk scoring; dynamic risk control

## 1. INTRODUCTION

The Middle East is closely tied to global oil supply, maritime transport routes, and energy security. When regional conflicts intensify, markets reprice the likelihood of supply contraction, shipping disruption, and higher risk premia; volatility in crude oil and related commodities then rises and is transmitted to Chinese commodity futures through dollar pricing, global risk appetite, and cross-market capital reallocation. For a trend-following Time Series Momentum (TSMOM) strategy, the key question is not whether a particular news item triggers an extreme loss on a single day, but whether the external risk state raises the probability of systematic left-tail losses over the next month.

The research target is consequently shifted from single-day tail risk event prediction to next-month systematic tail risk state identification. Geopolitical Risk (GPR) is a monthly variable, while daily tail events are sparse; attempting to forecast hyper-specific trading days primarily introduces severe statistical noise. This paper constructs a TSMOM portfolio from 42 continuous commodity futures contracts and, within a rolling out-of-sample (OOS) framework, examines whether market-risk variables, 'Global_ GPR', and Middle East Geopolitical Risk (ME-GPR) can identify next-month risk months.

The paper addresses three tasks. First, it relocates the object of analysis to monthly systematic risk states and traces the transmission channel from Middle East risk through energy, shipping, and risk appetite. Second, it compares the identification performance of Logistic, Random Forest (RF), and eXtreme Gradient Boosting (XGBoost) models. Third, under actual turnover and transaction-cost constraints, it tests whether monthly risk probabilities can be converted into risk budget adjustment signals. Dynamic risk control is used strictly as an empirical stress-test application and offers no guarantee of absolute returns.

# 2. LITERATURE REVIEW

## 2.1 Time Series Momentum and Commodity Futures Trend Strategies

Moskowitz et al. [1] document Time Series Momentum in futures markets, and Ming et al. [7] revisit this evidence in China's commodity futures market. Commodity futures have also long been discussed as strategic and tactical allocation assets, providing diversification benefits and exposure to macroeconomic and inflation-related risks, as shown by Erb and Harvey [5]. In this paper, however, commodity futures are not examined mainly as a return-enhancing allocation tool; instead, they provide the tradable cross-section through which TSMOM tail risk is identified. Table 1 summarizes the main literature streams and the incremental contribution of this paper.

## 2.2 Geopolitical Risk and Commodity Market Transmission

Caldara and Iacoviello [2] bring geopolitical risk into an empirical framework through the GPR index. Middle East risk may affect commodity volatility through energy, shipping, risk appetite, and dollar conditions, and Zheng et al. [8] provide related evidence for Chinese natural-resource commodity futures. Aloui et al. [9] further show that geopolitical risk affects commodity futures returns and downside risk through dynamic dependence and conditional value-at-risk channels, which is consistent with treating GPR as a risk-monitoring variable rather than a direct trading signal. Geopolitical risk can also affect spillovers among exchange, commodity, and equity markets, as documented by Hao et al. [10], implying that the effect of Middle East risk may be transmitted indirectly through dollar conditions and broader market-risk variables. This paper examines its risk-monitoring value only and does not make causal claims.

## 2.3 Tail Risk Identification, Machine Learning, and Dynamic Risk Control

Tail risk identification, risk scaling, and machine learning constitute the toolkit for trend-strategy risk management. Barroso and Santa-Clara [3] and Daniel and Moskowitz [4] show that momentum strategies may be exposed to crash-like states, while Liu et al. [6] emphasize asymmetry and tail risk in time series momentum strategies. This evidence suggests that left-tail losses cannot be fully understood from average return performance alone and supports the paper's focus on systematic tail risk states rather than on unconditional TSMOM profitability. Machine-learning and classification-evaluation studies, including Breiman [11], Chen and Guestrin [12], and Fawcett [13], provide tools for ranking and scoring risk states under sample imbalance, although complex models applied to monthly small samples remain prone to overfitting.

## 2.4 Literature Synthesis and Positioning of This Paper

Few existing studies directly examine whether Middle East GPR can identify monthly systematic tail risk in Chinese commodity futures trend strategies. This paper places GPR, market risk states, and TSMOM tail risk into a unified rolling forecast framework.

Table 1. Literature synthesis

| Literature stream | Representative studies | Main findings | Relation to this paper | Incremental contribution |
|---|---|---|---|---|
| Time Series Momentum | Moskowitz et al. [1]; Ming et al. [7]; Erb and Harvey [5] | Futures markets exhibit trend-following returns, but performance depends on market states. | Provides the TSMOM strategy framework. | Treats TSMOM as a risk identification object rather than a pure return strategy. |

Table 1. (continued) Literature synthesis

| Literature stream | Representative studies | Main findings | Relation to this paper | Incremental contribution |
|---|---|---|---|---|
| Momentum tail risk | Barroso and Santa-Clara [3]; Daniel and Moskowitz [4]; Liu et al. [6] | Momentum strategies may crash in extreme states, and risk scaling can reduce tail exposure. | Supports the risk budget adjustment logic. | Recasts tail events as monthly systematic risk months. |
| GPR and commodity markets | Caldara and Iacoviello [2]; Zheng et al. [8]; Aloui et al. [9]; Hao et al. [10] | Geopolitical risk affects energy, shipping, risk appetite, and commodity volatility. | Provides the basis for external risk variables. | Constructs Middle East GPR, MENA-GPR, and Shock variables and embeds them into the TSMOM risk framework. |
| Machine-learning risk forecasting | Breiman [11]; Chen and Guestrin [12]; Fawcett [13] | RF and XGBoost capture nonlinear ranking; ROC/AUC is suitable for classification evaluation. | Provides the basis for model comparison and evaluation metrics. | Emphasizes Logistic as the main model under monthly small samples, with nonlinear models as auxiliary checks. |

Taken together, these studies form a coherent chain for the present research design. Commodity futures provide a diversified and tradable asset universe [5], while the TSMOM literature shows that trend-following signals can be constructed across futures contracts [1,7]. The momentum-crash and tail-risk literature further suggests that the main risk of such strategies lies not only in average returns but also in asymmetric downside states [3,4,6]. At the same time, the GPR literature indicates that geopolitical shocks can be transmitted to commodity markets through energy prices, exchange-rate conditions, risk appetite, and cross-market spillovers [2,8,9,10]. Finally, machine-learning and classification-evaluation studies provide tools for ranking and scoring risk states under sample imbalance [11,13]. This paper connects these strands by asking whether Middle East GPR and market-risk variables can jointly identify next-month systematic tail risk in a Chinese commodity futures TSMOM strategy.

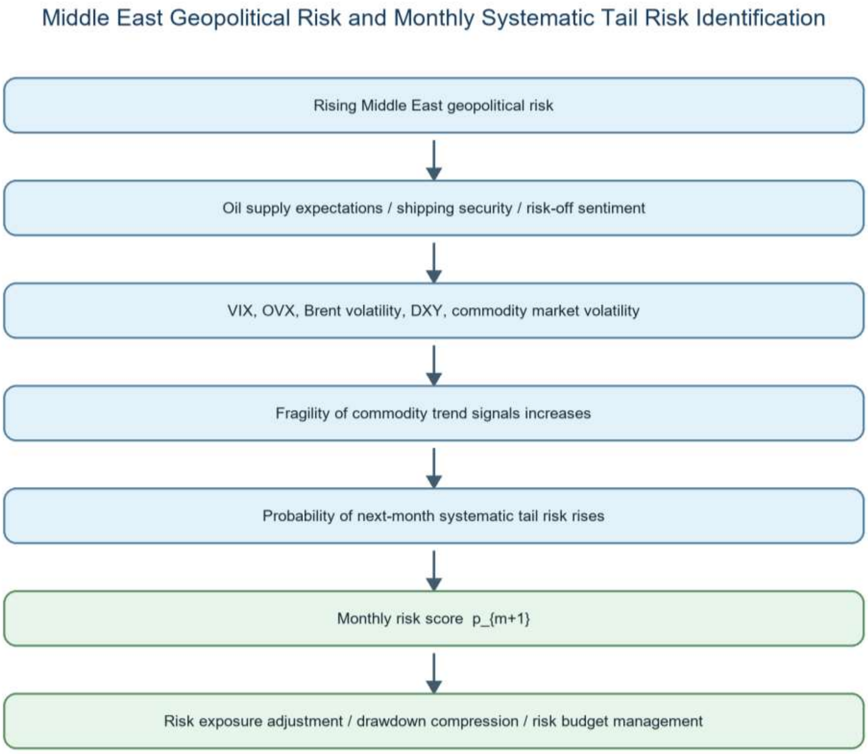
## 3. THEORETICAL MECHANISM AND RESEARCH HYPOTHESES

### 3.1 Theoretical Mechanism

The proposed transmission mechanism is illustrated in Figure 1. A rise in Middle East geopolitical risk first disturbs oil supply expectations, shipping security, and global risk-off sentiment; these disturbances are then reflected in the CBOE Volatility Index (VIX), the CBOE Crude Oil Volatility Index (OVX), Brent volatility, the U.S. Dollar Index (DXY), and commodity market volatility. Once market risk conditions deteriorate, commodity trend signals become fragile and prone to sudden reversals, and the probability that a TSMOM strategy enters a risk month increases. The risk-scoring model then produces $$p_{m+1}$$, which supports exposure reduction, drawdown compression, and risk budget management.

This mechanism should not be read as meaning that ‘ME_GPR’ directly causes strategy crashes. A more defensible interpretation is that ‘ME_GPR’ may be mapped into the tail risk state of trend strategies through market-risk variables. It is mainly an explanatory and monitoring variable, not an independent trading signal.

Figure 1. Mechanism linking Middle East geopolitical risk to monthly systematic tail risk identification




### 3.2 Research Hypotheses

- H1: Rising Middle East geopolitical risk affects commodity market risk states through energy, shipping, and risk-appetite channels.
- H2: Middle East GPR shocks may be mapped into commodity market risk states through energy, shipping, and risk-appetite channels, but their warning information may be absorbed or proxied by market variables such as VIX, OVX, Brent volatility, DXY, and commodity market volatility.
- H3: In monthly tail risk identification, a structurally simple and interpretable Logistic model is more suitable as the main risk-scoring tool than complex nonlinear models.
- H4: Monthly risk probabilities can serve as risk budget adjustment signals and help reduce strategy drawdowns and tail risk exposure.

## 4. RESEARCH DESIGN

### 4.1 Sample and Data

The main sample covers the period from March 26, 2018 to December 31, 2025, and includes 42 continuous contracts of Chinese commodity futures. Because contracts enter the market at different dates, the design adopts an unbalanced panel. Settlement prices are taken when available; otherwise, close prices substitute for them. The main variables and their processing methods are reported in Table 2. The GPR data come from the Caldara-Iacoviello country-level GPR file. ‘GPR’, ‘GPRC_EGY’, ‘GPRC_ISR’, ‘GPRC_SAU’, ‘GPRC_TUR’, and ‘GPRC_TUN’ enter the construction of ‘Global_ GPR’, ‘ME_GPR’, ‘MENA_GPR’, ‘dME_ GPR’, ‘dMENA_ GPR’, and ‘Shock’; the MENA term refers to Middle East and North Africa Geopolitical Risk (MENA-GPR), and ‘Shock’ denotes a positive month-to-month risk-change indicator. The current file does not contain ‘GPRC_IRN’ or ‘GPRC_IRQ’, so ‘ME_GPR’ should be interpreted as an available-country proxy for Middle East and neighboring-region risk rather than a complete coverage of all core conflict countries.

External market-risk variables include the CBOE Volatility Index (VIX), the CBOE Crude Oil Volatility Index (OVX), Brent crude oil prices, and the U.S. Dollar Index (DXY). VIX measures global risk appetite, OVX captures implied volatility in the oil market, Brent corresponds to the global crude pricing benchmark and Middle East supply shocks, and

DXY reflects global financial conditions. These variables are compiled from public market data and aggregated monthly before entering the model.

Table 2. Variables and data definitions

| Variable type | Variable name | Frequency | Processing | Use |
|---|---|---|---|---|
| GPR variable | Global_GPR | Monthly | Standardized within training window | Global geopolitical risk state |
| GPR variable | ME_GPR | Monthly | Equal-weighted average of standardized Egypt, Israel, Saudi Arabia, and Turkey GPR | Middle East risk level |
| GPR variable | MENA_GPR | Monthly | Adds Tunisia to ME_GPR | Regional robustness risk proxy |
| GPR variable | dME_GPR | Monthly | Monthly change in ME_GPR | Speed of risk escalation |
| GPR variable | Shock | Monthly | dME_ GPR exceeds the 90th percentile within the training window | Rapid risk escalation |
| Market risk | VIX | Daily to monthly | Monthly mean/end-of-month value | Global risk appetite |
| Market risk | OVX | Daily to monthly | Monthly mean/end-of-month value | Oil implied volatility |
| Market risk | Brent | Daily to monthly | Return and volatility | International oil price shock |
| Market risk | DXY | Daily to monthly | Monthly return | Global financial conditions |
| Strategy state | commodity_market_vol | Daily to monthly | Rolling volatility of commodity market returns | Internal market risk |
| Strategy state | strategy_ret_lag1 | Monthly | Lagged monthly strategy return | Momentum strategy state |
| Strategy state | strategy_vol | Daily to monthly | 20-day/60-day strategy volatility | Strategy risk level |
| Strategy state | strategy_mdd_lag1 | Monthly | Lagged monthly maximum drawdown | Recent drawdown pressure |

### 4.2 TSMOM Strategy

For each commodity, the cumulative return over the past $J$ trading days is computed on trading day $t$. A positive cumulative return generates a long position on the next trading day; a negative cumulative return generates a short position. The holding period is one day, and the portfolio is rebalanced daily. $J=60$ is the baseline specification, while $J=250$ serves as the robustness specification. Portfolio weights rely on inverse-volatility weighting from the standard deviation of daily returns over the previous 60 trading days and are normalized across all available commodities each day.

### 4.3 Monthly Systematic Tail Risk Labels

Three monthly risk labels are constructed. The one-day tail risk month label equals one if at least one $J=60$ daily tail risk day occurs in the target month. The two-day tail risk month label requires at least two such tail risk days in the target month. The monthly return 10% left-tail label equals one if the $J=60$ monthly strategy return in the target month falls below the 10th percentile of monthly returns estimated within the rolling training window. A daily tail risk day is defined as a $J=60$ raw strategy daily return below the 5th percentile of daily returns estimated within the rolling training window. All thresholds are estimated from the rolling training window; full-sample thresholds are avoided.

### 4.4 Time Alignment and Look-Ahead Bias Control

Time alignment follows a 'feature_ month=m' to 'target_ month=m+1' design. GPR variables, market-risk variables, and strategy-state variables all come from 'feature_ month' or earlier. GPR standardization parameters, the 'Shock' threshold, daily tail thresholds, monthly-return tail thresholds, and model parameters are re-estimated in each rolling training window; the test month is mapped using only training-window parameters to avoid look-ahead bias. 'Shock' equals one only when 'dME_ GPR' is positive and exceeds the 90th percentile of positive changes within the training window, thereby avoiding false signals caused by declines in geopolitical risk.

### 4.5 Rolling Out-of-Sample Prediction

For the main sample, the initial training period ends in December 2021, and the out-of-sample prediction period runs from January 2022 to December 2025. The robustness setting moves the initial training cutoff to December 2020 and takes January 2021 to December 2025 as the out-of-sample period. The long-sample robustness test draws on public continuous-contract data from 2012 to 2025, an unbalanced panel, an initial training period from January 2012 to December 2016, and an out-of-sample period from January 2017 to December 2025. Because early commodity coverage remains uneven, the long-sample results enter only as robustness evidence and do not replace the main sample.

### 4.6 Feature Sets and Models

The feature sets include Market-risk benchmark, Global GPR, Middle East GPR, GPR only, and Full features. The Market-risk benchmark incorporates VIX, OVX, 'Brent_ return', 'Brent_ vol', 'DXY_ return', 'commodity_ market_ vol', 'strategy_ret_lag1', 'strategy_vol_20d', 'strategy_vol_60d', 'strategy_mdd_lag1', and 'tail_ days_ in_ feature_ month'. Global GPR adds 'Global_ GPR' to this benchmark. Middle East GPR adds 'ME_GPR', 'dME_ GPR', and 'Shock'. GPR only comprises 'Global_ GPR', 'ME_GPR', 'MENA_GPR', 'dME_ GPR', 'dMENA_ GPR', and 'Shock'. Full features combine market-risk variables and GPR variables.

The models include a historical-frequency benchmark, Logistic, Random Forest (RF), and eXtreme Gradient Boosting (XGBoost). The historical-frequency benchmark takes the share of risk months in the training window as the predicted probability. Logistic is the main model in the text because the monthly sample is limited, its structure is simple, its output is interpretable, and its overfitting risk is relatively low. RF and XGBoost enter only as auxiliary evidence on nonlinear ranking information.

### 4.7 Evaluation Metrics

Evaluation metrics include Area Under the Curve (AUC), Average Precision (AP), Brier Score, Precision, Recall, and F1. AUC measures ranking ability; AP evaluates Precision-Recall performance under class imbalance; Brier Score measures probability calibration error; Precision, Recall, and F1 compare warning performance under the 0.3 and 0.5 probability thresholds.

### 4.8 Dynamic Risk Control and Transaction Costs

Monthly dynamic risk control predicts the 'target_ month' risk probability at the end of 'feature_ month' and sets the Exposure for the entire 'target_ month'. The Exposure mapping rule is reported in Table 3.

Table 3. Monthly risk probability and Exposure mapping

| Risk probability interval | Exposure |
|---|---|
| $p < 0.3$ | 1.0 |
| $0.3 \le p < 0.6$ | 0.7 |

Table 3. (continued) Monthly risk probability and Exposure mapping

| Risk probability interval | Exposure |
| --- | --- |
| $0.6 \le p < 0.8$ | 0.5 |
| $p \ge 0.8$ | 0.2 |

Transaction costs are calculated using changes in actual final target weights:

$$\text{final\_weight\_}\{i, t\} = \text{Exposure\_}\{\text{month}(t)\} \times \text{base\_TSMOM\_weight\_}\{i, t\}$$
$$\text{turnover\_}t = 0.5 \times \Sigma\_ i \, |\text{final\_weight\_}\{i, t\} - \text{final\_weight\_}\{i, t\text{-}1\}| \quad (1)$$

final_ weight_ {i, t} denotes the target weight of commodity i after applying the monthly risk-control exposure, and turnover_t denotes portfolio turnover generated by changes in final target weights. Cost specifications include gross, 2 basis points (bp), and 5bp. Dynamic risk control is evaluated only over the out-of-sample prediction period and does not mix in training-period returns.

## 5. EMPIRICAL RESULTS AND ANALYSIS

### 5.1 TSMOM Strategy and Sample Base

Table 4. Main sample and strategy specification diagnostics

| Item | Description |
| --- | --- |
| Main sample period | 2018-03-26 to 2025-12-31 |
| Commodities | 42 continuous contracts of Chinese commodity futures |
| Panel structure | Unbalanced panel; each commodity enters when available |
| Main strategy specification | J=60 TSMOM, with J=250 for robustness |
| Weighting method | Inverse 60-day volatility weighting, normalized daily |
| Transaction cost | Based on actual final-weight changes; gross, 2bp, and 5bp are reported |

Table 4 reports the main sample and strategy specification diagnostics. TSMOM is the object of tail risk analysis here, not a strategy whose profitability is assumed. The 42-commodity unbalanced panel and inverse-volatility weighting reduce sector concentration, while daily rebalancing also raises turnover. Cost results call for cautious interpretation.

### 5.2 Monthly Systematic Tail Risk Label Statistics

Table 5. Monthly systematic tail risk label statistics

| Label | Sample | OOS months | Risk months | Risk-month ratio |
| --- | --- | --- | --- | --- |
| One-day tail risk month | Main sample | 48 | 23 | 47.92% |
| Two-day tail risk month | Main sample | 48 | 10 | 20.83% |
| Monthly return 10% left-tail | Main sample | 48 | 2 | 4.17% |
| One-day tail risk month | Robustness | 60 | 34 | 56.67% |

Table 5. (continued) Monthly systematic tail risk label statistics

| Label | Sample | OOS months | Risk months | Risk-month ratio |
|---|---|---|---|---|
| Two-day tail risk month | Robustness | 60 | 17 | 28.33% |
| Monthly return 10% left-tail | Robustness | 60 | 5 | 8.33% |

Note: Because the monthly return 10% left-tail label has very few out-of-sample risk months, it serves only as a stricter supplementary tail definition and not as the main label.

Table 6. Rolling prediction sample-size statistics

| Sample | Label | Training start | Initial cutoff | OOS period | Initial training months | OOS months | OOS risk months |
|---|---|---|---|---|---|---|---|
| Main sample | One-day tail risk month | 2018-03 | 2021-12 | 2022-01-2025-12 | 46 | 48 | 23 |
| Main sample | Two-day tail risk month | 2018-03 | 2021-12 | 2022-01-2025-12 | 46 | 48 | 10 |
| Main sample | Monthly return 10% left-tail | 2018-03 | 2021-12 | 2022-01-2025-12 | 46 | 48 | 2 |
| Robustness | One-day tail risk month | 2018-03 | 2020-12 | 2021-01-2025-12 | 34 | 60 | 34 |
| Robustness | Two-day tail risk month | 2018-03 | 2020-12 | 2021-01-2025-12 | 34 | 60 | 17 |
| Robustness | Monthly return 10% left-tail | 2018-03 | 2020-12 | 2021-01-2025-12 | 34 | 60 | 5 |

Tables 5 and 6 report the monthly label statistics and rolling prediction sample-size design. In the main sample, the one-day tail risk month label appears 23 times across 48 out-of-sample months; in the robustness sample, it appears 34 times across 60 months. Relative to daily tail risk days, monthly labels mitigate sparsity and are better aligned with the monthly frequency of GPR variables.

### 5.3 Monthly Model Prediction Results

Table 7. Logistic feature-set comparison for one-day tail risk months

| Sample | Feature set | AUC | AP | Brier | Risk months/months | ΔAUC | ΔAP |
|---|---|---|---|---|---|---|---|
| Main sample | Market-risk benchmark | 0.701 | 0.668 | 0.225 | 23/48 | 0.000 | 0.000 |
| Main sample | Global GPR | 0.677 | 0.644 | 0.239 | 23/48 | -0.024 | -0.024 |
| Main sample | Middle East GPR | 0.563 | 0.550 | 0.295 | 23/48 | -0.137 | -0.118 |
| Main sample | Full features | 0.567 | 0.559 | 0.300 | 23/48 | -0.134 | -0.109 |
| Robustness | Market-risk benchmark | 0.730 | 0.780 | 0.214 | 34/60 | 0.000 | 0.000 |

Table 7. (continued) Logistic feature-set comparison for one-day tail risk months

| Sample | Feature set | AUC | AP | Brier | Risk months/months | ΔAUC | ΔAP |
|---|---|---|---|---|---|---|---|
| Robustness | Global GPR | 0.713 | 0.766 | 0.226 | 34/60 | -0.017 | -0.014 |
| Robustness | Middle East GPR | 0.650 | 0.714 | 0.263 | 34/60 | -0.079 | -0.066 |
| Robustness | Full features | 0.640 | 0.700 | 0.268 | 34/60 | -0.089 | -0.080 |

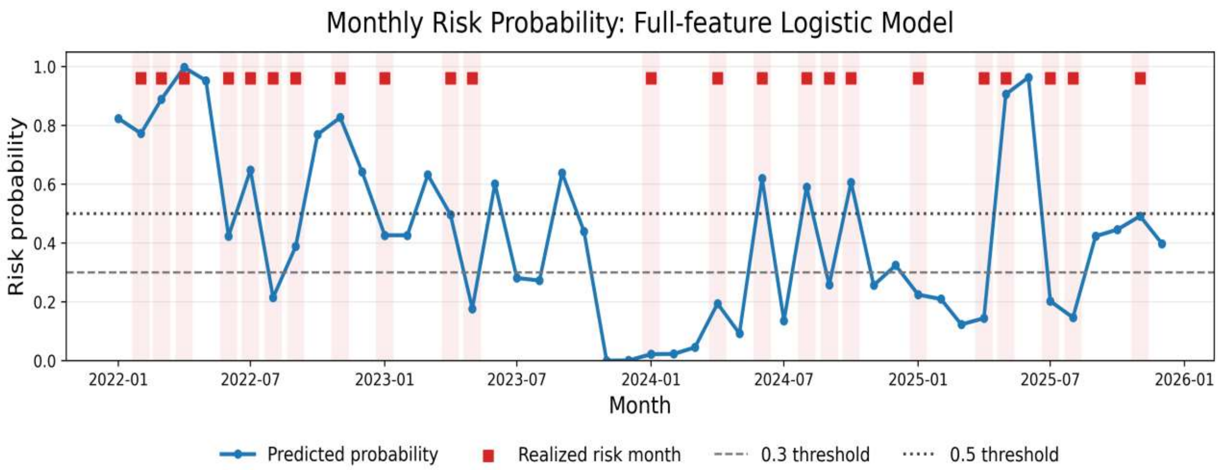
Note: The Market-risk benchmark is the most stable statistical forecasting benchmark; Full features provide a mechanism-oriented risk monitoring illustration with Middle East GPR included, and should not be interpreted as the statistically optimal model.

Table 7 reports the Logistic feature-set comparison for the one-day tail risk month label. The Market-risk benchmark is more stable in statistical prediction. The role of the Full features model is to demonstrate a mechanism-oriented risk monitoring framework after 'ME_GPR' is included. Figure 2 plots the monthly risk probability generated by the Full-feature Logistic model. Later reliance on this model is motivated by mechanism illustration rather than by superior predictive metrics over the Market-risk benchmark.

In the 2022-2025 main sample, the Market-risk benchmark Logistic model records an AUC of 0.701, AP of 0.668, and Brier Score of 0.225. The AUC values for Global GPR, Middle East GPR, and Full features are 0.677, 0.563, and 0.567, respectively. In the 2021-2025 robustness sample, the Market-risk benchmark Logistic model records an AUC of 0.730, AP of 0.780, and Brier Score of 0.214; the AUC values for Global GPR, Middle East GPR, and Full features are 0.713, 0.650, and 0.640, respectively.

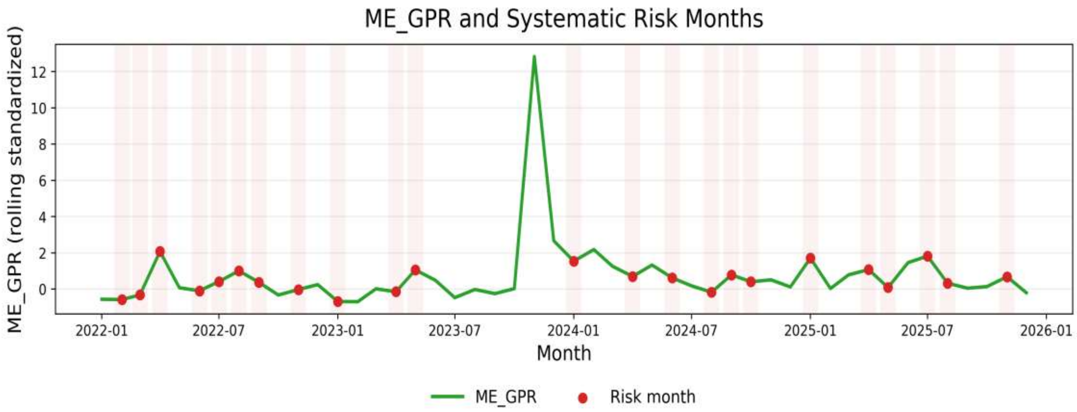
This does not imply that 'ME_GPR' is irrelevant. A more defensible reading is that, before geopolitical shocks enter Chinese commodity futures, part of their information has already been proxied by OVX, VIX, Brent volatility, DXY, and commodity market volatility. 'ME_GPR' is a low-frequency monthly macro indicator with lag, noise, and collinearity. When it enters the Logistic model together with high-frequency market-implied risk variables, out-of-sample performance may deteriorate. The empirical evidence points in that direction. GPR is more suitable for explaining risk sources and transmission channels than for replacing market variables as a trading signal.

Figure 2. Monthly risk probability from the Full-feature Logistic model




Note: Figure 2 serves exclusively as a conceptual diagnostic tool for mapping GPR-infused transmission channels, distinct from the statistically optimized baseline model.

Figure 3. Middle East GPR and systematic risk months




Note: Figure 3 compares ME-GPR with realized systematic risk months. Risk-month markers provide a time-series comparison and do not imply a univariate causal relationship.

### 5.4 Logistic and Nonlinear Model Comparison

Table 8. Main-sample model ranking performance

| Feature set | Model | AUC | AP | Brier |
|---|---|---|---|---|
| Market-risk benchmark | Historical frequency | 0.411 | 0.454 | 0.254 |
| Market-risk benchmark | Logistic | 0.706 | 0.674 | 0.226 |
| Market-risk benchmark | Random Forest | 0.597 | 0.568 | 0.251 |
| Market-risk benchmark | XGBoost | 0.520 | 0.503 | 0.279 |
| Full features | Historical frequency | 0.411 | 0.454 | 0.254 |
| Full features | Logistic | 0.577 | 0.579 | 0.281 |
| Full features | Random Forest | 0.557 | 0.519 | 0.263 |
| Full features | XGBoost | 0.517 | 0.500 | 0.288 |

Note: Random Forest and XGBoost enter only as auxiliary evidence on nonlinear ranking information.

Table 9. Main-sample warning-threshold performance

| Feature set | Model | Precision@0.3 | Recall@0.3 | F1@0.3 | Precision@0.5 | Recall@0.5 | F1@0.5 |
|---|---|---|---|---|---|---|---|
| Market-risk benchmark | Historical frequency | 0.479 | 1.000 | 0.648 | 0.467 | 0.913 | 0.618 |
| Market-risk benchmark | Logistic | 0.579 | 0.957 | 0.721 | 0.591 | 0.565 | 0.578 |
| Market-risk benchmark | Random Forest | 0.465 | 0.870 | 0.606 | 0.650 | 0.565 | 0.605 |

Table 9. (continued) Main-sample warning-threshold performance

| Feature set | Model | Precision@0.3 | Recall@0.3 | F1@0.3 | Precision@0.5 | Recall@0.5 | F1@0.5 |
|---|---|---|---|---|---|---|---|
| Market-risk benchmark | XGBoost | 0.459 | 0.739 | 0.567 | 0.500 | 0.478 | 0.489 |
| Full features | Historical frequency | 0.479 | 1.000 | 0.648 | 0.467 | 0.913 | 0.618 |
| Full features | Logistic | 0.469 | 0.652 | 0.545 | 0.588 | 0.435 | 0.500 |
| Full features | Random Forest | 0.488 | 0.870 | 0.625 | 0.529 | 0.391 | 0.450 |
| Full features | XGBoost | 0.500 | 0.739 | 0.596 | 0.533 | 0.348 | 0.421 |

Note: The high Recall of the historical-frequency benchmark mainly comes from a wide warning range and does not indicate effective forecasting ability.

Tables 8 and 9 report the main-sample model ranking and warning-threshold performance. Random Forest and XGBoost can deliver higher recall in selected labels or samples, but their performance is unstable. Monthly observations are limited, and tail labels are imbalanced, so complex nonlinear models are vulnerable to sample-specific noise. The Logistic model remains more transparent and stable for risk scoring.

Table 10. Rolling coefficient directions of core Logistic variables in the main sample

| Feature ID | Theoretical sign | Market benchmark mean | Market benchmark positive frequency | Full-feature mean | Full-feature positive frequency |
|---|---|---|---|---|---|
| VIX_mean | + | 0.268 | 100.0% | 0.115 | 83.3% |
| OVX_mean | + | -0.168 | 27.1% | -0.076 | 29.2% |
| Brent_vol | + | 0.302 | 100.0% | 0.304 | 97.9% |
| commodity_market_vol | + | 0.675 | 100.0% | 0.621 | 100.0% |
| strategy_vol_20d | + | 0.832 | 100.0% | 0.853 | 100.0% |
| strategy_mdd_lag1 | + | -0.580 | 0.0% | -0.455 | 0.0% |
| tail_days_in_feature_month | Ambiguous | -0.832 | 0.0% | -0.799 | 0.0% |
| Global_GPR | + | | | 0.157 | 83.3% |
| ME_GPR | + | | | -0.278 | 0.0% |
| MENA_GPR | + | | | -0.718 | 0.0% |
| Shock | + | | | 0.414 | 100.0% |

Note: The table summarizes coefficient directions from rolling-window Logistic models. A single static p-value is not reported. Directional results help interpret the risk-scoring mechanism and do not constitute causal identification.

Table 10 summarizes the rolling coefficient directions of core Logistic variables in the main sample. Coefficient directions also clarify the mechanism. ‘VIX_ mean’, ‘Brent_ vol’, ‘commodity_ market_ vol’, and ‘strategy_vol_20d’ are positive in most rolling windows, indicating that global risk appetite, oil volatility, and the strategy’s own volatility state are broadly associated with higher risk probabilities. 'ME_GPR' and 'Shock' are less stable in sign. This is consistent with the evidence that Middle East GPR is better interpreted as a risk-state variable and monitoring input than as a robust standalone predictor.

### 5.5 Dynamic Risk Control and Transaction Costs

Table 11 reports the monthly dynamic risk-control results, the constant-exposure benchmark, and transaction-cost specifications. Dynamic risk control is not designed to search for the highest-return strategy. It stress-tests whether monthly risk probabilities can be mapped into risk budgets. To avoid conflating general exposure reduction with active timing ability, this paper adds a constant-exposure benchmark: the raw TSMOM daily returns and target weights are multiplied by the average out-of-sample Exposure of the monthly risk-control strategy, and transaction costs are deducted using the same daily final-weight-difference method. The average Exposure reported in the dynamic risk-control table is calculated from the same out-of-sample Exposure sequence used in the Exposure distribution table, ensuring consistency between performance evaluation and risk-budget allocation.

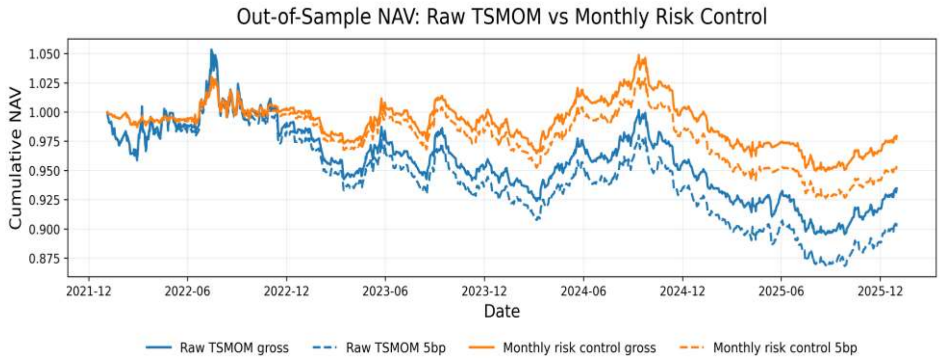
Table 11. Monthly dynamic risk control, constant-exposure benchmark, and transaction costs

| Sample | Strategy | Cost | Annual return | Annual volatility | Maximum drawdown | Average Exposure | Annual turnover |
|---|---|---|---|---|---|---|---|
| Main sample | Raw TSMOM | gross | -1.74% | 5.16% | -15.01% | 1.000 | 16.88 |
| Main sample | Raw TSMOM | 5bp | -2.57% | 5.16% | -17.17% | 1.000 | 16.88 |
| Main sample | Constant exposure | gross | -1.23% | 3.71% | -10.98% | 0.725 | 12.13 |
| Main sample | Constant exposure | 5bp | -1.83% | 3.71% | -12.61% | 0.725 | 12.13 |
| Main sample | Monthly risk control | gross | -0.54% | 3.66% | -9.47% | 0.725 | 13.77 |
| Main sample | Monthly risk control | 5bp | -1.22% | 3.66% | -10.16% | 0.725 | 13.77 |
| Robustness | Raw TSMOM | gross | -0.67% | 5.78% | -17.89% | 1.000 | 17.16 |
| Robustness | Raw TSMOM | 5bp | -1.52% | 5.79% | -20.54% | 1.000 | 17.16 |
| Robustness | Constant exposure | gross | -0.41% | 3.84% | -12.17% | 0.680 | 11.40 |
| Robustness | Constant exposure | 5bp | -0.97% | 3.84% | -14.06% | 0.680 | 11.40 |
| Robustness | Monthly risk control | gross | 1.00% | 3.66% | -9.47% | 0.680 | 12.93 |
| Robustness | Monthly risk control | 5bp | 0.35% | 3.66% | -10.16% | 0.680 | 12.93 |

In the main sample, raw TSMOM has an annualized return of -1.74% and a maximum drawdown of -15.01% under the gross specification. The constant-exposure benchmark, at the same average Exposure, has an annualized return of -1.23% and a maximum drawdown of -10.98%. Monthly risk control further changes these values to -0.54% and -9.47%. Under the 5bp cost specification, the constant-exposure benchmark records an annualized return of -1.83% and a maximum drawdown of -12.61%, while monthly risk control records -1.22% and -10.16%. Annualized turnover is daily average turnover multiplied by 252. These results suggest that the monthly warning model not only reduces the overall risk budget but also retains some timing value in tail risk exposure reduction. Returns remain negative, so the exercise should not be presented as a standalone return-generating trading strategy.

In the robustness sample, the constant-exposure benchmark records an annualized return of -0.97% and a maximum drawdown of -14.06% after 5bp costs; monthly risk control records 0.35% and -10.16%. The risk probability contains information for exposure-reduction timing. This is supplementary evidence only and should not be elevated into a claim of persistent profitability.

Figure 4. Out-of-sample Net Asset Value (NAV) comparison: raw TSMOM and monthly risk control




Note: Figure 4 compares the out-of-sample Net Asset Value (NAV) paths of raw TSMOM and monthly risk control. NAV is calculated only over the out-of-sample prediction period.

### 5.6 Exposure Distribution

Table 12. Exposure distribution under monthly risk control

| Sample | Average Exposure | Minimum | Maximum | Exposure=1 | Exposure=0.7 | Exposure=0.5 | Exposure=0.2 |
|---|---|---|---|---|---|---|---|
| Main sample | 0.725 | 0.2 | 1.0 | 16 | 23 | 3 | 6 |
| Robustness | 0.680 | 0.2 | 1.0 | 18 | 24 | 8 | 10 |

Table 12 reports the Exposure distribution under monthly risk control. In the main sample, the average Exposure is 0.725, and the numbers of months with Exposure=1, 0.7, 0.5, and 0.2 are 16, 23, 3, and 6, respectively. In the robustness sample, the average Exposure is 0.680, with corresponding counts of 18, 24, 8, and 10. The model reduces exposure in some high-risk months, and the predicted probability can be converted into a risk budget signal.

### 5.7 Long-Sample Robustness Test

The long-sample test extends the historical window and re-examines the cross-cycle applicability of the monthly risk identification framework in an unbalanced panel with gradually expanding commodity coverage. It does not replace the 2018-2025 main sample. Table 13 reports the long-sample discontinuity diagnosis. The results should be read as a robustness check rather than as the main inference.

The diagnosis shows that the decline in the number of valid commodities in 2014 is not caused by a broad absence of raw price data, but mainly arises from the 60-day volatility and valid strategy weight construction steps. The sharp decline in valid commodities mainly occurs in the 60-day volatility and valid strategy weight construction steps, with many trading days in that year falling below the 'min_active_assets=10' threshold. The long-sample result is accordingly treated only as directional weak robustness evidence and does not replace the 2018-2025 main sample.

Table 13. Long-sample discontinuity diagnosis

| Year | Raw price mean | Return-available mean | 60-day volatility-available mean | Strategy-valid mean | Days below 10 commodities |
|---|---|---|---|---|---|
| 2012 | 20.75 | 20.65 | 15.11 | 15.11 | 61 |
| 2013 | 23.65 | 23.30 | 15.68 | 15.68 | 72 |
| 2014 | 27.12 | 26.52 | 7.90 | 7.90 | 163 |
| 2015 | 30.88 | 30.45 | 10.77 | 10.77 | 125 |
| 2016 | 31.73 | 31.53 | 28.25 | 28.25 | 0 |

Table 14. Long-sample Logistic feature-set comparison for one-day tail risk months

| Sample | Feature set | AUC | AP | Brier | Risk months/months | ΔAUC | ΔAP |
|---|---|---|---|---|---|---|---|
| Long sample | Market-risk benchmark | 0.722 | 0.653 | 0.213 | 41/104 | 0.000 | 0.000 |
| Long sample | Global GPR | 0.707 | 0.645 | 0.216 | 41/104 | -0.015 | -0.008 |
| Long sample | Middle East GPR | 0.725 | 0.655 | 0.215 | 41/104 | 0.003 | 0.002 |
| Long sample | GPR only | 0.572 | 0.448 | 0.242 | 41/104 | -0.150 | -0.205 |
| Long sample | Full features | 0.717 | 0.638 | 0.216 | 41/104 | -0.005 | -0.015 |

Table 14 reports the long-sample Logistic feature-set comparison for the one-day tail risk month label. The long-sample results are broadly consistent with the main sample. The Market-risk benchmark Logistic model records an AUC of 0.722, AP of 0.653, and Brier Score of 0.213. The AUC values for Middle East GPR and Full features are 0.725 and 0.717, respectively; GPR only has an AUC of 0.572. Middle East GPR is slightly higher than the Market-risk benchmark in the long sample, but the margin is very small, and GPR only remains clearly weaker. This cannot support the claim that GPR is consistently superior to market variables.

Table 15. Long-sample dynamic risk control and transaction costs

| Strategy | Cost | Annual return | Annual volatility | Maximum drawdown | Average Exposure | Annual turnover |
|---|---|---|---|---|---|---|
| Raw TSMOM | gross | 1.50% | 7.73% | -15.91% | 1.000 | 15.30 |
| Raw TSMOM | 5bp | 0.72% | 7.73% | -17.62% | 1.000 | 15.30 |
| Monthly risk control | gross | 1.27% | 5.87% | -14.34% | 0.663 | 11.63 |
| Monthly risk control | 5bp | 0.68% | 5.87% | -15.03% | 0.663 | 11.63 |

Table 15 reports the long-sample dynamic risk-control and transaction-cost results. In the long sample, raw TSMOM records an annualized return of 1.50% and a maximum drawdown of -15.91% under the gross specification; monthly risk control records 1.27% and -14.34%. Under the 5bp cost specification, raw TSMOM records an annualized return of 0.72% and a maximum drawdown of -17.62%, while monthly risk control records 0.68% and -15.03%. The long-sample result again points to drawdown compression rather than robust return enhancement.

### 5.8 Chapter Summary

The evidence in this chapter points to the same conclusion: market-risk variables identify next-month risk months relatively stably; Middle East GPR has mechanism-explanation and risk-monitoring value, but its incremental predictive ability is unstable after market-risk variables are controlled for. Monthly risk control can reduce risk exposure and maximum drawdown, while return improvement is sensitive to transaction costs.

# 6. DISCUSSION

## 6.1 Prediction Frequency and Market Information Absorption

This study sets the prediction frequency at the monthly level not only to match the release cycle of macro GPR indicators, but also to reflect the slow-moving nature of cross-market transmission in geopolitical risk. GPR shocks usually permeate commodity markets gradually through energy prices, shipping risk, risk-off sentiment, and capital allocation, whereas VIX, OVX, Brent, DXY, and commodity market volatility absorb this information more quickly. Market prices are updated by actual trading every day; GPR is a monthly macro indicator and moves more slowly by construction. The stronger statistical performance of the Market-only model does not negate the mechanism value of GPR; it indicates that the incremental predictive information in GPR may be partly substituted by market variables.

## 6.2 Model Choice Under Small-Sample Constraints

Monthly observations are limited and tail risk labels are imbalanced, making complex nonlinear models prone to overfitting. Logistic is structurally simple, interpretable, and relatively stable in probability output, making it more suitable as a risk-scoring tool. RF and XGBoost are useful auxiliary checks but should not serve as the basis for the main conclusion.

## 6.3 Boundary Between Risk Identification and Trading Returns

The simplified TSMOM strategy is rebalanced daily, and transaction costs erode returns. In the turnover calculation, $t$ and $t-1$ refer to adjacent trading days, and costs are based on daily changes in final target weights rather than month-end snapshots. The constant-exposure benchmark demonstrates that part of the drawdown compression indeed comes from overall risk budget reduction. Yet in both the main and robustness samples, monthly risk control still reduces drawdown relative to the constant-exposure benchmark, indicating that risk probabilities retain some information for exposure-reduction timing. The main-sample return after 5bp costs remains negative. The paper does not prove persistent profitability; it demonstrates that risk signals can help reduce exposure and drawdown.

## 6.4 Methodological Value and Practical Implications

From an application perspective, the framework provides risk-management input rather than direct buy-sell signals. Commodity Trading Advisor (CTA) strategies may incorporate monthly risk scores into stage-based exposure reduction rules. Commodity allocation should not interpret a single GPR indicator in isolation, but should monitor states in which GPR deteriorates jointly with VIX, OVX, Brent, DXY, and commodity volatility. Future research should also combine macro risk variables with market-state variables instead of treating macro indicators as direct trading signals.

## 6.5 Limitations

The current design has clear limitations. GPR is monthly, and the available country-level data do not cover Iran or Iraq. The monthly sample size is limited, which is why Logistic is used as the main model and complex nonlinear models are not treated as core evidence. Although the long sample extends the time span, the volatility window and valid-commodity threshold create representativeness breaks. Given these limitations, the results should be understood as an initial test of a risk identification framework rather than strict causal identification.

# 7. CONCLUSIONS AND FUTURE RESEARCH

## 7.1 Conclusions

This paper shifts the research object from single-day tail risk events to monthly systematic tail risk states. It constructs TSMOM strategies from 42 continuous contracts of Chinese commodity futures and evaluates the identification ability of market-risk variables, Middle East GPR, and model-based risk scores within a rolling out-of-sample framework. Under the current sample and variable settings, the Market-risk benchmark Logistic model performs relatively stably. By contrast, the incremental predictive ability of Middle East GPR is unstable, although it remains useful for explaining risk sources and market transmission channels.

Monthly dynamic risk control reduces exposure and drawdown, and it also retains some timing value relative to the constant-exposure benchmark. After transaction costs are deducted, however, it should not be interpreted as a trading scheme with persistent profitability. The 2012-2025 long-sample results provide only directional weak robustness evidence: the risk identification framework has some extensibility, but the unbalanced panel and valid-commodity discontinuities prevent it from replacing the main-sample conclusion. The contribution of this paper is consequently better understood as a methodological framework for 'Middle East geopolitical risk-commodity futures trend strategies-monthly systematic tail risk identification,' rather than a directly validated stable-return trading scheme.

### 7.2 Future Research

Future research can proceed along two lines: more granular data and more realistic portfolio implementation. On the data side, higher-frequency Middle East event data, shipping congestion measures, tanker freight rates, and news-text indicators could help compensate for the limited short-horizon information in monthly GPR variables. On the portfolio side, monthly risk scores can be embedded into more complete CTA portfolios while incorporating margin, slippage, and liquidity constraints, which would bring the analysis closer to real investment management.

## REFERENCES

[1] Moskowitz T J, Ooi Y H, Pedersen L H. Time series momentum[J]. Journal of Financial Economics, 2012, 104(2): 228-250.
[2] Caldara D, Iacoviello M. Measuring geopolitical risk[J]. American Economic Review, 2022, 112(4): 1194-1225.
[3] Barroso P, Santa-Clara P. Momentum has its moments[J]. Journal of Financial Economics, 2015, 116(1): 111-120.
[4] Daniel K, Moskowitz T J. Momentum crashes[J]. Journal of Financial Economics, 2016, 122(2): 221-247.
[5] Erb C B, Harvey C R. The strategic and tactical value of commodity futures[J]. Financial Analysts Journal, 2006, 62(2): 69-97.
[6] Liu Z, Lu S, Wang S. Asymmetry, tail risk and time series momentum[J]. International Review of Financial Analysis, 2021, 78: 101938.
[7] Ming L, Song W, Dong M. Revisiting time series momentum in China's commodity futures market: Evidence on sources of momentum profits[J]. Economic Modelling, 2023, 128: 106522.
[8] Zheng D, Zhao C, Hu J. Impact of geopolitical risk on the volatility of natural resource commodity futures prices in China[J]. Resources Policy, 2023, 83: 103568.
[9] Aloui R, Ben Jabeur S, Rezgui H, Ben Arfi W. Geopolitical risk and commodity future returns: Fresh insights from dynamic copula conditional value-at-risk approach[J]. Resources Policy, 2023, 85: 103873.
[10] Hao X, Ma Y, Pan D. Geopolitical risk and the predictability of spillovers between exchange, commodity and stock markets[J]. Journal of Multinational Financial Management, 2024, 73: 100843.
[11] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
[12] Chen T, Guestrin C. XGBoost: A scalable tree boosting system[C]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 785-794.
[13] Fawcett T. An introduction to ROC analysis[J]. Pattern Recognition Letters, 2006, 27(8): 861-874.