energies

MDPI

Article

Dynamic Connectedness and Spillover-Based Machine Learning for Energy-Market Risk Identification: Evidence from U.S. Energy Markets

Junlong Ti [1,*], Hsing Hung Chen [2] and Yinchenyi Feng [1]

1 The Institute for Sustainable Development, Macau University of Science and Technology, Taipa, Macau 999078, China; 3250008788@student.must.edu.mo
2 Faculty of Business, Macau University of Science and Technology, Macau 999078, China; hhchen@must.edu.mo
* Correspondence: 3250000684@student.must.edu.mo

Abstract

Cross-market risk transmission in U.S. energy markets has become increasingly complex as fossil fuel prices, electricity markets, and clean energy financial exposure respond differently to stress episodes. Identifying whether dynamic spillover information contains forward-looking diagnostic value is therefore important for energy market risk monitoring. This study examines a daily six-market U.S. energy return panel covering WTI crude oil, Henry Hub natural gas, Brent crude oil, RBOB gasoline, PJM West electricity, and CELS clean-energy equity exposure from 2016 to 2025. We first estimate time-varying total, directional, and net connectedness using a TVP-VAR-DY framework and then transform the resulting connectedness measures into spillover-based features for supervised high-DSV20-state classification. The results show that energy-market connectedness is clearly time-varying, with crude oil benchmarks occupying central positions and market-level net spillover roles changing across market conditions. Under the retained label-80 Random Forest specification, connectedness-based features provide moderate diagnostic value for identifying future high-DSV20 states. Net WTI, Net Henry Hub, and Net CELS are the most informative spillover-role variables. Additional validation checks indicate that the evidence is best interpreted as support for diagnostic risk monitoring rather than as a high-accuracy forecasting system. The findings highlight the usefulness of dynamic connectedness measures as transparent inputs for energy-market risk assessment.

Keywords: energy-market connectedness; TVP-VAR-DY; spillover-based features; machine learning; high-DSV20 states; diagnostic risk monitoring

Check for updates

Academic Editor: Thanh Nguyen

Received: 14 May 2026
Revised: 10 June 2026
Accepted: 15 June 2026
Published: 18 June 2026

1. Introduction

The linkage among fossil-energy markets, electricity markets, and clean-energy financial assets is reshaping the formation and transmission of energy market risk. Energy security pressure, fuel price repricing, electricity market volatility, and energy transition investment increasingly occur at the same time, making it less appropriate to study crude oil, natural gas, refined fuels, electricity, and clean-energy assets as isolated markets. Instead, these segments are embedded in a more complex cross-market risk transmission system. Existing evidence shows that systemic risk and connectedness between dirty and clean-energy markets vary across market states and display stronger nonlinear and asymmetric features under tail-risk conditions [1]. Dynamic transmission structures are also observed within clean-energy markets, where different segments may act as transmitters or receivers during risk diffusion [2]. In addition, the dynamic dependence between energy-related ETFs and fossil energy commodities indicates that energy-market shocks can spread across financial assets, commodity prices, and risk-expectation channels [3,4]. Energy market risk analysis should therefore move beyond the volatility of individual price series and examine how cross-market shocks are formed, transmitted, and accumulated into system-level risk.

Connectedness measures provide a direct way to describe this transmission process. In studies of energy-market risk transmission, prices and returns are often treated as direct reflections of market dynamics. However, information frictions, liquidity differences, institutional heterogeneity, and heterogeneous shock-absorption capacities make it difficult to capture the full diffusion process by relying only on single-market price information. Dynamic connectedness offers a framework for measuring both the intensity and the direction of cross-market transmission. The total connectedness index captures the aggregate strength of system-wide spillovers, directional spillovers distinguish outward transmission from inward reception, and net spillovers indicate whether a market behaves more as a transmitter or a receiver at a given point in time. Recent evidence on carbon-energy-green finance systems, energy futures markets, and higher-order moment risk spillovers in global energy markets also shows that risk transmission among energy-related markets is not stable over time, but is affected by policy uncertainty, climate attention, geopolitical risk, and market stress conditions [5–7]. In this sense, connectedness measures do not merely reflect static correlations across markets; they transform energy-risk transmission into observable and comparable dynamic system states.

However, the use of connectedness measures in existing energy market studies remains largely retrospective. Most studies examine whether system connectedness increases during particular stress episodes, whether a specific asset becomes a net transmitter, or whether stronger tail dependence exists across markets. Less attention has been paid to whether measured dynamic spillover states can be used as diagnostic inputs for identifying future high-risk conditions. At the same time, machine-learning methods have been widely applied to volatility forecasting for crude oil, natural gas, and clean-energy markets, showing advantages in handling nonlinear dependence and high-dimensional information in energy-market risk management [8–10]. Yet the inputs in these studies are usually prices, volatility measures, technical indicators, or other market variables, while cross-market transmission structures are less often incorporated as economically interpretable machine-learning features. This leaves an important gap: dynamic connectedness measures are mostly used for the ex-post interpretation of energy-market risk transmission, while their role as spillover-based features for diagnostic high-risk-state classification remains insufficiently examined.

This study addresses this gap by linking the TVP-VAR-DY connectedness measurement with supervised high-DSV20-state classification in the U.S. energy-market system. The empirical analysis uses six daily U.S. energy-market series: WTI crude oil, Brent crude oil, Henry Hub natural gas, RBOB gasoline, PJM West electricity, and CELS clean-energy equity exposure. This six-market design covers crude-oil benchmarks, natural gas, refined fuel, regional electricity market conditions, and clean-energy equity exposure, thereby combining conventional energy commodities, power-market dynamics, and energy-transition-related financial exposure. CELS is included to capture clean-energy equity-market exposure associated with the energy-transition segment, rather than to represent a physical energy commodity. The analysis first estimates time-varying connectedness in the energy-market system and extracts total connectedness, directional spillovers, and market-level net spillovers. These measured connectedness outputs are then transformed into spillover-based machine-learning features, including total connectedness, market-level

https://doi.org/10.3390/en19122895

net spillover roles, lagged total connectedness, and short-run TCI variability. Finally, the machine-learning stage evaluates whether these spillover states contain diagnostic information for future high-DSV20 states over a 20-day horizon. Model evaluation considers not only ranking metrics, but also threshold selection, positive-class detection, and the trade-off between false positives and missed high-risk states [11,12].

The main findings can be summarized as follows. First, total connectedness in the U.S. energy-market system is clearly time-varying, indicating that cross-market risk transmission shifts between calmer conditions and intensified transmission regimes. Second, directional and net spillover results show that crude-oil benchmarks occupy central positions in the energy system, while the net roles of natural gas, electricity, refined fuel, and clean-energy equity exposure are more state-dependent. Third, the retained label-80 Random Forest specification provides moderate diagnostic value for identifying future high-DSV20 states. Net WTI, Net Henry Hub, and Net CELS are the most informative spillover-role variables, suggesting that changing market roles contain relevant information for risk monitoring. The evidence should be interpreted as diagnostic rather than as a high-accuracy forecasting system, since the retained classifier still involves false positives and missed high-risk states.

This study contributes to the existing literature in three ways. First, it extends the energy-market connectedness analysis from ex-post spillover measurement to diagnostic high-risk-state classification. Existing studies usually use connectedness measures to describe how shocks are transmitted across energy markets, carbon markets, clean finance, and energy-related assets [13–15]. This study further examines whether measured spillover states can serve as informative inputs for identifying future high-risk conditions. Second, it converts TVP-VAR-DY connectedness outputs into economically interpretable spillover-based machine-learning features. In this way, the machine-learning inputs are directly linked to energy-market transmission roles rather than to generic technical indicators. Third, it evaluates energy-market risk identification with explicit attention to threshold choice, positive-class detection, and false-positive/false-negative trade-offs. This design clarifies the informational value and limitations of connectedness-based features for energy-market risk monitoring.

2. Literature Review

2.1. Energy-Market Connectedness and Volatility Spillovers

Existing studies increasingly treat crude oil, natural gas, electricity, and energy-related financial assets as components of an interconnected spillover system rather than as an isolated price series. Cross-market evidence shows that energy shocks can be transmitted through both commodity and financial channels, embedding U.S. energy assets in broader volatility networks [16,17]. This system view is also supported by energy-firm network evidence, which shows that shocks can spread not only through spot and futures markets but also through energy equities [18]. From an upstream–downstream commodity-chain perspective, transmission roles also differ across markets and horizons [19].

Evidence from conventional energy, electricity, clean energy, and carbon markets further indicates that energy spillovers are state-dependent and asymmetric. Return and volatility transmission among natural gas, crude oil, electricity, and carbon markets varies across regions, horizons, and market states. Clean-energy and conventional energy assets also exhibit asymmetric transmission, while pandemic shocks, energy-security shocks, and geopolitical risk can reshape connectedness structures [20,21]. With the expansion of clean energy, carbon trading, and green finance, transition-related assets have become part of the spillover system. The linkage between clean energy and fossil energy tends to strengthen during crisis periods [22], and time–frequency spillovers among carbon, fossil energy, and clean energy further show that transition risk enters the energy-market transmission process [23]. The relationship between green finance and traditional energy markets also changes across market states [24]. Recent research on hybrid renewable microgrids further shows that clean-energy transition is associated with multi-energy system design, regional operating conditions, and uncertainty in energy-supply strategies. This evidence reinforces the relevance of including clean-energy exposure when examining energy-risk monitoring and cross-market transmission [25].

Methodologically, connectedness measures provide a unified framework for this literature. Forecast-error-variance-decomposition-based spillover measures identify the direction of transmission at both the system and market levels [26]. Network extensions further make it possible to interpret pairwise market roles and transmission structures within the system [27]. Frequency–domain connectedness distinguishes short- and long-horizon propagation, providing tools for analyzing energy-market spillovers across different horizons [28,29]. Energy-market spillovers are therefore better understood as time-varying system variables rather than fixed correlations.

2.2. Time-Varying Connectedness Methods and TVP-VAR-DY Applications

The spillover evidence reviewed in Section 2.1 raises a core measurement issue: cross-market volatility transmission should be represented as a system-level structure rather than as simple bilateral correlations. Forecast-error-variance-decomposition-based connectedness measures convert shocks in a multivariate system into total and directional spillover indices, thereby identifying the sources and recipients of transmission across markets [26,30]. Generalized impulse-response and generalized variance-decomposition approaches further reduce the influence of variable ordering on shock attribution [31,32]. On this basis, network connectedness extensions make it possible to identify pairwise market roles and transmission structures within the system [27].

Static full-sample models and rolling-window approaches are less able to capture transmission changes during stress periods or gradual structural shifts. Time-varying-parameter VAR models allow coefficients and shock covariance to adjust over time, making them better suited to changing financial and energy-market structures [33,34]. The TVP-VAR connectedness framework further applies this logic to dynamic spillover measurement, allowing TCI, directional spillovers, and net spillovers to be estimated as time-varying system variables rather than local averages imposed by a fixed rolling window [35].

Subsequent studies show that connectedness varies not only over time but also across horizons, quantiles, and market states. Frequency–domain connectedness separates short- and long-horizon transmission, while quantile and time–frequency evidence shows that spillovers among energy, metal, and carbon markets can become stronger or change direction under extreme market conditions [28,36]. Recent energy and commodity applications also apply TVP-VAR or related dynamic decomposition methods to uncertainty, oil shocks, clean-energy assets, and geopolitical risk, showing that time-varying connectedness outputs can describe changing transmission structures within energy systems [37]. Much of this literature, however, remains focused on measuring and interpreting connectedness itself, leaving open whether measured spillover states contain forward-looking information for identifying future high-risk conditions.

2.3. Machine-Learning Risk Identification and Early-Warning Models

The preceding discussion treats dynamic connectedness outputs as measured spillover states. A related risk-identification literature asks whether observable state variables can be translated into warning signals for future crises, distress, or high-risk conditions. In this setting, machine learning is better understood as a supervised classification framework rather than as a general forecasting tool. Banking-crisis and financial-crisis studies typically use macro-financial, balance-sheet, or market variables to identify future warning states and evaluate model performance against out-of-sample crisis realizations [38,39]. More recent work uses flexible algorithms to handle nonlinear relationships and time dependence, but its contribution lies in the early-warning design rather than in claiming that crises can be predicted with certainty [40,41].

Financial distress, credit risk, and market-risk studies provide a more specific precedent for classifying high-risk states. Credit scoring, bankruptcy prediction, and bank-failure models usually define default, failure, or high-risk events as binary outcomes and use machine-learning classifiers when predictor effects may be nonlinear or interactive [42,43]. Tree-based and boosting methods have also been applied to bank-failure and extreme market-risk prediction, where the objective is to separate high-risk observations from normal observations rather than to explain structural causality [44,45]. Machine-learning studies in energy economics further show that energy variables can serve not only as price-forecasting targets but also as informational inputs for risk-state identification [46]. Work on crude-oil futures and energy-risk prediction similarly emphasizes that the relevant information lies in movements in risk or volatility states, rather than in the price level itself [47,48].

Evaluation is central to high-risk identification because the positive class is often rare or unevenly distributed. Evidence from imbalanced credit-scoring and banking-crisis settings shows that standard accuracy may obscure weak detection of the high-risk class, making it necessary to distinguish ranking performance from threshold-dependent operating performance [49,50]. ROC-AUC and average precision assess whether a model ranks risky observations ahead of lower-risk observations, while precision, recall, F1, balanced accuracy, and threshold-based metrics describe the classification tradeoff after a classification rule is fixed [51]. Precision–recall analysis is especially relevant when the positive class is limited, because it directly captures the tradeoff between missed high-risk states and false positives [52].

This literature provides a basis for moving from dynamic spillover measurement to supervised risk identification, but it also leaves a more specific gap. Existing financial-crisis and market-risk models often rely on macro-financial indicators, credit variables, accounting ratios, or general market-risk predictors, while energy-market machine-learning studies more often use prices, volatility, or energy-risk features. Less attention has been paid to dynamic connectedness variables and net spillover roles as economically interpretable machine-learning features. Interpretable risk-prediction work shows that feature design can improve the economic readability of a model, but feature importance should not be equated with causal mechanisms [53]. The remaining gap can therefore be stated more narrowly: whether dynamic connectedness variables can serve as interpretable spillover-based features for identifying future high-risk states.

2.4. Research Gap and Contribution

Energy-market connectedness research has shown that volatility transmission extends across conventional energy, clean energy, green finance, and related financial assets. Cross-asset evidence indicates that crude oil, natural gas, coal, equity, and currency markets can form interconnected volatility systems [16]. Linkages between clean energy and fossil energy also become stronger during financial, oil-price, and pandemic stress episodes [22]. Energy-commodity spillovers differ between normal and stressed market states, and the relationship between green finance and traditional energy markets is also state-dependent [19,24]. The remaining question is therefore not whether energy spillovers exist, but whether these measured spillover states can be used for forward-looking risk identification.

Connectedness methods make this question empirically operational. Directional connectedness measures decompose volatility transmission into total connectedness, transmitted spillovers, received spillovers, and net spillovers, while network extensions further reveal pairwise market linkages and market roles [26,27]. The TVP-VAR connectedness framework allows TCI, directional spillovers, and net spillovers to evolve over time rather than remain tied to full-sample or fixed-window estimates [35]. This means that TCI and net spillover roles can be treated not only as post-estimation interpretation measures, but also as dynamic spillover states for subsequent risk-identification designs.

The machine-learning early-warning literature provides the classification framework for this design. Banking-crisis and financial-crisis studies usually define future distress as a supervised warning problem, in which observable indicators are used to identify future crisis or high-risk states [38,39]. Evaluation studies further show that, when high-risk states are unevenly distributed, threshold choice, precision-recall tradeoffs, and operating-point diagnostics are more informative for warning performance than accuracy alone [51,52].

The research gap lies at the intersection of these strands. Existing studies have rarely examined whether dynamic connectedness outputs contain forward-looking information for identifying future high-risk states. Horizon-specific connectedness suggests that spillover states may contain information beyond static correlations and full-sample averages [28]. Interpretable machine-learning research also shows that economically meaningful predictors can improve the readability of risk identification [47,53]. This study therefore transforms TVP-VAR-DY connectedness outputs into spillover-based machine-learning features and uses future high-DSV20 states as the risk-identification target. The focus is the link between dynamic spillover measurement and supervised risk identification, with the interpretation kept diagnostic rather than causal.

3. Data and Variables

The empirical analysis is based on a daily six-market U.S. energy return panel spanning 5 January 2016 to 31 December 2025. After return construction and complete-case alignment, the final balanced panel contains 2180 daily observations. The market coverage includes WTI crude oil, Henry Hub natural gas, Brent crude oil, RBOB gasoline, PJM West electricity, and CELS clean-energy equity index exposure.

The corresponding source symbols are DCOILWTICO for WTI crude oil, DHHNGSP for Henry Hub natural gas, DCOILBRENTEU for Brent crude oil, DRGASLA for RBOB gasoline, PJM WH Real Time Peak for PJM West electricity, and NASDAQCELS for CELS. The WTI, Henry Hub, Brent, and RBOB series use FRED-style identifiers corresponding to EIA energy price series. The CELS series is obtained from the Nasdaq official website and is used to represent clean-energy equity-market exposure. The PJM West electricity series is obtained from the EIA wholesale electricity history page. Specifically, the annual EIA files are downloaded year by year, the PJM West observations are extracted and merged, and the data-preparation procedure reads the market-date field and the average peak price field to construct the final electricity return series.

All variables enter the empirical analysis as daily log returns rather than raw price levels, with the return definition provided in Section 4.1. During sample construction, the raw level series are first aligned by date across the six market calendars, and daily log returns are then computed after this alignment step. The final balanced panel is obtained through complete-case filtering across the six return columns, retaining only dates for which all six market returns are available. No forward filling, backward filling, or interpolation is used. Therefore, non-trading days and cross-market calendar differences are handled through date alignment and complete-case filtering. The subsequent TVP-VAR-DY connectedness estimation and machine-learning classification analysis are conducted using this common daily return panel. Table 1 summarizes the energy-market variables, source symbols, data sources, and variable roles. Table 2 reports the descriptive statistics of the six daily energy-market return series.

Table 1. Energy-market variables and data sources.

Market Segment	Series	Source Symbol	Data Source	Variable Role
U.S. crude-oil benchmark	WTI crude oil	DCOILWTICO	FRED-style EIA series	U.S. crude-oil benchmark
International crude-oil benchmark	Brent crude oil	DCOILBRENTEU	FRED-style EIA series	Global crude-oil benchmark
Natural gas	Henry Hub natural gas	DHHNGSP	FRED-style EIA series	U.S. natural-gas benchmark
Refined fuel	RBOB gasoline	DRGASLA	FRED-style EIA series	Refined petroleum product
Electricity	PJM West electricity	PJM WH Real Time Peak	EIA wholesale electricity history	Regional electricity-market price
Clean-energy equity	CELS clean-energy index	NASDAQCELS	Nasdaq official website	Clean-energy equity-market exposure

Table 2. Descriptive statistics of energy-market returns.

Variable	Observations	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
WTI crude oil	2180	0.001	0.0309	−0.2814	0.4258	1.6384	43.2718
Brent crude oil	2180	0.001	0.0284	−0.2564	0.412	1.2734	39.9286
Henry Hub natural gas	2180	0.0011	0.0736	−1.0251	1.4328	3.2454	102.3041
RBOB gasoline	2180	0.0004	0.0438	−0.4493	0.4838	0.062	25.9106
PJM West electricity	2180	−0.0067	0.2416	−4.8966	4.8018	−0.1073	153.7186
CELS clean-energy equity	2180	0.0005	0.0215	−0.1558	0.1314	−0.1732	7.1731

Notes: This table reports descriptive statistics for the daily log returns of the six energy-market variables over the balanced sample from 5 January 2016 to 31 December 2025. The final balanced panel contains 2180 complete daily observations. The return definition is provided in Section 4.1. No forward filling, backward filling, or interpolation is used in constructing the balanced return panel.

The mean returns are close to zero, whereas volatility and tail behavior differ considerably across markets. PJM West electricity and Henry Hub natural gas show particularly large standard deviations, indicating stronger short-run fluctuations than crude-oil and clean-energy equity series. All six return distributions exhibit high kurtosis, suggesting pronounced tail behavior. These patterns motivate the use of a dynamic connectedness framework to capture cross-market shock transmission.

4. Methodology

4.1. Return Construction and Empirical Inputs

The empirical system uses the balanced six-market daily panel described in Section 3. To place markets with different units and price scales into a common dynamic framework, the aligned price or level series are transformed into daily log returns before connectedness measurement. For market i at date t, the return series is defined as

$$r_{i,t} = \ln(P_{i,t}) - \ln(P_{i,t-1}) \quad (1)$$

where $P_{i,t}$ denotes the observed daily price or index level. This transformation reduces the persistence associated with raw price levels and yields return inputs that are more suitable for modeling cross-market shock transmission in a multivariate dynamic system.

The resulting balanced daily return panel ensures that each observation contains a complete six-market return vector. This common return panel provides the empirical input for the TVP-VAR-DY connectedness framework in the next subsection.

4.2. TVP-VAR-DY Connectedness Framework

Shock transmission across energy markets is unlikely to remain stable over time. Episodes of market stress, fuel substitution, demand repricing, and cross-asset repricing may alter both the strength and the direction of spillovers. To capture this time-varying dependence structure, the analysis employs a time-varying-parameter vector autoregression combined with the Diebold–Yilmaz connectedness approach. This specification uses a true TVP-VAR-DY framework rather than a rolling-window VAR, allowing the coefficient and covariance dynamics to evolve over the sample period.

Let y_t denote the vector of daily returns for the six-market system. The TVP-VAR representation is written as

$$y_t = B_{1,t}y_{t-1} + \cdots + B_{p,t}y_{t-p} + u_t, \quad \mathrm{E}(u_t) = 0, \mathrm{Var}(u_t) = \Sigma_t \tag{2}$$

where $B_{\ell,t}$, $\ell = 1, \ldots, p$, are time-varying coefficient matrices and Σ_t is the time-varying covariance matrix of innovations. The lag order is selected using an AIC-based procedure with a maximum lag of five. This specification allows both dynamic propagation channels and shock covariance to evolve over time, which is consistent with a connectedness environment that does not remain fixed across the sample period.

To translate the estimated TVP-VAR system into connectedness measures, the analysis uses the generalized forecast error variance decomposition (GFEVD). For forecast horizon H, the contribution of shocks from market j to the forecast-error variance of market i at date t is given by

$$\theta_{ij,t}^{(H)} = \frac{\sigma_{jj,t}^{-1} \sum_{h=0}^{H-1} \left(e_i' A_{h,t} \Sigma_t e_j\right)^2}{\sum_{h=0}^{H-1} e_i' A_{h,t} \Sigma_t A_{h,t}' e_i} \tag{3}$$

where $A_{h,t}$ is the time-varying impulse-response coefficient matrix at horizon h, e_i and e_j are selection vectors, and $\sigma_{jj,t}$ is the j-th diagonal element of Σ_t. Under this convention, $\theta_{ij,t}^{(H)}$ measures the share of market i's forecast-error variance attributable to shocks originating from market j. The baseline connectedness calculation uses $H = 20$.

Because the raw GFEVD shares do not necessarily sum to one across each row, they are row-normalized as

$$\overset{\sim}{\theta}{}_{ij,t}^{(H)} = \frac{\theta_{ij,t}^{(H)}}{\sum_{l=1}^{N} \theta_{il,t}^{(H)}} \tag{4}$$

where $N = 6$. This normalization ensures that the contributions to each market's forecast-error variance sum to one. The normalized matrix $\overset{\sim}{\theta}{}_t^{(H)} = [\overset{\sim}{\theta}{}_{ij,t}^{(H)}]$ provides the empirical basis for the total, directional, and net connectedness measures defined in the next subsection. In this framework, off-diagonal elements capture cross-market shock transmission, while diagonal elements capture own-market contributions to forecast-error variance.

4.3. Total, Directional, and Net Spillovers

Based on the row-normalized GFEVD matrix defined in the previous subsection, the first summary measure is the total connectedness index (TCI), which captures the aggregate intensity of cross-market spillovers in the system:

$$TCI_t = 100 \times \frac{1}{N} \sum_{i \neq j}^{N} \overset{\sim(H)}{\theta}_{ij,t} \tag{5}$$

Here, $\overset{\sim(H)}{\theta}_{ij,t}$ denotes the normalized spillover share, and N is the number of markets. TCI_t summarizes the extent to which forecast-error variance is explained by cross-market shocks rather than by own-market innovations.

Directional spillovers decompose this aggregate measure into market-level transmission and reception components. The spillovers transmitted by market i to the rest of the system are defined as

$$TO_{i,t} = 100 \times \sum_{j \neq i} \overset{\sim(H)}{\theta}_{ji,t} \tag{6}$$

while the spillovers received by market i from the rest of the system are defined as

$$FROM_{i,t} = 100 \times \sum_{j \neq i} \overset{\sim(H)}{\theta}_{ij,t} \tag{7}$$

These two measures separate outward shock transmission from inward shock reception at each point in time. Net spillovers summarize the difference between the two directional components:

$$NET_{i,t} = TO_{i,t} - FROM_{i,t} \tag{8}$$

A positive value indicates that market i acts more as a transmitter than as a receiver at date t, whereas a negative value indicates that it acts more as a receiver. In this study, net spillovers are interpreted as time-varying market-role variables rather than fixed market labels. This formulation supports the interpretation of connectedness roles in the empirical results and provides the market-level spillover variables used later in the machine-learning feature set.

4.4. Future High-DSV20 State Definition

The supervised-learning target is constructed from a forward-looking downside-risk measure at the system level. Let $r_{i,t}$ denote the return of market i at date t, where $i = 1, \ldots, N$ and $N = 6$. The system return is defined as the equal-weighted average return across the six markets:

$$r_{sys,t} = \frac{1}{N} \sum_{i=1}^{N} r_{i,t}, N = 6. \tag{9}$$

Based on this system return, the 20-day downside variation measure is defined as the cumulative squared negative system return over the subsequent 20 trading days:

$$DSV20_t = \sum_{h=1}^{20} \left[\min\left(r_{sys,t+h}, 0 \right) \right]^2. \tag{10}$$

This construction captures the magnitude of future downside movements in the aggregate energy-market system. It is forward-looking by design because it measures realized downside risk over the next 20-day horizon. Importantly, $DSV20_t$ is used only as the supervised-learning target and is not included in the input feature set.

To convert this continuous downside-risk measure into a binary high-risk-state label, the analysis compares $DSV20_t$ with a shifted expanding historical quantile threshold. For a given threshold level q, the binary label is defined as

$$Y_t^{(q)} = 1\left(DSV20_t > Q_{q,t-1} \right). \tag{11}$$

where $Q_{q,t-1}$ denotes the expanding historical quantile of DSV20 computed using only observations available before date t. The one-period shift ensures that the threshold does not use the current or future DSV20 value when assigning the label at date t.

The threshold set is $q \in \{0.75, 0.80, 0.85\}$. The 80% threshold is retained as the baseline definition of a future high-DSV20 state, while the 75% and 85% thresholds are used as threshold-robustness checks. This design treats label 80 as the focal empirical specification rather than as a theoretically unique cutoff. The classification task therefore evaluates whether connectedness-based features help identify future high-downside-risk states under a transparent and consistently defined target construction.

4.5. Spillover-Based Feature Construction

After defining the future high-DSV20 target, the next step is to construct the explanatory features used in the classification exercise. The machine-learning stage does not rely on generic technical indicators. Instead, its inputs are derived directly from the TVP-VAR-DY connectedness outputs defined in Sections 4.2 and 4.3. This design allows the classification exercise to examine whether dynamic spillover information contains diagnostic value for identifying future high-downside-risk states.

The baseline feature set contains four types of connectedness variables. The first is the total connectedness index, TCI_t, which captures the overall intensity of cross-market spillovers in the six-market system. The second group consists of market-level net spillovers,

$$NET_{i,t} = TO_{i,t} - FROM_{i,t}, \tag{12}$$

for WTI crude oil, Henry Hub natural gas, Brent crude oil, RBOB gasoline, PJM West electricity, and CELS. These variables capture whether each market acts more as a transmitter or receiver of spillovers at date t.

The third feature captures persistence in the connectedness environment. It is defined as the first lag of total connectedness:

$$tci_{total_{lag1_t}} = TCI_{t-1}. \tag{13}$$

This variable reflects whether the previous day's system-wide connectedness condition carries information for the current risk-monitoring exercise.

The fourth feature captures short-horizon connectedness instability. It is measured based on the rolling five-day standard deviation of TCI:

$$tci_{total_{roll5_{std_t}}} = \mathrm{sd}(TCI_t, TCI_{t-1}, TCI_{t-2}, TCI_{t-3}, TCI_{t-4}). \tag{14}$$

The five-observation rolling window is used as a short recent-history proxy for recent fluctuations in system-wide connectedness. It is computed over the most recent five available observations in the balanced common return panel, rather than over a fixed Monday-to-Friday calendar week. Therefore, holidays and non-trading days do not cascade the calculation forward; they are handled through the date alignment and complete-case filtering described in Section 3. Unlike the DSV20 target, this feature is constructed only from contemporaneous and historical connectedness information available at date t. It therefore captures recent instability in the spillover environment without using future downside-risk information.

The final spillover-based feature vector is

$$X_t = \left(TCI_t, NET_{WTI,t}, NET_{HH,t}, NET_{Brent,t}, NET_{RBOB,t}, NET_{PJM,t}, NET_{CELS,t}, tci\_total\_lag1_t, tci\_total\_roll5\_std_t\right). \tag{15}$$

This fixed feature set links the econometric connectedness framework to the downstream classification design. It is used to assess whether the measured spillover intensity, market-level spillover roles, persistence, and short-term connectedness instability help identify future high-DSV20 states.

4.6. Machine-Learning Classification and Evaluation

The machine-learning stage evaluates whether the spillover-based feature vector X_t helps identify future high-DSV20 states. The classification task is therefore defined as a diagnostic high-risk-state classification problem rather than a general-purpose forecasting system. The baseline target is $Y_t^{(0.80)}$, corresponding to the 80% shifted expanding quantile threshold defined in Section 4.4. The 75% and 85% labels are used to assess threshold robustness.

For each classifier, the model maps the spillover-based feature vector X_t into an estimated probability of a future high-risk state:

$$\hat{p}_t = P(Y_t = 1 \mid X_t). \tag{16}$$

Here, Y_t denotes the baseline high-DSV20-state label under the 80% threshold.

The predicted class label is obtained by comparing the estimated probability with a validation-selected threshold τ_m:

$$\hat{Y}_t^{(m)} = 1\left(\hat{p}_t^{(m)} \geq \tau_m\right). \tag{17}$$

The evaluation follows a fixed chronological split to preserve the time ordering of the data. The training window runs from 12 January 2016 to 12 September 2022, the validation window runs from 21 October 2022 to 4 March 2024, and the test window starts on 15 April 2024. Although the original return panel extends to 31 December 2025, the effective evaluated test observations end earlier because the forward DSV20 target requires 20 subsequent observations.

The training sample is used only for model fitting. The validation sample is used for probability-threshold selection, and the test sample is reserved for the final out-of-sample evaluation. Thresholds are selected based on the validation set using three criteria: maximizing F1, maximizing balanced accuracy, and maximizing recall subject to a minimum precision requirement when feasible. The selected threshold can be expressed as

$$\tau_m^* = \underset{\tau \in T}{argmax} M_{val}(\tau), \tag{18}$$

where $M_{val}(\tau)$ denotes the validation-set objective evaluated at threshold τ.

Model performance is evaluated using the ROC-AUC, average precision (AP), F1, precision, recall, and balanced accuracy. Precision and recall are defined as

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}. \tag{19}$$

The F1 score is the harmonic mean of precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{20}$$

Balanced accuracy is defined as the average of sensitivity and specificity:

$$Balanced\ Accuracy = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right). \tag{21}$$

The ROC-AUC and AP summarize ranking performance, while F1, precision, recall, and balanced accuracy evaluate the threshold-dependent classification performance. Balanced accuracy is particularly useful because the high-risk labels are not evenly distributed across classes.

The classifier comparison is conducted under the same fixed feature set and chronological split. Random Forest is retained as the main nonlinear classifier because it can capture nonlinear interactions among connectedness variables while requiring limited distributional assumptions. Logistic regression and other standard classifiers are used as comparison models to assess whether the diagnostic signal is specific to one modeling choice. Simple benchmark rules, including the majority-class benchmark, are also reported to provide a reference point.

The results are interpreted cautiously: stronger values indicate that connectedness features contain diagnostic information for future high-risk-state classification, but they do not imply a standalone market-timing tool or deterministic forecasting system. To address the possibility that a single chronological split may understate time-series instability or leakage concerns from overlapping forward labels, additional validation checks are reported in Appendix C. These include target-horizon robustness, benchmark and ablation comparisons, and expanding-window validation with a 20-observation embargo.

5. Dynamic Spillover Results

5.1. Total Connectedness over Time

Figure 1 reports the dynamic total connectedness index for the six-market energy system. The index varies markedly over the sample period, ranging from 22.6 to 79.2 between January 2016 and December 2025, with an average of approximately 35.7. This variation indicates that system-wide spillovers are not constant over time. Instead, the energy system shifts between relatively calm periods and episodes of intensified cross-market transmission. Because the TCI summarizes the share of forecast-error variance attributable to cross-market shocks, these movements suggest that the degree of interdependence among oil, gas, gasoline, electricity, and clean-energy equity markets changes materially across the sample.

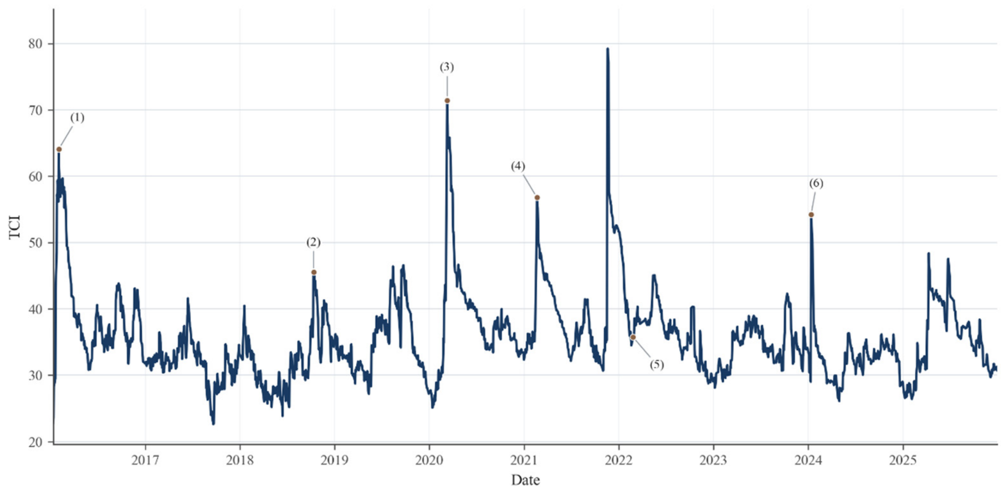

Figure 1. Dynamic total connectedness index for the six-market energy system. Notes: The numbered markers refer to major stress episodes: (1) 2016 oil-market stress; (2) 2018 oil-price correction; (3) 2020 COVID-19 and oil-price war; (4) 2021 Texas freeze; (5) Russia–Ukraine war period; and (6) 2024 connectedness spike. The labels provide temporal references rather than a formal event-study design.

The key finding is that total connectedness is time-varying rather than a fixed background feature of the U.S. energy system. This motivates the subsequent analysis of how aggregate connectedness is distributed across directional and net spillover roles.

5.2. Directional and Net Spillover Structure

Directional spillovers show how the aggregate connectedness documented above is allocated across individual markets. TO spillovers measure shocks transmitted from one market to the rest of the system, FROM spillovers measure shocks received from other markets, and NET spillovers summarize the difference between the two. A positive NET value indicates that a market acts more as a transmitter at date t, while a negative value indicates a stronger receiver position. Figure 2 presents market-level directional and net spillover dynamics for the six markets. Figure 3 and Figure 4 separately display FROM and TO spillovers, while Figure 5 summarizes these dynamics through net spillover roles.

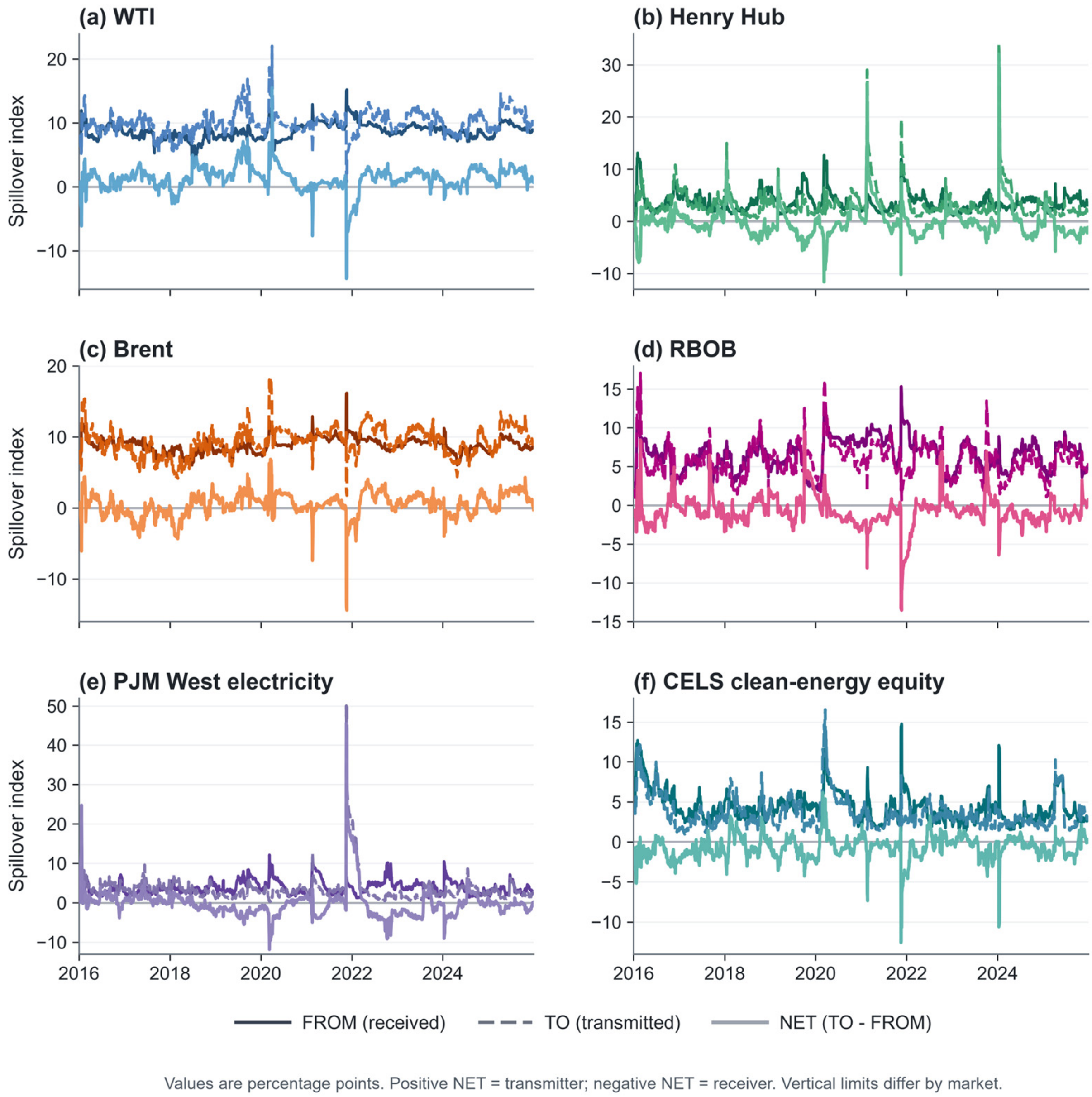

Figure 2. Market-level directional and net spillover dynamics.

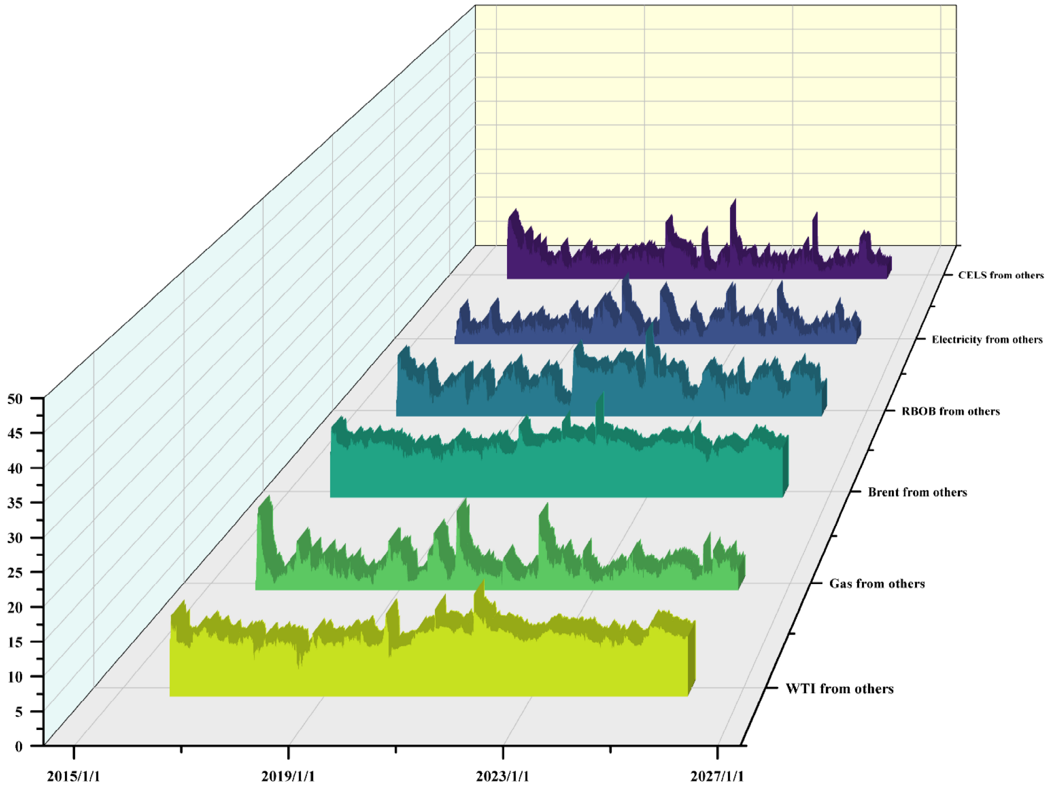

Figure 3. Directional FROM spillovers received from other markets.

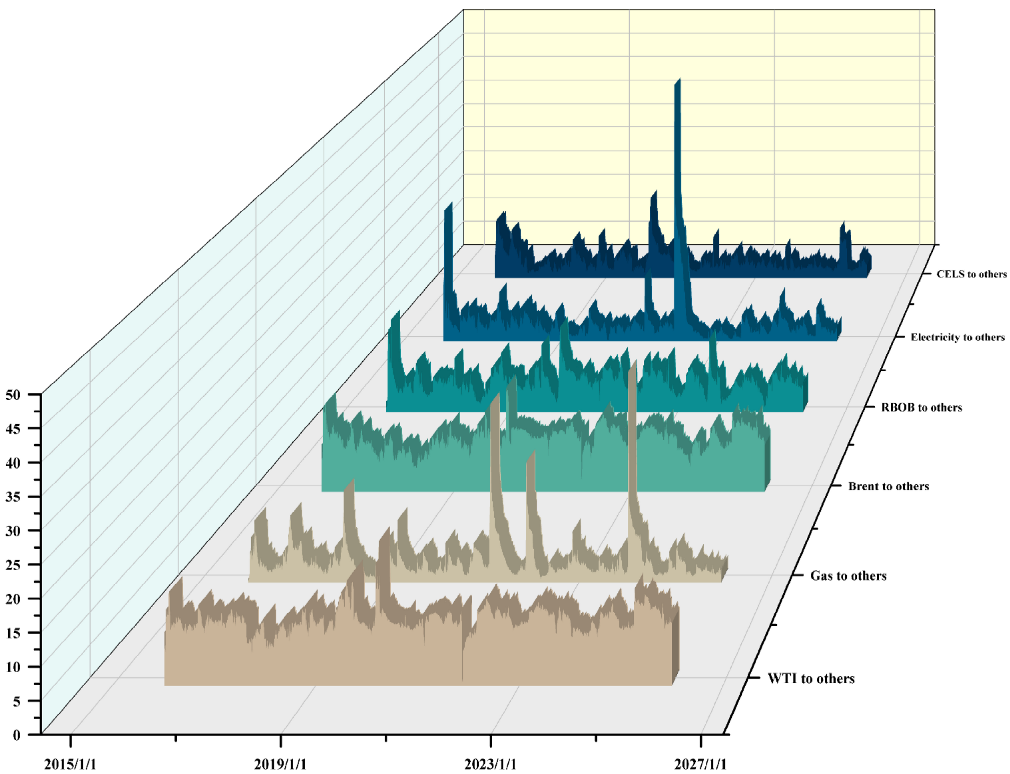

Figure 4. Directional TO spillovers transmitted to other markets.

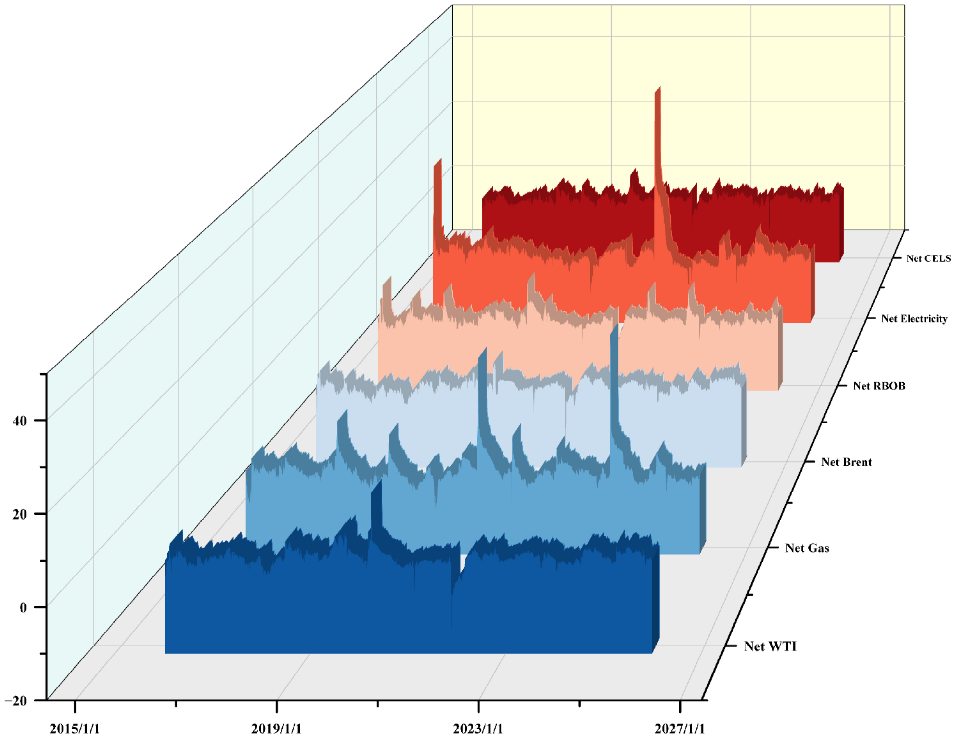

Figure 5. Net directional spillovers across the six markets.

The directional evidence shows that crude-oil-linked markets occupy central positions in the spillover system. WTI and Brent display relatively high outward and inward spillover levels, consistent with their role as broad pricing benchmarks for the wider energy complex. Their directional measures also vary substantially over time, indicating that crude-oil markets are not fixed transmitters or receivers but switch in intensity as market conditions change. RBOB gasoline shows a pronounced receiving component, linking refined-fuel dynamics to the broader crude-oil and energy-market system. By contrast, Henry Hub natural gas, PJM West electricity, and CELS clean-energy equity generally display lower average directional spillovers, although their time variation remains economically relevant during stress episodes.

The net spillover measures add a market-role dimension to the directional evidence. WTI is positive through much of the sample and has a positive average net position, suggesting that it more often contributes to system-wide transmission than it absorbs spillovers from the rest of the system. Brent is also positive on average, but its net role changes more frequently. RBOB gasoline and CELS clean-energy equity are negative on average, indicating that they more often receive shocks from the system than transmit them back. Henry Hub natural gas and PJM West electricity have average net positions close to zero but display wide time-varying ranges, suggesting state-dependent roles rather than fixed market identities.

The key finding is that spillover roles are dynamic rather than permanent market labels. Crude-oil benchmarks tend to sit near the center of the system, while refined fuel, electricity, natural gas, and clean-energy equity exhibit more state-dependent transmission and reception patterns. This role information provides the economic basis for the spillover-based feature set used later to identify future high-DSV20 states.

5.3. From Spillover Measurement to Diagnostic Features

The shift from spillover measurement to feature construction directly follows from the preceding results. TCI captures the time-varying intensity of system-wide transmission, while net spillovers capture changing transmitter and receiver positions at the market level. Lagged TCI and short-term TCI variability add information on persistence and recent instability in aggregate connectedness. Together, these variables form a spillover-based feature set that summarizes connectedness intensity, market roles, persistence, and short-horizon instability.

These features are not generic technical indicators added after the connectedness analysis. They are measured outputs from the TVP-VAR-DY system and retain a direct link to the estimated spillover structure. Their role in the next section is therefore diagnostic: the classification analysis evaluates whether observed spillover states contain useful information for identifying future high-DSV20 conditions over the subsequent 20-day horizon. This interpretation does not treat the spillover variables as structural causes of future risk episodes.

6. Machine-Learning Identification of Future High-DSV20 States

6.1. Model Comparison and Retained Specification

Section 6 evaluates whether the spillover-based feature set constructed from the TVP-VAR-DY outputs contains diagnostic information for future high-DSV20 states. The comparison is conducted across the 75%, 80%, and 85% label settings using six test-sample metrics: ROC-AUC, average precision (AP), F1, precision, recall, and balanced accuracy. The ROC-AUC and AP evaluate ranking performance before a probability threshold is imposed, whereas F1, precision, recall, and balanced accuracy evaluate the binary classification rule after threshold selection.

Figure 6 reports the model-performance comparison across label settings and classifiers. The heatmap displays the test-sample values of ROC-AUC, AP, F1, precision, recall, and balanced accuracy, while the full numerical results underlying the figure are reported in Appendix C, Table A5. The results indicate that model performance depends on the definition of the high-risk state. The 75% label defines a broader high-risk class, while the 85% label imposes a stricter and less frequent high-risk definition. The 80% label is retained as the baseline empirical specification because it provides an intermediate high-risk-state definition between these two alternatives. The 75% and 85% labels are used as threshold-robustness checks rather than as competing theoretical definitions.

Under the baseline label-80 setting, the retained Random Forest classifier obtains ROC-AUC of 0.662 and AP of 0.621 on the held-out test set. Its threshold-dependent performance is also above the naive benchmark, with F1 of 0.622, precision of 0.573, recall of 0.681, and balanced accuracy of 0.611. These values indicate moderate discriminatory ability rather than high-accuracy classification. The result supports the interpretation that connectedness features contain useful diagnostic information for identifying future high-DSV20 states, but the retained specification should not be interpreted as a standalone forecasting system.

The comparison with simpler benchmarks supports a cautious incremental-value interpretation. The majority-class benchmark cannot identify positive high-risk states, while the default logistic regression model shows limited threshold-dependent performance. The tuned balanced logistic model improves recall, but this comes with weaker ranking performance than the retained Random Forest specification. The retained Random Forest therefore provides the most balanced diagnostic trade-off among the reported comparison models when ranking metrics and threshold-dependent metrics are considered together.

The following subsection examines the precision–recall profile and the validation-selected operating threshold for the retained label-80 Random Forest specification. This step is necessary because AP summarizes ranking quality, whereas the final high-risk-state classification depends on the probability threshold used to convert predicted scores into binary labels.

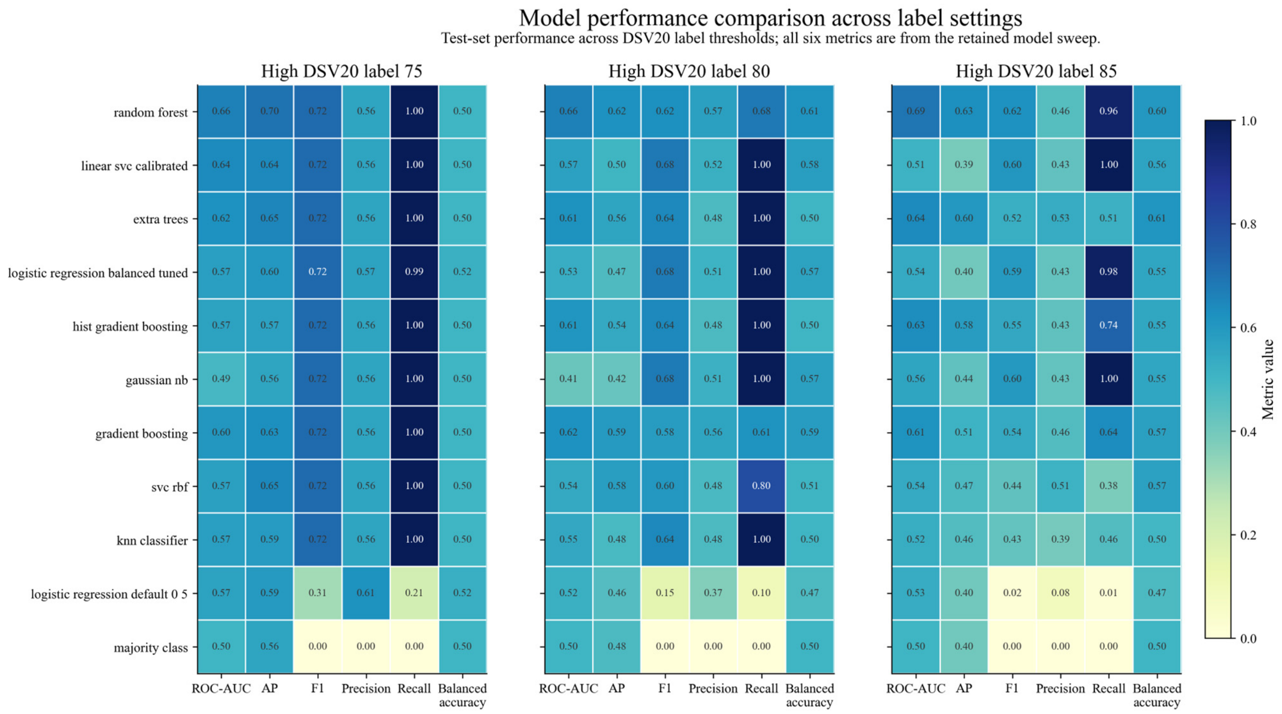

Figure 6. Model performance comparison across label settings and classifiers.

6.2. Precision–Recall Evidence and Threshold Selection

Precision–recall evidence is useful because the empirical task is to identify future high-DSV20 states rather than to maximize overall classification accuracy. When the positive class is unevenly distributed, accuracy can appear acceptable even if the model retrieves too few high-risk observations. Figure 7 therefore focuses on the retained label-80 specification and shows how the candidate classifiers trade precision against recall across probability thresholds. The horizontal reference line reports the test positive rate, providing a baseline for assessing whether the precision–recall curve adds information beyond the unconditional frequency of high-risk states.

For the retained Random Forest specification, the marked operating point corresponds to a validation-selected threshold of approximately 0.297, with precision of 0.573 and recall of 0.681. This threshold is selected before test-set evaluation and is therefore interpreted as the operating threshold for the retained specification rather than as a threshold tuned on the test set. At this operating point, precision remains above the unconditional test positive rate, while recall indicates that the model identifies a meaningful share of future high-DSV20 states.

The evidence supports a diagnostic risk-monitoring interpretation. The retained classifier improves positive-class identification relative to the baseline frequency benchmark, but it still involves a visible trade-off between missed high-risk states and false positives. The selected operating point should therefore be interpreted as a practical classification rule within the reported design, not as an accuracy-maximizing or universally optimal threshold.

Appendix A, Figure A1 and Table A1 report the threshold diagnostics behind the selected operating point. The test-set confusion matrix reported later evaluates the same retained operating rule out of sample. This structure links the precision–recall evidence in Figure 7 to the classification-error profile while keeping the interpretation focused on risk-monitoring performance rather than on strong forecasting claims.

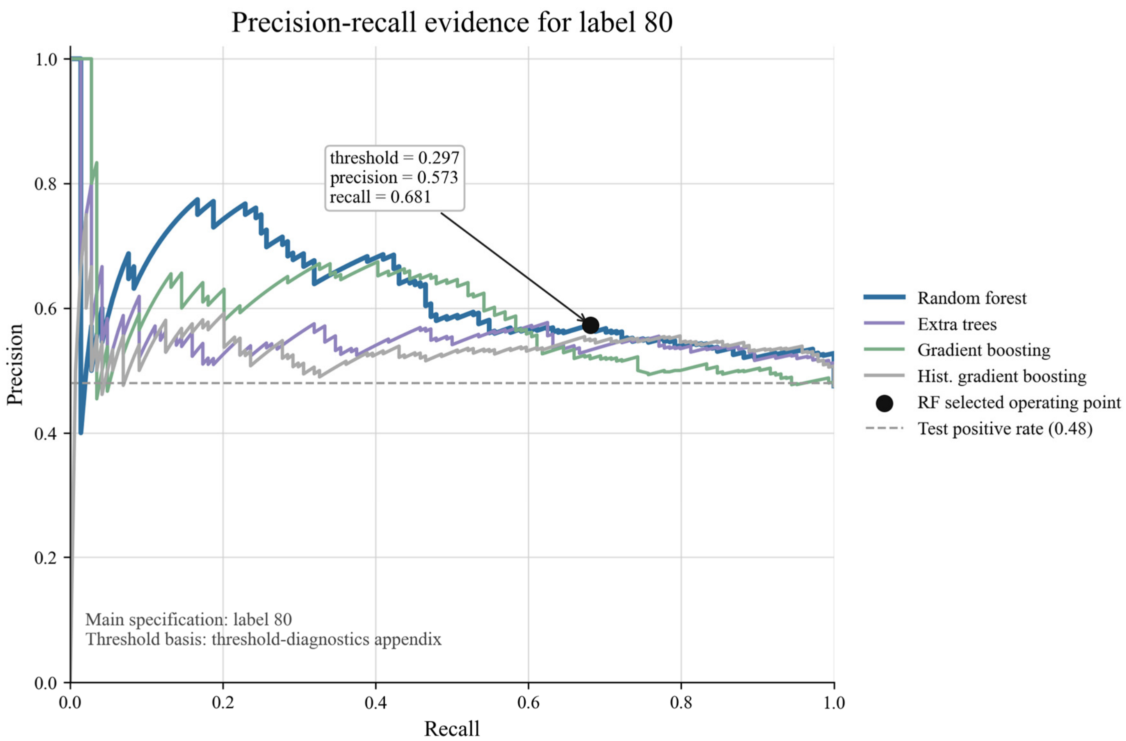

Figure 7. Precision–recall curve and selected operating point for the label-80 Random Forest specification.

6.3. Interpreting the Retained Random Forest Through Spillover Features

Figure 8 reports permutation importance for the net-spillover component of the retained label-80 Random Forest specification. The purpose is to identify which market-level spillover-role variables contribute most to the classifier's out-of-sample score within the reported design. These variables are not generic technical indicators introduced independently of the connectedness analysis. They are derived from the TVP-VAR-DY net spillover measures and summarize whether each market acts more as a net transmitter or a net receiver within the estimated energy-market system.

The ranking indicates that the diagnostic contribution is concentrated in a small set of spillover-role variables. Net WTI has the largest positive permutation importance, suggesting that variation in the net transmission role of the U.S. crude-oil benchmark contains the strongest incremental information for the retained classifier. Net Henry Hub is the second most important feature, indicating that the natural-gas spillover role also contributes to the identification of future high-DSV20 states. Net CELS provides a smaller but still positive contribution, linking clean-energy equity exposure to the diagnostic information set.

The remaining net-spillover variables provide weaker or less stable incremental information in this specification. Net Brent is close to zero, while Net PJM West and Net RBOB are slightly negative in the permutation exercise. These values should not be interpreted as negative economic effects. In permutation importance, a near-zero or negative value means that randomly shuffling the variable does not reduce the retained model's score in a stable way, which can occur when the variable is noisy, redundant, or less informative under the selected specification.

The interpretation is therefore diagnostic rather than causal. Figure 8 identifies which measured spillover-role variables contribute to out-of-sample classification performance for the retained classifier, but it does not establish a structural mechanism behind future high-DSV20 states. The results suggest that net spillover roles, especially those of WTI, Henry Hub, and CELS, contain useful information for risk monitoring within the reported machine-learning design.

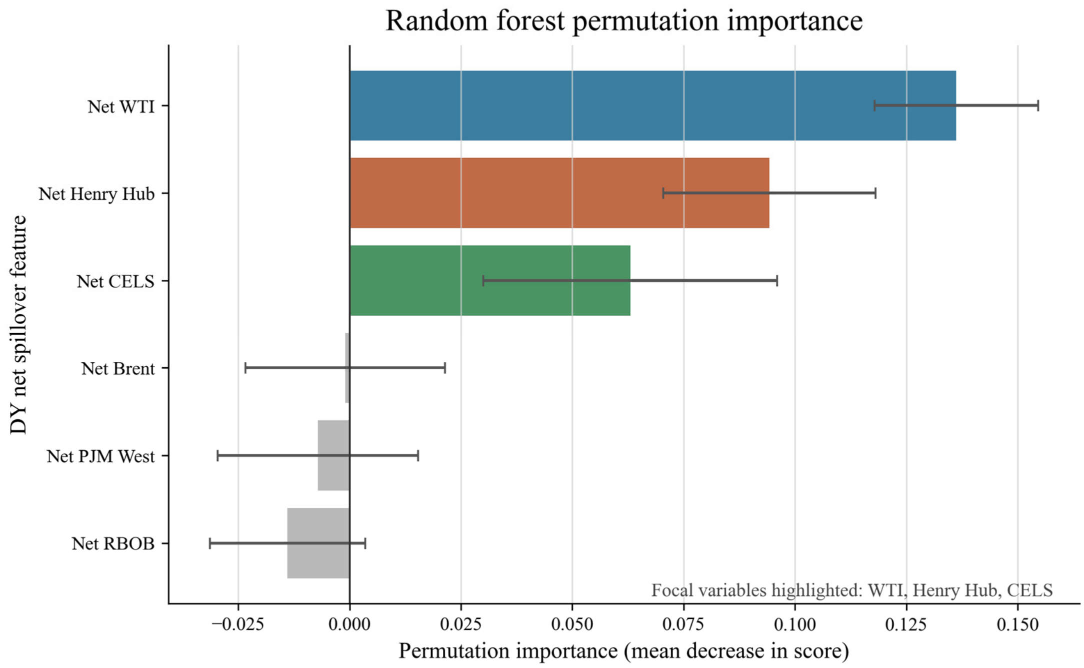

Figure 8. Permutation importance of net-spillover features for the retained label-80 Random Forest specification.

6.4. Spillover Dynamics as an Interpretation Bridge

Figure 9 links the retained spillover features back to the connectedness dynamics from which they are constructed. Figure 8 identifies the net-spillover variables that contribute most to the retained classifier, while Figure 9 shows how these variables evolve over time alongside the total connectedness index. This step is useful because the selected predictors are not external technical indicators; they are measured spillover roles embedded in the TVP-VAR-DY connectedness system documented in Section 5.

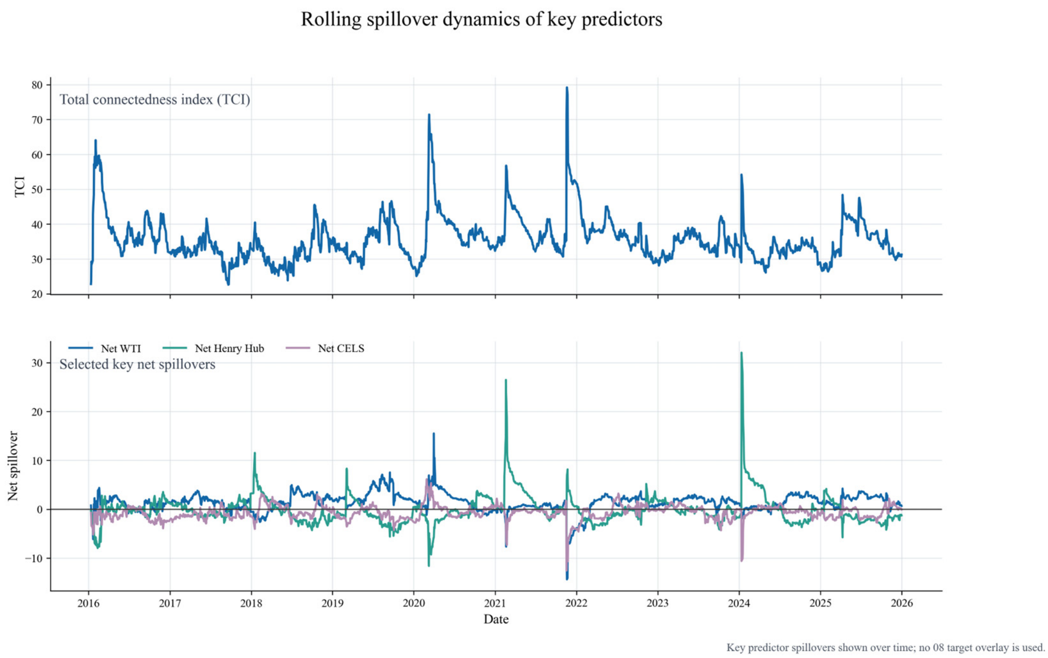

Figure 9. Rolling spillover dynamics for TCI and selected net spillovers.

The upper panel shows that aggregate connectedness changes substantially over time, with repeated shifts between calmer and more elevated transmission regimes. The lower panel shows that Net WTI, Net Henry Hub, and Net CELS also vary over time rather than remaining fixed market labels. Net WTI is positive through much of the sample, consistent with a recurrent net-transmitter role for the U.S. crude-oil benchmark. Net Henry Hub exhibits sharp reversals, indicating a more state-dependent gas-market spillover role. Net CELS is negative for extended periods but also shows short-lived positive episodes, suggesting that clean-energy equity exposure can shift between receiver and transmitter positions.

Figure 9 therefore provides a descriptive economic interpretation of the retained predictors. The variables highlighted by the Random Forest summarize changing roles within the connectedness system, while the interpretation remains diagnostic and descriptive rather than causal. The figure shows that the retained features are embedded in the evolving spillover environment, while the classification analysis evaluates whether these measured spillover states help identify future high-DSV20 conditions.

6.5. Classification Error Structure at the Selected Operating Point

Figure 10 reports the test-set classification outcomes for the retained label-80 Random Forest specification at the same validation-selected operating threshold marked in Figure 7. The confusion matrix contains 98 true positives, 46 false negatives, 73 false positives, and 86 true negatives. In this setting, true positives are future high-DSV20 observations correctly classified as high-risk states, while false negatives are high-DSV20 observations not classified as high risk. False positives are lower-risk observations classified as high risk, and true negatives are lower-risk observations correctly classified as low risk.

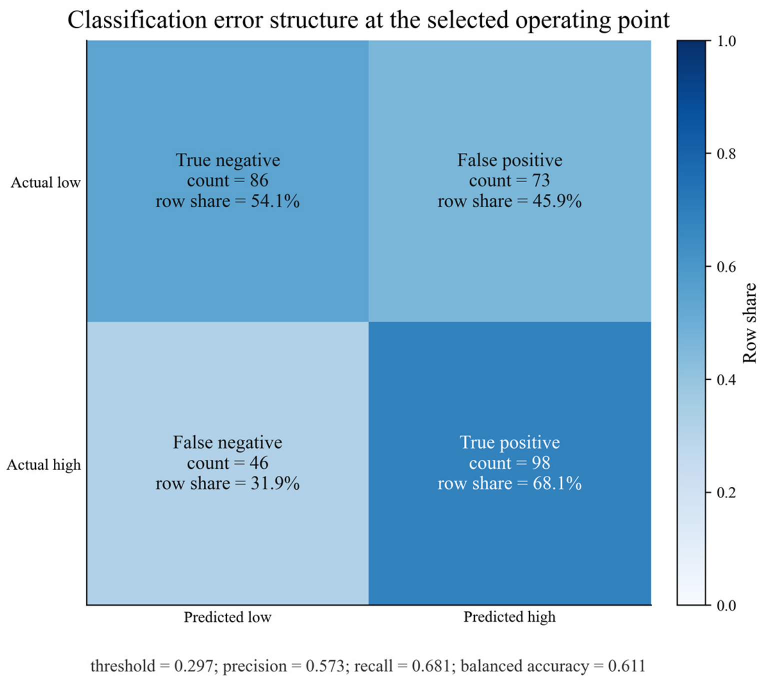

Figure 10. Confusion matrix for the retained label-80 Random Forest operating point.

The matrix is consistent with the precision–recall evidence reported above. At the selected operating point, the classifier identifies a meaningful share of future high-DSV20 states, with recall of 0.681. Precision is 0.573, indicating that a non-trivial share of predicted high-risk classifications does not correspond to realized high-DSV20 observations. Balanced accuracy is 0.611, showing that the retained rule improves on a no-skill benchmark but remains far from a high-accuracy classification rule.

The error structure therefore supports a cautious diagnostic interpretation. A lower threshold would identify more future high-DSV20 states but would also increase false positives, whereas a higher threshold would reduce false positives at the cost of more missed high-risk states. Figure 10 makes this operating trade-off explicit. The retained classifier should therefore be interpreted as a risk-monitoring tool that provides useful but imperfect classification information, rather than as a precise forecasting system.

Appendix A, Figure A1 and Table A1 document the threshold diagnostics behind this operating point, while Figure 10 evaluates the retained rule on the held-out test set. This separation keeps validation-stage threshold selection distinct from the out-of-sample classification assessment.

6.6. Robustness and Sensitivity Checks

The preceding subsections retain the label-80 Random Forest specification as the focal empirical configuration. Additional checks examine whether the classification evidence is sensitive to the feature grouping, input treatment, DY-parameter settings, target-horizon design, benchmark features, and time-series validation. Appendix B reports the original robustness and sensitivity checks, including feature-group robustness, winsorized-input sensitivity, and DY-parameter sensitivity. Appendix C reports additional validation and model diagnostics, including the full numerical metrics underlying Figure 6, target-horizon robustness, benchmark and ablation checks, and expanding-window validation with a 20-observation embargo.

Table 3 provides a compact summary of the feature-group and input-sensitivity checks reported in Appendix B, Tables A2 and A3. These checks are distinct from the model-comparison evidence in Figure 6 and Appendix C, Table A5.

Table 3. Compact robustness and sensitivity summaries for the retained label-80 Random Forest specification.

Robustness Dimension	Specification	AP	Balanced Accuracy	Interpretation
Feature group	Total-only	0.331	0.532	Contains some connectedness information, but weaker than net-only
Feature group	Net-only	0.490	0.585	Strongest among the reported feature groups
Feature group	Total plus net	0.433	0.556	Adds information, but does not exceed net-only
Input sensitivity	Baseline input	0.439	0.551	Baseline input comparison point
Input sensitivity	Winsorized input	0.401	0.582	Operating performance remains comparable after winsorization

The feature-group comparison supports the role of net-spillover information in the retained specification. As shown in Appendix B, Table A2, the net-only feature group has the highest AP and balanced accuracy among the reported feature-group checks, while the total-only and total-plus-net specifications provide weaker or less consistent performance. This pattern is consistent with the interpretation that market-level net spillover roles contain useful diagnostic information for identifying future high-DSV20 states. The result should not be interpreted as a universal feature-selection claim, because the comparison is conditional on the retained label definition, model class, feature construction, and evaluation protocol.

The winsorized-input check provides a separate input-sensitivity assessment. Appendix B, Table A3 shows that, after winsorization, balanced accuracy remains comparable to the baseline input case, although AP declines. This indicates that the retained interpretation is not driven solely by unprocessed extreme input values, while also confirming that positive-class ranking performance is sensitive to input treatment. Appendix B, Table A4 reports DY-parameter sensitivity evidence and should be read as a parameter-sensitivity check rather than as a definitive robustness proof across all connectedness-design choices.

The additional validation checks in Appendix C further qualify the baseline evidence. Appendix C, Table A6 shows that the DSV20 target horizon provides the most balanced diagnostic performance among the DSV10, DSV20, DSV30, and DSV60 specifications, while the longer and shorter horizons should not be overstated as uniformly supportive alternatives. Appendix C, Table A7 compares the connectedness-based specification with returns-only, realized-volatility-only, and combined-control benchmarks; the results support incremental diagnostic value from connectedness features rather than broad dominance across all feature designs. Appendix C, Table A8 reports expanding-window validation with a 20-observation embargo. The fold-based evidence suggests that the fixed chronological split is not obviously over-optimistic, but it also shows time-series variation in classification performance.

Overall, the robustness and validation evidence supports a cautious diagnostic risk-monitoring interpretation. The retained specification provides useful but imperfect information for identifying future high-DSV20 states. The evidence should therefore be interpreted as support for connectedness-based diagnostic classification, not as a claim of full robustness across all design choices or as evidence of a high-accuracy forecasting system.

7. Discussion and Implications

7.1. Economic Interpretation of Dynamic Spillover Roles

The evidence in Sections 5 and 6 suggests that U.S. energy-market connectedness is time-varying and role-dependent. The diagnostic information used by the retained classifier is concentrated in a limited set of measured spillover-role variables. In the retained label-80 Random Forest specification, Net WTI, Net Henry Hub, and Net CELS provide the most informative spillover signals. This pattern is economically meaningful because these variables are constructed from TVP-VAR-DY connectedness outputs and summarize changing net transmitter and receiver positions rather than generic technical covariates.

These selected spillover roles correspond to distinct positions within the energy-market system. Net WTI captures shifts in the net transmission role of the U.S. crude-oil benchmark and reflects its central position in broader energy-market repricing. Net Henry Hub reflects the role of the natural-gas market, whose transmission and reception patterns can change sharply across fuel-substitution pressures and regional energy conditions. Net CELS extends the interpretation beyond fossil-fuel markets by showing that clean-energy equity exposure can also enter the connectedness environment through changing net spillover roles.

Figures 8 and 9 provide complementary evidence for this interpretation. Figure 8 identifies the main diagnostic contributors within the retained classifier, while Figure 9 shows that these variables are embedded in a time-varying connectedness environment. The interpretation is therefore diagnostic and descriptive rather than causal. The results suggest that measured spillover roles contain useful information for risk monitoring, but they do not imply that these variables structurally cause future high-DSV20 states.

7.2. Implications for Spillover-Based Risk Identification

The results broaden the informational use of connectedness measures in energy-market risk analysis. In much of the connectedness literature, spillover measures are used mainly as retrospective summaries of how shocks propagate through the system. The evidence in this study suggests that measured spillover roles can also be used as inputs for forward-looking high-risk-state classification. Because the features are constructed from TVP-VAR-DY outputs, the retained classifier does not rely on external technical indicators. Instead, it evaluates whether measured spillover intensity, market-level net roles, persistence, and recent connectedness instability contain diagnostic information for future high-DSV20 states.

The confusion-matrix evidence is central to this interpretation because it makes the operating consequences of the retained classification rule visible. The classifier identifies a meaningful share of future high-DSV20 states, but it also generates false positives and misses some high-risk episodes. These classification errors are not secondary details; they are the practical expression of the precision–recall trade-off. A lower threshold would identify more high-DSV20 states at the cost of broader false positives, while a higher threshold would reduce false positives at the cost of more missed high-risk states.

The retained specification should therefore be read as a diagnostic risk-monitoring configuration under the reported design, not as an error-free decision rule. Its contribution lies in translating dynamic spillover information into an operational classification signal that can support monitoring of energy-market-risk conditions. This interpretation is consistent with the moderate performance reported in Section 6: the model provides useful but imperfect diagnostic information, rather than a deterministic or high-accuracy forecasting system.

7.3. Limitations and Scope of Interpretation

The retained evidence should be interpreted as diagnostic rather than causal. The permutation importance in Figure 8 identifies which spillover-role variables contribute most to the retained classifier's out-of-sample performance, but it does not show that these variables structurally cause future high-DSV20 states. Figure 9 provides an economic interpretation of the selected variables by linking them back to time-varying connectedness dynamics, but it should not be read as a structural mechanism test.

This boundary also applies to specification choices. The label-80 setting is retained as the baseline empirical specification because it provides an intermediate high-risk-state definition between the broader 75% threshold and the stricter 85% threshold. The Random Forest model is retained because it provides the most balanced diagnostic trade-off within the reported comparison, not because it is universally preferred across all possible designs. The additional validation checks in Appendix C examine alternative target horizons, benchmark feature sets, and expanding-window validation, but they do not eliminate all model-design uncertainty.

The operating interpretation is also conditional on the selected probability threshold. As Figures 7 and 10, Appendix A, Figure A1 and Table A1 show, threshold choice shifts the balance between false positives and false negatives. This does not weaken the diagnostic interpretation; rather, it defines the scope of the retained classification rule. The results support the use of dynamic spillover measures as informative inputs for energy-market risk monitoring, while leaving room for alternative thresholds, model classes, connectedness specifications, and forecast horizons in future research.

8. Conclusions

This study examines whether dynamic spillover information can support future high-risk-state identification in the U.S. energy market. Using a daily six-market return panel covering crude oil, natural gas, gasoline, electricity, and clean-energy equity exposure, the analysis combines a TVP-VAR-DY connectedness measurement with supervised high-DSV20-state classification. The framework first estimates time-varying total, directional, and net spillovers, and then evaluates whether these measured spillover states contain diagnostic information for future downside-risk conditions.

The results show that U.S. energy-market connectedness is clearly time-varying. Total connectedness shifts between calmer periods and episodes of intensified cross-market transmission, indicating that system-wide spillover intensity is not a fixed background feature. Directional and net spillover results further show that market roles are dynamic rather than permanent. Crude-oil benchmarks, especially WTI, tend to occupy central positions in the system, while natural gas, electricity, refined fuel, and clean-energy equity exposure display more state-dependent transmitter and receiver roles.

The machine-learning evidence suggests that connectedness-based features provide useful but moderate diagnostic value for identifying future high-DSV20 states. Under the retained label-80 Random Forest specification, the model improves on naive benchmarks and identifies a meaningful share of high-risk observations, but the confusion-matrix evidence also shows non-trivial false positives and missed high-risk states. The retained specification should therefore be interpreted as a diagnostic risk-monitoring tool rather than as a high-accuracy forecasting system or deterministic decision rule. The most informative variables are concentrated in net spillover roles, particularly Net WTI, Net Henry Hub, and Net CELS, suggesting that changing market roles contain relevant information for risk monitoring.

Several limitations define the scope of the findings. The results are conditional on the six-market sample, the DSV20 target construction, the selected threshold design, the retained classifier, and the sample split. The spillover variables should not be interpreted as structural causes of future high-risk states. Instead, they provide empirically measured signals from the connectedness system. Future research can extend this framework by considering alternative market coverage, additional macro-financial controls, other target horizons, and different validation designs. Overall, the study shows that dynamic spillover measures can enrich energy-market risk monitoring when interpreted cautiously and transparently.

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

Funding: This research received no external funding.

Data Availability Statement: Raw data were obtained from publicly available sources; processed data are available from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT by OpenAI (GPT-5.5 Thinking) to assist with English-language polishing, wording refinement, formatting suggestions, and cover-letter preparation. The tool was not used to generate empirical results, perform data analysis, create datasets, or draw scientific conclusions. The authors reviewed and edited all outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Threshold Diagnostics

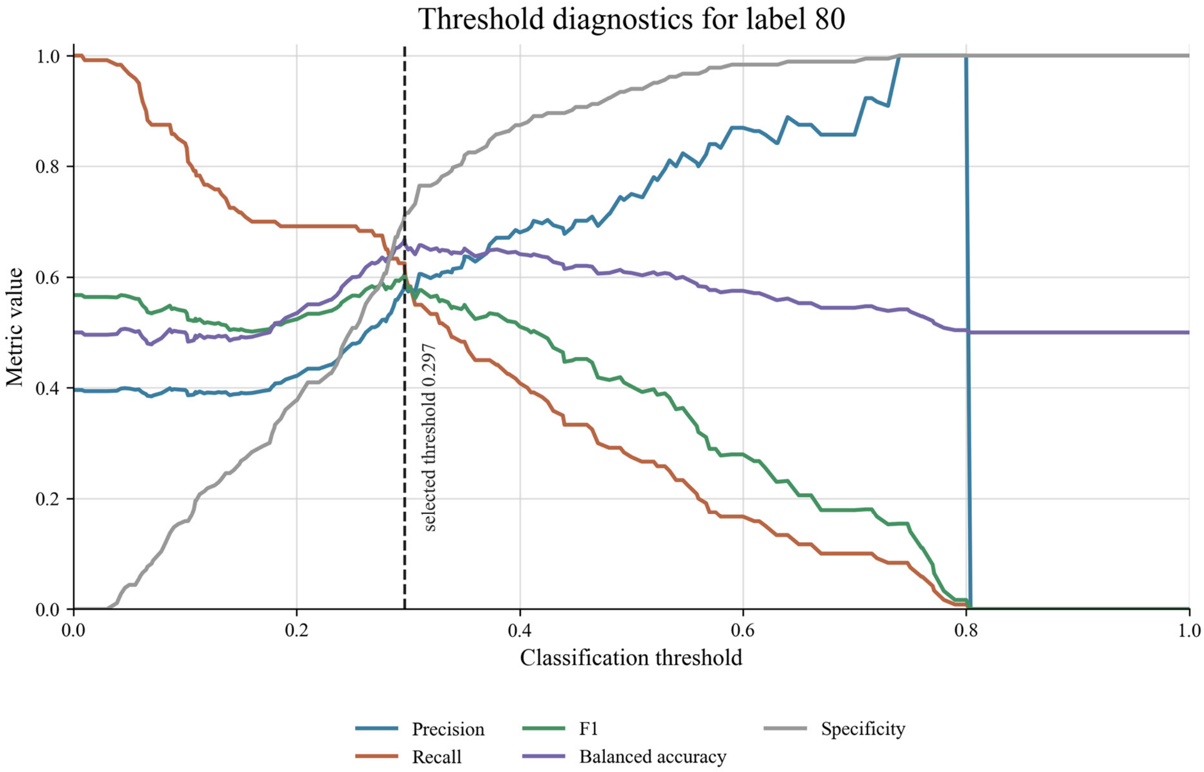

Figure A1. Threshold diagnostics for label-80. Note: This figure reports threshold diagnostics for the retained label-80 Random Forest specification. It supports the validation-selected precision–recall operating point and the subsequent confusion-matrix interpretation. The figure is used as diagnostic support rather than as a main-result figure.

Table A1. Threshold grid for label-80.

Threshold	Precision	Recall	F1	Bal. Acc.	Specificity	MCC	Pred. Pos. Rate
0.000	0.396	1.000	0.567	0.500	0.000	0.000	1.000
0.050	0.399	0.967	0.564	0.505	0.044	0.026	0.960
0.100	0.396	0.842	0.539	0.500	0.158	0.000	0.842
0.150	0.391	0.717	0.506	0.492	0.268	−0.017	0.726
0.200	0.421	0.692	0.524	0.534	0.377	0.070	0.650
0.250	0.480	0.692	0.567	0.600	0.508	0.197	0.571
0.297	0.586	0.625	0.605	0.668	0.710	0.332	0.422
0.300	0.574	0.583	0.579	0.650	0.716	0.298	0.403
0.350	0.630	0.483	0.547	0.649	0.814	0.316	0.304
0.400	0.681	0.408	0.510	0.641	0.874	0.325	0.238
0.450	0.702	0.333	0.452	0.620	0.907	0.301	0.188
0.500	0.750	0.275	0.402	0.607	0.940	0.298	0.145
0.550	0.818	0.225	0.353	0.596	0.967	0.302	0.109
0.600	0.870	0.167	0.280	0.575	0.984	0.277	0.076

Table A1. *Cont.*

Threshold	Precision	Recall	F1	Bal. Acc.	Specificity	MCC	Pred. Pos. Rate
0.650	0.875	0.117	0.206	0.553	0.989	0.231	0.053
0.700	0.857	0.100	0.179	0.545	0.989	0.208	0.046
0.750	1.000	0.075	0.140	0.537	1.000	0.216	0.030
0.800	1.000	0.008	0.017	0.504	1.000	0.071	0.003
0.850	0.000	0.000	0.000	0.500	1.000	0.000	0.000
0.900	0.000	0.000	0.000	0.500	1.000	0.000	0.000
0.950	0.000	0.000	0.000	0.500	1.000	0.000	0.000
1.000	0.000	0.000	0.000	0.500	1.000	0.000	0.000

Note: This table reports representative threshold-diagnostic results for the retained label-80 Random Forest specification. Rows are shown at 0.05 threshold increments, together with the validation-selected operating threshold when distinct. The table is used to document the threshold-selection process; the held-out test-set performance at the retained threshold is reported in Figure 10.

Appendix B. Robustness and Sensitivity Checks

Table A2. Feature-group robustness for the retained label-80 Random Forest specification.

Feature Group	Specification	Threshold	ROC-AUC	AP	F1	Precision	Recall	Bal. Acc.
total-only	default	0.500	0.560	0.331	0.318	0.325	0.310	0.526
total-only	max F1	0.413	0.560	0.331	0.448	0.299	0.897	0.525
total-only	max bal. acc.	0.414	0.560	0.331	0.450	0.303	0.874	0.532
total-only	max recall ($p \geq 0.30$)	0.365	0.560	0.331	0.442	0.289	0.943	0.504
net-only	default	0.500	0.664	0.490	0.248	0.722	0.149	0.563
net-only	max F1	0.163	0.664	0.490	0.466	0.346	0.713	0.585
net-only	max bal. acc.	0.163	0.664	0.490	0.466	0.346	0.713	0.585
net-only	max recall ($p \geq 0.30$)	0.056	0.664	0.490	0.460	0.303	0.954	0.535
total-plus-net	default	0.500	0.594	0.433	0.405	0.451	0.368	0.594
total-plus-net	max F1	0.358	0.594	0.433	0.400	0.288	0.655	0.501
total-plus-net	max bal. acc.	0.454	0.594	0.433	0.402	0.344	0.483	0.556
total-plus-net	max recall ($p \geq 0.30$)	0.269	0.594	0.433	0.459	0.308	0.897	0.543

Note: This table compares total-only, net-only, and total-plus-net feature groups for the retained label-80 Random Forest specification. The results are reported as Appendix B sensitivity evidence and are not used as a universal feature-selection claim.

Table A3. Winsorized-input sensitivity for the retained label-80 Random Forest specification.

Input Version	Specification	Threshold	ROC-AUC	AP	F1	Precision	Recall	Bal. Acc.
baseline	default	0.500	0.595	0.439	0.241	0.619	0.149	0.556
baseline	max F1	0.160	0.595	0.439	0.450	0.316	0.782	0.551
baseline	max bal. acc.	0.160	0.595	0.439	0.450	0.316	0.782	0.551
baseline	max recall ($p \geq 0.30$)	0.060	0.595	0.439	0.456	0.297	0.989	0.522
winsorized	default	0.500	0.583	0.401	0.147	0.875	0.080	0.538
winsorized	max F1	0.120	0.583	0.401	0.475	0.335	0.816	0.582
winsorized	max bal. acc.	0.120	0.583	0.401	0.475	0.335	0.816	0.582
winsorized	max recall ($p \geq 0.30$)	0.120	0.583	0.401	0.475	0.335	0.816	0.582

Note: This table reports winsorized-input sensitivity checks for the retained label-80 Random Forest specification. Winsorization is used as an input-sensitivity check and does not remove extreme observations from the main empirical design.

Table A4. DY-parameter sensitivity checks.

DY Setting	Specification	Threshold	ROC-AUC	AP	F1	Precision	Recall	Bal. Acc.
baseline	default	0.500	0.422	0.238	0.081	0.098	0.069	0.407
baseline	max F1	0.170	0.422	0.238	0.296	0.238	0.391	0.443
baseline	max bal. acc.	0.150	0.422	0.238	0.310	0.243	0.425	0.446
baseline	max recall ($p \geq 0.30$)	0.060	0.422	0.238	0.399	0.277	0.713	0.481
$p = 2$	default	0.500	0.618	0.396	0.170	0.474	0.103	0.529
$p = 2$	max F1	0.060	0.618	0.396	0.466	0.323	0.839	0.565
$p = 2$	max bal. acc.	0.200	0.618	0.396	0.405	0.333	0.517	0.550
$p = 2$	max recall ($p \geq 0.30$)	0.060	0.618	0.396	0.466	0.323	0.839	0.565
H = 10	default	0.500	0.440	0.244	0.068	0.083	0.057	0.401
H = 10	max F1	0.120	0.440	0.244	0.382	0.281	0.598	0.491
H = 10	max bal. acc.	0.120	0.440	0.244	0.382	0.281	0.598	0.491
H = 10	max recall ($p \geq 0.30$)	0.090	0.440	0.244	0.384	0.273	0.644	0.477
d = 0.98	default	0.500	0.410	0.236	0.114	0.151	0.092	0.442
d = 0.98	max F1	0.010	0.410	0.236	0.450	0.292	0.977	0.512
d = 0.98	max bal. acc.	0.530	0.410	0.236	0.109	0.167	0.080	0.459
d = 0.98	max recall ($p \geq 0.30$)	0.420	0.410	0.236	0.149	0.162	0.138	0.425

Note: This table reports DY-parameter sensitivity checks across alternative lag, forecast-horizon, and discount-factor settings. The evidence is reported as Appendix B sensitivity evidence rather than as main-text robustness proof.

Appendix C. Additional Validation and Model Diagnostics

Table A5. Full model-performance metrics across label settings and classifiers.

High-DSV20 Label	Classifier	Threshold	ROC-AUC	AP	F1	Precision	Recall/ Sensitivity	Balanced Accuracy
75	Majority class	—	0.500	0.558	0.000	0.000	0.000	0.500
75	Logistic regression default 0.5	0.500	0.565	0.591	0.310	0.614	0.207	0.521
75	Logistic regression balanced tuned	0.266	0.574	0.598	0.721	0.566	0.994	0.516
75	Random Forest	0.000	0.663	0.697	0.716	0.558	1.000	0.500
75	Extra Trees	0.000	0.618	0.651	0.716	0.558	1.000	0.500
75	Gradient Boosting	0.000	0.604	0.626	0.716	0.558	1.000	0.500
75	Histogram Gradient Boosting	0.000	0.565	0.567	0.716	0.558	1.000	0.500
75	SVC RBF	0.000	0.572	0.648	0.716	0.558	1.000	0.500
75	Linear SVC calibrated	0.000	0.638	0.643	0.716	0.558	1.000	0.500
75	KNN classifier	0.000	0.575	0.587	0.716	0.558	1.000	0.500
75	Gaussian NB	0.000	0.487	0.557	0.716	0.558	1.000	0.500
80	Majority class	—	0.500	0.475	0.000	0.000	0.000	0.500
80	Logistic regression default 0.5	0.500	0.517	0.458	0.154	0.368	0.097	0.473
80	Logistic regression balanced tuned	0.307	0.530	0.469	0.676	0.511	1.000	0.566
80	Random Forest	0.297	0.662	0.621	0.622	0.573	0.681	0.611
80	Extra Trees	0.000	0.615	0.557	0.644	0.475	1.000	0.500
80	Gradient Boosting	0.301	0.618	0.590	0.583	0.557	0.611	0.585
80	Histogram Gradient Boosting	0.000	0.607	0.541	0.644	0.475	1.000	0.500
80	SVC RBF	0.123	0.542	0.579	0.599	0.479	0.799	0.506
80	Linear SVC calibrated	0.249	0.569	0.500	0.682	0.518	1.000	0.579
80	KNN classifier	0.000	0.545	0.482	0.644	0.475	1.000	0.500
80	Gaussian NB	0.059	0.414	0.421	0.679	0.514	1.000	0.572
85	Majority class	—	0.500	0.399	0.000	0.000	0.000	0.500
85	Logistic regression default 0.5	0.500	0.534	0.399	0.015	0.083	0.008	0.474
85	Logistic regression balanced tuned	0.375	0.539	0.399	0.593	0.426	0.975	0.551
85	Random Forest	0.085	0.688	0.628	0.619	0.457	0.959	0.600
85	Extra Trees	0.346	0.637	0.599	0.523	0.534	0.512	0.608
85	Gradient Boosting	0.156	0.611	0.508	0.537	0.464	0.636	0.574
85	Histogram Gradient Boosting	0.000	0.630	0.583	0.546	0.434	0.736	0.549
85	SVC RBF	0.285	0.539	0.473	0.436	0.511	0.380	0.569
85	Linear SVC calibrated	0.207	0.511	0.388	0.602	0.431	1.000	0.560
85	KNN classifier	0.232	0.525	0.462	0.426	0.394	0.463	0.495
85	Gaussian NB	0.063	0.558	0.435	0.598	0.426	1.000	0.552

Note: This table reports held-out test-set ROC-AUC, average precision (AP), F1, precision, recall/sensitivity, and balanced accuracy across the 75%, 80%, and 85% high-DSV20-state labels and candidate classifiers. The table provides the full numerical results underlying Figure 6. The retained baseline result is the label-80 Random Forest specification.

Table A6. Target-horizon robustness: DSV10, DSV20, DSV30, and DSV60.

Target Horizon	Test Positive Rate	Threshold	ROC-AUC	AP	F1	Precision	Recall	Balanced Accuracy
DSV10	0.505	0.000	0.508	0.496	0.671	0.505	1.000	0.500
DSV20	0.475	0.297	0.662	0.621	0.622	0.573	0.681	0.611
DSV30	0.526	0.136	0.646	0.697	0.714	0.564	0.974	0.570
DSV60	0.643	0.000	0.518	0.726	0.782	0.643	1.000	0.500

Note: This table reports held-out test-set Random Forest results when the supervised-learning downside-risk target horizon is varied from 10 to 60 observations. The connectedness-feature construction is kept unchanged. DSV20 is retained as the baseline because it provides the most balanced diagnostic performance.

Table A7. Benchmark and ablation comparison.

Feature Set	Threshold	ROC-AUC	AP	F1	Precision	Recall	Balanced Accuracy
Connectedness-only	0.297	0.662	0.621	0.622	0.573	0.681	0.611
Total-connectedness-only	0.069	0.524	0.487	0.640	0.486	0.938	0.519
Net-connectedness-only	0.275	0.580	0.594	0.549	0.519	0.583	0.546
Returns-only	0.000	0.585	0.561	0.644	0.475	1.000	0.500
Realized-volatility-only	0.141	0.615	0.536	0.665	0.500	0.993	0.547
Connectedness + returns	0.283	0.658	0.607	0.623	0.544	0.729	0.588
Connectedness + realized volatility	0.138	0.712	0.609	0.656	0.488	1.000	0.525
Full controls	0.148	0.709	0.624	0.656	0.488	1.000	0.525

Note: This table reports held-out test-set Random Forest results under alternative feature sets for the baseline high-DSV20 label. The comparison evaluates whether connectedness-based features provide incremental diagnostic value relative to lagged returns, realized-volatility features, and combined-control specifications.

Table A8. Expanding-window validation with 20-observation embargo.

Fold	Model	Train Window	Validation Window	Test Window	Threshold	ROC-AUC	AP	F1	Precision	Recall	Balanced Accuracy
1	Random Forest	7 April 2016 to 27 September 2021	1 November 2021 to 15 November 2022	20 December 2022 to 1 December 2023	0.274	0.532	0.350	0.503	0.385	0.723	0.603
1	Logistic regression balanced	7 April 2016 to 27 September 2021	1 November 2021 to 15 November 2022	20 December 2022 to 1 December 2023	0.002	0.528	0.365	0.473	0.310	1.000	0.500
2	Random Forest	7 April 2016 to 14 January 2022	16 February 2022 to 2 March 2023	4 April 2023 to 14 March 2024	0.000	0.645	0.460	0.489	0.324	1.000	0.500
2	Logistic regression balanced	7 April 2016 to 14 January 2022	16 February 2022 to 2 March 2023	4 April 2023 to 14 March 2024	0.265	0.615	0.414	0.525	0.366	0.926	0.579
3	Random Forest	7 April 2016 to 19 May 2022	29 June 2022 to 14 June 2023	21 July 2023 to 23 July 2024	0.000	0.810	0.850	0.576	0.405	1.000	0.500
3	Logistic regression balanced	7 April 2016 to 19 May 2022	29 June 2022 to 14 June 2023	21 July 2023 to 23 July 2024	0.255	0.814	0.768	0.608	0.449	0.941	0.579
4	Random Forest	7 April 2016 to 1 September 2022	12 October 2022 to 27 September 2023	30 October 2023 to 26 November 2024	0.000	0.860	0.865	0.632	0.462	1.000	0.500
4	Logistic regression balanced	7 April 2016 to 1 September 2022	12 October 2022 to 27 September 2023	30 October 2023 to 26 November 2024	0.295	0.785	0.698	0.710	0.568	0.948	0.664
5	Random Forest	7 April 2016 to 21 December 2022	31 January 2023 to 9 January 2024	9 February 2024 to 17 March 2025	0.474	0.550	0.670	0.506	0.703	0.395	0.598
5	Logistic regression balanced	7 April 2016 to 21 December 2022	31 January 2023 to 9 January 2024	9 February 2024 to 17 March 2025	0.337	0.769	0.740	0.781	0.640	1.000	0.667
6	Random Forest	7 April 2016 to 5 April 2023	15 May 2023 to 29 April 2024	11 June 2024 to 16 July 2025	0.385	0.603	0.663	0.562	0.604	0.526	0.550
6	Logistic regression balanced	7 April 2016 to 5 April 2023	15 May 2023 to 29 April 2024	11 June 2024 to 16 July 2025	0.369	0.556	0.567	0.757	0.612	0.991	0.607
7	Random Forest	7 April 2016 to 24 July 2023	22 August 2023 to 9 September 2024	16 October 2024 to 19 November 2025	0.210	0.614	0.543	0.571	0.485	0.696	0.560
7	Logistic regression balanced	7 April 2016 to 24 July 2023	22 August 2023 to 9 September 2024	16 October 2024 to 19 November 2025	0.427	0.503	0.419	0.632	0.478	0.935	0.569

Note: This table reports the full fold-level results from expanding-window validation with a 20-observation embargo. Each fold includes Random Forest and balanced Logistic Regression evaluated on a later held-out test window. The results complement the fixed chronological split and assess whether the baseline evidence is sensitive to time-series validation design and overlapping forward labels.

References

1. Syuhada, K.; Hakim, A.; Suprijanto, D. Assessing systemic risk and connectedness among dirty and clean energy markets from the quantile and expectile perspectives. *Energy Econ.* **2024**, *129*, 107261. [CrossRef]
2. Cocca, T.; Gabauer, D.; Pomberger, S. Clean energy market connectedness and investment strategies: New evidence from DCC-GARCH R2 decomposed connectedness measures. *Energy Econ.* **2024**, *136*, 107680. [CrossRef]
3. Ji, H.; Naeem, M.; Zhang, J.; Tiwari, A.K. Dynamic dependence and spillover among the energy related ETFs: From the hedging effectiveness perspective. *Energy Econ.* **2024**, *136*, 107681. [CrossRef]
4. Kayani, U.; Hasnaoui, A.; Khan, M.; Zahoor, N.; Nawaz, F. Analyzing fossil fuel commodities' return spillovers during the Russia and Ukraine crisis in the energy market. *Energy Econ.* **2024**, *135*, 107651. [CrossRef]
5. Qi, S.; Pang, L.; Li, X.; Huang, L. The dynamic connectedness in the "carbon-energy-green finance" system: The role of climate policy uncertainty and artificial intelligence. *Energy Econ.* **2025**, *143*, 108241. [CrossRef]
6. Hu, L.; Song, M.; Wen, F.; Zhang, Y.; Zhao, Y. The impact of climate attention on risk spillover effect in energy futures markets. *Energy Econ.* **2025**, *141*, 108044. [CrossRef]
7. Xie, Q.; Bi, Y.; Xi, Y.; Xu, X. The impact of geopolitical risk on higher-order moment risk spillovers in global energy markets. *Energy Econ.* **2025**, *144*, 108292. [CrossRef]
8. Lyócsa, Š.; Todorova, N. Forecasting of clean energy market volatility: The role of oil and the technology sector. *Energy Econ.* **2024**, *132*, 107451. [CrossRef]
9. Li, Z.; Ma, F.; Lu, X. Financial risk management innovation in energy market: Evidence from a machine learning hybrid model. *Energy Econ.* **2025**, *144*, 108360. [CrossRef]
10. Tiwari, A.K.; Sharma, G.D.; Rao, A.; Hossain, M.R.; Dev, D. Unraveling the crystal ball: Machine learning models for crude oil and natural gas volatility forecasting. *Energy Econ.* **2024**, *134*, 107608. [CrossRef]
11. Tan, B.; Gan, Z.; Wu, Y. The measurement and early warning of daily financial stability index based on XGBoost and SHAP: Evidence from China. *Expert Syst. Appl.* **2023**, *227*, 120375. [CrossRef]
12. Zhang, T.; Zhu, W.; Wu, Y.; Wu, Z.; Zhang, C.; Hu, X. An explainable financial risk early warning model based on the DS-XGBoost model. *Financ. Res. Lett.* **2023**, *56*, 104045. [CrossRef]
13. Zhou, S.; Yuan, D.; Zhang, F. Multiscale systemic risk spillovers in Chinese energy market: Evidence from a tail-event driven network analysis. *Energy Econ.* **2025**, *142*, 108151. [CrossRef]
14. Lin, L.; Jiang, Y.; Zhou, Z. Asymmetric spillover and network connectedness of policy uncertainty, fossil fuel energy, and global ESG investment. *Appl. Energy* **2024**, *368*, 123432. [CrossRef]
15. Wang, J.; Mishra, S.; Sharif, A.; Chen, H. Dynamic spillover connectedness among green finance and policy uncertainty: Evidence from QVAR network approach. *Energy Econ.* **2024**, *131*, 107330. [CrossRef]
16. Asadi, M.; Roubaud, D.; Tiwari, A.K. Volatility spillovers amid crude oil, natural gas, coal, stock, and currency markets in the US and China based on time and frequency domain connectedness. *Energy Econ.* **2022**, *109*, 105961. [CrossRef]
17. Ding, A. Information and volatility linkages across energy and financial markets. *Aust. J. Manag.* **2019**, *44*, 594–613. [CrossRef]
18. Xie, Q.; Luo, C.; Cong, X.; Wang, X. Volatility connectedness and its determinants of global energy stock markets. *Econ. Syst.* **2024**, *48*, 101193. [CrossRef]
19. Chiappari, M.; Scotti, F.; Flori, A. Market responses to spillovers in the energy commodity markets: Evaluating short-term vs. long-term effects and business-as-usual vs. distressed phases. *Int. Rev. Financ. Anal.* **2024**, *96*, 103665. [CrossRef]
20. Tiwari, A.K.; Abakah, E.J.A.; Doğan, B.; Adekoya, O.B.; Wohar, M. Asymmetric spillover effects in energy markets. *Int. Rev. Econ. Financ.* **2024**, *92*, 470–502. [CrossRef]
21. Charteris, A.; Obojska, L.; Szczygielski, J.J.; Brzeszczyński, J. Energy market connectedness: A tale of two crises. *Energy Econ.* **2025**, *153*, 108787. [CrossRef]
22. Umar, M.; Farid, S.; Naeem, M.A. Time-frequency connectedness among clean-energy stocks and fossil fuel markets: Comparison between financial, oil and pandemic crisis. *Energy* **2022**, *240*, 122702. [CrossRef]
23. Ding, Q.; Huang, J.; Zhang, H. Time-frequency spillovers among carbon, fossil energy and clean energy markets: The effects of attention to climate change. *Int. Rev. Financ. Anal.* **2022**, *83*, 102222. [CrossRef]
24. Wu, R.; Li, B.; Qin, Z. Spillovers and dependency between green finance and traditional energy markets under different market conditions. *Energy Policy* **2024**, *192*, 114263. [CrossRef]
25. Kilinc-Ata, N.; Proskuryakova, L.N. Modeling Hybrid Renewable Microgrids in Remote Northern Regions: A Comparative Simulation Study. *Energies* **2025**, *18*, 5827. [CrossRef]
26. Diebold, F.X.; Yilmaz, K. Better to give than to receive: Predictive directional measurement of volatility spillovers. *Int. J. Forecast.* **2012**, *28*, 57–66. [CrossRef]
27. Diebold, F.X.; Yılmaz, K. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *J. Econom.* **2014**, *182*, 119–134. [CrossRef]

28. Baruník, J.; Křehlík, T. Measuring the frequency dynamics of financial connectedness and systemic risk. *J. Financ. Econom.* **2018**, *16*, 271–296. [CrossRef]
29. Yi, Q.; Jiang, Y. Time-frequency spillovers between carbon, fossil fuels, and clean energy markets: New insights from the TVP-VAR framework. *Energy* **2025**, *323*, 135737. [CrossRef]
30. Diebold, F.X.; Yilmaz, K. Measuring financial asset return and volatility spillovers, with application to global equity markets. *Econ. J.* **2009**, *119*, 158–171. [CrossRef]
31. Koop, G.; Pesaran, M.H.; Potter, S.M. Impulse response analysis in nonlinear multivariate models. *J. Econom.* **1996**, *74*, 119–147. [CrossRef]
32. Pesaran, H.H.; Shin, Y. Generalized impulse response analysis in linear multivariate models. *Econ. Lett.* **1998**, *58*, 17–29. [CrossRef]
33. Primiceri, G.E. Time varying structural vector autoregressions and monetary policy. *Rev. Econ. Stud.* **2005**, *72*, 821–852. [CrossRef]
34. Cogley, T.; Sargent, T.J. Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Rev. Econ. Dyn.* **2005**, *8*, 262–302. [CrossRef]
35. Antonakakis, N.; Chatziantoniou, I.; Gabauer, D. Refined measures of dynamic connectedness based on time-varying parameter vector autoregressions. *J. Risk Financ. Manag.* **2020**, *13*, 84. [CrossRef]
36. Chen, J.; Liang, Z.; Ding, Q.; Liu, Z. Quantile connectedness between energy, metal, and carbon markets. *Int. Rev. Financ. Anal.* **2022**, *83*, 102282. [CrossRef]
37. Gao, W.; Wei, J.; Zhang, H.; Zhang, H. The higher-order moments connectedness between rare earth and clean energy markets and the role of geopolitical risk: New insights from a TVP-VAR framework. *Energy* **2024**, *305*, 132280. [CrossRef]
38. Beutel, J.; List, S.; von Schweinitz, G. Does machine learning help us predict banking crises? *J. Financ. Stab.* **2019**, *45*, 100693. [CrossRef]
39. Samitas, A.; Kampouris, E.; Kenourgios, D. Machine learning as an early warning system to predict financial crisis. *Int. Rev. Financ. Anal.* **2020**, *71*, 101507. [CrossRef]
40. Bluwstein, K.; Buckmann, M.; Joseph, A.; Kapadia, S.; Şimşek, Ö. Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *J. Int. Econ.* **2023**, *145*, 103773. [CrossRef]
41. Tölö, E. Predicting systemic financial crises with recurrent neural networks. *J. Financ. Stab.* **2020**, *49*, 100746. [CrossRef]
42. Le, H.H.; Viviani, J.-L. Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. *Res. Int. Bus. Financ.* **2018**, *44*, 16–25. [CrossRef]
43. Lessmann, S.; Baesens, B.; Seow, H.-V.; Thomas, L.C. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Eur. J. Oper. Res.* **2015**, *247*, 124–136. [CrossRef]
44. Carmona, P.; Climent, F.; Momparler, A. Predicting failure in the US banking sector: An extreme gradient boosting approach. *Int. Rev. Econ. Financ.* **2019**, *61*, 304–323. [CrossRef]
45. Ren, T.; Li, S.; Zhang, S. Stock market extreme risk prediction based on machine learning: Evidence from the American market. *N. Am. J. Econ. Financ.* **2024**, *74*, 102241. [CrossRef]
46. Ghoddusi, H.; Creamer, G.G.; Rafizadeh, N. Machine learning in energy economics and finance: A review. *Energy Econ.* **2019**, *81*, 709–727. [CrossRef]
47. Guo, Y.; Zhang, S.; Liu, Y. Research on risk features and prediction of China's crude oil futures market based on machine learning. *Front. Energy Res.* **2022**, *10*, 741018. [CrossRef]
48. Luo, J.; Klein, T.; Walther, T.; Ji, Q. Forecasting realized volatility of crude oil futures prices based on machine learning. *J. Forecast.* **2024**, *43*, 1422–1446. [CrossRef]
49. Brown, I.; Mues, C. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Syst. Appl.* **2012**, *39*, 3446–3453. [CrossRef]
50. Coffinet, J.; Kien, J.-N. Detection of rare events: A machine learning toolkit with an application to banking crises. *J. Financ. Data Sci.* **2019**, *5*, 183–207. [CrossRef]
51. Sarlin, P. On policymakers' loss functions and the evaluation of early warning systems. *Econ. Lett.* **2013**, *119*, 1–7. [CrossRef]
52. Saito, T.; Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **2015**, *10*, e0118432. [CrossRef] [PubMed]
53. Tang, P.; Tang, T.; Lu, C. Predicting systemic financial risk with interpretable machine learning. *N. Am. J. Econ. Financ.* **2024**, *71*, 102088. [CrossRef]