Research Square

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

Measuring Nonlinear Spillovers among Carbon Allowances, Fuel Futures and Energy Equities with a Hybrid LSTM and TVP-VAR Estimator

Ozan Nadirgil

onadirgil@proton.me

Independent https://orcid.org/0009-0003-5584-0488

Research Article

Keywords: EU emission allowances; Spillover networks; Recurrent neural networks; Generalized variance decomposition; Energy stocks; Covariance estimation

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11575608/v1

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

Additional Declarations: The authors declare no competing interests.

Measuring Nonlinear Spillovers among Carbon Allowances, Fuel Futures and Energy Equities with a Hybrid LSTM and TVP-VAR Estimator

Ozan Nadirgil

Abstract

Stricter climate policy and repeated market turmoil have made shock transmission across carbon allowances, fuel futures and energy equities a practical concern for asset managers and regulators. This paper builds a two-part estimator. An LSTM recurrent network models the conditional expectation of each day's returns, while a covariance computed over a moving window of network errors substitutes for the exponentially weighted covariance of a VAR with time-varying parameters (TVP-VAR). Spillover indicators from this hybrid are contrasted with output from a conventional TVP-VAR for EU allowances, Brent, natural gas, coal and five equity indices covering fossil-fuel and renewable-power firms between January 2020 and August 2024. The hybrid yields a lower average system-wide spillover index (TCI; 14.9% versus 20.6%, HAC t-statistic −15.8), practically the same peak in early March 2020, and a wider gap between the pandemic crash and 2022-2023 (ratio 3.23 versus 2.12). The hierarchy of net senders and net absorbers of shocks hardly changes (Spearman $\rho = 0.93$). Nasdaq clean energy shares and Shanghai energy shares lead the network, US small-cap energy shares and coal absorb most shocks, and EU allowances are the least connected series. Because both components of the estimator change at once and its network learns from the first four fifths of the sample, the level gap is read as a property of the estimator rather than as proof of sharper stress detection.

Keywords: EU emission allowances; Spillover networks; Recurrent neural networks; Generalized variance decomposition; Energy stocks; Covariance estimation

1. Introduction

Pricing carbon is a central lever for decarbonizing the economy. By charging emitters for every tonne of CO_2, an allowance scheme lowers the relative cost of low-carbon technology and hastens the retirement of coal and gas capacity. A binding allowance price should therefore co-move with fuel prices, with the valuation

of fossil-fuel producers and with the prospects of renewable energy firms, so the way shocks spread through these assets concerns both investors who hold them side by side and regulators who read the allowance price as a policy signal. Attention to such links has grown since emission allowances, clean energy shares and energy commodities became common building blocks of diversified and sustainability-oriented portfolios.

Several economic channels connect these assets. Fuel switching from coal to gas in power generation alters the demand for allowances, so fuel price movements pass into allowance prices [1,2]. Oil prices interact with allowances through industrial activity and portfolio rebalancing, with consequences for hedging decisions [3,4]. Allowance prices and renewable-firm valuations both reflect the expected stringency of future climate rules [5], and spillovers have been found to run mainly from clean energy shares towards allowances [6]. Broader evidence confirms that carbon prices are tied to energy and financial assets in ways that shift across regimes and frequencies [7,8].

Most of this evidence rests on the spillover index built on variance decompositions [9-11], which attributes the forecast uncertainty of every market to shocks arising elsewhere in the system. In its time-varying form [12], this index is obtained from a VAR whose coefficients and error covariance are revised recursively through Kalman filtering with forgetting factors [13], so that the sample is used in full without any estimation window. At each date, though, its conditional expectation stays linear in lagged returns. Daily returns on allowances and fuels display heavy tails, volatility clustering and possibly nonlinear serial dependence [14], which has encouraged combining spillover measures with neural networks [15-17]. Research on return prediction, by contrast, shows that machine learning models extract only a small predictable component from daily financial returns and overfit easily [18-20], so any improvement credited to a neural filter must be weighed against these risks.

The estimator proposed here, called LSTM-TVP-VAR in the figures and simply the hybrid below, departs from the standard model in two respects. First, the VAR's linear forecast gives way to an LSTM recurrent network [21] that predicts tomorrow's returns nonlinearly. Second, the covariance of forecast errors that enters the generalized variance decomposition is computed over a moving window of network errors

rather than by exponential smoothing. All hyperparameters, window length included, are selected through a grid search with chronological cross-validation over the earliest 80% of observations. The estimator is taken to daily data on EU allowances (EUA), Brent, natural gas, coal and five energy share indices between January 2020 and August 2024, a span containing the pandemic crash, Europe's gas crisis and Russia's invasion of Ukraine, and its output is set against a conventional TVP-VAR.

Three findings stand out. The hybrid index is lower on average and smoother from day to day than its linear counterpart, while both reach nearly identical maxima in early March 2020. The ranking of net senders and net absorbers is almost unchanged across estimators, with Nasdaq clean energy shares and Shanghai energy shares leading and EUA as the least connected market. Finally, the level gap is statistically robust, yet because the hybrid modifies its mean equation and its covariance at once while its network is fitted on most observations, it cannot by itself be credited to nonlinearity. Section 2 surveys related work, Section 3 presents data and estimators, Section 4 discusses the empirical evidence, while Section 5 offers concluding remarks.

2. Related Literature and Research Gap

A first body of research asks what drives carbon prices. Energy prices, equity returns, fund flows, weather and broader commodity prices all appear among the drivers [22,23], and explainable boosting models reveal strong interactions among them [22]. Cross-market work began with linear systems. A VAR covering allowance prices and share prices of large European power utilities reports sizeable information flows in both directions [24], and quantile VAR estimates for crude oil, allowances and grain futures show that aggregate spillovers grow sharply when markets are under strain [25]. For China, a time-varying VAR finds that coal prices and uncertainty about climate policy both feed into the national allowance price [26], while spillover indices combined with copulas for iron ore, carbon, scrap, freight rates and steel shares reveal linkages that react to Beijing's campaign against excess steel capacity [27]. Earlier European evidence documents copula dependence between allowances and oil [3], volatility transmission from fossil fuels to allowances [1] and information flows in both directions linking carbon with energy prices [2].

A second body relies on volatility models and quantile techniques. Volatility transmission among ESG indices, green-bond indices, new-energy shares and allowances is asymmetric; ESG indices together with green-bond indices act as principal risk sources, while new-energy shares and allowances absorb risk [28]. Analysing policy uncertainty alongside energy and allowance prices across quantiles and frequencies, another study finds that long-horizon components prevail overall, whereas medium-horizon spillovers gain weight in extreme states [29]. Quantile VAR estimates spanning the pandemic and Russia's invasion of Ukraine show that links among fossil fuels, power and allowances change with market states, with oil and carbon acting as core senders and power and coal as absorbers [30]. Copula and frequency-domain evidence for renewable energy shares and allowances likewise describes dependence of a nonlinear kind, concentrated at particular horizons [6], and volatility links between EU allowances and commodity and financial assets vary with US policy uncertainty [8].

A third body refines the spillover framework itself. The original index relied on rolling VARs with orthogonalized shocks [9]; generalized variance decompositions [31,32] removed its dependence on variable ordering [10], and a network reading followed [11]. Its version with time-varying parameters [12], together with pairwise, frequency-domain, quantile and joint extensions [33-37], has become a routine tool. For Europe, time-varying estimates indicate that Brexit, the Green Deal and the pandemic altered relations linking allowances, equities and renewables, while long-run links stay close to zero [38]. In China, roughly seventy percent of all connectedness linking international oil with domestic new-energy supply chains originates inside those chains [39], while bank shares from G-17 economies are bound most tightly to climate, renewable and technology factors in crises [40]. Optimized time-varying VAR extensions applied to technology and green shares, exchange-traded funds, commodities and crypto and fiat currencies point to sizeable hedging potential for commodities and currencies and a dominant short-horizon component [41].

A fourth and more recent body marries econometric models to deep neural networks. Frequency-domain TVP-VAR measures have been coupled with LSTM networks to study technology and renewable shares over 2018-2023, revealing weak links between artificial intelligence firms and renewable firms and a prevalence

of short-horizon spillovers [15]. LSTM forecasts added to a DCC-GARCH model sharpen predicted correlations linking Latin American equities to US stocks during the pandemic [16]. A graph-based neural network fed by time-varying VAR output suggests that linear benchmarks understate how foreign oil and monetary disturbances reach Gulf financial markets [17], and a GARCH-LSTM design with wavelet-based spillovers applied to US conventional and green equities finds that climate risk spreads nonlinearly and unevenly across horizons [42]. Table 1 summarizes these studies.

Ref.	Year	Method and sample	Main finding
[24]	2019	VAR; EUA prices and returns on large European utility shares	Carbon prices exchange substantial information with power-sector share returns.
[7]	2020	Spillover index; EUA alongside energy, equity, bond and non-energy commodity markets	Carbon links mainly with equities, energy and non-energy commodities, hardly with bonds; carbon volatility shows three breaks.
[6]	2021	Frequency-domain spillovers plus copulas; EUA with renewable-energy share indices	EUA absorbs spillovers from renewable energy shares; dependence is nonlinear.
[27]	2021	Spillover index plus copulas; iron ore, carbon, scrap, freight rates, Chinese steel shares	Links react to Chinese overcapacity policy and to the pandemic.
[8]	2021	Volatility spillovers in time and frequency domains; EU carbon with commodity and financial markets	Carbon volatility links differ by frequency and respond to US policy uncertainty.
[23]	2022	Wavelet-based quantile regressions; carbon prices versus green-bond prices	Strong two-way dependence, above all in extreme quantiles.
[25]	2023	Quantile VAR; crude oil, allowances, grain futures	Aggregate spillovers far exceed normal-state levels in extreme states.
[26]	2023	Time-varying VAR; price of coal, uncertainty about climate policy, Chinese allowance price	Coal and policy uncertainty both transmit to carbon, more strongly near policy announcements.
[38]	2023	Time-varying VAR; European allowances, equities, renewables	Carbon moves renewables over short horizons; long-horizon links are negligible.
[22]	2024	Gradient boosting with SHAP; drivers of Chinese allowance prices	Macro variables, energy prices and policy uncertainty dominate, interacting nonlinearly.
[28]	2024	Asymmetric transmission of volatility; ESG indices, allowance prices, new-energy shares, green-bond indices	ESG plus green-bond indices transmit risk; allowance plus new-energy indices absorb it.
[29]	2024	Quantile spillovers across frequencies; carbon and fuel prices with uncertainty about policy	Long-horizon parts dominate; medium horizons gain weight in extreme states.
[39]	2024	Time-varying VAR in time and frequency; new-energy supply chain in China and global crude	Some 70% of connectedness arises inside the chain; oil acts mostly over medium frequencies.
[30]	2024	Quantile VAR; fossil fuels, power, allowances around the pandemic and Russia's invasion of Ukraine	Spillovers vary by market state; oil and carbon send, power and coal receive.
[40]	2024	Time-varying VAR; innovation, renewables, climate change, G-17 bank shares	Links tighten in crises; innovation indices lead.
[15]	2024	Frequency-domain time-varying VAR combined with LSTM; shares in artificial intelligence, big data, quantum computing and renewables	Weak AI-renewable links; short-horizon spillovers prevail.
[16]	2024	DCC-GARCH plus LSTM; Latin American and US equities during the pandemic	LSTM forecasts sharpen predicted dynamic correlations.
[17]	2025	Time-varying VAR feeding a graph-based causal shock model; Gulf markets facing foreign oil and monetary shocks	Nonlinear and causal transmission is understated by linear benchmarks.

Ref.	Year	Method and sample	Main finding
[41]	2025	Optimized time-varying VAR extensions with Granger causality; green and technology shares, ETFs, commodities, crypto and fiat currencies	Commodities, crypto and fiat currencies offer hedging potential; short-horizon transmission dominates.
[42]	2026	GARCH-LSTM with wavelet spillovers; US conventional and green equities, climate risk indices	Climate risk travels nonlinearly and differently across horizons.

Table 1 Overview of related studies

Taken together, this literature shows that cross-asset linkages attract sustained attention but also reveals clear methodological limits. Full-sample VARs impose constant coefficients and linear dynamics. GARCH-type models describe conditional heteroskedasticity well, yet they need additional structure to deliver directional, pairwise spillovers in systems with many assets. Time-varying VAR connectedness lets coefficients and covariances evolve without any rolling estimation, which explains its popularity, but its expected-return equation is linear at every date and its covariance follows exponentially weighted updating governed by a decay parameter that researchers fix in advance. Hybrid designs address nonlinearity, yet most report gains inside their estimation samples and rarely disentangle what the nonlinear mean adds from what the altered covariance estimator adds. To the author's knowledge, no earlier study applies such a hybrid jointly to emission allowances, fossil fuels, fossil-fuel shares and low-carbon shares.

Four contributions follow. First, LSTM-based forecasts take the place of the linear expected-return equation in a time-varying VAR, and their errors feed the generalized variance decomposition. Second, error covariance is estimated over a moving window of these errors, with window length chosen by Gaussian likelihood on held-out folds rather than by a decay constant fixed in advance. Third, every tuning parameter comes out of an exhaustive grid with chronological cross-validation confined to training observations, and the full grid appears in Table 7. Fourth, EUA, oil, gas, coal, fossil-fuel shares and low-carbon shares are analysed jointly over 2020-2024, and the paper documents which features of the spillover index change with the estimator and which remain intact.

3. Data, Estimators and Evaluation Design

3.1. Sample and Variables

The sample combines daily futures prices for EUA, Brent, coal and natural gas with daily closes of six widely followed fossil-fuel and low-carbon equity indices (Table 2). All series, covering 02.01.2020 through 30.08.2024, were downloaded via Investing.com and aligned on shared trading dates, which leaves 1,202 observations per series. Table 2 describes price levels, so its CV (standard deviation over mean) measures how widely each price ranged around its sample mean rather than how volatile daily returns were. By this yardstick, coal (64.9%) and gas (51.8%) moved over the widest relative range, reflecting the fuel price surge of 2021-2022, whereas both broad low-carbon indices, MSCIC (22.0%) and SPC (24.5%), moved over the narrowest. Kurtosis appears in excess form, so negative values indicate level distributions flatter than a normal one.

Series	Code	N	Mean	SD	Median	Skewness	Excess kurtosis	CV (%)
EU emission allowances	EUA	1,202	62.02	23.56	68.50	−0.48	−1.15	38.0
S&P SmallCap 600 Energy	SP600	1,202	294.21	99.52	317.33	−0.60	−0.87	33.8
Shanghai Stock Exchange Energy	SSEE	1,202	1,561.89	370.94	1,673.56	0.03	−1.26	23.7
MSCI World Energy	MSCIE	1,202	199.57	51.97	218.99	−0.42	−1.23	26.0
Nasdaq CELS index (green energy)	NASC	1,202	646.98	204.50	660.64	0.06	−0.75	31.6
S&P Global Clean Energy	SPC	1,202	1,182.31	290.10	1,206.32	0.32	0.07	24.5
MSCI ACWI Clean Energy	MSCIC	1,202	204.95	45.16	220.77	−0.52	−0.98	22.0
Brent crude futures	OIL	1,202	75.10	20.95	79.32	−0.42	−0.06	27.9
Natural gas futures	GAS	1,202	3.66	1.90	2.88	1.35	0.87	51.8
Coal futures	COAL	1,202	175.43	113.89	141.05	1.05	−0.11	64.9

Table 2 Price-level statistics of daily data, Investing.com

Once calendars are aligned, prices are converted into continuously compounded daily returns. Stationarity is examined with two complementary tests, namely Dickey-Fuller regressions in augmented form (ADF) [43], whose null is nonstationarity, and a KPSS test [44], which reverses both hypotheses. According to Table 3, every price series fails to reject nonstationarity under ADF and rejects stationarity under KPSS, whereas every return series rejects nonstationarity and is compatible with stationarity at 5%. All later steps therefore work with log returns.

Series	ADF, prices	KPSS, prices	ADF, returns	KPSS, returns	Integration order
EUA	−1.730	3.550	−37.194	0.258	I(1)
SP600	−1.191	4.196	−35.213	0.062	I(1)
SSEE	−0.888	4.566	−35.322	0.075	I(1)
MSCIE	−0.977	4.577	−33.675	0.108	I(1)
NASC	−1.873	0.976	−35.647	0.406	I(1)
SPC	−1.784	1.063	−32.217	0.243	I(1)
MSCIC	−1.151	4.508	−32.921	0.101	I(1)
OIL	−1.397	2.706	−31.987	0.088	I(1)
GAS	−1.805	0.861	−36.290	0.187	I(1)
COAL	−1.376	1.430	−33.279	0.288	I(1)

Table 3 Diagnostics for unit roots and stationarity

Note: ADF takes nonstationarity as its null, KPSS takes stationarity. Verdicts use 5% significance. Prices are I(1) and returns I(0) for every series.

Next, BDS statistics [45] are computed for each return series, with embedding dimension m ranging from 2 to 6, to check whether observations are i.i.d. In Table 4, nine of ten series reject this hypothesis with p-values below 0.01 for every dimension. NASC is the only borderline case. Its statistic is significant only at 10% for $m = 2$ but exceeds the 1% critical value once $m \geq 3$, so its returns are also judged serially dependent. BDS, however, tests the i.i.d. hypothesis rather than any specific nonlinear model. Rejections can stem from residual linear autocorrelation, from conditional heteroskedasticity or from nonlinearity in expected returns [45,46], and here the statistic is computed on raw returns rather than on residuals from a linear filter. These results therefore justify flexible models able to accommodate nonlinear dependence, but they do not by themselves establish that expected returns depend nonlinearly on past returns.

Series	Dimension 2	Dimension 3	Dimension 4	Dimension 5	Dimension 6	Verdict (5%)
EUA	2.5855*** (0.0097)	3.4249*** (0.0006)	3.7846*** (0.0002)	4.2760*** (<0.0001)	4.8530*** (<0.0001)	Reject
SP600	3.3157*** (0.0009)	5.7882*** (<0.0001)	9.7631*** (<0.0001)	13.3647*** (<0.0001)	18.7109*** (<0.0001)	Reject
SSEE	4.2658*** (<0.0001)	5.2658*** (<0.0001)	6.7609*** (<0.0001)	7.9738*** (<0.0001)	9.9105*** (<0.0001)	Reject
MSCIE	4.9230*** (<0.0001)	6.2407*** (<0.0001)	8.2612*** (<0.0001)	10.5484*** (<0.0001)	13.1381*** (<0.0001)	Reject
NASC	1.7785* (0.0753)	3.6755*** (0.0002)	6.3033*** (<0.0001)	7.8912*** (<0.0001)	10.5787*** (<0.0001)	Reject from dimension 3
SPC	3.7188*** (0.0002)	4.6823*** (<0.0001)	4.7875*** (<0.0001)	5.2723*** (<0.0001)	5.5407*** (<0.0001)	Reject
MSCIC	5.1951*** (<0.0001)	6.5516*** (<0.0001)	8.3903*** (<0.0001)	10.4984*** (<0.0001)	12.8744*** (<0.0001)	Reject
OIL	6.5611*** (<0.0001)	8.8913*** (<0.0001)	10.9862*** (<0.0001)	12.2925*** (<0.0001)	14.3829*** (<0.0001)	Reject

Series	Dimension 2	Dimension 3	Dimension 4	Dimension 5	Dimension 6	Verdict (5%)
GAS	2.4364** (0.0148)	4.0953*** (<0.0001)	6.2861*** (<0.0001)	8.7013*** (<0.0001)	12.4046*** (<0.0001)	Reject
COAL	10.6225*** (<0.0001)	11.7215*** (<0.0001)	12.4958*** (<0.0001)	13.7551*** (<0.0001)	15.4523*** (<0.0001)	Reject

Table 4 BDS statistics by series and embedding dimension

Note: BDS z-statistics, with p-values shown in parentheses. Three asterisks denote significance below 1%, two below 5% and one below 10%. Rejection signals serial dependence of some form, whether linear or nonlinear.

Pairwise correlations appear in Table 5 and variance-inflation factors in Table 6. MSCIE and MSCIC are almost perfectly correlated ($r = 0.995$), each with a VIF near 105, whereas all remaining VIFs lie between 1.01 and 1.07 and all remaining correlations stay below 0.20 in absolute terms. Multicollinearity does not bias least-squares coefficients, but it inflates their sampling variance; moreover, within variance decompositions two near-duplicate series split what is essentially one shock between them, which raises measured connectedness mechanically. MSCIC is therefore dropped and MSCIE, whose coverage is broader, is kept, so nine series remain.

	EUA	SP600	SSEE	MSCIE	NASC	SPC	MSCIC	OIL	GAS	COAL
EUA	1	.006	.001	−.019	−.024	.057	−.015	−.025	.014	−.011
SP600	.006	1	.028	.014	.135	.017	.013	.037	.028	.023
SSEE	.001	.028	1	.058	.007	.050	.062	−.067	−.035	−.025
MSCIE	−.019	.014	.058	1	.017	.176	.995	−.087	.003	.001
NASC	−.024	.135	.007	.017	1	.047	.024	.033	−.020	.049
SPC	.057	.017	.050	.176	.047	1	.191	−.062	.039	.014
MSCIC	−.015	.013	.062	.995	.024	.191	1	−.082	.008	−.005
OIL	−.025	.037	−.067	−.087	.033	−.062	−.082	1	−.036	.138
GAS	.014	.028	−.035	.003	−.020	.039	.008	−.036	1	.036
COAL	−.011	.023	−.025	.001	.049	.014	−.005	.138	.036	1

Table 5 Correlations between return series

Note: Pearson coefficients computed on returns in logs; leading zeros are omitted.

Series	R^2 against remaining returns	Tolerance ($1 - R^2$)	VIF	Collinearity
EUA	0.010	0.990	1.01	Negligible
SP600	0.022	0.978	1.02	Negligible
SSEE	0.013	0.987	1.01	Negligible
MSCIE	0.991	0.009	105.41	Severe
NASC	0.029	0.971	1.03	Negligible
SPC	0.062	0.938	1.07	Negligible
MSCIC	0.991	0.009	105.95	Severe

Series	R^2 against remaining returns	Tolerance ($1 - R^2$)	VIF	Collinearity
OIL	0.040	0.960	1.04	Negligible
GAS	0.009	0.991	1.01	Negligible
COAL	0.029	0.971	1.03	Negligible

Table 6 Variance-inflation diagnostics by series

Note: Mean VIF equals 21.96 across all ten series and 1.03 once MSCIE and MSCIC are left out.

3.2. Benchmark Connectedness Model

TVP-VAR-JC, the benchmark, follows the spillover approach with time-varying parameters [12]. Coefficients and error covariance of a VAR evolve over time and are estimated through Kalman filtering, where forgetting factors replace explicit modelling of state noise [13]. Because coefficients are revised recursively at every date, no estimation window is needed and no observation is discarded, which is the main practical advantage over the rolling regressions behind the original index [10]. On any given date, however, expected returns are linear in lagged returns, while error covariance averages past outer products of forecast errors with exponentially declining weights. Results therefore depend on lag order, on the forgetting parameter λ, on the decay parameter κ and on the prior that starts the filter. With one lag, the model reads

$$y_t = \Phi_t y_{t-1} + u_t, u_t \sim N(0, \Sigma_t) \quad (1)$$

$$vec(\Phi_t) = vec(\Phi_{t-1}) + v_t, v_t \sim N(0, Q_t) \quad (2)$$

$$\Sigma_t = \kappa \Sigma_{t-1} + (1-\kappa)\hat{u}_t \hat{u}_t' \quad (3)$$

Here y_t denotes the $k \times 1$ return vector, Φ_t is a k-by-k coefficient matrix, u_t collects errors with covariance Σ_t, and $\hat{u}_t$ denotes the prediction error formed with predicted coefficients. Forgetting enters through state covariance, $Q_t = (\lambda^{-1} - 1) P_{t-1|t-1}$, where $P_{t-1|t-1}$ is the filtered covariance of $vec(\Phi_{t-1})$; values of λ and κ near one discount old information slowly. Inverting the VAR yields its moving-average representation

$$y_t = \sum_{\ell=0}^{\infty} \Psi_{\ell,t} u_{t-\ell}, \Psi_{0,t} = I_k, \Psi_{\ell,t} = \Phi_t^{\ell} \quad (4)$$

Generalized impulse responses [31,32] then deliver what fraction of i's H-step prediction-error variance a shock in market j explains,

$$\theta_{ij,t}(H)=\frac{\sigma_{jj,t}^{-1}\sum_{\ell=0}^{H-1}\left(e_i'\Psi_{\ell,t}\Sigma_t e_j\right)^2}{\sum_{\ell=0}^{H-1}e_i'\Psi_{\ell,t}\Sigma_t\Psi_{\ell,t}'e_i} \tag{5}$$

where e_i is column i of I_k and $\sigma_{jj,t}$ is diagonal element j of Σ_t. Since generalized shocks are correlated, shares within each row do not sum to one and are rescaled,

$$\tilde{\theta}_{ij,t}(H)=\frac{\theta_{ij,t}(H)}{\sum_{j=1}^{k}\theta_{ij,t}(H)},\ \sum_{j=1}^{k}\tilde{\theta}_{ij,t}(H)=1,\ \sum_{i,j=1}^{k}\tilde{\theta}_{ij,t}(H)=k \tag{6}$$

Directional, net, system-wide and pairwise measures follow from these rescaled shares:

$$FROM_{i,t}=100\sum_{j=1,j\neq i}^{k}\tilde{\theta}_{ij,t}(H) \tag{7}$$

$$TO_{j,t}=100\sum_{i=1,i\neq j}^{k}\tilde{\theta}_{ij,t}(H) \tag{8}$$

$$NET_{i,t}=TO_{i,t}-FROM_{i,t} \tag{9}$$

$$TCI_t=\frac{100}{k}\sum_{i=1}^{k}\sum_{j=1,j\neq i}^{k}\tilde{\theta}_{ij,t}(H) \tag{10}$$

$$PCI_{ij,t}=2\frac{\tilde{\theta}_{ij,t}(H)+\tilde{\theta}_{ji,t}(H)}{\tilde{\theta}_{ii,t}(H)+\tilde{\theta}_{ij,t}(H)+\tilde{\theta}_{ji,t}(H)+\tilde{\theta}_{jj,t}(H)} \tag{11}$$

FROM records which part of *i*'s prediction-error variance originates in other markets, TO records how much market *j* passes on to others, and NET subtracts FROM from TO, so a positive value marks a net sender. TCI averages, over markets, that part of prediction-error variance crossing market boundaries. By construction it cannot exceed $100-100/k$, which equals 88.9% for nine series, and row normalization tends to raise measured connectedness when shocks are strongly correlated [47]. PCI lies between 0 and 1 and gauges how strong the bilateral link between *i* and *j* is [33].

3.3. Hybrid Estimator

LSTM networks belong to the recurrent family and were designed to retain information across long sequences [21]. Each cell couples gates for input, output and forgetting with a memory state carried from one step to the next, so the network can represent nonlinear, state-dependent responses to past inputs. These properties have made LSTMs a favourite deep learning tool for financial time series [18,20]. The same

flexibility becomes a liability in this setting. Daily returns carry little signal relative to noise, and heavily parameterized networks overfit unless architecture and training are tuned carefully [19].

Compared with TVP-VAR-JC, the hybrid changes two elements. The first change concerns expected returns, which an LSTM now produces in place of linear VAR algebra,

$$\hat{y}_t = f\left(y_{t-1}, y_{t-2}, \ldots, y_{t-p}; \omega\right) \tag{12}$$

with f denoting the trained network, ω its weights and p how many lagged days enter as inputs. Its forecast errors are

$$\eta_t = y_t - \hat{y}_t \tag{13}$$

The second change concerns covariance. Replacing the exponentially weighted recursion of Eq. (3), the hybrid uses a sample covariance over the latest W network errors,

$$\Sigma_t^{(W)} = \frac{1}{W-1} \sum_{\tau=0}^{W-1} \left(\eta_{t-\tau} - \acute{\eta}_t\right)\left(\eta_{t-\tau} - \acute{\eta}_t\right)', \tag{14}$$

where $\acute{\eta}_t$ averages errors inside the window. Because both components change together, any gap separating both spillover indices reflects their joint effect.

Network weights ω are estimated by back-propagation with early stopping [48], whereas hyperparameters (batch size, input length p, hidden units, optimizer step size) are picked from a grid. Every candidate ζ_m in grid Z is trained and evaluated through five chronological folds, each fold training on earlier data and validating on the block that follows [49,50]. Only training observations, i.e. the first four fifths of the sample, are involved. The selected candidate minimizes average validation squared error,

$$MSE^{CV}\left(\zeta_m\right) = \frac{1}{F} \sum_{f=1}^{F} \frac{1}{\left|V_f\right|} \sum_{t \in V_f} \left\| y_t - \hat{y}_t\left(\zeta_m\right) \right\|^2 \tag{15}$$

where $F = 5$ is the fold count and V_f holds fold f's validation dates. Window length W is tuned analogously. Each candidate W is judged by how well a covariance from those W errors preceding date t explains, under a Gaussian density, the error realized at t; the value with the smallest average negative log density is retained,

$$NLL(W) = \frac{1}{F}\sum_{f=1}^{F}\frac{1}{|V_f|}\sum_{t\in V_f}\frac{1}{2}\left[k\ln(2\pi) + \ln\left|\Sigma_{t-1}^{(W)}\right| + \eta_t{}'\left(\Sigma_{t-1}^{(W)}\right)^{-1}\eta_t\right] \tag{16}$$

Short windows give noisy covariance estimates, whereas long windows adapt slowly when volatility and co-movement shift. This criterion balances both effects and picks $W = 130$ trading days.

Hybrid connectedness then follows from Eqs. (5) to (11), with $\Sigma_t^{(W)}$ replacing exponentially weighted covariance and with moving-average coefficients implied by the hybrid specification. Every later step, from row normalization in Eq. (6) to pairwise links in Eq. (11), is identical across estimators.

Fitting takes place only once, on training observations. Network errors for roughly the first four fifths of the sample, up to about the third quarter of 2023, are therefore fitted values, and only errors for the final fifth are genuine forecasts on unseen data. Since connectedness is reported for every date, the hybrid index is a descriptive, full-information measure rather than one that could have been computed as events unfolded. Moreover, because data begin in January 2020, a complete window of 130 errors exists only from early July 2020, so readings for the first months of both indices rest on short estimation histories.

Figure 1 outlines the workflow. Data are first screened for distributional features, unit roots, serial dependence and collinearity. Returns are then split chronologically into training (80%) and test (20%) partitions, hyperparameters of both estimators are tuned on training observations, two sets of spillover measures follow, and statistical tests compare both indices.

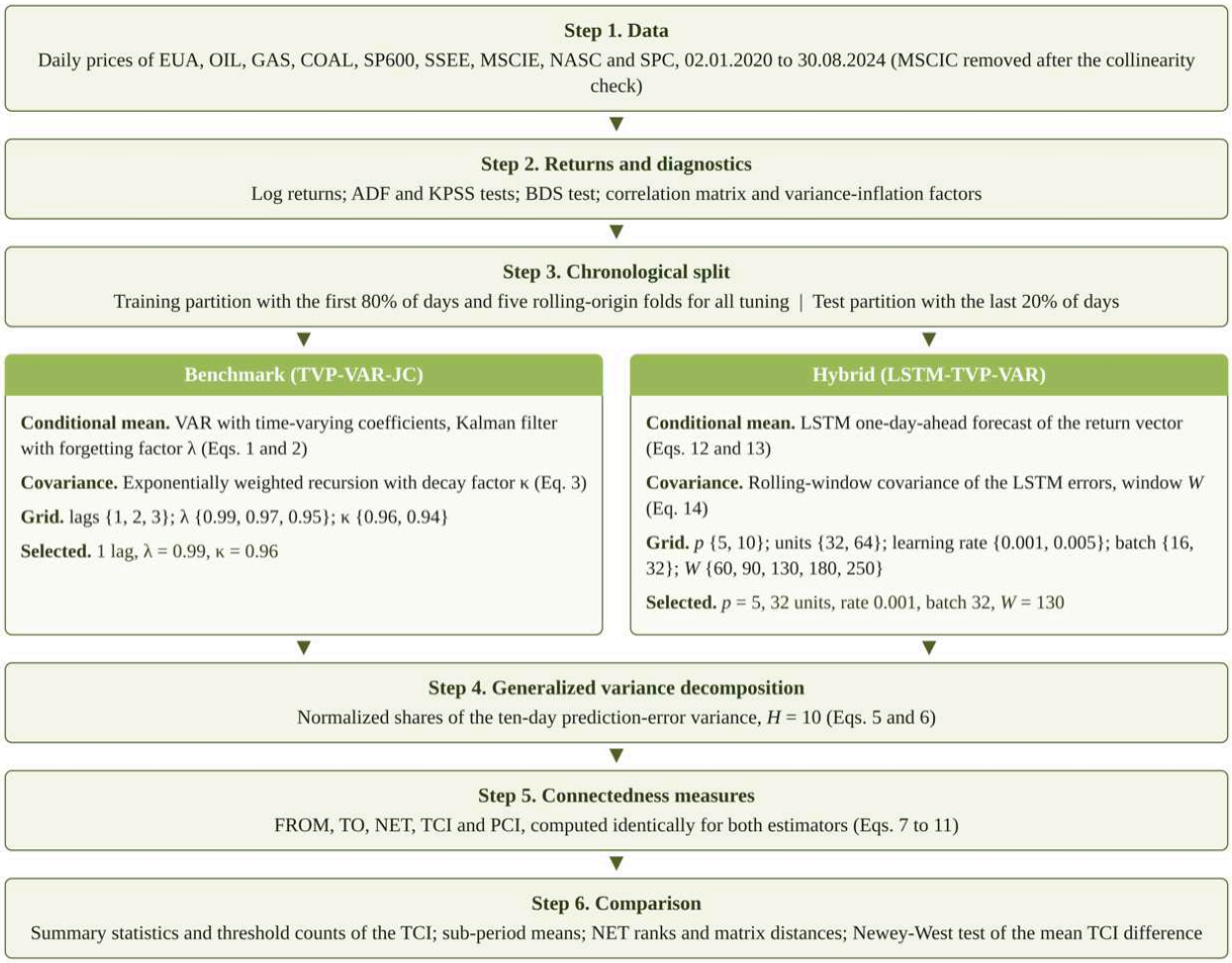


Figure 1 Workflow comparing benchmark TVP-VAR-JC with the hybrid estimator

3.4. Tuning and Comparison

Table 7 lists the full grid, selected values and the criterion behind each parameter. An exhaustive grid was preferred because, with few candidates, it tries every combination, yields deterministic and fully reportable outcomes, and fits naturally with chronological validation folds in which no candidate ever sees later data. In larger search spaces random search is usually more efficient [51], but with sixteen network configurations and five window lengths exhaustive search is entirely feasible. For TVP-VAR-JC, lag order ranges from one to three, λ takes 0.95, 0.97 or 0.99, while κ takes 0.94 or 0.96; mean squared one-day prediction error on identical folds selects one lag, λ at 0.99 together with κ at 0.96.

The chosen network uses five lagged days, 32 hidden units, step size 0.001 and batches of 32 observations. Its validation MSE equals 0.914 on a standardized scale, against 1.000 when tomorrow's return

is predicted by its training-sample average. A drop of 8.6% in squared prediction error counts as large by daily forecasting standards [19], and it arises within the training period, so it is not read as evidence of exploitable predictability. Criteria in Table 7 are also measured in different units for the two stages and cannot be compared across rows.

Stage	Parameter	Candidates	Chosen	Method	Criterion	Best score
TVP-VAR-JC	Lag order	One to three	1	Grid search	One-day MSE	0.000805
TVP-VAR-JC	Forgetting λ	0.95 or 0.97 or 0.99	0.99	Grid search	One-day MSE	0.000805
TVP-VAR-JC	Decay κ	0.94 or 0.96	0.96	Grid search	One-day MSE	0.000805
LSTM	Input length p (lagged days)	5 or 10	5	Grid search	One-day MSE, standardized	0.9140
LSTM	Hidden units	32 or 64	32	Grid search	One-day MSE, standardized	0.9140
LSTM	Step size	0.001 or 0.005	0.001	Grid search	One-day MSE, standardized	0.9140
LSTM	Batch size	16 or 32	32	Grid search	One-day MSE, standardized	0.9140
LSTM	Epochs	Upper limit 50	≤ 50	Early stopping	Loss plateau	n/a
Hybrid covariance	Window length W (days)	250, 180, 130, 90 or 60	130	Grid search	Gaussian NLL, held-out folds	−20.6874

Table 7 Tuning grid and chosen settings

Note: Every search uses five chronological folds with expanding training blocks inside the training partition; no observation from the 20% test partition enters selection. Parameters tuned jointly share one best score. Scores of both stages are on different scales (raw versus standardized returns).

Four comparisons are made between both spillover indices. They cover summary statistics and threshold exceedances of each TCI, average levels in two sub-periods, agreement between net positions and between average spillover matrices, and a formal test of whether mean TCI differs. Given strong persistence in each TCI, this test uses a HAC long-run variance in the manner of Newey and West [52]. Truncation is set at six lags, near the familiar rule of thumb $4(T/100)^{2/9}$ [53], with repetitions for every lag between 5 and 44.

4. Results

The evidence unfolds in three steps, covering the spillover structure for TVP-VAR-JC (Section 4.1), the structure produced by the hybrid (Section 4.2), and a comparison of both indices (Section 4.3), which ends with the HAC test.

4.1. Benchmark Spillover Structure

For TVP-VAR-JC, estimation uses one lag, λ at 0.99 together with κ at 0.96, and a ten-day horizon. Table 8 gives its normalized decomposition averaged over time, where row *i*, column *j* holds what fraction of *i*'s ten-day prediction-error variance stems from shocks in market *j*, and Figure 2 draws average bilateral links as a network. Average TCI equals 20.57%. Put differently, about one fifth of a typical market's ten-day prediction-error variance traces back to shocks from the remaining eight markets, and roughly four fifths to its own shocks.

In net terms, NASC (5.45 points) and SSEE (2.47) send most, followed at some distance by OIL (1.06), MSCIE (0.58) and SPC (0.57). COAL (−5.18) and SP600 (−4.42) absorb most, while EUA (−0.39) and GAS (−0.15) sit near balance. Gross figures add nuance. MSCIE and NASC have the highest TO values (24.8% each), yet MSCIE also receives almost as much as it sends, which leaves it with a small net position. Two channels dominate bilaterally, namely the two-way link joining MSCIE and SPC (7.67% and 7.57%) and the flow from NASC into SP600 (8.72%), while OIL keeps moderate links with most other markets. EUA receives least from the system (16.5%) and transmits second-least (16.1%).

Receiver \ Source	EUA	SP600	SSEE	MSCIE	NASC	SPC	OIL	GAS	COAL	FROM
EUA		1.83	1.99	2.38	2.66	1.79	2.36	2.03	1.50	16.53
SP600	2.09		2.54	2.41	8.72	2.12	2.98	2.12	1.47	24.45
SSEE	1.88	2.40		2.26	2.12	1.92	3.17	1.90	2.07	17.72
MSCIE	2.31	2.55	2.67		2.05	7.67	3.33	2.00	1.65	24.23
NASC	1.64	3.80	2.12	2.42		2.10	2.67	2.64	1.95	19.35
SPC	1.77	2.70	2.03	7.57	2.29		2.42	2.16	1.58	22.52
OIL	2.33	2.40	3.86	3.43	2.38	2.46		2.00	3.45	22.30
GAS	2.01	2.20	2.32	1.81	2.30	2.39	2.78		1.69	17.51
COAL	2.11	2.15	2.65	2.53	2.28	2.64	3.65	2.52		20.54
TO	**16.14**	**20.03**	**20.19**	**24.81**	**24.80**	**23.09**	**23.37**	**17.36**	**15.36**	
NET	**−0.39**	**−4.42**	**2.47**	**0.58**	**5.45**	**0.57**	**1.06**	**−0.15**	**−5.18**	**TCI 20.57**

Table 8 Time-averaged spillover matrix, TVP-VAR-JC (%)

Note: Row i, column j gives what percentage of i's ten-day prediction-error variance, averaged over time, is attributed to shocks in market j. Row sums give FROM, column sums TO, and NET = TO − FROM. Own shares are omitted.

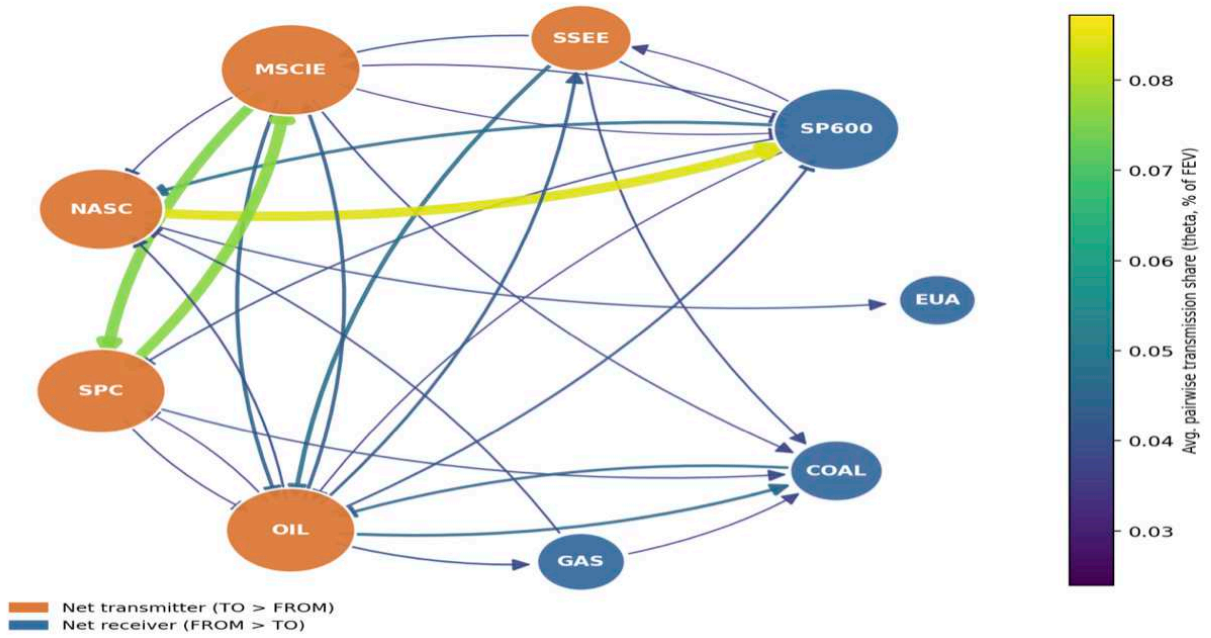


Figure 2 Time-averaged bilateral spillover network, TVP-VAR-JC. Orange nodes send on balance and blue nodes absorb; node size reflects gross TO plus FROM; only the strongest 40% of bilateral links are drawn, coloured by average share.

4.2. Hybrid Spillover Structure

The hybrid uses tuned settings from Table 7, while Table 9 together with Figure 3 presents its time-averaged decomposition. Average TCI falls to 14.86%, 5.7 points below TVP-VAR-JC, so the hybrid traces roughly one seventh of a typical market's ten-day prediction-error variance to other markets.

Net positions resemble those under TVP-VAR-JC. NASC (3.49) and SSEE (1.42) remain leading senders, followed by OIL (1.15), SPC (0.69), GAS (0.66) and MSCIE (0.34), whereas SP600 (−4.26), COAL (−3.04) and EUA (−0.44) absorb on balance. Bilaterally, the MSCIE and SPC pair (7.28% and 6.99%) and the flow from NASC into SP600 (7.28%) again dominate, and OIL retains moderate links with most markets. EUA has both the lowest TO and the lowest FROM among all nine series (TO 10.6%, FROM 11.0%).

Receiver \ Source	EUA	SP600	SSEE	MSCIE	NASC	SPC	OIL	GAS	COAL	FROM
EUA		1.09	1.11	1.60	1.81	1.23	1.76	1.26	1.15	11.01
SP600	1.15		1.74	1.77	7.28	1.15	2.78	1.54	0.87	18.29
SSEE	1.01	1.83		1.16	1.44	2.23	1.99	1.29	1.02	11.97
MSCIE	1.57	1.84	1.40		1.20	7.28	1.81	1.50	1.08	17.68

Receiver \ Source	EUA	SP600	SSEE	MSCIE	NASC	SPC	OIL	GAS	COAL	FROM
NASC	1.03	2.72	1.35	1.90		1.52	1.99	2.01	2.23	14.75
SPC	1.17	1.61	2.15	6.99	1.34		2.37	0.96	1.30	17.89
OIL	1.79	2.01	2.37	1.98	1.68	2.26		1.83	2.48	16.40
GAS	1.27	1.55	1.64	1.09	1.37	1.01	2.32		1.18	11.45
COAL	1.57	1.39	1.64	1.51	2.11	1.89	2.52	1.71		14.35
TO	**10.57**	**14.03**	**13.39**	**18.01**	**18.24**	**18.58**	**17.55**	**12.11**	**11.30**	
NET	**−0.44**	**−4.26**	**1.42**	**0.34**	**3.49**	**0.69**	**1.15**	**0.66**	**−3.04**	**TCI 14.86**

Table 9 Time-averaged spillover matrix, hybrid estimator (%)

Note: Layout as in Table 8.

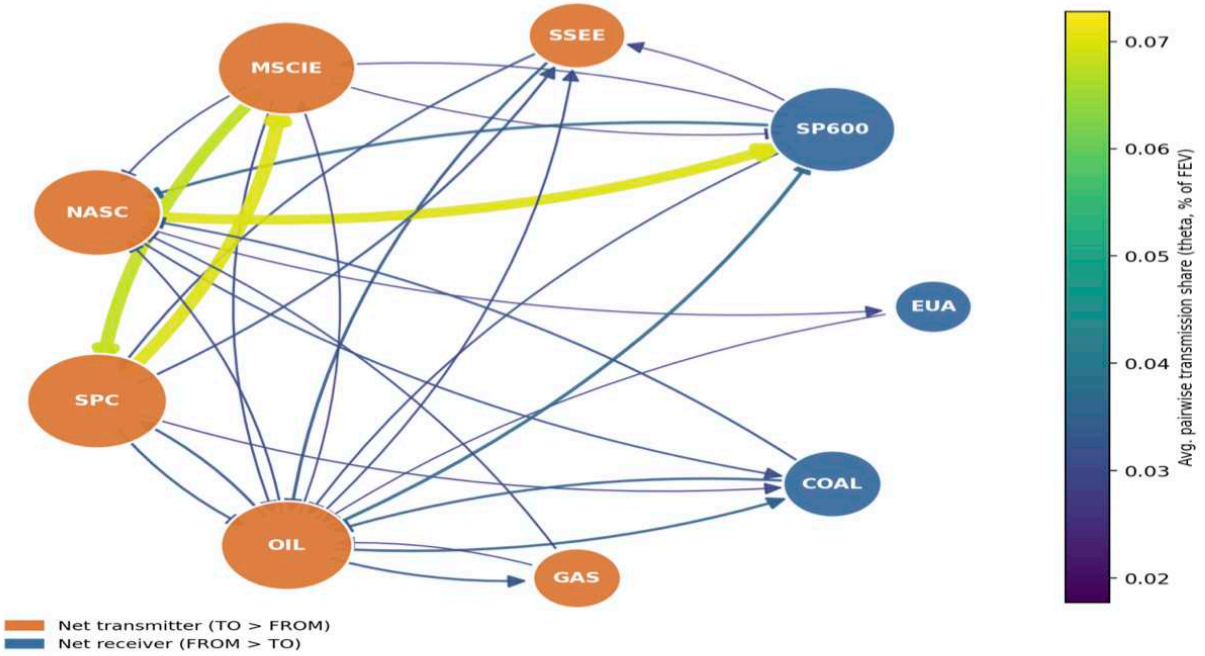


Figure 3 Time-averaged bilateral spillover network, hybrid estimator. Layout as in Figure 2.

4.3. Comparing Both Indices

Table 10 compares both TCI series over 1,196 days for which each is available. The hybrid shows a lower mean (14.86% against 20.57%) yet a larger standard deviation (8.28 against 6.99), hence a much higher ratio of dispersion to level (0.557 against 0.340). Minima (8.84% and 8.41%) and maxima (88.50% and 88.59%) are similar in size, although Figure 4 reveals that the lowest benchmark readings fall on the first days of January 2020, while the filter is still being initialized. Both maxima lie just below the theoretical ceiling of 88.9% and occur in early March 2020, within the first weeks of the sample, when each estimator relies on

short histories, so their agreement is weaker evidence than it might appear. The hybrid index is also smoother from one day to the next. Its average absolute daily move is about half as large (0.50 against 0.98 points), and its lag-one autocorrelation is higher (0.933 against 0.910). Such behaviour is what equal weights over 130 days imply when set against exponential weights with κ at 0.96, whose half-life is roughly 17 trading days.

Statistic	TVP-VAR-JC	Hybrid
Mean (%)	20.57	14.86
Standard deviation (%)	6.99	8.28
Dispersion relative to mean	0.340	0.557
Minimum (%)	8.41	8.84
Maximum (%)	88.59	88.50
Lag-one autocorrelation	0.910	0.933
Average absolute daily move (points)	0.977	0.502

Table 10 Descriptive comparison of both TCI series

Table 11 adds further comparisons. The hybrid exceeds 50% on 20 days, against 5 for TVP-VAR-JC, but exceeds 30% on fewer days (60 against 101). Its distribution is more skewed (4.54 against 3.09) and more heavy-tailed (excess kurtosis 24.23 against 16.33). The hybrid therefore spends less time at intermediate levels and more time either near a low base or at very high values. Whether such episodes coincide with genuine market stress cannot be judged from the index alone, because that would require an external stress gauge, such as implied volatility, plus predictive tests on unseen data, neither of which belongs to the present design. Threshold counts are therefore descriptive. They do not measure correct or false alarms.

Both indices are strongly but not perfectly related. They correlate at 0.795, so TVP-VAR-JC accounts for 63% of hybrid variation, and regressing hybrid readings against TVP-VAR-JC gives a slope of 0.94. Mean and median gaps are similar (−5.71 and −5.82 points), pointing to a fairly stable downward shift rather than a gap driven by a handful of outlying days. Still, the average absolute gap (6.46 points) exceeds the absolute mean gap, so on some days the ordering of both indices is reversed.

Statistic	Value
Hybrid mean minus TVP-VAR-JC mean (points)	−5.71
Median daily gap (points)	−5.82

Statistic	Value
Average absolute daily gap (points)	6.46
Root mean squared daily gap (points)	7.61
Average absolute gap relative to TVP-VAR-JC (%)	30.43
Correlation of levels	0.795
R^2 from regressing hybrid readings against TVP-VAR-JC	0.633
Slope from that regression	0.943
Covariance	45.98
Ratio of standard deviations, hybrid over TVP-VAR-JC	1.19
Skewness, TVP-VAR-JC versus hybrid	3.09 versus 4.54
Excess kurtosis, TVP-VAR-JC versus hybrid	16.33 versus 24.23
Days with TCI over 50%, TVP-VAR-JC versus hybrid	5 versus 20
Days with TCI over 30%, TVP-VAR-JC versus hybrid	101 versus 60

Table 11 Additional TCI comparisons

Table 12 contrasts two sub-periods. Between January and April 2020, a window that includes the pandemic crash, both indices average about the same (39.98% for TVP-VAR-JC and 41.55% for the hybrid). Over 2022-2023, averages drop to 18.86% under TVP-VAR-JC and 12.86% under the hybrid, so dividing the crash average by the 2022-2023 average gives 2.12 for TVP-VAR-JC and 3.23 for the hybrid. Two qualifications apply. First, 2022-2023 is not calm in an absolute sense, since it contains the height of Europe's gas crisis and the opening phase of Russia's war against Ukraine. Connectedness was nevertheless low, indicating that large commodity shocks in those years were mostly market-specific, and this span serves here only as a low-connectedness reference. Second, the larger ratio of the hybrid stems entirely from its lower 2022-2023 level, not from a higher crash level, and the crash window overlaps the start-up phase of both estimators. Correlation-based measures also rise mechanically when volatility increases [54], so ratios between high- and low-volatility periods blend changes in transmission with changes in volatility.

Period	TVP-VAR-JC (%)	Hybrid (%)
Pandemic crash, January to April 2020, average	39.98	41.55
Peak day, 03.03.2020	88.59	88.50
Reference span 2022-2023, average	18.86	12.86
Crash average divided by reference average	2.12	3.23

Table 12 TCI during the pandemic crash and over 2022-2023

Two features of the hybrid may explain its lower level in the reference period. Some common movement of returns may be absorbed by the network's expected-return forecast, leaving weaker cross-correlation in its

errors. Furthermore, equal weighting over 130 days keeps the weight of each past observation constant until it drops out, whereas exponential weighting discounts it gradually. The present design changes both elements together and cannot tell them apart. An ablation altering one component at a time could.

On 3 March 2020 both indices reach their maxima (88.59% and 88.50%). Agreement on this date is reassuring given such different estimation mechanisms, but both values sit within half a point of the 88.9% ceiling and fall in a period when the hybrid's rolling covariance relies on fewer than 130 errors, so the peak level calls for caution.

Figure 4 plots both indices. They follow the same broad episodes, with a sharp spike in early 2020 when both approach 88%. The figure also reveals that the hybrid opens above 60% during the opening weeks of January 2020, before the pandemic reached financial markets, whereas the benchmark opens near 8%; both patterns reflect start-up effects rather than market conditions, and they also enter the January to April 2020 averages in Table 12. From mid-2020 onwards TVP-VAR-JC fluctuates at a higher level, mostly between 15% and 25%, and moves more from day to day, whereas the hybrid settles mostly between 10% and 17% with occasional moderate increases that stay well below spikes recorded by TVP-VAR-JC.

In short, the hybrid delivers a lower and smoother baseline with fewer intermediate readings, while TVP-VAR-JC reacts faster to new information around a higher level. Which profile is preferable depends on purpose. A smoother index is easier to read when monitoring broad regimes, whereas a faster one may suit the detection of incipient stress better. Showing that either index anticipates stress would require predictive tests, left here for future work.

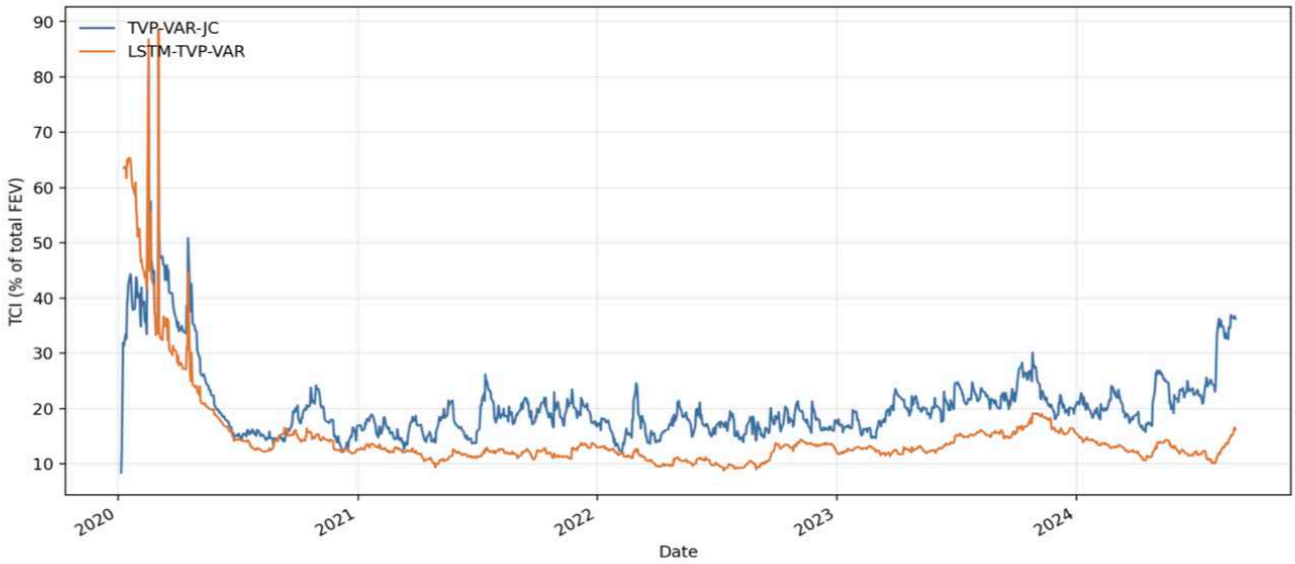


Figure 4 System-wide connectedness over time, both estimators

Table 13 compares net positions. Five series send on balance under both estimators (NASC, SSEE, OIL, MSCIE and SPC) and three absorb on balance (EUA, SP600 and COAL). GAS alone changes sign, from −0.15 under TVP-VAR-JC to 0.66 under the hybrid, and both values are small. EUA has a small negative NET in both cases (−0.39 and −0.44). It is not the series closest to balance, but its gross connectedness, TO plus FROM, is lowest, which makes it the least connected node in both networks.

Series	TVP-VAR-JC	Hybrid	Role
NASC	5.45	3.49	Sender
SSEE	2.47	1.42	Sender
OIL	1.06	1.15	Sender
MSCIE	0.58	0.34	Sender
SPC	0.57	0.69	Sender
GAS	−0.15	0.66	Absorber under TVP-VAR-JC, sender under hybrid
EUA	−0.39	−0.44	Absorber
SP600	−4.42	−4.26	Absorber
COAL	−5.18	−3.04	Absorber

Table 13 NET positions by estimator (percentage points)

Table 14 measures how close both average decompositions are. Their difference has a Frobenius norm of 0.068 in share units, an average absolute size of 0.73 points across all 72 cells off the diagonal (0.65 when the empty diagonal is included), and a maximum of 1.52 points for MSCIE's variance share attributed to OIL. Under either estimator, MSCIE and SPC form the tightest bilateral pair. Together with a Spearman coefficient

of 0.93 between both sets of NET values, these figures indicate that the hybrid scales most cross-market shares down, by about 28% on average, while leaving their ranking largely intact.

Statistic	Value
Frobenius norm of the difference, share units	0.0684
Average absolute gap, cells outside the diagonal (points)	0.73
Average absolute gap, all cells (points)	0.65
Largest absolute gap (points)	1.52
Cell with largest gap	MSCIE ← OIL
Strongest bilateral link, either estimator	MSCIE ↔ SPC
Spearman coefficient between NET vectors	0.93 ($p < 0.001$)

Table 14 Distance between time-averaged spillover matrices

Whether mean TCI differs is examined through a *t*-statistic whose standard error remains valid under heteroskedasticity and serial correlation [52]. Rather than treating each day as an independent draw, Newey-West weighting adds autocovariances up to truncation lag L to the sampling variance of an average, widening the standard error whenever a series is positively autocorrelated. Given how persistent both TCI series are, an unadjusted test would overstate how precisely their gap is estimated.

With $L = 6$, the HAC standard error for the mean gap is 0.362, about 2.5 times its naive i.i.d. counterpart of 0.146, which corresponds to a variance roughly 6.2 times larger. This yields a *t*-value of −15.77 ($p < 0.0001$), compared with −39.26 without correction, and a 95% confidence interval from −6.43 to −5.00 points. This conclusion holds for every truncation lag between 5 and 44. Because the hybrid index is built from overlapping 130-day windows, its serial dependence may extend beyond 44 lags, and longer truncation lags would offer a stricter check. The level gap is thus statistically robust. It shows that both estimators place connectedness at systematically different levels, but it cannot reveal which level lies closer to the true extent of spillovers.

Statistic	Value
Aligned trading days	1,196
Hybrid mean minus TVP-VAR-JC mean (points)	−5.714
Newey-West truncation lag L	6
HAC standard error	0.362

Statistic	Value
t-value	−15.77
p-value	< 0.0001
95% confidence interval (points)	−6.43 to −5.00
Naive i.i.d. standard error, for comparison	0.146
Naive t-value, for comparison	−39.26
HAC over naive standard error	2.48
Implied variance ratio	≈ 6.2

Table 15 Newey-West test for a gap in mean TCI

5. Conclusion and Implications

Climate policy and a run of extreme market events have altered how shocks move among carbon, fuel and energy equity markets. Whether these links are linear or not, and whether they change between calm and turbulent periods, matters for investors who want portfolios able to withstand shocks and for regulators who use the allowance price as a policy signal without wishing to destabilize financial markets. This paper addressed both questions from a methodological and an empirical angle.

Methodologically, the paper combined an LSTM forecast of expected returns and a moving-window error covariance inside a spillover framework with time-varying parameters, and tuned all parameters by chronological cross-validation on training observations. The hybrid produced a lower, smoother system-wide index than TVP-VAR-JC, peaked on the same March 2020 date and preserved the ranking of senders and absorbers. Its level gap is statistically robust. However, because the network was fitted on most observations and both components were changed together, this gap documents a property of the estimator rather than a demonstrated improvement in measuring spillovers.

Empirically, both estimators agree on a clear hierarchy. Nasdaq clean energy shares and Shanghai energy shares transmit most return shocks on balance; S&P Global Clean Energy, MSCI World Energy and oil transmit modestly; US small-cap energy shares, coal and EU allowances absorb shocks on balance; and gas is close to balance. EUA has the weakest links of all, which suggests that carbon prices were largely insulated from short-term shocks in fuel and energy shares over 2020-2024, notwithstanding their pivotal place in climate regulation.

Comparing estimators also matters for anyone who monitors markets with spillover indices. The larger crash-to-reference ratio of the hybrid arises from a lower reference level, not from a stronger crisis reading, and threshold counts without an external stress gauge cannot establish superior early warning. Before either index serves as an early-warning tool, its predictive ability should be tested on unseen data and against gauges such as implied volatility.

For investors, weak links between EUA and other markets point to diversification potential in portfolios holding fuel and energy equity exposures. Absorbing shocks on balance, however, does not make an asset a hedge or a safe haven, which requires low or negative co-movement in falling markets [55], so hedge ratios and portfolio back-tests would be needed to confirm this potential [4]. Net transmission by clean energy shares indicates that these firms now belong to the mainstream energy risk network rather than forming a segment apart.

For policymakers, the weak, receiving position of EUA suggests that Europe's carbon market did not lead price formation within energy markets during the sample period. Its price appears driven chiefly by allowance-market fundamentals, which calls for closer study of allowance supply, the Market Stability Reserve and trading liquidity in the EU scheme.

Several limitations qualify these conclusions and point to further work. Because the network was trained once on the first four fifths of the sample, a re-estimation scheme using only past data at each date would yield an index computable as events unfold. An ablation changing the mean equation and covariance one at a time would reveal what drives the level gap. Early 2020 falls within the start-up phase of both estimators and could be excluded or handled with a pre-sample. All prices come from one public provider, and European, American and Chinese closing times are not synchronized, which can create leads and lags that a flexible network may exploit. Frequency-domain and quantile extensions [34,35] could reveal whether nonlinear filtering matters more in distribution tails or at particular horizons.

Overall, embedding an LSTM in a spillover model with time-varying parameters provides an open, reproducible alternative estimator that keeps the main features of the spillover network intact while changing the level and smoothness of the system-wide index. As climate risk and financial risk become more intertwined, carefully validated spillover tools will remain valuable for research and for evidence-based policy.

Declarations

Funding: This research received no specific grant from any funding agency.

Declaration of generative AI and AI-assisted technologies: During the preparation of this work the author used generative AI tools for editing and formatting.

References

[1] Zhang, Y.-J., Sun, Y.-F., 2016. The dynamic volatility spillover between European carbon trading market and fossil energy market. Journal of Cleaner Production 112, 2654-2663. https://doi.org/10.1016/j.jclepro.2015.09.118

[2] Ji, Q., Zhang, D., Geng, J.-B., 2018. Information linkage, dynamic spillovers in prices and volatility between the carbon and energy markets. Journal of Cleaner Production 198, 972-978. https://doi.org/10.1016/j.jclepro.2018.07.126

[3] Reboredo, J.C., 2013. Modeling EU allowances and oil market interdependence. Implications for portfolio management. Energy Economics 36, 471-480. https://doi.org/10.1016/j.eneco.2012.10.004

[4] Balcilar, M., Demirer, R., Hammoudeh, S., Nguyen, D.K., 2016. Risk spillovers across the energy and carbon markets and hedging strategies for carbon risk. Energy Economics 54, 159-172. https://doi.org/10.1016/j.eneco.2015.11.003

[5] Dutta, A., Bouri, E., Noor, M.H., 2018. Return and volatility linkages between CO2 emission and clean energy stock prices. Energy 164, 803-810.

[6] Hanif, W., Arreola Hernandez, J., Mensi, W., Kang, S.H., Uddin, G.S., Yoon, S.-M., 2021. Nonlinear dependence and connectedness between clean/renewable energy sector equity and European emission allowance prices. Energy Economics 101, 105409. https://doi.org/10.1016/j.eneco.2021.105409

[7] Tan, X., Sirichand, K., Vivian, A., Wang, X., 2020. How connected is the carbon market to energy and financial markets? A systematic analysis of spillovers and dynamics. Energy Economics 90, 104870. https://doi.org/10.1016/j.eneco.2020.104870

[8] Adekoya, O.B., Oliyide, J.A., Noman, A., 2021. The volatility connectedness of the EU carbon market with commodity and financial markets in time- and frequency-domain: The role of the U.S. economic policy uncertainty. Resources Policy 74, 102252. https://doi.org/10.1016/j.resourpol.2021.102252

[9] Diebold, F.X., Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. The Economic Journal 119 (534), 158-171. https://doi.org/10.1111/j.1468-0297.2008.02208.x

[10] Diebold, F.X., Yilmaz, K., 2012. Better to give than to receive: Predictive directional measurement of volatility spillovers. International Journal of Forecasting 28 (1), 57-66. https://doi.org/10.1016/j.ijforecast.2011.02.006

[11] Diebold, F.X., Yilmaz, K., 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. Journal of Econometrics 182 (1), 119-134. https://doi.org/10.1016/j.jeconom.2014.04.012

[12] Antonakakis, N., Chatziantoniou, I., Gabauer, D., 2020. Refined measures of dynamic connectedness based on time-varying parameter vector autoregressions. Journal of Risk and Financial Management 13 (4), 84. https://doi.org/10.3390/jrfm13040084

[13] Koop, G., Korobilis, D., 2014. A new index of financial conditions. European Economic Review 71, 101-116. https://doi.org/10.1016/j.euroecorev.2014.07.002

[14] Cont, R., 2001. Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance 1 (2), 223-236. https://doi.org/10.1080/713665670

[15] Ghaemi Asl, M., Ben Jabeur, S., Nammouri, H., Bel Hadj Miled, K., 2024. Dynamic connectedness of quantum computing, artificial intelligence, and big data stocks on renewable and sustainable energy. Energy Economics 140, 108017.

[16] Chung, V., Espinoza, J., Mansilla, A., 2024. Analysis of financial contagion and prediction of dynamic correlations during the COVID-19 pandemic: A combined DCC-GARCH and deep learning approach. Journal of Risk and Financial Management 17 (12), 567. https://doi.org/10.3390/jrfm17120567

[17] Morshed, A., 2025. Graph neural networks and explainable spillovers: Global monetary and oil shocks in GCC financial markets. Economies 13 (11), 308. https://doi.org/10.3390/economies13110308

[18] Fischer, T., Krauss, C., 2018. Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research 270 (2), 654-669. https://doi.org/10.1016/j.ejor.2017.11.054

[19] Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33 (5), 2223-2273. https://doi.org/10.1093/rfs/hhaa009

[20] Sezer, O.B., Gudelek, M.U., Ozbayoglu, A.M., 2020. Financial time series forecasting with deep learning: A systematic literature review: 2005-2019. Applied Soft Computing 90, 106181. https://doi.org/10.1016/j.asoc.2020.106181

[21] Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation 9 (8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[22] Lei, H., Xue, M., Liu, H., Ye, J., 2024. Unveiling the driving patterns of carbon prices through an explainable machine learning framework: Evidence from Chinese emission trading schemes. Journal of Cleaner Production 438, 140697. https://doi.org/10.1016/j.jclepro.2024.140697

[23] Ren, X., Li, Y., Wen, F., Lu, Z., 2022. The interrelationship between the carbon market and the green bonds market: Evidence from wavelet quantile-on-quantile method. Technological Forecasting and Social Change 179, 121611.

[24] Ji, Q., Xia, T., Liu, F., Xu, J.-H., 2019. The information spillover between carbon price and power sector returns: Evidence from the major European electricity companies. Journal of Cleaner Production 208, 1178-1187. https://doi.org/10.1016/j.jclepro.2018.10.167

[25] Wei, Y., Wang, Y., Vigne, S.A., Ma, Z., 2023. Alarming contagion effects: The dangerous ripple effect of extreme price spillovers across crude oil, carbon emission allowance, and agriculture futures markets. Journal of International Financial Markets, Institutions and Money 88, 101821. https://doi.org/10.1016/j.intfin.2023.101821

[26] Yan, W., Cheung, A., 2023. The dynamic spillover effects of climate policy uncertainty and coal price on carbon price: Evidence from China. Finance Research Letters 53, 103400. https://doi.org/10.1016/j.frl.2022.103400

[27] Ma, Y., Wang, J., 2021. Time-varying spillovers and dependencies between iron ore, scrap steel, carbon emission, seaborne transportation, and China's steel stock prices. Resources Policy 74, 102254. https://doi.org/10.1016/j.resourpol.2021.102254

[28] Wu, R., Qin, Z., 2024. Asymmetric volatility spillovers among new energy, ESG, green bond and carbon markets. Energy 292, 130504. https://doi.org/10.1016/j.energy.2024.130504

[29] Jiang, W., Dong, L., Liu, X., Zou, L., 2024. Volatility spillovers among economic policy uncertainty, energy and carbon markets: The quantile time-frequency perspective. Energy 307, 132683. https://doi.org/10.1016/j.energy.2024.132683

[30] Ye, Y., Lin, B., Que, D., Cai, S., Wang, C., 2024. COVID-19, the Russian-Ukrainian conflict and the extreme spillovers between fossil energy, electricity, and carbon markets. Energy 311, 133399. https://doi.org/10.1016/j.energy.2024.133399

[31] Koop, G., Pesaran, M.H., Potter, S.M., 1996. Impulse response analysis in nonlinear multivariate models. Journal of Econometrics 74 (1), 119-147. https://doi.org/10.1016/0304-4076(95)01753-4

[32] Pesaran, H.H., Shin, Y., 1998. Generalized impulse response analysis in linear multivariate models. Economics Letters 58 (1), 17-29. https://doi.org/10.1016/S0165-1765(97)00214-0

[33] Gabauer, D., 2021. Dynamic measures of asymmetric & pairwise connectedness within an optimal currency area: Evidence from the ERM I system. Journal of Multinational Financial Management 60, 100680.

[34] Baruník, J., Křehlík, T., 2018. Measuring the frequency dynamics of financial connectedness and systemic risk. Journal of Financial Econometrics 16 (2), 271-296. https://doi.org/10.1093/jjfinec/nby001

[35] Ando, T., Greenwood-Nimmo, M., Shin, Y., 2022. Quantile connectedness: Modeling tail behavior in the topology of financial networks. Management Science 68 (4), 2401-2431.

[36] Lastrapes, W.D., Wiesen, T.F.P., 2021. The joint spillover index. Economic Modelling 94, 681-691. https://doi.org/10.1016/j.econmod.2020.02.010

[37] Balcilar, M., Gabauer, D., Umar, Z., 2021. Crude oil futures contracts and commodity markets: New evidence from a TVP-VAR extended joint connectedness approach. Resources Policy 73, 102219. https://doi.org/10.1016/j.resourpol.2021.102219

[38] Qiu, L., Chu, L., Zhou, R., Xu, H., Yuan, S., 2023. How do carbon, stock, and renewable energy markets interact: Evidence from Europe. Journal of Cleaner Production 407, 137106. https://doi.org/10.1016/j.jclepro.2023.137106

[39] Deng, X., Xu, F., 2024. Connectedness between international oil and China's new energy industry chain: A time-frequency analysis based on TVP-VAR model. Energy Economics 140, 107954. https://doi.org/10.1016/j.eneco.2024.107954

[40] Younis, I., Shah, W.U., Missaoui, I., Tang, X., 2024. Exploring interconnectedness between climate change, renewable energy, technological innovation, and G-17 banking stock markets. Journal of Cleaner Production 449, 141667.

[41] Nadirgil, O., 2025. Dynamic transmissions between green and technology stocks, ETFs, commodities, crypto and fiat currencies throughout the global turbulences. Journal of Cleaner Production 494, 145002. https://doi.org/10.1016/j.jclepro.2025.145002

[42] Almajali, A., Almajali, O., Alqaralleh, H., 2026. Assessing climate risk impact on financial markets: A GARCH-Wavelet-Spillover approach to green and traditional assets. The North American Journal of Economics and Finance 84, 102620. https://doi.org/10.1016/j.najef.2026.102620

[43] Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74 (366), 427-431. https://doi.org/10.1080/01621459.1979.10482531

[44] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics 54 (1-3), 159-178. https://doi.org/10.1016/0304-4076(92)90104-Y

[45] Brock, W.A., Dechert, W.D., Scheinkman, J.A., LeBaron, B., 1996. A test for independence based on the correlation dimension. Econometric Reviews 15 (3), 197-235. https://doi.org/10.1080/07474939608800353

[46] Hsieh, D.A., 1991. Chaos and nonlinear dynamics: Application to financial markets. The Journal of Finance 46 (5), 1839-1877. https://doi.org/10.1111/j.1540-6261.1991.tb04646.x

[47] Wiesen, T.F.P., Beaumont, P.M., Norrbin, S.C., Srivastava, A., 2018. Are generalized spillover indices overstating connectedness? Economics Letters 173, 131-134.

[48] Prechelt, L., 1998. Early stopping - but when?, in: Orr, G.B., Müller, K.-R. (Eds.), Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science 1524. Springer, Berlin, Heidelberg, pp. 55-69. https://doi.org/10.1007/3-540-49430-8_3

[49] Bergmeir, C., Benítez, J.M., 2012. On the use of cross-validation for time series predictor evaluation. Information Sciences 191, 192-213. https://doi.org/10.1016/j.ins.2011.12.028

[50] Bergmeir, C., Hyndman, R.J., Koo, B., 2018. A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis 120, 70-83. https://doi.org/10.1016/j.csda.2017.11.003

[51] Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. Journal of Machine Learning Research 13, 281-305.

[52] Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703-708. https://doi.org/10.2307/1913610

[53] Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. The Review of Economic Studies 61 (4), 631-653. https://doi.org/10.2307/2297912

[54] Forbes, K.J., Rigobon, R., 2002. No contagion, only interdependence: Measuring stock market comovements. The Journal of Finance 57 (5), 2223-2261. https://doi.org/10.1111/0022-1082.00494

[55] Baur, D.G., Lucey, B.M., 2010. Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. Financial Review 45 (2), 217-229. https://doi.org/10.1111/j.1540-6288.2010.00244.x