

Learned Attention or Mechanical Flattening? Diagnosing Index Gains in Spillover Networks

Ozan Nadirgil

onadirgil@gmail.com

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

Research Article

Keywords: Financial Contagion, Frequency Connectedness, Graph Attention Network, Return Transmissions, Systemic Risk

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11353782/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.

Learned Attention or Mechanical Flattening? Diagnosing Index Gains in Spillover Networks

Abstract

Learned components are increasingly added to spillover models, and inside the connectedness measure, the most visible evidence for it is a change in the index. This study asks what such a change can show. Return transmission is measured across thirteen series from technology, clean energy, cryptocurrency, commodity and macro-financial groups, using daily data from January 2018 to July 2026. A graph attention layer is placed on the frequency-domain variance decomposition of a time-varying parameter VAR estimated over rolling windows (GAT-FC-TVP-VAR). In the benchmark, total connectedness averages 55.06% and peaks at 66.28% in March 2020, and about 60% of the transmitted variance is absorbed within a trading week. The network is segmented, with a mean transmission share of 0.229 within the cryptocurrency cluster against 0.014 within the commodity cluster, and the global equity benchmark, the technology index and the clean-energy index are the main net transmitters. The attention layer raises the index by 3.72 percentage points, and the gap survives a moving-block bootstrap. An ablation shows that this gain says little about the layer. A one-parameter shrinkage towards the uniform matrix reproduces it exactly with a lower held-out error, and the layer's own training objective prefers no adjustment at all. The attention differs from shrinkage only in the direction in which it redistributes own-variance shares, a property that a scalar attention matrix also reproduces. The study offers a map of a five-group transmission network and a simple benchmark that any connectedness study with a learned component should report.

Keywords: Financial Contagion; Frequency Connectedness; Graph Attention Network; Return Transmissions; Systemic Risk.

1. Introduction

1.1. Background

Connectedness measures built on forecast error variance decompositions have become the standard way of describing how shocks move between markets (Diebold and Yilmaz, 2009, 2012, 2014). Their appeal is practical. A single vector autoregression (VAR) yields a directional measure for every pair of markets and a total index for the system, and once the decomposition is taken to the frequency domain each of these can be split by investment horizon (Baruník and Křehlík, 2018). The markets to which these tools are applied have changed as well. Technology and clean-energy equities, cryptocurrencies and energy commodities have each attracted a large literature on hedging and diversification, and all of them were hit by the pandemic crash of March 2020, the monetary tightening of 2022 and the collapses of Terra-LUNA and FTX in the same year. Whether such shocks spread across these asset classes or stay within them matters both for diversification and for the monitoring of systemic risk.

The methods have moved on too. Time-varying, frequency, quantile and regularised versions of the connectedness framework are now routine, and a more recent line of work adds learned components, usually neural networks, to models of volatility and spillovers. This study examines one such component. It places a graph attention layer on top of a frequency-domain connectedness model and asks a narrow question: does the attention-weighted decomposition describe the transmission structure better than the linear decomposition it replaces? The answer turns out to be largely negative, and the reasons for it are the main subject of the paper.

1.2. Previous Research

The framework of Diebold and Yilmaz (2009, 2012) measures spillovers through the generalised forecast error variance decomposition (GFEVD) of a VAR (Koop et al., 1996; Pesaran and Shin, 1998), and Diebold and Yilmaz (2014) show that the resulting matrix can be read as a weighted, directed network. The original approach tracks time variation with rolling windows. Antonakakis et al. (2020) replace the rolling window with a time-varying parameter VAR (TVP-VAR) in the tradition of Primiceri (2005), estimated with the Kalman filter and the forgetting factors of Koop and Korobilis (2014), which avoids both the loss of observations and the arbitrary choice of window length. Baruník and Křehlík (2018) decompose the same measures by frequency, and Chatziantoniou et al. (2023) combine the two extensions in a time-varying parameter frequency connectedness model. Other refinements deal with dimensionality through regularised estimation (Demirer et al., 2018) and with tail behaviour through quantile VARs (Ando et al., 2022), while Billio et al. (2012) offer a Granger-causality alternative to variance decompositions. All of these versions share one feature that matters here. Each entry of the connectedness matrix is a fixed function of the estimated coefficients and the innovation covariance, even when both are allowed to change over time. This is the property that the attention layer studied in this paper is designed to relax.

Applications to the markets studied here have produced a fairly consistent picture. Early work found cryptocurrencies largely isolated from conventional financial and economic assets (Corbet et al., 2018), although Symitsi and Chalvatzis (2018) report one-directional return and volatility spillovers and two-way shock spillovers between Bitcoin and energy and technology companies, and Ji et al. (2019) find cryptocurrencies increasingly integrated into a broad commodity network. Studies of green and environmental assets give a more mixed picture. Duan et al. (2023) find only weak volatility transmission from cryptocurrencies to traditional and green assets, with clean cryptocurrencies less connected than dirty ones. Hassan et al. (2022) find no strong spillover from cryptocurrency environmental attention to commodities, green bonds or environment-related stocks, although their quantile regressions suggest that most of these assets are adversely affected by it in some market conditions. Yousaf et al. (2024) find substantial return and volatility spillovers between green cryptocurrencies and green stocks in a TVP-VAR, with green-stock indices as net transmitters and green cryptocurrencies as net receivers, and Naeem and Karim (2021) report that green financial assets, clean-energy assets in particular, are effective hedges for Bitcoin. Wang et al. (2024) find energy tokens and

cryptocurrencies clustered together and acting as net receivers in extreme conditions, with renewable-energy stocks more interconnected with one another at the median quantile.

A second group of studies relates cryptocurrencies to equities, commodities and policy uncertainty. Narayan and Kumar (2024) combine several network methods with a Bayesian graphical VAR and find cryptocurrencies negatively related to other asset classes and only weakly connected to them during the pandemic. Bhanja et al. (2023) find Bitcoin relatively isolated from equities and precious metals, with connectedness driven mainly by short horizons and stronger in downturns. Gaies et al. (2024) find that cryptocurrencies can hedge equity risk in the short term and under stable conditions, and Younis et al. (2024), using a TVP-VAR, find decentralised finance assets strongly connected with Islamic equity markets and spillovers that rise during crises. Policy and event studies add detail. Helmi et al. (2023) show that the response of cryptocurrency markets to news about central bank digital currencies varies markedly over time, and Nasir et al. (2025) use cross-quantilogram and quantile time-frequency methods to show that green cryptocurrencies, energy tokens and exchange tokens respond to crypto policy uncertainty differently across quantiles and horizons. Esparcia et al. (2024) combine multiplicative component GARCH estimates of intraday volatility with a TVP-VAR and find that the FTT token was the main net transmitter of volatility shocks after the FTX bankruptcy. Outside cryptocurrency markets, Mai et al. (2022) find spillovers among Chinese carbon markets much stronger during the pandemic than before or after it, and Wu and Hu (2024) use a quantile VAR to show that spillovers between climate policy uncertainty and physical and financial assets in China depend on market conditions.

Three regularities run through this literature, and all three reappear in the results below. Cryptocurrencies tend to form a block of their own that is weakly tied to other assets, transmission is concentrated at short horizons (Le et al., 2021; Bhanja et al., 2023), and connectedness rises in crises. What the literature rarely does is to place technology, clean energy, cryptocurrencies, commodities and the main macro-financial factors in one system. Most of the studies above cover two or three of these groups, so the link between, for example, cryptocurrencies and clean-energy equities can end up being measured without the global equity factor on which both of them load.

A separate line of work adds learned components to models of volatility and spillovers. Araya et al. (2024) feed generalised autoregressive conditional heteroscedasticity (GARCH) volatility estimates into convolutional and recurrent networks, Liu et al. (2025) add a long short-term memory (LSTM) network to the Realized GARCH model, and Sahiner (2024) combines dynamic conditional correlations with an LSTM to build an early-warning system for contagion. Graph neural networks are a natural candidate for multi-asset problems (Wu et al., 2021), and graph attention networks (Veličković et al., 2018) have been used to learn the firm relations behind momentum spillovers (Cheng and Li, 2021) and to forecast returns on international stock indices (Lee et al., 2025). Zhang et al. (2025) use graph neural networks to forecast realised volatility with spillover effects. They find that nonlinear spillovers improve accuracy at horizons of up to a week, that information from multi-hop neighbours does not clearly help, and that a quasi-likelihood loss works better than mean squared error. Closest to the present study, Morshed (2025) initialises a graph neural network from

TVP-VAR estimates, benchmarks it against the Diebold and Yilmaz index and presents its learned edges as an explainable spillover network.

Two difficulties arise when a learned component is placed inside the connectedness measure itself. The first is evaluation. A forecasting model can be judged against realised outcomes, but a re-weighted variance decomposition has no observable counterpart, and the most visible evidence of its effect is a change in the connectedness index. That evidence is weak. The index is one minus the average own-variance share, so any transformation that moves weight off the diagonal raises it, whether or not the transformation is informative. The second difficulty is interpretation. Attention weights are often read as statements about which inputs matter, but work in natural language processing has shown that they need not support that reading (Jain and Wallace, 2019; Wiegrefe and Pinter, 2019). Neither difficulty is taken up in the connectedness literature reviewed above, which leaves open whether a learned layer in this position improves the measurement or merely changes it.

1.3. Contribution

This study builds a frequency connectedness model with time-varying parameters (FC-TVP-VAR) for thirteen daily return series from January 2018 to July 2026 and adds a graph attention layer that re-weights the variance decomposition window by window. The resulting GAT-FC-TVP-VAR is compared with the linear benchmark in two ways. The usual way looks at the level of the connectedness index and tests the difference. The less usual way is an ablation that replaces the learned attention with uniform weights, an untrained network, permuted asset labels, a time-shuffled training sequence, a one-parameter shrinkage towards the uniform matrix and a scalar attention matrix. The usual comparison favours the attention layer. The index rises by 3.72 percentage points, the gap appears in 1,860 of 1,884 windows, and it survives a moving-block bootstrap. The ablation reverses that reading. A single shrinkage parameter reproduces the gain exactly, at a lower held-out error than the trained layer, and the loss on which the layer is trained is minimised by leaving the decomposition unchanged.

The paper contributes in three ways. It provides a map of return transmission across five asset groups in a single system, showing a dense cryptocurrency block, an isolated commodity block, transmission concentrated within a trading week and a net-transmitting core formed by the global equity benchmark and the two equity sector indices. It proposes a simple benchmark for any connectedness study that adds a learned component, namely that the gain in the index should be set against the gain from a shrinkage towards the uniform matrix calibrated to the same index, since that gain can be had without learning anything. And it identifies what does set the attention layer apart from shrinkage. The attention redistributes own-variance shares in the opposite direction, taking the most from markets that are already the least self-contained, but this property follows from the bilinear-softmax form of the layer and not from its estimated parameters. The specification problems that the ablation exposes, including redundant attention heads and an almost trivial training target, are reported as findings and not set aside.

Section 2 describes the data and the two models, Section 3 reports the results and the ablation, and Section 4 concludes.

2. Data and Analytical Framework

2.1. Data

The data are daily closing prices of fourteen series in five groups (technology, clean energy, cryptocurrency, commodity and energy, and global risk and macro-financial factors) from 18 January 2018 to 30 July 2026, all obtained from Investing.com. In the commodity and energy group, OIL is the West Texas Intermediate crude oil futures contract traded on NYMEX, GOLD is the spot gold price in U.S. dollars per troy ounce, and GAS is the NYMEX natural gas futures contract. The technology and clean-energy series are index levels, and the cryptocurrencies are quoted in U.S. dollars. The macro-financial group contains the MSCI ACWI IMI as a broad global equity benchmark, the U.S. Dollar Index, the ten-year U.S. Treasury note futures price and the CBOE Volatility Index. The Treasury position enters through the futures price instead of the yield, so that every series is a log return and the variance decomposition is expressed in a single unit. The volatility index enters as a log change for the same reason.

Cryptocurrencies trade every day, the other markets do not, and the non-cryptocurrency markets follow different holiday calendars. The sample is therefore restricted to days on which all fourteen markets were open. Of the 2,338 dates on which at least one non-cryptocurrency market traded, 2,135 meet this condition, which gives $T = 2,134$ log returns per series. No observation is interpolated or carried forward, so the panel contains no artificial zero returns. The largest share of exactly zero daily returns is 1.9% for the Treasury futures contract and 1.6% for the dollar index, both quoted to two decimal places, and every other series is below 0.5%. This rule is stricter than carrying prices across non-trading days, and it avoids the spurious comovement that carried prices would create at the short horizons the frequency decomposition is meant to separate.

The start date is set by the data. The Ethereum series used here begins on 18 January 2018, and the rolling VAR needs a balanced panel. The sample still covers two full cryptocurrency cycles, the volatility shock of February 2018, the pandemic crash of March 2020, the liquidity expansion of 2020 and 2021, the tightening of 2022 together with the collapses of Terra-LUNA and FTX, and the recovery that followed. This is the kind of regime variation that a time-varying specification is meant to handle.

Table 1 reports descriptive statistics with the full names and short codes used throughout, and Table 2 reports the Jarque–Bera test (Jarque and Bera, 1987). Every series rejects normality at the 1% level. This rules out Gaussian innovations but says nothing by itself about nonlinearity in the conditional mean. Table 3 reports the augmented Dickey–Fuller (ADF) test (Dickey and Fuller, 1979) with a constant and with the lag order chosen by the Akaike information criterion from a maximum of twenty-six lags. Critical values and p-values are simulated under the null at the sample size used here instead of being read from an asymptotic table. The

unit-root null is rejected at the 1% level for every log-return series. In levels it is rejected only for the volatility index ($ADF = -4.524$, $p = 0.0001$), which is mean-reverting by construction, and the other thirteen price series are integrated of order one. The volatility index nevertheless enters in log changes to keep the system homogeneous in units, a choice that is revisited among the limitations in Section 4.

Conditional heteroscedasticity is tested with the ARCH Lagrange multiplier (ARCH-LM) test of Engle (1982) at 1, 5 and 10 lags. The test regresses squared residuals on their own lags, and rejection of its null of constant conditional variance indicates volatility clustering.

Table 4 shows rejection almost everywhere. At 5 and 10 lags the null is rejected at the 1% level for every series. At one lag the only exception is natural gas, which rejects at the 10% level but not at 5% ($p = 0.060$). Volatility clustering is therefore a feature of every market in the panel, which is consistent with information arriving in bursts and with shocks crossing markets through time-varying covariances.

Taken together, the Jarque–Bera and ARCH-LM results rule out a Gaussian, constant-variance benchmark and justify re-estimating the innovation covariance over time. Neither test is a test of linearity in the conditional mean, however, so neither shows on its own that a nonlinear correction is needed. That link is examined directly with the BDS test of Brock et al. (1996) on the residuals of the estimated system, reported in Table 6.

Variable	Code	Mean	Median	Std. Dev.	Skewness	Excess kurtosis
TECHNOLOGY						
NASDAQ 100 Technology	NAST	0.000667	0.001295	0.018986	-0.2064	4.9287
MSCI World Information Technology	MSCIT †	0.000764	0.001586	0.015893	-0.2938	7.0664
CLEAN ENERGY						
NASDAQ Clean Edge Green Energy	NASC	0.000391	0.001196	0.024099	-0.2068	2.9371
S&P Global Clean Energy	SPC	0.000282	0.000271	0.016873	-0.4306	6.4021
CRYPTOCURRENCY						
Bitcoin	BTC	0.000829	0.000644	0.040281	-0.9569	12.8726
Ethereum	ETH	0.000310	0.000061	0.054138	-0.7270	10.6765
BNB	BNB	0.001750	0.001514	0.054002	-0.4133	15.7692
COMMODITY AND ENERGY						
Crude oil futures (WTI, NYMEX)	OIL	0.000126	0.001518	0.033012	-0.4713	92.8040

Variable	Code	Mean	Median	Std. Dev.	Skewness	Excess kurtosis
TECHNOLOGY						
Gold spot (USD per troy ounce)	GOLD	0.000529	0.000753	0.010644	-0.6821	7.1410
Natural gas futures (NYMEX)	GAS	-0.000039	0.000358	0.043446	-1.1859	28.4238
GLOBAL RISK AND MACRO FACTORS						
MSCI ACWI IMI	ACWI	0.000328	0.000751	0.009874	-1.0757	15.5651
US Dollar Index	DXY	0.000046	0.000000	0.004172	-0.1745	1.6992
US 10-year Treasury note futures	ZN	-0.000056	0.000000	0.003603	0.1388	1.9904
CBOE Volatility Index	VIX	0.000157	-0.007917	0.082159	1.2950	8.1932

Table 1 Descriptive Statistics

(Log returns of daily closing prices, $N = 2,134$ observations from 19 January 2018 to 30 July 2026. Rows are grouped by cluster. The Code column gives the short identifier used throughout the paper. † marks the series excluded after the VIF screen of Table 5; it is reported in Tables 1 to 5 so that the exclusion can be verified, and does not enter any subsequent estimate.)

Code	JB statistic	p-value	Skewness	Excess kurtosis
NAST	2175.16	<0.0001	-0.2064	4.9287
MSCIT †	4470.71	<0.0001	-0.2938	7.0664
NASC	782.27	<0.0001	-0.2068	2.9371
SPC	3710.42	<0.0001	-0.4306	6.4021
BTC	15059.50	<0.0001	-0.9569	12.8726
ETH	10323.34	<0.0001	-0.7270	10.6765
BNB	22171.49	<0.0001	-0.4133	15.7692
OIL	765881.91	<0.0001	-0.4713	92.8040
GOLD	4699.68	<0.0001	-0.6821	7.1410
GAS	72336.90	<0.0001	-1.1859	28.4238
ACWI	21953.53	<0.0001	-1.0757	15.5651
DXY	267.55	<0.0001	-0.1745	1.6992
ZN	359.13	<0.0001	0.1388	1.9904
VIX	6565.25	<0.0001	1.2950	8.1932

Table 2 Jarque–Bera (JB) Test Results

(The test examines the null hypothesis of normality on the basis of sample skewness and excess kurtosis. The kurtosis columns in Tables 1 and 2 report excess kurtosis. All series display highly significant non-normality,

consistent with fat-tailed, leptokurtic financial return distributions. Non-normality of the unconditional distribution is implied by conditional heteroscedasticity even under a linear conditional mean, so the test motivates a specification that does not rely on Gaussian innovations rather than a non-linear conditional mean as such. † marks the series excluded after the VIF screen of Table 5.)

Code	ADF (level)	p-value	ADF (log return)	p-value	Order of integration
NAST	0.2639	0.9762	-51.9381	<0.0001	I(1)
MSCIT †	0.8266	0.9946	-50.1668	<0.0001	I(1)
NASC	-1.7989	0.3828	-47.8754	<0.0001	I(1)
SPC	-2.0099	0.2850	-43.0573	<0.0001	I(1)
BTC	-1.3090	0.6289	-48.7180	<0.0001	I(1)
ETH	-2.1510	0.2270	-49.1648	<0.0001	I(1)
BNB	-1.4661	0.5520	-11.6662	<0.0001	I(1)
OIL	-2.1322	0.2341	-30.9232	<0.0001	I(1)
GOLD	0.4100	0.9835	-45.8958	<0.0001	I(1)
GAS	-2.3564	0.1564	-48.7287	<0.0001	I(1)
ACWI	0.7798	0.9939	-14.2893	<0.0001	I(1)
DXY	-2.2362	0.1952	-45.1082	<0.0001	I(1)
ZN	-0.6290	0.8618	-47.0906	<0.0001	I(1)
VIX	-4.5240	0.0001	-16.1285	<0.0001	I(0)

Table 3 ADF Results

(The test is computed with a constant and with the lag order chosen by the Akaike Information Criterion. Critical values and p-values are obtained from 120,000 Monte Carlo replications of the Dickey–Fuller null at the sample size used here; the simulated 1% critical value is -3.44 for the price levels and -3.43 for the returns, which agrees with the response surface of MacKinnon (1996) to within 0.01. Every log-return series is stationary. Among the levels only the volatility index is stationary, which reflects its mean-reverting con-

struction rather than a property of an asset price. † marks the series excluded after the VIF screen of Table 5.)

Code	LM (lag 1)	p-value	LM (lag 5)	p-value	LM (lag 10)	p-value
NAST	243.75	<0.0001	374.28	<0.0001	386.22	<0.0001
MSCIT †	237.08	<0.0001	380.45	<0.0001	410.88	<0.0001
NASC	47.66	<0.0001	354.55	<0.0001	372.14	<0.0001
SPC	94.13	<0.0001	377.53	<0.0001	390.33	<0.0001
BTC	8.94	0.0028	34.97	<0.0001	36.21	<0.0001
ETH	25.29	<0.0001	56.77	<0.0001	58.48	<0.0001
BNB	54.78	<0.0001	108.57	<0.0001	125.28	<0.0001
OIL	613.56	<0.0001	893.82	<0.0001	917.73	<0.0001
GOLD	51.25	<0.0001	212.61	<0.0001	223.72	<0.0001
GAS	3.53	0.0602	219.23	<0.0001	227.17	<0.0001
ACWI	171.26	<0.0001	590.62	<0.0001	661.81	<0.0001
DXY	131.52	<0.0001	227.83	<0.0001	267.18	<0.0001
ZN	93.32	<0.0001	238.66	<0.0001	243.61	<0.0001
VIX	70.90	<0.0001	97.15	<0.0001	104.97	<0.0001

Table 4 ARCH-LM Test Results

($LM = n \cdot R^2$ is distributed as $\chi^2(q)$ under the null of no ARCH effects, where q is the number of lags. The null is rejected at the 1% level in every cell except GAS at lag 1, where it is rejected at the 10% level only. † marks the series excluded after the VIF screen of Table 5.)

Multicollinearity among the log returns is assessed with variance inflation factors (VIF), reported in Table 5. The NASDAQ 100 Technology index and the MSCI World Information Technology index move al-

most identically ($r = 0.948$), and keeping both leaves the system ill-conditioned, with factors of 13.248 and 12.453 in the full fourteen-series panel. The series with the larger factor, MSCIT, is therefore removed from all later steps. The choice also makes substantive sense, since the global equity exposure that MSCIT shares with the ACWI is already represented by the ACWI itself, while NAST isolates the technology sector more sharply. Two points should be kept in mind. A variance inflation factor measures the precision of coefficients in a single-equation regression, whereas in a variance decomposition correlation between series is information and not a nuisance. The exclusion is therefore justified by the numerical conditioning of the rolling estimation, not by any bias in the connectedness estimates. The final column of Table 5 also shows that the exclusion is decisive. The factor for NAST falls from 12.453 to 5.010 and that for the ACWI from 6.158 to 5.023, while the factors for the remaining series change little, the largest further reduction being that of NASC, from 5.872 to 5.766. That value is the largest in the thirteen-series panel and well below the conventional threshold of ten.

#	Variable	VIF (n = 14)	VIF (n = 13)	Severity	Interpretation
1	MSCI World Information Technology (MSCIT)	13.248	excluded	Severe	Near-duplicate of NAST ($r = +0.948$). Excluded so that the technology cluster is not represented twice by the same underlying portfolio.
2	NASDAQ 100 Technology (NAST)	12.453	5.010	Severe	Severe inflation driven by MSCIT ($r = +0.948$); falls to 5.01 once the duplicated index is removed.
3	MSCI ACWI IMI (ACWI)	6.158	5.023	High	Shares a common global equity factor with MSCIT ($r = +0.872$). Tolerable: the coefficients stay identified, their standard errors widen.
4	NASDAQ Clean Edge Green Energy (NASC)	5.872	5.766	High	Shares a clean-energy sector factor with SPC ($r = +0.824$). Tolerable: the coefficients stay identified, their standard errors widen.
5	S&P Global Clean Energy (SPC)	3.847	3.846	Moderate	Moderate co-movement with NASC ($r = +0.824$), consistent with a common clean-energy factor. No remedial action required.
6	Ethereum (ETH)	3.598	3.597	Moderate	Moderate co-movement with BTC ($r = +0.823$), consistent with a common cryptocurrency factor. No remedial action required.
7	Bitcoin (BTC)	3.336	3.336	Moderate	Moderate co-movement with ETH ($r = +0.823$), consistent with a common cryptocurrency factor. No remedial action required.
8	CBOE Volatility Index (VIX)	2.143	2.123	Moderate	Negative co-movement with MSCIT ($r = -0.699$), the usual inverse relation between implied volatility and equity returns. No remedial action required.
9	BNB (BNB)	2.123	2.123	Moderate	Moderate co-movement with ETH ($r = +0.704$), consistent with a common cryptocurrency factor. No remedial action required.
10	US Dollar Index (DXY)	1.490	1.478	Low	Largely independent of the rest of the system (strongest pairwise correlation -0.436 with GOLD).
11	Gold spot (USD per troy ounce) (GOLD)	1.325	1.325	Low	Largely independent of the rest of the system (strongest pairwise correlation -0.436 with DXY).
12	US 10-year Treasury note futures (ZN)	1.268	1.268	Low	Largely independent of the rest of the system (strongest pairwise correlation -0.341 with DXY).

#	Variable	VIF (n = 14)	VIF (n = 13)	Severity	Interpretation
13	Crude oil futures (WTI, NYMEX) (OIL)	1.094	1.093	Low	Largely independent of the rest of the system (strongest pairwise correlation +0.224 with ACWI).
14	Natural gas futures (NYMEX) (GAS)	1.013	1.013	Low	Largely independent of the rest of the system (strongest pairwise correlation +0.091 with OIL).

Table 5 VIF Results

(VIF calculated via OLS auxiliary regressions with a constant added. $VIF = 1 / (1 - R_i^2)$, where R_i^2 is the proportion of the variation in series i explained by all remaining series. The VIF (n = 14) column reports the full system; the VIF (n = 13) column reports the system estimated after the exclusion of MSCIT.)

The attention layer is motivated by the possibility that a linear specification leaves systematic structure in its residuals. This possibility is tested with the BDS test of Brock et al. (1996), applied to the residuals of a VAR(1) for the thirteen retained log-return series over the full sample, which is the specification that underlies the rolling decompositions. The null is that each residual series is independent and identically distributed. The alternative is left unspecified, so the test responds to neglected linear structure, to conditional heteroscedasticity and to nonlinear dependence alike. Standardised statistics are reported for embedding dimensions $m = 2$ to 5 at a distance of one standard deviation. Panel A of Table 6 tests the raw residuals, and Panel B tests the same residuals after dividing them by the conditional standard deviation of a GARCH(1,1) model (Bollerslev, 1986), which separates dependence due to volatility clustering from dependence of other kinds.

Series	m = 2	m = 3	m = 4	m = 5
Panel A: raw VAR(1) residuals				
NAST	7.67***	9.71***	11.22***	12.83***
NASC	5.54***	7.63***	8.79***	10.06***
SPC	4.37***	6.20***	7.07***	7.82***
BTC	5.01***	5.78***	6.22***	7.03***
ETH	4.69***	5.15***	5.61***	5.94***
BNB	9.12***	10.17***	10.08***	10.77***
OIL	8.95***	9.79***	10.34***	10.99***
GOLD	2.63***	4.74***	6.14***	7.54***
GAS	6.19***	8.52***	9.79***	11.15***
ACWI	9.10***	10.50***	11.28***	12.29***
DXY	2.25**	2.35**	3.22***	3.53***

Series	m = 2	m = 3	m = 4	m = 5
Panel A: raw VAR(1) residuals				
ZN	5.82***	7.53***	8.46***	9.38***
VIX	7.00***	7.71***	8.09***	7.99***
Panel B: residuals standardised by a GARCH(1,1) conditional standard deviation				
NAST	1.90*	2.75***	2.81***	2.58***
NASC	0.63	0.78	0.58	0.53
SPC	-0.38	-0.08	-0.12	-0.16
BTC	1.15	1.16	1.16	1.48
ETH	0.89	0.88	1.21	1.43
BNB	2.19**	2.27**	1.66*	1.42
OIL	1.67*	1.38	0.84	0.37
GOLD	-1.61	-1.08	-0.91	-0.68
GAS	0.91	1.69*	1.72*	1.97**
ACWI	1.10	0.95	0.88	0.78
DXY	-0.02	-0.73	-0.39	-0.47
ZN	0.63	0.41	-0.19	-0.75
VIX	1.03	1.14	1.38	1.15

Table 6 BDS Test Results on VAR Residuals

*(BDS z-statistics for embedding dimensions $m = 2$ to 5 at a distance threshold of one sample standard deviation, computed on the residuals of the VAR(1) system fitted to the thirteen retained log-return series. ***, ** and * denote rejection of the null hypothesis of independent and identical distribution at the 1, 5 and 10% levels respectively. Panel B applies the same test to residuals standardised by the conditional standard deviation of a GARCH(1,1) model fitted to each residual series.)*

Panel A rejects independence at the 1% level in 50 of the 52 series-dimension combinations. The two exceptions are the dollar index at $m = 2$ and $m = 3$, where the rejection is at the 5% level. The statistics are large in every group and rise with the embedding dimension, and they are largest for the technology index (7.67 to 12.83), the global equity benchmark (9.10 to 12.29) and crude oil (8.95 to 10.99). The linear VAR therefore leaves structure in its residuals, and this is the empirical premise of the attention-based correction.

Panel B shows where most of that structure comes from. After GARCH standardisation the statistics collapse. Only 6 of the 52 combinations remain significant at the 5% level, and for ten of the thirteen series no embedding dimension is significant at that level. Conditional heteroscedasticity thus accounts for most of the

departure from independence, in line with the ARCH-LM results in Table 4. Three series keep some dependence after filtering, namely the technology index for $m = 3$ and above (2.58 to 2.81), the exchange token BNB at $m = 2$ and $m = 3$ (2.19 and 2.27) and natural gas at $m = 5$ (1.97). Dependence that survives a volatility filter suggests discrete, event-driven behaviour more than smooth variance dynamics.

Two consequences follow for the empirical design. The dependence that remains after volatility filtering is real but limited. It is specific to a few series and moderate in size, which is the setting in which a learned correction could add information without merely re-estimating a variance process that a GARCH model already captures. And because most of the raw rejection reflects conditional heteroscedasticity, any gain from the attention step has to be attributed with care, between the volatility-clustering component that a linear decomposition averages away and the nonlinear component that survives filtering. Section 3.4 returns to this question.

2.2. Analytical Framework

The evidence in Section 2.1 points to volatility clustering, fat tails and some residual nonlinearity, all of which strain a constant-parameter linear model of cross-market transmission. The framework used here has two layers. The first is a frequency connectedness model estimated on rolling windows, which lets coefficients and covariances change over time. The second is a graph attention layer that takes the variance decomposition of each window and re-weights its entries with a single learned attention matrix, so that the weight of each transmission channel can depend on the wider exposure of the two markets it connects. Figure 1 summarises the steps.

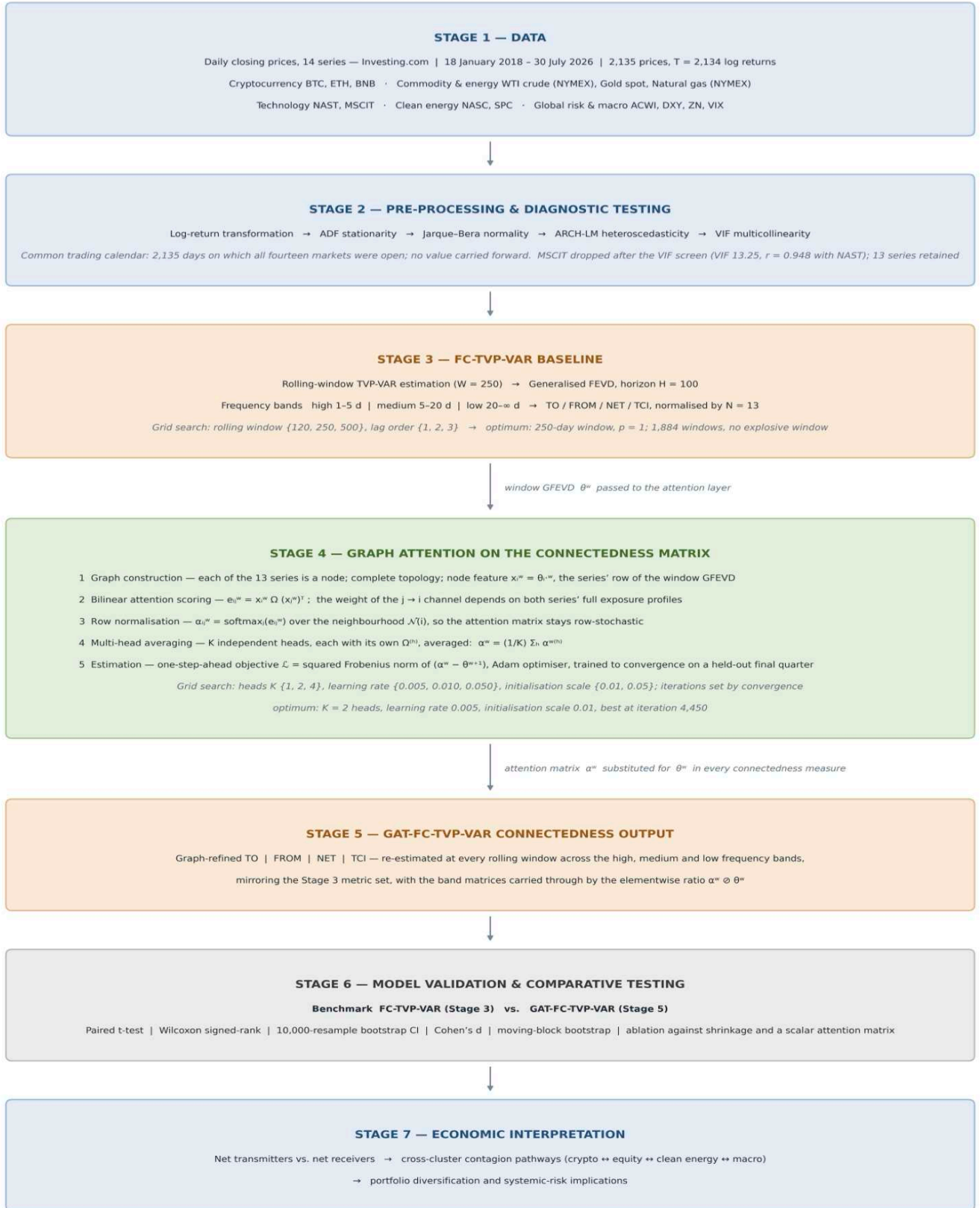


Figure 1 Methodological Framework

2.2.1. FC-TVP-VAR

The FC-TVP-VAR combines a VAR whose parameters change over time with the frequency decomposition of Baruník and Křehlík (2018), so that spillovers can be split by horizon as well as by direction. Time variation is obtained by re-estimating the coefficient and covariance matrices on rolling windows of fixed

length, and the label TVP-VAR is used here in this broad sense. It does not refer to the Kalman-filter estimator with forgetting factors (Koop and Korobilis, 2014; Antonakakis et al., 2020), which is an alternative device for the same purpose, so the specification has no forgetting or decay factor to report. Its role is played by the window length, chosen on one-step-ahead forecast error from the grid in Table 9, which selects 250 trading days. Gradual parameter drift is accommodated in this way, while discrete structural breaks are only approximated, since no jump component is specified. Following Antonakakis et al. (2020), the model is written as in Equation (1), where s_t and ϵ_t are $N \times 1$ vectors, α_t is the $N \times N$ time-varying coefficient matrix for lag $n = 1, \dots, p$, p is the lag order, and Σ is the time-varying variance-covariance matrix. The normality assumption in Equation (1) is the one under which the generalised impulse responses of Pesaran and Shin (1998) are derived. The rolling least-squares estimator does not require it, so the Jarque–Bera rejections in Table 2 bear on the interpretation of the decomposition but not on the consistency of the coefficient estimates.

$$s_t = \sum_{n=1}^p \alpha_{nt} s_{t-n} + \epsilon_t, \epsilon_t \sim N(0, \Sigma_t) \quad (1)$$

Equation (2) gives the GFEVD of Koop et al. (1996) and Pesaran and Shin (1998), where $\tilde{\omega}_{ijt}^g(H)$ is the contribution of a shock in variable j to the forecast error variance of variable i at forecast horizon H . The normalisation over the N source variables makes these contributions sum to one for each variable i at each point in time.

$$\tilde{\omega}_{ijt}^g(H) = \frac{\omega_{ijt}^g(H)}{\sum_{j=1}^N \omega_{ijt}^g(H)} \quad (2)$$

Following Baruník and Křehlík (2018), Equation (2) is taken to the frequency domain, which gives a spectral representation of the GFEVD. Equation (3) defines the decomposition on a frequency band D , where $\tilde{\omega}_{ijt}^g(D)$ is the share of the forecast error variance of variable i that is due to shocks in variable j within that band. Summing over bands partitions total connectedness into short-, medium- and long-run components.

$$\tilde{\omega}_{ijt}^g(D) = \frac{\omega_{ijt}^g(D)}{\sum_{j=1}^N \omega_{ijt}^g(D)} \quad (3)$$

Directional and total measures follow Diebold and Yilmaz (2012). Equation (4) gives total directional connectedness TO, the share of the forecast error variance of all other variables that is due to shocks in variable j . Equation (5) gives FROM, the share of the forecast error variance of variable i that is due to shocks in all other variables, and Equation (6) gives the net measure NET. Positive NET values identify net transmitters and negative values net receivers of return shocks. The total connectedness index (TCI) and its rolling counterpart for window w are given in Equations (7) and (8), which express the same quantity as a proportion and as a percentage. Throughout the paper the TCI is normalised by the number of variables N , as in Diebold and Yilmaz (2012), and not by $N - 1$ as in the corrected index of Antonakakis et al. (2020). With $N = 13$, every TCI reported below is the sum of the off-diagonal variance shares divided by 13.

$$TO_{jt}(D) = \sum_{i=1, i \neq j}^N \tilde{\omega}_{ijt}^g(D) \quad (4)$$

$$FROM_{it}(D) = \sum_{j=1, j \neq i}^N \tilde{\omega}_{ijt}^g(D) \quad (5)$$

$$NET_{it}(D) = TO_{it}(D) - FROM_{it}(D) \quad (6)$$

$$TCI_t(H) = \sum_D TCI_t(D) = \sum_D \left(N^{-1} \sum_{i=1}^N (TO_{it}(D)) \right) = \sum_D \left(N^{-1} \sum_{i=1}^N (FROM_{it}(D)) \right) \quad (7)$$

$$TCI_w(H) = \left(N^{-1} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \tilde{\omega}_{ijw}^g(H) \right) \times 100 \quad (8)$$

2.2.2. Graph Attention Network FC-TVP-VAR (GAT-FC-TVP-VAR)

The FC-TVP-VAR splits transmission by frequency, but it rests on a linear moving-average representation of the system. Every entry of its decomposition is a fixed function of the estimated coefficients and the innovation covariance, so the model cannot give more weight to the channels that dominate in a given regime. A graph attention network offers a way to do so. The system is represented as a graph with a complete, fixed topology in which each market is a node, and learnable attention coefficients decide how much weight each node places on every other. Here the attention operates neither on returns nor on an auxiliary forecast but on the connectedness matrix itself. In every rolling window the frequency-domain decomposition from the FC-TVP-VAR step is read as a complete directed graph over the thirteen markets, and the attention layer re-weights each transmission channel in that graph. The object being learned is therefore the transmission structure, which is also the object of interest. The layer is designed to add three features to the FC-TVP-VAR:

- (i) channel weights that adapt to changing market conditions;
- (ii) channel weights that depend on the full exposure profiles of both markets, so that exposure through third markets enters the weight of a direct link; and
- (iii) a forward-looking criterion, under which the attention weights are chosen to anticipate the decomposition of the next window and not to reproduce the current one. Whether these features improve the measurement is an empirical question, taken up in Section 3.4. Robustness to the conditional heteroscedasticity and fat tails documented in Section 2.1 comes from re-estimating the innovation covariance in each window of the FC-TVP-VAR step, not from the attention layer, whose objective in Equation (14) is an unweighted mean squared error.

At each rolling window w the FC-TVP-VAR step produces the $N \times N$ generalised forecast error variance decomposition θ^w , whose (i, j) entry is the share of the H -step forecast error variance of market i that is due to shocks in market j and whose rows sum to one. This matrix is read as a directed graph $G^w = (V, E)$ over the thirteen markets retained after the VIF screen. The edge set carries no window subscript because the topology is complete and fixed; only the attention weights vary across windows. Node i is represented in window w by its own row of the matrix, $x^{w_i} = \theta^{w_i} \in \mathbb{R}^N$, which is the full profile of where the forecast error variance of market

i comes from. Two properties of this choice matter later. The representation is scale-free, because every node vector lies on the unit simplex, so no standardisation is needed and no node can dominate the attention scores through its own variance. It is also an estimated, system-level object, so the attention layer works on the estimated transmission structure and not on the noise from which that structure is estimated.

Attention layer. The unnormalised attention score of the ordered pair (i, j) is the bilinear form in Equation (9), where $\Omega \in \mathbb{R}^{N \times N}$ is a learnable attention matrix shared by all edges of the graph.

$$e_{ij}^w = x_i^w \Omega (x_j^w)^T = \sum_{k=1}^N \sum_{l=1}^N \theta_{ik}^w \Omega_{kl} \theta_{jl}^w \quad (9)$$

Equation (9) departs from the additive attention of Veličković et al. (2018) and follows the bilinear, or 'general', score of Luong et al. (2015). The difference has a practical consequence. Brody et al. (2022) show that the additive score ranks the neighbours of every node in the same order, whereas under a bilinear score the ranking depends on the receiving node. Because the score is bilinear in two full rows of the decomposition, the parameter Ω weights the coincidence of the exposure of market i to market k with the exposure of market j to market l. The weight of the $j \rightarrow i$ channel therefore depends not on that pair alone but on how the exposure profiles of the two markets interact across the whole system, and two markets that are both exposed to a third receive a different weight from two markets that are not. The GFEVD entry for $j \rightarrow i$ already contains indirect propagation through the dynamics of the system, but it does so through a fixed linear map. The bilinear score adds a second, learned route through which common exposure of the form $k \rightarrow i$ and $k \rightarrow j$, for example to the global equity benchmark, can raise or lower the weight of the direct link.

The scores are normalised over the neighbourhood $\mathcal{N}(i)$, which under the complete topology is the whole set of markets, by the row softmax in Equation (10). The attention matrix therefore keeps the row-stochastic property of the decomposition it replaces and can still be read as a set of variance shares.

$$\alpha_{ij}^w = \text{softmax}_j(e_{ij}^w) = \frac{\exp(e_{ij}^w)}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik}^w)} \quad (10)$$

As in Veličković et al. (2018), K attention heads are estimated in parallel, each with its own matrix $\Omega^{(h)}$, and their normalised outputs are averaged as in Equation (11). Several heads are meant to stabilise estimation and to let different heads attend to different patterns. The grid search selects $K = 2$ (Table 10), although Section 3.4 shows that the heads do not in fact differ.

$$\alpha^w = K^{-1} \sum_{h=1}^K \alpha^{w(h)} \quad (11)$$

Injection into the connectedness measures. The attention matrix replaces the decomposition in every connectedness measure, as in Equation (12), so the directional and total indices of Section 2.2.1 are computed on $\tilde{\theta}^w$ instead of θ^w .

$$\tilde{\theta}^w = \alpha^w \quad (12)$$

The frequency-band matrices of the Baruník–Křehlík decomposition are carried through by the element-wise ratio in Equation (13), where $\odot$ and $\oslash$ denote elementwise multiplication and division.

$$\tilde{\theta}_D^w = \theta_D^w \odot (\alpha^w \oslash \theta^w) \quad (13)$$

Equation (13) transfers the attention-implied level of each channel to every frequency band while keeping the spectral allocation that the linear decomposition assigns to that channel. This is an approximation. It assumes that the attention layer rescales a channel in the same proportion at all frequencies instead of moving weight between them. The near-identical band shares of the two models in Table 14 follow from Equation (13) and are not independent evidence about the frequency structure.

Estimation. The attention matrices are estimated by minimising the one-step-ahead prediction error of the connectedness structure in Equation (14), where M is the number of rolling windows and $\|\cdot\|$ denotes the Frobenius norm.

$$L(\Omega) = (M - 1)^{-1} \sum_{w=1}^{M-1} \|\alpha^w - \theta^{w+1}\|_F^2 \quad (14)$$

The criterion is forward-looking by design. The layer is asked which configuration of channels in window w best anticipates the configuration observed in window $w + 1$, so the weights are chosen for what they predict about the transmission structure, not for how well they reproduce it in the same window. Optimisation uses Adam (Kingma and Ba, 2015). The number of attention heads K , the learning rate and the initialisation scale are chosen by grid search on the prediction error over a held-out final quarter of the window sequence, in the same way as for the baseline model. The number of iterations is not a tuning parameter. The layer is trained to convergence, and the iteration at which the held-out error is lowest is recorded as an outcome. Table 10 reports the grid and the selected values. Two features of the criterion deserve comment. It is an unweighted squared error and so corresponds to a homoscedastic Gaussian criterion. Weighting it by the estimated covariance, or using a quasi-likelihood loss of the kind that Zhang et al. (2025) find superior for volatility forecasting, would align the layer more closely with the evidence of Section 2.1. It is also not a demanding benchmark. As Section 3.4 shows, the naive forecast that the decomposition does not change between windows has a lower error than the trained layer, so Equation (14) should be read as a device for identifying the attention weights and not as evidence of predictive superiority.

In design terms, the GAT-FC-TVP-VAR differs from the FC-TVP-VAR in three respects. The learned attention matrix $\alpha^w = [\alpha^w_{ij}]$ is intended as a time-varying measure of which cross-market channels carry transmission in each window, whereas the FC-TVP-VAR fixes every entry mechanically through the estimated coefficients and covariance. The bilinear score of Equation (9) lets the weight of each direct channel depend on the exposure profiles of both endpoints. And because a single estimated Ω is applied to a new decompo-

sition in every window, the attention weights change with market conditions without a parametric form being imposed on them. Section 3.4 examines how much of this design survives contact with the data.

Among neural architectures, a graph-based layer suits the connectedness problem because the problem is relational. Feed-forward and convolutional networks treat assets as exchangeable feature vectors, and recurrent networks model sequential dependence but not the structure among assets, which they can capture only through dense cross-asset inputs. The specification used here has a single shared matrix Ω per head, so the number of parameters is N^2 per head and does not grow with the number of windows or the length of the sample, and the cost of each window is dominated by two $N \times N$ matrix products. Neither is a concern for thirteen markets. In a much larger panel the parameter count would have to be controlled, for instance through a low-rank factorisation $\Omega = UV^T$ with $U, V \in \mathbb{R}^{N \times r}$, which keeps the bilinear interaction at a cost linear in N for a fixed rank r . The average attention matrix for the full sample is reported in Figure 4. Whether persistent high-weight edges can be read as evidence about transmission, as attention weights often are, is one of the questions the ablation in Section 3.4 is designed to answer.

3. Results and Discussion

3.1. FC-TVP-VAR Results

The FC-TVP-VAR is estimated with the configuration in Table 9, chosen by one-step-ahead mean squared forecast error (MSFE) over rolling windows of 120, 250 and 500 trading days and lag orders one to three. The selected model uses a 250-day window and one lag, which gives 1,884 rolling windows from 17 January 2019 to 30 July 2026. The companion matrix is stable in every window, with a median spectral radius of 0.290 and a maximum of 0.576, so the generalised decomposition is well defined throughout and no stabilising adjustment is needed. The longer window is what secures this. With 120-day windows the same system produces occasional near-explosive estimates and a higher forecast error.

The TCI averages 55.06% over the full sample (Table 8). Its minimum, 47.76%, falls in March 2024 and its maximum, 66.28%, on 17 March 2020, in the first weeks of the pandemic sell-off, when every market in the panel moved together and the distinction between asset classes briefly disappeared. The twelve highest readings in the sample all lie between 13 March and 1 April 2020. By calendar year the index averages 61.8% in 2020 and 60.1% in 2023, against 51.9% in 2021 and 50.5% in 2024, so the system alternates between periods of near-synchronised movement and periods in which the groups separate again.

The frequency decomposition assigns 0.0554 of the total to the low-frequency band (20 days and longer), 0.1657 to the medium band (5 to 20 days) and 0.3259 to the high band (1 to 5 days), or 10.1%, 30.3% and 59.6% of their sum. Connectedness in this system is therefore mainly a short-horizon phenomenon, with about 60% of the transmitted variance absorbed within a trading week. Le et al. (2021) report the same pattern for fintech, cryptocurrency and green-bond volatility, and Bhanja et al. (2023) for equities, precious metals and

Bitcoin. Because the decomposition horizon is 100 days, the low-frequency band is resolved well beyond the 20-day cycle that defines it, so the small long-horizon contribution reflects the data and not a truncated spectral representation. The three bands sum to 0.5470, against a total of 0.5506. The two figures are not algebraically identical. The total is the time-domain generalised decomposition, which weights each forecast horizon equally, while the bands integrate the normalised spectral decomposition with equal weight across frequencies. They coincide only for a white-noise system, and here they differ by 0.36 percentage points, or 0.65% of the total.

Table 7 and Figure 2(a) give the net positions. The global equity benchmark ACWI (NET = +0.2618), the clean-energy index NASC (+0.1406), the technology index NAST (+0.1198) and the two largest cryptocurrencies, ETH (+0.0494) and BTC (+0.0224), are net transmitters, and every other series is a net receiver. The ordering makes economic sense. The broad equity index is the common factor on which the rest of the system loads, and the two equity sector indices share that role. The commodities sit at the other end. Crude oil is the largest net receiver (−0.1404), and natural gas is the most isolated series in the panel, with only 0.127 of its forecast error variance coming from the rest of the system and 0.873 from its own shocks. The dollar index (−0.1096) and the Treasury futures contract (−0.0949) are also net receivers, which fits their role as pricing factors and not as sources of independent shocks in this sample. Le et al. (2021) report a similar split for volatility, with global equity indices among the net transmitters and the dollar, oil and gold among the net receivers.

The panel is clearly segmented. The mean pairwise entry is 0.2285 within the cryptocurrency group and 0.1981 within the clean-energy pair, against 0.0144 within the commodity group and between 0.008 and 0.037 for every pairing of a cryptocurrency with a commodity or a macro-financial series. Among the sector groups, the technology and clean-energy indices are the most tightly linked pair (0.1428 and 0.1315 in the two directions), which reflects the overlap in the constituents of the two indices at least as much as a distinct transmission channel. Cryptocurrencies therefore form a dense block that exchanges little with the rest of the system, in line with Corbet et al. (2018), Duan et al. (2023) and Narayan and Kumar (2024), and commodities remain the natural diversifier in this panel. The result differs from that of Ji et al. (2019), who find cryptocurrencies increasingly integrated into a broad commodity network. Their entropy-based measure, sample period and commodity set all differ from those used here, and the comparison suggests that the integration they document does not extend to daily return transmission in a system that also contains the global equity factor.

Series	TO	FROM	NET	FROM (low)	FROM (med)	FROM (high)
NAST	0.8448	0.7250	0.1198	0.0731	0.2186	0.4328
NASC	0.8721	0.7315	0.1406	0.0724	0.2176	0.4397
SPC	0.6487	0.6666	-0.0179	0.0682	0.2028	0.3916
BTC	0.6557	0.6333	0.0224	0.0626	0.1885	0.3811
ETH	0.6901	0.6407	0.0494	0.0638	0.1916	0.3841
BNB	0.5399	0.5899	-0.0500	0.0565	0.1727	0.3589
OIL	0.1654	0.3058	-0.1404	0.0357	0.1015	0.1637
GOLD	0.3273	0.4084	-0.0812	0.0408	0.1217	0.2402
GAS	0.0536	0.1267	-0.0732	0.0137	0.0375	0.0736
ACWI	1.0148	0.7530	0.2618	0.0756	0.2266	0.4493
DXY	0.3901	0.4997	-0.1096	0.0540	0.1591	0.2738
ZN	0.3056	0.4005	-0.0949	0.0372	0.1147	0.2427
VIX	0.6495	0.6764	-0.0268	0.0665	0.2012	0.4053
TCI		0.5506		0.0554	0.1657	0.3259

Table 7 FC-TVP-VAR Results

(FROM is the row sum of the decomposition excluding the diagonal, TO is the column sum excluding the diagonal, and $NET = TO - FROM$. The last three columns give each series' FROM score in the low-, medium- and high-frequency band. These are obtained from the spectral decomposition while the FROM column is computed in the time domain, so the three band scores sum to the total only up to the difference between the two, which reaches 0.013 for the dollar index and is at most 0.006 for every other series (see Section 3.1). The final row reports the corresponding total connectedness index, computed as the mean FROM score across the $N = 13$ series.)

TCI statistic	Value
Mean TCI (rolling, full period)	55.06%
Minimum TCI	47.76%
Maximum TCI	66.28%
Low-frequency band (20–∞ days)	5.54%
Medium-frequency band (5–20 days)	16.57%
High-frequency band (1–5 days)	32.59%
Sum of the three bands	54.70%

Table 8 TCI Summary Statistics

(The first three rows summarise the rolling-window TCI series. The final four rows report the frequency-band decomposition. The three band contributions sum to 0.5470 while the full-period mean of the rolling index is 0.5506; the two are computed in the frequency and in the time domain respectively and are not algebraically identical, as explained in Section 3.1. TCI is normalised by $N = 13$ throughout.)

Window (days)	Lag (p)	1-step MSFE	Result
120	1	1.226679	
120	2	1.449174	
120	3	1.766066	
250	1	1.124963	selected
250	2	1.216067	
250	3	1.331437	
500	1	1.155620	
500	2	1.206163	
500	3	1.269287	

Table 9 Parameter Optimisation Results

(One-step-ahead mean squared forecast error of the rolling VAR over the window and lag grid. The selected configuration is the one with the lowest error.)

3.2. GAT-FC-TVP-VAR Results

The GAT-FC-TVP-VAR is applied to the same thirteen series. As described in Section 2.2.2, the attention layer operates on the window-by-window decomposition from the FC-TVP-VAR step, not on the returns, and the resulting attention matrix replaces that decomposition in the rolling connectedness calculations. Training uses the Adam optimiser (Kingma and Ba, 2015). Table 10 reports the grid and shows, for each dimension, whether it changes the solution or only the number of iterations needed to reach it. The lag order and the window length, both chosen on forecast error, matter. The other three dimensions do not. The learning rate changes the iteration count by a factor of six but moves the held-out error by less than 1%, the initialisation scale is immaterial, and the number of attention heads has no effect at all, a result discussed in Section 3.4. Training stops at convergence, and the selected configuration reaches its lowest held-out error after 4,450 iterations.

Parameter	Grid range	Selected value	Effect at convergence
VAR lag order (p)	{1, 2, 3}	1	Material: selected on 1-step MSFE
Rolling window (days)	{120, 250, 500}	250	Material: selected on 1-step MSFE
Attention heads (K)	{1, 2, 4}	2	None: identical solutions (see Section 3.4)
Learning rate	{0.005, 0.010, 0.050}	0.005	Speed only: changes the iteration count, not the optimum
Initialisation scale	{0.01, 0.05}	0.01	None: spread across the grid is 5.8e-06 in validation MSFE
Training iterations	to convergence (see note)	4,450	Outcome of the run, not a grid point
GFEVD horizon (H)	fixed	100	-
Frequency grid	fixed	200	-

Table 10 Optimisation Results

(The attention layer is trained to convergence rather than for a fixed number of iterations: training stops when the relative fall in the objective over fifty consecutive iterations drops below 10^{-6} , or when the held-out error has risen for six consecutive checks. The iteration count is therefore an outcome of the run. The final column records whether a grid dimension changes the solution or only the path to it; the evidence for the entries is in Section 3.4.)

Under the attention-weighted decomposition the system is highly integrated, with a mean TCI of 58.78% (Table 11). The minimum, 49.29%, falls in March 2022 and the maximum, 76.12%, on 17 March 2020, the same window in which the benchmark peaks (Table 12). The attention-weighted index exceeds the benchmark in all but 24 of the 1,884 windows. What that difference does and does not show is the subject of Section 3.4. On its own it is a statement about the average own-variance share, which falls from 0.4494 to 0.4122.

The shape of the frequency decomposition hardly changes. The attention-weighted bands contribute 0.0598, 0.1787 and 0.3493 at the low, medium and high frequencies, or 10.2%, 30.4% and 59.4% of their sum, against 10.1%, 30.3% and 59.6% for the benchmark. This is a consequence of the construction and not a finding. Equation (13) allocates the attention-implied level of each channel across the bands in the proportions that the linear decomposition assigns to that channel, so the band shares are inherited and only their common level changes. The comparison between the two models is therefore about the level of each channel, not about its distribution across horizons.

Series	TO	FROM	NET	FROM (low)	FROM (med)	FROM (high)
NAST	0.9197	0.7889	0.1308	0.0796	0.2388	0.4705
NASC	0.8915	0.7914	0.1001	0.0789	0.2372	0.4754
SPC	0.7082	0.7088	-0.0006	0.0734	0.2185	0.4169

Series	TO	FROM	NET	FROM (low)	FROM (med)	FROM (high)
BTC	0.6992	0.6874	0.0118	0.0673	0.2031	0.4170
ETH	0.7104	0.7037	0.0067	0.0696	0.2095	0.4246
BNB	0.5802	0.6204	-0.0402	0.0594	0.1814	0.3796
OIL	0.2118	0.3019	-0.0902	0.0367	0.1030	0.1623
GOLD	0.3357	0.4213	-0.0856	0.0436	0.1297	0.2480
GAS	0.0910	0.1162	-0.0252	0.0123	0.0336	0.0703
ACWI	0.9361	0.8127	0.1234	0.0828	0.2476	0.4824
DXY	0.4472	0.5365	-0.0893	0.0599	0.1762	0.3004
ZN	0.3543	0.4090	-0.0547	0.0397	0.1208	0.2485
VIX	0.7559	0.7430	0.0128	0.0742	0.2233	0.4455
TCI		0.5878		0.0598	0.1787	0.3493

Table 11 GAT-FC-TVP-VAR Transmission Results

(FROM is the row sum of the decomposition excluding the diagonal, TO is the column sum excluding the diagonal, and $NET = TO - FROM$. The last three columns give each series' FROM score in the low-, medium- and high-frequency band; the three band scores sum to the total. The final row reports the corresponding total connectedness index, computed as the mean FROM score across the $N = 13$ series.)

The total-band results in Table 11 and Figure 2(b) show the same division as the benchmark, with one change. The technology index NAST ($NET = +0.1308$), the global equity benchmark ACWI ($+0.1234$), the clean-energy index NASC ($+0.1001$), BTC ($+0.0118$) and ETH ($+0.0067$) remain net transmitters, and the volatility index VIX joins them ($+0.0128$) after being a marginal net receiver in the benchmark. Crude oil, the dollar index and gold remain the largest net receivers. Twelve of the thirteen series keep their rank by net transmission to within two positions, the exception being VIX, which moves up three places, and twelve keep the sign of their net score. The attention step re-levels the system without reordering it.

TCI statistic	Value
Mean TCI (rolling, full period)	58.78%
Minimum TCI	49.29%
Maximum TCI	76.12%
Low-frequency band (20– ∞ days)	5.98%
Medium-frequency band (5–20 days)	17.87%
High-frequency band (1–5 days)	34.93%

TCI statistic	Value
Sum of the three bands	58.78%

Table 12 TCI Summary

(For the attention-weighted model the rolling-window mean and the sum of the three band contributions coincide at 58.78%, because Equation (13) allocates the bands in proportion to the attention matrix. The corresponding quantities differ by 0.36 percentage points for the FC-TVP-VAR in Table 8, where the bands are obtained from the spectral decomposition rather than by allocation.)

The strongest bilateral links are inside the cryptocurrency group, with BTC–ETH at 0.2810 and 0.2791 in the two directions, ETH–BNB at 0.2154 and BTC–BNB at 0.2050. The clean-energy pair NASC–SPC (0.2100) and the link between the technology index and the global benchmark (0.1899 and 0.1888) come next. The commodity markets have much lower gross scores throughout. Natural gas transmits 0.091 and receives 0.116, an order of magnitude below the cryptocurrency pairs.

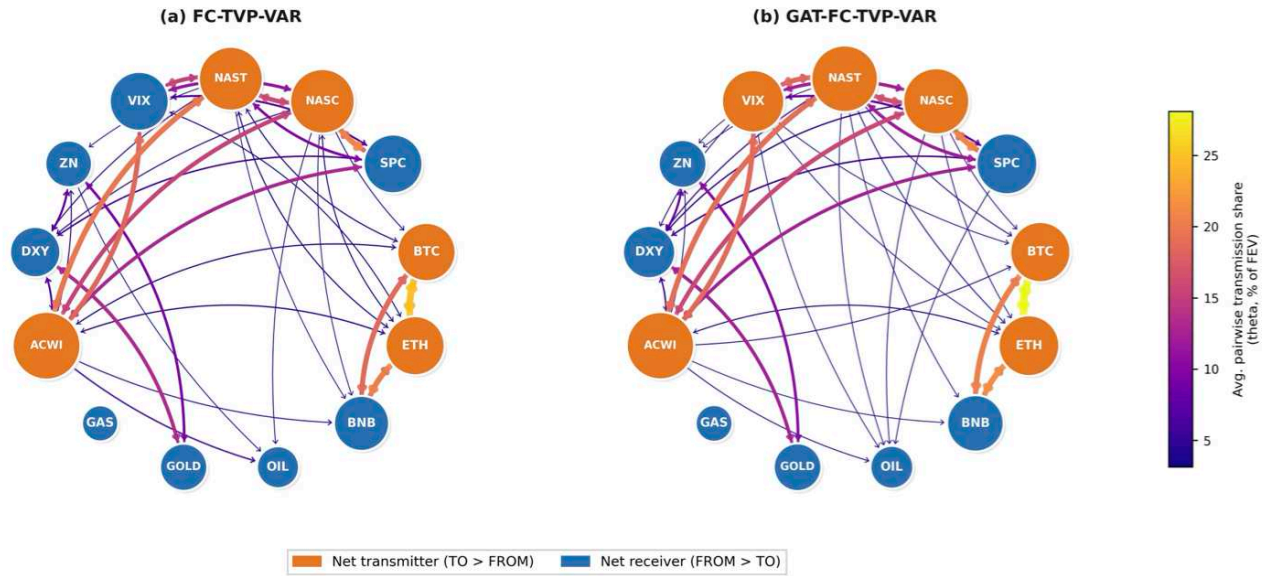


Figure 2 Transmission networks: (a) FC-TVP-VAR, (b) GAT-FC-TVP-VAR

(Average pairwise transmission network over the rolling windows. The strongest 38% of pairwise links are shown. Edge colour and width give the average pairwise transmission share as a percentage of forecast error variance, on a scale common to both panels; node size is gross TO + FROM involvement, also on a common scale. Orange nodes are net transmitters (TO > FROM), blue nodes net receivers.)

The concentration of connectedness at high frequencies holds for every series, so it is a property of the whole sample and not of any one group. In the FROM columns of Table 11 no series has a low-frequency score above 0.10, and every series except natural gas has a high-frequency score above 0.15.

3.3. Comparative Results

Relative to the FC-TVP-VAR, the GAT-FC-TVP-VAR has a mean TCI 3.72 percentage points higher, an increase of 6.76%. The gap is positive in 1,860 of the 1,884 windows and reaches 10.17 percentage points at its widest, in March 2020. In the 24 windows where it is negative it never falls below −0.32 percentage points. The selected configuration uses two attention heads, a learning rate of 0.005 and an initialisation scale of 0.01, although, as Table 10 records and Section 3.4 shows, none of these choices changes the solution.

Statistic	FC-TVP-VAR	GAT-FC-TVP-VAR	Δ (GAT – FC)	% change
Mean TCI (full period)	0.5506	0.5878	+0.0372	+6.76%
Minimum TCI	0.4776	0.4929	+0.0153	+3.21%
Maximum TCI	0.6628	0.7612	+0.0984	+14.85%
Low-frequency band	0.0554	0.0598	+0.0044	+7.95%
Medium-frequency band	0.1657	0.1787	+0.0130	+7.82%
High-frequency band	0.3259	0.3493	+0.0234	+7.19%

Table 13 TCI Comparison

(Mean, minimum and maximum TCI are computed over the rolling windows. The mean values are the totals over all frequency bands reported in Tables 8 and 12. % Change is Δ expressed as a percentage of the FC-TVP-VAR value.)

Figure 3 plots the two indices over time. They move together throughout, with a positive and fairly stable gap. Both start 2019 at around 53% and 56%, jump in the first quarter of 2020 to their sample maxima as the pandemic sell-off synchronises every market in the panel, and fall back in 2021 as the groups separate again during the post-pandemic liquidity expansion. Part of the abrupt fall in March 2021 is mechanical, since it comes when the observations of March 2020 leave the 250-day estimation window. A second elevated episode runs through 2023, when the index averages 60.1% for the benchmark and 65.7% for the attention-weighted model, a period that includes the banking stress of March 2023 and the tightening cycle. The lowest readings are in 2024, at 50.5% and 53.3% on annual average, and both indices rise again through the first half of 2026. The gap is widest in the high-connectedness episodes, which is what a multiplicative re-levelling of the off-diagonal mass would produce.

Table 13 compares the summary statistics. The mean TCI rises from 0.5506 to 0.5878, the minimum from 0.4776 to 0.4929 and the maximum from 0.6628 to 0.7612. The band contributions in Table 14 rise in proportion, by 0.0044 at low, 0.0130 at medium and 0.0234 at high frequencies, and the band shares stay within a tenth of a percentage point of the benchmark.

Band	FC-TVP-VAR	Share	GAT-FC-TVP-VAR	Share	Δ TCI
Low-frequency (20– ∞ days)	0.0554	10.13%	0.0598	10.17%	+0.0044
Medium-frequency (5–20 days)	0.1657	30.29%	0.1787	30.40%	+0.0130
High-frequency (1–5 days)	0.3259	59.58%	0.3493	59.43%	+0.0234
Total	0.5470	100.00%	0.5878	100.00%	+0.0408

Table 14 Frequency-band Contribution (% of total TCI)

(Shares are computed against the sum of the three bands in the same column, that is 0.5470 for the FC-TVP-VAR and 0.5878 for the GAT-FC-TVP-VAR. Δ TCI in the final row is the difference between the two totals and therefore equals the sum of the three band differences.)

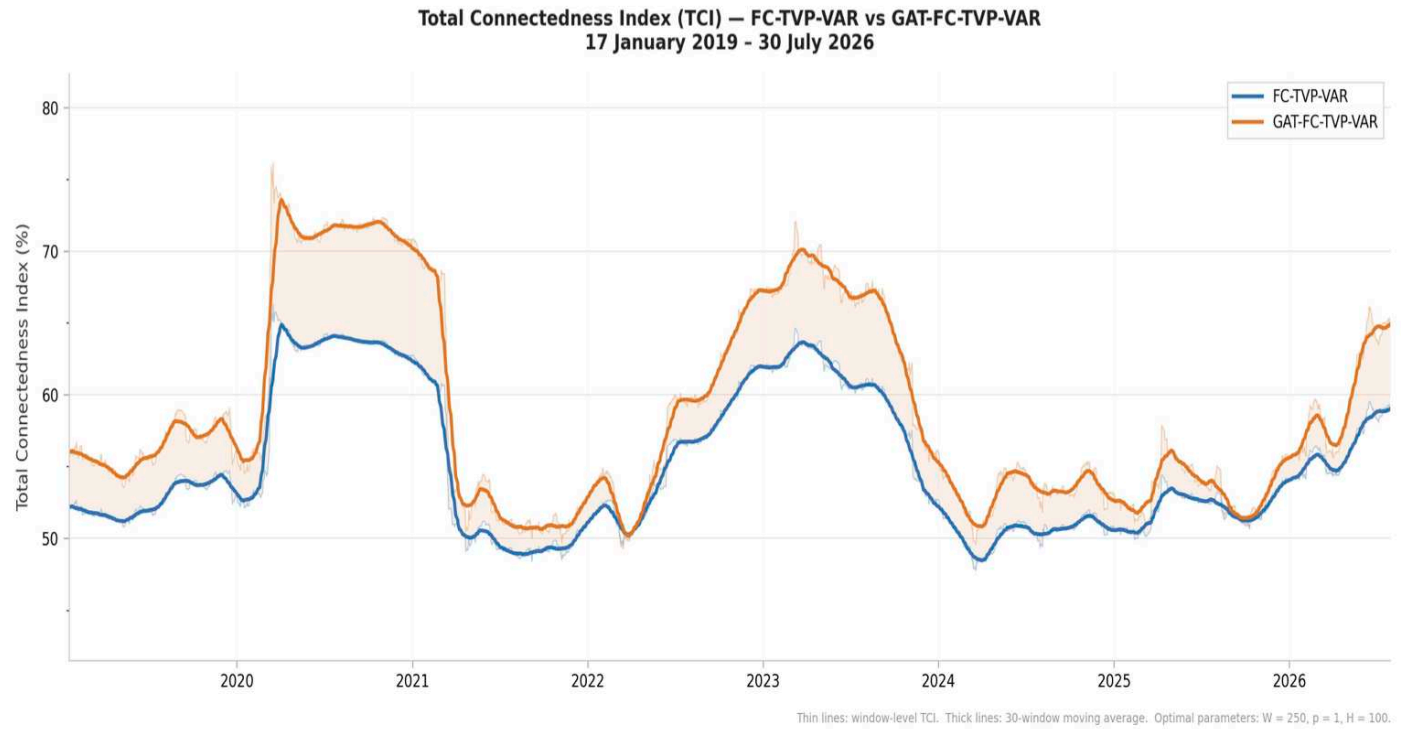


Figure 3 TCI Over Time

At the level of individual series, FROM scores rise most where the benchmark already recorded the highest connectedness, namely for the volatility index (6.67 percentage points of own-variance share released), the technology index NAST (6.39), ETH (6.30), the clean-energy index NASC (5.99) and the global equity benchmark ACWI (5.97). Natural gas and crude oil move the other way, with their own-variance shares rising by 1.05 and 0.39 percentage points. The attention step is therefore not a uniform correction. It widens the cross-sectional dispersion of own-variance shares, making connected series more connected and isolated ones more isolated. Section 3.4 asks how far this redistribution distinguishes the learned layer from one-parameter alternatives.

The per-series differences between the two models are tested with three procedures and one effect-size measure, all on the same thirteen paired FROM scores from the total band (Table 15). The paired t-statistic is 4.899 ($p = 0.0004$), the Wilcoxon (1945) signed-rank statistic is $W = 4$ ($p = 0.0017$), and the bootstrap confidence interval for the mean difference (Efron, 1979) is $[0.0224, 0.0510]$, entirely above zero. Cohen's d is 1.359, a large effect by conventional standards (Cohen, 1988).

Test	Type	Statistic	p-value	95% CI (Δ)	Result
Paired t-test	Parametric, paired	$t(12) = 4.899$	$p = 0.0004$	[0.0207, 0.0538]	Significant ***
Wilcoxon signed-rank	Non-parametric, paired	$W = 4.000$	$p = 0.0017$	-	Significant ***
Bootstrap CI	10,000 resamples	Mean $\Delta = 0.0372$	-	[0.0224, 0.0510]	Excludes zero
Effect size (Cohen d)	Standardised mean difference	$d = 1.359$	-	-	Large effect
Moving-block bootstrap, L = 10	10,000 replications	Mean $\Delta = 3.71$ pp	$p < 0.0001$	[3.39, 4.04]	Excludes zero
Moving-block bootstrap, L = 20	10,000 replications	Mean $\Delta = 3.70$ pp	$p < 0.0001$	[3.26, 4.16]	Excludes zero
Moving-block bootstrap, L = 40	10,000 replications	Mean $\Delta = 3.70$ pp	$p < 0.0001$	[3.08, 4.34]	Excludes zero
Moving-block bootstrap, L = 60	10,000 replications	Mean $\Delta = 3.68$ pp	$p < 0.0001$	[2.95, 4.45]	Excludes zero

Table 15 Statistical Significance Test Results

(The first four rows are computed on per-series FROM scores, total band, $n = 13$. The paired t-statistic is reported for the GAT-FC-TVP-VAR minus FC-TVP-VAR difference. The moving-block bootstrap rows are computed on the window-level difference in the total connectedness index and are robust to its serial dependence; L is the block length in windows.)

These four quantities are computed on the same thirteen series, which are cross-sectionally dependent by construction, so they are not independent pieces of evidence. They show that the difference is systematic across the cross-section and not driven by one or two series. The time dimension needs separate treatment. The difference between the two indices is highly persistent, with a first-order autocorrelation of 0.997 across the 1,884 windows, and a test that treated windows as independent would be badly oversized. The moving-block bootstrap of Künsch (1989), reported in the last four rows of Table 15, deals with this directly. The mean difference is 3.72 percentage points, and the 95% interval excludes zero at every block length, widening from [3.39, 4.04] at the shortest block to [2.95, 4.45] at the longest. The gap between the two models is therefore not an artefact of serial dependence.

None of this shows that the attention layer has found a better description of the transmission structure. The tests establish that the difference between the two indices is systematic and not sampling noise, but they say nothing about whether the direction of the difference is informative. Section 3.4 addresses that question.

3.4. Ablation of the Attention Mechanism and the Learned Attention Weights

A higher connectedness index is not in itself evidence that the attention layer has learned anything. The index is one minus the mean own-variance share, so any transformation that moves weight off the diagonal raises it. The completely uninformative matrix, in which every entry equals $1/N$, gives $1 - 1/N = 92.31\%$, far above either model. The ablation in Table 16 is designed to separate what the layer has learned from what the index rewards mechanically. Each row replaces the learned attention matrix with a stated alternative and re-computes the connectedness measures on the same decompositions.

Specification	TCI (%)	Δ TCI (pp)	Mean own-share	1-step MSFE	Corr. with benchmark	Top-20 edge stability
FC-TVP-VAR benchmark (no attention)	55.06	+0.00	0.4494	0.000003	1.000	0.885
Learned attention (as specified)	58.78	+3.72	0.4122	0.000558	0.937	0.885
Uniform attention (no learning)	92.31	+37.25	0.0769	0.019194	-	1.000
Untrained GAT (random initialisation)	92.31	+37.25	0.0769	0.019192	-0.139	0.861
Permuted asset labels	66.50	+11.44	0.3350	0.002518	0.876	0.866
Attention trained on a time-shuffled sequence	59.43	+4.37	0.4057	0.000750	-	-
Linear shrinkage, λ fitted on the loss (0.000)	55.06	+0.00	0.4494	0.000003	1.000	0.885
Linear shrinkage, λ matched to the attention (0.100)	58.78	+3.72	0.4122	0.000193	1.000	0.885
Scalar attention matrix ($\Omega = c \cdot I$, c matched to the attention, $c = 10.03$)	58.78	+3.72	0.4122	0.001722	0.936	0.882

Table 16 Ablation of the Attention Mechanism

(Every row replaces the learned attention matrix by the stated alternative and recomputes the connectedness measures on the same decompositions. Δ TCI is the change relative to the FC-TVP-VAR benchmark in the first row. The 1-step MSFE is the held-out error of the specification in predicting the next window's decomposition. Corr. with benchmark is the cross-sectional correlation between the resulting NET vector and that of the benchmark; top-20 edge stability is the share of a window's twenty strongest links that also appear in the full-sample ranking. In the final row the learnable matrix Ω is replaced by the scalar $c \cdot I$, with c set so that the resulting index matches the one the learned layer implies.)

Two results stand out immediately. Uniform attention and the untrained network are degenerate, not neutral, placebos. With a small initialisation scale the untrained scores are close to zero and the softmax returns almost equal weights, so both specifications give exactly 92.31%, the value of the uninformative matrix. Their held-out error is about 34 times that of the trained layer and more than three orders of magnitude above that of the unchanged decomposition. Applying the learned weights after permuting the asset labels raises the index to 66.50%, well above the learned value, with a held-out error about four and a half times larger. The

index and the objective therefore point in opposite directions, and the index alone cannot rank the specifications.

The decisive comparison is with linear shrinkage. Shrinking the decomposition towards the uniform matrix with one parameter λ , chosen on the same held-out loss that the attention layer minimises, gives $\lambda = 0.000$. The loss-optimal amount of shrinkage is zero, and the benchmark decomposition is its own best predictor of the next window's decomposition, with a held-out error of 3×10^{-6} against 5.6×10^{-4} for the trained layer. Setting λ instead to the value that reproduces the attention-implied index, $\lambda = 0.100$, matches the learned TCI to four decimal places. It does so with a held-out error of 1.9×10^{-4} , about a third of that of the trained layer, the same top-twenty edge stability of 0.885, and a correlation of 1.000 with the benchmark net vector, which shrinkage merely rescales. On the objective the layer is trained to minimise, and on the index it was expected to improve, one scalar does better. Shrinkage of this kind is familiar from covariance estimation (Ledoit and Wolf, 2004), where it is valued for improving conditioning. Here it serves as the benchmark that any learned re-weighting has to beat.

Three further checks narrow down what the layer has fitted. First, it does not use the ordering of the windows. Retrained on a randomly permuted sequence of windows, it gives a mean index of 59.43%, slightly above the 58.78% obtained with the true ordering, so what the attention captures is a cross-sectional regularity of the average decomposition and not a temporal one. Second, the solution does not depend on the random seed. Five seeds give the same index to within 0.001 percentage points and the same held-out error to six decimal places. Third, the number of attention heads does not matter. Across the grid, the spread of held-out errors within a given head count, which is driven by the learning rate, is 5.6×10^{-6} , while the spread between head counts is 4×10^{-8} , two orders of magnitude smaller. The reason is structural. Each head minimises the same objective independently and the heads are then averaged, and nothing in the specification, such as a separate subspace for each head or a learned combination of heads, would make them differ. At convergence they coincide, and averaging identical matrices returns the same matrix. Multi-head attention as specified here is therefore redundant, and the single-head model is the accurate description of what is estimated.

Figure 4 reports the learned weights, which are far from uniform. Individual weights range from 0.0008 to 0.975, against a uniform value of 0.0769, and the mean entropy of an attention row is 1.770 nats, against 2.565 for a uniform row and 1.715 for the benchmark decomposition. The attention rows are thus slightly more diffuse than the rows they replace but still highly concentrated. Panel (a) shows a clear block structure. The cryptocurrency block is dense and close to symmetric, with the strongest internal weights on the BTC–ETH pair (0.281 and 0.279), the clean energy pair and the technology benchmark pair form two further blocks, and the three commodities exchange almost no weight with any other node. The attention therefore reproduces the segmentation that the linear decomposition finds. The net positions are also largely preserved.

The correlation between the two net vectors is 0.937, and the only change of sign is that of VIX, which moves from a marginal net receiver (-0.0268) to a marginal net transmitter ($+0.0128$), while the clean-energy index SPC moves from a marginal net receiver to a value indistinguishable from zero.

The layer is nevertheless not the same object as the shrinkage, and the difference lies in the cross-section. Shrinkage towards the uniform matrix necessarily compresses the cross-section of own-variance shares, and the change it makes to each series' own-variance share correlates -1.000 with that series' initial own-share, by construction. The attention does the opposite. The change it makes correlates $+0.945$ with the initial own-share, so it takes the most variance from the series that were already the least self-contained and returns a little to the two most isolated commodities. The two adjustments correlate -0.945 with each other, and the equality of the two indices is an equality of averages reached by opposite cross-sectional operations.

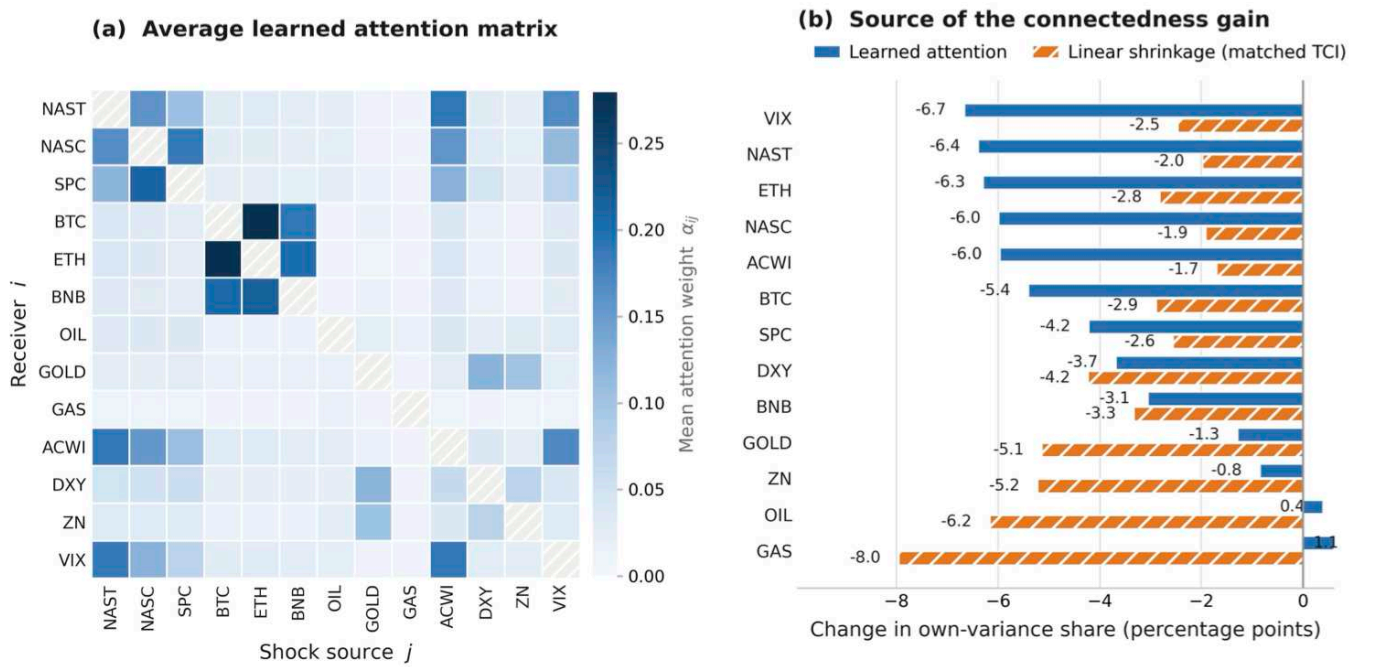


Figure 4 Learned Attention Weights and the Source of the Connectedness Gain
 (Panel (a) reports the attention weights averaged over the rolling windows; the diagonal is the own-variance share and is excluded and hatched so that the cross-series structure is visible. Panel (b) reports the change in each series' own-variance share relative to the FC-TVP-VAR benchmark, for the learned attention and for the one-parameter linear shrinkage calibrated to the same total connectedness index. The two profiles are negatively correlated at -0.945 : the shrinkage takes the most own-variance from the series that hold the most, while the attention takes it from the series that hold the least.)

This direction is not, however, specific to the estimated matrix. Replacing Ω with the scalar matrix $c \cdot I$ removes every cross-term from the bilinear score and leaves the softmax acting on the inner products of the exposure profiles alone. Choosing c so that the resulting index matches the attention-implied one gives $c = 10.03$ and the last row of Table 16. This degenerate specification reproduces the index to two decimal places, the off-diagonal weights with a cross-sectional correlation of 0.952 with the learned ones, the own-variance shares with a correlation of 0.995, and the direction of the redistribution, $+0.950$ against $+0.945$ for the learned layer. The redistribution that separates attention from shrinkage is therefore a property of the bilinear-softmax

form and not of the N^2 estimated parameters. What those parameters add is narrower and measurable. The held-out error falls from 1.7×10^{-3} for the matched scalar to 5.6×10^{-4} for the learned matrix, a threefold reduction, although both remain two orders of magnitude above the 3×10^{-6} of the unchanged decomposition. The estimated weights do something that a single number does not, but none of it shows up in the connectedness index, which one parameter reproduces in two different ways, once as a shrinkage intensity and once as a scale.

Three conclusions follow, and they qualify the comparative results of Section 3.3 without overturning them. First, the 3.72-percentage-point increase in the index is not evidence of a better description of the transmission structure. A single shrinkage parameter reproduces it exactly with a lower held-out error, and the layer's own training objective prefers no adjustment at all. Any study that reports a connectedness gain from a learned component should report this comparison, because the gain is available without learning. Second, the attention redistributes own-variance shares in the direction opposite to shrinkage. This asymmetry is a real property of the mechanism, but it does not come from estimation, since a scalar attention matrix reproduces it. What estimating the N^2 weights adds is a threefold reduction in held-out error relative to the matched scalar, and nothing that the connectedness index reveals. Third, the architecture is over-specified for what it estimates. The heads are redundant, the initialisation is irrelevant, and the learning rate affects only the number of iterations. The evidence supports a single-head bilinear re-weighting trained to convergence and nothing more elaborate.

4. Conclusions

This study measures return transmission across thirteen series from technology, clean energy, cryptocurrency, commodity and macro-financial groups, using daily data from January 2018 to July 2026, and asks whether a learned attention layer placed on a frequency-domain connectedness model improves on the linear decomposition it replaces.

Three empirical findings stand out. The system is highly and persistently connected. The benchmark index averages 55.06% and never falls below 47.76%, and about 60% of the transmitted variance is absorbed within a trading week, so transmission in this panel is mainly a short-horizon phenomenon. The network is also clearly segmented. Cryptocurrencies form a dense block with a mean internal transmission share of 0.229, against 0.014 within the commodity group, and the global equity index, the technology index and the clean-energy index NASC are the main net transmitters. Finally, connectedness depends strongly on the regime. It peaks at 66.28% in March 2020, stays elevated through 2023 and reaches its sample minimum in 2024.

The methodological finding is of a different kind and is probably the more useful one. The attention-weighted decomposition raises the index by 3.72 percentage points, and the difference is robust to the serial dependence of the window sequence. A one-parameter shrinkage towards the uniform matrix reproduces the whole gain with a lower held-out error, and the loss on which the layer is trained prefers no adjustment at all.

Because the index increases mechanically with any flattening of the decomposition, an index gain reported without this comparison establishes nothing. What separates the attention from shrinkage is the direction of its redistribution. It widens the dispersion of own-variance shares where shrinkage necessarily narrows it, and the two adjustments correlate -0.945 . That asymmetry follows from the bilinear-softmax form and not from the estimated weights, since a scalar attention matrix reproduces it, and estimation adds only a threefold reduction in held-out error that the index does not show. The diagnostic, not the index gain, is what the framework contributes.

The results also have practical implications. For investors, the segmentation of the network means that commodities, natural gas in particular, still offer diversification against shocks that originate in equity and cryptocurrency markets, whereas holding several cryptocurrencies diversifies little, because shocks travel quickly within that dense block. The concentration of transmission within a trading week, and of about nine tenths of it within a month, suggests that hedges need to be adjusted at short horizons, since most of the spillover has run its course before a monthly rebalancing. For risk monitoring, the role of the global equity benchmark and the two equity sector indices as the main net transmitters points to the equity factor as the channel through which a shock would most likely reach the rest of the system. For researchers, the ablation offers a simple and inexpensive check. Before a learned component is credited with improving a connectedness measure, its effect on the index should be compared with that of a shrinkage calibrated to the same index and with that of the same functional form with a scalar parameter.

Several limitations bound these conclusions and point to further work. The attention layer is trained to predict the next window's decomposition, and with a 250-day rolling window consecutive decompositions share 249 of their 250 observations. The target is therefore almost constant and the identity map is close to optimal. An objective defined on a quantity that the decomposition is meant to anticipate, such as the realised covariance of the following period, would be a sharper test of the mechanism and is the natural next step. Multi-head attention is redundant as specified, because the heads minimise the same objective independently and are then averaged; assigning each head a distinct frequency band or subsample would give the parameter a meaning. Time variation comes from rolling windows, and a Kalman-filter TVP-VAR with forgetting factors (Antonakakis et al., 2020) would avoid both the loss of the first window and the mechanical effects that appear when extreme observations leave the window. The volatility index is stationary in levels but enters in log changes for homogeneity of units, which over-differences an $I(0)$ series.

ABBREVIATIONS			
Augmented Dickey–Fuller	ADF	Long Short-Term Memory	LSTM
Autoregressive conditional heteroscedasticity Lagrange multiplier test	ARCH-LM	Mean Squared Forecast Error	MSFE
Brock–Dechert–Scheinkman–LeBaron test	BDS	Net total directional connectedness	NET
Frequency Connectedness Time-Varying Parameter Vector Autoregression	FC-TVP-VAR	Ordinary Least Squares	OLS
Generalised Autoregressive Conditional Heteroscedasticity	GARCH	Percentage points	pp
Generalised Forecast Error Variance Decomposition	GFEVD	Time-Varying Parameter Vector Autoregression	TVP-VAR
Graph Attention Network	GAT	Total Connectedness Index	TCI
Graph Attention Network Frequency Connectedness Time-Varying Parameter Vector Autoregression	GAT-FC-TVP-VAR	Variance Inflation Factor	VIF
Jarque–Bera	JB	Vector Autoregression	VAR

REFERENCES

- 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.
- 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.
- Araya, H. T., Aduda, J., & Berhane, T. (2024). A hybrid GARCH and deep learning method for volatility prediction. *Journal of Applied Mathematics*, 2024, 6305525.
- Baruník, J., & Křehlík, T. (2018). Measuring the frequency dynamics of financial connectedness and systemic risk. *Journal of Financial Econometrics*, 16(2), 271–296.
- Bhanja, N., Shah, A. A., & Dar, A. B. (2023). Aggregate, asymmetric and frequency-based spillover among equity, precious metals, and cryptocurrency. *Resources Policy*, 80, 103145.

- Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3), 535–559.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.
- 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.
- Brody, S., Alon, U., & Yahav, E. (2022). How attentive are graph attention networks? In *International Conference on Learning Representations (ICLR)*.
- Chatziantoniou, I., Gabauer, D., & Gupta, R. (2023). Integration and risk transmission in the market for crude oil: New evidence from a time-varying parameter frequency connectedness approach. *Resources Policy*, 84, 103729.
- Cheng, R., & Li, Q. (2021). Modeling the momentum spillover effect for stock prediction via attribute-driven graph attention networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1), 55–62.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Corbet, S., Meegan, A., Larkin, C., Lucey, B., & Yarovaya, L. (2018). Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, 165, 28–34.
- Demirer, M., Diebold, F. X., Liu, L., & Yilmaz, K. (2018). Estimating global bank network connectedness. *Journal of Applied Econometrics*, 33(1), 1–15.
- 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.
- 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.
- 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.
- 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.
- Duan, K., Zhao, Y., Urquhart, A., & Huang, Y. (2023). Do clean and dirty cryptocurrencies connect with financial assets differently? The role of economic policy uncertainty. *Energy Economics*, 127(Part A), 107079.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26.

- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.
- Esparcia, C., Escribano, A., & Jareño, F. (2024). Assessing the crypto market stability after the FTX collapse: A study of high frequency volatility and connectedness. *International Review of Financial Analysis*, 94, 103287.
- Gaies, B., Chaâbane, N., & Bouzouita, N. (2024). Navigating the storm: Time-frequency quantile dependence and non-linear causality between crypto-currency market volatility and financial instability. *The Quarterly Review of Economics and Finance*, 93, 43–70.
- Hassan, M. K., Hasan, M. B., Halim, Z. A., Maroney, N., & Rashid, M. M. (2022). Exploring the dynamic spillover of cryptocurrency environmental attention across the commodities, green bonds, and environment-related stocks. *The North American Journal of Economics and Finance*, 61, 101700.
- Helmi, M. H., Çatık, A. N., & Akdeniz, C. (2023). The impact of central bank digital currency news on the stock and cryptocurrency markets: Evidence from the TVP-VAR model. *Research in International Business and Finance*, 65, 101968.
- Jain, S., & Wallace, B. C. (2019). Attention is not explanation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 3543–3556). Association for Computational Linguistics.
- Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163–172.
- Ji, Q., Bouri, E., Roubaud, D., & Kristoufek, L. (2019). Information interdependence among energy, cryptocurrency and major commodity markets. *Energy Economics*, 81, 1042–1055.
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.
- Koop, G., & Korobilis, D. (2014). A new index of financial conditions. *European Economic Review*, 71, 101–116.
- Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1), 119–147.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3), 1217–1241.
- Le, T.-L., Abakah, E. J. A., & Tiwari, A. K. (2021). Time and frequency domain connectedness and spill-over among fintech, green bonds and cryptocurrencies in the age of the fourth industrial revolution. *Technological Forecasting and Social Change*, 162, 120382.

- Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365–411.
- Lee, T. K., Choi, I., & Kim, W. C. (2025). Symmetry-aware graph neural approaches for data-efficient return prediction in international financial market indices. *Symmetry*, 17(9), 1372.
- Liu, C., Wang, C., Tran, M.-N., & Kohn, R. (2025). A long short-term memory enhanced realized conditional heteroskedasticity model. *Economic Modelling*, 142, 106922.
- Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1412–1421). Association for Computational Linguistics.
- MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, 11(6), 601–618.
- Mai, T., Foley, A. M., McAleer, M., & Chang, C. (2022). Impact of COVID-19 on returns-volatility spillovers in national and regional carbon markets in China. *Renewable and Sustainable Energy Reviews*, 169, 112861.
- Morshed, A. (2025). Graph neural networks and explainable spillovers: Global monetary and oil shocks in GCC financial markets. *Economies*, 13(11), 308.
- Naeem, M. A., & Karim, S. (2021). Tail dependence between bitcoin and green financial assets. *Economics Letters*, 208, 110068.
- Narayan, S., & Kumar, D. (2024). Unveiling interconnectedness and risk spillover among cryptocurrencies and other asset classes. *Global Finance Journal*, 62, 101018.
- Nasir, R. M., He, F., & Yousaf, I. (2025). Relationship of green cryptocurrencies, energy tokens, centralized and decentralized exchange tokens with crypto policy uncertainty. *Research in International Business and Finance*, 75, 102743.
- Pesaran, H. H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1), 17–29.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3), 821–852.
- Sahiner, M. (2024). Volatility spillovers and contagion during major crises: An early warning approach based on a deep learning model. *Computational Economics*, 63(6), 2435–2499.
- Symitsi, E., & Chalvatzis, K. J. (2018). Return, volatility and shock spillovers of Bitcoin with energy and technology companies. *Economics Letters*, 170, 127–130.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. In *International Conference on Learning Representations (ICLR)*.

- Wang, X., Liu, J., & Xie, Q. (2024). Quantile frequency connectedness between energy tokens, crypto market, and renewable energy stock markets. *Heliyon*, 10(3), e25068.
- Wiegrefe, S., & Pinter, Y. (2019). Attention is not not explanation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 11–20). Association for Computational Linguistics.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83.
- Wu, G., & Hu, G. (2024). Asymmetric spillovers and resilience in physical and financial assets amid climate policy uncertainties: Evidence from China. *Technological Forecasting and Social Change*, 208, 123701.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24.
- Younis, I., Du, A. M., Gupta, H., & Shah, W. U. (2024). Dynamic spillover effects and interconnectedness of DeFi assets, commodities, and Islamic stock markets during crises. *International Review of Financial Analysis*, 96(Part A), 103563.
- Yousaf, I., Cui, J., & Ali, S. (2024). Dynamic spillover between green cryptocurrencies and stocks: A portfolio implication. *International Review of Economics & Finance*, 96(Part B), 103661.
- Zhang, C., Pu, X., Cucuringu, M., & Dong, X. (2025). Forecasting realized volatility with spillover effects: Perspectives from graph neural networks. *International Journal of Forecasting*, 41(1), 377–397.