The Journal of Finance and Data Science 10 (2024) 100142

Contents lists available at ScienceDirect

The Journal of Finance and Data Science

journal homepage: www.keaipublishing.com/en/journals/jfds

Research Article

# Tail-driven portfolios: Unveiling financial contagion and enhancing risk management

Tingyu Qu

*Department of Physics, National University of Singapore, 119077, Singapore*

ARTICLE INFO

*Keywords:*
Tail risk
Financial contagion
Comovement

ABSTRACT

In financial markets, tail risks, representing the potential for substantial losses, bear significant implications for the formulation of effective risk management strategies. Yet, there exists a notable gap in understanding the interconnectedness within the global market, particularly when analysing time-series tail data. This study introduces a reliable method for identifying events indicative of tail transitions in financial time-series data. The investigation suggests consistent patterns governing extreme events across diverse industries and different time periods, suggestive of the financial contagion in tail risks. Importantly, time-series tail slopes in specific stocks emerge as viable predictors of price fluctuations in others. These findings offer valuable insights for portfolio diversification and risk mitigation in the interconnected financial market.

## 1. Introduction

The interconnectedness in global financial markets plays a pivotal role in shaping contemporary investment strategies and risk management (Giudici et al., 2020; Minoiu et al., 2015; Raddant and Kenett, 2021; Sun and Chan-Lau, 2017; Wang et al., 2021). The interplay among different financial assets can absorb shocks, leading to an overall resilience; however, the high connectivity can also amplify shocks, creating greater fragility (Martinez-Jaramillo et al., 2019). Financial contagion, namely, how distress in one financial sector triggers the collapse of other sectors (Allen and Gale, 2000; Pericoli and Sbracia, 2003), becomes particularly evident during extreme events (Aloui et al., 2011; Bae et al., 2003). Over the past two decades, instances of global contagion have become more pronounced, as observed in events such as the Global Financial Crisis (2007–2008) (Kenourgios and Dimitriou, 2015), European Sovereign Debt Crisis (2010–2012) (Beirne and Fratzscher, 2013), Brexit (2016) (Belke et al., 2018), and COVID-19 Pandemic (2020) (Akhtaruzzaman et al., 2021). These occurrences serve as tangible manifestations of systemic failure and, consequently, are of substantial value and interest for research.

Comovement in stock returns is a key indicator for the financial contagion among diverse assets (Billio et al., 2012; Diebold and Yilmaz, 2014; Forbes and Chinn, 2004). While one mainstream approach focuses on assessing the comovement of entire markets using aggregated datasets, for example, in the form of stock market indices, the evolving global linkages over time can pose challenges in untangling various influences on the comovement of stock markets (Baur and Jung, 2006; Lopes and Migon, 2002). In fact, the nuanced nature of stock market comovement exhibits significant time variation, often featuring specific lags between distinct assets (Jeon and Von Furstenberg, 1990; Barberis et al., 2005; Reboredo et al., 2017). To uncover dependencies between stocks in different countries by sector, it is necessary to test whether stocks within a specific sector exhibit higher correlations than those from randomly chosen sectors. This entails a careful examination of dependencies at the asset level and their roles on financial contagion dynamics in time series.

*E-mail address:* ty.qu@nus.edu.sg.

Peer review under the responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.jfds.2024.100142
Received 19 January 2024; Received in revised form 21 September 2024; Accepted 21 October 2024

Time-series tail data serves as a valuable font of insights into systemic risks and contagion effects (Bae et al., 2003; Guo et al., 2021). Unlike the normal distribution model, which may overlook extreme movements located several standard deviations away from the mean, measuring the tail of the distribution offers a robust method to unveil these rare events (Kelly and Jiang, 2014; McNeil and Frey, 2000). This capability is instrumental in comprehending the dynamics of the financial market and enhancing the forecasting of sudden price fluctuations (Gofman, 2017). Despite the potential significance of time-series tail data, there is a scarcity of research delving into the intricate interdependencies arising from tail events across diverse financial instruments and markets. Bridging this gap is imperative for achieving a more nuanced comprehension about global financial systems and refining risk management strategies. A number of well-known estimators have been developed for the evaluation of tail exponent, such as Hill, Pickands and Dekkers–Einmahl–de Haan estimators (Dekkers et al., 1989; Hill, 1975; Pickands, 1975). These estimators operate under the assumption that that the exceedance of a random variable $P(X \geq x) \sim L(x) \cdot x^{-k}$, where $k \geq 0$ is the tail exponent or the shape parameter in the power-law form and $L(x)$ is a slowly-varying function (Embrechts et al., 2013). While these estimators have proven effective in various contexts, applying them to financial time-series data introduces several challenges (Clauset et al., 2009). Large sample sizes, the discretization nature of extreme events, and the uncertainty in extracting higher moments due to multiple fitting parameters pose limitations on the applicability of tail estimation in financial contexts (Francq and Zakoian, 2022; Li and Mykland, 2015; Matsui et al., 2013). To address these challenges, simpler power-law models ($P(X \geq x) \sim x^{-k}$) have gained prominence for their ease of fitting and effectiveness in probing the timing of extreme events across various stocks (Alstott et al., 2014; Gabaix et al., 2003; Gillespie, 2014; Qu et al., 2022).

Building upon the aforementioned advancements, this paper proposes a tail-based portfolio that uses a simple power-model to detect extreme events. The tail-correlated portfolio is an extension of the methodology grounded in analysing tail events, emphasizing pairs of stocks that demonstrate comovements during tail events. This means that the decline in the tail of one stock can serve as a precursor for a ripple effect in another stock. The tailored portfolio strategy is thus a new approach in identifying contagion dynamics and predictive analytics in financial markets. The remaining of the paper is constructed as the following. The data source and the methodology are described in Section 2. In particular, I present the results of using wavelet analysis for filtering out noises from the time-series tail slope data. I reveal a universality in the occurrence of extreme events among various stocks in Section 3. The contagion behaviours among various stocks and in various industries are discussed in Section 4, with the indication by lag correlations in the tails. In Section 5, I show the statistical probability of financial distress anticipated by the tail risks and then demonstrate my time-series tail model as a key feature for price forecasting by a decision tree model (with XGBoost), in terms of both root-mean-square deviation and the lag in the predication. I summarise my findings and raise an outlook in Section 6.

## 2. Data source and methodology

To cover diverse situations, I select 174 representative stocks, which includes all the 11 sectors set by the Global Industry Classification Standard – with nearly close weighting for the number of stocks in each sector to minimize the bias of study in a specific sector (See Appendix A). These stocks range from small cap to mega cap in terms of their market values and are exchanged in different countries (See Appendix A). Another important criterion is that the selected stock needs to have at least 20 years' history since its initial public offering, such that there are enough data for analysis and comparison throughout a long period. I extract the data from Yahoo Finance (https://finance.yahoo.com). Daily adjusted closing price $p(t)$ is used as the input, which screens out the abrupt changes of the price caused by dividends, stock splits, and new stock offerings. The gross return $Q(t,h)$ of $p(t)$ is defined as $Q(t,h) = p(t+h) - p(t)$, where $h$ is the lead time.

The methodology for calculating the time-series tail slope can refer to the author's former investigation (Qu et al., 2022). Concisely, by scanning a moving window with fixed size ($T$) in time, the tail section in a probability distribution $P(x)$ follows a power-law principle with the form $\sim x^{-k(t)}$ (where $x$ and $k$ refer to the gross return and tail slope, respectively) at the forward time $t$, resulting in a linear form of $k(t)$. For a reliable estimation of $k(t)$, I define the tail section with both upper and lower thresholds for $P(x)$, with $P^* = P(x > x^*)$ and $P^{**} = P(x > x^{**})$, respectively (See schematics in Fig. 1a). The tail section is then determined by the domain where $x^* < x < x^{**}$. The upper limit $P^*$ is crucial in separating the tail section (that reflects the extreme events) from the main body (the non-extreme events) in the probability distribution. Equally important is to handle the outliers in tail distribution, therefore, I define a lower limit $P^{**} = P(x > x^{**}) \geq \frac{n}{T-h}$, where $n$ is a constant that determines the cut-off of the under-sampled events. I have shown that the method is applicable to both fat-tailed and thin-tailed distributions and the transition in $k(t)$ can be associated with an extreme event in the previous work (Qu et al., 2022).

Here, I show the optimal parameters for the evaluation of $k(t)$. I target at daily gross return by setting $h = 1$ working day to reflect the price change with high sensitivity. The window size $T$ is set as 4 years to cover sufficient points for fitting the tail slope. The upper limit $P^*$ is set as 0.1 such that in this regime, $ln(P)$ vs. $ln(x)$ shows a linear relation (See the example in Fig. 1b). I find that $n = 10$ is effective to rule out the outliers in tail section.

I then perform wavelet convolution to identify the transitions in $k(t)$, with the formulae as

$$CWT_{k(t)}(a,b) = \int_{-\infty}^{\infty} k(t) \cdot \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dt$$

$$CWT'_{k(t)} = CWT_{k(t)} \Big/ k(t)$$

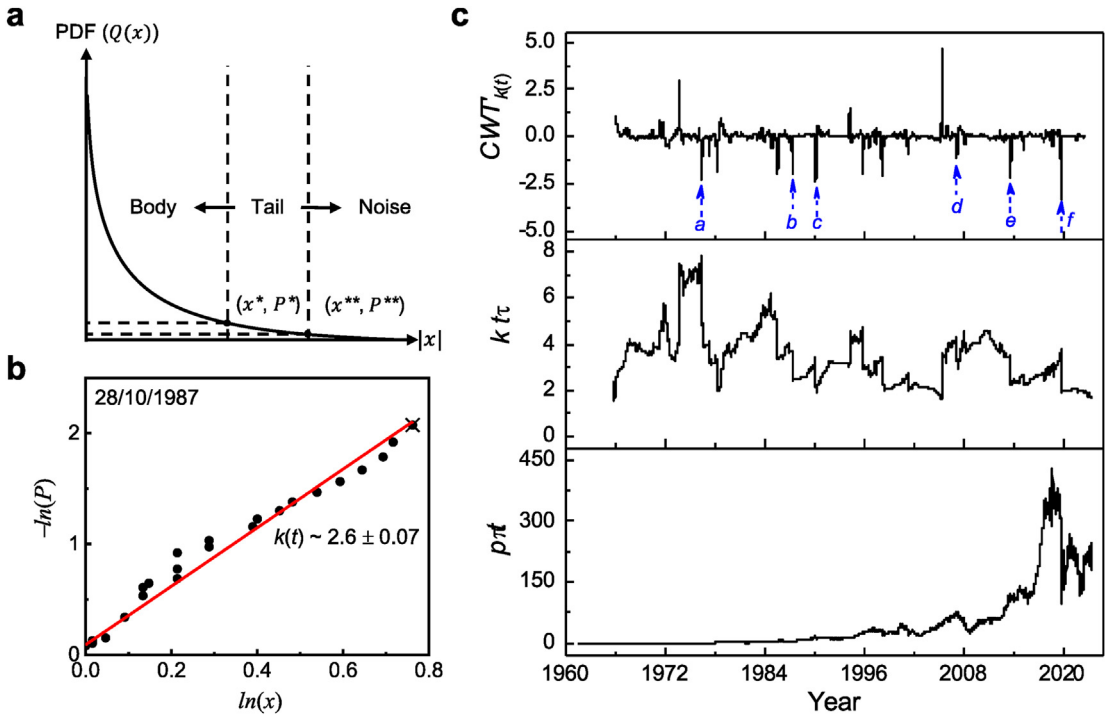


**Fig. 1. a,** Definition of the tail in the probability density function. **b,** An example demonstrating the extraction of the tail slope ($k$) of Boeing's stock on Black Monday by a least square method. The fitting is shown by the red linear curve. The cross refers to the lower threshold for the tail-slope estimation. **c**, Illustration of tail transformation of time-series financial data based on the adjusted close price ($p(t)$) of Boeing's stock (bottom panel). The transformation from $p(t)$ to $k(t)$ is shown in the middle panel. A number of financial distresses are identified by the sharp declines in $CWT'_{k(t)} - t$ curves (top panel), namely, (a) Dan-Air Boeing 707 crash in 1977; (b) Black Monday in 1987; (c) The rise of Airbus in late 1980s; (d) Global financial crisis in 2007∼2008; (e) Malaysia Airlines Flight 370 disappearance in 2014; (f) Covid-19 in 2020.

where $k(t)$ is the input signal of time-series tail-slope data, $\psi$ is the wavelet function, $a$ is a constant that scales the frequency and $b$ is a translation factor that provides time localization (Daubechies, 1992).

It is noted that the normalization term $CWT'_{k(t)}$ is necessary for resolving the transitions in time extension by minimizing the noises introduced by the convolution. The choice of wavelet type ($\psi$) (by comparing Gaussian, Mexican hat, and Morlet wavelets) and the values of $a$ can be optimized by performing the autocorrelation analysis to exam the lag between $k(t)$ and $CWT'_{k(t)}$ (Ghaderpour et al., 2021; Grané and Veiga, 2010). A quick decrease in autocorrelation to nearly zero with a small lag via Gaussian wavelet suggests it is the most effective to probe the step-like tail transitions and increase signal-to-noise ratio (See details shown in Appendix B). In addition, tuning the scale factor can adjust the trade-off between time and frequency localization in the analysis. In my case, $a$ is set as 40 working days to effectively filter out the noises.

Using Boeing as an example, I compare the $p(t) - t$, $k(t) - t$ and $CWT'_{k(t)} - t$ curves in Fig. 1c. I just name a few significant events (Dan-Air Boeing 707 crash in 1977; Black Monday in 1987; The rise of Airbus in late 1980s; Global financial crisis in 2007∼2008; Malaysia Airlines Flight 370 disappearance in 2014; Covid-19 in 2020) detected by the sharp transitions in $CWT'_{k(t)} - t$ curve, which shows the effectiveness of the method in probing the tail risks. Importantly, the tail model transforms inconspicuous price changes in Boeing to notable spikes in tails (especially during 1960s–1990s), allowing viable analysis of the patterns in those extreme events.

## 3. Universality in the tail risks

I now shift my focus to investigating common features among stocks through the analysis of tail transitions. Despite the diversity among the studied stocks, a noteworthy universality emerges in the number of tail risks, governed by a critical scaling law. The scaling factor, denoted as $k_s$, plays a pivotal role in the analysis as it determines the threshold for identifying tail events. To clarify, $k_s$ represents the critical scaling level. I then introduce $n_k$, which represents the number of significant tail events observed per unit of time during the study. In other words, $n_k$ is a measure of the frequency of extreme tail transitions (where $k(t)$ falls below $k_s$) and serves as an indicator of tail risk magnitude. Understanding the importance of $k_s$ is essential as it directly influences the observations on the interconnectedness of tail risks in global financial markets. With this understanding, I proceed to interpret my findings. I plot $n_k - k_s$ curves encompassing various sectors in Fig. 2a and different exchange markets in Fig. 2b. The alignment between the number of tail events and $k_s$ is consistent across diverse sectors and exchange markets. Strikingly, in the regime where substantial noise is filtered out ($k_s \leq -0.5$), the data

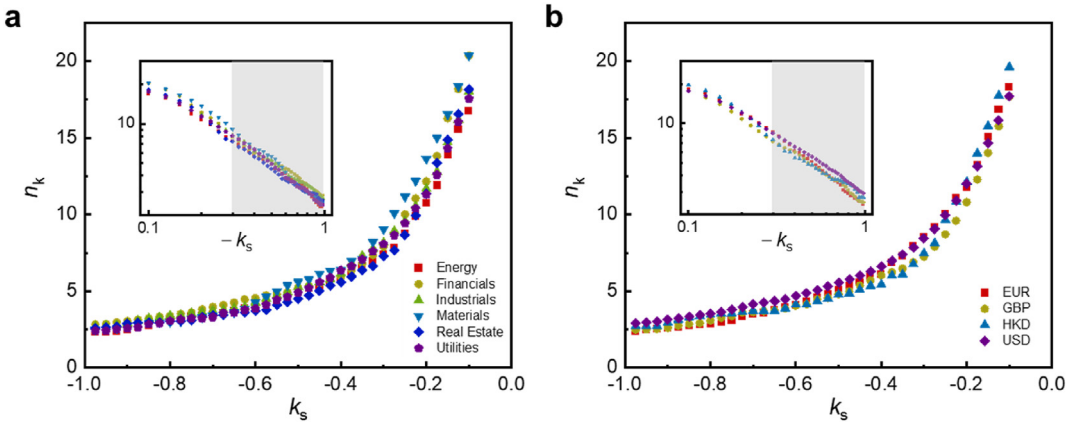


**Fig. 2.** Universality of the number of tail events in various sectors (**a**) and exchange markets (**b**). Inset: $\log(n_k)$vs.$\log(-k_s)$, showing a nearly linear relation with $k_s > -0.2$ (as shown in the shaded region).

collapse onto a universal linear relation $\log(n_k) - \log(k_s)$ (See the insets), meaning $n_k \sim {k_s}^{\alpha}$, where $\alpha$ is an exponent. This suggests that the tail risks are closely interconnected in the diverse financial market and there is likely a universal law governing the occurrence times with the risks. However, in the regime where $k_s > -0.2$, the data from different sectors or exchange markets tend to diverge. This divergence is reasonable, as below a certain scaling threshold, the fraction of noises in the wavelet transform becomes noticeable, affecting the accurate determination of actual tail transitions. This finding agrees with other relevant investigations (Mantegna and Stanley, 1995; Lux and Marchesi, 1999).

## 4. Financial contagion

As the scaling law in the $n_k - k_s$ curves demonstrate the interconnectedness in global financial markets, I further investigate the financial contagion based on the tails. I detect a coupled behaviour between the two tail transitions from two different stocks under the condition of $d_1 \leq t_{k1} - t_{k2} \leq d_2$, where $t_{k1}$ and $t_{k2}$ refer to the timing of the tail transitions in two stocks, respectively; $d_1$ and $d_2$ are the constraint in time length. I avoid looking at the case $t_{k1} = t_{k2}$ because I aim at finding a certain lag between the two transitions in order to tell the contagion from one stock to the other.

I then count the number of coupled tail transitions to evaluate the interconnectivity in the pair of stocks. For example, Fig. 3a plots the $CWT'_{k(t)} - t$ curves of General Electric (GE) and JPMorgan Chase & Co (JPM), which shows a number of coupled sharp tail declines in the two companies. It is intriguing to note that GE consistently lags behind JPM by approximately within three weeks concerning nine detectable tail transitions since the start-up of GE. Remarkably, the inverse scenario is not observed, indicating an asymmetry in the lag correlation between GE and JPM. I provide detailed instances to illustrate the price changes for both companies during the detected tail transitions in Fig. 3b.

It is a compelling question whether the identified lag correlation between GE and JM is reasonable in the context of actual finance market. Indeed, there exists a financial courtship between JPM and GE. JPM played a pivotal role in providing financial support and coordinating the merger process for the foundation of GE in 1982 (DeLong, 1991). Subsequent collaborations between the two entities are evident in many aspects, including JPM's capital support for GE's advancements in new energy management and digital technologies. This interdependence between the two entities implies a similar market cycle, which is indeed reflected by their tail correlations. Importantly, the model suggests that tail risks detected in JPM may serve as precursors for comparable risks in GE, and therefore offers valuable insights to mitigate potential losses associated with investments in GE based on the tail declines in JPM. Another illustration of lagged comovement is evident in the relationship between Chevron Corporation (CVX) and Expeditors International of Washington (EXPD), where distress in CVX can serve as an antecedent indicator for EXPD (See Appendix C for details).

It is worthwhile to note that the tail slope, unlike raw prices or gross returns, is non-dimensional. This makes it independent of the absolute scale or magnitude of the financial instrument being studied. For instance, even though JPM and GE may exhibit vastly different prices, the tail slope captures risks in a similar range, unless an extreme event disrupts the market. This allows for comparison across different stocks or asset classes without needing to adjust for price levels or units.

I now validate the model across different asset classes and time periods. First, to visualize the interconnectedness across diverse sectors and exchange currencies, I construct 2D mapping and network plots that illustrate the strength of connections based on the number of coupled tail risks (See Fig. 4). The 2D heat mapping (Fig. 4a and b) reveals that stocks grouped within the same sector (e.g., Consumer Staples, Energy, or Finance) or the same exchange market (e.g., CNY, GBP, or SGD) tend to exhibit a high connectivity in terms of the occurrence of extreme events. However, strong interconnections also emerge between different sectors (Consumer Staples, Energy, Finance, and Industrials) and exchange markets (GBP, EUR, and USD), as evident in the color-coded counts. In addition, the network plots (Fig. 4c and d) highlight directional influences, shedding light on how the contagion flows from one sector to another

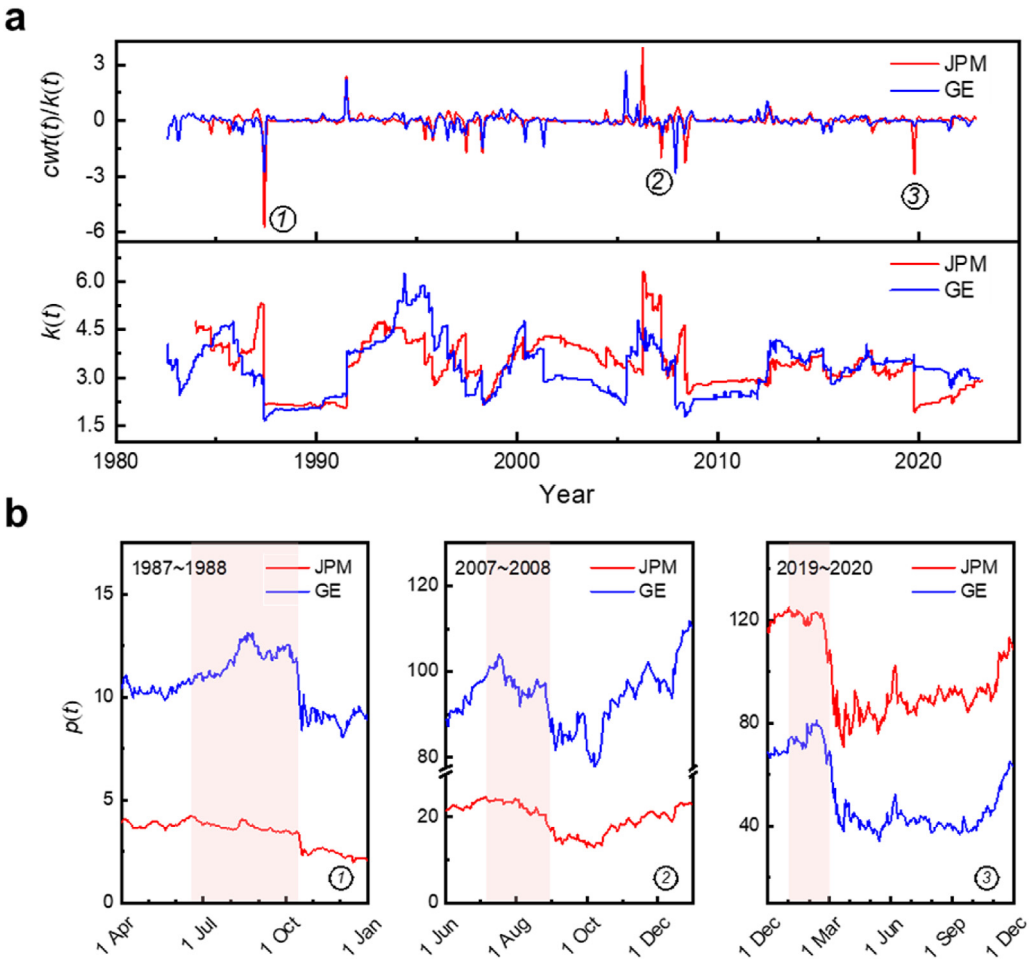


**Fig. 3.** **a**, Time-series tail curves with labelled co-transitions. **b**, Price comovements between JPM and GE during identified timings (labelled as (1) Black Monday, (2) Global financial crisis, and (3) Covid-19), as detected by co-tail declines (if $k(t)$ drops below $-0.2$). Notably, in all three instances, JPM exhibited a price decline preceding the response from GE, as depicted by the shaded regions.

(based on a certain number of comoved tail events with a lag time). This directional influence is discernible in connections such as those from Consumer Staples to Energy and from GBP to USD, revealing the temporal sequence of tail risk transmissions. Contrary to the conventional belief that tail risks might be confined to specific sectors or exchange markets, the network analysis suggests a more pervasive and global contagion pattern. This observation aligns with recent studies employing alternative approaches (Raddant and Kenett, 2021), emphasizing that tail risks are not limited to specific sectors or exchange markets but rather exhibit global contagiousness.

Second, I examine comovement patterns in tail risks across different periods. Fig. 5 presents the percentage of studied stocks exhibiting tail slopes below specified thresholds over time. The time-series plot reveals two prominent clusters of financial distress: in 1987 (Black Monday) and 2020 (Covid-19). The Black Monday crash is clearly detected by the model, with almost 90% of the 174 studied stocks experiencing sharp tail drops below 3. Similarly, during the Covid-19 pandemic, a large number of stocks exhibited tail risks, with the effects persisting for over three years. These findings suggest that during economic crises, financial contagion spreads rapidly across sectors, resulting in correlated tail transitions across multiple industries. Other contagion events such as Asian financial crisis in 1997 and global financial crisis in 2008 are also clearly detected in the plot.

## 5. Price forecasting

Transitioning to price forecasting, I now evaluate how tail declines can serve as useful indicators for predicting future price declines. I first plot the distribution of the normalized price return before $(Q(t-h,h)/p(t-h))$ and after $(Q(t,dt)/p(t))$ all the detected tail declines (See Fig. 6a and b). Subsequently, I look into the expectation value of the distribution, denoted as $E(p)$, to exam the consequence of price changes following tail declines. Here, $E(p\uparrow)$ signifies the expectation of an increase in price turn after the identified tail declines, while $E(p\downarrow)$ denotes the expectation of a decrease in price return after the detected tail declines. A distinct shift towards $E(p\downarrow)$ can be

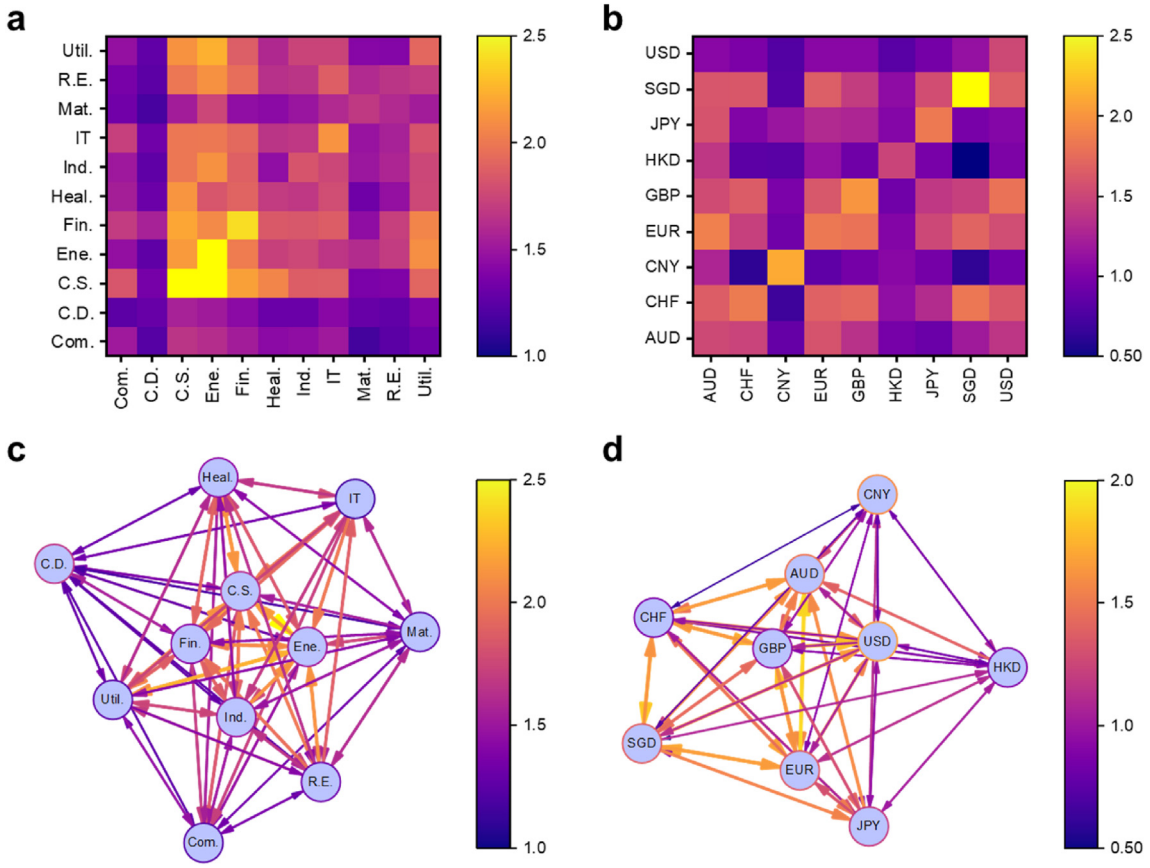


**Fig. 4.** Network visualizations of interconnected global financial tail risks across different markets. Stocks are organized by sectors (**a**) and exchange currencies (**b**). The sector/currency names, in abbreviated form, on the horizontal axis correspond to those on the vertical axis. The figures utilize color-coded counts to illustrate detected comovements in tails, where comovements involve objects on the vertical axis leading those on the horizontal axis within a 7∼14 days' duration. The value assigned to each block is computed as the average of all values within the same sector/currency, as illustrated in Appendix D. The network plots corresponding to sectors (**a**) and currencies (**b**) are depicted in (**c**) and (**d**), respectively.

seen in Fig. 6a, highlighting a noticeable trend of price decline following tail declines. However, in Fig. 6b, the difference between $|E(p\downarrow)|$ and $E(p\uparrow)$ is marginal, suggesting that solely looking at the tail declines of the stock itself is not timely to anticipate its price volatility. Drawing from the insights gained about financial contagion through tail risks, I leverage tail transitions as precursors for forecasting price volatility over a specified time lag ($dt$). This forms the basis of what I term a "tail-correlated portfolio", where pairs of stocks exhibiting high comovements in tail events are selectively filtered for enhanced risk anticipation (See the portfolio in Appendix E). The tangible outcome of this strategy is demonstrated in Fig. 6c, showcasing an enriched probability density of $E(p\downarrow)$. This presents a more nuanced perspective for anticipating price volatility driven by the portfolio with tail correlations.

In the comparison of my model with other stochastic processes, I assess $E(p)$ as a function of $dt$ based on three different decision models, namely, a random stochastic process (i), a price-driven decision process (ii), and a tail-driven decision process (iii) (For detailed information, please refer to Appendix E). As shown in Fig. 7, $E(p)$ exhibits a gradual growth with the augmentation of $d_t$ in both (i) and (ii). This aligns with the general growth of $p(t)$ throughout the timeline in most stocks, indicating the probability of a following price decline cannot be adequately represented by the preceding price decline. In stark contrast, within the tail-driven decision model, $E(p)$ declines as $dt$ increases up to two weeks, implying a robust correlation between the tail decline and consequent price decline. Also, $E(p)$ remains the negative sign up to six weeks, though after that it follows the overall economic growth. More importantly, using a tail-correlated portfolio (iii-b) allows to obtain a more negative $E(p)$ (down to $-$ 2.8% as the minimum) than the tail-driven decision model without incorporating correlated pairs of stocks (iii-a). This strongly implies that the comovement behaviour observed in tail transitions can effectively be leveraged to mitigate financial risks and anticipate price volatility in the near future.

Finally, I highlight the role of time-series tail data as a crucial feature for machine learning in price forecasting. Here, I choose the open-source software library, XGBoost (Brownlee, 2016; Chen and Guestrin, 2016) to build a decision tree model (See illustrations in Appendix F) to forecast the price with $h$ days in advance (See Fig. 8a). I use $p(t)$ as the output label and input a number of features for the training, including the commonly used technical indicators ($p(t-h)$, the standard deviation ($\sigma_{p(t-h)}$), the moving average ($\overline{p(t-h)}$) and the exponential moving average in a rolling window). Also, I incorporate the tail risk indicator $k(t)$ as an additional feature. The model comprises ten layers, and I partition the dataset into training and test samples with a ratio around 7/3. Interestingly, $k(t)$ emerges as the

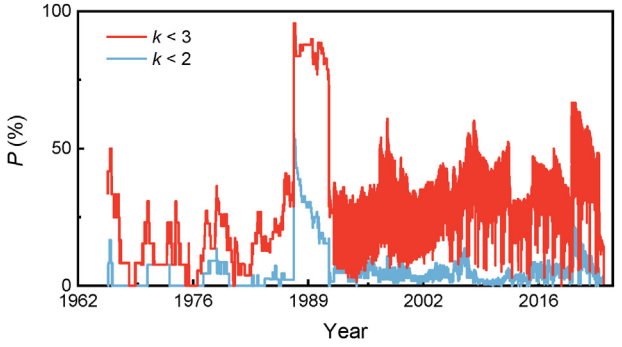


**Fig. 5.** Time-series plot of the percentage of stocks exhibiting tail risks across different periods. The plot tracks the percentage of stocks below specific tail slope thresholds (red line for $k < 3$ and blue line for $k < 2$) over time. Peaks in the plot correspond to periods of significant financial stress, such as the Black Monday in 1987 and the Covid-19 in 2020, where a large proportion of stocks experienced extreme tail events.

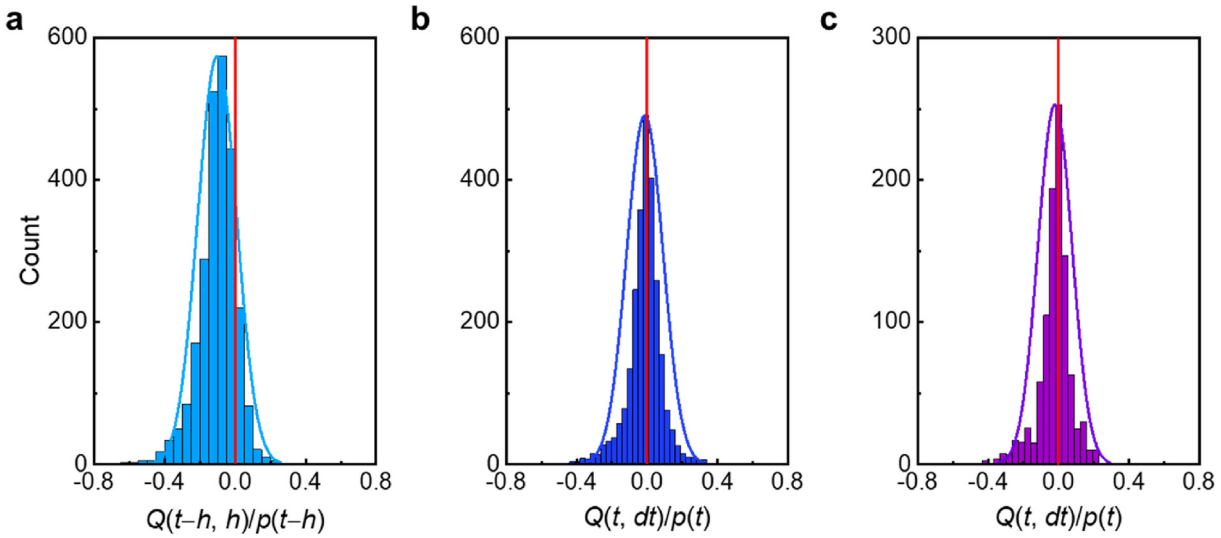


**Fig. 6.** Frequency histogram of the normalized price return before the identified tail declines, denoted as $Q(t-h,h)/p(t-h)$ (depicted in (**a**)), and after the detected tail declines, represented as $Q(t,dt)/p(t)$, illustrated in (**b**) with all detected tail declines and in (**c**) with only the detected tail declines from the tail-correlated portfolio. The author sets $h = dt = 14$ days.

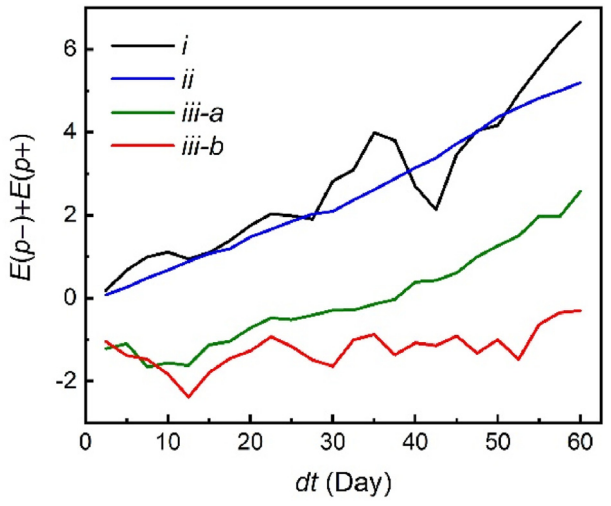


**Fig. 7.** Expectation of normalized price return plotted against time lag ($dt$) for different decision processes: (i) a random stochastic process, (ii) a price-driven decision process, (iii-a) a tail-driven decision process, and (iii-b) a tail-driven decision process based on a tail-correlated portfolio.

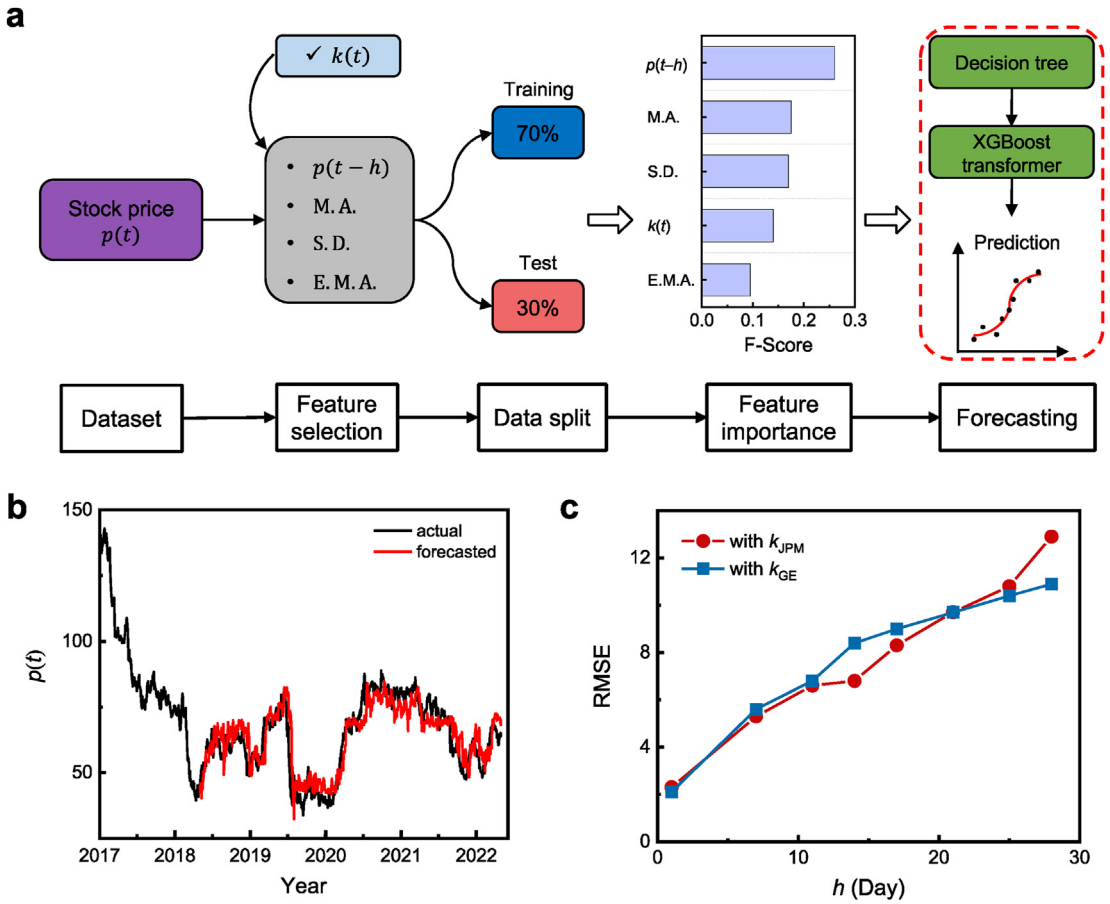


**Fig. 8.** Time-series stock price forecasting using an XGBoost decision-tree machine learning model. (**a**) Forecasting pipeline: Four primary technical indicators (lagged price, standard deviation, moving average, and exponential moving average) alongside the tail risk indicator are utilized as features. The training/test dataset ratio is set at 7/3, and F-scores rank the importance of features. (**b**) Forecasted vs. actual prices for GE stock with a lead time of three weeks. (**c**) Comparison of RMSE values in forecasting with tail data from GE and JPM.

fourth most important feature after the training/test optimization, which implies its potential in capturing price movements. An illustrative example in Fig. 8b showcases the forecasted price $p(t)$ for GE, with the incorporation of $k(t)$ from JPM as the additional feature, demonstrating its utility in predicting price changes with a lead time of three weeks since the year 2018.

Expanding on the insights gained from portfolio analysis, I delve into the impact of utilizing tail risks from JPM as precursors for forecasting stock price in GE. By introducing $k(t)$ from JPM as the tail risk indicator alongside standard technical indicators in GE, I assess the deviation for price forecasting as a function of lead time. Root mean square error (RMSE) is employed for evaluation due to its sensitivity to outliers (compared to mean absolute error) and its interpretability (compared to mean square error) because it shares the same units as the predicted variables. Therefore, I show direct comparison of the RMSE between forecasts using tail data from GE and JPM (Fig. 8c). Between a lead time of 12–20 days, the forecasting for GE stock prices using JPM's tail data shows greater accuracy compared to using GE's own tail data. This result hints cross-sector tail correlations in forecasting price movements, supporting that tail risks in one sector can provide predictive signals for another. The finding also highlights the value of incorporating $k(t)$ for more accurate predictions via the decision-tree model, especially during extreme events like the Covid-19 pandemic, and also for achieving improved long-term forecasting outcomes.

## 6. Summary and outlook

In summary, I developed a time-series tail-risk portfolio that detects the financial distress and the consequent risk contagion across different asset classes and different periods. Through wavelet analysis, there could exist a universal law governing the tail risks in global financial market, particularly pronounced at critical scaling levels. Importantly, I find that specific group of stocks can serve as precursors for effectively anticipating the tail transitions in other groups, which helps improve the accuracy of volatility forecasting for select stocks through machine learning methodologies. My work not only contributes to the construction of tail-based risk management portfolios but also underscores the significance of tails as pivotal features for forecasting extreme events. This dual perspective offers valuable insights for investors and financial analysts, providing enhanced tools for managing risk and predicting extreme market conditions.

Despite the above findings, I draw four limitations in this research. First, the use of daily price data, as opposed to more granular minute or second data, might limit the model's suitability for very short-term price forecasting. Future research could explore the implications of utilizing higher-frequency data. Second, while the main focus lies in comparing machine learning models with and without tail features, it is important to note that the parameters in the models are not fully optimized. Future studies could delve deeper into feature engineering to refine the predictive accuracy of the models. Third, the tail correlation analysis assumes a simple power-law in data distribution, which is intentional for straightforward effectiveness, but it might sacrifice certain accuracy in specific scenarios. The choice of the threshold of tail slope, once set, is applied for all datasets, therefore, still subjective. The trade-off between simplicity and the comprehensiveness of the model requires more investigations in future. Finally, the forecasting relies on price rather than return, which may overemphasize historical data. Nevertheless, price forecasting can offer values to equity analysts who need to study the target price for individual stocks.

## Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

I would like to acknowledge the comments and suggestions of the editor-in-chief and reviewers. I also thank Arnold Doray for the fruitful discussions.

## Appendix A. Information of the stock data

**Table A1**
List of stocks with ID (ticker symbols), sector and exchange currency.

| Stock ID | Sector | Currency | Stock ID | Sector | Currency | Stock ID | Sector | Currency |
|---|---|---|---|---|---|---|---|---|
| 000002.SZ | Real Estate | CNY | BT-A.L | Communication Services | GBP | MSFT | Information Technology | USD |
| 000538.SZ | Healthcare | CNY | C | Financials | USD | MSI | Communication Services | USD |
| 000625.SZ | Consumer Discretionary | CNY | C6L.SI | Industrials | SGD | MU | Information Technology | USD |
| 000651.SZ | Consumer Discretionary | CNY | CNI.AX | Real Estate | AUD | NDX1.DE | Materials | EUR |
| 000831.SZ | Materials | CNY | CVX | Energy | USD | NESN.SW | Consumer Staples | CHF |
| 0016.HK | Real Estate | HKD | D05.SI | Financials | SGD | NFLX | Communication Services | USD |
| 0347.HK | Materials | HKD | DBK.DE | Financials | EUR | NG.L | Utilities | GBP |
| 0386.HK | Energy | HKD | DD | Materials | USD | NKE | Consumer Staples | USD |
| 0700.HK | Information Technology | HKD | DG.PA | Industrials | EUR | NOC | Industrials | USD |
| 0728.HK | Communication Services | HKD | DIS | Communication Services | USD | NOKIA.HE | Communication Services | EUR |
| 0855.HK | Utilities | CNY | DTE.DE | Communication Services | EUR | NOVN.SW | Healthcare | CHF |
| 0992.HK | Information Technology | HKD | DUK | Utilities | USD | NVDA | Information Technology | USD |
| 2388.HK | Financials | HKD | DY | Communication Services | USD | NYT | Communication Services | USD |
| 2897.T | Consumer Staples | JPY | EBAY | Consumer Discretionary | USD | O | Real Estate | USD |
| 3382.T | Consumer Staples | JPY | EL | Consumer Discretionary | USD | O39.SI | Financials | SGD |
| 4186.T | Information Technology | JPY | ELM.L | Materials | GBP | OR.PA | Consumer Discretionary | EUR |
| 4911.T | Consumer Discretionary | JPY | ENB | Utilities | USD | ORA.PA | Communication Services | EUR |
| 5401.T | Materials | JPY | EOAN.DE | Utilities | EUR | ORCL | Information Technology | USD |
| 600009.SS | Industrials | CNY | EQT | Utilities | USD | ORG.AX | Utilities | AUD |
| 600029.SS | Industrials | CNY | EXPD | Industrials | USD | PCAR | Industrials | USD |
| 600221.SS | Industrials | CNY | EXPE | Consumer Discretionary | USD | PEP | Consumer Staples | USD |
| 600519.SS | | CNY | F | | USD | PFE | Healthcare | USD |

(*continued on next page*)

**Table A1** (*continued*)

| Stock ID | Sector | Currency | Stock ID | Sector | Currency | Stock ID | Sector | Currency |
|---|---|---|---|---|---|---|---|---|
| | Consumer Discretionary | | | Consumer Discretionary | | | | |
| 600600.SS | Consumer Staples | CNY | FDX | Industrials | USD | PG | Consumer Staples | USD |
| 600606.SS | Real Estate | CNY | FJTSY | Communication Services | USD | QCOM | Communication Services | USD |
| 600688.SS | Energy | CNY | FLEX | Industrials | USD | RELX | Information Technology | USD |
| 6370.T | Utilities | JPY | FMG.AX | Materials | AUD | RIO.L | Materials | GBP |
| 6951.T | Information Technology | JPY | FORTUM.HE | Utilities | EUR | ROG.SW | Healthcare | CHF |
| 6952.T | Information Technology | JPY | FWRD | Industrials | USD | RR.L | Consumer Discretionary | GBP |
| 7203.T | Consumer Discretionary | JPY | GD | Industrials | USD | RWE.DE | Utilities | EUR |
| 7267.T | Consumer Discretionary | JPY | GE | Utilities | USD | SAN.PA | Healthcare | EUR |
| 7731.T | Information Technology | JPY | GFC.PA | Real Estate | EUR | SBUX | Consumer Staples | USD |
| 7951.T | Consumer Discretionary | JPY | GFF | Financials | USD | SCHN.SW | Industrials | CHF |
| 7974.T | Consumer Discretionary | JPY | GIS | Consumer Staples | USD | SFR.L | Materials | GBP |
| 900942.SS | Consumer Discretionary | CNY | GME | Consumer Discretionary | USD | SGRO.L | Real Estate | GBP |
| 9202.T | Industrials | JPY | GPK | Materials | USD | SHEL | Energy | USD |
| 9984.T | Financials | JPY | HOLN.SW | Materials | CHF | SIE.DE | Information Technology | EUR |
| A | Information Technology | USD | HPQ | Information Technology | USD | SONY | Information Technology | USD |
| AAPL | Information Technology | USD | HSBA.L | Financials | USD | SPG | Real Estate | USD |
| ADBE | Information Technology | USD | HTHIY | Information Technology | USD | SSE.L | Energy | GBP |
| AE | Utilities | USD | HXL | Materials | USD | STAN.L | Financials | GBP |
| AEO | Consumer Discretionary | USD | IBM | Information Technology | USD | TCOM | Consumer Discretionary | USD |
| AFX.DE | Healthcare | EUR | IMO | Energy | USD | TKA.DE | Materials | EUR |
| AIR | Industrials | USD | INTC | Information Technology | USD | TLS.AX | Communication Services | AUD |
| AMD | Information Technology | USD | ISRG | Healthcare | USD | TSCO.L | Consumer Staples | GBP |
| AMGN | Healthcare | USD | JD.L | Consumer Staples | GBP | TSM | Information Technology | USD |
| AMZN | Consumer Discretionary | USD | JPM | Financials | USD | TTE.PA | Energy | EUR |
| ASML | Information Technology | USD | KER.PA | Consumer Discretionary | EUR | U11.SI | Financials | SGD |
| BA | Industrials | USD | KO | Consumer Staples | USD | UBSG.SW | Financials | CHF |
| BARC.L | Financials | GBP | LHA.DE | Consumer Staples | EUR | UHR.SW | Consumer Discretionary | CHF |
| BAYN.DE | Healthcare | EUR | LISN.SW | Consumer Staples | CHF | UL | Consumer Staples | USD |
| BBWI | Consumer Discretionary | USD | LMT | Industrials | USD | UPS | Industrials | USD |
| BIO | Healthcare | USD | LONN.SW | Healthcare | CHF | UVV | Communication Services | USD |
| BKNG | Consumer Discretionary | USD | MAR | Real Estate | USD | VOW3.DE | Consumer Discretionary | EUR |
| BLND.L | Real Estate | GBP | MBG.DE | Consumer Discretionary | EUR | WMT | Consumer Staples | USD |
| BMW.DE | Consumer Discretionary | EUR | MC.PA | Consumer Discretionary | EUR | X | Materials | USD |
| BNP.PA | Financials | EUR | MCD | Consumer Staples | USD | XOM | Energy | USD |
| BP.L | Energy | GBP | MMM | Industrials | USD | Z74.SI | Communication Services | SGD |
| BRK-A | Financials | USD | MS | Financials | USD | ZURN.SW | Financials | CHF |

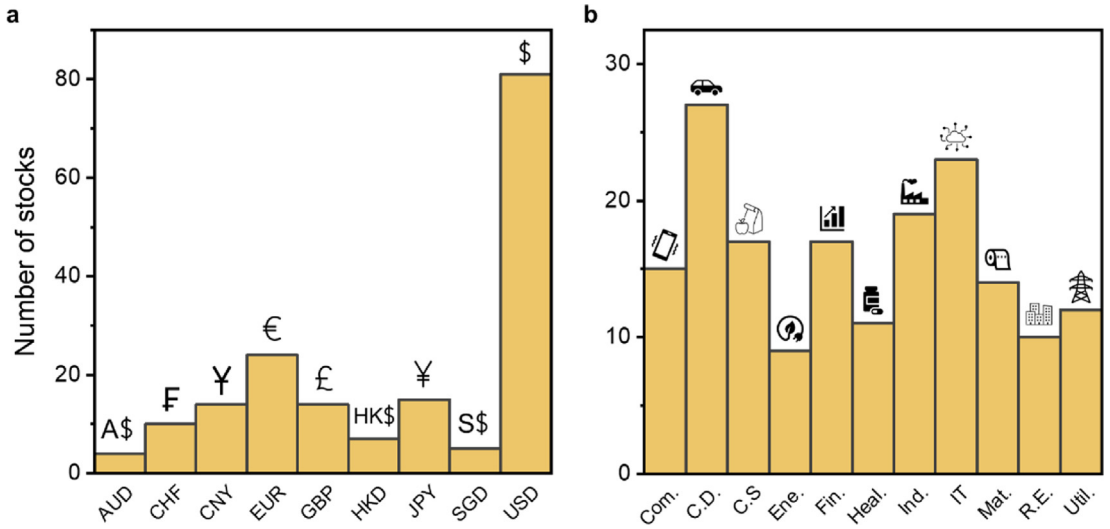


**Fig. A2.** Number of stocks classified by sector and exchange currency. Sectors include: Communication Services (Com.), Consumer Discretionary (C.D.), Consumer Staples (C.S.), Energy (Ene.), Financials (Fin.), Healthcare (Heal.), Industrials (Ind.), Information Technology (IT), Materials (Mat.), Real Estate (R.E.) and Utilities (Util.). Exchange currencies include: Austrian Dollar (AUD), Swiss Franc (CHF), Chinese Yuan (CNY), Euro (EUR), British Pound (GBP), Hong Kong Dollar (HKD), Japanese Yen (JPY), Singapore Dollar (SGD) and United States Dollar (USD).

## Appendix B. Evaluation of the wavelet forms

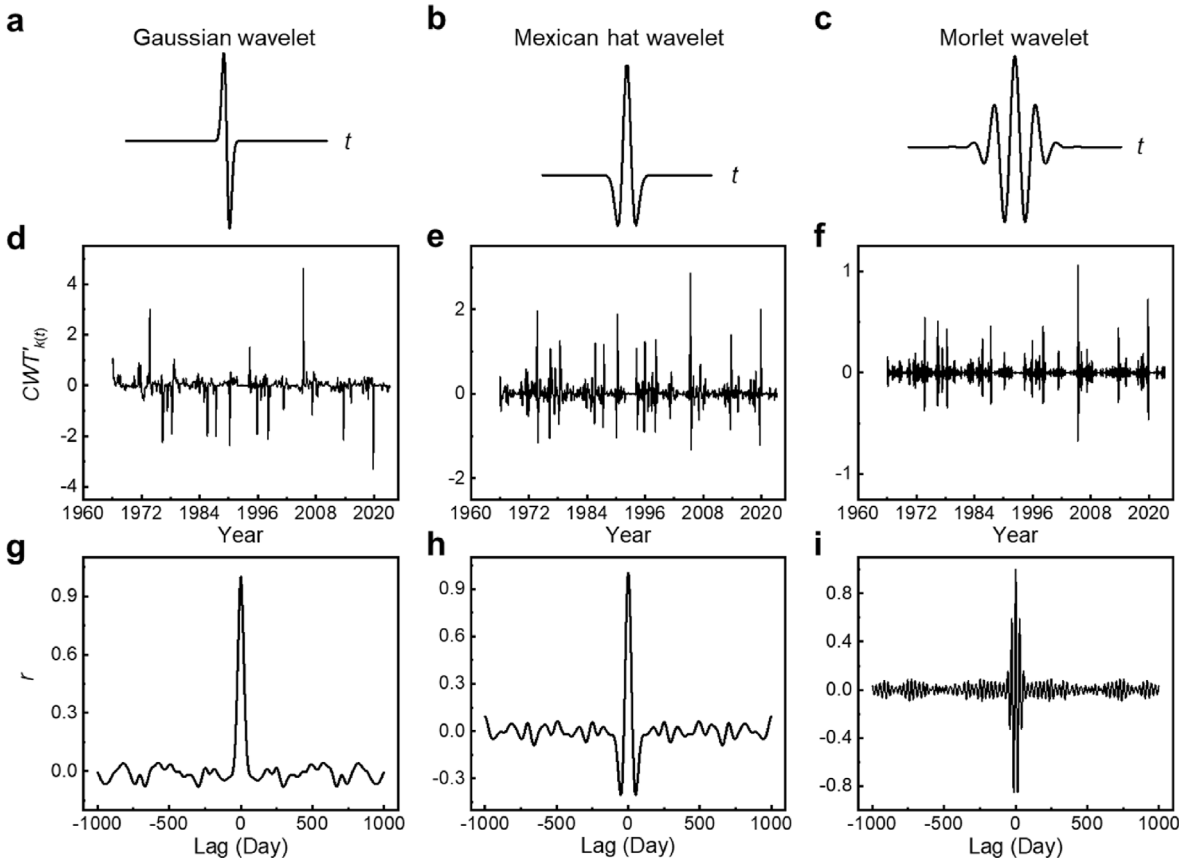


**Fig. B.** Comparative analysis of three wavelet functions (Gaussian, Mexican Hat, Morlet) for extracting transitions in time-series tail data. Wavelet forms: The subfigures (**a**–**c**) in the figure illustrate the three different wavelet functions – Gaussian, Mexican Hat, and Morlet. Each subfigure provides

a visual representation of the shape and characteristics of the respective wavelet function. Convolution results: The subfigures (**d**–**f**) show the convolution of the time-series tail signal $k(t)$ after applying the wavelet transformation using the Gaussian, Mexican Hat, and Morlet wavelet functions, respectively. The convolution process involves the integration of the wavelet function with $k(t)$ at different time points, highlighting the time periods of interest and potential transitions. Autocorrelations: The subfigures (**g**–**i**) present the autocorrelations corresponding to the convolution results displayed in (**d**–**f**). Autocorrelation measures the similarity between a signal and a delayed version of itself. In this context, it helps assess the effectiveness of each wavelet function in capturing and highlighting transitions in the time-series tail data. Comparing (**g**), (**h**), and (**i**), it is evident that the Gaussian wavelet is the most effective, demonstrating a quick decrease in autocorrelation to nearly zero with a small lag.

## Appendix C. Time-series tail curves with labelled co-transitions

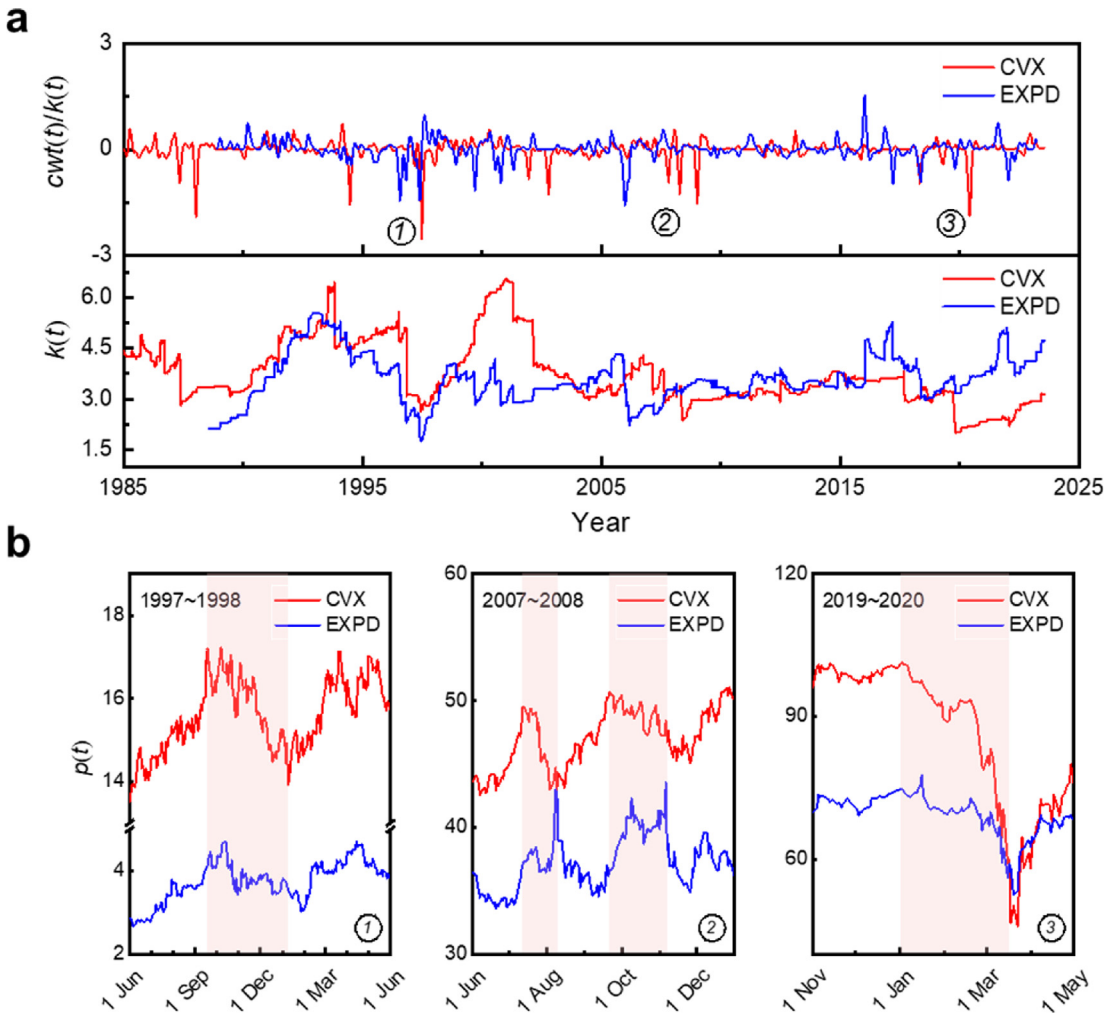


**Fig. C.** **a**, Time-series tail curves with labelled co-transitions. **b**, Price comovements between CVX and EXPD during identified timings (labelled as (1) Asian financial crisis, (2) Global financial crisis, and (3) Covid-19), as detected by co-tail declines (if $k(t)$ drops below $-0.2$). Notably, in all three instances, CVX exhibited a price decline preceding the response from EXPD, as depicted by the shaded regions.

## Appendix D. Interconnected financial tail risks among the stocks

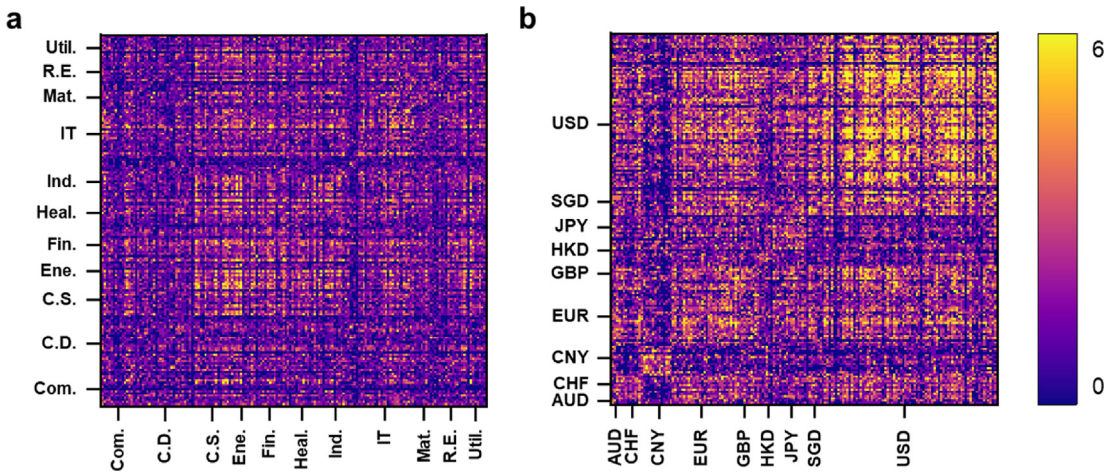


**Fig. D.** 2D heat maps illustrating interconnected global financial tail risks. These visualizations portray interconnected global financial tail risks through 2D heat maps. Stocks are categorized by sectors (**a**) and exchange currencies (**b**). The horizontal axis represents sector/currency names in abbreviated form, aligning with those on the vertical axis. The figures use a color-coded representation of detected comovements, showing occurrences where entities on the vertical axis precede those on the horizontal axis within a duration of 7–14 days.

## Appendix E. The tail-driven portfolio for price forecasting

**Table E1**
Portfolio selection. The portfolio comprises pairs of stocks denoted as A-B. These pairs are selected based on a significant number of comovements (>5 since IPO) during tail declines, identified when $k(t)$ drops below $-0.2$. Importantly, the tail transition in stock (A) consistently precedes that in stock (B), with a defined duration of 7–21 days. The paper discusses specific pairs, namely, "JPM-GE" and "CVX-EXPD".

| Pair of stocks | No. tail comovements | Pair of stocks | No. tail comovements | Pair of stocks | No. tail comovements |
|---|---|---|---|---|---|
| 000625.SZ-600688.SS | 6 | XOM-BNP.PA | 7 | JPM-GE | 8 |
| 6952.T-000625.SZ | 6 | BRK-A-WMT | 8 | U11.SI-GE | 7 |
| 900942.SS-000625.SZ | 7 | C-NVDA | 6 | GFC.PA-IBM | 6 |
| 000625.SZ-WMT | 6 | C-SPG | 6 | RIO.L-GFC.PA | 7 |
| 4911.T-600221.SS | 7 | C6L.SI-D05.SI | 6 | GFC.PA-STAN.L | 9 |
| 5401.T-7203.T | 6 | CVX-EXPD | 9 | GPK-WMT | 8 |
| 600606.SS-MU | 6 | CVX-O39.SI | 7 | INTC-PEP | 8 |
| DBK.DE-900942.SS | 6 | TSCO.L-CVX | 7 | UHR.SW-ISRG | 7 |
| 9202.T-DTE.DE | 6 | D05.SI-JPM | 6 | JPM-KER.PA | 6 |
| AMGN-9984.T | 7 | D05.SI-MSI | 6 | JPM-KO | 8 |
| GFC.PA-A | 6 | D05.SI-PEP | 7 | JPM-MS | 6 |
| BKNG-AAPL | 8 | D05.SI-PG | 6 | JPM-STAN.L | 7 |
| AAPL-C | 9 | DBK.DE-EOAN.DE | 6 | KO-WMT | 10 |
| AAPL-DTE.DE | 7 | DBK.DE-TSCO.L | 6 | LMT-SPG | 6 |
| AAPL-FMG.AX | 6 | OR.PA-DTE.DE | 6 | LMT-TTE.PA | 6 |
| AAPL-GFC.PA | 7 | DUK-UL | 8 | MS-SIE.DE | 6 |
| AMGN-FJTSY | 6 | EL-ENB | 6 | NOC-XOM | 8 |
| U11.SI-AMZN | 6 | ENB-GFC.PA | 7 | PFE-UL | 8 |
| FLEX-ASML | 6 | XOM-EXPD | 7 | RIO.L-SIE.DE | 7 |
| BIO-ROG.SW | 6 | FJTSY-FLEX | 6 | SGRO.L-WMT | 7 |
| BIO-U11.SI | 8 | KO-FJTSY | 6 | SONY-TLS.AX | 6 |
| FJTSY-BKNG | 6 | FJTSY-STAN.L | 6 | SPG-UL | 6 |
| LMT-BNP.PA | 6 | JPM-FLEX | 7 | | |
| NKE-BNP.PA | 6 | ORCL-FLEX | 7 | | |

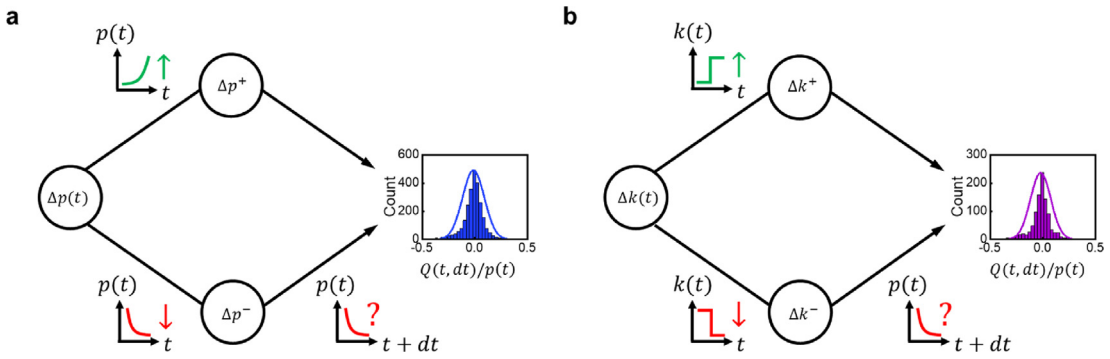


**Fig. E2.** Schematics describing (**a**) a price-driven decision process and (**b**) a tail-driven decision process. $\Delta p^+$ and $\Delta p^-$ represent the positive and negative price returns, respectively, after the lead-time period preceding the prediction for future price change (with lag time $dt$). Similarly, $\Delta k^+$ and $\Delta k^-$ represent the positive and negative tail transitions, respectively, after the lead-time period preceding the prediction for future price change (with lag time $dt$).

## Appendix F. Data analysis and scripts

The time-series datasets utilized in this study, comprising the adjusted closing prices of General Electric (GE) and JPMorgan Chase & Co. (JPM), are sourced from Yahoo Finance (https://finance.yahoo.com). The price forecasting employs five principal technical indicators – lagged price, standard deviation, moving average, exponential moving average, and Bollinger Bands – as fundamental learning features for forecasting GE stock prices. Additionally, the tail data, serving as a precursor feature, is derived from JPM stock prices. For a comprehensive understanding of the analytical processes, detailed examples and scripts are accessible on the project's GitHub repository: https://github.com/tyqu90/time-series-tail/tree/main.

## References

Akhtaruzzaman, M., Boubaker, S., Sensoy, A., 2021. Financial contagion during COVID–19 crisis. Finance Research Letters 38, 101604.
Allen, F., Gale, D., 2000. Financial contagion. Journal of Political Economy 108 (1), 1–33.
Aloui, R., Aïssa, M.S.B., Nguyen, D.K., 2011. Global financial crisis, extreme interdependences, and contagion effects: the role of economic structure? Journal of Banking & Finance 35 (1), 130–141.
Alstott, J., Bullmore, E., Plenz, D., 2014. powerlaw: a Python package for analysis of heavy-tailed distributions. PloS One 9 (1), e85777.
Bae, K.H., Karolyi, G.A., Stulz, R.M., 2003. A new approach to measuring financial contagion. The Review of Financial Studies 16 (3), 717–763.
Barberis, N., Shleifer, A., Wurgler, J., 2005. Comovement. Journal of Financial Economics 75 (2), 283–317.
Baur, D., Jung, R.C., 2006. Return and volatility linkages between the US and the German stock market. Journal of International Money and Finance 25 (4), 598–613.
Beirne, J., Fratzscher, M., 2013. The pricing of sovereign risk and contagion during the European sovereign debt crisis. Journal of International Money and Finance 34, 60–82.
Belke, A., Dubova, I., Osowski, T., 2018. Policy uncertainty and international financial markets: the case of Brexit. Applied Economics 50 (34–35), 3752–3770.
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.
Brownlee, J., 2016. XGBoost with python: Gradient Boosted Trees with XGBoost and Scikit-Learn. Machine Learning Mastery.
Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794.
Clauset, A., Shalizi, C.R., Newman, M.E., 2009. Power-law distributions in empirical data. SIAM Review 51 (4), 661–703.
Daubechies, I., 1992. Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics.
Dekkers, A.L., Einmahl, J.H., De Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. The Annals of Statistics 1833–1855.
DeLong, J.B., 1991. Did JP Morgan's men add value? An economist's perspective on financial capitalism. In: Inside the Business Enterprise: Historical Perspectives on the Use of Information. University of Chicago Press, pp. 205–250.
Diebold, F.X., Yılmaz, K., 2014. On the network topology of variance decompositions: measuring the connectedness of financial firms. Journal of Econometrics 182 (1), 119–134.
Embrechts, P., Klüppelberg, C., Mikosch, T., 2013. Modelling Extremal Events: for Insurance and Finance, 33. Springer Science & Business Media.
Forbes, K.J., Chinn, M.D., 2004. A decomposition of global linkages in financial markets over time. Review of Economics and Statistics 86 (3), 705–722.
Francq, C., Zakoian, J.M., 2022. Testing the existence of moments for GARCH processes. Journal of Econometrics 227 (1), 47–64.
Gabaix, X., Gopikrishnan, P., Plerou, V., Stanley, H.E., 2003. A theory of power-law distributions in financial market fluctuations. Nature 423 (6937), 267–270.
Ghaderpour, E., Pagiatakis, S.D., Hassan, Q.K., 2021. A survey on change detection and time series analysis with applications. Applied Sciences 11 (13), 6141.
Gillespie, C.S., 2014. Fitting heavy tailed distributions: the poweRlaw package. arXiv preprint arXiv:1407.3492.
Giudici, P., Sarlin, P., Spelta, A., 2020. The interconnected nature of financial systems: direct and common exposures. J. Bank. Finance 112, 105149.
Gofman, M., 2017. Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. Journal of Financial Economics 124 (1), 113–146.
Grané, A., Veiga, H., 2010. Wavelet-based detection of outliers in financial time series. Computational Statistics & Data Analysis 54 (11), 2580–2593.
Guo, Y., Li, P., Li, A., 2021. Tail risk contagion between international financial markets during COVID-19 pandemic. International Review of Financial Analysis 73, 101649.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics 1163–1174.
Jeon, B.N., Von Furstenberg, G.M., 1990. Growing international co-movement in stock price indexes. Quarterly Review of Economics and Business 30 (3), 15–31.
Kelly, B., Jiang, H., 2014. Tail risk and asset prices. The Review of Financial Studies 27 (10), 2841–2871.
Kenourgios, D., Dimitriou, D., 2015. Contagion of the Global Financial Crisis and the real economy: a regional analysis. Economic Modelling 44, 283–293.

Li, Y., Mykland, P.A., 2015. Rounding errors and volatility estimation. Journal of Financial Econometrics 13 (2), 478–504.
Lopes, H.F., Migon, H.S., 2002. Comovements and contagion in emergent markets: stock indexes volatilities. In: Case Studies in Bayesian Statistics: Volume VI. Springer, New York, pp. 285–300.
Lux, T., Marchesi, M., 1999. Scaling and criticality in a stochastic multi-agent model of a financial market. Nature 397 (6719), 498–500.
Mantegna, R.N., Stanley, H.E., 1995. Scaling behaviour in the dynamics of an economic index. Nature 376 (6535), 46–49.
Martinez-Jaramillo, S., Carmona, C.U., Kenett, D.Y., 2019. Interconnectedness and financial stability. Journal of Risk Management in Financial Institutions 12 (2), 168–183.
Matsui, M., Mikosch, T., Tafakori, L., 2013. Estimation of the tail index for lattice-valued sequences. Extremes 16, 429–455.
McNeil, A.J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance 7 (3–4), 271–300.
Minoiu, C., Kang, C., Subrahmanian, V.S., Berea, A., 2015. Does financial connectedness predict crises? Quantitative Finance 15 (4), 607–624.
Pericoli, M., Sbracia, M., 2003. A primer on financial contagion. Journal of Economic Surveys 17 (4), 571–608.
Pickands, I.I.I.,J., 1975. Statistical inference using extreme order statistics. The Annals of Statistics 119–131.
Qu, T., Mei, K.W., Doray, A., 2022. A simple method to detect extreme events from financial time series data. Machine Learning with Applications 10, 100415.
Raddant, M., Kenett, D.Y., 2021. Interconnectedness in the global financial market. Journal of International Money and Finance 110, 102280.
Reboredo, J.C., Rivera-Castro, M.A., Ugolini, A., 2017. Wavelet-based test of co-movement and causality between oil and renewable energy stock prices. Energy Economics 61, 241–252.
Sun, A.J., Chan-Lau, J.A., 2017. Financial networks and interconnectedness in an advanced emerging market economy. Quantitative Finance 17 (12), 1833–1858.
Wang, G.J., Yi, S., Xie, C., Stanley, H.E., 2021. Multilayer information spillover networks: measuring interconnectedness of financial institutions. Quantitative Finance 21 (7), 1163–1185.