# Drivers of firm-level tail dependence: A machine learning approach

Thomas Conlon[a,], John Cotter[a,], Ioannis Ropotos[a,b,*]

[a] *Department of Banking and Finance, University College Dublin, Michael Smurfit Graduate Business School, Carysfort Avenue, Blackrock, Co. Dublin, Ireland*
[b] *Dublin Business School, Dublin City University, Dublin D09, Ireland*

---

## Abstract

The paper studies the determinants of firm-level tail dependence of companies with respect to foreign markets using machine learning. We measure dependence for a comprehensive international set of firms using copulas and we find that left tail dependence is consistently stronger than right tail dependence with their gap widening in recessionary periods. We then apply random forest regressions to identify and characterize the factors that account for the total panel variation of tail risk. The World Uncertainty Index, the R2 integration measure and coskewness with respect to foreign markets are the most important determinants. For US firms individual ownership variables such as the number of total or foreign investors dominate the remaining firm-level characteristics in explaining tail dependence. Our results contribute to the understanding of crash risk in the modern global financial landscape with implications for asset managers.

*Keywords:* Firm-level tail dependence, Copulas, Determinants, Random forest regression, Machine learning, Shapley values
*JEL:* F30, G15

---

[⋆]The authors would like to acknowledge the support of Science Foundation Ireland under Grant Numbers 16/SPP/3347, 13/RC/2106_P2, 17/SPP/5447 and 18/CRT/6224.

[*]Corresponding author

*Email addresses:* conlon.thomas@ucd.ie (Thomas Conlon), john.cotter@ucd.ie (John Cotter), ioannis.ropotos@ucd.ie (Ioannis Ropotos), ioannis.ropotos@dcu.ie (Ioannis Ropotos)

## 1. Introduction

Markets suffered severe losses during the Global Financial Crisis of 2008 and again during the COVID-19 pandemic. Both episodes highlighted the importance of understanding the interdependence of international markets and the role of market participants in the transmission of shocks. This paper focuses on the exposure of individual firms to foreign market shocks. Firms around the world are vulnerable to extreme events originating abroad and by studying their response to such shocks, we can isolate cross-border interdependence and the potential for contagion. Measuring firm-level tail dependence on foreign markets is particularly relevant for asset managers who hold multi-country portfolios and are less concerned about isolated domestic meltdowns than systemic international risks that can threaten the whole portfolio. For these reasons, we wish to quantify the level of firm-level tail dependence across the world and identify the characteristics that are associated with elevated crash risk.[1]

We measure the tail dependence between a publicly traded stock and the corresponding foreign market index using the methodology of Chabi-Yo et al. (2018) for the period 2000-2023 for an international sample of 8,589 US companies and 25,039 non-US companies. Specifically, we fit the convex combination of the Clayton, Gaussian and Rotated Clayton copulas annually using daily returns from July of year t-1 to June of year t. This copula combination is very flexible in modelling both left (Clayton), right (Rotated Clayton) and no (Gaussian) tail dependence at the same time. Furthermore, it has the advantage of using the information of the whole joint distribution instead of the few observations found only in the tails. First, we calculate the copula-based left (LTD) and right (UTD) tail dependence coefficient and study their properties. The LTD (UTD) coefficient computed from the fitted copulas is the theoretical probability for a stock to experience the worst (best) return given that the market index also experiences its worst (best) return.

We then apply a random forest regression model on LTD and UTD and a representative dataset of firm- and country-specific characteristics in order to rank

[1]In line with the literature on tail dependence (De Jonghe, 2010, Laeven et al., 2016, Weiß et al., 2014, Ammann et al., 2023) we do not attempt to make any causal arguments in this paper.

the determinants of LTD. Our dataset includes value, profitability, investment, ownership and macroeconomic variables that have been shown to be linked with tail dependence in the literature. We expand on this literature in later sections. Our primary objective is to establish which determinants matter most in explaining firm-level tail dependence with the world. Thanks to the ability to digest multiple features at once, machine learning techniques help to identify the most important characteristics. For that reason, machine learning models have become very popular and they are now used widely in a plethora of finance applications (Georges and Pereira, 2021; Catullo et al., 2022; Arifovic et al., 2022; Cakici et al., 2023). Thus our choice to work with random forest regression is not arbitrary; we opt for it due to its ability to handle correlated variables and capture non-linear and interaction effects between our regressors. After we fit our model, we rank variables in terms of their importance using a variety of measures.

Our results can be summarized as follows. The time series equal- or value-weighted mean of the LTD and UTD coefficients has the same pattern across the US and non-US sample: left tail dependence is always higher than right tail dependence and their difference widens in recessions. However, this widening is the result of the increased levels of LTD rather than UTD with the latter being almost constant throughout the years. This finding is in line with Forbes and Rigobon (2002) who conclude that interdependence in the left tails is stronger during crises.

In the second part, we concentrate on the firm-level left tail dependence with results being qualitatively similar for right tail dependence. We find that the the R2 measure of Pukthuanthong and Roll (2009), the World Uncertainty Index of Ahir et al. (2022) and coskewness with respect to the corresponding foreign market index are the most important determinants of LTD across samples and variable importance measures. The World Uncertainty Index is a text-based measure of uncertainty and captures the crisis periods in which the dependence structure between a stock and markets change. It explains 6% and 4% of the panel variation of left tail dependence in the US and outside of the US, respectively. The R2 captures the dependence of a firm and foreign markets on the central part of their joint distribution while coskewness describes the behaviour of the stock return when the market return undergoes extreme deviations. The R2 explains 7% and 8% of the variation of LTD with coskewness explaining 6% and 7% in the US and non-US sample. After the World Uncertainty Index, R2 and coskewness, *Ownership* variables such as the number of institutional investors as well as foreign and total institutional ownership matter the most in explaining left tail dependence for US stocks only. Their variable importance ranges between 2% and 3.3% in both samples. For comparison, size has a score of 3.2% or 2% depending whether we examine US or non-US firms while the rest of the variables have a negligible contribution with scores of 2% or less. Recognizing the black-box characteristics of the random forest regression, we augment our analysis with Shapley values, providing additional confirmation of our key results.

We further study variables as groups. Aggregation of individual variables allows us to see which categories have the strongest relationship with firm tail dependence. The variables are grouped into eight broad categories that include macroeconomic (*Macro*), price and return related (*Market*), institutional ownership (*Ownership*), value (*Value*), investment (*Investment*) and profitability (*Profitability*) variables. Variables are categorized in groups based on their economic content; for example, market capitalization and coskewness are price and return based variables while the corporate default spread and the number of internet users capture the macroeconomic environment of the US. In general, the variables that belong to the same group are correlated but it can be the case that they do not proxy for the same informational content and thus aggregating them allows us to harvest all available information within a group.[2]

We find that left tail dependence in the US sample is driven primarily by *Market*, *Ownership* and *Macro* variables that are always the top 3 most important groups by a large margin. When these groups are excluded, the explanatory power of the model is reduced by 40%, 34% and 25%, respectively. In contrast, the results for the non-US sample show a more diffuse pattern of explanatory power. Beyond the consistently important *Market* group that includes the R2 and coskewness, the relative importance of the remaining variable groups is less clear. Institutional

[2]For example, Hou et al. (2020) classify anomalies into categories and then test their ability to be replicated using acommon framework. Their economic categorization of anomalies which we adopt for the *Value*, *Investment* and *Profitability* variables is consistent with statistical clustering and principle component analysis meaning that the information content of these anomalies is very similar but not exactly the same.

ownership is less influential in the non-US sample compared to its prominent role in the US. This might be due to the high prevalence and influence of institutional investors in the US; they hold more than 70% of the US stock market in contrast to 22% of the non-US market (see Figure 7). Our results hold under a battery of robustness checks.

The paper is organized as follows. Section 2 summarizes the related literature. Section 3 discusses the modelling of tail dependence, describes the random forest regression algorithm and introduces variable importance measures and Shapley values. In Section 4 we motivate and describe the variables that we use in our analysis. Section 5 documents our empirical findings. Finally we discuss several robustness checks in Section 6 and we conclude in Section 7.

## 2. Literature review

Systemic risk, and tail dependence in general, has received considerable attention in the literature in recent years especially after the disastrous events of the Global Financial Crisis (GFC) of 2008. The GFC highlighted the importance of understanding the interdependence of financial institutions and the potential for contagion across market sectors. For that reason, tail dependence and its determinants has been studied on the country- (Nguyen and Lambe, 2021; Beine et al., 2010), industry- (Chiu et al., 2015) and firm-level (De Jonghe, 2010; Weiß et al., 2014; Laeven et al., 2016). Studying tail dependence at the firm level is especially important when there are stocks that are more sensitive to the extreme shocks of local or foreign markets than others. It is therefore valuable for global investors to identify these stocks and their characteristics in order to reduce their likelihood of experiencing large losses. We contribute to the literature by measuring tail dependence at the firm level for all international companies and identify the characteristics that help explain their tail dependence with foreign markets.

Even though the literature documenting tail dependence and extreme events at the market level is not new (Longin and Solnik, 2001; Forbes and Rigobon, 2002; Poon et al., 2004), there is a more recent strand that studies its determinants. Beine et al. (2010) measure the impact of several bilateral characteristics of markets at both the left and right tail of the return distribution and find that their impact is asymmetrical; financial liberalization increases only left tail comovement while trade integration affects positively the whole distribution. Nguyen and Lambe (2021) characterize both the direction (does a tail event in country i cause a tail event in country j or the other way around?) and the determinants of tail risk in bilateral pairs of markets thus categorizing countries into tail risk drivers and receivers. They find that the size of the economy of the driver country is the strongest determinant of the country pair connectedness with a positive effect, followed by trade and capital linkages with the latter variables having a negative effect. There is also evidence of tail risk drivers at the industry level such as Chiu et al. (2015) who study the tail risk spillover from the financial to non-financial sectors in the US and find that sectors with high net debt financing and lower valuation and investment suffer the most in crisis periods. A more recent and similar study is that of Nguyen et al. (2021) who explore tail risk spillovers between US industries, highlighting the role of the customer-supply relations between the industries.

The literature on firm-level tail dependence and its determinants is concentrated on the banking and financial sector.[3] For example, De Jonghe (2010) and Laeven et al. (2016) find that size and tail beta are highly and positively correlated. De Jonghe (2010) notice the positive relation between ordinary betas and tail betas and conclude that comovement of a bank with a banking index at the center of their distribution extends to the tails. Weiß et al. (2014) find that profitability and value variables are relevant for explaining and transmitting tail risk across banks during a banking crisis. Anginer et al. (2018) study banks under a corporate governance framework and find that more shareholder friendly governance increases both firm-level and systemic risk in banks, particularly for large institutions and in countries with generous financial safety nets.

Outside of the financial sector, Cheng et al. (2023) show that the tail risk of firms is driven by the realized tail risk of their peer firms that are commonly owned by blockholder institutions. In a recent paper, Glossner et al. (2025) demonstrate

[3]The literature on measuring tail risk and tail dependence in the banking sector is vast (see for example Billio et al., 2012; Tobias and Brunnermeier, 2016; Girardi and Ergün, 2013; Karimalis and Nomikos, 2018; Yang et al., 2021; Torri et al., 2021).

that institutional investors exacerbated price declines during the Covid-19 crash, with higher institutional ownership associated with greater downside risk as investors faced redemptions, reduced portfolio risk, and engaged in broad-based selling. These findings highlight the role of institutions in the propagation of shocks in a network of firms and demonstrate the significance of ownership variables in influencing tail dynamics. At the macro level, it has been established that the dependence structure of markets changes during crisis periods (Longin and Solnik, 2001; Forbes and Rigobon, 2002) while trade integration increases comovement for all quantiles of the joint distribution of market returns (Beine et al., 2010). Even when firm-level tail dependence is estimated for non-financial firms, it is studied only in an asset pricing framework (Kelly and Jiang, 2014; Van Oordt and Zhou, 2016; Chabi-Yo et al., 2018) with no emphasis on its determinants.

The firm-level evidence has been primarily aimed at banks and financial institutions, and we add to the literature by focusing on additional issues. Our paper addresses two fundamental questions: i) how does left tail dependence vary across all publicly listed firms within or across countries and ii) which determinants can explain its panel variation? Identifying the characteristics of the local stocks that are more sensitive to extreme shocks of foreign markets is of fundamental importance for global investors who are averse to extreme losses. For example, Kelly and Jiang (2014) and Chabi-Yo et al. (2018) find that the US stocks with higher left tail dependence with respect to the US local market index have higher expected returns and vice versa. Weigert (2016) show that this crash sensitivity premium is not only a US phenomenon as it is present in 39 other countries besides the US. This view is consistent with the "safety first" framework of Roy (1952) and Barro (2006) in which investors require a premium to hold stocks that are more likely to crash when the market portfolio crashes and highlights the importance of left tail comovement in times of market turmoil.

We contribute to the existing literature which examines firm-level tail dependence in four ways. First, we extend the literature to a comprehensive global sample of publicly listed firms instead of limiting our sample to financial institutions only. It is true that the health of the banking sector is of the highest priority for regulators since banks have a fundamental role in the economy and their collapse have negative rippling effects (Global Financial Crisis of 2008) to real output. However, there is still a strong incentive to measure how extreme negative market returns affect all firms, regardless of whether they are financial or not, in order to identify the stocks most sensitive to crash risk and minimize the exposure to portfolio tail risk. For instance, asset managers who follow global investment strategies and hold non-financial firms in their portfolios can benefit from our analysis. They can closely monitor firms with elevated left tail dependence, integrate them into stress tests or set position limits in order to mitigate crash risk.

Second, we characterize the factors that determine left tail dependence on a US and non-US sample, separately and, as such, we shed light on whether the US is different from other markets. After the firm-level calculation of tail dependence, a natural question arises; what are the characteristics that drive left tail dependence? This question extends the literature on the determinants of tail dependence across countries, industries, and financial firms to a comprehensive firm-level analysis of all companies, offering immediate practical relevance. Identifying the characteristics linked to left-tail dependence enables investors and policymakers to pinpoint firms most likely to suffer severe losses when markets experience their worst returns. This insight allows asset managers to design more effective crash mitigation strategies and helps regulators monitor potential sources of firm-level fragility within the global market network.

Third, we employ a machine learning (ML) approach, namely random forest regression, to rank the determinants of left tail dependence of firms over time and across countries. Our machine learning approach is capable of handling correlated variables and non-linear effects. In conjunction with our representative dataset of characteristics, random forest regression provides new insights on the factors that matter the most in explaining the panel variation of tail dependence for US and non-US companies. Thus our work expands the literature on the application of advanced machine learning techniques in empirical Finance. In their seminal paper, Gu et al. (2020) applied machine learning algorithms in order to measure asset risk premiums popularizing their use in Finance. Since then researchers have used ML in a variety of cases. For example, Akbari et al. (2021) use random forest regression to rank determinants of country level financial and economic integration while Casabianca et al. (2022) use AdaBoost to rank determinants of banking crises. Georges and Pereira (2021) and Catullo et al. (2022) incorporate ML in their agent-based models with applications on market stability and forecasting sales, respectively. Other uses of ML include Arifovic et al. (2022) who employ ML in a high-frequency trading framework, Cakici et al. (2023) who study global stock return predictability, Han et al. (2024) who introduce ML augmented Fama-McBeth regressions and Chen et al. (2024) who apply neural networks in an innovative way to estimate asset pricing models for individual stock returns.[4]

Finally, we focus on the tail dependence that exists between a stock and foreign markets. Past studies measure the tail dependence of a firm with respect to a local or global index. We deviate from that framework in the sense that we explore the link between the tails of a firm and its corresponding foreign index in an effort to provide insights on how firms can be adversely affected by shocks outside of the firm's country.[5] This approach enables us to study how vulnerable firms are to international conditions regardless of the state of the local market. In general, local investors possess an informational advantage (Coval and Moskowitz, 2001) about their respective country markets, and as such, they should be aware of the tail dependence of the local stocks with the market. However, there is additional value in learning how extreme negative shocks originating outside of their home country propagate to the local equities. This is exactly the effect that we capture.

Our results suggest that market conditions as well as the integration of a firm with foreign markets and the activity of institutional investors are strongly associated with tail dependence. Even though the effect of the market conditions is not new (Forbes and Rigobon, 2002), we document how the dependence structure between local firms and foreign markets changes in crisis periods using copulas. The fact that high integration levels are positively correlated with high tail dependence levels implies that the dependence in the central part of the joint distribution extends naturally to the tails and specifically to the left tail. The rise of institutional

[4]See Kelly et al. (2023) for a more detailed survey on the use of ML techniques in financial applications and Giglio et al. (2022) for a survey on their use in asset pricing in particular.

[5]Previous studies calculate the tail dependence of a financial institution with respect to a local or global banking index. For example, De Jonghe (2010) and Laeven et al. (2016) use a local banking index while Weiß et al. (2014) calculate systemic risk using both a local and global index.

investors in global markets contributes to the increase of the firm-level tail dependence highlighting once more their role in the landscape of the modern financial world. Thus, through a fuller understanding of the determinants of tail dependence between local stocks and foreign markets, investors can make better ex-ante evaluations on their local equity portfolio's sensitivity to foreign shocks.

## 3. Methodology

### *3.1. Measuring firm-level tail dependence*

We measure the firm-level tail dependence of a local stock with a foreign market index that proxies non-local markets using the methodology of Chabi-Yo et al. (2018). It is a copula based method and it offers certain advantages in capturing tail dependence over other parametric and non-parametric methods. First, it allows for a flexible fit of combinations of basic parametric copulas to the bivariate distribution of the stock and the foreign market index in which the left (lower) and right (upper) tail dependence coefficients can be explicitly derived and estimated simultaneously. Second, the copula approach exploits the information from the whole joint distribution instead of a small number of return observations in the tail in comparison to non-parametric measures. This property allows us to model dependence using daily returns in a span of a year and update the copula parameters from one period to the next, thus capturing the dynamic nature of dependence.

Generally, basic bivariate copulas, such as those in the Gaussian or Archimedean family, do not allow for modelling both the left, right or no tail dependence at the same time. Thus Chabi-Yo et al. (2018) chose to work with convex combinations of copulas. In the same spirit, we use the combination of the Clayton-Gaussian-Rotated Clayton copula. The (Rotated) Clayton copula exhibits only (right) left tail dependence while the Gaussian copula exhibits no tail dependence at all. We focus on a single copula combination to make our results comparable across firms and years. The final form of the copula is

$$C(u, v; \Theta) = w_1 C_{Clayton}(u, v; \theta_1) + w_2 C_{Gaussian}(u, v; \theta_2) + w_3 C_{rClayton}(u, v; \theta_3) \quad (3.1)$$

where $\Theta$ is the set of the basic copula parameters $\theta_1, \theta_2, \theta_3$. The weights have to sum to 1, $w_1 + w_2 + w_3 = 1$ and satisfy $0 \leq w_1, w_2, w_3 \leq 1$. $C(u, v; \theta)$ denotes the cumulative density function (CDF) of a bivariate copula with parameters $\theta$ as a function of the uniformly distributed random variables $u, v$. The parameters $\theta_1$ and $\theta_3$ control the left and right tail dependence that Clayton copulas exhibit while $\theta_2$ is just the correlation coefficient for the fitted Gaussian copula. The weights, $w_1, w_2, w_3$, are representative of the dependence structure of the two random variables u, v. When $w_1$ ($w_3$) increases, u and v exhibit a dependence structure that is left (right) tail dominant while an increase of $w_2$ indicates a structure with weaker tail dependence.

Throughout we consider $X$ and $Y$ to be two random variables that correspond to the return of the local stock and the return of the respective foreign market index with joint distribution $F_{X,Y}(x, y)$ and marginals $F_X(x), F_Y(y)$. The conditional probabilities that capture the left and right tail dependence (hereafter LTD and UTD, respectively) in the case of the basic parametric copulas such as the Gaussian and Clayton, can be simplified (see McNeil et al., 2015 for a proof) to the following expressions only in terms of the bivariate copula C that models them:

$$LTD = \lim_{q \to 0^+} Pr\left(X < F_X^{-1}(q) \mid Y < F_Y^{-1}(q)\right) = \lim_{q \to 0^+} \frac{C(q, q)}{q} \tag{3.2}$$

$$UTD = \lim_{q \to 1^-} Pr\left(X > F_X^{-1}(q) \mid Y > F_Y^{-1}(q)\right) = \lim_{q \to 1^-} \frac{1 - 2q - C(q, q)}{1 - q} \tag{3.3}$$

The LTD and UTD can be calculated explicitly for the basic parametric copulas, and thus, once the parameters $\Theta$ of Eq. 3.1 are known, LTD and UTD for the convex combination of Clayton-Gaussian-Rotated Clayton are given as:

$$LTD = w_1 2^{-1/\theta_1} \text{ and } UTD = w_3 2^{-1/\theta_3} \tag{3.4}$$

The LTD and UTD of equation 3.4 will be our copula-based measure of left and right tail dependence, respectively. Note that the measured tail dependence is controlled essentially by two parameters; the weights of $w_1$ and $w_3$ and the Clayton copula parameters $\theta_1$ and $\theta_3$. This means that even though the weights assigned to the copulas may be the same, the copula parameters may not be and vice versa.

The estimation of Eq. 3.1 is a two-fold procedure. First the marginal distributions $X, Y$ of the stock and foreign market index are estimated by their empirical counterparts:

$$\hat{F}_X(x) = \frac{1}{n+1}\sum_{k=1}^{n} I_X(k \leq x) \text{ and } \hat{F}_Y(y) = \frac{1}{n+1}\sum_{k=1}^{n} I_Y(k \leq y) \tag{3.5}$$

where n is the number of valid daily return observations in the period of July of year t-1 to June of year t. We opt for the July-June scheme as per Fama and French (2015) in order to map accounting variables to returns. We require at least 150 non-missing observations of which no more than 80% of them are zero (stale returns) for the estimation of LTD and UTD. For US stocks, we proxy foreign markets with the Fama-French Developed Market index excluding the US. For non-US stocks, we use the US CRSP value-weighted index on the basis that the US is the largest tail risk driver in a global network of countries according to Nguyen and Lambe (2021). Specifically, they construct a network of directional tail risk connectedness for a large number of countries and they find that a tail event in the US leads to by far, the highest increase in the probability of causing a tail event in all other countries in the network.

The estimated empirical quantiles of $X, Y$ are, by definition, uniformly distributed in $[0, 1] \times [0, 1]$ and play the role of the random variables $U, V$ used in the definition of the copula combination in Eq. 3.1. The realizations of $U, V$ constitute the pseudo-observations. These pseudo-observations are used in minimizing the logarithm of the maximum likelihood function in order to find the parameters $\Theta = [w_1, w_2, w_3, \theta_1, \theta_2, \theta_3]$:

$$\hat{\Theta} = \underset{\Theta}{argmax} \sum_{i=1}^{n} \log c(u_i, v_i; \Theta) \tag{3.6}$$

where $c(u_i, v_i; \Theta)$ is the corresponding copula density function of $C(u_i, v_i; \Theta)$ in equation 3.1. The following constraints are used for the canonical maximum likelihood estimator (CMLE): $w_1 + w_2 + w_3 = 1$, $0 \leq w_1, w_2, w_3 \leq 1$, $0 \leq \theta_1, \theta_3 < \infty$, $-1 \leq \theta_2 \leq 1$.

### *3.2. Random forest regression*

We wish to distinguish the relative importance of a comprehensive list of variables for our measured tail-dependence coefficients without imposing strong theoretical priors. For that reason, we employ the random forest regression (RFR) of Breiman (2001) to determine which firm- or country-specific variables explain the panel variation of firm-level tail dependence and then rank these variables. RFR has been applied recently by Akbari et al. (2021) in the search for the drivers of economic and financial integration.

The RFR methodology offers several advantages over other machine learning algorithms or traditional econometric techniques. First, the tree based nature of RFR allows for nonlinear and complex relationships between explanatory and dependent variables that enhance its predictive power.[6] In essence, RFR implicitly models nonlinear interactions between variables without requiring these interactions to be specified in advance. This is a limitation in regularization techniques like LASSO or Ridge regression, where interaction terms must be manually constructed and included.

Second, RFR is remarkably robust to hyperparameter tuning, often delivering superior performance with minimal calibration compared to other machine learning techniques that are sensitive to parameter choices, such as neural networks or support vector machines (Athey and Imbens, 2019). Its structure also imposes sparsity by rarely using covariates that do not meaningfully improve model splits, thereby assigning zero weight to irrelevant variables without explicit penalization (Athey and Imbens, 2019). The inclusion of many irrelevant variables does not alter the RFR results and as such it is highly suitable for our study in which we may include variables that may not have any connection to left tail dependence.

Additionally, RFR avoids overfitting through its dual randomization process: it builds each tree using a random subsample of the data and selects a random subset of covariates at each split within each tree. These two steps ensure that different observations and different subsets of variables are used in building the

[6]For example Gu et al. (2020) in their seminal paper that popularized the use of machine learning techniques in Finance found that the nonlinearities in RFR in the form of interactions among explanatory variables, substantially improved predictions over traditional regression models.

tree, decreasing the likelihood of overfitting. They also serve a secondary purpose; the randomness induces sufficient variation in the trees that the average outcome $f_{rf}(x)$ of the RFR is relatively smooth across the values of x (Breiman, 2001). Together, these features make RFR an especially powerful and versatile machine learning technique for modeling complex and nonlinear relationships when the underlying model structure is unknown or too complex to specify through traditional econometric approaches.

Finally, RFR is inherently able to handle multicollinearity. Traditional econometric techniques, such as OLS, often suffer when predictor variables are highly correlated, resulting in inconsistent and inefficient coefficient estimates. Akbari et al. (2021) acknowledge this issue and adopt the RFR to uncover the drivers of financial and economic integration at the country level. RFR uses ensemble learning by averaging predictions across many decision trees, each trained on bootstrapped samples and random subsets of features. This process diffuses the influence of collinear variables across trees, naturally mitigating the adverse effects of multicollinearity.

The RFR algorithm involves three steps: i) fitting piece-wise linear relationships between dependent and explanatory variables using a decision tree, ii) bootstrapping the number of observations and the number of explanatory variables making the decision trees random and iii) averaging the fitted values from the ensemble of random trees. These steps are explained in detail below.

First, we draw a bootstrap sample of size $max\_samples$ from the training data X. The training data X is a matrix of $M \times K$ potential determinants of firm-level tail dependence. In our study, M = 65,103 firm-year observations for the US sample and M = 213,810 firm-year observations for the non-US sample while K=39 is the number of explanatory variables that drive tail dependence. We choose $max\_samples = 2/3$ and $max\_features = 39$ meaning that we randomly select only 2/3 of our original dataset with all 39 explanatory variables to start growing each Tree b. This sampling process ensures that each tree is trained on a unique dataset, which helps to reduce overfitting and improve generalizability.[7]

[7]Every time we build a Tree, we use randomly only 2/3 of the full sample. This randomness is controlled by the state of the random seed. We set the random state to 1 for our baseline

Then we grow a random-forest tree $T(X, \Theta^b)$ from the bootstrapped data, by recursively splitting the sample into homogeneous subsamples until the termination conditions apply. The process is terminated when either the number of observations in the final node is less than 10 or the number of observations in either of the two subsamples generated after a split is less than 5. The splitting of the sample at each node of the tree occurs at the optimal splitting point s such as $\min_s [MSE(y|x_k < s) + MSE(y|x_k \geq s)]$. MSE(.) denotes the mean squared error of a linear regression of y on X on each of two subsamples $R_1 = \{x_k < s\}$ and $R_2 = \{x_k \geq s\}$. At each node of the decision tree, the variable $x_k$ and the corresponding splitting point s that yield the lowest MSE are chosen. The search for s is over the K=39 candidate variables $x_k, k = 1, .., 39$ and over the values of each variable $x_k$. The least MSE approach ensures that any other partitioning $s^*$ based on other explanatory variables or based on different threshold levels will result in less homogeneous subsamples.

Once the maximum depth of the Tree b has been reached, the fitted value $\hat{y}$ is the average value of Y in the final node, $\hat{y} = f_b(X) = T(X, \Theta^b)$ where $\Theta^b$ is the set of parameters of the Tree. The piece-wise linear relationships applied by the splitting procedure allow the tree to accommodate a more general form of relationships, $f_b(X)$, between dependent and explanatory variables. We grow B=1000 trees in total by repeating the steps above for $b = 1, \ldots, B$. The process results in the ensemble $\{T(X, \Theta^b)\}_{b=1}^B$ and a prediction at a new point x in a regression setting is just $\hat{y} = f_{rf}(x) = \frac{1}{B}\sum_{b=1}^{B} T(x, \Theta^b)$. More details of the RFR implementation are presented in Appendix A.

### *3.2.1. Variable importance*

After we fit the random forest regression to the data, we rank variables using two different measures of importance. The first is the permutation test (PT) of Breiman (2001) in which we score variable j by the difference in prediction accuracy before and after permuting j. For each variable j, we generate a matrix $X_{perm,j}$ by permuting all data points of j. We then estimate the prediction error

results. When we set the random state to 3, 5, 7 or 11, results remain largely unchanged and they are available upon request.

$e_{perm,j}$ of the permuted model and repeat the process K=10 times generating K corrupted datasets $X_{perm,j,k}$ with their associated model errors $e_{perm,j,k}$. The PT score is calculated as the difference $VI_j = \frac{1}{K}\sum_{k=1}^{K}(e_{perm,j,k} - e_{orig})$. In essence the permutation process breaks the relation between variable j and the true outcome y, and as such, larger values of the PT score imply greater importance for variable j. The sum of the permutation test scores of all variables is normalized to equal 1.

The second approach is the reduction in predictive R2 from setting all values of variable j to zero, while holding the remaining model estimates fixed. The premise is that in the absence of important variables, the fit of the model will be significantly worse. Further details of these measures can be found in Appendix B.

### *3.3. Shapley values*

Despite the advantages of random forest regression as a machine learning method, it still remains a black box. To enhance the interpretability of the RFR methodology, we use Shapley values. Shapley (1953) used concepts from game theory to prove that there is a unique value that measures the marginal contribution of a player in a game and satisfies the axioms of efficiency, symmetry, null player, and additivity. More specifically, if there are p players who can form coalitions S with $|S|$ players each and $\nu(S)$ is a function that assigns the reward that coalition S can achieve in the game, then the Shapley value of player i can be estimated using the following equation:

$$\text{Shapley value}(i) = \sum_{S\subseteq\{1,\dots,p\}\setminus\{i\}} \frac{|S|!(p-|S|-1)!}{p!}\left(\nu(S\cup\{i\}) - \nu(S)\right) \tag{3.7}$$

where $(\nu(S\cup\{i\}) - \nu(S))$ is the marginal contribution of player i to coalition S as the difference of the reward $\nu(S\cup\{i\})$ when player i is in coalition S and the reward $\nu(S)$ when player i is absent from coalition S. The marginal contribution is then summed over all possible coalitions S that player i can participate multiplied by a weight that assigns the probability that player i is in coalation S. $p!$ is the number of ways to form a coalition of p players, $|S|!$ is the number of ways coalition S can form and $(p - |S| - 1)!$ is the number of ways $|S|$ players can join after player i has joined coalition $S \cup \{i\}$. In the context of machine learning, a Shapley value represents a "fair" allocation of the reward to a feature value in a predictive model, as it allocates across features the difference between a specific instance prediction and the average prediction (Lundberg and Lee, 2017, Molnar, 2020, Agarwal et al., 2023). Thus the marginal contribution of feature $X_j$, i.e. its Shapley value $\phi$, can be expressed as,

$$\phi_j(\hat{f}(x)) = \sum_{S \subseteq C(X) \setminus X_j} \frac{|S|!(p - |S| - 1)!}{p!} \left(\hat{f}_x(S \cup \{X_j\}) - \hat{f}_x(S)\right) \tag{3.8}$$

In this notation, $C(X) \setminus X_j$ is the set of all model configurations excluding variable $X_j$ and $\hat{f}$ is the trained model. Thus the set S contains a combination of up to p possible features except for feature $X_j$. x is the instance of the dataset to be explained and $\hat{f}_x(S)$ is the prediction of the model for the features included in S for instance x. In other words, $\phi_j(\hat{f}(x))$ is the marginal contribution of explanatory variable $X_j$ in the instance x of the dataset to the payout which is the difference in predicted value $\hat{f}_x(S \cup \{X_j\})$ of instance x when both variable $X_j$ and the variables in S are included and the average predicted value $\hat{f}_x(S)$ when only the variables S are included, weighted and summed over all possible feature value combinations S. We construct global Shapley values or GSVs for variable $X_j$ by averaging the absolute Shapley values of all instances x as

$$\Phi_j = \sum_{i=1}^{N} \frac{|\phi_j(\hat{f}(x_i))|}{N}, \tag{3.9}$$

where N is the number of instances in the dataset. The US sample has N = 65,103 instances while the non-US sample has N=213,810 instances. To compare feature importance measured by other methods, their relative importance is evaluated as

$$\Phi'_j = 100 \times \frac{\Phi_j}{\sum_{j=1}^{p} \Phi_j}, \tag{3.10}$$

where p=39 is the number of features. We refer to $\Phi'_j$ as a normalized Global Shapley Value (nGSV). nGSV is quoted as a percentage (%). Further details on the calculation of Shapley values can be found in Appendix C.

## 4. Data

We construct a representative list of firm characteristics sourced from the intersection of Compustat, CRSP (Datastream) and FactSet for the US (non-US) sample. We group variables into broad categories by adopting, altering and extending the group definition of Hou et al. (2020). The groups are *Market*, *Investment*, *Profitability*, *Value*, *Ownership* and *Macro*. Last, we gather daily return, price and volume data for non-US stocks from Thomson Reuters Datastream. All items are converted to US dollars and the list of all 39 variables along with their categorization into groups can be found in Table 1.

[Insert Table 1 here.]

The vast literature on firm-level tail dependence of banks has established that size plays an important role in explaining the variation of systemic risk. For example, De Jonghe (2010) find that size is the largest driver of tail beta with a positive effect. Similar conclusions are drawn from Laeven et al. (2016) who focus on the events of GFC. What is more, De Jonghe (2010) notice that ordinary betas and tail betas are highly correlated (in the 50% to 75% range) and conclude that "banks with large exposure to movements in the banking index in normal economic conditions will be more exposed to extreme movements as well". For that reason, we include the R2 of Pukthuanthong and Roll (2009) as a measure of the dependence of a local stock with its foreign market index on the full support of the joint distribution. We would also like to differentiate firm-specific tail risk from tail dependence and to do so, we include Value-at-Risk (var90) and expected shortfall (es90) at the 90% level. Then we group size, R2, var90 and es90 along with coskewness, illiquidity, volatility and momentum variables in the *Market* category.

Weiß et al. (2014) examine the effect of both firm- and country-specific variables on the systemic risk of banks during crises and find that the only firm variables that are relevant are the profitability and book-to-market ratio. Similarly, when Chiu et al. (2015) study the tail risk spillover from the financial sector to all other sectors in the US economy, they conclude that low investment and value industries are more likely to experience a price decline following a banking sector crisis. Thus we include a variety of *Profitability*, *Investment* and *Value* variables in our analysis since they might be also relevant for non-financial firms.

Next, we study the effect of institutional investors on the tail dependence of local stocks with foreign markets. Recently, Cheng et al. (2023) showed that the tail risk of firms is driven by the realized tail risk of their peer firms that are commonly owned by blockholder institutions (a blockholder entity owns at least 5% of the firm). They provide evidence that this common institutional blockholder (CIB) effect is one of the main channels through which tail risk propagates in the network of firms: tail risk increases after initiations of peer connections via CIB. This finding highlights the role of institutions on the relationship of firms at the extreme tails of their joint distribution. In a similar vein, Faias and Ferreira (2017) also establish that institutions act as agents of financial globalization by investing worldwide and, as such, firms with higher institutional ownership exhibit higher levels of comovement with global factors rather than local or industry factors. Even though they examine the role of investors on the central part of the distribution, we expect that such a link applies to the tails, too. Glossner et al. (2025) show that institutional investors amplified price declines during the COVID-19 crash, as higher institutional ownership increased downside risk due to redemptions, portfolio de-risking, and broad-based selling. In order to explore the multifaceted role of institutional investors, we include *Ownership* variables in our models.

Finally, we include global and country-specific macroeconomic variables. It is a stylized fact of the international finance literature that the dependence structure of markets changes during crisis periods (Longin and Solnik, 2001; Forbes and Rigobon, 2002). For that reason, we use the World Uncertainty Index (WUI) and the World Trade Uncertainty Index (WTUI) of Ahir et al. (2022) which are text based measures of global economic and trade uncertainty in order to capture this effect. We also use the Trade and Market capitalization of all public stocks over GDP for each country as measures of de jure economic openness and financial development, respectively. These variables are shown to explain market segmentation at the country level in Bekaert et al. (2011) while Beine et al. (2010) find that trade integration increases comovement across all quantiles of the joint distribution of market returns.

We keep only public traded firms with common shares and we require that these firms have no missing data for any of the variables used in our analysis. Our final US sample includes 65,103 firm-year observations in total for June of years 2000-2023. The final non-US sample includes 213,810 firm-years for June of years 2000-2023 spanning across 73 countries. Details on the construction of the dataset can be found in Appendix D.

## 5. Empirical findings

### *5.1. Firm-level tail dependence*

Figures 1a and 1b plot the cross-sectional equal- and value-weighted mean of the LTD and UTD for the US and non-US sample, respectively. The pattern is clear in all cases; i) left tail dependence is always stronger than right tail dependence and ii) the difference between the two becomes large in recessionary periods. The last finding is consistent with the conclusion of Forbes and Rigobon (2002) that there is increased interdependence during market crashes. The LTD measure captures exactly that interdependence on the left tail of the joint distribution of a local stock and its corresponding foreign market index. More specifically, the LTD values correspond to the conditional probability for a stock to experience its worst return given that foreign markets experience their worst returns in a given year.

[Insert Figure 1 here.]

For example in the equal-weighted (value-weighted) case, LTD rises over the value of 10% (15%) for the US sample in NBER classified recession periods with the most notable example the Global Financial Crisis period when left tail dependence reaches its peak at 16% (24%). This means that, in recessions, there is a probability of 10% or higher that the average US stock will crash when foreign markets crash. The same phenomenon is observed in the non-US sample in which the same crash probability rises above 8% or 10% at the end of recessions depending on whether we equal- or value-weight it and peaks at 9% or 15% during the GFC.

Generally, value-weighted estimates of LTD and UTD are higher than their corresponding equal-weighted values implying that larger firms are far more exposed to foreign shocks than smaller firms. It is also important to note that the widening of the difference between LTD and UTD during crises arises from the stark increase of LTD while right tail dependence is almost stable throughout the years. Detailed summary statistics of both left and right tail dependence values can be found in Table 2. A notable feature of the results is the 150-200% increase of the LTD dispersion as measured by its cross-sectional standard deviation in recessionary periods compared to normal times. This increased dispersion suggests larger heterogeneity for the exposure of firms to left tail events.

[Insert Table 2 here.]

The calculation of the tail dependence parameter requires the fit of the Clayton-Gaussian-Rotated Clayton copula combination and, as such, weights are assigned to each copula, for a given stock and year. The weights are representative of the dependence structure between the firm and the corresponding foreign market index. When the weight of the Clayton (Rotated Clayton) copula increases, the stock and the index exhibit a dependence structure that is left (right) tail dominant while an increase of the Gaussian weight indicates a structure of weaker tail dependence. The weights convey additional information over the single copula parameters: they indicate whether left, right or no tail dependence exist while the copula parameters indicate the strength of that tail dependence. Together, they dictate the type and strength of tail dependence. For example, the correlation of the copula weight w, and the corresponding parameter $\theta$ is -15% (-15%), 45% (44%) and -14% (-10%) for the Clayton, Gaussian and Rotated Clayton in the US (non-US) sample, respectively. The relatively low correlations mean that the weights and the single copula parameters are not the same and that their information content is different.

Figures 2a and 2b plot the equal-weighted average weights (%) that are assigned to the Clayton, Gaussian and Rotated Clayton copulas. We show that the average weight of the Rotated Clayton that captures the right tail dependence is the same across the years with a mean value of 25% for US and 20% for non-US stocks. However, the average weight assigned to the Gaussian copula that captures no dependence dominates that of the Clayton copula in non-crisis periods. In other words, the Clayton copula explains better the joint realizations of the stock returns and the corresponding index than the Gaussian copula in crises. However, the explanatory power of the Rotated Clayton remains the same on average, regardless of the state of the economy. This is direct evidence on how the dependence structure changes in crisis periods: from weak to strong left tail dependence.

[Insert Figure 2 here.]

Next we assess the persistence of the left tail dependence measure since only the sensitivity of stocks to market crashes is relevant to crash averse investors. Specifically, we are interested in whether the LTD of the previous period is related to the LTD of the current period. In other words, if a firm exhibits high left tail dependence with foreign markets in one period, should we expect it to behave the same way in the next? Persistence is measured as the relative frequency at which a stock is sorted into a LTD quintile portfolio in year t given that it was in portfolio i in year t-1. The rank 1 portfolio contains the 20% of stocks with the lowest LTD while rank 5 contains those with the highest LTD. If LTD is random, then LTD of year t-1 should not convey any information for the future LTD and thus it is equally likely for a stock to belong in one of the five LTD quantile portfolios in year t regardless of its previous ranking. This random pattern will translate to a LTD persistence of 20% for all quintile portfolios. Figures 3a and 3b plot the persistence of the copula-based LTD coefficient for the US and non-US sample, respectively. The persistence of the 5th quintile portfolio is evident since its value is always above 20% for both US and non-US firms. Its persistence, however, is extremely high for international stocks and it is almost always above 30% with a peak of 45% around the Global Financial Crisis. This high persistence is evidence of the importance of the US stock market for the rest of the world and a confirmation of the findings of Nguyen and Lambe (2021). In other words, we find that firms with the highest LTD exhibit the highest persistence. This implies that there exists a set of firms with certain non-transient characteristics that contribute to their systematically high left tail dependence. Similarly, the persistence of the 1st quintile portfolio of the lowest LTD stocks is high (above 20%) but it is not consistently as high as the persistence of the 5th quintile portfolio. This suggests that a set of firms with low LTD also exists, but its characteristics are not as well defined as those of the high LTD group.

[Insert Figure 3 here.]

Finally, we report the correlation of left tail dependence with all other variables in our dataset in Table 3. The rank order of the individual variables is indicative of the random forest regression results as we see below: i) the R2 integration measure and coskewness have the highest absolute correlation with left tail dependence and ii) *Ownership* variables dominate all others. Interestingly, firm size (log_me) has the same positive relationship with left tail dependence in both samples with a correlation value of 32%. The World Uncertainty Index (WUI) has by far the strongest linear relation with LTD among all other *Macro* variables.

[Insert Table 3 here.]

*5.2. Determinants of firm-level left tail dependence*

In this section, we now present the findings of our empirical analysis in terms of the random forest regression model and the measures of variable importance that we use to distinguish the dominant variables that explain firm-level left tail dependence of US and non-US stocks.[8] All individual variables are ranked in terms of the permutation test score and the change of R2 while results are presented for variable groups. In the permutation test, we randomly permute variable j thus breaking its relationship with LTD. We then apply the already fitted random forest regression model to the permuted dataset and take the difference in prediction accuracy before and after permuting j. The sum of the permutation test scores of all variables is standardized to equal 1. The change in R2 is the reduction in predictive R2 in the absence of a variable from the model and it is calculated by setting all values of that variable to zero, while holding the remaining model estimates fixed. The larger the values of the permutation test score and the change in R2 are, the more important a variable is in explaining LTD.

As powerful as random forest regression might be as a machine learning technique, it does not generate interpretable coefficients similar to those in the conventional regression framework. For that purpose, we augment the RFR analysis with linear regression specifications in which we examine whether the effect of a variable on tail dependence is negative or positive.

[8]Results for right tail dependence are qualitatively similar to those for left tail dependence and as such we do not report them here. They are available upon request.

*5.2.1. US results*

First, we present results for the US sample. Figure 4 shows the importance of individual variables and their groups on firm-level left tail dependence of US stocks based on the permutation test score and the change in R2. We find that the R2 measure of Pukthuanthong and Roll (2009) is the most important determinant of LTD and it explains 7% of this panel variation. The R2 captures the dependence of a US stock with the foreign market index in the central part of their joint distribution which means that this central dependence also extends to the tails. This relates to the findings of De Jonghe (2010) regarding the strong relation between ordinary and tail betas; firms that have a large exposure to foreign shocks in tranquil economic conditions will be more exposed to negative extreme movements during turbulent market conditions. After R2, the World Uncertainty Index (WUI) and coskewness matter the most in explaining the variation of LTD with both of them contributing slightly less than 6%. WUI measures the global market uncertainty and as such it captures crisis and non-crisis periods. Thus its high importance is not a surprising finding given the stylized fact that, in crisis periods, stocks and markets tend to crash together more often than in non-crisis periods with WUI signalling the transition between them. Coskewness, on the other hand, measures the comovement of the stock with the foreign market return squared and as such it describes how the stock return behaves when the market return undergoes extreme deviations. Thus a positive (negative) value of coskewness implies that, when the market return deviates from its mean, stock returns are positive (negative). If coskewness is positive, then LTD is weak.

[Insert Figure 4 here.]

The most interesting pattern that we observe in our analysis is the importance of *Ownership* variables such as the number of total and foreign institutional investors (io_num and fio_num). They rank at the 4th and 6th place according to their permutation test score values of 3.3%, respectively. The dominance of the *Ownership* variables, however, is not limited to the number of total and foreign investors; total and foreign institutional ownership (io and fio) as well as foreign common ownership (fco_mean) are also highly ranked explaining 2.6%, 2.6% and 3.3% of LTD variation, respectively. Only the total stock market capitalization and the total trade over the US GDP (Mcap_GDP and Trade_GDP) are almost on the same level of importance as the *Ownership* variables. Both Mcap_GDP and Trade_GDP are important drivers of the left tail dependence of US stocks with foreign markets with Trade_GDP being a de jure factor of economic openness and free flow of capital among countries. Finally, firm size (log_me) does not turn out to be the most influential variable for LTD since it is lagging behind the R2 measure, World Uncertainty Index, coskewness and number of total and foreign investors. Its permutation test score is only 3.2%.[9] Other determinant variables exert a much lesser influence on left tail dependence.[10]

When we repeat our analysis with groups, we find that *Market*, *Ownership* and *Macro* variables matter the most. When these groups are excluded, the explanatory power of the random forest regression model is reduced by 40%, 34% and 25%, respectively. *Market* variables that include the R2, coskewness and size (log_me) explain the greatest proportion of variation in LTD in RFR followed by *Ownership* variables. The dominance of institutional ownership related characteristics highlights the power of institutions as agents of globalization who affect prices as a result of their trading activity. Surprisingly, *Macro* variables are behind the previous two categories meaning that, even though market conditions matter, firm-specific characteristics play a more important role for the level of tail dependence in US.

Finally, we employ Shapley values to increase the interpretability of the random forest regression. We report the normalized Global Shapley Value of the 15 most important individual variables in Figure 5a. Shapley values confirm our previous

[9]The fact that firm size is not a prime driver of tail dependence aligns with the study of De Jonghe (2010), who found that its influence diminished once more relevant variables were included in the analysis. Notably, when the ordinary least squares (OLS) beta that captures dependence in the central part of the joint distribution, is added to their regression, the marginal importance of size decreased by 37%. In our setting, the R2 measure plays a similar role, with variables such as institutional ownership and coskewness further reducing size's explanatory power.

[10]We acknowledge that firms may use derivatives, insurance instruments, or cash buffers to hedge tail risk. In our analysis, we include two firm-level proxies for cash buffers: cash to total assets ("c") and cash flows to total liabilities ("c2d"). Our results indicate that neither of these variables accounts for the panel variation of firm-level tail dependence, both within and outside the US.

findings; the R2 measure, coskewness and the World Uncertainty Index contribute the most to the Random Forest Regression prediction. Size (log_me) is found to be the 5th most important variable while the ownership breadth as proxied by the number of institutional investors (io_num) is ranked 6th. Although Global Shapley Values highlight the strength of the relationship between left tail dependence and our variables of interest, they remain silent on the direction of the relationship. For that purpose we use the Beeswarm plot in Figure 5b. The plot confirms that the R2 measure has a positive effect on LTD while coskewness is negatively associated with LTD. When global economic uncertainty proxied by WUI increases so does firm-level tail dependence. Both size and ownership breadth are positively associated with LTD as low and high values correspond to negative and positive SHAP values, respectively.

[Insert Figure 5 here.]

*5.2.2. Non-US results*

Second, we present results for the non-US sample. Figure 6 shows the importance of individual variables and their groups on LTD of non-US stocks. We observe the following patterns: i) coskewness, the R2 integration measure and World Uncertainty Index are the top 3 most important variables and ii) *Ownership* variables do not emerge as the primary firm-level determinants of tail dependence within the global sample. This finding contrasts sharply with the US sample where institutional investors play a major role in shaping crash risk dynamics. Interestingly, left tail dependence does not depend at all on firm size (log_cap) in the global dataset in contrast to the US dataset where size (log_me) ranks very high. The remaining determinant variables have little impact on LTD.

[Insert Figure 6 here.]

When we focus on groups, we find that *Market* variables with the inclusion of the R2 and coskewness are the most important drivers of left tail dependence in the international sample. It is also clear that the *Investment* variables matter the least no matter which test is used. Apart from the *Market* and *Investment* groups, the other variable groups contribute equally to tail dependence, highlighting a balanced distribution of explanatory power. For instance, the permutation test scores for the *Profitability*, *Value*, and *Ownership* groups are closely clustered at 16%, 15.8%, and 15%, respectively. Similarly, when assessing variable importance through changes in $R^2$, the *Macro*, *Profitability*, and *Ownership* groups exhibit nearly identical effects on left-tail dependence. This pattern diverges from the US sample, where *Ownership* variables are more influential, suggesting that institutional ownership exerts a more pronounced effect on crash risk within the US context than in global markets.

Part of the explanation on why institutional investors matter more within the US rather than outside might be their prevalence in the recent years. Institutions hold more than 70% of the US total market capitalization of stocks since 2004 while they hold only 22% of the non-US market (see Figure 7). Thus the representative investor within the US has become the representative institution. Their stock preferences and their trading activity have an immediate impact on US asset prices as well as their propensity to crash when foreign markets crash.

[Insert Figure 7 here.]

Shapley values complement our non-US analysis. We report the normalized Global Shapley Value of the 15 most important individual variables in Figure 8a. Shapley values confirm our previous findings; the R2 measure, coskewness and the World Uncertainty Index contribute the most to the Random Forest Regression prediction. Contrary to the permutation test and the change in R2 measures of variable importance, Shapley values depict a different picture for the rest of the variables. First, the state of the economy of the country the firm belongs to, plays an important role. The sum of exports and imports (Trade_GDP) as well as the stock market capitalization over GDP (Mcap_GDP) are ranked 4th and 5th, respectively. Second, we observe that market based variables such as the best daily return in a year (max), expected shortfall (es90), momentum (mom6m) and skewness dominate other individual variables. Interestingly, it is foreign ownership (fio) and foreign common ownership (fco_mean) that belong to the top 15 most important features that explain tail dependence.

[Insert Figure 8 here.]

Next, we employ the Beeswarm plot in Figure 8b to illustrate the direction and magnitude of the relationships between our explanatory variables and left tail dependence (LTD). The plot reveals that the R2 measure has a positive association with LTD, whereas coskewness is negatively related. Additionally, increases in global economic uncertainty, proxied by the World Uncertainty Index (WUI), are associated with higher firm-level tail dependence outside the US. The effects of Trade to GDP on LTD appear ambiguous, while Stock market capitalization to GDP (Mcap_GDP) shows a clear positive impact. Moreover, the best daily return in a year, expected shortfall, and momentum are all positively linked to LTD. Lastly, both foreign ownership and foreign common ownership exhibit positive influences on tail dependence.

### *5.3. Institutional investors and left tail dependence*

In this section, we study the effect of ownership of different types of institutional investors on left tail dependence inside and outside of the US. The literature on institutional ownership consistently highlights the significant heterogeneity among investors (Koijen et al., 2024, Glossner et al., 2025) and as such we wish to distinguish between investor groups. For that reason we follow Koijen et al. (2024) and we group institutional investors into investment advisors, hedge funds, long-term investors, private banking, and brokers. As investment advisors are a large group, we further split them into four subgroups: large-passive, small-passive, small-active, and large-active. At the end of June of each year, we first split the investment advisors into two groups by wealth (i.e. assets under management), so that the total wealth is equal across the groups. Within each wealth group, we further split the investment advisors into two subgroups at the median of the active share. The active share which is a slightly modified version of the active share of Cremers and Petajisto (2009) is defined as the total share of the investor's portfolio that deviates from the market weights. The active share is one-half times the sum of the absolute differences between the portfolio weights and the market weights over the set of stocks that are in the investor's investment universe.

Following the methodology of Koijen et al. (2024) institutional ownership is decomposed into 8 components and we study the effect of each component in a panel regression of left tail dependence against the different components. In other words, we regress LTD against the ownership of different institutional investors in a panel setting with time and country fixed effects in an effort to isolate the firm-level effect of ownership on left tail dependence. Table 4 summarizes the regression results. Holdings of large-passive investment advisors are consistently positively associated with firm-level left tail dependence across all samples of stocks. Small-active investment advisors have the same positive effect but their influence is weaker as suggested by the lower magnitude of the coefficients compared to their large-passive counterparts. Increased ownership by hedge funds, brokers and long-term investors is also associated with higher values of LTD both inside and outside of the US universe. However not all types of investors matter the same. Specifically, ownership by brokers has the strongest relationship with LTD.

[Insert Table 4 here.]

In a recent paper, Glossner et al. (2025) took advantage of the Covid-19 stock market crash to shed light on the risk-averse behaviour of institutional investors within the US during a crisis. They find that stocks with higher institutional ownership performed worse. Two key mechanisms contributed to this outcome: institutional investors faced a surge in redemptions while also seeking to reduce portfolio risk. Most types of investors reallocated their wealth towards financially robust firms, whereas hedge funds engaged in broad-based selling. Their findings suggest that institutional investors tend to exacerbate price declines when a tail risk materializes. Under this framework, the propensity of the firm to crash when the market crashes as proxied by LTD should be positively associated with institutional ownership variables. This is exactly what we find; i) higher values of ownership increase LTD and ii) almost all types of institutional investors contribute to the increase of LTD.

## 6. Robustness checks

In this section we discuss a series of additional tests, designed to demonstrate the robustness of our findings.

### *6.1. The effect of micro-cap stocks*

In this section, we explore the effect of micro-cap stocks on our analysis by excluding them. Even though micro-caps comprise the majority of the equity universe, their economic significance is trivial. We follow the definition of the most recent papers on asset pricing such as those of Hou et al. (2020) and Jensen et al. (2023) and we consider a stock to be a micro-cap when it belongs to the bottom 20% quantile of all stocks.

For the US sample, we sort stocks into micro-, and non micro-cap portfolios using the 20% NYSE breakpoint. We repeat the sorting procedure for the non-US sample using all international stocks. In both cases, the size quantiles for each June of year t are defined using all stocks in our original sample with available market cap data at June of year t. In other words, their size rankings are not determined by the subsetted sample that we use in our analysis.

Panel A of Table 5 reports the random forest regression results for the US and non-US stock universe when we exclude micro cap stocks from our analysis. The findings of Section 5.2 remain unchanged and thus our results are not driven by the inclusion of micro-caps.

[Insert Table 5 here]

### *6.2. The effect of financial firms*

Our baseline analysis includes all publicly traded firms, regardless of whether they are financial or not. Thus it is natural to examine how our results change when we exclude financial firms (SIC=6000-6999). Panel B of Table 5 reports the variable importance when financial firms are excluded from the random forest regression analysis with results remaining largely unchanged for both the US and non-US sample.

### *6.3. Choice of copulas*

To assess the sensitivity of our tail dependence estimates to copula choice, we re-estimate our model using two alternative copula combinations: (i) Clayton-Gaussian-Galambos and (ii) Rotated Clayton-Gaussian-Rotated Galambos. These combinations, alongside our baseline Clayton-Gaussian-Rotated Clayton specification, are among the most frequently selected in the empirical copula study of

Chabi-Yo et al. (2018), which evaluates 64 distinct copula configurations. Given their empirical prominence, our robustness checks focus exclusively on these three copula structures.

Panels A and B of Table 6 report the variable importance when the Clayton-Gaussian-Galambos and the Rotated Clayton-Gaussian-Rotated Galambos have been used, respectively. We report results for both the US and non-US sample separately. R2, coskewness, the World Uncertainty Index, Trade over GDP and institutional ownership variables are always prominent drivers of left tail dependence across samples and copula combinations. The *Market* group that includes the R2 measure and coskewness consistently dominates all other groups. While the *Ownership* group does not always rank as the second most important across all settings, it frequently exhibits levels of importance comparable to other prominent groups. In particular, aside from the *Macro* and *Investment* groups which are consistently found at the top and bottom of the rankings respectively, the remaining groups show relatively similar importance scores. This implies that across most of sample-copula combinations, *Ownership* variables are nearly as influential as those in the *Macro*, *Profitability*, and *Value* groups highlighting their influence on tail dependence.

[Insert Table 6 here]

### *6.4. Tail dependence in crisis vs non-crisis periods*

Our sample allows us to study whether the relationship between tail dependence and firm- and macro-level characteristics to firm integration, changes in crisis periods. Table 7 reports the crisis periods that might have affected the integration dynamics. We use four major crises; the Dot-Com bubble, the Great Recession of 2007, the Euro area sovereign debt crisis and Covid-19. The years that lie outside of the crisis periods of Table 7 define the non-crisis periods. For our 2000-2023 sample, the crisis periods amount to 8 years. It is important to note that our firm-year observations do not follow calendar years; instead, each year in our sample spans from July of year t–1 to June of year t, following the convention of Fama and French (1993). The firm-years affected by crisis periods reflect this convention. For example, the first crisis firm-year 2001 corresponds to the period of July 2000 to June 2001.

[Insert Table 7 here]

We split our sample into crisis and non-crisis periods and we apply the RFR algorithm in each subsample. Panels A and B of Table 8 report the most important drivers of tail dependence for the two subsamples. *Market* variables such as the R2 measure of integration of Pukthuanthong and Roll (2009) and coskewness dominate all other variables in both crisis and non-crisis periods. When the markets are distressed macroeconomic variables are the second most important drivers of tail dependence within and outside US. For example, the World Uncertainty Index (WUI) ranks very high in terms of variable importance in a consistent manner. *Ownership* variables are ranked in the third place behind the *Market* and *Macro* groups. In normal market times, firm-level ownership variables have the second largest effect both individually and collectively, at least within the US. Outside of the US, *Profitability* variables emerge as important determinants of left tail dependence. Overall, we find that the tail dependence dynamics do not change significantly between crisis and non-crisis periods.

[Insert Table 8 here]

### *6.5. Developed vs Emerging markets*

Risk dynamics can be different between Developed and Emerging markets due to political risk and institutions, financial development or information variables (Bekaert et al., 2011, Akbari et al., 2020). Thus we wish to test if our results are consistent between Developed and Emerging Markets. For that purpose, we use the MSCI Market classification as of April 2025 (`https://www.msci.com/our-solutions/indexes/market-classification`) to characterize the set of Developed countries. We consider the countries that do not belong to the set of Developed Markets (DMs) as Emerging Markets (EMs). Thus we split our sample of non-US stocks into two subsamples: i) Developed Markets and ii) Emerging Markets and repeat our RFR analysis in each subsample.

Table 9 reports the RFR results for Developed and Emerging Markets. *Market* variables such as coskewness and R2 dominate all other groups as in the baseline results. The World Uncertainty Index (WUI) and the World Trade Uncertainty Index (WTUI) are primary drivers of firm-level tail dependence in both DMs and

EMs. *Ownership*, *Macro*, *Profitability* and *Value* variables are the most important groups after *Market* variables. However, their relative ranking is difficult to determine since they all have similar permutation scores and change in R2 values. Overall, there are no large differences between DMs and EMs.

[Insert Table 9 here]

### *6.6. Tail dependence in the pre- and post-2010 era*

The Global Financial Crisis (GFC) of 2007–2009 stands as one of the most consequential events in modern financial history, triggering a systemic collapse that reshaped global markets, regulatory frameworks, and investor behavior. Given its profound impact, we divide the sample into a pre-2010 period that includes the events of the GFC and a post-2010 era to reflect the fundamentally altered landscape of global finance that emerged in its wake.

Table 10 reports the RFR results for the pre- and post-2010 era for both US and non-US stocks. *Market* variables such as the R2 measure of integration and coskewness dominate all other variables across all samples. The results for the pre- and post-2010 subperiod are very similar within the US and non-US sample. In particular, aside from *Investment* group which is consistently found at the bottom of the rankings, the remaining groups show relatively similar importance scores. This means that across time and stock samples, *Ownership* variables are nearly as influential as those in the *Macro*, *Profitability*, and *Value* groups highlighting their influence on firm-level tail dependence.

[Insert Table 10 here]

### *6.7. Bootstrapped confidence intervals for RFR model*

To assess the sensitivity of our variable importance metrics, we applied a non-parametric bootstrap procedure. Specifically, we draw $N = 100$ bootstrap samples with replacement from the original US and non-US datasets. For each sample, we compute the variable importance of individual variables and groups using the permutation test and the change in R2 metric. This process yields an empirical sampling distribution for each predictor's and group's importance, from which we report the bootstrap mean and 95% confidence interval, calculated as the 2.5th and 97.5th percentiles. Figure 9 and 10 plots the bootstrapped estimates for the US and non-US sample, respectively. Results remain largely unchanged.

[Insert Figures 9 and 10 here]

## 7. Conclusion

Even though firm-level tail dependence has been studied extensively in the banking literature, it has not been fully explored for firms outside of banks and financial institutions. In this paper, we provide insights on what determines left tail dependence between a local firm and its corresponding foreign market index regardless of the state of the local market. To that end, we first estimate a measure of firm-level tail dependence using the copula based methodology of Chabi-Yo et al. (2018) in a representative international sample of stocks for the period 2000-2023. We then combine that measure with a list of firm characteristics and macroeconomic variables to uncover the factors that characterize firm-level left tail dependence. We employ the random forest regression model to distinguish between variables that matter in explaining the panel variation of left tail dependence and those that do not in both a US and non-US sample.

We rank the variables in terms of importance using the permutation test and the change in R2. More specifically, we find that the World Uncertainty Index and World Trade Uncertainty Index of Ahir et al. (2022) along with the R2 measure of Pukthuanthong and Roll (2009) and the coskewness of local stocks with respect to their corresponding foreign market index are the most important determinants of left tail dependence for US and non-US stocks. Interestingly, *Ownership* variables such as the number of total and foreign institutional owners as well as the total and foreign institutional ownership dominate all other variables only within the US. The importance of *Ownership* observed within the US market which constitutes 50% of the share of the global stock market capitalization highlights the power of institutions as agents of globalization to the extent that they trade internationally and contribute to the increased exposure of local firms to foreign shocks in the left tail of their joint distribution. When we categorize variables into groups, we find that *Market* groups are the largest drivers of left tail dependence across US and non-US samples.

Our results suggest that market conditions as well as the integration of a firm with foreign markets and the activity of institutional investors are the most important drivers of left-tail dependence. Thus, we provide insights on the determinants of tail dependence between local stocks and foreign markets that investors can exploit to make better evaluations on their stock portfolio's sensitivity to foreign shocks.

Even though the effect of the market conditions is not new (Forbes and Rigobon, 2002), we document how the dependence structure between local firms and foreign markets changes in crisis periods using copulas. The fact that high integration levels are positively correlated with high tail dependence levels, implies that the dependence in the central part of the joint distribution extends naturally to the tails and specifically to the left tail. The rise of the institutional investors in global markets contributes to the increase of the firm-level tail dependence highlighting once more their role in the landscape of the modern financial world. Thus, through a fuller understanding of the determinants of tail dependence between local stocks and foreign markets, investors can make better ex-ante evaluations on their local equity portfolio's sensitivity to foreign shocks. Of course, further research will be necessary to establish the causation channels of the evidence presented in this paper.

Our findings offer a clear framework for identifying firms that are most exposed to crash risk. These firms tend to exhibit high coskewness, meaning their returns are more likely to decline sharply when the market falls. They also show greater global integration, making them more susceptible to cross-border shocks, and have high levels of institutional ownership or are embedded in common-ownership networks, which can accelerate the transmission of distress across markets, especially within the US. Importantly, the vulnerability of these firms intensifies during periods of heightened global uncertainty, as proxied by high values of the World Uncertainty Index. By focusing on these characteristics, namely coskewness, integration, institutional ownership, and the level of global uncertainty, investors and policymakers can identify a set of firms that are more likely to experience severe losses given that markets also experience extreme negative returns. Asset managers seeking to reduce stock price crash risk should monitor these firms closely, incorporate them into targeted stress tests, consider position limits, and explore dynamic hedging or diversification strategies that explicitly account for their high left-tail dependence. This can help asset managers design more robust crash mitigation strategies and enable regulators to monitor these potential sources of firm-level fragility.

### References

Agarwal, S., Muckley, C. B., and Neelakantan, P. (2023). Countering racial discrimination in algorithmic lending: A case for model-agnostic interpretation methods. *Economics Letters*, 226:111117.

Ahir, H., Bloom, N., and Furceri, D. (2022). The world uncertainty index. Technical report, National Bureau of Economic Research.

Akbari, A., Ng, L., and Solnik, B. (2020). Emerging markets are catching up: Economic or financial integration? *Journal of Financial and Quantitative Analysis*, 55(7):2270–2303.

Akbari, A., Ng, L., and Solnik, B. (2021). Drivers of economic and financial integration: A machine learning approach. *Journal of Empirical Finance*, 61:82–102.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Ammann, M., Moerke, M., Prokopczuk, M., and Würsig, C. M. (2023). Commodity tail risks. *Journal of Futures Markets*, 43(2):168–197.

Anginer, D., Demirguc-Kunt, A., Huizinga, H., and Ma, K. (2018). Corporate governance of banks and financial stability. *Journal of Financial Economics*, 130(2):327–346.

Anton, M. and Polk, C. (2014). Connected stocks. *The Journal of Finance*, 69(3):1099–1127.

Arifovic, J., He, X.-z., and Wei, L. (2022). Machine learning and speed in high-frequency trading. *Journal of Economic Dynamics and Control*, 139:104438.

Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1):685–725.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics*, 121(3):823–866.

Beine, M., Cosma, A., and Vermeulen, R. (2010). The dark side of global integration: Increasing tail dependence. *Journal of Banking & Finance*, 34(1):184–192.

Bekaert, G., Harvey, C. R., Lundblad, C. T., and Siegel, S. (2011). What segments equity markets? *The Review of Financial Studies*, 24(12):3841–3890.

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Cakici, N., Fieberg, C., Metko, D., and Zaremba, A. (2023). Machine learning goes global: Cross-sectional return predictability in international stock markets. *Journal of Economic Dynamics and Control*, 155:104725.

Casabianca, E. J., Catalano, M., Forni, L., Giarda, E., and Passeri, S. (2022). A machine learning approach to rank the determinants of banking crises over time and across countries. *Journal of International Money and Finance*, 129:102739.

Catullo, E., Gallegati, M., and Russo, A. (2022). Forecasting in a complex environment: Machine learning sales expectations in a stock flow consistent agent-based simulation model. *Journal of Economic Dynamics and Control*, 139:104405.

Chabi-Yo, F., Ruenzi, S., and Weigert, F. (2018). Crash sensitivity and the cross section of expected stock returns. *Journal of Financial and Quantitative Analysis*, 53(3):1059–1100.

Chen, L., Pelger, M., and Zhu, J. (2024). Deep learning in asset pricing. *Management Science*, 70(2):714–750.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794.

Cheng, C. A., Xie, J., and Zhong, Y. (2023). Common institutional blockholders and tail risk. *Journal of Banking & Finance*, 148:106723.

Chiu, W.-C., Peña, J. I., and Wang, C.-W. (2015). Industry characteristics and financial risk contagion. *Journal of Banking & Finance*, 50:411–427.

Coval, J. D. and Moskowitz, T. J. (2001). The geography of investment: Informed trading and asset prices. *Journal of Political Economy*, 109(4):811–841.

Cremers, K. M. and Petajisto, A. (2009). How active is your fund manager? A new measure that predicts performance. *The Review of Financial Studies*, 22(9):3329–3365.

De Jonghe, O. (2010). Back to the basics in banking? A micro-analysis of banking system stability. *Journal of Financial Intermediation*, 19(3):387–417.

Faias, J. A. and Ferreira, M. A. (2017). Does institutional ownership matter for international stock return comovement? *Journal of International Money and Finance*, 78:64–83.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Ferreira, M. A. and Matos, P. (2008). The colors of investors' money: The role of institutional investors around the world. *Journal of Financial Economics*, 88(3):499–533.

Forbes, K. J. and Rigobon, R. (2002). No contagion, only interdependence: measuring stock market comovements. *The Journal of Finance*, 57(5):2223–2261.

Georges, C. and Pereira, J. (2021). Market stability with machine learning agents. *Journal of Economic Dynamics and Control*, 122:104032.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1):3–42.

Giglio, S., Kelly, B., and Xiu, D. (2022). Factor models, machine learning, and asset pricing. *Annual Review of Financial Economics*, 14(1):337–368.

Girardi, G. and Ergün, A. T. (2013). Systemic risk measurement: Multivariate garch estimation of covar. *Journal of Banking & Finance*, 37(8):3169–3180.

Glossner, S., Matos, P., Ramelli, S., and Wagner, A. F. (2025). Do institutional investors stabilize equity markets in crisis periods? Evidence from covid-19. *Management Science*.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Han, Y., He, A., Rapach, D. E., and Zhou, G. (2024). Cross-sectional expected returns: new Fama–MacBeth regressions in the era of machine learning. *Review of Finance*, 28(6):1807–1831.

Hou, K., Xue, C., and Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, 33(5):2019–2133.

Jensen, T. I., Kelly, B., and Pedersen, L. H. (2023). Is there a replication crisis in finance? *The Journal of Finance*, 78(5):2465–2518.

Karimalis, E. N. and Nomikos, N. K. (2018). Measuring systemic risk in the European banking sector: A copula covar approach. *The European Journal of Finance*, 24(11):944–975.

Kelly, B. and Jiang, H. (2014). Tail risk and asset prices. *The Review of Financial Studies*, 27(10):2841–2871.

Kelly, B., Xiu, D., et al. (2023). Financial machine learning. *Foundations and Trends® in Finance*, 13(3-4):205–363.

Koijen, R. S., Richmond, R. J., and Yogo, M. (2024). Which investors matter for equity valuations and expected returns? *Review of Economic Studies*, 91(4):2387–2424.

Laeven, L., Ratnovski, L., and Tong, H. (2016). Bank size, capital, and systemic risk: Some international evidence. *Journal of Banking & Finance*, 69:S25–S34.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, 2(1):2522–5839.

Lundberg, S. M., Erion, G. G., and Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton University Press.

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu. com.

Nguyen, L. H. and Lambe, B. J. (2021). International tail risk connectedness: Network and determinants. *Journal of International Financial Markets, Institutions and Money*, 72:101332.

Nguyen, L. H., Nguyen, L. X., and Tan, L. (2021). Tail risk connectedness between us industries. *International Journal of Finance & Economics*, 26(3):3624–3650.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Poon, S.-H., Rockinger, M., and Tawn, J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *The Review of Financial Studies*, 17(2):581–610.

Pukthuanthong, K. and Roll, R. (2009). Global market integration: An alternative measure and its application. *Journal of Financial Economics*, 94(2):214–232.

Roy, A. D. (1952). Safety first and the holding of assets. *Econometrica: Journal of the Econometric Society*, pages 431–449.

Shapley, L. S. (1953). A value for n-person games.

Tobias, A. and Brunnermeier, M. K. (2016). Covar. *The American Economic Review*, 106(7):1705.

Torri, G., Giacometti, R., and Tichỳ, T. (2021). Network tail risk estimation in the European banking system. *Journal of Economic Dynamics and Control*, 127:104125.

Van Oordt, M. R. and Zhou, C. (2016). Systematic tail risk. *Journal of Financial and Quantitative Analysis*, 51(2):685–705.

Weigert, F. (2016). Crash aversion and the cross-section of expected stock returns worldwide. *The Review of Asset Pricing Studies*, 6(1):135–178.

Weiß, G. N., Bostandzic, D., and Neumann, S. (2014). What factors drive systemic risk during international financial crises? *Journal of Banking & Finance*, 41:78–96.

Yang, H., Cai, J., Huang, L., and Marcus, A. J. (2021). Bank stocks, risk factors, and tail behavior. *Journal of Empirical Finance*, 63:203–229.

# Appendices

## A. Random forest regression algorithm

We use the RandomForestRegressor class of the scikit-learn Python package to run random forest regressions. As in Akbari et al. (2021), we also follow Geurts et al. (2006) in setting the hyper-parameters of RFR. The RFR algorithm is described below:

1. Draw a bootstrap sample of size $max\_samples$ from the training data X. We choose $max\_samples = 2/3$ meaning that we randomly select only 2/3 of our original dataset to start building each Tree b.

2. Grow a random-forest tree $T(X, \Theta^b)$ to the bootstrapped data, by recursively repeating the following steps for each node of the tree, until the maximum depth ($max\_depth$) is reached. The maximum depth is reached when the samples of the final node is less than $min\_sample\_split = 10$ or either of the sub-samples left the split is less than $min\_samples\_leaf = 5$.

   (a) Select $max\_features$ variables at random from the K variables. We follow the convention of Geurts et al. (2006) and set $max\_features = K$ which in our case is 39.

   (b) Pick the best variable/split-point among the K candidate variables. For the kth explanatory variable, we find the optimal splitting point s such that

   $$\min_s \left[MSE(y|x_k < s) + MSE(y|x_k \geq s)\right] \tag{A.1}$$

   where MSE(.) denotes the mean squared error of a linear regression of y on X ($criterion$ = "squared error"). At each node of the decision tree, the variable $x_k$ and the corresponding splitting point s that yield the lowest MSE are chosen.

   (c) Split the node into two daughter nodes.

(d) Once the maximum depth of the Tree has been reached, the fitted value $\hat{y}$ is the average value of Y in the final node, $\hat{y} = f_b(X) = T(X, \Theta^b)$

3. Steps 1 and 2 creates the Tree $T(X, \Theta^b)$ where $\Theta^b$ contains the information of all the Tree parameters used. Repeating those steps for $b = 1, \ldots, B$ results in the ensemble $\{T(X, \Theta^b)\}_{b=1}^{B}$. A prediction at a new point x in a regression setting is just

$$\hat{y} = f_{rf}(x) = \frac{1}{B}\sum_{b=1}^{B} T(x, \Theta^b) \tag{A.2}$$

The random state of the RFR algorithm has been set to 1 (random_state=1). The RFR parameters are summarized in Table 11.

[Insert Table 11 here.]

## B. Definition of variable importance measures

### *B.1. Permutation test*

Once our model is trained, we can estimate the importance score for each of the explanatory variables using the permutation test of Breiman (2001). The premise of the test is that the fitted values show the largest sensitivity to changes in the most important variables. Thus our score is the difference in prediction accuracy before and after permuting the explanatory variables. This approach is known as "Mean Decrease Accuracy" method.

If $\hat{f}$ is our trained model, X our variable matrix, y the target vector and $L = L(y, \hat{f})$ is our prediction accuracy measure, then we can estimate the error of the original model as $e_{orig} = L(y, \hat{f}(X))$. Our choice for L is the mean squared error, $L(y, \hat{f}(X)) = E\left[y - \hat{f}(X)\right]^2$. For each variable j, we generate matrix $X_{perm,j}$ by permuting all data points of j. This permutation breaks the relation between variable j and the true outcome y. We then estimate the prediction error $e_{perm,j} = L(y, \hat{f}(X_{perm,j}))$ of the permuted model and repeat the process K times generating K corrupted datasets $X_{perm,j,k}$. Finally, we calculate the variable importance as the difference $VI_j = \frac{1}{K}\sum_{k=1}^{K}(e_{perm,j,k} - e_{orig})$ for K=10. The scores are standardized so that they sum up to one and all variables are ranked based on that score. The higher the value of $VI_j$, the more important that variable must be in explaining y since the prediction error increases. The permutation test is our primary variable importance measure in RFR modelling.

### *B.2. Change in R2*

We measure the importance of variable j by setting its value to zero and compute the difference between the R2 of the original data matrix and the R2 of the one with zeros in column j keeping everything else fixed. The larger the change in R2 is, the more important variable j must be since the fit of the model worsens. When we apply this method for variable group g, we set to zero all variables j that belong to g, $j \in g$, to zero and compute the difference in R2 again. This method popularized by Gu et al. (2020) is also used as a complementary measure to the permutation test.

## C. Details on the calculation of Shapley values

All possible coalitions (sets) S of feature values have to be evaluated with and without the jth feature to calculate the exact Shapley value, $\phi_j(\hat{f}(x)) = \sum_{S \subseteq C(X) \setminus X_j} \frac{|S|!(p - |S| - 1)!}{p!} \left( \hat{f}_x(S \cup \{X_j\}) - \hat{f}_x(S) \right)$. The exact solution to this problem becomes infeasible as the number of possible coalitions increases exponentially with the number of features. For that purpose, we employ the TreeSHAP method of Lundberg et al. (2018) which is a variant of the SHAP (SHapley Additive exPlanations) of Lundberg and Lee (2017). TreeSHAP is fine tuned for tree-based machine learning models such as decision trees, random forests, and gradient-boosted trees.

We employ the SHAP Python package of Lundberg et al., 2020. For the implementation of TreeSHAP, we train random forests through the XGBRegressor class of the XGBoost library (Chen and Guestrin, 2016). The XGBoost library offers higher computational gains over the scikit-learn library. Table 12 summarizes the parameters of the XGBRegressor class. The parameters are chosen so that the implementation of the random forest regression is consistent throughout the paper.

[Insert Table 12 here.]

## D. Data construction

Our primary dataset is Compustat North America for US companies and Compustat Global for international companies. Each company in Compustat is identified by "GVKEY". The US dataset is merged with CRSP utilizing the Compustat-CRSP link table. Thus each 'GVKEY' is matched with a valid "PERMCO" and possibly multiple "PERMNO"s. We merge the Compustat datasets with FactSet. Specifically, for each "GVKEY", we assign the equivalent FactSet company identifier "FACTSET_ENTITY_ID". We keep only companies that have been identified in both Compustat and FactSet, i.e. each "GVKEY" is assigned successfully to a "FACTSET_ENTITY_ID". For the North American dataset that contains both US and Canadian companies, the merge with FactSet is implemented through the CRSP-FactSet table; each CRSP security identified by "PERMNO" is matched with a FactSet security identified by "FSYM_ID" and the corresponding firm entity denoted by "FACTSET_ENTITY_ID". For the Global dataset, we use the most recent "ISIN" that Compustat provides in order to merge with FactSet. The merged Compustat-FactSet-CRSP dataset constitutes our initial sample of US stocks; 14,843 companies or 133,407 firm-year observations for June of year 2000 through June of year 2023. The corresponding international dataset consists of 46,849 non-US firms or 674,352 firm-year observations.

We obtain data for market capitalization, trading volume, and daily returns from CRSP for US and Canadian companies and from Datastream/Eikon for global companies. We convert the Return Index (RI) of an international stock to US dollars and we keep up to 9 decimal points after the conversion. We extract income statement and balance sheet variables from Compustat. We extract the World Uncertainty Index (WUI) and the World Trade Uncertainty Index (WTUI) from the website of Ahir et al. (2022). We obtain Trade over GDP (%) and the stock market capitalization over GDP (%) for all countries except Taiwan from the World Bank Development Indicators website. In the case that GDP for a country-year observation is available but its total stock market capitalization (MCAP) is not, we impute the missing value of MCAP over GDP using our own calculations. Data for some country-years can still be missing and as such we omit the corresponding firm-years from our analysis.

Finally, we keep only public traded firms with common shares and we require that these firms have no missing data for any of the variables used in our analysis. Effectively, we limit our CRSP dataset with stocks that have a "SHRCD" of 10 or 11 that trade in NYSE, Nasdaq and AMEX (now known as NYSE American) with "EXCHCD" of 1, 2 or 3. The final US dataset contains 8,589 companies/"GVKEY"s or 65,103 firm-year obsevations from June of 2000 to June of 2023. The non-US dataset is comprised of 25,039 companies/"GVKEY"s or 213,810 firm-years observations for the 2000-2023 period spanning across 73 countries. Table 13 summarizes the representation of countries in our final international sample of stocks.

[Insert Table 13 here.]

## E. FactSet ownership

We source institutional holdings data from FactSet, which compiles this information from regulatory filings, stock exchange disclosures, company annual reports, and interviews with fund managers. We use the method of Ferreira and Matos (2008) to aggregate holdings by 13f reporting entities and fund-level holdings at the FactSet institution level and carry forward past reports. We then follow the methodology of Koijen et al. (2024) and we categorise institutional investors based on FactSet's entity sub-types into five groups: investment advisers, long-term investors, hedge funds, private banking institutions, and brokers. Table 14 provides the mapping between FactSet entity sub-types and our institution types.

[Insert Table 14 here.]

**Table 1.** Potential determinants of firm-level tail dependence

| | | | | | |
|---|---|---|---|---|---|
| | **Market** | | | **Profitability** | |
| 1 | illiquidity | Amihud (2002)'s illiquidity measure | 20 | prof | Gross profitability over book equity |
| 2 | R2 | Pukthuathong & Roll (2009)'s integration measure | 21 | roc | Size + long-term debt - total assets to cash |
| 3 | Coskewness | Coskewness of local stock with foreign market index | 22 | roe | Income before extraordinary items to lagged book equity |
| 4 | log_me/log_cap | Log of market capitalization | 23 | roic | Return on invested capital |
| 5 | mom6m | Return from 6 to 2 months before end of June | 24 | s2c | Sales to cash |
| 6 | total_vol | Total daily return volatility | | | |
| 7 | max | Maximum daily return in a year | | **Value** | |
| 8 | skewness | Skewness of daily stock returns | 25 | a2me | Total assets to market cap |
| 9 | var90 | Value-at-risk at the 90% level | 26 | btm | Book to market ratio |
| 10 | es90 | Expected shortfall at the 90% level | 27 | c | Cash to total assets |
| | | | 28 | c2d | Cash flow to total liabilities |
| | **Ownership** | | 29 | dSo | Log change in split-adjusted shares outstanding |
| 11 | io | Total institutional ownership | 30 | debt2p | Total debt to market cap |
| 12 | fco_mean | Average of Anton and Polk (2014)'s FCAP_(i,j) over foreign stocks j | 31 | e2p | Income before extraordinary items to market cap |
| 13 | fio | Foreign institutional ownership | 32 | sales_g | Sales growth |
| 14 | fio_num | Number of foreign institutional owners | 33 | sat | Sales to total assets |
| 15 | io_hhi | Herfindahl–Hirschman index of io | 34 | cto | Sales to lagged total assets |
| 16 | io_num | Number of total institutional owners | 35 | ipm | Pre-tax income over sales |
| | | | | | |
| | **Investment** | | | **Macroeconomy** | |
| 17 | inv | Percentage change in total assets | 36 | WUI | World uncertainty index of Ahir et al. (2022) |
| 18 | dceq | Percentage change in book equity | 37 | WTUI | World trade uncertainty index of Ahir et al. (2022) |
| 19 | ivc | Change in inventory over average total assets | 38 | Mcap_GDP | Total market capitalization of local market over GDP |
| | | | 39 | Trade_GDP | The sum of exports and imports over GDP |

*Notes:* The table lists the variables we consider as potential determinants of firm-level tail dependence in our analysis by category. Variables are categorized in groups based on their economic content. In general, the variables that belong to the same group are correlated but it can be the case that they do not proxy for the same informational content and thus aggregating them allows us to harvest all available information within a group. Variables are defined annually for each June. We use balance-sheet data from the fiscal year ending in year t-1 for June of year t as per Fama and French (2015) convention. Only for Factset data, we use information about institutional holdings for June of year t from the past quarter that ends in March of year t. WUI and WTUI are sourced from `https://worlduncertaintyindex.com/` while Mcap_GDP and Trade_GDP from the World Bank Development Indicators website (with the exception of Taiwan).

**Table 2.** Summary statistics of firm-level tail dependence for the US and non-US sample

**US sample**

| | | LTD | | | | | UTD | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Date | Firms | Mean | St.Dev. | 25% | Median | 75% | Mean | St.Dev. | 25% | Median | 75% |
| 200006 | 3740 | 4.51 | 5.39 | 0.00 | 2.32 | 7.78 | 2.58 | 3.85 | 0.00 | 0.19 | 4.34 |
| 200106 | 3533 | 4.65 | 5.68 | 0.00 | 2.17 | 7.86 | 5.28 | 6.71 | 0.00 | 2.32 | 8.81 |
| 200206 | 3430 | 4.42 | 5.25 | 0.00 | 2.42 | 7.76 | 5.90 | 6.34 | 0.04 | 4.21 | 9.54 |
| 200306 | 3264 | 9.17 | 7.91 | 1.35 | 8.09 | 14.76 | 5.48 | 6.27 | 0.00 | 3.49 | 9.42 |
| 200406 | 3250 | 6.71 | 7.23 | 0.12 | 4.57 | 10.99 | 3.23 | 4.64 | 0.00 | 0.03 | 5.81 |
| 200506 | 3196 | 3.00 | 4.31 | 0.00 | 0.25 | 5.20 | 3.07 | 4.48 | 0.00 | 0.19 | 5.33 |
| 200606 | 3149 | 3.64 | 5.32 | 0.00 | 0.24 | 6.10 | 6.64 | 6.87 | 0.00 | 5.00 | 11.37 |
| 200706 | 3016 | 9.76 | 7.96 | 1.91 | 9.26 | 15.54 | 5.06 | 6.08 | 0.00 | 2.38 | 8.95 |
| 200806 | 2896 | 6.72 | 7.09 | 0.01 | 4.96 | 11.17 | 5.68 | 6.23 | 0.00 | 3.72 | 9.88 |
| 200906 | 2655 | 17.26 | 10.29 | 9.06 | 17.97 | 25.29 | 11.10 | 8.24 | 4.00 | 10.97 | 17.17 |
| 201006 | 2693 | 14.26 | 10.90 | 4.97 | 13.17 | 22.14 | 5.55 | 6.73 | 0.00 | 2.77 | 9.53 |
| 201106 | 2636 | 8.25 | 8.10 | 0.26 | 6.47 | 13.62 | 8.87 | 9.00 | 0.00 | 6.78 | 14.82 |
| 201206 | 2523 | 15.67 | 11.73 | 6.11 | 13.93 | 23.93 | 8.11 | 9.18 | 0.00 | 5.01 | 13.38 |
| 201306 | 2478 | 7.18 | 7.69 | 0.01 | 4.90 | 11.89 | 8.07 | 8.11 | 0.00 | 6.32 | 13.54 |
| 201406 | 2413 | 6.29 | 6.74 | 0.00 | 4.13 | 10.85 | 5.15 | 6.16 | 0.00 | 2.77 | 8.77 |
| 201506 | 2344 | 8.51 | 7.81 | 0.81 | 7.40 | 13.85 | 5.24 | 6.13 | 0.00 | 2.81 | 9.64 |
| 201606 | 2357 | 12.72 | 10.52 | 3.31 | 11.10 | 20.37 | 4.91 | 5.62 | 0.00 | 2.88 | 8.63 |
| 201706 | 2165 | 5.50 | 6.00 | 0.00 | 3.87 | 9.47 | 6.09 | 6.45 | 0.01 | 4.38 | 10.44 |
| 201806 | 2168 | 10.81 | 9.18 | 1.85 | 9.98 | 17.09 | 2.83 | 4.02 | 0.00 | 0.27 | 5.06 |
| 201906 | 2251 | 8.19 | 7.67 | 0.52 | 6.75 | 13.27 | 7.83 | 7.41 | 0.15 | 6.54 | 12.95 |
| 202006 | 2114 | 24.01 | 7.48 | 19.94 | 24.54 | 29.10 | 18.20 | 9.79 | 11.02 | 18.95 | 25.72 |
| 202106 | 2165 | 9.96 | 7.95 | 2.87 | 9.25 | 15.64 | 7.92 | 7.23 | 0.46 | 6.74 | 13.39 |
| 202206 | 2224 | 12.03 | 8.85 | 4.58 | 11.73 | 18.17 | 12.03 | 9.39 | 3.81 | 11.06 | 18.92 |
| 202306 | 2443 | 8.48 | 8.13 | 0.08 | 7.06 | 14.11 | 11.42 | 10.02 | 1.97 | 9.69 | 18.79 |

**non-US sample**

| | | LTD | | | | | UTD | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Date | Firms | Mean | St.Dev. | 25% | Median | 75% | Mean | St.Dev. | 25% | Median | 75% |
| 200006 | 1273 | 3.05 | 4.55 | 0.00 | 0.34 | 5.07 | 2.89 | 4.38 | 0.00 | 0.20 | 4.75 |
| 200106 | 2553 | 4.27 | 5.78 | 0.00 | 1.34 | 7.27 | 3.48 | 5.10 | 0.00 | 0.26 | 5.97 |
| 200206 | 3226 | 4.71 | 4.97 | 0.03 | 3.44 | 7.72 | 3.42 | 5.16 | 0.00 | 0.58 | 5.41 |
| 200306 | 3684 | 4.70 | 5.75 | 0.00 | 2.36 | 7.80 | 3.69 | 5.38 | 0.00 | 0.71 | 5.95 |
| 200406 | 4539 | 5.59 | 5.91 | 0.44 | 3.95 | 8.91 | 1.78 | 3.33 | 0.00 | 0.00 | 2.39 |
| 200506 | 5139 | 2.85 | 4.19 | 0.00 | 0.11 | 4.89 | 1.98 | 3.58 | 0.00 | 0.00 | 2.84 |
| 200606 | 5989 | 2.79 | 4.38 | 0.00 | 0.23 | 4.47 | 2.31 | 3.92 | 0.00 | 0.00 | 3.72 |
| 200706 | 7170 | 5.54 | 6.06 | 0.01 | 3.83 | 9.29 | 2.70 | 4.21 | 0.00 | 0.00 | 4.61 |
| 200806 | 7857 | 3.44 | 5.49 | 0.00 | 0.15 | 5.42 | 2.69 | 4.25 | 0.00 | 0.01 | 4.39 |
| 200906 | 7601 | 9.21 | 9.11 | 0.86 | 7.01 | 14.54 | 5.09 | 5.76 | 0.00 | 3.29 | 8.68 |
| 201006 | 8603 | 6.89 | 7.39 | 0.31 | 4.80 | 11.02 | 3.66 | 5.64 | 0.00 | 0.06 | 5.81 |
| 201106 | 9287 | 4.31 | 5.71 | 0.00 | 1.34 | 7.35 | 4.12 | 6.22 | 0.00 | 0.45 | 6.62 |
| 201206 | 9871 | 10.79 | 9.63 | 2.81 | 8.83 | 16.13 | 5.59 | 7.13 | 0.00 | 2.54 | 9.35 |
| 201306 | 10559 | 4.52 | 5.39 | 0.00 | 2.47 | 7.69 | 2.54 | 4.49 | 0.00 | 0.00 | 3.83 |
| 201406 | 11264 | 2.83 | 4.53 | 0.00 | 0.06 | 4.49 | 2.73 | 4.12 | 0.00 | 0.09 | 4.56 |
| 201506 | 11514 | 4.44 | 5.34 | 0.00 | 2.60 | 7.30 | 2.82 | 4.71 | 0.00 | 0.00 | 4.42 |
| 201606 | 11621 | 10.04 | 7.19 | 4.59 | 9.41 | 14.37 | 4.23 | 5.13 | 0.00 | 2.14 | 7.18 |
| 201706 | 12186 | 2.81 | 4.43 | 0.00 | 0.10 | 4.59 | 2.57 | 3.96 | 0.00 | 0.01 | 4.33 |
| 201806 | 12128 | 6.36 | 6.58 | 0.08 | 4.64 | 10.68 | 2.81 | 4.40 | 0.00 | 0.01 | 4.72 |
| 201906 | 13099 | 5.55 | 6.32 | 0.00 | 3.52 | 9.41 | 5.10 | 5.99 | 0.00 | 2.84 | 8.98 |
| 202006 | 13042 | 16.82 | 9.42 | 10.00 | 17.17 | 23.08 | 9.10 | 6.47 | 3.99 | 8.79 | 13.38 |
| 202106 | 13075 | 4.07 | 5.47 | 0.00 | 0.94 | 7.13 | 3.66 | 5.10 | 0.00 | 0.61 | 6.39 |
| 202206 | 14103 | 5.71 | 6.32 | 0.00 | 3.88 | 9.55 | 5.06 | 6.67 | 0.00 | 1.93 | 8.52 |
| 202306 | 14427 | 3.92 | 5.04 | 0.00 | 1.12 | 6.95 | 4.24 | 6.20 | 0.00 | 0.26 | 7.10 |

*Notes:* The table reports the left and right tail dependence coefficient (LTD and UTD respectively) of the Clayton-Gaussian-Rotated Clayton copula between a stock and its corresponding foreign market index for the US and non-US sample. The copulas are estimated with daily returns from July of year t-1 to June of year t. The corresponding foreign market index is the Fama-French Developed market index excluding the US and the CRSP value-weighted index for the US and non-US stocks, respectively. We report the mean, standard deviation and 25%, 50% (median) and 75% quantile of the distribution of LTD and UTD for each period.

**Table 3.** Correlations of firm-level left tail dependence

| US sample | | | | non-US sample | | | |
|---|---|---|---|---|---|---|---|
| Variable | Correlation | Variable | Correlation | Variable | Correlation | Variable | Correlation |
| R2 | 0.49 | e2p | 0.04 | R2 | 0.49 | e2p | 0.04 |
| Coskewness | -0.41 | es90 | -0.03 | Coskewness | -0.41 | es90 | -0.03 |
| fco_mean | 0.34 | Trade_GDP | 0.03 | fco_mean | 0.34 | Trade_GDP | 0.03 |
| log_me | 0.32 | sat | -0.03 | WUI | 0.17 | sat | -0.03 |
| WUI | 0.27 | c2d | 0.03 | log_me | 0.32 | c2d | 0.03 |
| io | 0.27 | Mom6m | 0.03 | io | 0.27 | Mom6m | 0.03 |
| io_num | 0.26 | debt2p | -0.03 | io_num | 0.26 | debt2p | -0.03 |
| fio_num | 0.24 | cto | -0.02 | fio_num | 0.24 | cto | -0.02 |
| fio | 0.18 | c | -0.02 | fio | 0.18 | c | -0.02 |
| operpro | 0.11 | s2c | -0.01 | operpro | 0.11 | s2c | -0.01 |
| roic | 0.10 | op | 0.01 | roic | 0.10 | op | 0.01 |
| Mcap_GDP | 0.09 | ipm | 0.01 | Mcap_GDP | 0.09 | ipm | 0.01 |
| skewness | -0.08 | ivc | 0.00 | skewness | -0.08 | ivc | 0.00 |
| btm | -0.08 | prof | 0.00 | btm | -0.08 | prof | 0.00 |
| max | -0.06 | dceq | 0.00 | max | -0.06 | dceq | 0.00 |
| illiquidity | -0.06 | roe | 0.00 | illiquidity | -0.06 | roe | 0.00 |
| total_vol | -0.06 | inv | 0.00 | total_vol | -0.06 | inv | 0.00 |
| var90 | -0.05 | roc | 0.00 | var90 | -0.05 | roc | 0.00 |
| io_hhi | -0.04 | WTUI | 0.00 | io_hhi | -0.04 | WTUI | 0.00 |
| dSo | -0.04 | sales_g | 0.00 | dSo | -0.04 | sales_g | 0.00 |
| a2me | -0.04 | | | a2me | -0.04 | | |

*Notes:* The table reports the correlation of the left tail dependence coefficient of the Clayton-Gaussian-Rotated Clayton copula between a stock and its corresponding foreign market index with all other variables in our sample. The copulas are estimated with daily returns from July of year t-1 to June of year t. The corresponding foreign market index is the Fama-French Developed market index excluding the US and the CRSP value-weighted index for the US and non-US stocks, respectively. Variables are sorted in descending order based on the absolute magnitude of their correlation with left tail depedence.

**Table 4.** Ownership of different types of institutional investors and left tail dependence

| | US sample | | non-US sample | | All stocks | |
|---|---|---|---|---|---|---|
| Variable | (1) | (2) | (3) | (4) | (5) | (6) |
| All investors | 0.016*** | | 0.038*** | | 0.034*** | |
| | -0.001 | | -0.002 | | -0.001 | |
| log_me | 1.042*** | 1.033*** | 0.466*** | 0.420*** | 0.615*** | 0.569*** |
| | -0.021 | -0.023 | -0.01 | -0.01 | -0.009 | -0.009 |
| Large-passive IA | | 0.031*** | | 0.114*** | | 0.073*** |
| | | -0.005 | | -0.008 | | -0.004 |
| Large-active IA | | 0.004 | | 0.053*** | | 0.055*** |
| | | -0.005 | | -0.011 | | -0.004 |
| Small-passive IA | | -0.01 | | 0.034*** | | 0.011** |
| | | -0.007 | | -0.007 | | -0.005 |
| Small-active IA | | 0.023*** | | 0.008*** | | 0.013*** |
| | | -0.003 | | -0.003 | | -0.002 |
| Private Banking | | -0.016*** | | 0.01 | | -0.003 |
| | | -0.005 | | -0.017 | | -0.005 |
| Hedge funds | | 0.009** | | 0.078*** | | 0.014*** |
| | | -0.004 | | -0.022 | | -0.004 |
| Brokers | | 0.086*** | | 0.266 | | 0.172*** |
| | | -0.023 | | -0.172 | | -0.026 |
| Long-term investors | | 0.038*** | | 0.105*** | | 0.080*** |
| | | -0.014 | | -0.011 | | -0.008 |
| Other firm-level controls | x | x | x | x | x | x |
| Year Fixed Effects | x | x | x | x | x | x |
| Country Fixed Effects | - | - | x | x | x | x |
| Observations | 65336 | 65336 | 213577 | 213577 | 278913 | 278913 |
| $R^2$ | 0.345 | 0.346 | 0.330 | 0.331 | 0.337 | 0.338 |

*Notes:* Table reports the regression results of left tail dependence (LTD) against different types of institutional ownership after controlling for size, year and country fixed effects as well as other firm-level variables. The firm-level variables that we do not report are statistically and economically insignificant and they include illiquidity, profitability (prof), book-to-market ratio (btm), cash to assets (c) and cash flows to total liabilities (c2d). We follow the methodology of Koijen et al. (2024) and we group group institutional investors into investment advisors (IA), hedge funds, long-term investors, private banking, and brokers. Because investment advisors are a large group, we further split them into four subgroups: large-passive, small-passive, small-active, and large-active. The sample spans from June of 2000 to June of 2023. Robust standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

**Table 5.** Importance of determinants of firm-level tail dependence - Robustness checks: micro-cap stocks and financials

**Panel A: Effect of micro-cap stocks**

| US sample | | | | | | | | Non-US sample | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Permutation test | | | | Change in R2 | | | | Permutation test | | | | Change in R2 | | | |
| Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value |
| R2 | 0.069 | Market | 0.321 | WUI | 0.106 | Market | 0.409 | Coskewness | 0.082 | Market | 0.344 | Coskewness | 0.111 | Market | 0.462 |
| Coskewness | 0.067 | Macro | 0.162 | R2 | 0.105 | Ownership | 0.385 | R2 | 0.079 | Profitability | 0.158 | R2 | 0.107 | Macro | 0.239 |
| WUI | 0.063 | Value | 0.157 | Coskewness | 0.074 | Macro | 0.261 | WUI | 0.044 | Value | 0.156 | WUI | 0.097 | Ownership | 0.225 |
| fco_mean | 0.035 | Ownership | 0.155 | fco_mean | 0.069 | Value | 0.182 | WTUI | 0.036 | Ownership | 0.146 | Trade_GDP | 0.064 | Profitability | 0.224 |
| Mcap_GDP | 0.034 | Profitability | 0.154 | fio_num | 0.068 | Profitability | 0.177 | Trade_GDP | 0.031 | Macro | 0.136 | io_hhi | 0.061 | Value | 0.196 |
| WTUI | 0.032 | Investment | 0.052 | io | 0.066 | Investment | 0.048 | es90 | 0.028 | Investment | 0.060 | illiquidity | 0.060 | Investment | 0.062 |
| Trade_GDP | 0.032 | | | Trade_GDP | 0.065 | | | fio | 0.026 | | | WTUI | 0.060 | | |
| fio_num | 0.030 | | | io_num | 0.064 | | | max | 0.026 | | | max | 0.058 | | |
| io_num | 0.027 | | | log_me | 0.063 | | | fio_num | 0.026 | | | total_vol | 0.057 | | |
| es90 | 0.027 | | | fio | 0.061 | | | fco_mean | 0.026 | | | es90 | 0.056 | | |
| fio | 0.026 | | | WTUI | 0.053 | | | total_vol | 0.025 | | | sat | 0.055 | | |
| var90 | 0.025 | | | io_hhi | 0.051 | | | Mcap_GDP | 0.025 | | | var90 | 0.052 | | |
| illiquidity | 0.024 | | | Mcap_GDP | 0.048 | | | io_num | 0.024 | | | cto | 0.050 | | |
| log_me | 0.023 | | | prof | 0.047 | | | var90 | 0.024 | | | s2c | 0.050 | | |
| total_vol | 0.023 | | | c | 0.045 | | | io | 0.023 | | | Mcap_GDP | 0.048 | | |

**Panel B: Effect of financial firms**

| US sample | | | | | | | | Non-US sample | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Permutation test | | | | Change in R2 | | | | Permutation test | | | | Change in R2 | | | |
| Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value |
| R2 | 0.067 | Market | 0.319 | WUI | 0.098 | Market | 0.385 | Coskewness | 0.084 | Market | 0.348 | Coskewness | 0.114 | Market | 0.468 |
| Coskewness | 0.058 | Ownership | 0.168 | R2 | 0.097 | Ownership | 0.305 | R2 | 0.080 | Profitability | 0.158 | R2 | 0.109 | Macro | 0.237 |
| WUI | 0.057 | Profitability | 0.161 | Coskewness | 0.068 | Macro | 0.230 | WUI | 0.043 | Value | 0.157 | WUI | 0.095 | Ownership | 0.224 |
| fco_mean | 0.034 | Value | 0.152 | io | 0.059 | Profitability | 0.191 | WTUI | 0.036 | Ownership | 0.145 | Trade_GDP | 0.065 | Profitability | 0.224 |
| Mcap_GDP | 0.033 | Macro | 0.149 | Trade_GDP | 0.058 | Value | 0.171 | Trade_GDP | 0.030 | Macro | 0.133 | io_hhi | 0.063 | Value | 0.198 |
| fio_num | 0.033 | Investment | 0.050 | fco_mean | 0.057 | Investment | 0.047 | es90 | 0.029 | Investment | 0.059 | WTUI | 0.060 | Investment | 0.061 |
| io_num | 0.032 | | | io_num | 0.055 | | | fio_num | 0.026 | | | es90 | 0.060 | | |
| WTUI | 0.030 | | | fio_num | 0.055 | | | fio | 0.026 | | | max | 0.058 | | |
| illiquidity | 0.030 | | | fio | 0.053 | | | fco_mean | 0.025 | | | total_vol | 0.057 | | |
| log_me | 0.029 | | | log_me | 0.052 | | | max | 0.025 | | | cto | 0.054 | | |
| Trade_GDP | 0.029 | | | Mcap_GDP | 0.048 | | | total_vol | 0.025 | | | sat | 0.053 | | |
| fio | 0.026 | | | WTUI | 0.047 | | | Mcap_GDP | 0.024 | | | var90 | 0.053 | | |
| io | 0.026 | | | prof | 0.045 | | | io_num | 0.024 | | | s2c | 0.051 | | |
| es90 | 0.025 | | | es90 | 0.044 | | | var90 | 0.023 | | | illiquidity | 0.051 | | |
| var90 | 0.024 | | | sat | 0.042 | | | io | 0.022 | | | a2me | 0.051 | | |

*Notes:* Panel A reports the variable importance when we exclude micro-cap stocks from the sample. A stock is considered micro-cap when it belongs to the bottom 20% quantile of all stocks. Panels B reports the variable importance when we exclude financial firms (SIC 6000-6999). Results for random forest regression (RFR) are presented for individual variables (column "Variable") and groups (column "Group") along with their associated variable importance measure (column "Value") for the US and non-US sample.

**Table 6.** Importance of determinants of firm-level tail dependence - Robustness checks: choice of copula

**Panel A: Clayton-Gaussian-Galambos copula**

| US sample | | | | | | | | Non-US sample | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Permutation test | | | | Change in R2 | | | | Permutation test | | | | Change in R2 | | | |
| Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value |
| Coskewness | 0.075 | Market | 0.310 | Coskewness | 0.099 | Market | 0.351 | Coskewness | 0.070 | Market | 0.334 | Coskewness | 0.128 | Market | 0.438 |
| WUI | 0.049 | Ownership | 0.163 | WUI | 0.089 | Ownership | 0.269 | R2 | 0.043 | Profitability | 0.156 | WUI | 0.101 | Macro | 0.269 |
| Trade_GDP | 0.038 | Value | 0.162 | Trade_GDP | 0.065 | Macro | 0.228 | WUI | 0.045 | Value | 0.156 | R2 | 0.080 | Profitability | 0.232 |
| R2 | 0.037 | Profitability | 0.160 | R2 | 0.064 | Profitability | 0.211 | WTUI | 0.040 | Ownership | 0.147 | WTUI | 0.074 | Ownership | 0.224 |
| Mcap_GDP | 0.033 | Macro | 0.153 | io | 0.061 | Value | 0.197 | es90 | 0.030 | Macro | 0.146 | Trade_GDP | 0.072 | Value | 0.209 |
| WTUI | 0.033 | Investment | 0.052 | fco_mean | 0.058 | Investment | 0.054 | Trade_GDP | 0.031 | Investment | 0.062 | io_hhi | 0.067 | Investment | 0.064 |
| fco_mean | 0.032 | | | fio | 0.055 | | | max | 0.027 | | | max | 0.066 | | |
| fio_num | 0.031 | | | io_num | 0.054 | | | Mcap_GDP | 0.028 | | | es90 | 0.062 | | |
| io_num | 0.029 | | | fio_num | 0.054 | | | fio | 0.028 | | | total_vol | 0.060 | | |
| illiquidity | 0.029 | | | WTUI | 0.054 | | | total_vol | 0.027 | | | cto | 0.056 | | |
| fio | 0.027 | | | io_hhi | 0.054 | | | fio_num | 0.027 | | | var90 | 0.056 | | |
| log_me | 0.027 | | | roc | 0.053 | | | fco_mean | 0.026 | | | sat | 0.054 | | |
| es90 | 0.027 | | | total_vol | 0.053 | | | var90 | 0.025 | | | s2c | 0.052 | | |
| total_vol | 0.026 | | | Mcap_GDP | 0.052 | | | io_num | 0.025 | | | fco_mean | 0.051 | | |
| io | 0.025 | | | log_me | 0.051 | | | dceq | 0.024 | | | Mcap_GDP | 0.051 | | |

**Panel B: Rotated Galambos-Gaussian-Rotated Clayton copula**

| US sample | | | | | | | | Non-US sample | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Permutation test | | | | Change in R2 | | | | Permutation test | | | | Change in R2 | | | |
| Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value |
| Coskewness | 0.064 | Market | 0.310 | Coskewness | 0.082 | Market | 0.351 | WUI | 0.051 | Market | 0.261 | total_vol | 0.095 | Market | 0.392 |
| Trade_GDP | 0.038 | Profitability | 0.173 | WTUI | 0.069 | Ownership | 0.269 | Coskewness | 0.043 | Profitability | 0.176 | WUI | 0.094 | Macro | 0.270 |
| Mcap_GDP | 0.037 | Value | 0.160 | io_num | 0.069 | Macro | 0.228 | WTUI | 0.041 | Value | 0.172 | Trade_GDP | 0.089 | Profitability | 0.256 |
| WUI | 0.035 | Ownership | 0.156 | Trade_GDP | 0.068 | Profitability | 0.211 | R2 | 0.040 | Ownership | 0.171 | max | 0.088 | Value | 0.236 |
| WTUI | 0.035 | Macro | 0.145 | io | 0.067 | Value | 0.197 | Mcap_GDP | 0.035 | Macro | 0.151 | es90 | 0.080 | Ownership | 0.231 |
| es90 | 0.031 | Investment | 0.056 | log_me | 0.066 | Investment | 0.054 | fio | 0.030 | Investment | 0.071 | var90 | 0.078 | Investment | 0.078 |
| io_num | 0.031 | | | WUI | 0.065 | | | fio_num | 0.030 | | | WTUI | 0.075 | | |
| R2 | 0.031 | | | fio_num | 0.064 | | | fco_mean | 0.030 | | | io_hhi | 0.069 | | |
| fio_num | 0.030 | | | Mcap_GDP | 0.063 | | | io_num | 0.030 | | | Mcap_GDP | 0.066 | | |
| total_vol | 0.029 | | | fio | 0.061 | | | io | 0.027 | | | s2c | 0.063 | | |
| illiquidity | 0.029 | | | fco_mean | 0.059 | | | illiquidity | 0.027 | | | sat | 0.063 | | |
| log_me | 0.029 | | | roc | 0.057 | | | dceq | 0.027 | | | R2 | 0.063 | | |
| fco_mean | 0.028 | | | R2 | 0.056 | | | Trade_GDP | 0.024 | | | a2me | 0.062 | | |
| var90 | 0.027 | | | cto | 0.054 | | | sales_g | 0.024 | | | cto | 0.062 | | |
| io | 0.026 | | | io_hhi | 0.054 | | | inv | 0.024 | | | illiquidity | 0.061 | | |

*Notes:* Panels A and B report the variable importance when we estimate left tail dependence (LTD) using the Clayton-Gaussian-Galambos and Rotated Galambos-Gaussian-Rotated Clayton copula, respectively. Results for random forest regression (RFR) are presented for individual variables (column "Variable") and groups (column "Group") along with their associated variable importance measure (column "Value") for the US and non-US sample.

**Table 7.** Outline of the 2000-2023 crisis periods

| Crisis | Start of crisis period | End of crisis period | Duration (months) | Firm-years affected |
|---|---|---|---|---|
| Dot-Com bubble | Mar 2001 | Nov 2001 | 8 | 2001, 2002 |
| Global Financial Crisis of 2007 | Dec 2007 | Jun 2009 | 18 | 2008, 2009 |
| Euro area sovereign debt crisis | May 2010 | Jun 2013 | 26 | 2011, 2012, 2013 |
| Covid-19 | Feb 2020 | Apr 2020 | 2 | 2020 |

*Notes:* The table lists the major crises for the period of 2000-2023 along with start and end dates as well as the firm-years affected. The dates for the Dot-com bubble, the Global Financial Crisis of 2007 and Covid-19 are retrieved from US NBER crisis periods while the dates for the Euro area sovereign debt crisis are retrieved from the European Central Bank (see De Jonghe, 2010). It is important to note that our firm-year observations do not follow calendar years; instead, each year in our sample spans from July of year t–1 to June of year t, following the convention of Fama and French, 1993. The firm-years affected by crisis periods reflect this convention. For example, the first crisis firm-year 2001 corresponds to the period of July 2000 to June 2001.

**Table 8.** Importance of determinants of firm-level tail dependence - Robustness checks: crisis vs non-crisis periods

**Panel A: Crisis periods**

| US sample | | | | | | | | Non-US sample | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Permutation test | | | | Change in R2 | | | | Permutation test | | | | Change in R2 | | | |
| Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value |
| WUI | 0.087 | Market | 0.309 | WUI | 0.091 | Market | 0.327 | Coskewness | 0.086 | Market | 0.335 | WUI | 0.104 | Market | 0.409 |
| R2 | 0.065 | Macro | 0.196 | R2 | 0.074 | Macro | 0.314 | R2 | 0.085 | Macro | 0.192 | R2 | 0.093 | Macro | 0.307 |
| Coskewness | 0.054 | Ownership | 0.161 | Trade_GDP | 0.070 | Ownership | 0.268 | WUI | 0.079 | Ownership | 0.144 | Coskewness | 0.088 | Ownership | 0.217 |
| Trade_GDP | 0.041 | Value | 0.145 | Coskewness | 0.046 | Profitability | 0.157 | WTUI | 0.056 | Profitability | 0.137 | WTUI | 0.062 | Profitability | 0.175 |
| WTUI | 0.035 | Profitability | 0.144 | io_num | 0.045 | Value | 0.134 | Mcap_GDP | 0.029 | Value | 0.136 | Trade_GDP | 0.054 | Value | 0.155 |
| fio_num | 0.033 | Investment | 0.046 | es90 | 0.044 | Investment | 0.033 | Trade_GDP | 0.028 | Investment | 0.056 | max | 0.048 | Investment | 0.051 |
| Mcap_GDP | 0.032 | | | fio_num | 0.040 | | | fio_num | 0.027 | | | es90 | 0.047 | | |
| io_num | 0.032 | | | total_vol | 0.039 | | | max | 0.026 | | | io_hhi | 0.046 | | |
| fco_mean | 0.031 | | | fio | 0.039 | | | io_num | 0.025 | | | var90 | 0.042 | | |
| illiquidity | 0.029 | | | WTUI | 0.039 | | | fio | 0.025 | | | total_vol | 0.042 | | |
| es90 | 0.029 | | | io | 0.039 | | | es90 | 0.024 | | | a2me | 0.041 | | |
| log_me | 0.027 | | | log_me | 0.038 | | | io | 0.024 | | | s2c | 0.039 | | |
| fio | 0.026 | | | fco_mean | 0.037 | | | dceq | 0.022 | | | debt2p | 0.039 | | |
| total_vol | 0.025 | | | sat | 0.037 | | | fco_mean | 0.022 | | | illiquidity | 0.038 | | |
| io | 0.024 | | | max | 0.037 | | | illiquidity | 0.022 | | | fio_num | 0.036 | | |

**Panel B: Non-crisis periods**

| US sample | | | | | | | | Non-US sample | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Permutation test | | | | Change in R2 | | | | Permutation test | | | | Change in R2 | | | |
| Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value |
| R2 | 0.060 | Market | 0.323 | WUI | 0.089 | Market | 0.344 | Coskewness | 0.066 | Market | 0.316 | Coskewness | 0.094 | Market | 0.396 |
| Coskewness | 0.059 | Ownership | 0.176 | R2 | 0.085 | Ownership | 0.302 | R2 | 0.062 | Profitability | 0.175 | WUI | 0.094 | Profitability | 0.239 |
| WUI | 0.046 | Value | 0.159 | Mcap_GDP | 0.079 | Macro | 0.217 | Trade_GDP | 0.035 | Value | 0.170 | R2 | 0.085 | Macro | 0.214 |
| io_num | 0.035 | Profitability | 0.158 | Coskewness | 0.070 | Profitability | 0.199 | WUI | 0.035 | Ownership | 0.151 | Trade_GDP | 0.074 | Ownership | 0.213 |
| log_me | 0.035 | Macro | 0.131 | log_me | 0.060 | Value | 0.182 | WTUI | 0.032 | Macro | 0.127 | io_hhi | 0.066 | Value | 0.209 |
| illiquidity | 0.033 | Investment | 0.052 | io_num | 0.059 | Investment | 0.050 | fco_mean | 0.027 | Investment | 0.062 | total_vol | 0.063 | Investment | 0.069 |
| fco_mean | 0.033 | | | io | 0.056 | | | fio_num | 0.026 | | | es90 | 0.060 | | |
| fio_num | 0.032 | | | fio | 0.056 | | | fio | 0.026 | | | s2c | 0.059 | | |
| Trade_GDP | 0.031 | | | WTUI | 0.055 | | | io_num | 0.026 | | | var90 | 0.059 | | |
| fio | 0.029 | | | fco_mean | 0.053 | | | Mcap_GDP | 0.025 | | | roc | 0.059 | | |
| Mcap_GDP | 0.029 | | | fio_num | 0.052 | | | illiquidity | 0.025 | | | max | 0.058 | | |
| io | 0.027 | | | Trade_GDP | 0.050 | | | var90 | 0.025 | | | Mcap_GDP | 0.056 | | |
| WTUI | 0.025 | | | prof | 0.046 | | | es90 | 0.024 | | | WTUI | 0.056 | | |
| total_vol | 0.024 | | | cto | 0.044 | | | total_vol | 0.024 | | | cto | 0.055 | | |
| es90 | 0.024 | | | sat | 0.044 | | | prof | 0.024 | | | sat | 0.054 | | |

*Notes:* Panels A and B report the variable importance when we split our sample into crisis and non-crisis periods (see Table 7 for a definition of the crisis periods). Results for both general-to-specific modelling (GETS) and random forest regression (RFR) are presented for individual variables (column "Variable") and groups (column "Group") along with their associated variable importance measure (column "Value").

**Table 9.** Importance of determinants of firm-level tail dependence - Robustness checks: Developed vs Emerging Markets

**Panel A: Non-US sample**

| Developed Markets | | | | | | | | Emerging Markets | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Permutation test | | | | Change in R2 | | | | Permutation test | | | | Change in R2 | | | |
| Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value |
| Coskewness | 0.080 | Market | 0.348 | Coskewness | 0.103 | Market | 0.458 | Coskewness | 0.086 | Market | 0.346 | Coskewness | 0.111 | Market | 0.461 |
| R2 | 0.078 | Profitability | 0.161 | R2 | 0.099 | Macro | 0.242 | R2 | 0.078 | Value | 0.157 | R2 | 0.100 | Ownership | 0.211 |
| WUI | 0.047 | Value | 0.15 | WUI | 0.088 | Ownership | 0.227 | WUI | 0.043 | Profitability | 0.154 | WUI | 0.096 | Profitability | 0.211 |
| WTUI | 0.037 | Ownership | 0.148 | Trade_GDP | 0.071 | Profitability | 0.217 | WTUI | 0.036 | Ownership | 0.147 | Trade_GDP | 0.058 | Macro | 0.210 |
| Trade_GDP | 0.035 | Macro | 0.142 | es90 | 0.063 | Value | 0.185 | fio | 0.029 | Macro | 0.134 | io_hhi | 0.056 | Value | 0.189 |
| es90 | 0.031 | Investment | 0.053 | max | 0.063 | Investment | 0.049 | Trade_GDP | 0.028 | Investment | 0.064 | WTUI | 0.055 | Investment | 0.066 |
| io_num | 0.027 | | | total_vol | 0.060 | | | max | 0.028 | | | fco_mean | 0.054 | | |
| total_vol | 0.027 | | | io_hhi | 0.058 | | | Mcap_GDP | 0.027 | | | illiquidity | 0.053 | | |
| fio_num | 0.026 | | | WTUI | 0.058 | | | fco_mean | 0.027 | | | max | 0.051 | | |
| io | 0.026 | | | Mcap_GDP | 0.056 | | | fio_num | 0.026 | | | cto | 0.051 | | |
| var90 | 0.025 | | | var90 | 0.054 | | | illiquidity | 0.025 | | | sat | 0.049 | | |
| fio | 0.025 | | | s2c | 0.049 | | | total_vol | 0.023 | | | s2c | 0.049 | | |
| fco_mean | 0.024 | | | sat | 0.049 | | | es90 | 0.023 | | | roc | 0.049 | | |
| illiquidity | 0.024 | | | cto | 0.048 | | | io_num | 0.023 | | | log_me | 0.048 | | |
| log_me | 0.023 | | | illiquidity | 0.045 | | | e2p | 0.022 | | | io | 0.048 | | |

*Notes:* Panels A reports the variable importance when we split our non-US sample into Developed and Emerging Markets. we use the MSCI Market classification as of April 2025 (`https://www.msci.com/our-solutions/indexes/market-classification`) to characterize the set of Developed countries. We consider the countries that do not belong to the set of Developed Markets as Emerging Markets. The list of countries and their classification can be found in Table 13. Results for random forest regression (RFR) are presented for individual variables (column "Variable") and groups (column "Group") along with their associated variable importance measure (column "Value").

**Table 10.** Importance of determinants of firm-level tail dependence - Robustness checks: pre- and post-2010 era

**Panel A: Pre-2010 era**

| US sample | | | | | | | | Non-US sample | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Permutation test | | | | Change in R2 | | | | Permutation test | | | | Change in R2 | | | |
| Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value |
| WUI | 0.071 | Market | 0.312 | WUI | 0.090 | Market | 0.341 | R2 | 0.083 | Market | 0.335 | R2 | 0.099 | Market | 0.394 |
| R2 | 0.065 | Macro | 0.176 | R2 | 0.079 | Ownership | 0.285 | Coskewness | 0.053 | Profitability | 0.163 | WUI | 0.073 | Profitability | 0.218 |
| Coskewness | 0.047 | Ownership | 0.164 | io_num | 0.052 | Macro | 0.236 | WUI | 0.044 | Value | 0.161 | WTUI | 0.067 | Macro | 0.214 |
| Mcap_GDP | 0.041 | Value | 0.152 | Coskewness | 0.050 | Profitability | 0.178 | Trade_GDP | 0.032 | Ownership | 0.154 | Coskewness | 0.066 | Ownership | 0.209 |
| io_num | 0.033 | Profitability | 0.147 | log_me | 0.050 | Value | 0.165 | WTUI | 0.030 | Macro | 0.128 | Trade_GDP | 0.065 | Value | 0.197 |
| illiquidity | 0.033 | Investment | 0.048 | io | 0.049 | Investment | 0.043 | fio_num | 0.030 | Investment | 0.060 | illiquidity | 0.060 | Investment | 0.059 |
| WTUI | 0.033 | | | Trade_GDP | 0.048 | | | io_num | 0.029 | | | Mcap_GDP | 0.056 | | |
| fio_num | 0.033 | | | sat | 0.048 | | | max | 0.028 | | | io_hhi | 0.054 | | |
| Trade_GDP | 0.032 | | | WTUI | 0.047 | | | var90 | 0.027 | | | max | 0.053 | | |
| log_me | 0.030 | | | fio | 0.047 | | | total_vol | 0.027 | | | sat | 0.051 | | |
| fco_mean | 0.029 | | | fio_num | 0.046 | | | es90 | 0.027 | | | s2c | 0.050 | | |
| fio | 0.027 | | | Mcap_GDP | 0.046 | | | fio | 0.027 | | | a2me | 0.049 | | |
| es90 | 0.026 | | | fco_mean | 0.043 | | | io | 0.026 | | | cto | 0.048 | | |
| var90 | 0.026 | | | cto | 0.042 | | | log_me | 0.025 | | | total_vol | 0.046 | | |
| io | 0.026 | | | es90 | 0.041 | | | Mom6m | 0.024 | | | var90 | 0.046 | | |

**Panel B: Post-2010 era**

| US sample | | | | | | | | Non-US sample | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Permutation test | | | | Change in R2 | | | | Permutation test | | | | Change in R2 | | | |
| Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value | Variable | Value | Group | Value |
| Coskewness | 0.077 | Market | 0.325 | Trade_GDP | 0.119 | Market | 0.365 | Coskewness | 0.092 | Market | 0.350 | Coskewness | 0.121 | Market | 0.470 |
| R2 | 0.056 | Profitability | 0.161 | Coskewness | 0.083 | Ownership | 0.288 | R2 | 0.076 | Profitability | 0.160 | R2 | 0.099 | Macro | 0.223 |
| WUI | 0.046 | Ownership | 0.156 | R2 | 0.083 | Macro | 0.230 | WUI | 0.047 | Value | 0.154 | WUI | 0.088 | Ownership | 0.222 |
| Trade_GDP | 0.038 | Value | 0.156 | WUI | 0.068 | Profitability | 0.191 | WTUI | 0.041 | Macro | 0.144 | Trade_GDP | 0.060 | Profitability | 0.214 |
| WTUI | 0.035 | Macro | 0.152 | fio | 0.053 | Value | 0.167 | Trade_GDP | 0.030 | Ownership | 0.136 | io_hhi | 0.060 | Value | 0.192 |
| Mcap_GDP | 0.034 | Investment | 0.050 | io_num | 0.052 | Investment | 0.044 | es90 | 0.029 | Investment | 0.056 | total_vol | 0.059 | Investment | 0.054 |
| log_me | 0.031 | | | io | 0.052 | | | fio | 0.026 | | | es90 | 0.058 | | |
| io_num | 0.031 | | | fco_mean | 0.051 | | | Mcap_GDP | 0.026 | | | WTUI | 0.058 | | |
| illiquidity | 0.030 | | | log_me | 0.051 | | | total_vol | 0.025 | | | max | 0.056 | | |
| fio_num | 0.029 | | | fio_num | 0.049 | | | fio_num | 0.025 | | | var90 | 0.051 | | |
| fco_mean | 0.029 | | | debt2p | 0.044 | | | max | 0.025 | | | sat | 0.051 | | |
| es90 | 0.026 | | | WTUI | 0.044 | | | var90 | 0.023 | | | a2me | 0.050 | | |
| fio | 0.025 | | | prof | 0.043 | | | illiquidity | 0.022 | | | cto | 0.050 | | |
| io | 0.024 | | | cto | 0.043 | | | io_num | 0.022 | | | illiquidity | 0.049 | | |
| total_vol | 0.024 | | | sat | 0.043 | | | roe | 0.022 | | | s2c | 0.047 | | |

*Notes:* Panels A and B report the variable importance when we split our sample into a pre- and a post-2010 subsample. The pre-2010 subperiod encompasses the events of the Global Financial Crisis of 2007-09 while the post-2010 subperiod signifies the emergence of the new post-GFC era. Results for both general-to-specific modelling (GETS) and random forest regression (RFR) are presented for individual variables (column "Variable") and groups (column "Group") along with their associated variable importance measure (column "Value").

**Table 11.** Random Forest Regression parameters

| **Parameter** | **Description** | **Value** |
|---|---|---|
| n_estimators | The number of trees in the forest. | 1000 |
| criterion | The function to measure the quality of a split. | squared_error |
| max_depth | The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples. | None |
| min_samples_split | The minimum number of samples required to split an internal node. | 10 |
| min_samples_leaf | The minimum number of samples required to be at a leaf node. A split point at any depth will only be considered if it leaves at least min_samples_leaf training samples in each of the left and right branches. | 5 |
| max_features | The number of features to consider when looking for the best split. | 39 |
| bootstrap | If False, the whole dataset is used to build each tree. If True, only a part of the dataset will be used. | True |
| random_state | Controls both the randomness of the bootstrapping of the samples used when building trees (if bootstrap=True) and the sampling of the features to consider when looking for the best split at each node (max_features≤39). | 7 |
| max_samples | If bootstrap is True, the fraction of samples to draw from X to train each base estimator. | 0.66 or 2/3 |

*Notes:* Table summarizes the Random Forest Regression parameters. We use the RandomForestRegressor class of Pedregosa et al. (2011) to build our Trees.

**Table 12.** XGBoost Regressor parameters

| **Parameter** | **Description** | **Value** |
|---|---|---|
| n_estimators | The number of trees in the forest. | 1000 |
| min_child_weight | The minimum number of samples required to be at a leaf node. A split point at any depth will only be considered if it leaves at least min_samples_leaf training samples in each of the left and right branches. | 5 |
| max_depth | The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples. | None |
| learning_rate | Must be 1 for random forest behavior. | 1 |
| subsample | Sample size for bootstrapping. | 0.632 |
| colsample_bynode | The number of features to consider when looking for the best split. If 1, then all features are used. | 1 |
| reg_lambda | Controls the L2 regularization. If 0, there is no L2 regularization. | 0 |
| reg_alpha | Controls the L1 regularization. If 0, there is no L1 reguralization. | 0 |
| tree_method | Controls the approximation algorithm that trains trees. | hist |
| random_state | Controls both the randomness of the bootstrapping of the samples used when building trees (if bootstrap=True) and the sampling of the features to consider when looking for the best split at each node (max_features ¡ 39). | 1 |

*Notes:* Table summarizes the Random Forest Regression parameters using the XGBoost library (Chen and Guestrin (2016)). We use the Regressor class of XGBoost to build our Trees.

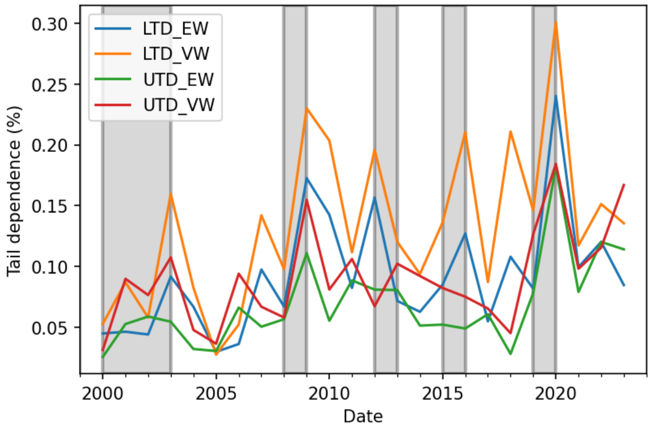
**Table 13.** Country and firm coverage of the non-US dataset

| Country | iso2 | Firms | Market cap | Market | Country | iso2 | Firms | Market cap | Market |
|---|---|---|---|---|---|---|---|---|---|
| China | CN | 4783 | 9516.07 | EM | Viet Nam | VN | 251 | 64.19 | EM |
| Japan | JP | 3428 | 4885.12 | DM | Portugal | PT | 41 | 61.37 | DM |
| France | FR | 698 | 2736.53 | DM | Greece | GR | 172 | 46.48 | EM |
| India | IN | 1179 | 2295.68 | EM | Colombia | CO | 22 | 40.1 | EM |
| United Kingdom | GB | 1414 | 1773.29 | DM | Argentina | AR | 42 | 30.36 | EM |
| Germany | DE | 608 | 1658.85 | DM | Morocco | MA | 30 | 25.71 | EM |
| Korea, Republic of | KR | 1641 | 1565.63 | EM | Czechia | CZ | 9 | 24.6 | EM |
| Taiwan, Province of China | TW | 1385 | 1493.82 | EM | Bermuda | BM | 24 | 22.07 | EM |
| Switzerland | CH | 202 | 1397.33 | DM | Peru | PE | 24 | 18.98 | EM |
| Netherlands | NL | 135 | 917.61 | DM | Romania | RO | 52 | 18.85 | EM |
| Hong Kong | HK | 872 | 914.65 | DM | United States | US | 37 | 18.6 | DM |
| Australia | AU | 1143 | 913.99 | DM | Pakistan | PK | 267 | 14.55 | EM |
| Sweden | SE | 559 | 583.86 | DM | Hungary | HU | 23 | 14.07 | EM |
| Denmark | DK | 130 | 525.09 | DM | Egypt | EG | 67 | 11.18 | EM |
| Spain | ES | 162 | 517.88 | DM | Jordan | JO | 13 | 9.09 | EM |
| United Arab Emirates | AE | 41 | 484.38 | EM | Bangladesh | BD | 49 | 8.16 | EM |
| Italy | IT | 382 | 457.57 | DM | Iceland | IS | 12 | 7.13 | EM |
| Saudi Arabia | SA | 136 | 422.07 | EM | Bahrain | BH | 4 | 6.36 | EM |
| Thailand | TH | 446 | 376.4 | EM | Slovenia | SI | 20 | 6.3 | EM |
| Mexico | MX | 86 | 322.06 | EM | Croatia | HR | 49 | 6.17 | EM |
| Indonesia | ID | 309 | 287.28 | EM | Kenya | KE | 22 | 5.48 | EM |
| Brazil | BR | 200 | 281.93 | EM | Sri Lanka | LK | 56 | 4.46 | EM |
| Norway | NO | 231 | 237.16 | DM | Panama | PA | 1 | 4.35 | EM |
| South Africa | ZA | 243 | 217.02 | EM | Lithuania | LT | 26 | 3.78 | EM |
| Malaysia | MY | 772 | 212.1 | EM | Estonia | EE | 19 | 3.32 | EM |
| Belgium | BE | 106 | 211.91 | DM | Senegal | SN | 1 | 2.46 | EM |
| Finland | FI | 175 | 199.26 | DM | Malta | MT | 15 | 2.42 | EM |
| Singapore | SG | 429 | 176.96 | DM | Bulgaria | BG | 32 | 1.95 | EM |
| Israel | IL | 368 | 141.64 | EM | Tunisia | TN | 15 | 1.71 | EM |
| Philippines | PH | 106 | 124.01 | EM | Cyprus | CY | 23 | 1.65 | EM |
| Turkey | TR | 207 | 123.79 | EM | Serbia | RS | 11 | 1.14 | EM |
| Ireland | IE | 49 | 101.7 | DM | Palestine, State of | PS | 1 | 1.02 | EM |
| Poland | PL | 422 | 94.15 | EM | Gabon | GA | 1 | 0.89 | EM |
| Luxembourg | LU | 35 | 90.27 | DM | Ukraine | UA | 11 | 0.36 | EM |
| New Zealand | NZ | 104 | 78.71 | DM | Slovakia | SK | 4 | 0.18 | EM |
| Austria | AT | 62 | 78.4 | DM | Latvia | LV | 10 | 0.15 | EM |
| Chile | CL | 81 | 71.35 | EM | | | | | |

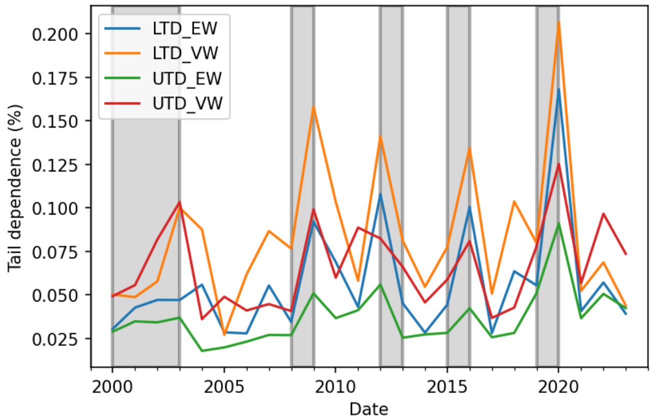
*Notes:* Table summarizes the country and firm coverage of our final non-US dataset for the 2000-2023 period. "iso2" is the 2-digit ISO code of the country, "Firms" is the number of unique companies from the corresponding country and "Market cap" is the total market capitalization of all stocks in a country in billions USD as of June of 2023. "Market" denotes the characterization of the country as a Developed Market (DM) or as an Emerging Market (EM). We use the MSCI Market classification as of April 2025 (`https://www.msci.com/our-solutions/indexes/market-classification`) to characterize the set of Developed Markets. We consider the countries that do not belong to the set of DMs as Emerging Markets. Countries are ranked in descending order based on the total stock market capitalization in June of 2023.

**Table 14.** Institution type classification according to FactSet entity sub-type

| Institution type | FactSet entity sub-type | Institution type | FactSet entity sub-type |
|---|---|---|---|
| Investment Advisor (IA) | IA: Investment Advisor | Brokers | MM: Market Maker |
| | IC: Investment Company | | BM: Bank Management Division |
| | RE: Research Firm | | IB: Investment Banking |
| | PP: Real Estate Manager | | ST: Stock Borrowing/Lending |
| | SB: Subsidiary Branch | | BR: Broker |
| | MF: Mutual Fund Manager | | |
| | ML: Master Ltd part | Long-term investors | FO: Foundation/Endowment Manager |
| | | | SV: Sovereign Wealth Manager |
| Private Banking | CP: Corporate Portfolio | | IN: Insurance Company |
| | CU: Custodial | | PF: Pension funds |
| | FY: Family Office | | GV: 'Government/Fed/local/agency |
| | PB: Private Banking | | |
| | VC: Venture Capital/Pvt Equity | | |
| | | | |
| Hedge Funds | FH: Fund of Hedge Funds Manager | | |
| | FU: Fund | | |
| | FS: Fund Distributor | | |
| | HF: Hedge Fund Company | | |
| | AR: Arbitrage | | |



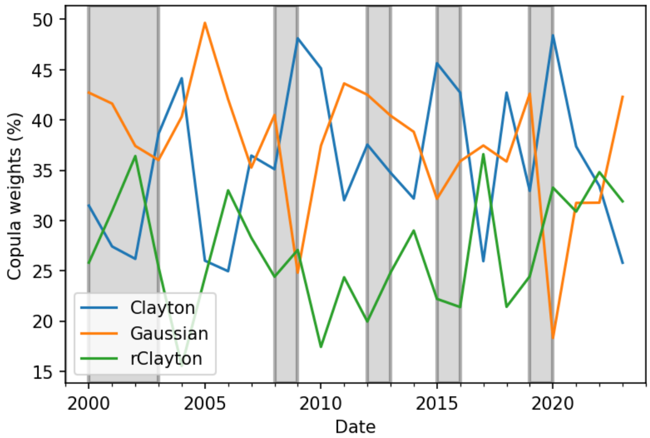

**(a)** EW and VW average LTD and UTD for US stocks



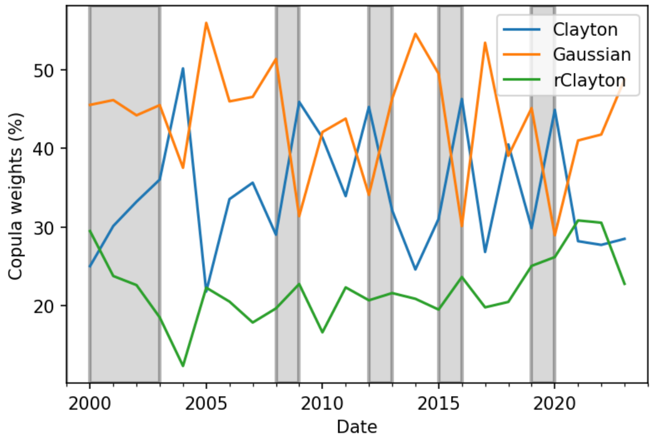

**(b)** EW and VW average LTD and UTD for non-US stocks

**Figure 1.** Firm-level tail dependence means

*Notes:* Figures 1a and 1b shows the equal- and value-weighted (EW and VW) average value of the LTD and UTD for all US and non-US stocks in our sample. The LTD and UTD measures are estimated annually from the daily returns of a stock and its corresponding foreign market index using the fitted Clayton-Gauss-Rotated Clayton copula. Gray shaded areas correspond to NBER recession periods for US and for Developed Markets excluding the US in the US and non-US sample, respectively.




**(a)** EW average copula weights for US stocks




**(b)** EW average copula weights for non-US stocks

**Figure 2.** Firm-level dependence structure

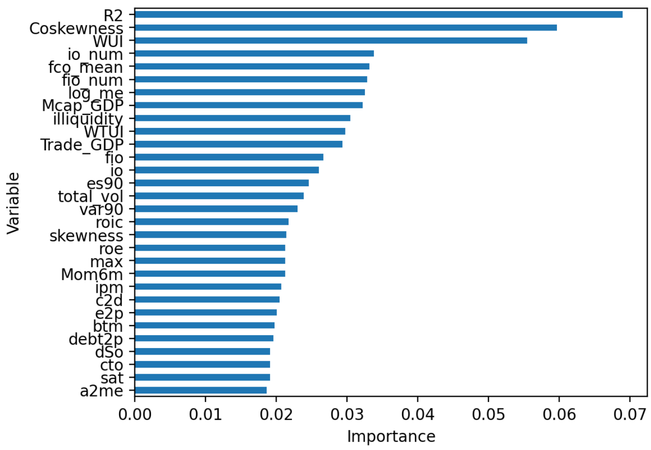
*Notes:* Figures 2a and 2b show the equal-weighted average weights that are assigned to the Clayton, Gaussian and Rotated Clayton (rClayton) copulas in the estimation process. The weights, $w_{Clayton}$, $w_{Gaussian}$, $w_{rClayton}$, are representative of the dependence structure of the stock and the corresponding foreign market index. When $w_{Clayton}$ ($w_{rClayton}$) increases, the stock and the index exhibit a dependence structure that is left (right) tail dominant. On the contrary, an increase of $w_{Gaussian}$ indicates a structure of weaker left and right tail dependence. Gray shaded areas correspond to NBER recession periods for US and for Developed Markets excluding the US in the US and non-US sample, respectively.


**(a)** Persistence of LTD for US stocks

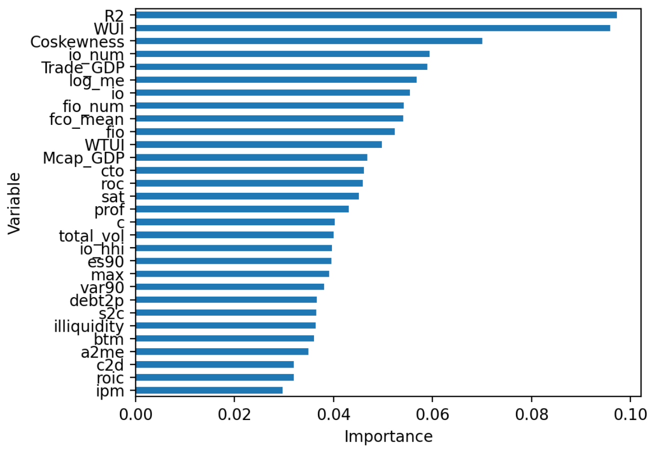
**(b)** Persistence of LTD for non-US stocks

**Figure 3.** Year-to-year persistence of left tail dependence

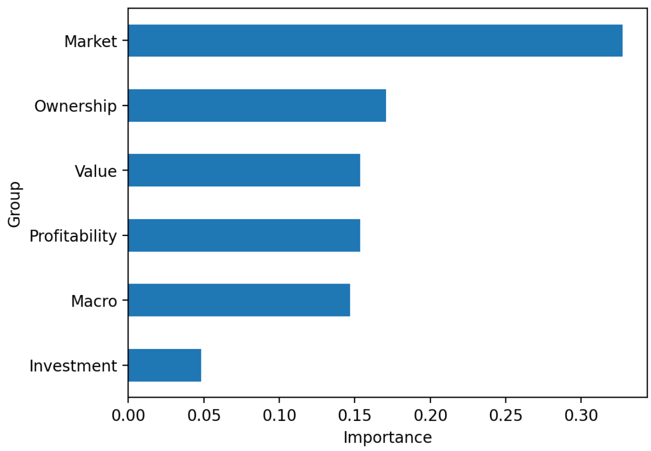
*Notes:* Figures 3a and 3b plot the persistence of LTD for US and non-US stocks. Persistence is measured as the relative frequency at which a stock is sorted into a LTD quintile portfolio i in year t given that it was in same portfolio i in year t-1. The rank 1 portfolio contains the 20% stocks with the lowest LTD while rank 5 contains those with the highest LTD. For example, a value of 43% for the rank 5 portfolio in the US in year 2015 means that 43% of stocks that belonged to the rank 5 portfolio in year 2014 remained in it in 2015. Gray shaded areas correspond to NBER recession periods for US and for Developed Markets excluding the US in the US and non-US sample, respectively.



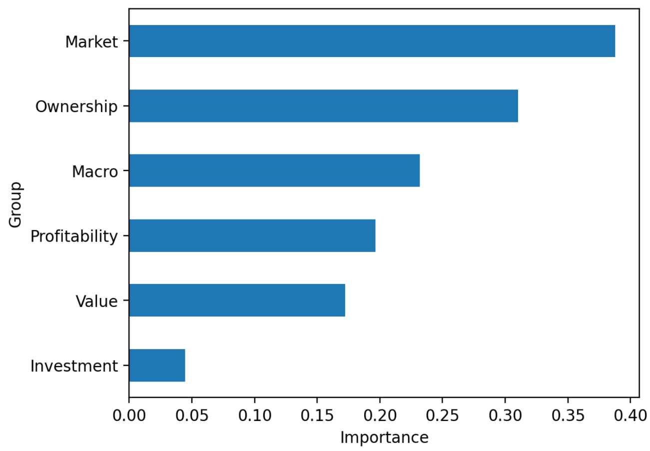

**(a)** Permutation test of LTD for variables




**(b)** Change in R2 of LTD for variables



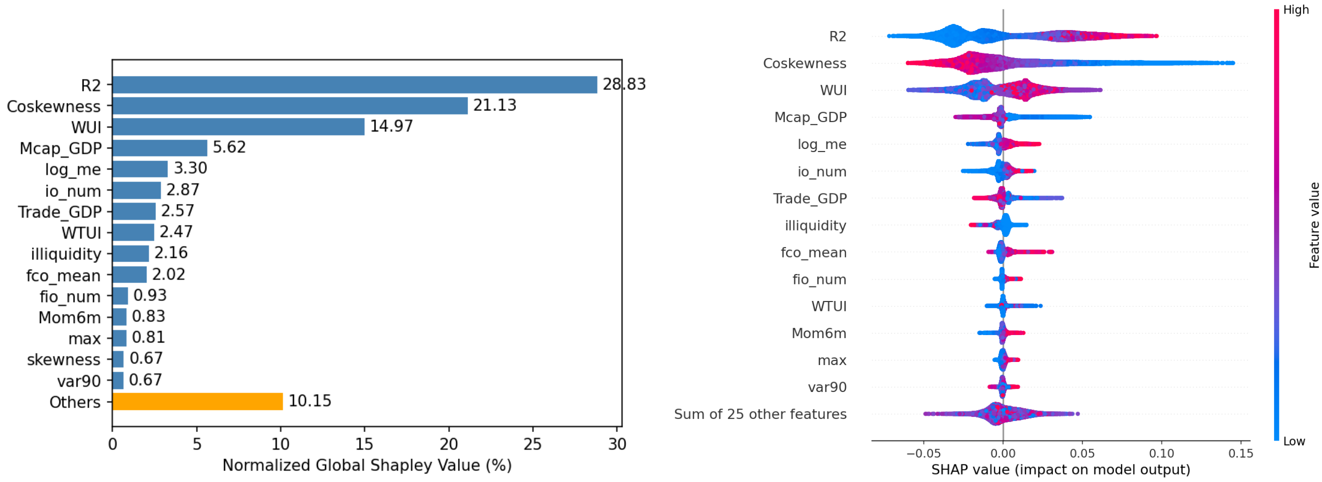

**(c)** Permutation test of LTD for groups




**(d)** Change in R2 of LTD for groups

**Figure 4.** Importance of determinants of firm-level left tail dependence for the US sample

*Notes:* Figure shows the importance of individual variables and their groups as determinants of firm-level left tail dependence of US stocks with their corresponding foreign market index. Variable importance is calculated from both the permutation test and the change in R2 of the random forest regression model. The value of the permutation test for variable j is the average difference in prediction accuracy before and after permuting j in the data matrix. The sum of the permutation test scores of all variables is normalized to equal 1. The change in R2 corresponds to the reduction in predictive R2 from setting all values of variable j to zero, while holding the remaining model estimates fixed. The higher the permutation test score and the change of R2 are for variable j, the more important that variable is in explaining left tail dependence.

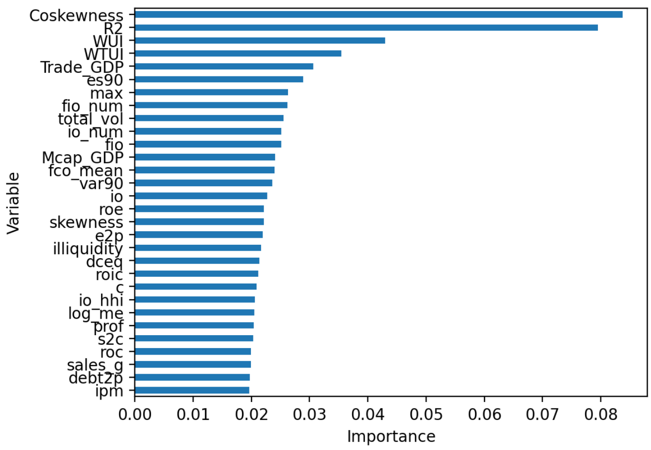



**(a)** Normalized Global Shapley Values for US stocks

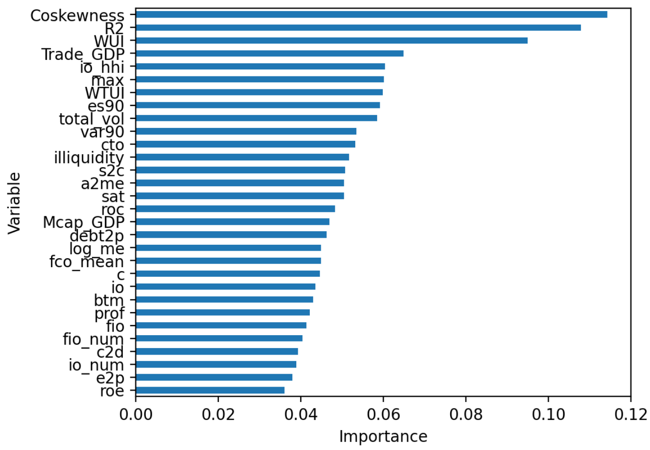
**(b)** Beeswarm for US stocks

**Figure 5.** Shapley values for US sample

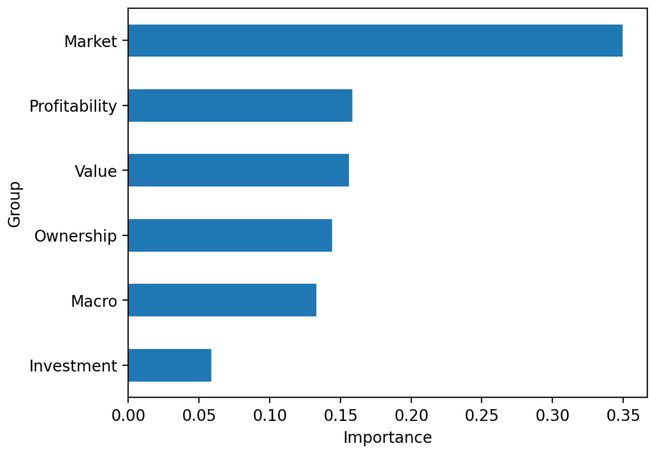
*Notes:* Figures 5a and 5b plot the normalized Global Shapley Values (GSVs) of the 15 most important individual variables and the corresponding Beeswarm of those variables for the US sample, respectively. The Global Shapley Value $\Phi_j$ of feature j is calculated as the average absolute mean Shapley value of that feature over all instances of the dataset. To compare feature importance measured by other methods, we normalize the GSV of feature j by dividing with the sum of the GSVs of all features: $\Phi'_j = 100 \times \frac{\Phi_j}{\sum_{j=1}^{p} \Phi_j}$. The normalized GSV is quoted in percentage terms (%). The Beeswarm shows how the predicted value of an instance changes as the value of a single feature is varied. For example, if low values of a feature (blue color) transition to high values (red color) from left to right, then the feature has a positive impact on the Shapley value and as such, it is positively correlated with left tail dependence. The Shapley values are calculated using the TreeSHAP methodology of Lundberg et al. (2018) which is fine tuned to tree-based machine learning models such as random forests.



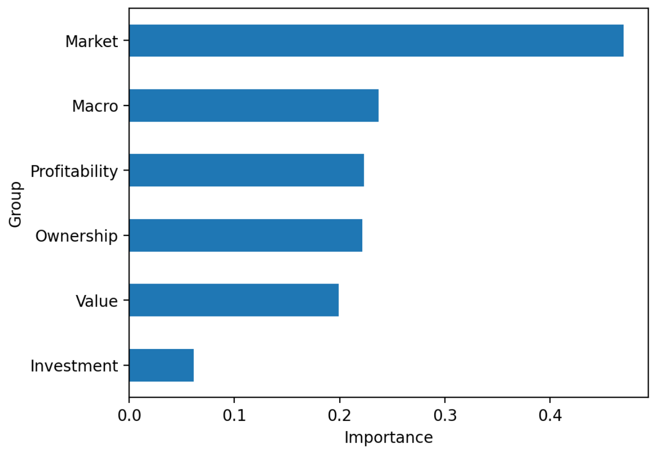

**(a)** Permutation test of LTD for variables




**(b)** Change in R2 of LTD for variables




**(c)** Permutation test of LTD for groups




**(d)** Change in R2 of LTD for groups

**Figure 6.** Importance of determinants of firm-level left tail dependence for the non-US sample

*Notes:* Figure shows the importance of individual variables and their groups as determinants of firm-level left tail dependence of non-US stocks with their corresponding foreign market index. Variable importance is calculated from both the permutation test and the change in R2 of the random forest regression model. The value of the permutation test for variable j is the average difference in prediction accuracy before and after permuting j in the data matrix. The sum of the permutation test scores of all variables is normalized to equal 1. The change in R2 corresponds to the reduction in predictive R2 from setting all values of variable j to zero, while holding the remaining model estimates fixed. The higher the permutation test score and the change of R2 are for variable j, the more important that variable is in explaining left tail dependence.

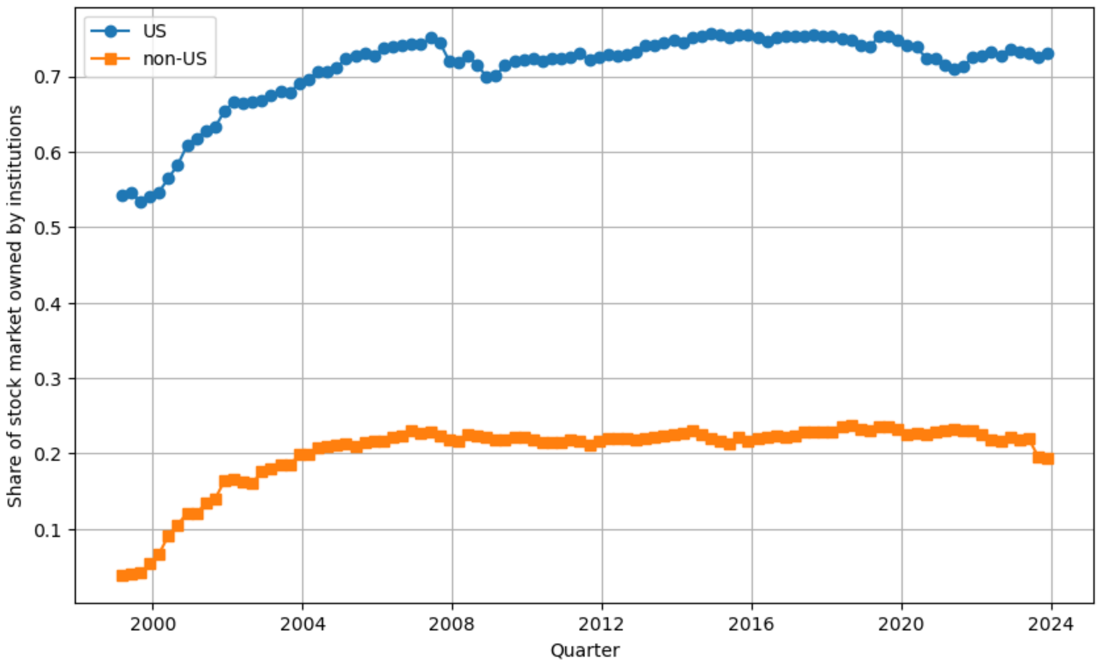
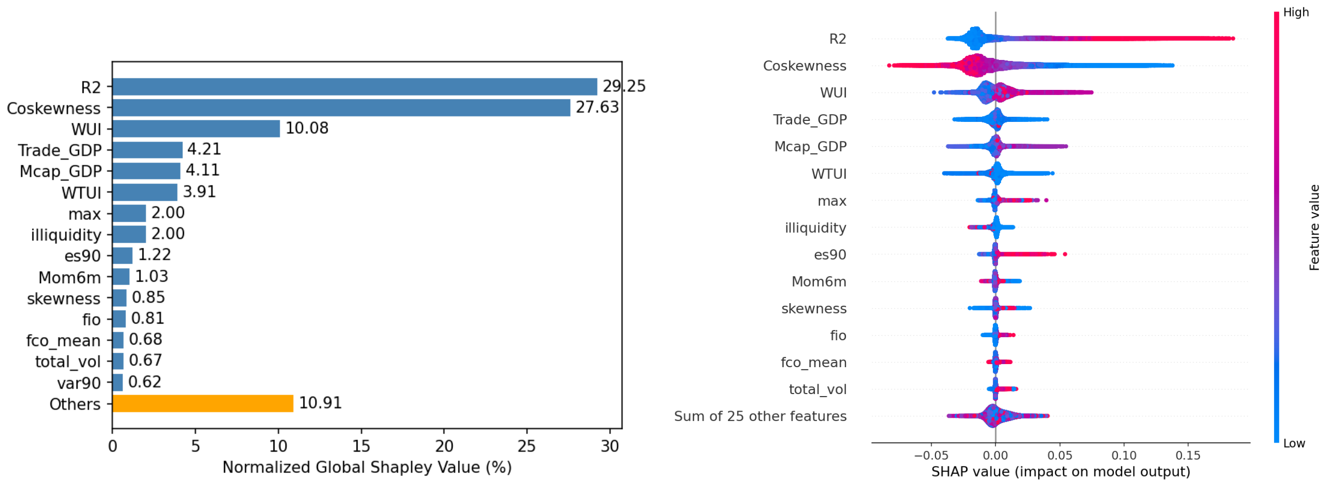



**Figure 7.** Fraction of the stock market capitalization held by institutional investors

*Notes:* Figure plots the fraction of the total market capitalization held by institutional investors for the US and non-US market. Institutional ownership data from March 1999 to December 2023 are sourced from FactSet and processed following the methodology of Ferreira and Matos (2008).

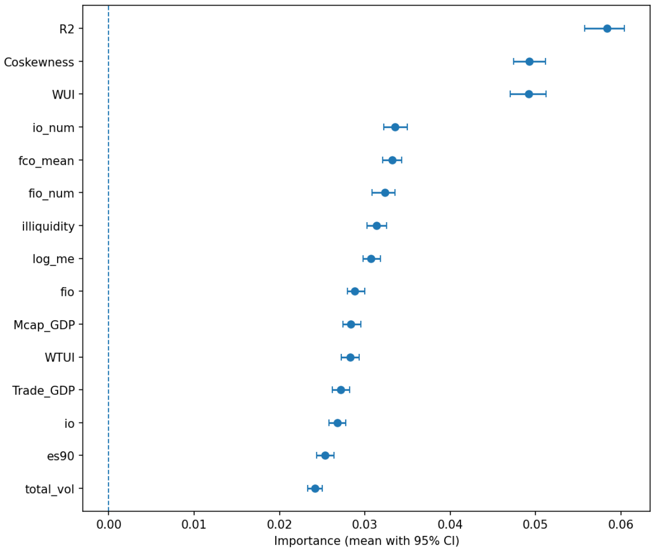



**(a)** Global Shapley Values for non-US stocks

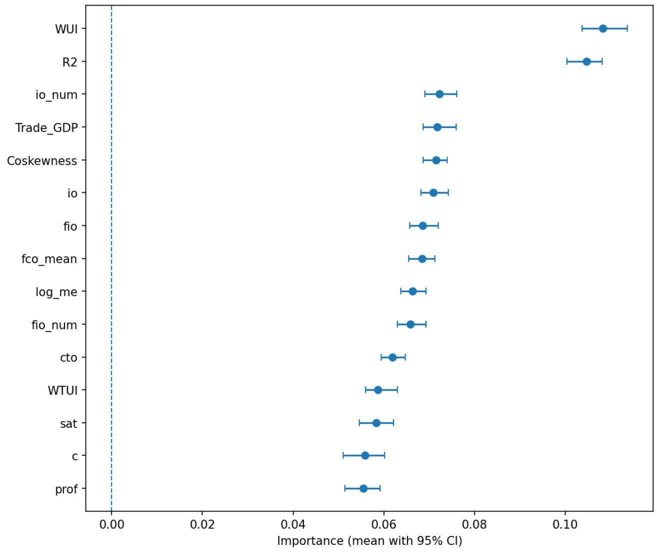
**(b)** Beeswarm for non-US stocks

**Figure 8.** Shapley values for non-US sample

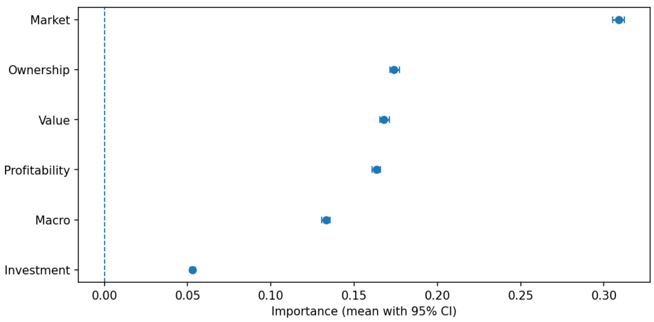
*Notes:* Figures 8a and 8b plot the normalized Global Shapley Values (GSVs) of the 15 most important individual variables and the corresponding Beeswarm of those variables for the non-US sample, respectively. The Global Shapley Value $\Phi_j$ of feature j is calculated as the average absolute mean Shapley value of that feature over all instances of the dataset. To compare feature importance measured by other methods, we normalize the GSV of feature j by dividing with the sum of the GSVs of all features: $\Phi'_j = 100 \times \frac{\Phi_j}{\sum_{j=1}^{p} \Phi_j}$. The normalized GSV is quoted in percentage terms (%). The Beeswarm shows how the predicted value of an instance changes as the value of a single feature is varied. For example, if low values of a feature (blue color) transition to high values (red color) from left to right, then the feature has a positive impact on the Shapley value and as such, it is positively correlated with left tail dependence. The Shapley values are calculated using the TreeSHAP methodology of Lundberg et al. (2018) which is fine tuned to tree-based machine learning models such as random forests.



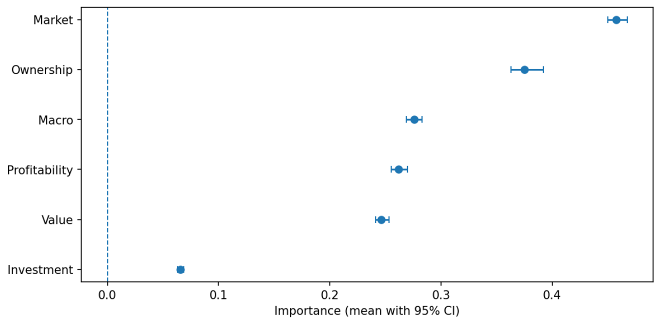

**(a)** Permutation test of LTD for variables




**(b)** Change in R2 of LTD for variables



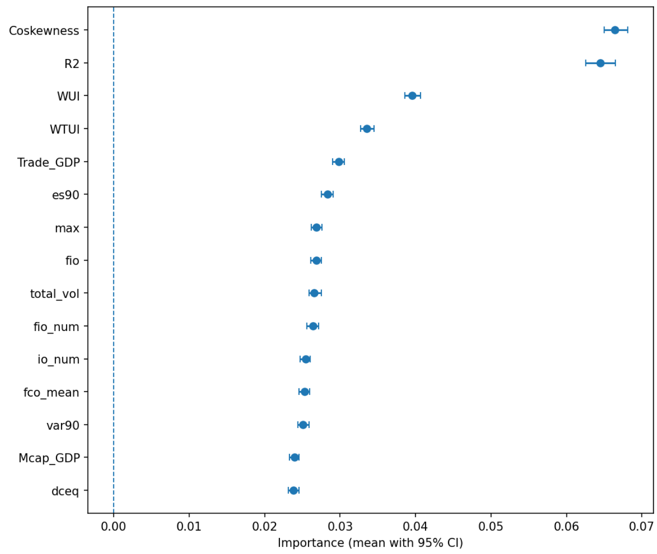

**(c)** Permutation test of LTD for groups



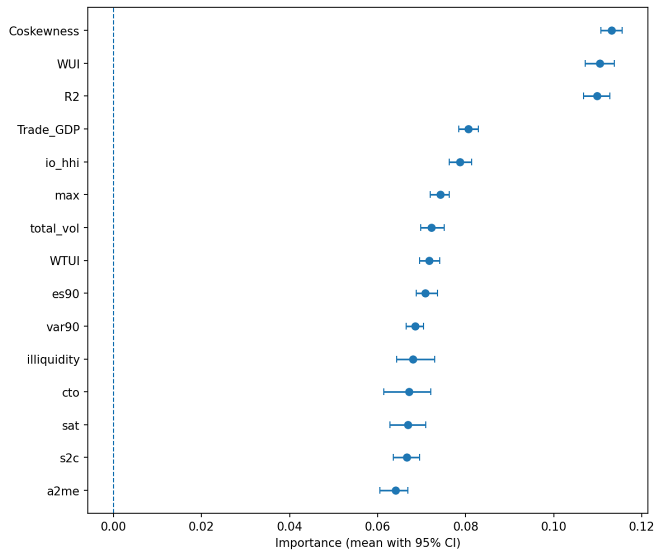

**(d)** Change in R2 of LTD for groups

**Figure 9.** Bootstrapped values of importance of determinants of firm-level left tail dependence for the US sample

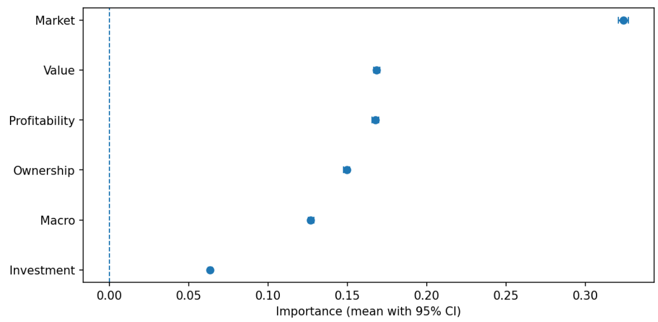
*Notes:* Figure shows the bootstrapped values of variable importance of individual variables and their groups. Specifically, We draw $N = 100$ bootstrap samples with replacement from the original US dataset, compute variable importance metrics for each, and summarize the resulting distributions by their mean and 95% percentile-based confidence intervals. Variable importance is calculated from both the permutation test and the change in R2 of the random forest regression model. The value of the permutation test for variable j is the average difference in prediction accuracy before and after permuting j in the data matrix. The sum of the permutation test scores of all variables is normalized to equal 1. The change in R2 corresponds to the reduction in predictive R2 from setting all values of variable j to zero, while holding the remaining model estimates fixed. The higher the permutation test score and the change of R2 are for variable j, the more important that variable is in explaining left tail dependence.



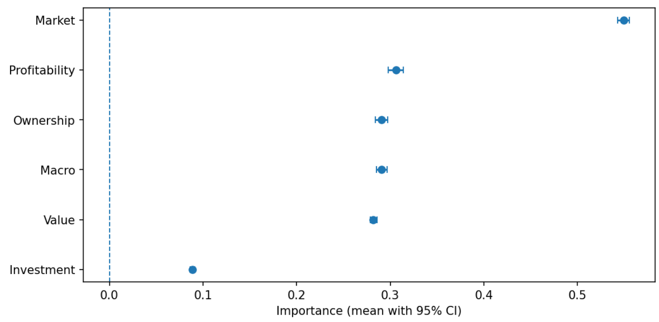

**(a)** Permutation test of LTD for variables




**(b)** Change in R2 of LTD for variables




**(c)** Permutation test of LTD for groups




**(d)** Change in R2 of LTD for groups

**Figure 10.** Bootstrapped values of importance of determinants of firm-level left tail dependence for the non-US sample

*Notes:* Figure shows the bootstrapped values of variable importance of individual variables and their groups. Specifically, We draw $N = 100$ bootstrap samples with replacement from the original non-US dataset, compute variable importance metrics for each, and summarize the resulting distributions by their mean and 95% percentile-based confidence intervals. Variable importance is calculated from both the permutation test and the change in R2 of the random forest regression model. The value of the permutation test for variable j is the average difference in prediction accuracy before and after permuting j in the data matrix. The sum of the permutation test scores of all variables is normalized to equal 1. The change in R2 corresponds to the reduction in predictive R2 from setting all values of variable j to zero, while holding the remaining model estimates fixed. The higher the permutation test score and the change of R2 are for variable j, the more important that variable is in explaining left tail dependence.