Firm-level Tail Dependence and its Determinants

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

[a] *Michael Smurfit Graduate Business School, University College Dublin, Dublin, Ireland*

Abstract

We measure the tail dependence of international companies with respect to foreign markets at the firm level using copulas. We observe that interdependence, in and outside US, increases in crises while left tail dependence is always stronger than right tail dependence with their difference widening in recessionary periods. We then characterize the factors that account for the total panel variation of firm-level tail dependence using the random forest regression framework. The World Uncertainty Index, the R-square integration measure and coskewness with respect to foreign markets are the most important determinants. Individual *Ownership* variables such as the number of total or foreign investors dominate the remaining firm-level characteristics in explaining tail dependence. When we categorize our variables into groups, we find that *Market*, *Ownership* and *Macro* (*Profitability*) variables matter the most in the US (non-US) sample.

Keywords: Firm-level tail dependence, Copulas, Determinants, Random forest regression

1. Introduction

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

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

[*] Corresponding author

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

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). This is especially true at the firm level where 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 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 extensive but it is concentrated on the banking and financial sector. For example, De Jonghe (2010) constructs the tail beta of a European bank with respect to a regional banking index and then studies its relationship with the type of banks' revenue generating activities and other characteristics such as size and loan to assets. Weiß et al. (2014) create an international dataset of banks and then assess which bank- or country-specific characteristics contribute more to the local or global systemic risk during financial crises. In a similar vein, Laeven et al. (2016) focus on the cross-section of banks' systemic risk during the Global Financial Crisis and find that risk increases across size and decreases across the bank's capital. Even when tail dependence for non-financial firms is measured, 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 focused on 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 of examining firm-level tail dependence in three 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 or the recent collapse of Sillicon Valley Bank and Credit Suisse of 2023) to real output. However, there is still a strong incentive for investors to measure how extreme negative market returns affect all firms, regardless of whether they are financial or not, in order to identify these sensitive stocks and minimize their portfolio tail 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 of the determinants of tail dependence between countries, industries and financial firms. These previous studies on the determinants employ standard regression techniques. In addition to regressions, we augment our analysis with a machine learning approach, namely random forest regression, to rank the determinants of left tail dependence of firms over time and 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.

Third, 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. 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 what we try to capture.

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 2000-2019 for both the US and non-US sample. 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. The corresponding foreign market index is the Fama-French Developed Market index excluding the US and the US CRSP value-weighted index for US and non-US stocks, respectively. 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.

In the second part of our analysis, we apply a random forest regression model on LTD and a representative dataset of firm- and country-specific characteristics in order to rank 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 choice to work with random forest regression is not random; we opt for it due to its ability to handle correlated variables and capture non-linear and interaction effects between our regressors. Our primary objective is to establish which determinants matter most in explaining firm-level tail dependence with the world and, for that purpose, 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 World Uncertainty Index of Ahir et al. (2022), the R-square measure of Pukthuanthong and Roll (2009) 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 R-square 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 R-square explains 6% and 8% of the variation of LTD with coskewness explaining 6% and 6.5% in the US and non-US sample. After the World Uncertainty Index, R-square 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 both US and non-US stocks. Their variable importance ranges between 2% and 3.5% in both samples. For comparison, size has a score of 3% or 2.5% 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.

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.

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 36%, 30% and 24%, respectively. Similarly, *Market*, *Ownership* and *Profitability* groups drive left tail dependence in the non-US sample with corresponding reductions in explanatory power of 40%, 24%, and 23%. The *Market* group that includes the R-square, coskewness and firm size is the primary driver of tail dependence followed by the *Ownership* group. The dominance of ownership related firm-level characteristics highlights the increasingly important role of institutions that trade internationally and affect prices. Only in the US, macroeconomic variables such as the World Uncertainty Index and the World Trade Uncertainty Index matter more than the other categories. However, the importance of the *Macro* group fades in the international sample of firms where *Profitability* characteristics are emphasized.

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 left-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 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.

The paper is organized as follows. Section 2 discusses the modelling of tail dependence. In Section 3 we motivate and describe the variables that we use in our analysis. Section 4 describes the random forest regression algorithm and the methods used to determine the most important drivers of tail dependence. Section 5.1 documents the characteristics of our measure of firm-level tail dependence and its correlation with other variables. Section 5.2 contains the main empirical findings where the determinants of tail dependence are presented. Finally we discuss several robustness checks in Section 6 and we conclude in Section 7.

2. 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 and 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

$$C(u, v; \Theta) = w_1 C_{Clayton}(u, v; \theta_1) + w_2 C_{Gaussian}(u, v; \theta_2) + w_3 C_{rClayton}(u, v; \theta_3) \quad (2.1)$$

where Θ 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 θ as a function of the uniformly distributed random variables u, v. The parameters θ_1 and θ_3 control the left and right tail dependence that Clayton copulas exhibit while θ_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:

$$LTD = \lim_{q \to 0^+} Pr\left(X < F_X^{-1}(q) \mid Y < F_Y^{-1}(q)\right) = \lim_{q \to 0^+} \frac{C(q,q)}{q} \tag{2.2}$$

$$UTD = \lim_{q \to 1^-} Pr\left(X > F_X^{-1}(q) \mid Y > F_Y^{-1}(q)\right) = \lim_{q \to 1^-} \frac{1 - 2q - C(q,q)}{1-q} \tag{2.3}$$

The LTD and UTD can be calculated explicitly for the basic parametric copulas, and thus, once the parameters Θ of equation 2.1 are known, LTD and UTD for the convex combination of Clayton-Gaussian-Rotated Clayton are given as:

$$LTD = w_1 2^{-1/\theta_1} \text{ and } UTD = w_3 2^{-1/\theta_3} \tag{2.4}$$

The LTD and UTD of equation 2.4 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 θ_1 and θ_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 equation 2.1 is a two-fold procedure. First the marginal distributions X, Y of the stock and foreign market index are estimated by their empirical counterparts:

$$\hat{F}_X(x) = \frac{1}{n+1} \sum_{k=1}^{n} I_X(k \leq x) \text{ and } \hat{F}_Y(y) = \frac{1}{n+1} \sum_{k=1}^{n} I_Y(k \leq y) \tag{2.5}$$

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 US and Canada are the largest tail risk drivers 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 and Canada 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 Equation 2.1. 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 $\Theta = [w_1, w_2, w_3, \theta_1, \theta_2, \theta_3]$:

$$\hat{\Theta} = \underset{\Theta}{argmax} \sum_{i=1}^{n} \log c(u_i, v_i; \Theta) \tag{2.6}$$

where $c(u_i, v_i; \Theta)$ is the corresponding copula density function of $C(u_i, v_i; \Theta)$ in equation 2.1. 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. Data

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 R-square 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, R-square, 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. 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 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 36 variables along with their categorization into groups can be found in Table 1.

[Insert Table 1 here]

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 55,744 firm-year observations in total with 3710 firms in June of year 2000 and 2137 firms in 2019. The final non-US sample includes 108,891 firm-years with 1029 international firms in June of year 2000 and 9147 firms in 2019.

4. 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.

Random forest regression offers several advantages over conventional linear regression models. Its main advantage is its ability to handle highly correlated variables as well as non-linear interactions between independent and dependent variables. Multicollinearity biases the coefficients and t-statistics of the corresponding linear models and thus the importance of variables can be masked. Akbari et al.

(2021) acknowledge this issue and adopt the RFR to uncover the drivers of financial and economic integration at the country level. Furthermore, RFR is based on a random sampling and averaging procedure which reduces the model's sensitivity to noise and outliers. Excluding part of the data and the explanatory variables when building each tree also corrects implicitly for the over-fitting problem. For those reasons, random forest regression is applied to our list of variables in order to find the determinants of firm-level left tail dependence across the world. The details of the RFR implementation are presented in Appendix A.

4.1. Variable importance

After we fit the random forest regression in the data, we rank variables using two different measures of importance. The first is the permutation test of Breiman (2001) in which we score variable j by the difference in prediction accuracy before and after permuting j. 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. This technique identifies the information content of each input determinant relative to all other determinants. For that reason, the sum of the permutation test scores of all variables is normalized to equal 1. The second 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.

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.

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 international sample in which the same crash probability of non-US stocks 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 values 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 with 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 θ is -15% (-15%), 43% (42%) and -14% (-10%) for the Clayton, Gaussian and Rotated Clayton in the US (outside of the US), 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. 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.

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% 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. On the contrary, the same pattern does not emerge for the 1st quintile portfolio of the lowest LTD stocks.

Finally, we report the correlation of left tail dependence with all other variables in our dataset in Table 3. The ranking order of the individual variables is indicative of the random forest regression results as we see below: i) the R-square integration measure (R2) and Coskewness have the highest absolute correlation with left tail dependence and ii) *Ownership* variables dominate all others. Interestingly, firm size (log_me) is more correlated (31%) with LTD in the US sample whereas this positive relationship is weakened (log_cap has a correlation of 15%) in the international sample of stocks. 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 true variables that explain firm-level left tail dependence of US and non-US stocks.[1] 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

[1]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.

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.

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 World Uncertainty Index (WUI) of Ahir et al. (2022) is the most important determinant of LTD and it explains 6% of its panel variation. 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. After WUI, Coskewness and R2 matter the most in explaining the variation of LTD with both of them contributing slightly less than 6%. The R-Square measure of Pukthuanthong and Roll (2009) 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. 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 (negative), then LTD (UTD) is weak.

The most interesting pattern that we observe in our analysis is the importance of *Ownership* variables such as the number of foreign and total institutional investors (fio_num and io_num). They rank at the 5th 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%, 3% and 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 World Uncertainty Index, R-square measure, coskewness and number of total and foreign investors. Its permutation test score is only 3.2%. Other determinant variables exert a much lesser influence on left tail dependence.

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 36%, 30% and 24%, respectively. *Market* variables that include the R-square, 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 complement our random forest regression analysis with linear regressions using the top 15 most important variables as ranked by the permutation test score. The left panel of Table 4 reports the OLS coefficients for the US sample. The signs are intuitive. More specifically, WUI and R2 have a positive and significant effect on LTD: a 1% increase of WUI and R2 increases left tail dependence by 1.5% and 0.23%, respectively. Coskewness, on the other hand, affects LTD negatively since, by definition, a positive value of coskewness implies a weak left tail dependence and this fact is reflected in the negative coefficient. All the *Ownership* variables (io_num, fio, fco_mean, io) are positively correlated with LTD but the magnitude of their coefficients is not large. As far as the role of firm size (log_me) is concerned, a 1% annual return is related to a 0.5% increase in the probability that the stock will decline in price if foreign markets also experience a decline in the far left tail.

5.2.2. non-US results

Second, we present results for the non-US sample. Figure 5 shows the importance of individual variables and their groups on LTD of non-US stocks. Unlike in the case of the US, there are differences in the ranking of variables and groups depending on whether the permutation test or the change in R2 is used. However, despite these differences, there are some common patterns: i) the R-Square integration measure, Coskewness, World Uncertainty Index and World Trade Uncertainty Index are the top 4 most important variables and ii) *Ownership* variables are dominating places 5 to 9 based on the permutation test with their dominance weakening according to change in R2 values. Interestingly, left tail dependence does not depend at all on firm size (log_cap) in the global dataset in contrast to the US sample where size (log_me) ranks very high. The remaining determinant variables have little impact on LTD.

When we focus on groups, we find that *Market* variables with the inclusion of the R-square and Coskewness are the most important drivers of left tail dependence in the international sample. The difference of the value of the permutation test score for *Profitability*, *Value* and *Ownership* groups is very small (17%, 16.8% and 16% respectively) and thus we use the change in R2 results for inference. In the latter case, *Ownership* variables clearly dominate *Profitability* and *Value* variables as it was the case with US firms. Thus, *Market*, *Ownership* and *Profitability* variables drive left tail dependence outside of the US with corresponding reductions in the explanatory power of the model of 40%, 24%, and 23%. In contrast to the US sample, *Macro* variables are unimportant as a category in the non-US universe. Even though WUI and WTUI matter, the country-specific total stock market capitalization and trade over GDP variables exert no effect on the LTD of international firms with respect to the US market.

The right panel of Table 4 reports the OLS coefficients of a regression of LTD on the 15 most important variables selected by RFR using the permutation test score for non-US stocks. The OLS results resemble those of the US: WUI and R2 have a positive and significant effect on LTD with coefficients 0.7% and 0.5%, respectively. The relation between Coskewness and LTD is weakened in the international sample with a -15% coefficient compared to -20% in the US. The same weak but positive relation pattern holds for the *Ownership* variables (fio, fco_mean, io_num, io, io_hhi). Firm size (log_cap) still exerts a positive but weaker effect on LTD for international firms: a 1% annual return is translated to a 0.1% increase on left tail dependence in contrast to the 0.5% increase in the US.

6. Robustness checks

We discuss several robustness checks. Our results are robust when we control for small-cap stocks, financial firms or the random seed of the random forest regression model.

6.1. The effect of micro-cap stocks

The higher value-weighted average levels of left tail dependence of Figure 1 suggest that larger firms are more exposed to foreign shocks than the average firm represented by the equal-weighted mean of LTD. This is supportive of our results being linked with firm size. Even though micro-cap stocks comprise the majority of the equity universe, their economic significance is trivial. In this section, we explore their effect on our analysis by excluding them. We follow the definition of the most recent papers on asset pricing such as those of Hou et al. (2020) and Jensen et al. (2022) 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.

Figure 6 shows the random forest regression results for the US and non-US stock universe when we exclude small cap stocks from our analysis. The findings of Section 5.2 remain unchanged and thus our results are not driven by firm size.

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). Figure 7 plots 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. Effect of the random seed in random forest regression

Every time we build a Tree for our random forest, we use randomly only 2/3 of the full sample. This randomness is controlled by the state of the random seed and in this section we assess its effect on the random forest regression algorithm results. For the baseline results of Section 5.2, we set the random state to 1. Then we proceed to set the random state to 3, 5, 7 and 11 and run the RFR model again. For brevity we report results only for the random seed=3 case in Figure 8 but results remain largely unchanged for the other cases and are available upon request.

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 2000-2019. 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 R-square 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. The importance of *Ownership* is observed inside and outside of US and it 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*, *Ownership* and *Macro* (*Profitability*) groups are the largest drivers of left tail dependence in the US (non-US) sample.

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.

References

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. (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.

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

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.

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

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.

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.

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. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

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

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

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

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

Jensen, T. I., Kelly, B. T., and Pedersen, L. H. (2022). Is there a replication crisis in finance? *The Journal of Finance, Forthcoming.*

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

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.

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

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.

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.

Appendices

A. Random forest regression algorithm

We use the Pedregosa et al. (2011) package to run random forest regressions. The RFR algorithm is described below:[2]

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 36.

   (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.

[2] The notation of the RandomForestRegressor class of the scikit-learn package is used.

(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 Θ^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).

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 generic and as such it is applicable to both GETS and RFR models. It 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 generic and applicable to both GETS and RFR models and it is used as a complementary measure to the overall contribution and permutation test.

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	Pukthuanthong and Roll (2009)'s integration measure	21	roc	Size + longterm 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			
15	io_hhi	Herfindahl–Hirschman index of io		**Macro**	
16	io_num	Number of total institutional owners	33	WUI	World uncertainty index of Ahir et al. (2022)
			34	WTUI	World trade uncertainty index of Ahir et al. (2022)
	Investment		35	Mcap_GDP	Total market capitalization of local market over GDP
17	inv	Percentage change in total assets	36	Trade_GDP	The sum of exports and imports over GDP
18	dceq	Percentage change in book equity			
19	ivc	Change in inventory over average total assets			

Notes: The table lists the variables we consider as potential determinants of firm-level tail dependence in our analysis by category. 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%
2000	3710	4.55	5.52	0.00	2.22	7.83	2.61	3.79	0.00	0.30	4.43
2001	3636	4.65	5.75	0.00	2.15	7.84	5.27	6.82	0.00	2.28	8.84
2002	3518	4.34	5.17	0.00	2.31	7.58	5.87	6.37	0.03	4.16	9.47
2003	3340	9.13	7.96	1.29	8.05	14.77	5.51	6.31	0.00	3.47	9.42
2004	3276	6.80	7.09	0.18	4.81	11.21	3.20	4.65	0.00	0.02	5.80
2005	3213	2.97	4.31	0.00	0.18	5.19	3.07	4.50	0.00	0.21	5.24
2006	3157	3.56	5.29	0.00	0.15	6.01	6.69	6.74	0.00	5.18	11.49
2007	2985	9.79	8.02	1.91	9.31	15.63	4.87	5.99	0.00	2.03	8.64
2008	2889	6.73	7.15	0.01	4.95	11.13	5.48	6.07	0.00	3.58	9.59
2009	2652	17.13	10.36	8.72	17.76	25.17	11.05	8.11	4.13	10.77	17.15
2010	2660	14.01	10.88	4.70	12.76	22.01	5.35	6.64	0.00	2.44	9.29
2011	2555	8.20	8.09	0.17	6.56	13.49	8.73	8.92	0.00	6.55	14.65
2012	2482	15.95	11.77	6.27	14.21	24.36	8.08	9.26	0.00	4.85	13.49
2013	2400	7.12	7.66	0.01	4.81	11.77	7.99	8.05	0.00	6.13	13.30
2014	2330	6.12	6.63	0.00	4.01	10.48	5.18	6.20	0.00	2.77	8.89
2015	2268	8.79	8.05	0.97	7.66	14.25	5.32	6.09	0.00	3.12	9.68
2016	2303	12.94	10.67	3.47	11.45	20.67	4.92	5.63	0.00	2.78	8.77
2017	2103	5.46	6.00	0.00	3.78	9.51	5.86	6.29	0.01	3.99	9.97
2018	2130	10.82	9.20	1.85	9.75	17.26	2.80	4.01	0.00	0.24	5.00
2019	2137	8.23	7.81	0.56	6.64	13.24	7.78	7.46	0.14	6.36	12.86

non-US sample

		LTD					UTD				
Date	Firms	Mean	St.Dev.	25%	Median	75%	Mean	St.Dev.	25%	Median	75%
2000	1029	3.13	4.86	0.00	0.32	4.89	2.79	4.24	0.00	0.22	4.45
2001	2099	4.22	5.95	0.00	1.17	6.87	3.51	5.21	0.00	0.20	5.89
2002	2425	4.34	4.90	0.00	2.71	7.30	3.44	5.26	0.00	0.43	5.48
2003	2534	4.66	5.79	0.00	2.21	7.79	4.30	5.78	0.00	1.76	6.87
2004	3240	5.71	6.04	0.62	4.07	8.95	1.83	3.35	0.00	0.00	2.56
2005	3124	2.90	4.23	0.00	0.08	5.07	2.23	3.86	0.00	0.00	3.42
2006	4219	2.82	4.46	0.00	0.24	4.31	2.34	3.97	0.00	0.00	3.48
2007	5149	5.72	6.16	0.02	4.14	9.47	2.93	4.37	0.00	0.03	5.03
2008	5787	3.46	5.55	0.00	0.12	5.44	2.73	4.32	0.00	0.01	4.50
2009	5348	8.78	9.45	0.32	6.06	13.90	5.43	5.94	0.00	3.81	9.02
2010	5848	6.75	7.51	0.18	4.38	10.84	3.55	5.75	0.00	0.00	5.39
2011	6099	3.84	5.70	0.00	0.55	6.17	4.23	6.46	0.00	0.50	6.80
2012	6473	9.96	9.49	2.20	7.64	14.74	5.19	7.14	0.00	1.66	8.45
2013	6687	4.72	5.49	0.00	2.74	8.00	2.57	4.73	0.00	0.00	3.61
2014	7364	2.85	4.62	0.00	0.03	4.41	2.85	4.17	0.00	0.24	4.84
2015	7771	4.60	5.50	0.00	2.83	7.43	2.90	4.82	0.00	0.00	4.54
2016	7856	9.93	7.25	4.44	9.24	14.24	4.70	5.27	0.00	2.98	7.77
2017	8299	2.88	4.56	0.00	0.08	4.61	2.41	3.75	0.00	0.02	3.99
2018	8393	7.25	6.77	0.66	5.91	12.00	3.00	4.54	0.00	0.02	5.16
2019	9147	5.53	6.37	0.00	3.27	9.56	5.88	6.26	0.00	4.37	10.04

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

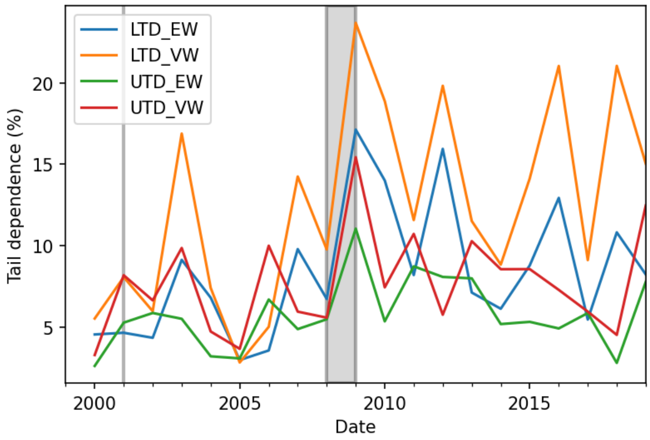
US sample				**non-US sample**			
Variable	Correlation	Variable	Correlation	Variable	Correlation	Variable	Correlation
R2	0.46	illiquidity	-0.04	R2	0.47	btm	-0.02
Coskewness	-0.35	Mom6m	0.04	Coskewness	-0.37	cto	-0.02
log_me	0.31	dSo	-0.04	io_num	0.22	sat	-0.01
fco_mean	0.26	e2p	0.03	UTD	0.22	Mcap_GDP	0.01
io	0.25	c2d	0.03	fio_num	0.22	roe	0.01
io_num	0.25	debt2p	-0.02	fio	0.21	ivc	-0.01
WUI	0.25	io_hhi	0.02	fco_mean	0.18	debt2p	0.01
UTD	0.24	sat	-0.02	io	0.18	inv	-0.01
fio_num	0.24	cto	-0.02	log_cap	0.15	c2d	0.01
fio	0.20	roe	0.02	WUI	0.14	WTUI	0.01
Trade_GDP	0.13	WTUI	0.01	var90	0.10	e2p	0.01
be	0.12	c	-0.01	be	0.09	roc	0.01
operpro	0.11	s2c	-0.01	es90	0.09	sales_g	0.00
Mcap_GDP	-0.10	op	0.01	operpro	0.08	op	0.00
roic	0.10	roc	-0.01	Trade_GDP	0.07	s2c	0.00
skewness	-0.08	ivc	0.00	c	-0.06	illiquidity	0.00
max	-0.08	ipm	0.00	total_vol	0.06	dSo	0.00
total_vol	-0.07	prof	0.00	roic	0.04	prof	0.00
btm	-0.07	dceq	0.00	Mom6m	0.04	dceq	0.00
es90	-0.06	inv	0.00	io_hhi	0.04	a2me	0.00
var90	-0.05	sales_g	0.00	skewness	-0.03	ipm	0.00
a2me	-0.04			max	0.02		

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.

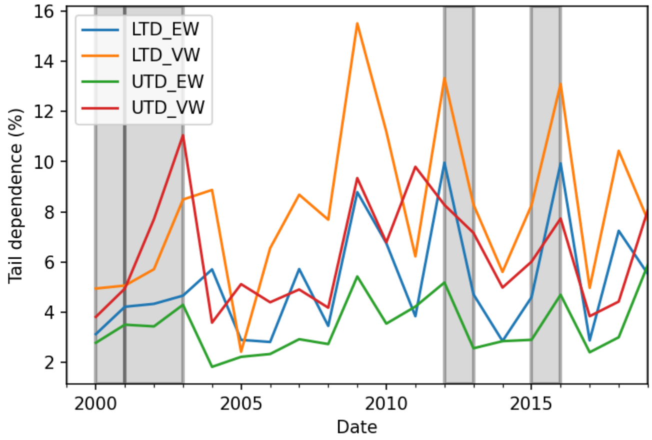
Table 4. OLS regression coefficients of left tail dependence

US sample			non-US sample		
Variable	Coefficient	Permutation test	Variable	Coefficient	Permutation test
WUI	1.481***	6.032	R2	0.545***	5.341
Coskewness	-20.378***	6.021	Coskewness	-14.754***	4.420
R2	0.234***	5.870	WUI	0.683***	2.825
Mcap_GDP	-0.020***	4.338	WTUI	-7.722***	2.100
io_num	0.001***	3.456	fio	0.026***	1.970
log_me	0.461***	3.299	fco_mean	0.03	1.944
WTUI	-36.919***	3.224	io_num	0	1.862
Trade_GDP	-0.018**	3.178	io	0.003	1.812
illiquidity	-0.033*	3.161	max	1.011	1.739
fio	0.048***	3.106	roe	0.000***	1.734
fco_mean	0.208***	3.021	debt2p	0	1.667
io	0.012***	2.703	io_hhi	-1.813**	1.644
es90	0.428***	2.468	log_cap	0.105***	1.614
var90	-0.237***	2.429	total_vol	0.023***	1.610
Obs	55,744		Obs	108,891	
r-squared	0.651		r-squared	0.594	

Notes: The table reports the OLS coefficients of the left tail dependence coefficient of the Clayton-Gaussian-Rotated Clayton copula between a stock and its corresponding foreign market index against the 15 most important variables as selected by random forest regression with the permutation test score. The copulas are estimated with daily returns from July of year t-1 to June of year t. fio_num is dropped from the OLS regressions due to its very high correlation of 97% with io_num. Errors are robust and *, ** and *** correspond to significance at the 10%, 5% and 1% level, respectively.



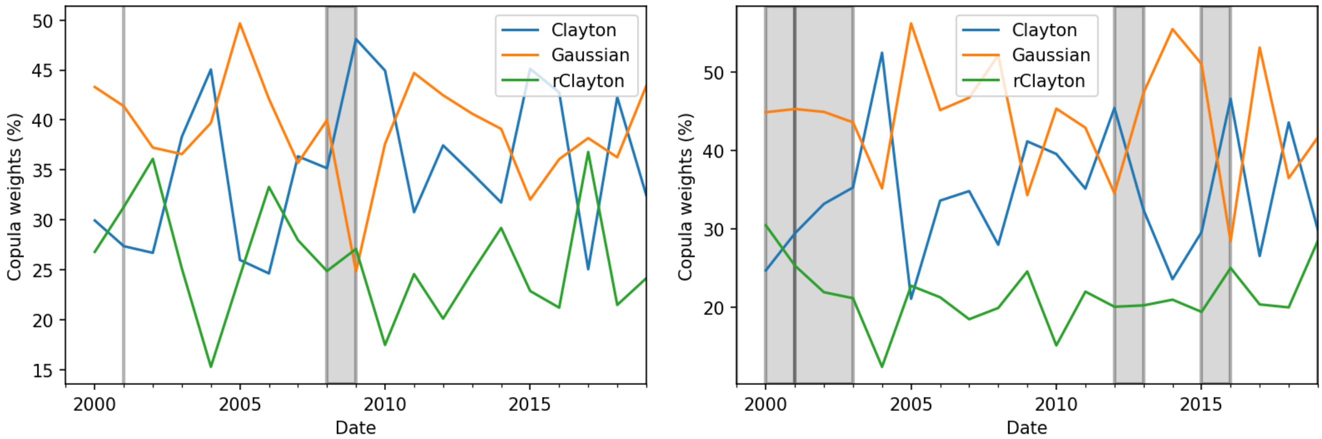
(a) EW and VW average LTD and UTD for US stocks



(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.

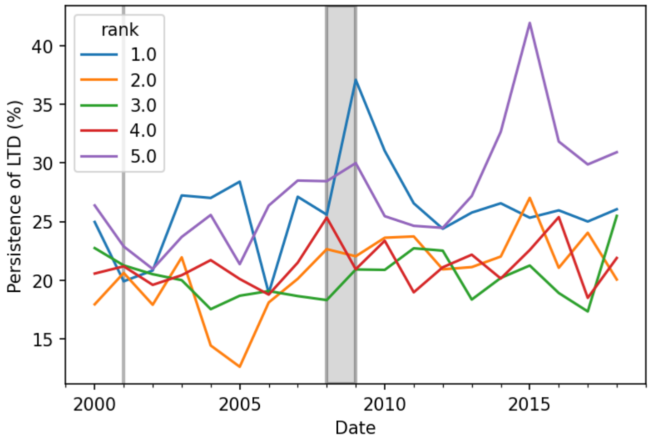


(a) EW average copula weights for US stocks

(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.

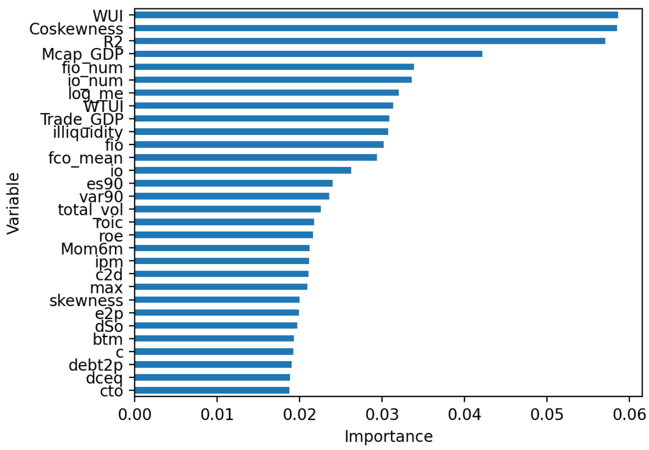


(a) Persistence of LTD for US stocks

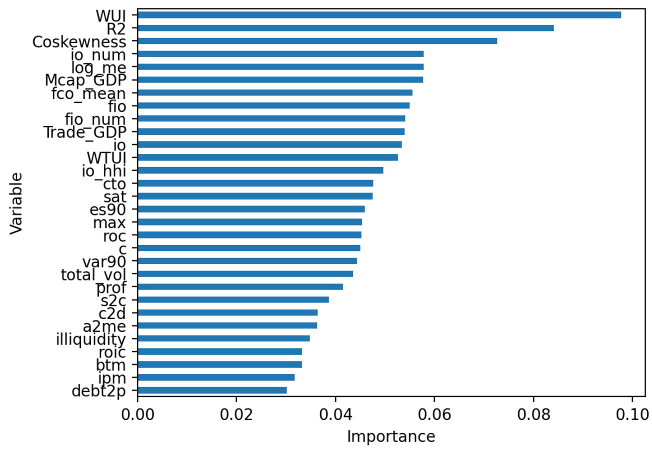
(b) Persistence of LTD for non-US stocks

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

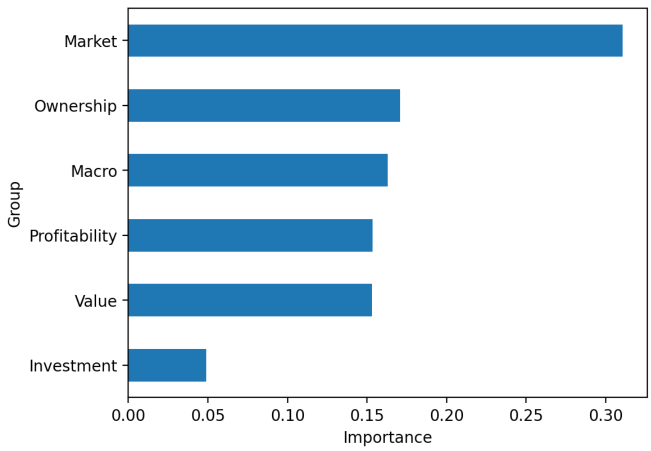
Notes: Figures 3a and 3b plot the persistence of LTD for US and non-US stocks. Persistence is measured as the relative frequency at which a stock is sorted into a LTD quintile portfolio i in year t given that it was in same portfolio i in year t-1. The rank 1 portfolio contains the 20% stocks with the 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.



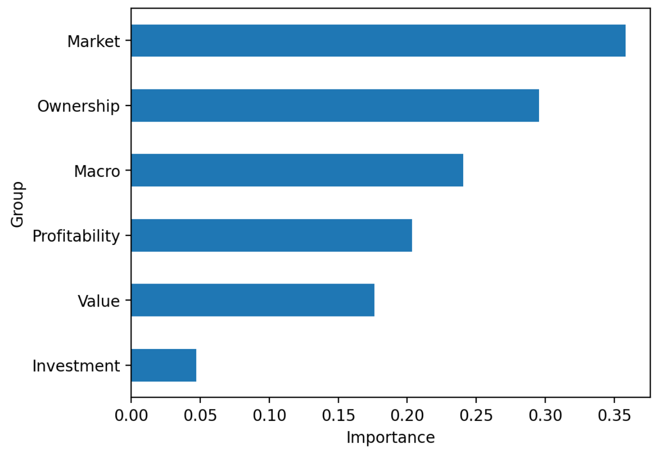
(a) Permutation test of LTD for variables



(b) Change in R2 of LTD for variables



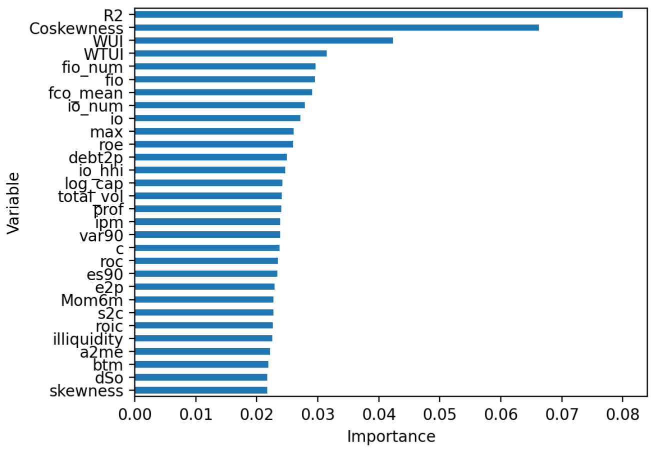
(c) Permutation test of LTD for groups



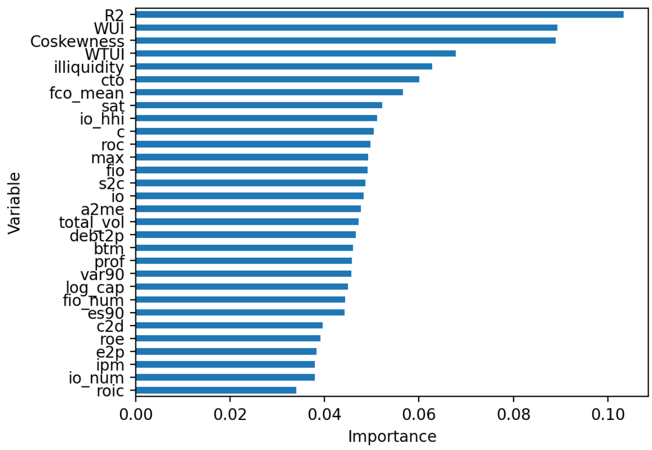
(d) Change in R2 of LTD for groups

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

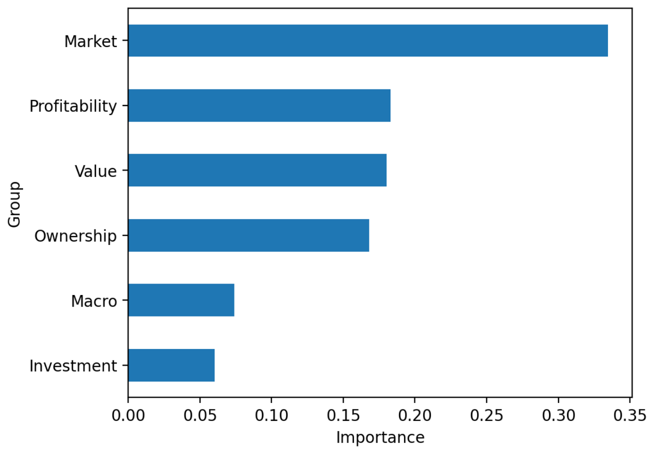
Notes: Figure shows the importance of individual variables and their groups as determinants of firm-level left tail dependence of US stocks with their corresponding foreign market index. Variable importance is calculated from both the permutation test and the change in R2 of the random forest regression 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.



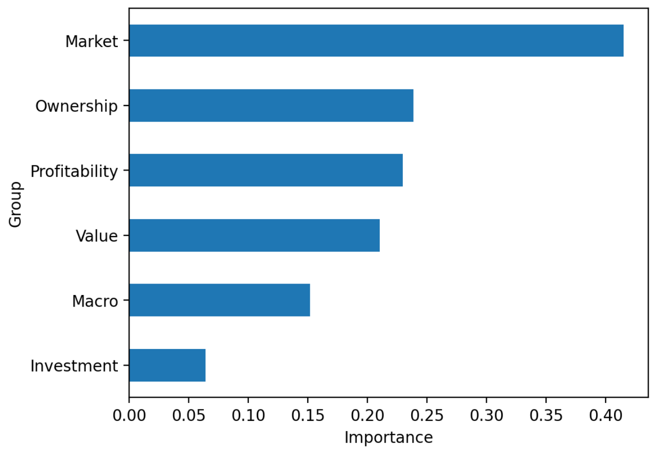
(a) Permutation test of LTD for variables



(b) Change in R2 of LTD for variables



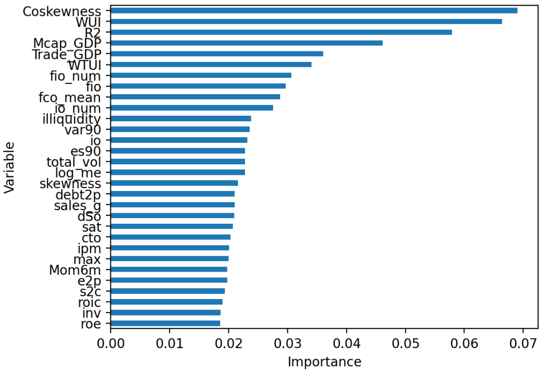
(c) Permutation test of LTD for groups



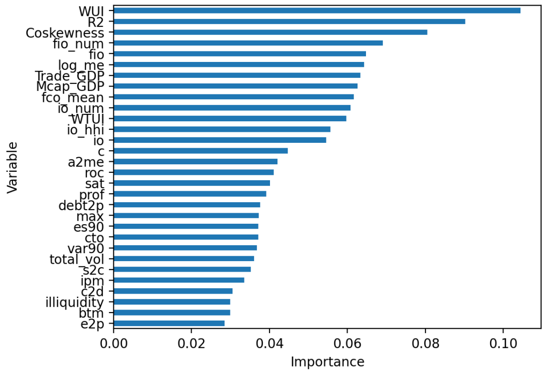
(d) Change in R2 of LTD for groups

Figure 5. Importance of determinants of firm-level left tail dependence for the non-US sample

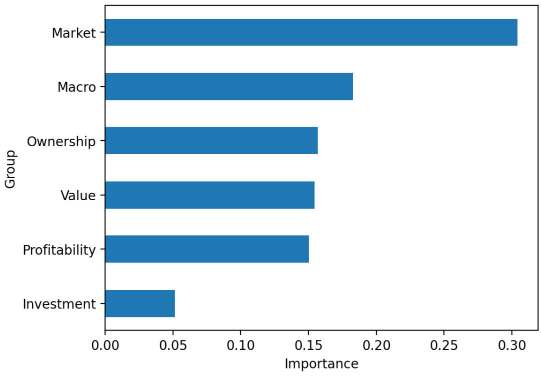
Notes: Figure shows the importance of individual variables and their groups as determinants of firm-level left tail dependence of non-US stocks with their corresponding foreign market index. Variable importance is calculated from both the permutation test and the change in R2 of the random forest regression 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.



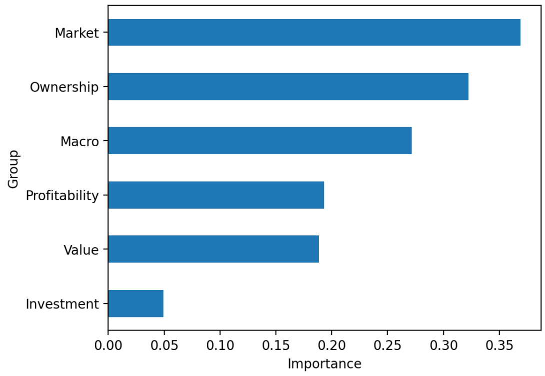
(a) Permutation test of LTD for variables in US



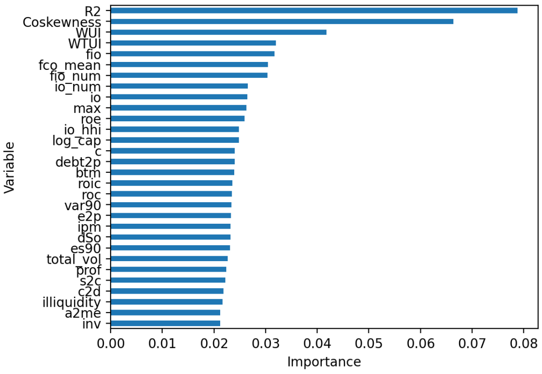
(b) Change in R2 of LTD for variables in US



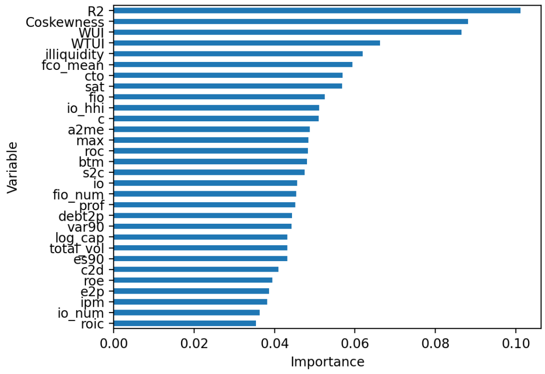
(c) Permutation test of LTD for groups in US



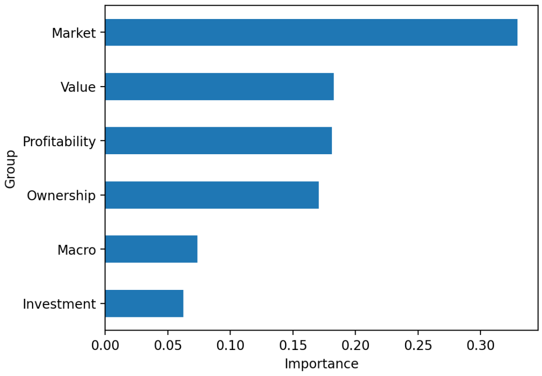
(d) Change in R2 of LTD for groups in US



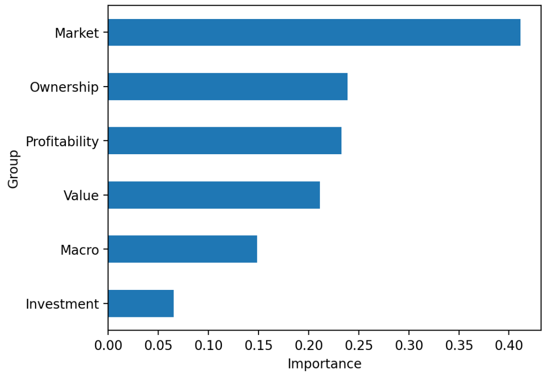
(e) Permutation test of LTD for variables in non-US



(f) Change in R2 of LTD for variables in non-US



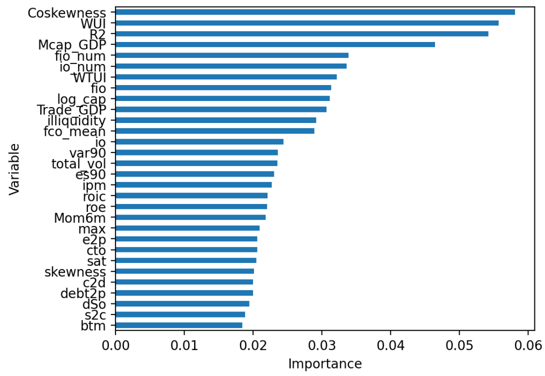
(g) Permutation test of LTD for groups in non-US



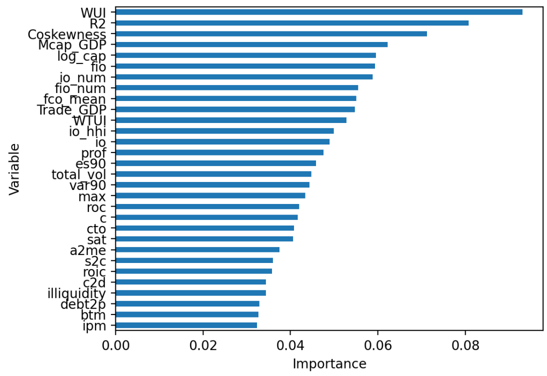
(h) Change in R2 of LTD for groups in non-US

Figure 6. Importance of determinants of firm-level left tail dependence when we exclude small-cap stocks

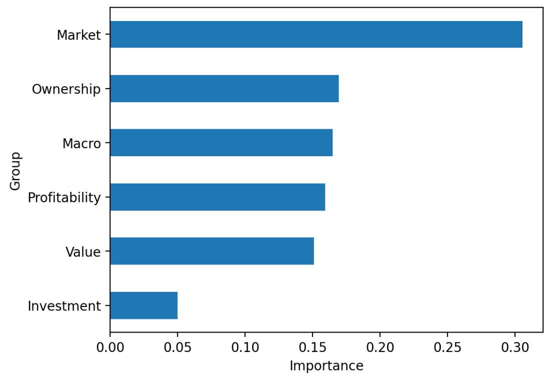
Notes: Figure shows the importance of individual variables and their groups as determinants of firm-level left tail dependence of US and non-US stocks with their corresponding foreign market index when we exclude small-cap stocks. Variable importance is calculated from both the permutation test and the change in R2 of the random forest regression model.



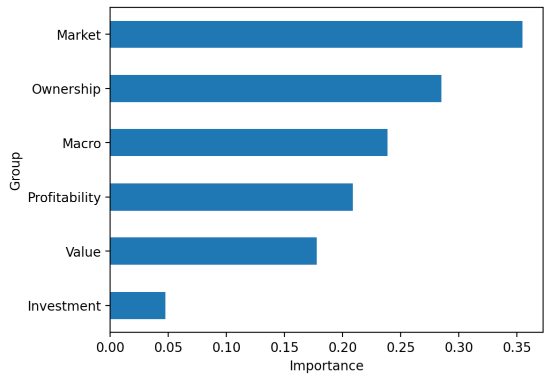
(a) Permutation test of LTD for variables in US



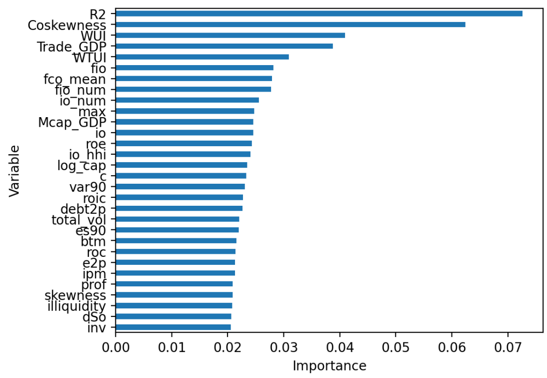
(b) Change in R2 of LTD for variables in US



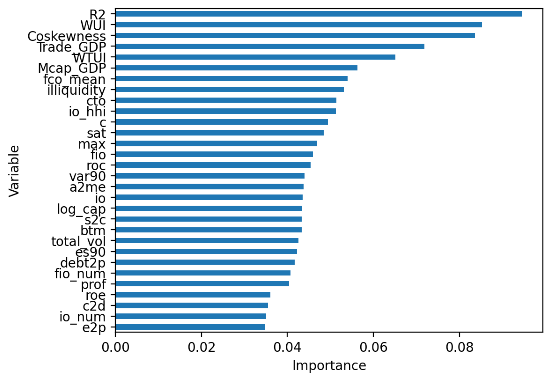
(c) Permutation test of LTD for groups in US



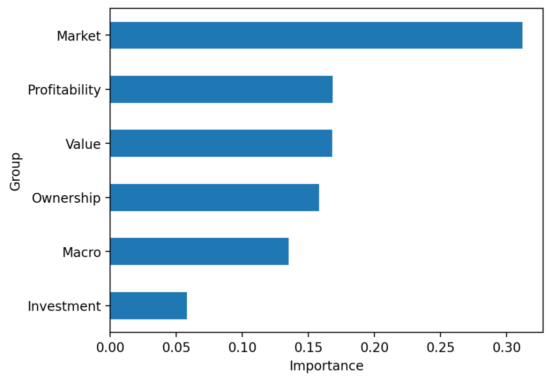
(d) Change in R2 of LTD for groups in US



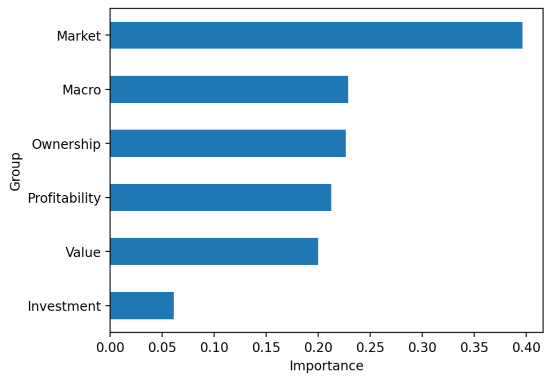
(e) Permutation test of LTD for variables in non-US



(f) Change in R2 of LTD for variables in non-US



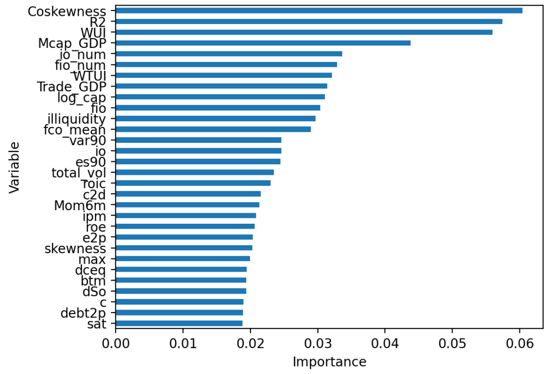
(g) Permutation test of LTD for groups in non-US



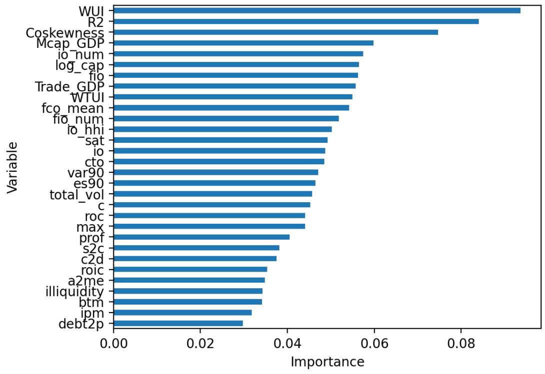
(h) Change in R2 of LTD for groups non-US

Figure 7. Importance of determinants of firm-level left tail dependence when we exclude financial firms

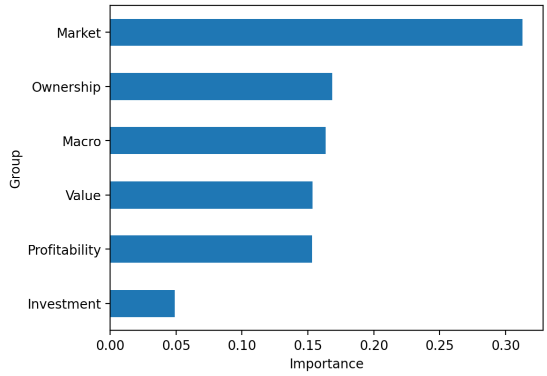
Notes: Figure shows the importance of individual variables and their groups as determinants of firm-level left tail dependence of US and non-US stocks with their corresponding foreign market index when we exclude financial firms. Variable importance is calculated from both the permutation test and the change in R2 of the random forest regression model.



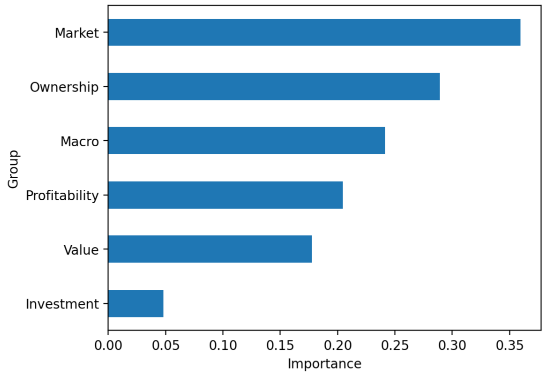
(a) Permutation test of LTD for variables in US



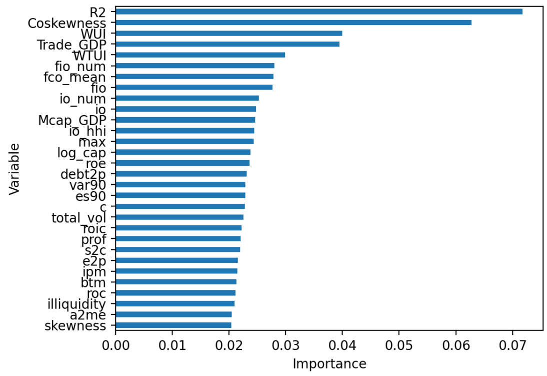
(b) Change in R2 of LTD for variables in US



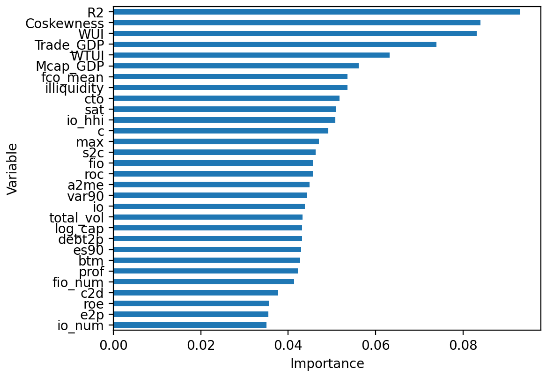
(c) Permutation test of LTD for groups in US



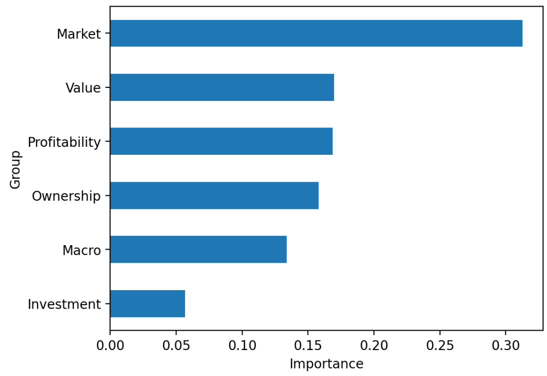
(d) Change in R2 of LTD for groups in US



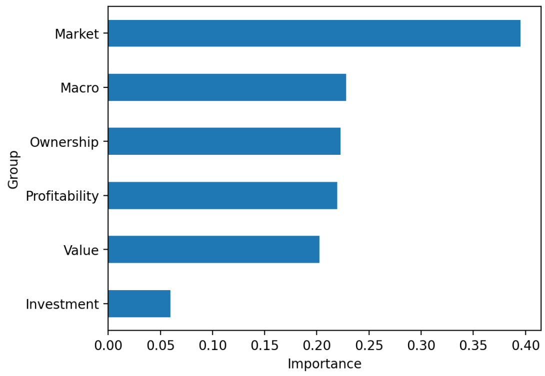
(e) Permutation test of LTD for variables in non-US



(f) Change in R2 of LTD for variables in non-US



(g) Permutation test of LTD for groups in non-US



(h) Change in R2 of LTD for groups in non-US

Figure 8. Importance of determinants of firm-level left tail dependence for random seed=3

Notes: Figure shows the importance of individual variables and their groups as determinants of firm-level left tail dependence of US and non-US stocks with their corresponding foreign market index when we set the random seed of the random forest regression algorithm to 3. Variable importance is calculated from both the permutation test and the change in R2 of the random forest regression model.