Trends and Key Determinants of Firm-level Integration

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

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

Abstract

We measure market integration at a firm-level for all US companies with the rest of the world. While we observe that integration increased through the years for the US as a whole, there are differences across firms according to their characteristics. Past research indicates that large firms, significant exporters and firms held primarily by institutional investors are more integrated. However, not all characteristics affect integration to the same degree. As such, we characterize the key factors that account for most of the total panel variation of firm-level integration. The corporate spread between BAA and AAA bond indices is the most important variable that determines the level of integration of a stock followed by size, institutional ownership and foreign sales. When we categorize our variables into groups, we find that *Macro*, *Market* and *Ownership* variables matter the most. In general, *Macro* variables are the primary drivers of US integration levels and have an effect that is larger than any firm characteristic.

Keywords: Firm-level integration, Determinants, General-to-specific modelling, Random forest regression
JEL: F30, G15

1. Introduction

Market integration across countries has been studied extensively over the past decades. The literature has established that the globalization trend at the end of

[⋆]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)

the 20th and start of the 21st century led to increased financial integration for both Developed (DM) and Emerging Markets (EM) (see for example Pukthuanthong and Roll, 2009 and Christoffersen et al., 2014). Furthermore, Bekaert et al. (2011) and Akbari et al. (2021) characterize the factors that drive country-level integration. However, integration at the firm-level has been overlooked with very few exceptions. Thus a fundamental question arises: how does integration vary across firms within the same country? Further, if there are cross-sectional differences, which determinants explain them?

The vast majority of papers in the related literature focus on the trend and determinants of integration at the country rather than at the firm level. There are two distinct strands; one that studies the trend of integration across markets and time (Longin and Solnik, 1995; Goetzmann et al., 2001; Obstfeld and Taylor, 2003; Quinn and Voth, 2008; Christoffersen et al., 2014; Bartram and Wang, 2015; Rangvid et al., 2016) and another that characterizes the drivers of that trend (Bracker et al., 1999; Pretorius, 2002; Forbes and Chinn, 2004; Chambet and Gibson, 2008; Bekaert et al., 2011; Bartram and Wang, 2015; Akbari et al., 2021). These studies have established that financial and economic integration of both Developed and Emerging markets (DMs and EMs, respectively) increased over the past few decades while international trade and equity market openness are two of the biggest drivers of country-level integration.

In contrast to the above literature, very few papers explicitly study integration at the firm-level and those that do, focus only on one dimension of the problem at a time. For example, both Brooks and Negro (2006) and Di Giovanni et al. (2018) examine the positive effect of the international activity of a firm on its integration with the world, as proxied by foreign sales and assets or the establishment of foreign subsidiaries. Specifically, Brooks and Negro (2006) estimate a latent factor model with global, country and industry factors for samples of stocks that are sorted based on the level of internationalization of the firm proxied by foreign sales, assets and income ratio. They find that stocks belonging to the top quartile comove more with global factors than those that belong to the bottom one. On the other hand Di Giovanni et al. (2018) use a proprietary dataset of both private and public French firms and establish that at the micro level, the correlation between a firm and a foreign country is positively related to trade and multinational linkages in terms of subsidiaries with the foreign country. Both studies offer strong evidence of the firm-level relation between integration and internationalization.

Separately, size has been found to have a positive effect on the integration of a firm with international markets. For example, Huang (2007) use the single-factor conditional asset pricing model of Harvey (1991) to test whether country large-, mid- and small-cap portfolios are priced globally and find that only the returns of large-cap stocks are related to the world price of covariance risk. This means that financial integration between markets is due to large cap stocks. In a similar vein, Eun et al. (2008) find that the correlation of small-cap stocks with other small-caps or large-caps is much lower than the correlation between large-caps. A size effect is also documented by Di Giovanni et al. (2017) who show that the top 100 largest firms contribute substantially to the business cycle comovement of France and its foreign partners.

In addition, institutional ownership as established by Faias and Ferreira (2017) and Anton and Polk (2014), also plays a crucial role on increasing international capital market integration. Faias and Ferreira (2017) sort stocks into low and high institutional ownership (IO) groups and fit a factor model with global, country and industry effects within each IO group. They show that the relative importance of global-to-country effects is higher in the high IO group where the marginal investor is most likely to be an institution. The higher global component of returns of high IO stocks is a testament of increased cross-country correlation and the power of institutions to contribute to the convergence of prices. Anton and Polk (2014) study the comovement of commonly owned stocks by constructing a foreign ownership portfolio that takes into account these common ownership linkages. Commonly owned stocks constitute an "investment habitat" and the foreign ownership portfolio behaves like a common factor for these stocks emphasizing the role of institutions on comovement. While the literature suggests that size, importing and exporting activity, as well as institutional ownership, dictate the relationship of a firm with international markets, no study has jointly assessed and disentangled the importance of those firm characteristics in explaining firm-level integration.

Identifying the characteristics of the stocks that are more integrated with foreign markets is of fundamental importance for global investors. Highly integrated stocks comove more with international markets and thus their benefits in terms of diversification will be less than that of the typical stock (see for example Conlon et al., 2024). As detailed, past research indicates that large firms and those that are significant exporters held mostly by institutions are more integrated with the world and, as such, these firms constitute poor global diversification choices. However, one should not expect that these characteristics exert the same influence on firm-level integration and it remains to be seen which matter more.

Our study makes an important contribution to the existing literature on market integration. It represents the first to explicitly investigate the financial integration of firms with respect to foreign markets along with the underlying country- and firm-specific characteristics that can explain its variation. Thus, our research strongly contrasts with prior literature that mostly focuses on the determinants of market-level integration only (Bekaert et al., 2011; Akbari et al., 2021) as we explore the role of both firm- and country-level variables in fitting the measured firm integration. Our primary objective is to establish which factors matter most; for instance, are macroeconomic variables which are country-specific more important in explaining the panel variation of firm-level integration or are firm characteristics equally important? Our framework of analysis provides insights on this question that are robust to model selection techniques.

First, we measure the integration of a firm with respect to foreign markets annually using the R-squared methodology of Pukthuanthong and Roll (2009) for all US domiciled and public traded stocks for the period 1999-2019. Specifically, we regress the daily returns of a firm in a year against the first 5 out-of-sample principal components of 10 international market indices excluding the US. The principal components proxy the foreign factors and the R-squared of that regression is our measure of integration.[1] By definition, the R-squared captures the percentage of the variation of a stock's returns in a year that is explained by the variation of the global factors' returns. The higher the adjusted R-squared is for a stock, the more integrated it is with the global economy.

[1] We also use simple correlations as an alternative to the R-squared measure with results being consistent. The correlation is computed between daily stock returns and the Fama-French world market factor excluding the US.

In the second part of our analysis, we apply the general-to-specific (GETS) algorithm of Bekaert et al. (2011) using the stock-level R-squared measure and a comprehensive dataset that includes 43 firm characteristics and 6 macroeconomic variables in order to determine the importance of each variable in explaining the panel variation of firm-level integration. With machine learning having many benefits and being widely used recently in finance (Akbari et al., 2021; Avramov et al., 2021; Olson et al., 2021), we also use random forest regression as an alternative variable selection technique to compliment GETS with consistent results. Our dataset includes strong candidates as determinants of firm-level integration such as size, foreign sales and foreign sales ratio as proxies for the exporting activity of the firm as well as total and foreign institutional ownership, among others. Macroeconomic variables such as the corporate spread between BAA and AAA bonds (Bekaert et al., 2011) and the dollar amount of US exports and imports to GDP (Bekaert and Harvey, 1995) are also included. We also include value, investment, profitability and intangibles variables, not all necessarily related to integration, for two reasons: i) to increase the hurdle for our model to select the most important variables and ii) to control for firm characteristics that have been found to generate anomalous returns. The associated portfolio returns of anomalous characteristics cannot be fully explained by well established factors (Hou et al., 2015) meaning that the R-squared of factor models against portfolios sorted by these characteristics will vary. We wish to control for this kind of variation in our R-squared framework.

Our primary objective is to establish which determinants matter most in explaining firm-level integration with the world and, for that purpose, we rank variables in terms of their importance using a variety of measures. In our empirical setting, the strength of the relationship between the variables strongly motivated by the literature and firm-level integration can be assessed jointly instead of unilaterally as it has been the case in previous studies. We assess the strength of the link between our variables and integration by applying the general-to-specific (GETS) algorithm of Bekaert et al. (2011). Bekaert et al. (2011) have used GETS for country-level characteristics only while we use it mainly for firm-level characteristics. GETS constitutes a "testing-down" procedure that, in multiple steps, eliminates variables with coefficient estimates that are not statistically significant, until it reaches to a parsimonious model with mostly significant regressors. It represents a class of linear models and, as such, it yields interpretable ordinary least square coefficients at every stage of its estimation including the final model. The high interpretability of this algorithm is its main advantage. To complement the shortcomings of linear modeling, we also use random forest regression techniques to account for both non-linear effects and multicollinearity issues with our findings remaining largely unchanged.

Our results can be summarized as follows. We find that the corporate spread is the most important determinant of the level of firm integration and it explains 32% of the total panel variation closely followed by size, which explains 31% of the R-squared variation and is the single most important firm-specific variable that determines the relation between a firm and the rest of the world. Total institutional ownership and foreign sales follow size with an explanatory power of 14% and 8% in the GETS model, respectively. Thus, even though a positive and significant effect on integration is established as expected for size (Di Giovanni et al., 2017; Eun et al., 2008), foreign sales (Brooks and Negro, 2006) and institutional ownership (Faias and Ferreira, 2017), there is a large difference between them in terms of their importance as determinants of firm-level integration.

We further study variables as groups. Aggregation of individual variables allows us to see which categories have the strongest relationship with firm integration. The variables are grouped into eight broad categories that include macroeconomic (*Macro*), price and return related (*Market*), export related (*Business*), institutional ownership (*Ownership*), value (*Value*), intangibles (*Intangibles*), 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]

[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

We find that integration is driven primarily by *Macro*, *Market* and *Ownership* variables that are always the top 3 most important groups by a large margin. Their overall contribution in the GETS model is 41%, 34% and 11%, respectively, while the *Business* group that includes foreign sales, contributes only 8% to the fit. The fact that *Business* variables, which are de jure measures of economic integration, always fall behind *Ownership* variables such as total or foreign institutional ownership, highlights the role of institutional investors as agents of globalization. In other words, we find that, besides the macroeconomic environment, larger market cap firms that are mostly owned by institutions and are big exporters, are more sensitive to global shocks than their counterparts. Findings are unchanged after examining the effect of various robustness adjustments.

The paper is organized as follows. In Section 2 we discuss how we measure integration at the firm level. Section 3 describes the data and the construction of variables used in our analysis. Section 3.1 describes the algorithms and the methods used to determine the most important determinants of integration. Section 4.1 documents the upward trend of integration in the US while Section 4.2 summarizes, motivates and examines the set of plausible explanatory variables of firm-level integration in an univariate setting. Section 4.3 presents our empirical findings in a multivariate framework while Sections 4.4 and 4.5 discuss our results. We discuss several robustness checks in Section 5 and we conclude in Section 6.

2. Measuring firm-level integration

We employ the R-squared methodology of Pukthuanthong and Roll (2009) to measure integration at the firm level. The R-squared$_{i,t}$ measure for firm i at period t is estimated on an annual frequency using daily returns from July of year t-1 to June of year t. The advantage of the July-June scheme is that it allows us to match our R-squared measure with the most recent accounting information for a firm as in Fama and French (1993).

Daily returns of time period t are regressed against the first M Principal Com-

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.

ponents (PCs) constructed by the Principal Component Vectors (PCVs) of the market return matrix of the previous period t-1. The return matrix includes the international market indices of the 10 most developed countries except the US. [3] The 10 countries are Australia, Canada, France, Germany, Hong Kong, Italy, Japan, Singapore, Switzerland, and the UK. We gather their daily return series data converted to US dollars from Thomson Reuters Datastream for 1977-2019. In the original paper of Pukthuanthong and Roll (2009), the daily return matrix of eastern countries is augmented by the lagged returns of western markets. In our case, lagged returns are not included in the return matrix X_t since the US market is the last trading market in a given day. In that framework, the M Principal Components as estimated from only international markets are treated as global factors that proxy the world economic environment.[4]

$$\vec{r}_{i,t} = F_t \vec{b}_{i,t} + c\vec{1} + \vec{e}_{i,t}, \tag{2.1}$$

where $\vec{r}_{i,t} \in \Re^{N_{i,t} \times 1}$ is a column vector of the returns of stock i at period t, c is the constant term, $\vec{1}$ is a $\Re^{N_{i,t} \times 1}$ vector of ones, $F_t \in \Re^{N \times M}$ is the principal component matrix and $\vec{e}_{i,t}$ is the error term. $N_{i,t}$ is the number of stock returns used in the regression and it varies across stocks i and time t. The matrix F_t is calculated as $F_t = X_t \tilde{P}_{t-1}$ where $\tilde{P}_{t-1} \in \Re^{K \times M}$ is the principal component vector matrix of the covariance matrix of the Market Index returns X_{t-1} of the previous period or the eigenvector matrix of $\left(X_{t-1} - \bar{X}_{t-1}\right)^T \left(X_{t-1} - \bar{X}_{t-1}\right)$ where $\bar{X}_{t-1}$ demeans the columns of $X_{t-1} \in \Re^{N_{t-1} \times K}$. The PCs that we construct are out-of-sample since we apply the PCVs of the previous period t-1 to the demeaned return market index

[3] In an alternative specification, we have also used all available MSCI indices to construct foreign global factors. When the MSCI indices are used, we find that the resulting R-squared measure gives similar findings, and has a 91% correlation with that calculated from just the 10 indices.

[4] We directly test whether our R-squared captures the effect of global rather than US market changes on US companies. Specifically, we regress each non-US equity index on US market returns and use the resulting residuals as adjusted market indices. The purified indices are, in turn, used to extract global factors and compute a revised R-squared measure. This approach effectively purges the influence of the US market from the international indices. We find that the correlation of the purified R-squared with the original R-squared is 84% meaning that the latter is largely unaffected by US market changes and that it is truly a measure of financial integration of US companies with the rest of the world.

matrix of the current period t. A valid R-squared measure is computed when the number of daily observations for a stock in a given year exceeds 50, $N_{i,t} \geq 50$.

By definition, the R-squared captures the percentage of the variation of a stock's returns in a year that is explained by the variation of the global factors' returns. Intuitively, the R-squared in that setting measures how much global factors affect a US stock by incorporating all international markets in a single regression. The measure has a simple and direct interpretation. The higher the adjusted R-squared is for a stock, the more integrated it is with the global economy as proxied by the out-of-sample principal components of the international indices. A value of 1 corresponds to total integration while a value of 0 to no integration. Since we use the adjusted R-squared to define firm-level integration instead of just the R-squared, no integration can correspond to negative values.

In principle, the global factors are not known. We proxy the global factors as the first five (M=5) principal components. This choice is guided by the data since five eigenvectors are typically enough to explain on average 85% of the variation in our set of equity index returns. Any omitted factor would have the effect of adjusting the overall level of integration, rather than impacting our company-level inference. Figure 1 shows the average explained variance ratio for the 10 largest by market cap country indices. [5]

[Insert Figure 1 here]

3. Data

We construct a comprehensive list of 43 firm-level characteristics sourced from the intersection of CRSP, Compustat and FactSet that we summarize in Table 1. We use a dataset that contains a variety of variables, not all necessarily related to integration, in order to increase the hurdle for our models to select the most important determinants of firm-level integration. For that purpose, we construct variables found in Hou et al. (2015) who use the most representative firm-

[5]For example, Pukthuanthong and Roll (2009) use the first 10 principal components that capture on average 90% of the total variation of the market return matrix. They also explain that they use the first 10 components since "it seems reasonable that 10 large industry groupings adequately capture most global shocks."

level anomalies in their tests. Our measure of integration is essentially a measure of goodness of fit of a naive international asset pricing model to US stock returns. Anomalous firm-specific variables and their associated long-short portfolio (anomaly) have been found to generate abnormal returns and low R-squared values against well established asset pricing factors (Hou et al., 2015). Thus these anomalous characteristics are best suited to be used as controls. The variables are further categorized into groups. The groups are *Market*, *Business*, *Ownership*, *Investment*, *Profitability* and *Intangibles*. We also expand their dataset by defining the *Business* group that includes foreign sales and the foreign sales ratio as well as the *Ownership* group that contains institutional ownership related variables. Size, foreign sales and institutional ownership variables are assumed to be the strongest determinants of firm-level integration based on the previous literature. We provide a detailed description of that literature in later sections.

Finally, we augment the 43 firm-level characteristics with 6 country-level variables from Bekaert et al. (2011) that are found to be important in explaining the panel variation of country-level integration in their study, for a total list of 49 variables. These variables include the default spread, the VIX and the Risk Aversion index of Bekaert et al. (2021) which is an updated version of the Risk Aversion index found in Bekaert et al. (2011). They have been used in the literature to capture innovations that forecast future changes in the investment opportunity set under the framework of the Intertemporal Capital Asset Pricing Model of Merton (1973). For example, Petkova (2006) used the innovations of the default spread among other state variables that describe investment opportunities in her study and found that these variables can explain better the cross-section of returns than the Fama-French HML and SMB factors. Bekaert et al. (2021) define the Risk Aversion index and relate it to changes in risk aversion which is one of the components of the asset pricing kernel based on consumption. The other country-specific variables that we include in our analysis are the number of Internet users and the total dollar amount of exports and imports over GDP and they both capture the globalization trend of information and trade. The list of all 49 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 source of institutional ownership data is FactSet in which data is available quarterly after March 1999 meaning that *Ownership* variables are available after June 1999. Thus our final sample includes 30,000 firm-year observations that span 20 years with 1700 firms in 1999 and 1200 firms in 2019.

3.1. Variable selection models

We wish to distinguish the relative importance of a comprehensive list of variables for our measured integration without imposing strong theoretical priors. For that reason, we employ two separate variable selection methods to determine which firm- or country-specific variables explain the panel variation of firm-level integration. The first approach is an OLS method in the form of a general-to-specific (GETS) algorithm that Bekaert et al. (2011) use to find the country variables that are most important in explaining market segmentation. The GETS model represents a general class of linear models and as such its results can be easily interpreted. The second method is the random forest regression (RFR) of Breiman (2001) that was recently applied by Akbari et al. (2021) in the search for the drivers of economic and financial integration. RFR has the advantage of handling better highly correlated variables while it allows for complex non-linear interactions between the candidate explanatory variables and our measure of integration.

3.1.1. General-to-specific modelling

Our aim is to find a parsimonious set of factors that best explain the variation in integration. For that purpose, we employ the general-to-specific search algorithm of Krolzig and Hendry (2001). The algorithm constitutes a "testing-down" process that in multiple steps eliminates variables with coefficient estimates that are not statistically significant, leading to a parsimonious model with mostly significant regressors. Both the intermediate testing models as well as the final parsimonious are linear models. Thus GETS represents a class of linear models and it yields interpretable OLS coefficients at every stage of its estimation. The interpretability inherited by the class of linear models is its main advantage. Appendix A provides a more detailed discussion of the test procedure.

After we select the variables from GETS, we run the following linear panel model with only those variables X:

$$\text{R-squared}_{i,t} = \alpha + \boldsymbol{\beta_i X_{i,t}} + \epsilon_{i,t} \tag{3.1}$$

where R-squared$_{i,t}$ is the R-squared of firm i at year t and $\boldsymbol{X_{i,t}}$ is a set of variables that have been selected by GETS.

3.1.2. Random forest regression

Our second option in search of the determinants of firm-level integration is the Random Forest Regression (RFR) of Breiman (2001). The main advantage of RFR over the general-to-specific modelling of the previous section 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 linear models of Section 3.1.1 and thus the importance of variables can be masked. Akbari et al. (2021) discuss this issue and adopt the RFR to uncover the drivers of financial and economic integration at the country level. 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, RFR is applied to our list of 49 variables in order to find the determinants of firm-level integration in the US for the period 1999-2019. The details of the RFR implementation are presented in Appendix B.

3.1.3. Variable importance

Even though our primary goal is to find a parsimonious model to explain the variation of integration across firms, we also wish to discover the most influential covariates by ranking them according to a measure of variable importance. In other words, we wish to find the variables that are the key determinants of firm-level integration. For that purpose, we consider three different notions of importance. The first is the overall contribution of variable j to the explained variation of the fitted integration R-squared measure of Bekaert et al. (2011). It captures the variation that each variable can explain over the total fitted variation of the explanatory power of the linear model. The second 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 permutation test score imply greater importance for variable j. The third is the reduction in R2 (explanatory power) 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. The first method is applicable only for linear models such as GETS while the last two are generic and they can be used for any model. The details of their definition can be found in Appendix C.

4. Empirical findings

4.1. US firm integration

The US integration with the world market is uncovered by aggregating the individual integration measure of all US incorporated and publicly traded stocks in our sample. The equal-weighted cross-sectional mean and standard deviation of integration of all US firms are plotted in Figure 2 which, upon a simple visual inspection, indicates the increasing integration of the US through the years.

[Insert Figure 2 here]

Apart from the increasing time trend of integration, two periods of inflated R-squared measures are also observed; the first period with a mean R-squared of 22% is the year 1987 when the market crashed in October in a black swan event; the second period when the mean R-squared is as high as 22% corresponds to the recent global financial crisis of 2007-09. The effect of the mortgage crisis lasts from 2008 to 2011. It is also clear that the standard deviation of integration across firms increases when the average level increases implying greater heterogeneity of the sensitivity of firms to foreign markets. In the next section we study how integration varies across firm characteristics and crisis periods.

4.2. Variables and univariate analysis

In order to give a flavour of the relationship between the variables and integration in a univariate setting, we concentrate on those variables that have been found to be important determinants of integration in the past. As such, we present a list of explanatory firm-specific or macroeconomic variables motivated by the literature and their expected relationship with firm-level integration. To establish how firm characteristics affect integration, we study the cross-sectional variation of the integration series across the dimension of portfolios based on those characteristics in a univariate setting. More specifically, at the end of June of year t, we sort stocks into quintile portfolios based on the value of a firm characteristic at June of year t. We then calculate and plot the equal-weighted average integration estimate of stocks within each of those portfolios. In other words, the integration of a characteristic portfolio in the period that ends on June of year t is defined as $\text{R-squared}_{s,t} = \sum_{i=1}^{N_{s,t}} \text{R-squared}_{i,t}/N_{s,t}$ where $N_{s,t}$ is the number of stocks in portfolio s and $R^2_{i,t}$ is the firm-level measure of integration. In each July-June period, the difference in average integration between the extreme portfolios is tested formally via the non-parametric two-tailed test of Welch (1947) with the corresponding p-values being estimated from permutations.

Some of the *Market* variables such as past returns as well as the *Investment*, *Profitability*, *Value* and *Intangibles* variables are not motivated by the integration literature at all. We have no reason to expect investment-related characteristics such as the annual percentage change in total assets or profitability-related characteristics such as gross profitability over the book-value of equity to explain firm-level integration. However, these characteristics serve two purposes in our analysis: (i) they are used as controls and (ii) they increase the hurdle for our models to select the most important variables in determining firm-level integration.

4.2.1. Size

Size is one of the most important characteristics that can determine the integration of a stock due to the inherent heterogeneity between small and large firms in many levels. Huang (2007) show that only large-caps are priced globally meaning that small-caps are not as financially integrated as large-caps. Eun et al. (2008) find that the correlation dynamics of small and large cap portfolios are different; the correlation of small-cap stocks with other small-caps or large-caps is much lower than the correlation between large-caps. Similarly, this difference mechanically distinguishes firms with respect to size when return-based measures of integration are used. However, this difference in integration between large and small firms might be only the manifestation of other underlying factors. These factors may be the exporting/importing activity of the firm or the increased levels of institutional ownership and trading of its shares due to media and analyst coverage.

For instance, Di Giovanni et al. (2017) use a proprietary French customs dataset of 1 million firms and find that the top 100 largest firms contribute substantially to the business cycle comovement of France and its foreign partners. These firms are more internationally connected to foreign countries through their exporting/importing activity or multinational linkages, namely, each firm is a subsidiary of a foreign multinational, or is itself a French parent with a subsidiary abroad. Were these direct linkages to be severed, the average correlation between the output growth of France and the GDP growth of its trading partners would fall by 0.10 and 25% of that is due to the top 100 firms. This finding highlights the importance of the real economic links of a firm and the markets that it operates.

There is also evidence that large stocks attract foreign institutional investors. Kang et al. (1997) use ownership data for Japanese incorporated stocks and find that there is a disproportionate allocation of foreign capital in large firms instead of smaller ones. This could be due to large cap stocks being more well-known mostly because of their exporting activity or liquidity. Ferreira and Matos (2008) confirm the findings of Kang et al. (1997) using an international dataset of equity holdings for 27 countries. Institutional investors operate globally and have the ability to change international prices, especially of the largest stocks. Even though size is linked with institutional ownership, there is enough evidence that the ownership channel is distinct and as such we consider it separately from size.

For the aforementioned reasons, we expect that larger firms would be more integrated with world markets. Figure 3 plots the equal-weighted integration time series of the quintile portfolios based on market capitalization. We highlight the Developed Markets (DMs) recession periods as defined by NBER with grey shaded areas. Although strongly overlapping, we choose DM recessions over US recessions because our global setting tells a story for both the US and DMs. The 1st quintile portfolio denoted by "Small" contains the smallest 20% of stocks by market capitalization while the 5th quintile denoted by "Large" contains the largest 20%.

The pattern is clear: integration increases with size and the difference between "Small" and "Large" portfolios is exacerbated in distress periods of Developed Markets (DMs) indicated by the grey areas in the graph. In more detail, the integration of small firms remains low and stable with its level never rising beyond 3.5% with the exception of 1987 and 2020. On the contrary, large firms experience a more turbulent pattern with high levels of integration after 2000. The two-tailed non-parametric test based on permutations rejects the equality of integration between small- and large-cap stocks at any conventional significance level after July 1980. The average difference in integration for the same period is 10% with huge spikes in NBER recession periods of up to 36% in magnitude.

[Insert Figure 3 here]

4.2.2. Foreign sales

Exporting (and importing) activity creates strong economic links between a firm and foreign markets and its importance in integration has been documented both at the firm and country level. Brooks and Negro (2006) showed that the more international a firm is, the more its returns co-move with a global factor. In their paper, internationalization is defined from the foreign sales ratio which is the ratio of sales generated abroad over the total sales of a firm and the foreign asset ratio which is the ratio of foreign assets over total assets. In their analysis, they regress the factor loading of the global factor against the Foreign Sales and Foreign Assets Ratio and establish a positive and statistically significant relation between the two.

In the same vein, Di Giovanni et al. (2018) establish that at the micro level, trade and multinational linkages with a particular foreign country are positively associated with the correlation between a firm and that foreign country. Their methodology allows them to estimate the impact of direct linkages on co-movement which would fall by about 0.098 or one-third of the observed average correlation of 0.291 in their sample of partner countries, if these linkages were to be severed.

We proxy the trading activity of the firm by its foreign sales (FS) and the foreign sales ratio (FSR) and we expect a positive relationship with integration. Figure 4 plots the equal-weighted integration time series of the "zero" and quartile portfolios based on foreign sales and its ratio. The "zero" portfolio contains all the firms that report a zero value for sales from abroad while the 1st and 4th quartile portfolio denoted by "Low" and "High" contains the bottom and top 25% of firms that report a positive FS value. The pattern for foreign sales and its ratio is strictly monotonic across the quartiles with integration increasing along the foreign sales dimension. More specifically, small exporters in terms of foreign sales are subject to a relatively less volatile and lower integration level that is mostly restricted to the 3%-5% range while integration for large exporters is above 10%, especially after 2000. The difference in integration level between the "Low" and "High" portfolios is larger in magnitude for FS than FSR but it is nevertheless statistically significant for both. For example, the average difference for foreign sales is 9% while that of foreign sales ratio is just 1%. Part of that difference can be explained by size. In unreported results, we observe a contraction of the difference between the "Low" and "High" FS portfolio when we residualize against size while the residualized FSR portfolios remain almost unchanged.

[Insert Figure 4 here]

4.2.3. Ownership variables

We also consider three dimensions of institutional ownership as potential determinants of firm-level integration. These dimensions are the total and foreign institutional ownership as well as a measure of common ownership. Each variable captures a different aspect of the ownership channel and as such it is studied separately. Specifically, we distinguish foreign from total investors while foreign common ownership is a more refined interlinkage measure between the firm and foreign markets.

Faias and Ferreira (2017) establish that cross-border institutional portfolio investment is a powerful force of international capital market integration and convergence of asset prices. These institutions invest worldwide as agents of financial globalization and as such we expect firms with high total institutional ownership to be more integrated. Specifically, they construct market-cap weighted portfolios of low and high institutional ownership and use a factor model to decompose the variance explained by industry, country and global shocks. The ratio of country to global variance is higher for the low ownership portfolio meaning that highly owned stocks are more sensitive to global rather than local shocks. Even though the effect of the number of investors on integration has not been directly observed, it is logical to assume that it would be positive.

Figure 5 plots the equal-weighted integration time series of the quintile portfolios based on total ownership. The 1st quintile portfolio denoted by “Low” contains the 20% stocks with least institutional ownership while the 5th quintile denoted by “High” contains the top 20% highly owned stocks. Stocks that have high levels of total ownership, exhibit high levels of integration and vice versa. The average integration level of minimally and highly owned firms is 3.6% and 13%, respectively with a large and significant difference between the two of 9.4% that is similar to the size effect. The total institutional ownership pattern remains unchanged when even we control for size by residualizing ownership against size.

[Insert Figure 5 here]

When Faias and Ferreira (2017) control for foreign institutional ownership, their results are stronger for firms with high foreign ownership; the global factor explains more of the variation of the portfolio return than the country factor for firms with high total ownership in the high foreign ownership quantile. Thus among similar stocks in terms of total ownership, the ones with high foreign ownership will be more sensitive to global shocks. They also use the addition of a firm to the MSCI All index as an exogenous shock that increases total ownership through foreign ownership and find that global betas increase significantly over local betas for those stocks. The last finding also points to a positive relation between foreign ownership and global firm-level integration.

Figure 6 plots the equal-weighted integration time series of the quintile portfolios based on foreign ownership. The 1st quintile portfolio denoted by “Low” contains the bottom 20% stocks by foreign ownership while the 5th quintile denoted by “High” contains the top 20%. Stocks that have high levels of foreign ownership exhibit high levels of integration and vice versa. The foreign ownership integration levels for the “High” and “Low” portfolio are similar to those of total ownership with an average and statistically significant difference of 12%. Results remain the same when we control for size.

[Insert Figure 6 here]

Bartram et al. (2015) construct a foreign ownership return for stock i that takes into account the institutional investment habitat that gives more weight to stock j that shares a large number of common owners with stock i, to stocks whose owners hold more of stock i and to stocks that are heavily invested in by stock i's owners. They find that this ownership return behaves like a common risk factor for stocks that belong to the habitat emphasizing the importance of the linkages created between stocks via foreign common ownership. In a similar vein Anton and Polk (2014) also emphasize the role of institutions in linking stocks that they commonly own. Their measure of interconnectedness is $FCAP_{i,j,t} = \sum_{f=1}^{F} \left(\frac{hcap_{i,t}^{f} + hcap_{j,t}^{f}}{ME_{i,t} + ME_{j,t}} \right)$ for mutual funds f that own both stocks i and j at end of the period t. $hcap_{i,t}^{f}$ and $ME_{i,t}$ are the market value of the holdings of fund f in stock i and the market value of the stock i, respectively. We modify this measure by considering all institutions f (not only mutual funds) and only stocks j that are foreign with respect to i. We choose to work with foreign stocks j since we are interested in US firm-level linkages with foreign markets and assets and as such it is natural to omit linkages with domestic assets. Then we average over all foreign stocks j that are commonly owned with i to define FCO_mean as $\frac{1}{J} \sum_{j=1, j \in foreign}^{J} FCAP_{i,j,t}$. Thus FCO_mean is a more refined version of Anton and Polk (2014)'s FCAP and it is designed to capture institutional ownership linkages between US and foreign equities.

Figure 7 plots the equal-weighted integration time series of the quintile portfolios based on foreign common ownership. The 1st quintile portfolio denoted by "Low" contains the bottom 20% stocks by foreign common ownership while the 5th quintile denoted by "High" contains the top 20%. It is clear that the effect of common ownership on integration is positive and monotonic as expected. The average integration of the "Low" and "High" portfolio is 2.7% and 14%, respectively and as such their significant difference is estimated at 11.3%. When we control for size, the foreign common ownership pattern remains the same.

[Insert Figure 7 here]

4.2.4. *Corporate spread, Risk aversion and VIX*

Market conditions in developed countries may drive capital flows and thus affect international valuation differentials (Fernandez-Arias, 1996). The corporate spread between BAA and AAA bonds as well as the VIX may affect capital flows as proxies for risk aversion or sentiments of world investors. High levels of the BBB-AAA spread and the VIX signal distress periods in which the correlation dynamics of markets changes (Forbes and Rigobon, 2002; Rodriguez, 2007). RiskAversion is the direct measure of risk-aversion in the US of Bekaert et al. (2021). All three indicators of business or financial cycle are positively related to country-level integration according to Bekaert et al. (2011) and as such they should also affect all firms in the US the same. Figure 8 shows the strong and positive relationship between the mean US integration level and corporate spread, VIX and risk aversion.

4.2.5. *Internet users and Trade*

The number of internet users and the total dollar amount of US exports and imports over the GDP are proxies for information and trade, respectively. The percentage of internet users reflects the general ease with which country's citizens can obtain information and thus recognize risks and improve risk sharing. On the other hand, trade/GDP is a de jure factor of economic openness and free flow of capital among countries. Trade/GDP takes into account the exporting and importing activity of both private and public firms in the US and thus it contains more information than the foreign sales data for individual public firms.[6] Both these variables have been found to be positively correlated to integration as suggested by Bekaert et al. (2011) and Akbari et al. (2021).

4.3. *Multivariate analysis*

In this section, we present the findings of our empirical analysis in terms of the GETS model and the measures of variable importance that we use to distinguish

[6]We have also used net exports of goods and services (NETEXP) of the US economy as an additional country level business variable and we found that our baseline results remain unchanged. In fact, the sum of US exports and imports over GDP is a more important determinant of the time-series variation of integration than NETEXP. Results for NETEXP are available upon request.

the true variables that explain firm-level integration. Our multivariate framework confirms the plausibility of the variables described in the previous section and then ranks them in terms of their explanatory power within our models.

Table 2 shows the results of the panel regression of the final GETS model whose overall explanatory power is 47%. The signs and significance of the preferred multivariate analysis are straightforward to interpret but they do not provide guidance on which variables are more important in explaining integration. As such we report the coefficient of the final regression with the variables selected by the GETS algorithm, the overall contribution of each variable to the fitted model by which we rank them, the values of the permutation test and the change in explanatory power as defined in Section 3.1.3. Figure 9 offers a visual representation of the importance of individual variables and their groups across all different measures.

[Insert Table 2 and Figure 9 here]

Our analysis reveals several interesting results. First, we find that macroeconomic variables have a large effect on the level of integration of all stocks in the US market and they are always ranked very highly in terms of importance. The corporate spread, also known as the default spread, is the single most important variable that can explain most of the time variation in R-squared. In the regression framework, when the daily (annual) default spread increases by 1% (16%), US integration with the world increases by almost 9%. The spread explains 32% of the fitted variation or alternatively it explains 32% of the temporal fitted variation. When its value is set to zero, the explanatory power of the model decreases by 1.3% while its score in the permutation test is 23%. In other words, when we permute the corporate spread values and thus break the relation with integration, the prediction error is the second largest capturing 23% of the sum of errors generated by permuting all the other variables. High corporate spread values signify periods of distress in financial markets in which the correlation between them increases. Rodriguez (2007) documented this phenomenon and concluded that there can be structural breaks in tail dependence in high volatility regimes. Forbes and Rigobon (2002) also study crisis periods including the 1987 US market crash and reach the same conclusion; interdependence structure changes in these periods.

VIX and the Bekaert et al. (2021) risk-aversion index behave the same way as corporate spread and their levels increase in times of distress. When VIX increases by one unit, integration increases by 0.24% but it decreases by 2.5% when the risk-aversion index does the same. Both the percentage of internet users as well as trade/GDP have a positive relationship with US integration with coefficients 0.13 and 0.16, respectively. Only private credit, which is a proxy for the financial development of the US, exhibits an unexpected negative sign (-0.19) in the final GETS model. Overall, the relationship of the *Macro* variables with firm-level integration is the expected one as confirmed by Bekaert et al. (2011) in their study of country-level determinants.

Second, we confirm that size (ME) is the most important firm characteristic that determines the level of integration of a firm with the world. Specifically, a 1% change in market capitalization increases the value of the R-squared measure of the stock by 1.7%. The importance of size is also the highest across all measures of variable importance as shown in Figure 9. For example, size by itself can explain 31% of the variation of the fitted firm-level integration while removing its effect by setting its value to zero for all firms drops the fit of the model by 2.6%. The change in explanatory power with respect to size is by far the highest that we observe in our GETS analysis followed only by corporate spread and total institutional ownership (1.3% and 1.2% respectively). The permutation test corroborates the significance of market capitalization with a score of 33% that is larger than that of any other variable. The relation between size and integration is positive and strong thus validating the findings of past literature.

The second most important firm-specific determinants of integration are either the total institutional ownership variables (IO) or the number of total investors (IO_num). When institutional investors increase their holdings of a stock by 1%, the integration of that stock increases by 0.08% confirming the positive effect that ownership has on integration. To put it differently, a one standard deviation increase of total institutional ownership induces, on average, an increase of 0.22 standard deviations in the R-squared. Institutional ownership is the second most important firm characteristic after size based on the metrics of overall contribution and change in explanatory power. Specifically, it explains 14% of the fitted R-squared variation while the fit of the model drops by 1.2% when we assume that none of the firms are held by institutions. The score of total ownership in the permutation test is very strong with the number of institutional investors, IO_num, ranking higher than total ownership.

Surprisingly, neither foreign institutional ownership (FIO) and the number of foreign investors (FIO_num) or foreign common institutional ownership (FCO_mean) can explain firm-level integration. Only foreign ownership was selected by the GETS algorithm but its contribution to the model is essentially non-existent. Both foreign ownership and the number of foreign investors rank low among the top 24 most important variables based on the other importance tests. Foreign common ownership shares a worse fate than the others as it is excluded by both the general-to-specific algorithm and random forest regression even though the literature supports a positive relation with integration.

Another interesting finding is that foreign sales (FS) and not the foreign sales ratio (FSR) is what matters more in explaining the total panel variation of integration. This suggests that the monotonic and positive pattern between firm-level comovement and the foreign sales ratio that is documented by Brooks and Negro (2006) is not the complete story: the absolute magnitude of exports matters more than the proportion of exports over total sales for integration. Based on our measure of overall contribution foreign sales is the third firm characteristic after size and institutional ownership and explains 7.2% of the fitted R-squared. When a firm increases its sales abroad by 1%, its R-squared increases by 0.3%. Alternatively, when exports rise by one standard deviation, integration with foreign markets rises by 0.12 standard deviations of R-squared on average. The sign of the effect is positive as expected by the literature on exporting firms. The importance of foreign sales fades when we measure it via the permutation test or the change in explanatory power where foreign sales is found to be less important with scores of 0.8% and 0.2% respectively.

When we study variables as groups, we find that firm-level integration is driven by three types of factors: *Macro*, *Market* and *Ownership* variables that are always the top 3 most important groups by a large margin. Their overall contribution in the GETS model is 41%, 34% and 11% (Figure 9b), respectively. The *Business* group ranks in the fourth place and explains 8% of the fitted R-squared variation while the other groups explain less than 2.5%. Thus we confirm that variables from other groups which are used only as controls, contribute little in explaining integration. Interestingly, *Business* variables always fall behind *Ownership* variables in terms of importance as shown in Figure 9 meaning that institutional investors play a crucial role in the convergence of international prices.

4.4. Institutional investors and integration

In this section, we investigate the relationship between institutional investors and firm-level integration by providing potential explanations for two of our main findings: i) why is foreign institutional ownership not as important as total institutional ownership and ii) how do institutional investors drive integration?

Financial integration is ultimately shaped by the actions of market participants as they implement their investment strategies and engage in trading. Foreign ownership (FIO) is defined as the fraction of shares of a US stock that non-US institutional investors hold, whereas total institutional ownership (IO) refers to holdings by both domestic (US-based) and foreign institutions. Figure 11 plots the aggregate market capitalization of US equities owned by domestic (US) and foreign (non-US) institutions. The pattern is clear; domestic institutional holdings are consistently at least five times larger than those of foreign institutions. This substantial disparity suggests that domestic institutions play a dominant role in shaping US asset prices.

[Insert Figure 11 here]

Thus, it is the trading behavior and capital allocation decisions of US-based institutions that primarily drive the integration of US firms with global financial markets. Their market presence allows them to more effectively transmit global shocks and trends into domestic pricing. On the contrary, foreign institutional investors hold relatively smaller positions and thus exert a more limited, second-order influence on US prices. This imbalance in market presence offers a plausible explanation for our empirical finding: total institutional ownership (IO) that is dominated by US institutions emerges as a significantly more important determinant of firm-level integration in the US than foreign institutional ownership (FIO).

The influence of institutional investors in global financial markets has grown substantially in recent decades. Since 2004, institutional investors have consistently held over 70% of the total market capitalization of US equities, and approximately 40% of global market capitalization. Thus, the representative investor is in fact, the representative institution. It is their trading activity that has an immediate impact on asset prices and drives integration. Both Faias and Ferreira, 2017 and Bartram et al. (2015) provide evidence that institutions tend to follow more internationally diversified strategies and that global institutions have the largest effect on cross-border comovement. To empirically examine their impact on market integration, we decompose total institutional ownership (IO) into three components: global (IO_G), regional (IO_R), and local (IO_L) ownership. This decomposition allows us to isolate the role of globally-oriented institutions in explaining the variation in firm-level integration of US stocks.

First, we follow the methodology of Bartram et al. (2015) to classify institutions into local, regional or global investors. We calculate for each institution the percentage of its holdings that are in a country and a region in a quarter. If the maximum average percentage of the holdings in a country over the previous 4 quarters (1 year) is more than 90% of the institution's total holdings, the institution is classified as a local investor. Otherwise, if the maximum average percentage in a region is more than 80%, it is classified as a regional investor. Otherwise, it is classified as a global investor.

Second, we examine the relationship between firm-level integration and institutional ownership by regressing R-squared on the ownership of global, regional, and local investors, while controlling for firm size and year fixed effects. Model (3) in Table 3 reveals a particularly compelling story: global institutional ownership has the strongest positive association with R-squared, followed by regional and then local ownership. Specifically, the estimated coefficients for global and regional institutions are 0.104 and 0.082, respectively which are both more than double the coefficient for local institutions (0.031). This pattern is consistent with the hypothesis of Faias and Ferreira, 2017 and the findings of Bartram et al. (2015) that investors are becoming global in nature and that these global investors matter the most in international comovement and thus integration.

[Insert Table 3 here]

4.5. Foreign sales, firm size and integration

Foreign sales, defined as the dollar amount of revenue generated outside the United States, represent one of the most direct measures of a firm's degree of internationalization (Brooks and Negro, 2006; Aggarwal et al., 2011; Di Giovanni et al., 2018). The link between a firm's operational internationalization and its financial integration becomes evident when considering the following. Basic asset pricing theory states that the price of a stock equals the discounted sum of expected future dividends. Consequently, stock returns reflect two components: i) innovations in expected future dividends and ii) innovations in the discount rate. Innovations in dividends are influenced by the extent of a firm's international operations. Firms operating globally are more exposed to global shocks, whereas firms with primarily domestic operations are more sensitive to local market shocks. Innovations in the discount rate are driven by the degree of financial market integration of the country the company is domiciled. Given that our sample is restricted to US-listed firms, we expect a uniform response across firms to discount rate innovations but differential responses to dividend innovations based on their level of internationalization. As a result, stocks of more (less) international firms should exhibit stronger (weaker) comovement with global markets, leading to higher (lower) levels of financial integration. This mechanism, originally documented by Brooks and Negro (2006) and Di Giovanni et al. (2018), is confirmed in our analysis.

We explore in detail the role of firm size on integration in Section 4.2.1.

5. Robustness checks

We discuss several robustness checks.

5.1. Random forest regression

In this section, we augment our analysis using the random forest regression model that is able to handle highly correlated variables and capture non-linear effects.

When we employ the random forest regression, we find that our results are very similar to those of general-to-specific modelling. However, the RFR technique does not generate interpretable coefficients as those in the conventional regression framework and as such we examine the RFR results under the prism of the variable importance measures only. The tree structure of RFR explains almost 90% of the total variation of R-squared in contrast to the 47% value of GETS. This almost double increase in the fit of RFR compared to GETS suggests that there are complex and possibly non-linear interactions between our plausible explanatory variables and our measure of firm-level integration that only RFR manages to capture.

Figure 10 shows the importance of individual variables and their groups based on the permutation test and the change in R2 with our focus on the former. Corporate spread is again found to be the most important determinant of integration that influences all US firms. Size (ME) remains one of the most important firm characteristics in explaining integration but it is ranked second with a score of 13% (Figure 10a). Surprisingly, the number of institutional owners (IO_num) is at the top ranking in both tests scoring 15% and 40% and it is only second to corporate spread (36% and 56%). Foreign sales (FS) is placed at the third (5%) or fourth place (10%) of firm characteristics that explain the R-squared variation behind both ME and IO_num.

When we repeat our analysis with variable groups, we find that *Macro*, *Market* and *Ownership* variables are the most important with the RFR results being in agreement with those obtained by GETS. Their permutation test scores are 41%, 18% and 17% (Figure 10b) while *Profitabiliy* ranks fourth and explains 6%.

[Insert Figure 10 here]

5.2. Correlation VS R-squared measure

We first assess the simple correlation as a substitute for our R-squared and find similar results. Correlation is the simplest and most popular measure of dependence between two variables and as such our results are benchmarked against it. Specifically, we re-estimate the GETS and RFR models using the correlation with the Fama-French world market factor[7] excluding the US as an alternative measure of integration of a firm with foreign markets. It is calculated only when

[7]The world market factor is found in the Fama/French Developed ex US 3 Factors file in their website `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

there are at least 50 daily returns available for a firm in the July-June period in the same vein as the R-squared.

Panel A of Table 4 reports the variable importance when correlation is used in the analysis for GETS and RFR. Results remain largely unchanged with the *Macro*, *Market* and *Ownership* variables being at the top. Their overall contribution in the GETS model is 50%, 30% and 11% and their permutation test scores in the RFR setting are 50%, 17% and 14% , respectively. Corporate spread is again the most important variable explaining 42% of the variation in GETS. Size (ME) and the number of foreign investors (FIO_num) prevail as the most important firm-specific characteristics in GETS and RFR, respectively. Specifically, ME explains 26% of the fitted variation of integration while FIO_num has a permutation score of 10% that is second only to that of corporate spread (41%). One noticeable difference between the correlation and R-squared results is the concentration of the macroeconomic variables at the top of the list in terms of importance, especially in the general-to-specific modelling.

[Insert Table 4 here]

5.3. Probability to default as a firm-level distress measure

The corporate spread between BAA and AAA bond indices remains the most important variable of firm-level integration in all specifications. The spread measures the difference in default between high- and grade-investment bonds. Thus a question is raised; will a firm-specific measure of probability to default be able to capture better the effect of the corporate-spread? In other words, is there a macro- or a micro-economic effect? We tackle this question by including the probability to default (PtD) as a plausible explanatory variable in our models. PtD is extracted from the Kealhofer-Merton-Vasicek model in which the market value of the shareholder's equity can be viewed as a call option on the total assets of the firm with a strike price equal to the debt obligations over a one-year horizon. Figure 12 shows that PtD is not even included in the first 30 most important variables in the RFR model. Thus changes in the macroeconomic environment signified by the increase of the corporate spread matter more in explaining integration than firm-level measures of distress.

[Insert Figure 12 here]

5.4. The effect of micro-cap stocks

Even though micro-cap stocks comprise the majority of the US 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 micro-cap when it belongs to the bottom 20% quantile of all NYSE stocks. In this section, we explore their effect on our analysis by excluding them.

Panel B of Table 4 reports the variable importance when micro-caps are excluded from the analysis for GETS and RFR. Micro-cap stocks do not alter qualitatively our results across models and variable importance measures. More specifically, size (ME) is still the most important important firm characteristic with institutional ownership variables ranking high across models. Overall, *Macro*, *Market* and *Ownership* variables hold their positions as the primary determinants of firm-level integration in the US.

5.5. The effect of financial firms

Finally, we examine how our results change when we exclude financial firms. Panel C of Table 4 reports the variable importance when financial firms are excluded from the analysis for GETS and RFR with results remaining largely unchanged. The same pattern of *Macro*, *Market* and *Ownership* variables being the most determinants of integration emerges suggesting that our results are not driven by the financial sector.

5.6. Additional macroeconomic variables

Motivated by the success of macroeconomic variables such as Corporate Spread in explaining the time series variation of integration, we incorporate additional key business cycle indicators proposed by Brogaard and Detzel (2015) into both the GETS and RFR models. Specifically, we include four variables from their study: (i) VAR: the realized variance of the CRSP value-weighted index; (ii) BILL: the yield on the 3-month Treasury bill; (iii) TERM: the term spread between the 10-year Treasury bond and the 3-month Treasury bill; and (iv) CFNAI: the Chicago Fed National Activity Index. These variables capture broader aspects of macroeconomic conditions and investor sentiment, providing a richer characterization of the business cycle.

The inclusion of the additional variables improves the explanatory power of our GETS model, increasing the R^2 from 47% to 52%. These business cycle indicators exert a strong influence in the linear model and they have the highest values of variable importance. The results from the RFR analysis depict a complementary picture. Under this framework, ownership breadth variables such as the number of total and foreign institutional investors (IO_num and FIO_num) consistently emerge as the most important drivers of firm-level integration; a finding that is consistent with our baseline results. Both GETS and RFR results are summarized in Panel A of Table 5.

[Insert Table 5 here]

5.7. Controlling for industry effects

To ensure the robustness of our findings, we further examine the potential influence of industry-level effects by implementing a two-stage process. In the first stage, we isolate the variables selected by the GETS algorithm. In the second stage, we estimate two panel regressions: one using only the GETS selected variables and another that includes both the GETS variables and industry fixed effects.[8] Industry classifications are based on the 10 Fama-French industry groupings. The regression results are summarized in Table 6. We show that the inclusion of industry dummies does not affect neither the magnitude nor the statistical significance of the estimated coefficients for the GETS selected variables. This finding confirms that our main conclusions hold even after controlling for industry effects.

[Insert Table 6 here]

5.8. Integration in crisis and non-crisis periods

Our sample allows us to study whether the dynamics of equation 3.1 which relates firm- and macro-level characteristics to firm integration, changes in crisis

[8]Given that the GETS algorithm performs a sequential elimination of variables based on individual statistical significance, incorporating industry fixed effects directly within the GETS framework is not feasible. Specifically, including industry dummies would require treating them as a group, which is impossible to do in the variable by variable testing structure of the GETS approach.

periods. Table 7 reports the crisis periods that might have affected the integration dynamics. We identify three major crises; the Dot-Com bubble, the Great Recession of 2007 and the Euro area sovereign debt crisis. The years that lie outside of the crisis periods of Table 7 define the non-crisis periods. For our 1999-2019 sample, the crisis periods amount to 7 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 both the GETS and RFR algorithms in each subsample. Panels B and C of Table 5 report the most important drivers of integration for the two subsamples. When the markets are distressed, macroeconomic variables such as VIX and Corporate Spread become even more important in explaining the variation in firm-level integration. Size (ME) is again the most important firm characteristic regardless of the condition of the markets. The RFR model that incorporates non-linear interactions between the variables, reveals that the breadth of institutional ownership as proxied by the number of total and foreign institutional investors, IO_num and FIO_num respectively, plays a more important role than size in both regimes. Overall, we find that the integration dynamics do not change significantly between crisis and non-crisis periods.

5.9. Firm-level foreign direct investment

Outward foreign direct investment (OFDI) provides a complementary picture to the level of the firm internationalization that is not captured by foreign sales. OFDI is typically reflected in a firm's balance sheet through foreign assets, including equity stakes in foreign subsidiaries, intercompany loans, and overseas property or equipment. Thus, we use foreign assets as a proxy for OFDI to capture the degree of a firm's international investment footprint. Specifically, we obtain data on the foreign assets to total assets ratio (FAR) from Refinitiv Eikon and calculate foreign assets (FA) by multiplying total assets by FAR. To assess the importance of FA and FAR in explaining integration, we augment our set of *Business* variables and re-estimate the Random Forest Regression model. Figure 13 presents the top 21 most important variables based on the RFR algorithm. FA ranks 13th based on the permutation test, while FAR ranks 18th based on the change in R^2. Although these OFDI proxies do not exhibit high explanatory power, they nonetheless contribute to firm-level integration.

[Insert Figure 13 here]

6. Conclusion

Even though integration at the country-level has been studied extensively in the literature, firm-level integration has not been fully explored. In this paper, we provide insights on what determines the relationship of firms with foreign markets. To that end, we first compute a measure of firm-level integration using the Pukthuanthong and Roll (2009) methodology for all US stocks for the period 1999-2019. We then combine that measure with a comprehensive dataset of 43 firm characteristics and 6 macroeconomic variables for a total of 49 variables to uncover the factors that characterize firm integration. General-to-specific modelling (GETS) and random forest regression (RFR) are employed to distinguish between variables that matter in explaining the time-series and cross-sectional variation of integration and those that do not.

First, we find that the integration of the average US stock has increased over the years. However, there is great heterogeneity in terms of the R-squared values of firms meaning that firms exhibit various degrees of integration. Second, we examine how integration varies across firm- and macro-level characteristics that have been shown to be linked with integration in the past literature. By doing so, we establish a relationship between integration as measured by the R-squared and size, foreign sales, foreign sales ratio, institutional ownership and macro-economic variables such as the corporate spread.

Finally, we rank the variables in terms of importance and find that results are consistent across the GETS and RFR techniques and across different measures of variable importance. More specifically, the US corporate spread between

BAA and AAA bond indices has the largest effect in the integration of all US stocks. However, this is not limited to corporate spread; macroeconomic variables in general matter more than firm characteristics. Furthermore, we find that size (ME) is the single most important firm characteristic that determines the level of integration of a firm with the rest of the world. Institutional ownership (IO) or the total number of institutional investors (IO_num) are the second most important characteristics after size that matter in explaining integration. This last finding highlights the power of institutions as agents of globalization that trade internationally and contribute to the convergence of prices and thus increase the financial integration between markets. Surprisingly, foreign sales, a de jure factor of economic integration that serves as a proxy for the exporting activity of a firm, comes third after size and institutional ownership.

When we categorize variables into groups, we find that *Macro*, *Market* and *Ownership* variables contribute the most in explaining the heterogeneity of integration across US firms, in that order. Thus, even though country-level conditions are the primary determinants of integration, our results suggest that firm-specific characteristics such as size and institutional ownership still play an equally important role at the granular level.

The evidence presented by the study holds various meaningful implications for firms and investors. For firms, our results highlight that higher levels of institutional ownership, greater visibility and liquidity as proxied by market capitalization and increased foreign sales translate to increased integration. Thus firms emphasizing transparency and engaging in international trading activities are more integrated and as such they enjoy lower costs of capital (Stulz, 1999, Hail and Leuz, 2009). A lower cost of capital translates to higher valuations since future cash flows are now discounted by a lower discount rate. For investors, the identification of the important drivers of integration offers a way to filter for the less or more integrated firms based on institutional ownership, size and foreign sales. Specifically, they may choose to allocate capital toward less integrated firms to capture diversification benefits by tilting their portfolio to smaller, less institutionally owned and domestically oriented firms. From a policy perspective, our findings suggest that promoting institutional investment by easing cross-border capital flows, and supporting firm-level international expansion can contribute to deeper financial market integration and thus lower costs of capital for companies.

References

Aggarwal, R., Berrill, J., Hutson, E., and Kearney, C. (2011). What is a multinational corporation? classifying the degree of firm-level multinationality. *International Business Review*, 20(5):557–577.

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.

Avramov, D., Li, M., and Wang, H. (2021). Predicting corporate policies using downside risk: A machine learning approach. *Journal of Empirical Finance*, 63:1–26.

Bartram, S. M., Griffin, J. M., Lim, T.-H., and Ng, D. T. (2015). How important are foreign ownership linkages for international stock returns? *The Review of Financial Studies*, 28(11):3036–3072.

Bartram, S. M. and Wang, Y.-H. (2015). European financial market dependence: An industry analysis. *Journal of Banking & Finance*, 59:146–163.

Bekaert, G., Engstrom, E. C., and Xu, N. R. (2021). The time variation in risk appetite and uncertainty. *Management Science*.

Bekaert, G. and Harvey, C. R. (1995). Time-varying world market integration. *the Journal of Finance*, 50(2):403–444.

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.

Bracker, K., Docking, D. S., and Koch, P. D. (1999). Economic determinants of evolution in international stock market integration. *Journal of Empirical Finance*, 6(1):1–27.

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

Breusch, T. S. and Pagan, A. R. (1980). The lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, 47(1):239–253.

Brogaard, J. and Detzel, A. (2015). The asset-pricing implications of government economic policy uncertainty. *Management Science*, 61(1):3–18.

Brooks, R. and Negro, M. D. (2006). Firm-level evidence on international stock market comovement. *Review of Finance*, 10(1):69–98.

Chambet, A. and Gibson, R. (2008). Financial integration, economic instability and trade structure in emerging markets. *Journal of International Money and Finance*, 27(4):654–675.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, pages 591–605.

Christoffersen, P., Errunza, V., Jacobs, K., and Jin, X. (2014). Correlation dynamics and international diversification benefits. *International Journal of Forecasting*, 30(3):807–824.

Conlon, T., Cotter, J., and Ropotos, I. (2024). Diversification with globally integrated us stocks. *Journal of International Financial Markets, Institutions and Money*, 90:101889.

De Marco, F. (2019). Bank lending and the european sovereign debt crisis. *Journal of Financial and Quantitative Analysis*, 54(1):155–182.

Di Giovanni, J., Levchenko, A. A., and Mejean, I. (2017). Large firms and international business cycle comovement. *American Economic Review*, 107(5):598–602.

Di Giovanni, J., Levchenko, A. A., and Mejean, I. (2018). The micro origins of international business-cycle comovement. *American Economic Review*, 108(1):82–108.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Eun, C. S., Huang, W., and Lai, S. (2008). International diversification with large-and small-cap stocks. *Journal of Financial and Quantitative Analysis*, pages 489–523.

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.

Fernandez-Arias, E. (1996). The new wave of private capital inflows: push or pull? *Journal of Development Economics*, 48(2):389–418.

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 Chinn, M. D. (2004). A decomposition of global linkages in financial markets over time. *Review of Economics and Statistics*, 86(3):705–722.

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.

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica: Journal of the Econometric Society*, pages 1293–1301.

Goetzmann, W. N., Li, L., and Rouwenhorst, K. G. (2001). Long-term global market correlations. Technical report, National Bureau of Economic Research.

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

Hail, L. and Leuz, C. (2009). Cost of capital effects and changes in growth expectations around us cross-listings. *Journal of Financial Economics*, 93(3):428–454.

Harvey, C. R. (1991). The world price of covariance risk. *The Journal of Finance*, 46(1):111–157.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705.

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

Huang, W. (2007). Financial integration and the price of world covariance risk: Large-vs. small-cap stocks. *Journal of International Money and Finance*, 26(8):1311–1337.

Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3):255–259.

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

Kang, J.-K. et al. (1997). Why is there a home bias? an analysis of foreign portfolio equity ownership in japan. *Journal of Financial Economics*, 46(1):3–28.

Krolzig, H.-M. and Hendry, D. F. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, 25(6-7):831–866.

Longin, F. and Solnik, B. (1995). Is the correlation in international equity returns constant: 1960–1990? *Journal of International Money and Finance*, 14(1):3–26.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, pages 867–887.

Obstfeld, M. and Taylor, A. M. (2003). Globalization and capital markets. In *Globalization in historical perspective*, pages 121–188. University of Chicago Press.

Olson, L. M., Qi, M., Zhang, X., and Zhao, X. (2021). Machine learning loss given default for corporate debt. *Journal of Empirical Finance*, 64:144–159.

Petkova, R. (2006). Do the fama-french factors proxy for innovations in predictive variables? *The Journal of Finance*, 61(2):581–612.

Pretorius, E. (2002). Economic determinants of emerging stock market interdependence. *Emerging Markets Review*, 3(1):84–105.

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

Quinn, D. P. and Voth, H.-J. (2008). A century of global equity market correlations. *American Economic Review*, 98(2):535–40.

Rangvid, J., Santa-Clara, P., and Schmeling, M. (2016). Capital market integration and consumption risk sharing over the long run. *Journal of International Economics*, 103:27–43.

Rodriguez, J. C. (2007). Measuring financial contagion: A copula approach. *Journal of Empirical Finance*, 14(3):401–423.

Stulz, R. M. (1999). Globalization of equity markets and the cost of capital.

Welch, B. L. (1947). The generalization of 'student's'problem when several different population variances are involved. *Biometrika*, 34(1-2):28–35.

Appendices

A. General-to-specific algorithm

The GETS algorithm takes the following form:

1. The starting point is the general unrestricted model (GUM) as defined in equation 3.1 and includes all the variables described in Section 3.

2. Of the full sample, 90% is retained, while the remaining 10% is set aside for out-of- sample testing. The following battery of tests is run on this 90% sample at the nominal size which is 1% for our specification:

   (a) normality of residuals (Jarque and Bera, 1980).

   (b) autocorrelation of residuals up to second order (χ^2 test, see Godfrey, 1978, Breusch and Pagan, 1980)

   (c) autocorrelated conditional heteroscedasticity (ARCH) up to second order (Engle, 1982).

   (d) in-sample stability test (first half of the sample against the second half, see Chow, 1960).

   (e) out-of-sample stability test of specification estimated against re-estimation using 10% of data points retained for the Chow (1960) test
   .

   If one of these tests is failed, it is eliminated from the battery in the following steps of the search path.

3. Each variable in the general model is ranked by the size of its t statistic, and the algorithm then follows m (in our case, 10) search paths. The first search path is initiated by eliminating the variable with the lowest (insignificant) t statistic from the GUM. The second follows the same process, but rather than eliminating the lowest, it eliminates the second lowest. This process is followed until reaching the mth search path that eliminates the mth-lowest variable. For each search path, the current specification then includes all remaining variables, and this specification is estimated by regression.

4. The current specification is then subjected to the full battery of tests of Stage 2, along with an F test, to determine whether the current specification is a valid restriction of the GUM. If any of these tests fails, the current search path is abandoned, and the algorithm jumps to the subsequent search path.

5. If the current specification passes the above tests, the variables in the current specification are once again ordered by the size of their t statistics, and the variable with the next-lowest t statistic is eliminated. This then becomes a potential current specification, which is subjected to the battery of tests. If any of these tests fails, the model reverts to the previous current specification, and the variable with the second-lowest (insignificant) t statistic is eliminated. Such a process is followed until a variable is successfully eliminated or until all insignificant variables have been attempted. If an insignificant variable is eliminated, Stage 5 is restarted with the current specification. This process is followed iteratively until either all insignificant variables have been eliminated or no more variables can be successfully removed.

6. Once no further variables can be eliminated, a potential terminal specification is reached. This specification is estimated using the full sample of data. If all variables are significant, it is accepted as the terminal specification. If any insignificant variables remain, these are eliminated as a group, and the new terminal specification is subjected to the battery of tests. If it passes these tests, it is the terminal specification; if it does not, the previous terminal specification is accepted.

7. Each of the m (in our case, ten) terminal specifications is compared, and if these are different, the final specification is determined using encompassing or an information criterion.

B. 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 49.

   (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

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

   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 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

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

C. Definition of variable importance measures

C.1. Overall contribution

To examine the contribution of each of the independent variables X_j to the overall variation of the fitted integration measure $\hat{R}^2_{i,t}$, we compute the following covariance for each explanatory variable j:

$$Cov\left(\hat{R}^2_{i,t}, \hat{\beta}_j x_{i,j,t}\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i}\sum_{t=1}^{T_i}\hat{\beta}_j\left(\hat{R}^2_{i,t} - \bar{\hat{R}}^2_{i,t}\right)(x_{i,j,t} - \bar{x}_j) \tag{C.1}$$

where $\bar{x}_j$ is the mean of variable $x_{i,j}$ across firms and time. We also compute the variance of the fitted $\hat{R}^2_{i,t}$ as:

$$Var\left(\hat{R}^2_{i,t}\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i}\sum_{t=1}^{T_i}\left(\hat{R}^2_{i,t} - \bar{\hat{R}}^2_{i,t}\right)^2 \tag{C.2}$$

where $\bar{\hat{R}}^2_{i,t} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i}\sum_{t=1}^{T_i}\hat{R}^2_{i,t}$.

Thus we can define the ratio of each covariance term of equation C.1 to the overall predicted market integration variance of C.2, $\frac{Cov\left(\hat{R}^2_{i,t}, \hat{\beta}_j x_{i,j,t}\right)}{Var\left(\hat{R}^2_{i,t}\right)}$ as a measure of the economic significance of each variable j.[9] The higher (and more positive) the value of the ratio is, the more important variable j is in explaining the fitted values of $R^2_{i,t}$. The ratio is generally positive but it can be negative for some variables. Let's assume that the coefficient $\hat{\beta}$ of variable X is found to be positive. Then, if variable X contributes to the overall fit of the model, we should expect that an increase in X and thus in the predicted value of $\hat{\beta}X$ should increase the predicted value of $\hat{R}^2$ since $\hat{\beta}$ is positive. This means that the covariance of $\hat{\beta}X$ and $\hat{R}^2$ is positive. If, on the other hand, an increase in X and thus in $\hat{\beta}X$,

[9]Note that $\sum_{j=1}^{N_{gets}}\frac{Cov\left(\hat{R}^2_{i,t}, \hat{\beta}_j x_{i,j,t}\right)}{Var\left(\hat{R}^2_{i,t}\right)} = 1$ since $Var\left(\hat{R}^2_{i,t}\right) = Cov\left(\hat{R}^2_{i,t}, \hat{R}^2_{i,t}\right)$ with $\hat{R}^2_{i,t} = \sum_{j=1}^{N_{gets}}\hat{\beta}_j x_{i,j,t}$.

yields a decrease in $\hat{R}^2$, then the expected positive relationship between X and R^2 is broken and the covariance becomes negative. In essence, variable X does a bad job in explaining the overall variation of R^2. This measure is particularly appealing to linear regression models such as GETS and it has been used by Bekaert et al. (2011) to rank the effect of country-level predictors of segmentation. We follow their paper and use it as our primary variable importance measure in GETS modelling.

C.2. 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=10 times generating K corrupted datasets $X_{perm,j,k}$. Finally, we calculate the variable importance as the difference $VI_j = \frac{1}{K}\sum_{k=1}^{K}(e_{perm,j,k} - e_{orig})$. 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.

C.3. Change in R2

We measure the importance of variable j by setting its value to zero and compute the difference between the R2 (explanatory power) 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 explanatory power 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 integration

	Market			**Profitability**	
1	ILLIQ	Amihud, 2002's illiquidity measure	25	ROC	Size + longterm debt - total assets to cash
2	Coskewness	Coskewness with CRSP value-weighted index	26	ROE	Income before extraordinary items to lagged BE
3	ME	Size	27	ROIC	Return on invested capital
4	Mom6m	Return from 6 to 2 months before end of June	28	S2C	Sales to cash
	Business			**Intangibles**	
5	FS	Foreign sales	29	AOA	Absolute value of operating accruals
6	FSR	Foreign sales ratio	30	OL	Costs of goods solds + SG&A to total assets
			31	Tan	Tangibility
	Ownership		32	OA	Operating accruals
7	IO	Total institutional ownership			
8	FCO_mean	Average of Anton and Polk, 2014's FCAP_(i,j) over foreign stocks j		**Value**	
9	FIO	Foreign institutional ownership	33	A2ME	Total assets to Size
10	FIO_num	Number of foreign institutional owners	34	BtM	Book to market ratio
11	IO_HHI	Herfindahl–Hirschman index of IO	35	C	Cash to AT
12	IO_num	Number of institutional owners	36	C2D	Cash flow to total liabilities
			37	dSo	Log change in split-adjusted shares outstanding
	Investment		38	Debt2P	Total debt to Size
13	INV	% change in AT	39	E2P	Income before extraordinary items to Size
14	dCEQ	% change in BE	40	NOP	Net payouts to Size
15	dPI2A	Change in PP&E and inventory over lagged AT	41	O2P	Operating payouts to market cap
16	IVC	Change in inventory over average AT	42	S2P	Sales to price
17	NOA	Net-operating assets over lagged AT	43	Sales_g	Sales growth
	Profitability			**Macroeconomy**	
18	ATO	Sales to lagged net operating assets	44	CorporateSpread	US corporate bond spread between BBB and AAA
19	dGM_Sales)	Δ (% change in gross margin and % change in sales)	45	Internet_100_jun	Number of internet users per 100 people
20	PCM	Sales minus costs of goods sold to sales	46	RiskAversion	US risk aversion index of Bekaert and Engstrom (2010).
21	PM	OI after depreciation over sales	47	PrivateCredit_GDP_jun	Private credit divided by gross domestic product
22	Prof	Gross profitability over BE	48	Trade_GDP_jun	The sum of US exports and imports over GDP
23	RNA	OI after depreciation to lagged net operating assets	49	VIX	Option volatility index
24	ROA	Income before extraordinary items to lagged AT			

Notes: The table lists the variables we consider as potential determinants of firm-level integration in our analysis. 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.

Table 2. Regression results of GETS model

Variable	Coefficient	Overall contribution	Permutation test	Change in R2
CorporateSpread	7.864***	0.324	0.232	0.013
ME	1.662***	0.313	0.334	0.026
IO	0.079***	0.139	0.017	0.012
FS	0.310***	0.072	0.008	0.002
Internet_100_jun	0.127***	0.068	0.012	0.005
VIX	0.243***	0.064	0.154	0.001
FIO_num	0.016***	0.058	0.007	0.000
OL	-1.405***	0.040	0.014	0.006
Coskewness	2.843***	0.021	0.003	0.003
Trade_GDP_jun	0.158***	0.018	0.011	0.001
PrivateCredit_GDP_jun	-0.186***	0.018	0.001	0.013
C	-0.077***	0.015	0.035	0.003
BtM	-4.637***	0.011	0.000	0.001
FSR	0.008***	0.009	0.001	0.000
AOA	-2.818***	0.009	0.000	0.001
ILLIQ	-0.110***	0.006	0.001	0.001
INV	0.011***	0.003	0.001	0.001
dPI2A	0.011***	0.002	0.001	0.001
Mom6m	-0.011***	0.001	0.005	0.002
PM	0.000**	0.001	0.000	0.000
FIO	0.001	0.000	0.000	0.000
ROC	0	0.000	0.000	0.000
NOA	-0.018***	0.000	0.001	0.001
PCM	-0.003	-0.001	0.000	0.000
C2D	-0.001	-0.001	0.000	0.000
NOP	-7.445***	-0.001	0.000	0.000
S2P	0.204**	-0.002	0.000	0.000
IO_HHI	-23.739***	-0.003	0.004	0.003
Tan	0.050***	-0.020	0.027	0.001
RiskAversion	-2.538***	-0.078	0.087	0.001
IO_num	-0.006***	-0.088	0.039	0.001

Notes: Table summarizes the results of the regression of the GETS model. Variables (first column) are ranked in a descending order on their overall contribution to the model (third column). Overall contribution is the ratio of the covariance of the fitted explanatory variable j with the fitted $\hat{R}^2_{i,t}$ over the variance of $\hat{R}^2_{i,t}$. The fourth and fifth columns summarize the variable importance of the regressors in terms of the permutation test of Breiman (2001) and the change in R2. In the permutation test, we score variable j by the difference in prediction accuracy before and after permuting j. The change in R2 is the reduction in predictive R2 from setting all values of variable j to zero, while holding the remaining model estimates fixed. Errors are robust and in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 3. Global, regional and local institutional ownership and integration

	R-squared		
Variable	(1)	(2)	(3)
Intercept	-5.838***	-4.332***	
	(0.094)	(0.094)	
IO	0.063***		
	(0.001)		
IO_G		0.143***	0.104***
		(0.007)	(0.005)
IO_R		0.188***	0.082***
		(0.003)	(0.003)
IO_L		-0.032***	0.031***
		(0.002)	(0.002)
logme	2.125***	1.868***	2.151***
	(0.026)	(0.023)	(0.02)
Year Fixed Effects	-	-	x
Observations	86609	86609	86609
R^2	0.268	0.309	0.568

Notes: Table reports the regression results of R-squared against different types of institutional ownership after controlling for size and year fixed effects. We follow the methodology of Bartram et al. (2015) to classify institutions into local, regional or global investors. We calculate for each institution the percentage of its holdings that are in a country and a region in a quarter. If the maximum average percentage of the holdings in a country over the previous 4 quarters (1 year) is more than 90% of the institution's total holdings, the institution is classified as a local investor. Otherwise, if the maximum average percentage in a region is more than 80%, it is classified as a regional investor. Otherwise, it is classified as a global investor. Thus total institutional ownership (IO) is decomposed into global (IO_G), regional (IO_G) and local (IO_L) ownership. The sample includes all firm-years observations with non-missing institutional ownership and market capitalization date from June of 1999 to June of 2020. Robust standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table 4. Importance of determinants of firm-level integration - Robustness checks

Panel A: Correlation VS R-squared measure

General-to-specific modelling												Random forest regression							
Overall contribution				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	Variable	Value	Group	Value
CorporateSpread	0.423	Macro	0.495	RiskAversion	0.263	Macro	0.727	RiskAversion	7.279	Market	6.630	CorporateSpread	0.408	Macro	0.507	CorporateSpread	0.632	Macro	0.746
ME	0.257	Market	0.298	CorporateSpread	0.258	Market	0.170	ME	6.334	Macro	6.435	FIO_num	0.103	Market	0.173	FIO_num	0.184	Ownership	0.241
IO	0.157	Ownership	0.115	VIX	0.172	Ownership	0.095	Trade_GDP_jun	2.420	Ownership	0.007	ME	0.083	Ownership	0.137	ILLIQ	0.135	Market	0.214
FIO_num	0.100	Business	0.070	ME	0.160	Business	0.005	VIX	1.874	Business	0.000	ILLIQ	0.073	Value	0.049	Trade_GDP_jun	0.105	Profitability	0.088
Internet_100_jun	0.094	Intangibles	0.015	IO_num	0.056	Value	0.002	CorporateSpread	1.767	Intangibles	-0.012	Trade_GDP_jun	0.047	Business	0.049	Internet_100_jun	0.073	Value	0.086
VIX	0.081	Value	0.008	Trade_GDP_jun	0.023	Intangibles	0.001	PrivateCredit_GDP_jun	0.322	Value	-0.016	FS	0.043	Profitability	0.046	FS	0.072	Business	0.078
Trade_GDP_jun	0.080	Profitability	0.000	FIO_num	0.020	Investment	0.001	Internet_100_jun	0.210	Investment	-0.016	Internet_100_jun	0.029	Investment	0.021	ME	0.069	Intangibles	0.043
FS	0.050	Investment	0.000	IO	0.016	Profitability	0.000	IO_num	0.062	Profitability	-0.019	IO_num	0.017	Intangibles	0.017	IO_num	0.028	Investment	0.038
Coskewness	0.031			Internet_100_jun	0.010			IO	0.049			Sales_g	0.011			PrivateCredit_GDP_jun	0.025		
FSR	0.020			Coskewness	0.006			FIO_num	0.007			PrivateCredit_GDP_jun	0.010			OL	0.022		
Tan	0.010			IO_HHI	0.003			Coskewness	-0.008			Coskewness	0.009			Prof	0.019		
ILLIQ	0.009			FS	0.002			Tan	-0.011			Prof	0.009			ATO	0.018		
BtM	0.007			FSR	0.002			FS	-0.013			Mom6m	0.008			Sales_g	0.018		
AOA	0.006			Mom6m	0.002			IO_HHI	-0.014			VIX	0.008			Coskewness	0.017		
S2P	0.002			ILLIQ	0.002			FSR	-0.014			ATO	0.007			dPI2A	0.015		

Panel B: Effect of micro-cap stocks

General-to-specific modelling												Random forest regression							
Overall contribution				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	Variable	Value	Group	Value
CorporateSpread	0.806	Macro	0.586	CorporateSpread	0.244	Macro	0.678	ME	7.428	Market	7.615	CorporateSpread	0.419	Macro	0.484	ME	0.748	Macro	0.813
VIX	0.169	Market	0.180	VIX	0.215	Market	0.159	RiskAversion	4.187	Macro	5.526	ME	0.095	Market	0.128	CorporateSpread	0.635	Market	0.787
ME	0.131	Business	0.087	RiskAversion	0.169	Ownership	0.068	VIX	2.090	Intangibles	0.116	FS	0.082	Ownership	0.084	FIO_num	0.280	Ownership	0.330
IO	0.086	Ownership	0.068	ME	0.152	Intangibles	0.049	CorporateSpread	1.558	Value	0.031	FIO_num	0.062	Business	0.082	OL	0.231	Intangibles	0.246
FS	0.063	Intangibles	0.050	IO_num	0.045	Value	0.036	Internet_100_jun	0.941	Business	0.007	Prof	0.026	Profitability	0.081	FS	0.165	Profitability	0.206
OL	0.060	Value	0.017	Internet_100_jun	0.043	Business	0.009	Trade_GDP_jun	0.601	Ownership	-0.014	Trade_GDP_jun	0.022	Value	0.064	Prof	0.080	Business	0.165
FIO_num	0.056	Investment	0.011	C	0.034	Investment	0.002	Tan	0.249	Investment	-0.032	Internet_100_jun	0.021	Investment	0.035	Trade_GDP_jun	0.060	Value	0.138
Internet_100_jun	0.039			Tan	0.029			IO_num	0.039	Profitability	-0.034	Sales_g	0.021	Intangibles	0.031	dPI2A	0.041	Investment	0.066
Coskewness	0.031			OL	0.019			IO	0.031			dPI2A	0.021			Internet_100_jun	0.038		
FSR	0.024			IO	0.011			C	0.029			PCM	0.017			Sales_g	0.037		
ILLIQ	0.019			FIO_num	0.008			OL	0.013			OL	0.015			ATO	0.036		
C	0.015			Trade_GDP_jun	0.006			FS	-0.018			Coskewness	0.014			PCM	0.031		
dPI2A	0.009			FS	0.005			FIO_num	-0.023			ATO	0.011			Coskewness	0.029		
Debt2P	0.004			Coskewness	0.003			FSR	-0.027			ILLIQ	0.010			C	0.024		
NOA	0.002			FSR	0.003			Coskewness	-0.028			FSR	0.010			IO_num	0.021		

Panel C: Effect of financials

General-to-specific modelling												Random forest regression							
Overall contribution				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	Variable	Value	Group	Value
CorporateSpread	0.590	Macro	0.400	ME	0.352	Macro	0.476	ME	8.003	Market	8.160	CorporateSpread	0.357	Macro	0.413	CorporateSpread	0.554	Macro	0.669
ME	0.227	Market	0.257	CorporateSpread	0.221	Market	0.361	RiskAversion	1.149	Macro	4.590	IO_num	0.157	Market	0.181	IO_num	0.406	Ownership	0.414
IO	0.162	Ownership	0.169	VIX	0.150	Ownership	0.077	VIX	0.923	Intangibles	0.052	ME	0.129	Ownership	0.178	ME	0.165	Market	0.247
FS	0.095	Business	0.101	RiskAversion	0.075	Intangibles	0.042	CorporateSpread	0.848	Value	-0.001	FS	0.053	Profitability	0.064	OL	0.123	Profitability	0.164
VIX	0.054	Intangibles	0.045	IO_num	0.045	Value	0.032	Trade_GDP_jun	0.814	Business	-0.018	ILLIQ	0.035	Business	0.053	FS	0.096	Intangibles	0.138
FIO_num	0.050	Value	0.025	C	0.031	Business	0.010	Internet_100_jun	0.176	Ownership	-0.021	Prof	0.020	Value	0.052	ILLIQ	0.067	Value	0.101
OL	0.043	Investment	0.003	Tan	0.027	Investment	0.001	Tan	0.146	Investment	-0.036	Internet_100_jun	0.020	Investment	0.028	Prof	0.058	Business	0.096
Coskewness	0.022	Profitability	0.001	IO	0.018	Profitability	0.000	IO	0.005	Profitability	-0.038	Trade_GDP_jun	0.018	Intangibles	0.022	Internet_100_jun	0.045	Investment	0.053
BtM	0.018			Internet_100_jun	0.016			IO_num	-0.001			Sales_g	0.016			Trade_GDP_jun	0.042		
AOA	0.014			OL	0.014			C	-0.003			dPI2A	0.016			ATO	0.035		
C	0.010			Trade_GDP_jun	0.014			OL	-0.015			PCM	0.012			dPI2A	0.031		
Internet_100_jun	0.008			FIO_num	0.010			FS	-0.026			OL	0.010			Sales_g	0.027		
ILLIQ	0.006			FS	0.009			FIO_num	-0.032			Coskewness	0.010			PCM	0.025		
FSR	0.006			Mom6m	0.005			Coskewness	-0.035			FSR	0.009			Coskewness	0.019		
dPI2A	0.003			IO_HHI	0.004			Mom6m	-0.036			ATO	0.009			FSR	0.017		

Notes: Panel A reports the variable importance when correlation between a stock and the Fama-French world market factor excluding the US is used instead of the R-squared measure. Panel B reports the variable importance when micro-cap stocks (bottom 20% NYSE quantile) are excluded from the sample. Panel C reports the variable importance when financial firms (SIC between 6000 and 6999) are excluded from the sample. 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 5. Importance of determinants of firm-level integration - Robustness checks

Panel A: Inclusion of additional business cycle variables such as VAR, TERM, BILL and CFNAI

General-to-specific modelling												Random forest regression							
Overall contribution				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	Variable	Value	Group	Value
VAR	0.361	Market	0.330	VAR	0.252	Macro	0.346	PrivateCredit_GDP_jun	11.537	Macro	22.090	IO_num	0.042	Macro	0.288	IO_num	5.735	Macro	41.461
CorporateSpread	0.352	Ownership	0.118	CFNAI	0.193	Market	0.130	ME	7.716	Market	7.910	FIO_num	0.038	Ownership	0.165	ME	5.313	Ownership	37.296
ME	0.295	Macro	0.110	CorporateSpread	0.184	Intangibles	0.016	VAR	2.436	Intangibles	0.057	ME	0.038	Value	0.149	FIO_num	5.031	Value	15.721
IO	0.134	Business	0.079	ME	0.127	Value	0.013	VIX	1.948	Value	-0.004	CorporateSpread	0.037	Profitability	0.142	FIO	4.972	Profitability	14.848
BILL	0.110	Intangibles	0.026	VIX	0.117	Ownership	0.012	CorporateSpread	1.915	Business	-0.021	BILL	0.035	Market	0.101	TERM	4.023	Market	9.686
FS	0.070	Value	0.011	BILL	0.025	Business	0.004	RiskAversion	0.723	Ownership	-0.021	ILLIQ	0.035	Investment	0.054	FCO_mean	3.962	Investment	6.328
OL	0.035	Investment	0.002	PrivateCredit_GDP_jun	0.023	Investment	0.000	CFNAI	0.293	Investment	-0.040	RiskAversion	0.033	Intangibles	0.054	CorporateSpread	3.919	Intangibles	5.586
Coskewness	0.027	Profitability	0.000	RiskAversion	0.018	Profitability	0.000	Tan	0.146	Profitability	-0.042	VIX	0.032	Business	0.029	PrivateCredit_GDP_jun	3.864	Business	2.961
PrivateCredit_GDP_jun	0.014			C	0.012			Internet_100_jun	0.141			VAR	0.030			RiskAversion	3.847		
C	0.011			Tan	0.011			BILL	0.040			FIO	0.030			VIX	3.539		
AOA	0.010			TERM	0.009			TERM	0.007			FS	0.029			Internet_100_jun	3.532		
FSR	0.009			IO	0.006			IO	-0.002			CFNAI	0.028			VAR	3.341		
FIO	0.007			Internet_100_jun	0.005			C	-0.004			TERM	0.027			IO	3.177		
ILLIQ	0.007			OL	0.005			OL	-0.021			FCO_mean	0.025			IO_HHI	3.069		
Internet_100_jun	0.006			IO_num	0.004			FS	-0.029			Trade_GDP_jun	0.023			FS	2.961		

Panel B: Crisis periods

General-to-specific modelling												Random forest regression							
Overall contribution				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	Variable	Value	Group	Value
CorporateSpread	0.631	Macro	0.421	RiskAversion	0.557	Macro	0.972	RiskAversion	97.600	Market	6.332	CorporateSpread	0.060	Macro	0.241	CorporateSpread	0.056	Ownership	0.381
VIX	0.548	Market	0.401	VIX	0.256	Market	0.019	VIX	33.843	Macro	3.339	VIX	0.050	Ownership	0.182	IO_num	0.055	Macro	0.375
ME	0.370	Ownership	0.091	CorporateSpread	0.127	Ownership	0.003	CorporateSpread	12.956	Intangibles	0.083	IO_num	0.045	Profitability	0.154	FIO_num	0.052	Profitability	0.162
Internet_100_jun	0.286	Business	0.073	Internet_100_jun	0.027	Intangibles	0.003	PrivateCredit_GDP_jun	10.034	Ownership	0.061	FIO_num	0.044	Value	0.153	ME	0.050	Value	0.158
FCO_mean	0.109	Value	0.006	ME	0.018	Value	0.002	Internet_100_jun	9.444	Value	0.040	ILLIQ	0.041	Market	0.108	VIX	0.049	Market	0.099
FS	0.073	Investment	0.004	PrivateCredit_GDP_jun	0.004	Business	0.001	ME	5.895	Business	0.027	RiskAversion	0.041	Investment	0.058	FIO	0.047	Investment	0.060
OL	0.028	Intangibles	0.003	C	0.002	Investment	0.001	Trade_GDP_jun	1.154	Investment	0.004	ME	0.040	Intangibles	0.052	RiskAversion	0.044	Intangibles	0.047
Coskewness	0.027	Profitability	0.000	FCO_mean	0.002	Profitability	0.000	Tan	0.154	Profitability	0.000	Trade_GDP_jun	0.034	Business	0.032	Internet_100_jun	0.041	Business	0.028
FIO	0.015			Trade_GDP_jun	0.002			FCO_mean	0.070			FIO	0.033			FCO_mean	0.040		
C	0.011			Tan	0.002			C	0.038			FS	0.032			PrivateCredit_GDP_jun	0.039		
INV	0.005			FS	0.001			FS	0.027			Internet_100_jun	0.030			Trade_GDP_jun	0.034		
ILLIQ	0.004			OL	0.001			OL	0.023			FCO_mean	0.028			ROC	0.034		
ATO	0.000			Coskewness	0.001			Coskewness	0.013			PrivateCredit_GDP_jun	0.026			IO	0.029		
dGM_SALE	0.000			INV	0.000			NOA	0.005			IO	0.020			FS	0.028		
NOA	-0.001			FIO_num	0.000			INV	0.004			FSR	0.019			IO_HHI	0.026		

Panel C: Non-crisis periods

General-to-specific modelling												Random forest regression							
Overall contribution				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	Variable	Value	Group	Value
ME	0.495	Market	0.501	ME	0.218	Macro	0.690	PrivateCredit_GDP_jun	30.400	Market	8.907	CorporateSpread	0.060	Macro	0.193	CorporateSpread	0.056	Ownership	0.398
CorporateSpread	0.461	Macro	0.300	CorporateSpread	0.212	Market	0.222	RiskAversion	11.354	Macro	5.937	VIX	0.050	Ownership	0.176	IO_num	0.055	Macro	0.274
VIX	0.103	Ownership	0.098	VIX	0.192	Intangibles	0.036	ME	8.985	Intangibles	0.141	IO_num	0.045	Value	0.174	FIO_num	0.052	Value	0.184
FCO_mean	0.078	Business	0.065	RiskAversion	0.161	Value	0.029	Trade_GDP_jun	6.005	Value	0.053	FIO_num	0.044	Profitability	0.161	ME	0.050	Profitability	0.158
IO	0.054	Value	0.018	PrivateCredit_GDP_jun	0.088	Ownership	0.018	VIX	2.265	Ownership	0.035	ILLIQ	0.041	Market	0.116	VIX	0.049	Market	0.111
OL	0.046	Intangibles	0.015	Trade_GDP_jun	0.039	Business	0.004	CorporateSpread	1.778	Business	0.010	RiskAversion	0.041	Investment	0.066	FIO	0.047	Investment	0.075
FS	0.046	Investment	0.003	C	0.028	Investment	0.001	Tan	0.275	Investment	-0.005	ME	0.040	Intangibles	0.060	RiskAversion	0.044	Intangibles	0.068
C	0.022	Profitability	0.000	Tan	0.025	Profitability	5.90E-05	C	0.051	Profitability	-0.007	Trade_GDP_jun	0.034	Business	0.033	Internet_100_jun	0.041	Business	0.036
FSR	0.019			OL	0.011			OL	0.023			FIO	0.033			FCO_mean	0.040		
FIO	0.016			FCO_mean	0.008			FCO_mean	0.016			FS	0.032			PrivateCredit_GDP_jun	0.039		
PrivateCredit_GDP_jun	0.012			IO_num	0.005			IO	0.003			Internet_100_jun	0.030			Trade_GDP_jun	0.034		
AOA	0.008			Mom6m	0.003			IO_num	0.001			FCO_mean	0.028			ROC	0.034		
ILLIQ	0.007			FS	0.003			FS	-0.001			PrivateCredit_GDP_jun	0.026			IO	0.029		
dPI2A	0.002			IO	0.003			Mom6m	-0.004			IO	0.020			FS	0.028		
IVC	0.001			FSR	0.001			FSR	-0.004			FSR	0.019			IO_HHI	0.026		

Notes: Panel A reports the variable importance when we include the VAR, BILL, TERM and CFNAI of Brogaard and Detzel, 2015 as additional business cycle variables. Panels B and C 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 6. Industry effects

	R-squared	
Variable	(1)	(2)
Intercept	-52.622***	
CorporateSpread	8.256***	8.104***
ME	2.546***	2.447***
IO	0.042***	0.046***
FS	0.239***	0.210***
Internet_100_jun	0.070***	0.066***
VIX	0.537***	0.536***
FIO_num	0.012***	0.011***
OL	-1.325***	-0.609***
Coskewness	2.077***	2.062***
Trade_GDP_jun	0.351***	0.345***
PrivateCredit_GDP_jun	0.028***	0.026***
C	-0.093***	-0.053***
BtM	-1.136*	-1.178*
FSR	0.012***	0.009***
AOA	-1.656***	-0.725*
ILLIQ	-0.108***	-0.105***
INV	0.005***	0.003**
dPI2A	0.006***	-0.003**
Mom6m	-0.013***	-0.013***
PM	0.000**	0
FIO	0.018	0.031
ROC	0	0
NOA	-0.007***	-0.002
PCM	-0.002	0
C2D	-0.002***	-0.001**
NOP	-14.554***	-12.839***
S2P	0.254***	0.152*
IO_HHI	-14.594***	-16.110***
Tan	0.095***	0.059***
RiskAversion	-4.433***	-4.368***
IO_num	-0.008***	-0.007***
Industry Fixed Effects	-	x
Observations	30440	30440
R^2	0.496	0.526

Notes: Table summarizes the results of the regression of the GETS model without industry effects (model (1)) and with industry effects (model (2)). Robust standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table 7. Outline of the 1999-2019 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

Notes: The table lists the major crises for the period of 1999-2019 along with start and end dates as well as the firm-years affected. The dates for the Dot-com bubble and the Global Financial Crisis of 2007 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 Marco, 2019). 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.

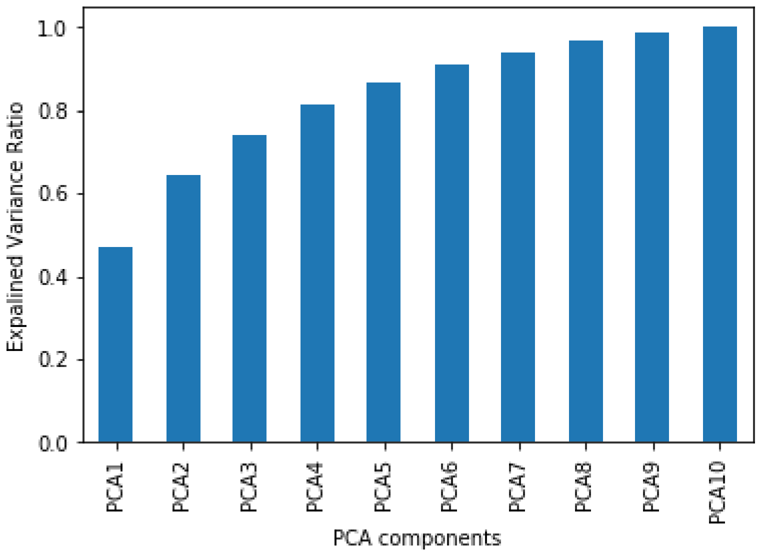


Figure 1. Cumulative explained variance ratio of the 10 PCA components.

Notes: The figure plots the cumulative explained variance ratio of the ten principal components as extracted from the largest market cap indices. The indices are from Australia, Canada, France, Germany, Hong Kong, Italy, Japan, Singapore, Switzerland and UK. On average the first five Principal Components can explain almost 85% of the total variance of the return matrix. We use the 85% bound as our decision making criterion.

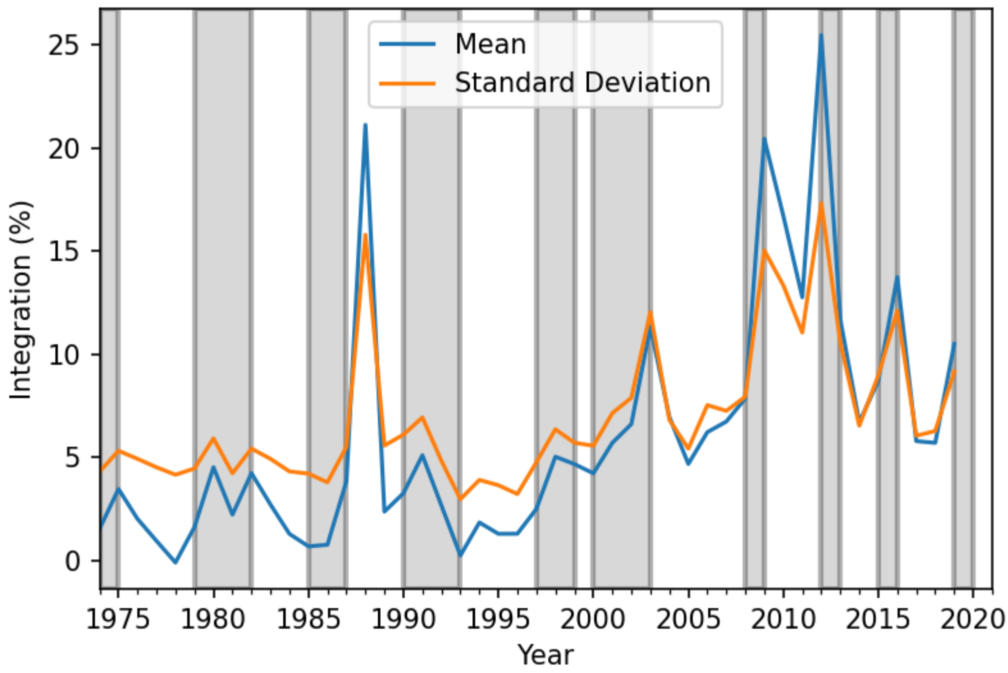


Figure 2. Global integration level of the US market for the period 1974-2019

Notes: Figure plots the mean (%) and standard deviation (%) of global integration for all US domiciled and traded stocks. Global integration is measured as the adjusted R-squared of a regression of principal components of foreign market country indices against the returns of a stock. The R-squared is computed on an annual calendar basis from 1974 to 2019 using daily returns. The gray shaded areas correspond to Developed Markets recessionary periods as defined by NBER.

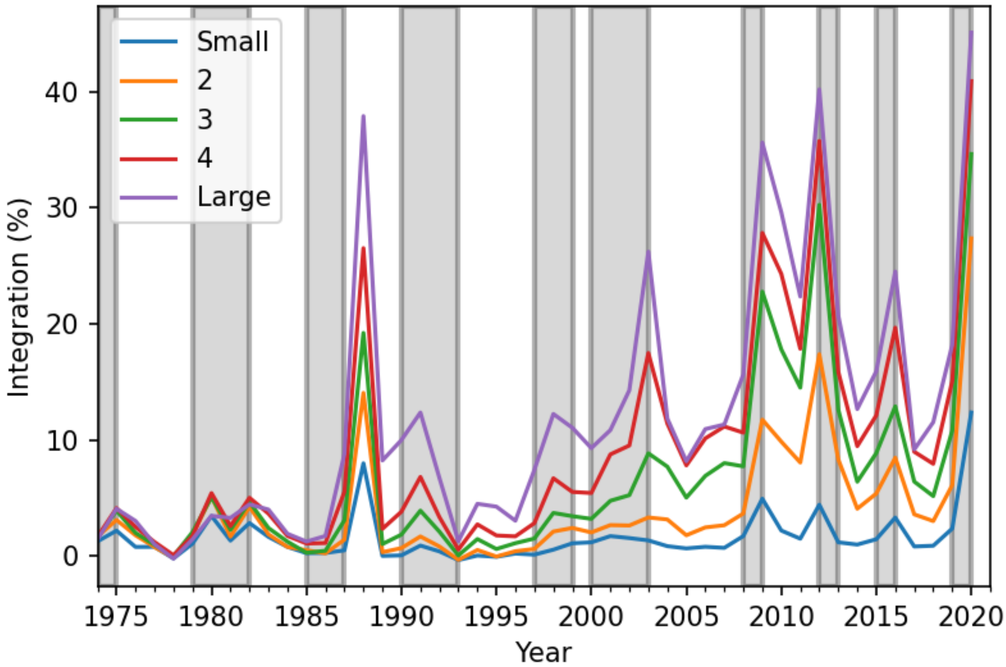
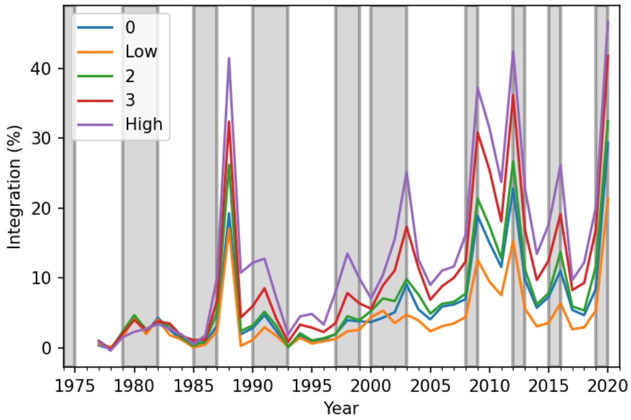
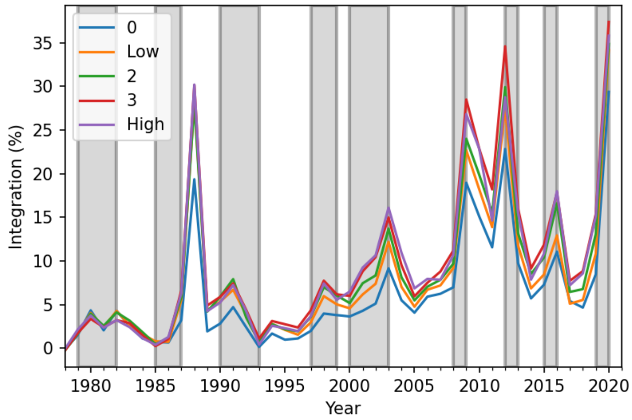


Figure 3. Integration time series across market cap based quintile portfolios.

Notes: Figure shows the equal-weighted integration time series across market cap quintile portfolios. At the end of each June, we sort stocks into quintile portfolios based on their current market capitalization (ME). We then calculate the mean integration estimate of stocks within each of these 5 portfolios. The 1st and 5th quintile portfolios denoted by "Small" and "Large" correspond to the smallest and largest by market cap stocks, respectively. The gray shaded areas correspond to Developed Markets recessionary periods as defined by NBER.



(a) Foreign sales



(b) Foreign sales ratio

Figure 4. Integration time series across foreign sales and foreign sales ratio based quintile portfolios

Notes: Figure 4a and 4b show the equal-weighted integration time series across foreign sales and foreign sales ratio quintile portfolios, respectively. At the end of each June, we sort stocks into quartile portfolios based on the value of their foreign sales (FS) and foreign sales ratio (FSR) for the current fiscal year, only when both FS and FSR are positive. Stocks with zero value of FS or FSR are included in portfolio 0. We then calculate the equal-weighted integration estimate of stocks within each of these 5 portfolios. The 1st and 4th quartile portfolios denoted by "Low" and "High" correspond to stocks with positive low and high FS or FSR, respectively. The gray shaded areas correspond to Developed Markets recessionary periods as defined by NBER.

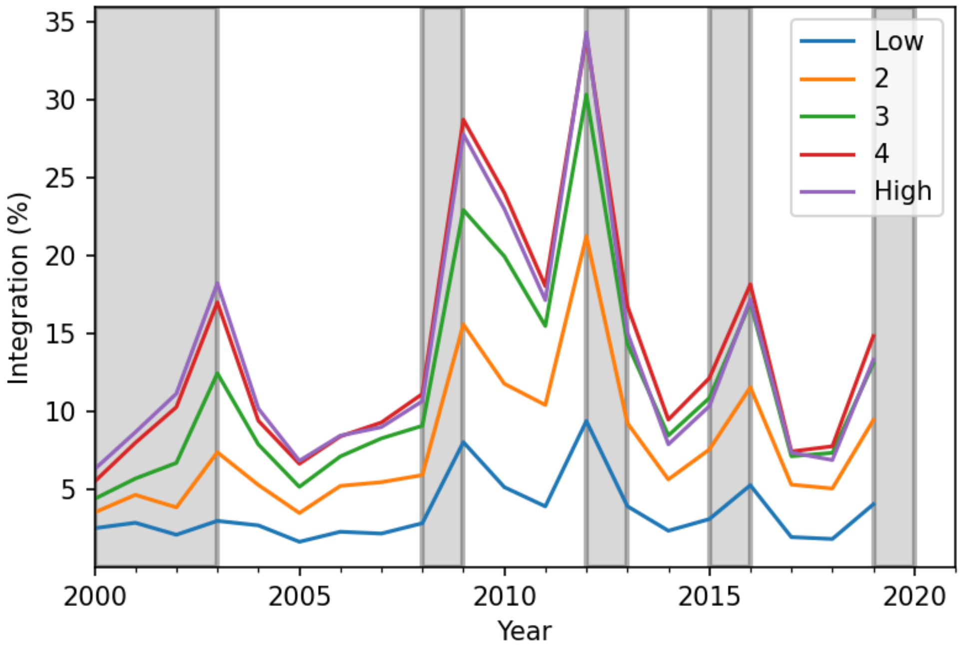


Figure 5. Integration time series across total institutional ownership based quintile portfolios

Notes: Figure shows the equal-weighted integration time series across IO portfolios. At the end of each June, we sort stocks into quintile portfolios based on the value of their total institutional ownership (IO) at June of the current year. We then calculate the equal-weighted integration estimate of stocks within each of those 5 portfolios. The 1st and 5th quintile portfolios denoted by "Low" and "High" correspond to stocks with low and high IO values. The data spans June 1999 to June 2019. The gray shaded areas correspond to Developed Markets recessionary periods as defined by NBER.

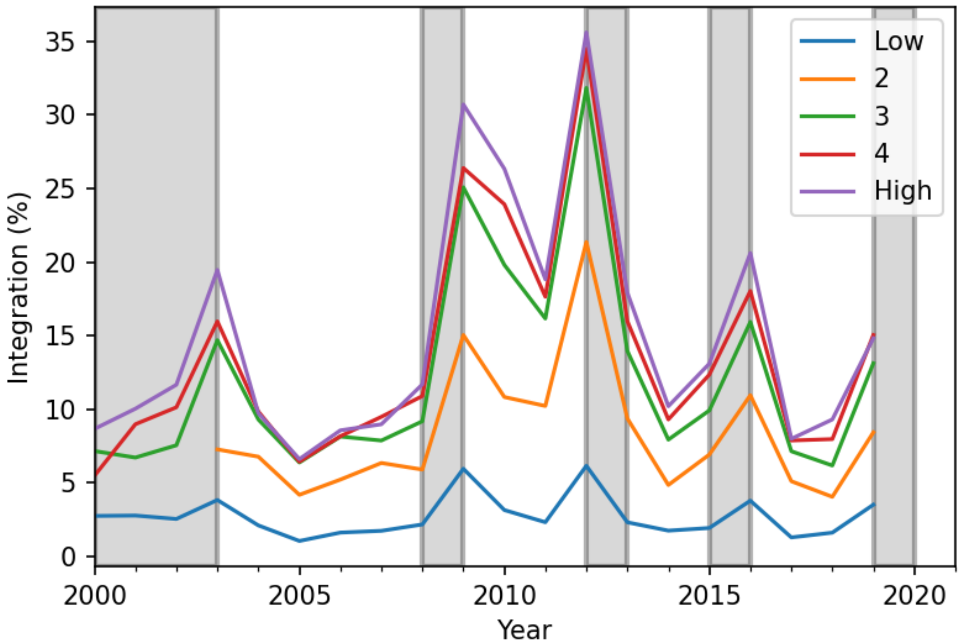


Figure 6. Integration time series across foreign institutional ownership based quintile portfolios

Notes: Figure shows the equal-weighted integration time series across FIO portfolios. At the end of each June, we sort stocks into quintile portfolios based on the value of their foreign institutional ownership (FIO) at June of the current year. We then calculate the equal-weighted integration estimate of stocks within each of those 5 portfolios. The 1st and 5th quintile portfolios denoted by "Low" and "High" correspond to stocks with low and high FIO values. The data spans June 1999 to June 2019. The line of portfolio No2 in Figure 6 is discontinued because there is not enough variation of FIO for the calculation of proper quintiles. The gray shaded areas correspond to Developed Markets recessionary periods as defined by NBER.

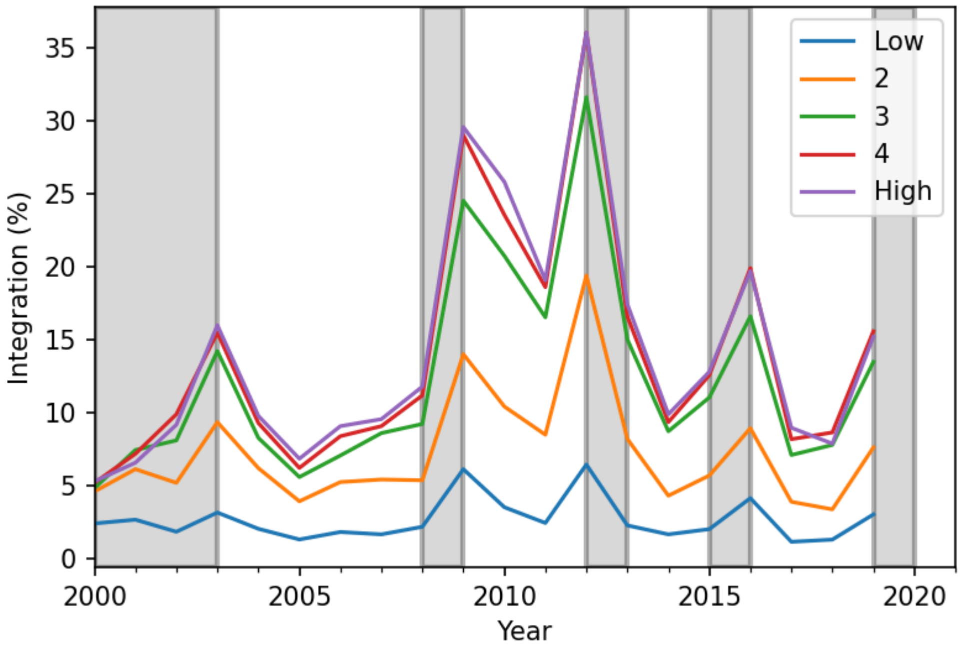
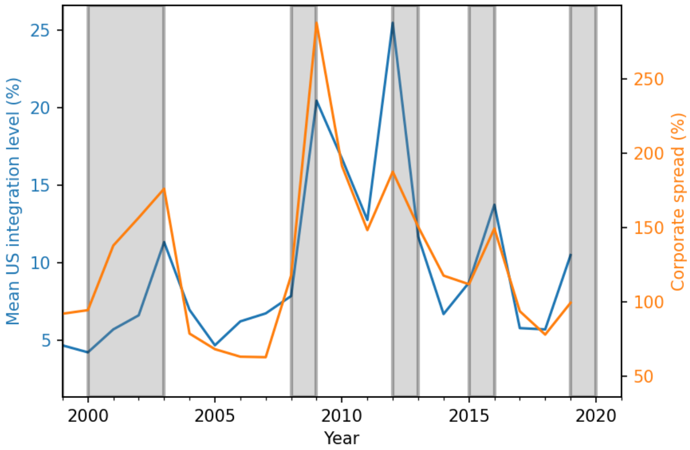
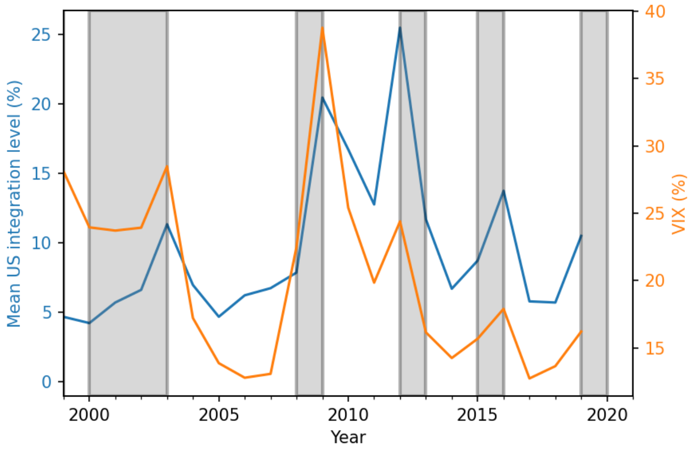


Figure 7. Integration time series across foreign common institutional ownership quintile portfolios

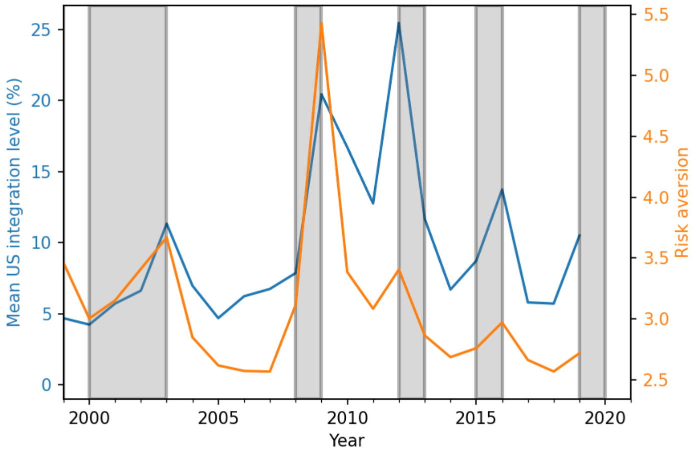
Notes: Figure shows the equal-weighted integration time series across FCO_mean portfolios. At the end of each June, we sort stocks into quintile portfolios based on the value of their foreign common institutional ownership (FCO_mean) at June of the current year. We then calculate the equal-weighted integration estimate of stocks within each of those 5 portfolios. The 1st and 5th quintile portfolios denoted by "Low" and "High" correspond to stocks with low and high FCO_mean values. The data spans June 1999 to June 2019. The gray shaded areas correspond to Developed Markets recessionary periods as defined by NBER.



(a) Corporate spread



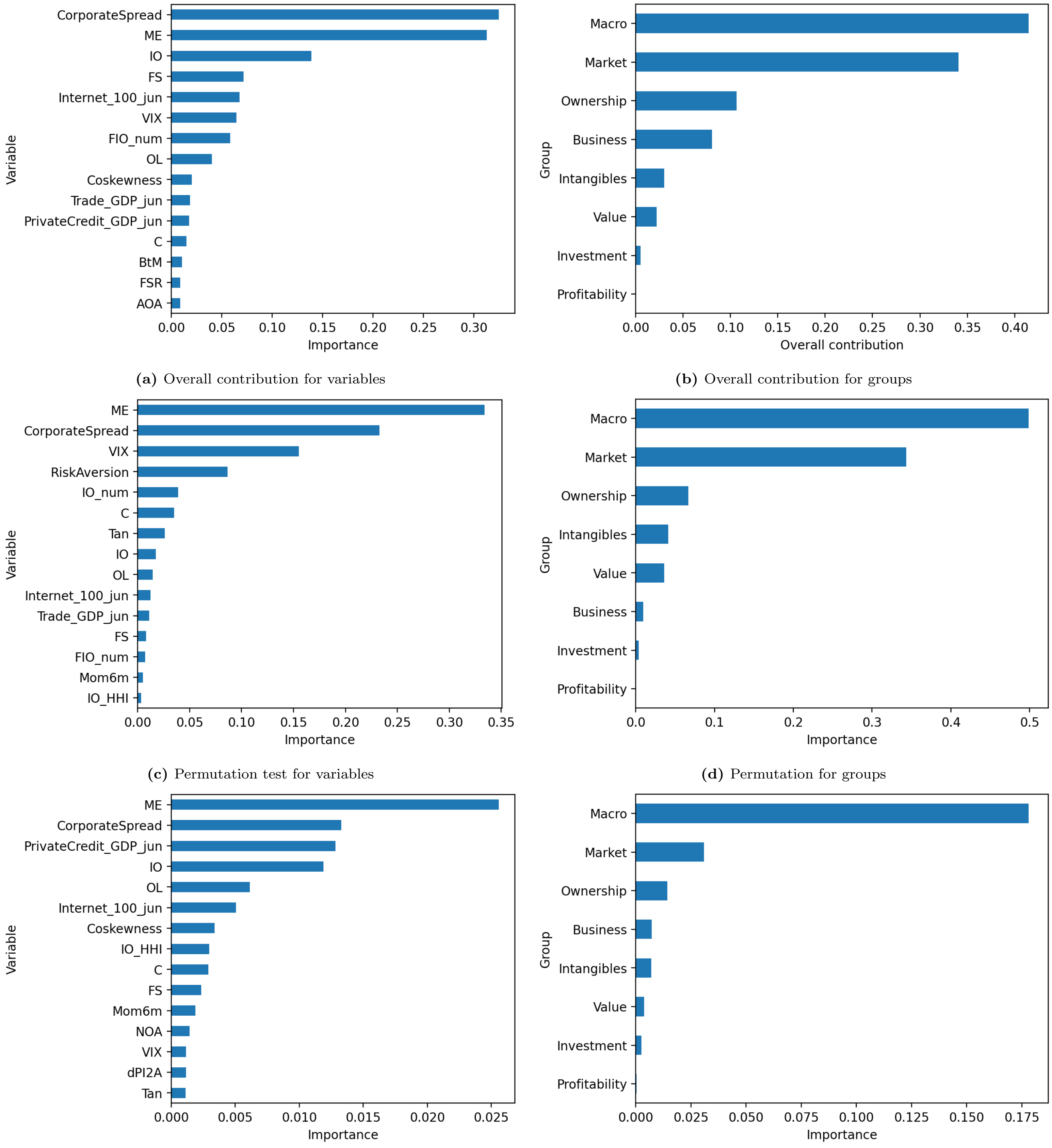
(b) VIX



(c) Risk aversion

Figure 8. Time series of mean US integration level with corporate spread, VIX and risk aversion

Notes: Figure plots the equal-weighted integration time series against corporate spread, VIX and risk aversion. Corporate spread is the spread between corporate US BAA and AAA bonds in a year. VIX is the option volatility index of CBOE. Risk aversion is the direct measure of risk-aversion in the US of Bekaert et al. (2021). The gray shaded areas correspond to Developed Markets recessionary periods as defined by NBER.



(a) Overall contribution for variables

(b) Overall contribution for groups

(c) Permutation test for variables

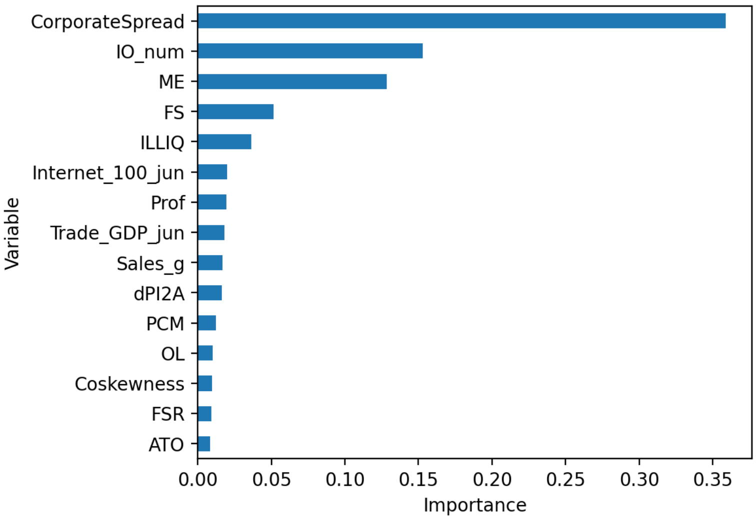
(d) Permutation for groups

(e) Change in R2 for variables

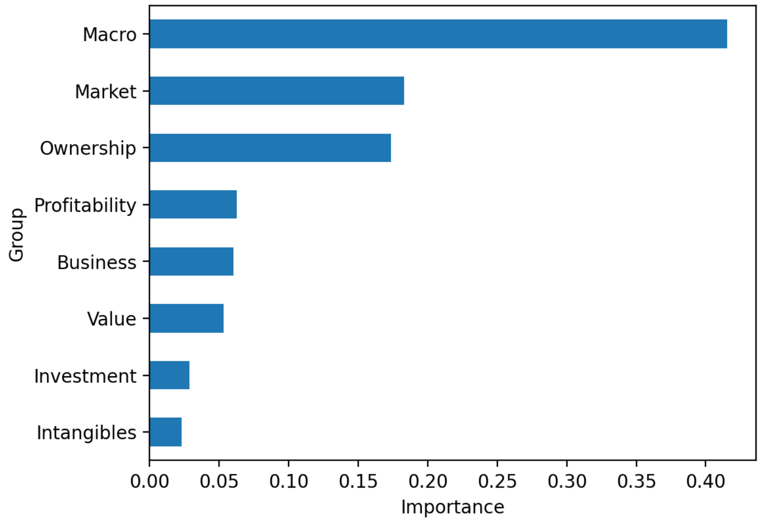
(f) Change in R2 for groups

Figure 9. Importance of determinants of firm-level integration in GETS

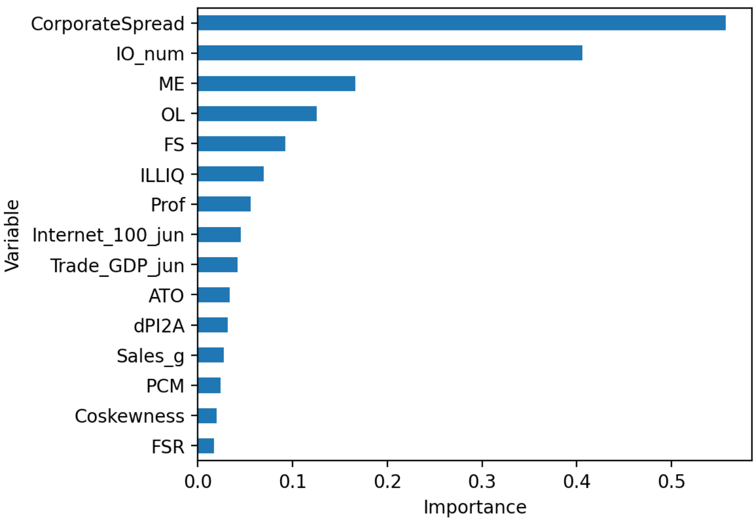
Notes: Figure shows the importance of individual variables and their groups in the GETS model based on the measures of importance of Section 3.1.3. Figures 9a and 9b plot the overall contribution, Figures 9c and 9d the results of the permutation test and Figures 9e and 9f the change in R2 for variables and groups, respectively. For brevity, we only report the 15 most important individual variables.



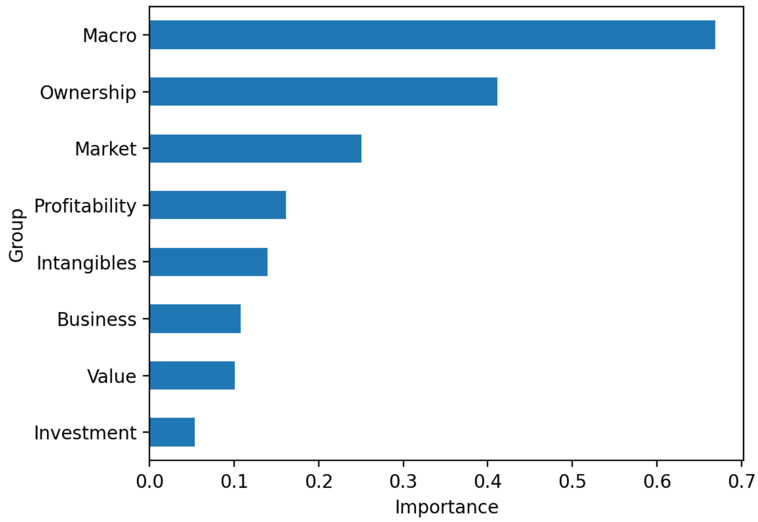
(a) Permutation for variables



(b) Permutation test for groups



(c) Change in R2 for variables



(d) Change in R2 for groups

Figure 10. Importance of determinants of firm-level integration in RFR

Notes: Figure shows the importance of the individual variables and their groups in the RFR based on the permutation test and the change in R2. In Figures 10a and 10c we plot the 15 most important variables.

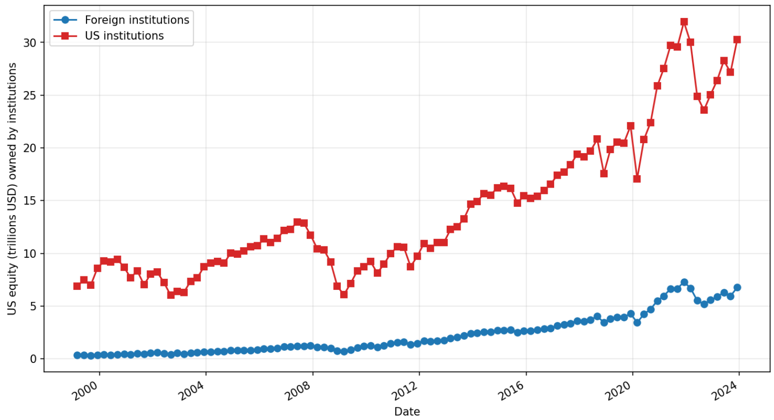
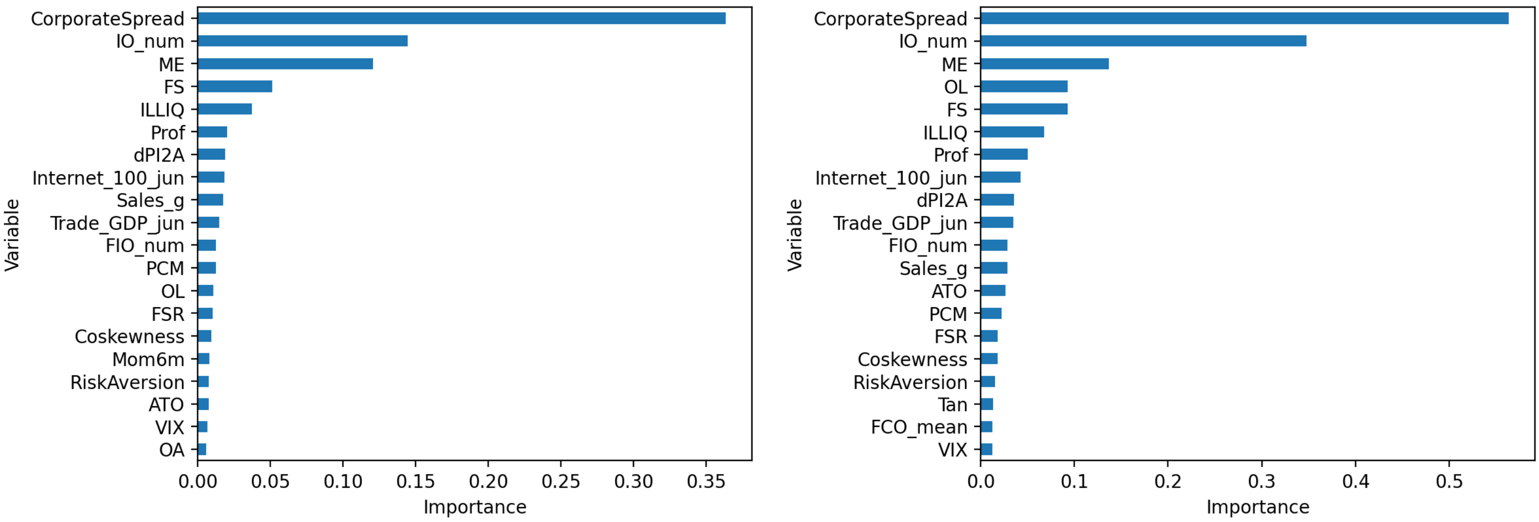


Figure 11. US market capitalization in trillions USD owned by foreign (non-US) and US institutional investors

Notes: Figure shows the market capitalization (trillions USD) of the US stock market that foreign and US institutional investors held from 1999Q1 to 2023Q4.

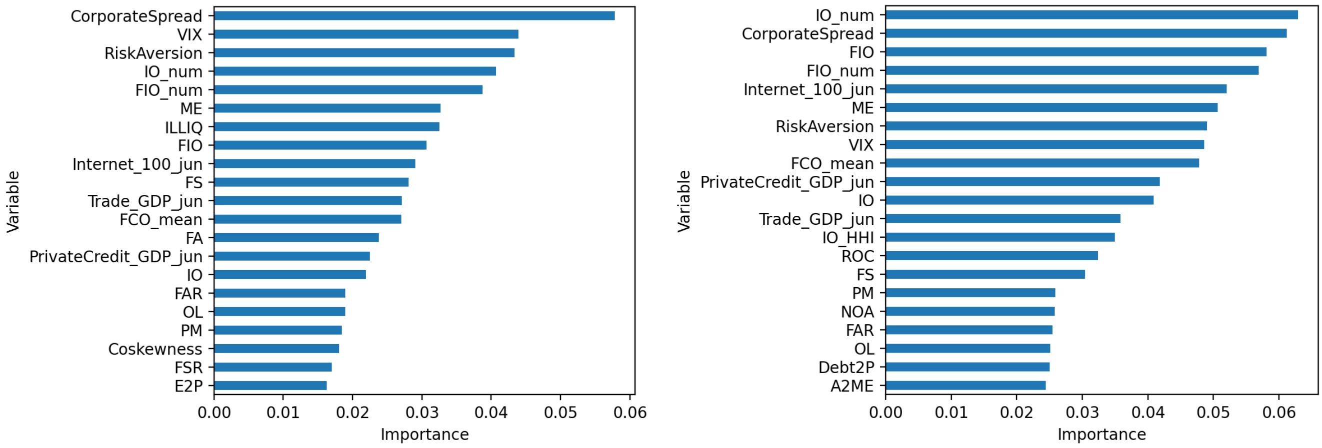


(a) Permutation for variables

(b) Change in R2 for variables

Figure 12. Importance of determinants of firm-level integration in RFR - Probability to default

Notes: Figure shows the importance of the individual variables and their groups in the RFR based on the permutation test and the change in R2. In Figures 12a and 12b we plot the 15 most important variables. We have included probability to default as a firm-level measure of distress.



(a) Permutation for variables

(b) Change in R2 for variables

Figure 13. Importance of determinants of firm-level integration in RFR - Outward foreign direct investment

Notes: Figure shows the importance of the individual variables and their groups in the RFR based on the permutation test and the change in R2. In Figures 13a and 13b we plot the 21 most important variables. We have included foreign assets (FA) and the foreign assets to total assets ratio (FAR) as a proxy for outward foreign direct investment of the firm.