Journal of Empirical Finance 61 (2021) 82–102

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Drivers of economic and financial integration: A machine learning approach[☆]

Amir Akbari [a,*], Lilian Ng [b], Bruno Solnik [c]

[a] *Ontario Tech University, 2000 Simcoe St N, Oshawa, ON L1G 0C5, Canada*
[b] *Schulich School of Business at York University, 4700 Keele Street, North York, M3J 1P3, Ontario, Canada*
[c] *HEC Paris and Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong*

ARTICLE INFO

JEL classification:
F15
F30
G15
E44

Keywords:
Determinants of market integration
Random Forest Regression
Machine learning

ABSTRACT

We propose a new approach to identifying drivers of economic and financial integration, separately, and across emerging and developed countries. Our advanced machine learning technique allows for nonlinear relationships, corrects for over-fitting, and is less prone to noise. It also can tackle a large number of highly correlated explanatory variables and controls for multicollinearity. Results suggest that general economic growth, increasing international trade, and contained population growth have helped emerging countries catch up to the level of the economic integration of developed countries. However, slow financial development and a high level of investment riskiness have hindered the speed of emerging countries' financial integration. Furthermore, the results suggest that integration is a gradual process and is not driven by cyclical or transitory events.

1. Introduction

There is now substantial evidence that countries worldwide have become increasingly integrated in the past decades, but that the pace of integration varies across developed (DEV) and emerging (EMG) countries.[1] A recent study by Akbari et al. (2020) (hereafter ANS, 2020) offers new significant insights on the dynamics of integration across the two types of markets and through time. ANS propose a simple metric that allows them to disentangle the two forms of integration (i.e., financial and economic integration). They employ the smooth-transition dynamic conditional correlation (STDCC) specification to analyze short- and long-term dynamics of integration using a sample of 39,202 firms from 41 countries worldwide. Their study shows that the levels of both forms of integration have increased across all countries since the start of their 1989–2015 sample period but have fallen after the global financial crisis and that DEV countries are more financially and economically integrated than their EMG counterparts. Since the global crisis, the gap in economic integration between EMG and DEV countries has narrowed dramatically and then converged toward the end of their sample period, whereas their financial integration gap remains fairly stable. While ANS's results are interesting and insightful, their study offers no further analysis of the plausible factors explaining the phenomena they have documented. The purpose of this study is to address this important issue.

[☆] We thank anonymous referees and participants at the Asian-Finance Association Meetings in Tokyo, Conference of Asia-Pacific Financial Markets in Seoul, China International Conference in Finance in Tianjin, China, Finance Down Under in Melbourne, Financial Management Association (FMA) Meetings in San Diego, FMA-Asia Pacific Conference in Hong Kong, Hong Kong University of Science and Technology, and Northern Finance Association (NFA) Meetings for many helpful comments and suggestions.

[*] Corresponding author.
E-mail addresses: amir.akbari@ontariotechu.ca (A. Akbari), lng@schulich.yorku.ca (L. Ng), solnik@ust.hk (B. Solnik).

[1] See, for e.g., Bekaert and Harvey (1995); Phylaktis and Ravazzolo (2002); Carrieri et al. (2007); Chambet and Gibson (2008); Bekaert et al. (2009); Pukthuanthong and Roll (2009); Bekaert et al. (2011); Carrieri et al. (2013) and Eiling and Gerard (2015).

https://doi.org/10.1016/j.jempfin.2020.12.005
Received 30 April 2020; Received in revised form 15 October 2020; Accepted 28 December 2020
Available online 13 January 2021

Given that theory does not predict the channels through which countries are integrated with the world market, existing studies typically resort to selecting an exhaustive list of macro variables that serve as proxies for the country and global factors that can explain a country's level of integration with the world market. Using different variable proxies and methodologies employed, some researchers show that a country's regulatory policies affect the level of its world market integration, while others find that a country's financial openness, financial liberalization, and economic advancements explain the evolution of the country's degree of integration. We relegate the review of this literature to Section 3.2 below.

One concern with prior studies is that their multivariate analyses are based on a large number of usually highly correlated variables and hence may give rise to the multicollinearity problem. More critically, using a broad set of correlated variables in regression models not only yields inconsistent and inefficient coefficient estimates but also poses challenges in the interpretation of these regression-based estimates. Standard econometric methods cannot deal with such a large set of proxy variables. One must typically engage in two steps by first preselecting a small number of variables and then investigating their relationships with the integration metric. To circumvent the multicollinearity and dimensionality problems, our study employs a powerful statistical tool, namely, the random forests regression (RFR) technique. RFR, initially introduced in Breiman (2001), is an ensemble machine learning method in the context of a multitude of decision trees. In a given tree, the RFR technique implements a series of piece-wise linear relations between candidate variables and the economic or financial integration metric. This technique allows for nonlinear and complex relationships between explanatory and dependent variables over the whole sample. Gu et al. (2020) find that the nonlinearities in RFR, especially in the form of complex interactions among explanatory variables, substantially improve predictions in their study over traditional regression models. The Breiman (2001) bootstrapping procedure is used to compute and rank the importance of each candidate variable. The variable importance measure is calculated by averaging the difference in out-of-sample errors before and after the permutation over all trees. Overall, RFR accommodates a more general form of relationships, including nonlinear relationships, between dependent and independent variables. It also corrects for over-fitting, which may result from a large set of explanatory variables often employed in determining integration drivers. Its advantages over existing variable-selection methodologies (e.g., the jackknife methodology and general-to-specific search algorithm) are that it takes into account the multicollinearity of variables and does not eliminate variables solely based on their level of statistical significance.

We exploit the RFR technique to examine plausible drivers of market integration on a broad cross-section of 21 DEV and 20 EMG markets worldwide for the 1989–2015 sample period. Our analysis employs a fairly exhaustive list of 30 variables, mainly drawn from the existing empirical literature.[2] Results show that variables that influence economic integration are quite different from those that affect financial integration. We find that a country's economic development plays the most important role in explaining economic integration. Economic development accounts for 45% of the time-series and cross-country variation in economic integration. Among the various proxies for economic development, GDP per capita is the strongest determinant of economic integration, followed by population growth rate. Information/openness and international trade are the next most important determinants of economic integration. Results for financial integration, however, differ markedly. The most significant determinants are the proxies for financial development, which contribute to 42% of the time-series and cross-country variation in financial integration. The size of a country's stock market, a proxy for the country's capital market development, plays a crucial role in explaining financial integration variation. The next most important determinants are the number of Internet users and the country's investment profile, a proxy for the risk of expropriation. The investment profile proxy allows investors to evaluate the investment riskiness, specifically in areas of expropriation, profits repatriation, and payment delays. Our finding implies that the development of information technology, access to information, and Internet investors' savvy influence the globalization of financial markets. Finally, our analyses suggest that cyclical variables do not significantly explain the dynamics of market integration in our sample. None of these variables has an importance score of above 2%, suggesting that short-term cyclical events have virtually no immediate and drastic effect on the global integration of economies and capital markets. Overall, our findings indicate that integration is a gradual process driven mainly by fundamental economic and financial variables rather than cyclical or transitory events.

Our study makes an important contribution to the existing literature on market integration. It represents the first to investigate the underlying country and global characteristics that can explain each form of integration and to provide insights on the varying integration gaps between EMG and DEV countries and through time. Our research vastly contrasts with prior literature that mostly focuses on the determinants of global market integration or of one form of integration, namely, financial integration. We find that cross-country differences in regulatory policies, institutional constraints, and the information environment drive the risk-pricing differences between these markets. In contrast, the rapid advancement in EMG countries' economic development, especially after the global financial crisis, has helped to speed up the pace of their economic integration with the world markets and thereby has narrowed their economic integration distance from that of their DEV peers.

Our work also expands the literature on the application of advanced machine learning techniques in empirical research in finance. For example, Khandani et al. (2010) use classification and regression trees to model consumer credit risk. Feng et al. (2020), Freyberger et al. (2020), and Kelly et al. (2019) propose modeling approaches, such as instrumented principal component analysis and adaptive group LASSO, to explain the cross section of average equity returns. Gu et al. (2020) perform a comparative analysis of machine learning methods for measuring risk premiums of stocks. They highlight the advantages of these methods over the linear regression-based technique in identifying factors that predict stock returns. Studying more than 30,000 firms from 1956 to 2016, they find the RFR as one of the best-performing methods, with highest predictive power.[3] We advance this line of research by applying the RFR technique to circumvent the multicollinearity issue in determining drivers of market integration in the international setting.

[2] See Section 3.2 for the details.

[3] We refer interested readers to Gu et al. (2020) for the comparative performance analysis of the other machine learning models.

The paper is organized as follows. Section 2 introduces the RFR technique. Section 3 describes the construction and estimation of economic and financial integration measures. Section 4 studies the channels through which country- and world-level variables influence economic and financial integration, and Section 5 concludes.

2. The random forests regression

In this section, we begin by describing the RFR technique that we employ in determining the variables that explain the dynamics of economic and financial integration between DEV and EMG markets, followed by a discussion of the advantages of using RFR over the approaches employed in the existing literature.[4]

2.1. RFR technique

RFR, initially introduced in Breiman (2001), is a powerful statistical tool that helps us overcome some of the regression-based variable selection procedure's pitfalls. RFR is an ensemble machine learning method that maximizes the information entropy in the context of a multitude of decision trees to evaluate the importance of the explanatory variables. The implementation of RFR involves the following steps. In the first step, a general form of a relationship is fitted between dependent and explanatory variables through a decision tree that involves fitting a series of piece-wise linear relationships between the two types of variables in subsamples of the data. In the next step, to reduce the variance of the fitted values, RFR implements a bootstrapping procedure on an ensemble of trees, hence the name *random forest*. We grow a large number of these decision trees. The fitted values for the dependent variable are estimated by averaging the fitted values of a random selection of the decision trees. Finally, RFR assigns an importance score for each of the explanatory variables by measuring the output's sensitivity to changes in that explanatory variable. Below, we describe these steps more formally.

2.1.1. Decision trees

In the first step, the explanatory variables are divided into homogeneous subsamples by recursive splitting. Then in each subsample, a piece-wise linear relationship is fitted between the dependent and explanatory variables. This relationship forms a decision tree.[5,6] The piece-wise linear relationships allow the decision tree to accommodate a more general form of relationships, $f_0(.)$, between dependent and explanatory variables throughout the whole sample:

$$\begin{aligned} y &= f_0(X) + u, \\ \mathbb{E}[u|X] &= 0, \\ u &\sim iid \end{aligned} \tag{1}$$

In our study, the dependent variables y is a vector of $M \times 1$ observations of economic or financial integration, across 21 DEV and 20 EMG countries for 27 years, from 1989 to 2015. That is, in our sample, there are $M = 41 \times 27$ country-year observations. The explanatory variables X is a matrix of $M \times K$ potential determinants of market integration. Our sample has $K = 30$ plausible explanatory variables for market integration (below, each explanatory variable in X is denoted with x_k). Countries and explanatory variables are described in Section 3.2.

To illustrate how a decision tree with several layers and branches is formed, in Fig. 1, we plot a schematic of a decision tree with three explanatory variables, $X = [x_1, x_2, x_3]$, and with a maximum of three layers. This is a simple example of one possible tree. In this example, the sample is split iteratively into seven subsamples (regions): R_1 to R_7. We do this by first dividing the sample based on the explanatory variable x_2 at the threshold level of tl_1. We next divide the subsample with $x_2 \le tl_1$, i.e., the left branch of the tree, into two subsequent subsamples based on x_3 at the threshold level of tl_2. For the sample with $x_2 > tl_1$ (the right branch of the tree), we divide the subsample based on x_1 at the threshold level of tl_3. We continue splitting the subsamples at the third branch in a similar manner. These steps partition the sample into seven non-overlapping regions, which are also known as leaf nodes of the tree. For example, region R_4 refers to the subsample where $x_2 > tl_1$ and $x1 \le tl_3$ and $x_2 \le tl_6$. In forming the decision tree, we may choose a smaller layer for some regions. For example, for region R_3, we divide the sample only twice.

RFR chooses the threshold levels tl_i optimally such that the observations in regions R_j are homogeneous. We discuss this process in Section 2.1.2. Once the homogeneous regions are identified, in each region, the dependent variables are fitted as their average in that region, conditional on being in that branch. This identifies the output of the linear decision tree in each region. The collection of these relationships over subsamples forms the decision tree.

More formally, a decision tree is a set of piece-wise linear relationships between the dependent and explanatory variables, in homogeneous subsamples of the input data. A decision tree, is identified by $T(X;\Theta)$, where Θ is the vector of tree parameters, including splitting the threshold levels as well as the parameters of piece-wise linear relationships in each region. Therefore, we denote the fitted values of the decision tree in Eq. (1) as $\hat{y} = f_0(X) = T(X;\Theta)$.

The collection of these "local" linear relationships over subsamples allows for nonlinear and complex relationships between the dependent and explanatory variables. Section 4.4 provides a visual example to illustrate how RFR implements non-linearity. Gu et al. (2020) find that nonlinearities in RFRs, especially in the form of complex interactions among explanatory variables, substantially improve predictions in their study over regression models.

[4] For recent analyses of theoretical properties of RFR models, please see Biau (2012), Wager et al. (2014), Scornet et al. (2015), Mentch and Hooker (2016), and Wager and Athey (2018).

[5] These subsamples are also known as "regions", which should not be misunderstood as a group of countries in our application.

[6] This step is similar to a quantile regression, albeit with more than four subsamples, which are optimally chosen to be homogeneous.

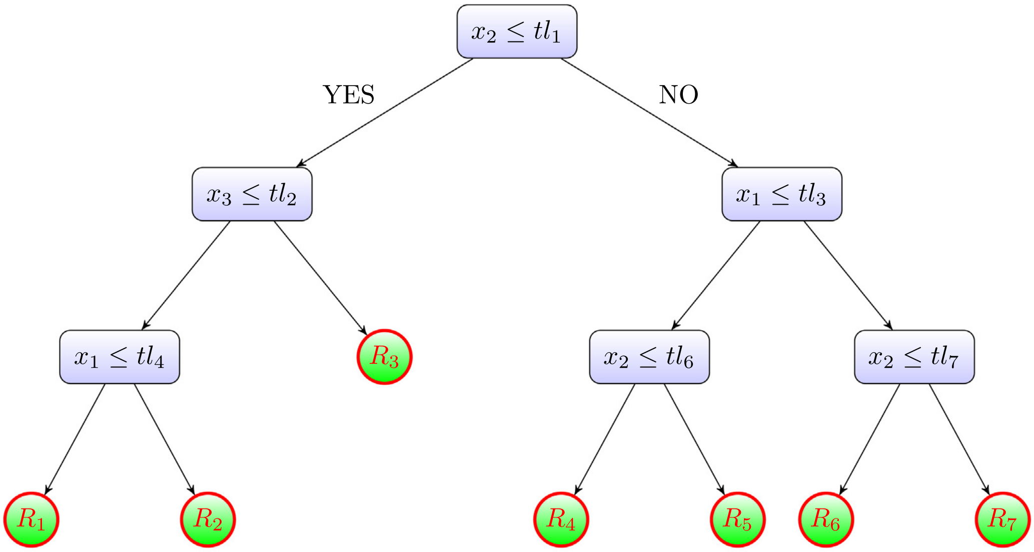

Fig. 1. Schematic of a decision tree: The figure shows an example of a decision tree with 3 explanatory variables: $X = [x_1, x_2, x_3]$. The splitting of the sample is done at threshold levels tl_1 to tl_7, which result in regions R_1 to R_7. In each region, the decision tree fits a relationship between the dependent variable, y, and explanatory variables, X.

2.1.2. Homogeneous regions

RFR chooses the optimal threshold levels, tl, to split the sample into homogeneous regions. This step is critical, because it helps with the performance and fitness of the piece-wise linear relationships in each region. The optimal (homogeneous) region's choice is made using the recursive binary split approach, based on minimizing the mean squared error (MSE) in the adjacent regions. More specifically, for the kth explanatory variable (shown as x_k), we find the splitting point, s, such that:

$$\min_{s} \quad \Big[\text{MSE}\left(y|x_k < s\right) + \text{MSE}\left(y|x_k \geq s\right)\Big]. \tag{2}$$

We repeat this procedure for all explanatory variables. At each branch of the decision tree, the variable k and the corresponding splitting point s that yield the lowest MSE are chosen. In the next step, we partition the sample into two subsamples based on x_k and the optimal threshold level $tl = s$. The least MSE approach ensures that any other partitioning based on other explanatory variables or based on different threshold levels will result in less homogeneous subsamples.

2.1.3. Random forest

To reduce the model output's variance, Breiman (2001) suggests a bootstrapping procedure on an ensemble of trees, hence the name, *Random Forest*. The procedure involves growing B trees on subsets of the input dataset and averaging each tree's outputs. Therefore the fitted values for the Random Forest, $\hat{y} = f_{rf}(X)$ is:

$$\hat{y} = f_{rf}(X) = \frac{1}{B}\sum_{b=1}^{B} T(X;\Theta^b) \tag{3}$$

where $T(X;\Theta^b)$ denotes the decision tree model for the bth tree, and $b \in \{1,\dots,B\}$.

If all the trees in the forest are uncorrelated, the variance of $\hat{y}$ depreciates with a reciprocal of forest size, B. In this case, by choosing a sufficiently large value for B, RFR increases the precision of the fitted values, without increasing the estimation bias. To alleviate the effect of correlation between each sampled tree, $T(X;\Theta^b)$, and the rest of the forest, the RFR approach involves drawing a random sample, Z, from the input data to form a training dataset and then selecting $m \leq K$ of the explanatory variables at random, prior to each split. This random sampling and averaging procedure reduce the model's sensitivity to noise and outliers. Excluding part of the data and the explanatory variables at each tree corrects for the over-fitting problem.[7]

In our analysis, we follow Geurts et al. (2006) in setting the hyper-parameters of the RFR. These parameters also govern how RFR evaluates importance measures for the training set. Specifically, we choose $m = K$, i.e., 30. We also set the sample size (number of observations in Z) equal to two-thirds of the full sample size (i.e., 738). Recall, our sample includes a panel of 21 DEV and 20 EMG markets from 1989 to 2015, resulting in 1107 annual observations. We choose to grow $B = 1000$ trees in each forest for measures of economic and financial integration. Finally, we have to decide on the potential depth of a tree and how fine is the splitting process.

[7] The over-fitting problem occurs in models that explain too closely or exactly a particular set of data (i.e., the training set), and fails to explain additional data (i.e., the test set) reliably. This results in poor out-of-sample prediction of the estimation method.

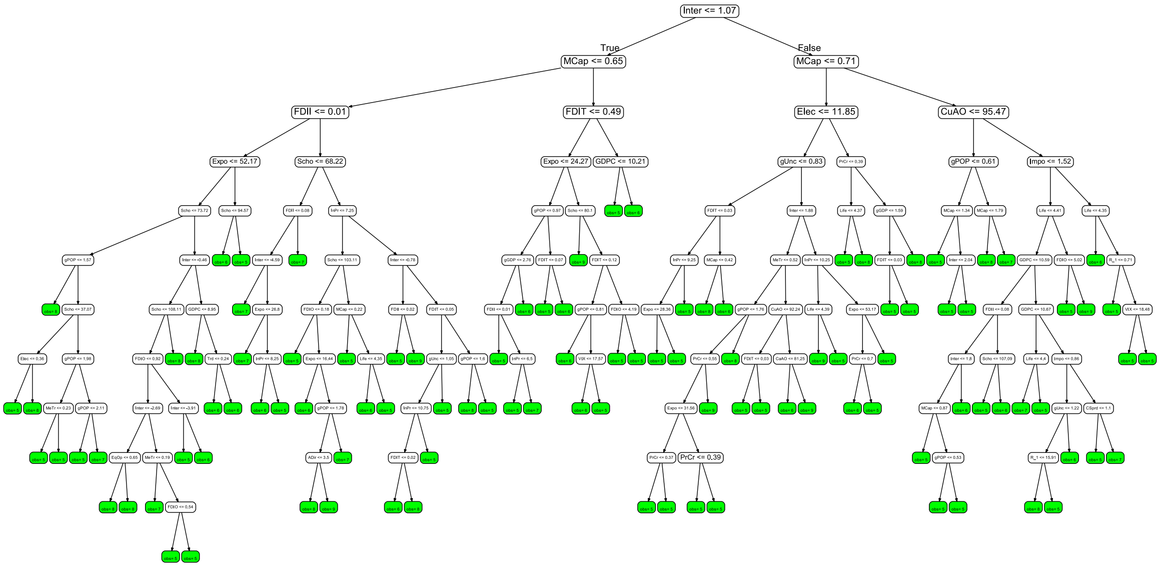

Fig. 2. A tree from the forest: The figure shows one of the 1000 trees of the forest for the drivers of the economic integration measure. At each node, it shows the chosen explanatory variable and the splitting threshold level. At each leaf node (in green), it shows the number of observations in that leaf node. For ease of reference, we shorten some of the variable names. *Economic Development Proxies*: Internet (Inter), GDPC (GDPC), Electricity (Elec), School (Scho), gPopulation (gPop), Life (Life); *Information/Openness Proxies*: Investment Profile (InPr), Anti-Director (ADir), Capital Account Openness (CaAO), Current Account Openness (CuAO), Financial Openness (FiOp), Law & Order (Law), Equity Mkt Openness (EqOp), IFRS (IFRS), Trade Openness (TrOp); *Financial Development Proxies*: Market Cap (MCap), Private Credit (PrCr); *International Trade*: Exports (Expo), Imports (Impo), Merchandise Trade (MeTr), Trade (Trd); *Foreign Direct Investment*: FDI Inflow (FDII), FDI Outflow (FDIO), FDI Total (FDIT); *Cyclical Proxies*: gGDP (gGDP), gUncertainty$_w$ (gUnc), R$_{-1}$ (R$_{-1}$), gGDP$_w$ (gGDP$_w$), Credit Spread (CSprd), VIX (VIX). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A deep tree will have fewer observations in the leaf nodes to compute the conditional fitted value of the dependent variable. In our implementation, we stop the splitting process either when further splitting does not yield to an improvement in MSE or when the number of observations in the leaf node reaches five.[8] Fig. 2 presents one of these trees for the measure of economic integration, as an example.

In Fig. 2, the nodes in white show the chosen explanatory variable and the splitting threshold level, based on the MSE criteria introduced in (2). The nodes in green (the leaf node of the tree) identify the homogeneous regions. In each green node, we report the number of observations. In this tree, the sample is divided in 104 regions.

The main analysis of the paper focuses on the variable importance. Therefore, except in Section 4.4, we do not consider a test set. In that section, we focus on predicative power of RFR and we compare the out-of-sample performance of RFR with LASSO and Ridge regressions. Please refer to Section 4.4 for details on training set and test set in these experiments.

2.1.4. Variable importance

Once we train our model, i.e., estimated the tree parameters, Θ, of all trees in the forest and computed the fitted values, $\hat{y}$, we can estimate the importance score for each of the explanatory variables. The premise is that the fitted values show the largest sensitivity with respect to changes in the most important variable. Therefore, to measure the importance of the kth variable, Breiman (2001) suggests computing the difference in prediction accuracy before and after permuting the explanatory variable x_k. Thus this approach is known as "Mean Decrease Accuracy" method.

More specifically, after estimating the tree parameters, the values of x_k are permuted among the training data. That is, only the values of the kth column of X are shuffled randomly such that its ith row is $(x_{i,1}, \ldots, x_{i,k-1}, x_{\pi_k(i),k}, x_{i,k+1}, \ldots, x_{i,K})$.[9] Then the out-of-sample error is computed similar to (3) on this perturbed data set. The variable's importance measure is calculated by averaging the difference in out-of-sample errors before and after the permutation over all trees. More formally, let $\mathbb{Z}^b$ be the out of sample (out-of-bag sample) for the bth tree of the forest. Then the importance score for the variable x_k in this tree is:

$$VI(x_k)^b = \frac{\sum_{\mathbb{Z}^b} y - \hat{y}^b}{|\mathbb{Z}^b|} - \frac{\sum_{\mathbb{Z}^b} y - \widehat{y_{\pi_k}}^{\,b}}{|\mathbb{Z}^b|} \tag{4}$$

where $|\mathbb{Z}^b|$ is the size of the out-of-bag sample, and $\widehat{y_{\pi_k}} = f(X_{\pi_k}, \Theta^b)$ is the fitted values after permuting x_k in the bth tree. The goal is to assess how much the permutation decreases the accuracy of the model. The permutation of unimportant variables has statistically

[8] Five or more observations per leaf node is a commonly used hyper-parameter in RFR research. Our results are robust to choosing a larger number of observations, as seen in Section 4.3.

[9] See Strobl et al. (2008) for a study of variable importance in random forests using conditional permutation.

no effect on the model accuracy, while permuting important variables significantly decreases it. The raw variable importance score for x_k is then computed as the average measure over all trees:

$$VI(x_k) = \frac{1}{B}\sum_{b=1}^{B} VI(x_k)^b \tag{5}$$

For comparison purposes, the measure is scaled by the standard deviation, $\frac{\hat{\sigma}}{\sqrt{B}}$, and reported in percentage. A variable importance close to zero indicates that the contribution of the variable x_k to the predictive accuracy of the estimation is negligible.

An alternative method for variable importance is the "Mean Decrease Impurity" approach. In a random forest, impurity is defined as the objective function in Eq. (2). In this approach, the importance is computed based on how much each variable decreases the weighted impurity in a tree. For a forest, the impurity reduces from each variable that can be averaged, and the variables are ranked according to this measure. This approach highlights the information entropy of each explanatory variable in the estimation. In practice, most of the time, the two approaches rank the important variables similarly.

2.2. Advantages of RFR

At this juncture, it is worthwhile to discuss the advantages of our RFR technique over those employed in prior research in the market integration literature, including the jackknife methodology and general-to-specific search algorithm used (e.g., BHLS 2011). Given that theory offers no guidance on which financial and economic factors would drive market integration, researchers typically start from an exhaustive list of variable candidates. They then adopt some techniques such as a general-to-specific algorithm to reduce the list to a much smaller number amenable to regression analysis by eliminating variables with insignificant coefficient estimates through multiple runs of regressions. Based on the resulting model with mainly significant variables, they conclude which variables are important in explaining global integration. However, many of these variables are highly correlated, and hence, this multicollinearity issue may mask some variables' importance. The pre-selection procedure, therefore, biases these statistical tests. Similarly, the jackknife methodology is also based on introducing randomness in the estimation process. It randomly selects candidate variables to determine whether or not they are statistically significant in multivariate regressions.

In contrast, our machine learning approach is based on maximizing information entropy and not minimizing the p-value of a regression coefficient. This method is robust to the inclusion of irrelevant or redundant explanatory variables, and the multicollinearity problem arises in panels with highly correlated explanatory variables. Also, RFR allows for nonlinear relationships between the explanatory and dependent variables, which most likely is the case for the drivers and market integration measures in the panel of DEV and EMG markets. Moreover, by implementing the bootstrapping approach, RFR achieves better precision by reducing variance through averaging the prediction of orthogonal trees. This method reduces the model's sensitivity to noise and outliers, challenging the regression-based analysis in international empirical research. Lastly, RFR also corrects for over-fitting, resulting from a broader set of explanatory variables, often observed in other similar methods, such as least squares. The bootstrapping approach in RFR, which selects $m \leq K$ of the explanatory variables at random, prior to each split, alleviates the over-fitting concern that arises from including potentially more parameters in the model than can be justified by the input data.

One should note that the RFR methodology cannot address causality between the explanatory and dependent variables but can provide evidence of relationships. This issue is common to all previous research looking at the relationship between integration and explanatory variables with regression models. To provide direct and robust evidence of causal linkages between the determinants and market integration metrics would require a (quasi)-natural experiment. Such experiments could adequately control for other factors affecting market integration but require identifying, for each determinant or group of determinants, several shocks that only affect a (large) group of countries but not the rest of the countries. This is an impossible task. In the absence of such sophisticated experiments, our contribution is to provide a more statistically sound approach over existing methodologies to identify the links between market integration and economic variables.

3. Data and variable construction

In this section, we describe the construction of all variables employed in the study and the sources of information that we use to construct them. Specifically, we construct the measures of economic and financial integration employed in ANS (2020) and a set of plausible explanatory variables for economic and financial integration, as drawn from the extant literature.

3.1. Measures of economic and financial integration

We strictly follow the ANS's (2020) approach to estimate the time-varying measures of economic and financial integration. We first compute firm-level cash flow revisions and risk pricing revisions and then employ these return components to estimate measures of economic and financial integration using a smooth-transition dynamic conditional correlation (STDCC) specification.

We decompose a firm's rate of equity return, R_t, into cash flow revision (CF_t) and risk pricing revision (RP_t). CF_t is estimated based on revisions in the present value of future expected dividends, $E_t d_{t+j}$:

$$PV_t = \sum_{j=1}^{\infty} \frac{E_t d_{t+j}}{(1 + r_{f,t+j})^j}, \tag{6}$$

$$CF_t = \frac{PV_t - PV_{t-1}}{P_{t-1}}, \tag{7}$$

where P_t is the price of asset at time t and $r_{f,t+j}$ is the term structure of riskfree rates.[10] The risk pricing (RP_t) revisions are computed using $RP_t = R_t - CF_t$, where R_t is the return of asset from period $t-1$ to t. For every t, we construct a value-weighted average of CF_ts and a value-weighted average of RP_ts of all available firms within a country as proxies for the country's cash flow revisions ($CF_{c,t}$) and risk pricing changes ($RP_{c,t}$), respectively. In a similar manner, we also construct the global market-weighted cash flow ($CF_{w,t}$) and risk pricing ($RP_{w,t}$) components.

We estimate the cash flow and risk pricing revision components at the firm-level for a cross-section of 21 developed market (DEV) and 20 emerging markets (EMG). Our sample includes 39,202 firms; 28,411 of them are from DEV and 10,791 from 20 EMG markets. This sample of firms intersects the I/B/E/S and Datastream databases with non-missing earnings forecasts, payout ratios, and monthly returns information. To mitigate possible returns data errors, we also apply the filters suggested by Ince and Porter (2006). In addition, we winsorize firm-level CF and RP estimates before we aggregate them at the country level to reduce the effect of returns errors. The distribution of the number of firms in our sample is reported in Table 1 by type of markets and by country. There is limited information on firms from emerging markets at the start of our sample period, but information becomes increasingly available by mid 1990s.

The measure of a country's financial integration (R^2_{Fin}) is the square of the correlation between its own risk pricing revision RP_c and the world risk pricing component RP_w, whereas its economic integration measure (R^2_{Econ}) is given by the square of the correlation between its own cash flow revision CF_c and its world counterpart CF_w. We employ Ohashi and Okimoto's (2016) model of smooth-transition dynamic conditional correlation (STDCC) to generate time-varying measures of a country's levels of economic and financial integration. The STDCC model has several advantages over Engle's (2002) dynamic conditional correlation model. STDCC allows both the unconditional correlation, or the stationary level of correlation, and the conditional correlation to be time-varying. In addition, it not only allows us to control for the volatility effect in estimating the market integration measure, but also has the ability to capture both short- and long-run dynamics of market integration.[11] Appendix contains the details and formulation of the STDCC approach.

Fig. 3 reproduces ANS's (2020) Fig. 1 that depicts the time-series dynamics of economic and financial integration. It shows that both DEV and EMG markets have become more integrated over the sample period, but that DEV markets experience higher levels of financial and economic integration with the world market than their EMG counterparts. The financial integration gap between DEV and EMG markets still remains large throughout the whole period, but their economic integration gap shows convergence between the two markets toward the end of 2015.

In the next section, we present the set of commonly studied explanatory variables that can explain cross-country and time-series variations in ANS's (2020) measures of economic and financial integration. We recognize that ANS's integration metrics do contain measurement errors. As these metrics are employed as dependent variables in the second step of the analysis (e.g., see Bekaert et al., 2011, Carrieri et al., 2013, and Lehkonen, 2014 for this two-step procedure), any errors We recognize that ANS's integration metrics do contain measurement errors. As these metrics are employed as dependent variables in the second step of the analysis (e.g., see Bekaert et al., 2011; Carrieri et al., 2013; Lehkonen, 2014 for this two-step procedure), any errors in the dependent variables would be incorporated in the disturbance term and will cause no problems to the estimation". in the dependent variables would be incorporated in the disturbance term and will cause no problems to the estimation.

3.2. Plausible explanatory variables for integration

Our study employs a fairly exhaustive list of 30 variables, which are mainly drawn from the existing empirical literature.[12] To facilitate our discussion below, we group the 30 variables into six broad, but not necessarily mutually exclusive, categories: (i) economic development, (ii) information/openness, (iii) financial development, (iv) international trade, (v) foreign direct investment, and (vi) cyclical (business and financial cycles) variables. Below we briefly describe the proxies for each category and relegate the construction of these variables and data sources to Table A.1; their cross-correlation matrix is in Table 2.

a. *Economic development*

The extent of a country's level of economic development offers a broad array of benefits that promote global integration. We employ the following proxies for development in different areas of an economy that potentially can have an impact on both economic and financial integration. (i) GDP per capita (GDPC) reflects a country's availability of economic and financial resources, as well as its efficient allocation of these resources within the economy (e.g., Bekaert, Harvey, and Lundblad, 2001; Love, 2003). Thus, it is the most general measure of the country's level of economic development. (ii) Better infrastructure is measured by the amount of electricity consumed per capita (Electricity). (iii) Other proxies include measures of human capital: the proportion of secondary school enrollment (School), the rate of population growth (gPopulation), and the average life expectancy (Life).

[10] Please refer to ANS (2020), Table A.1 for details of the estimation procedure.

[11] Akbari et al. (2020) provide strong arguments for the appropriateness of their integration measures and the choice of the STDCC correlations over rolling-window correlation estimates. The authors argue that the latter approach to measure market integration is highly sensitive to the window size and is prone to the conditional volatility bias, especially during financial crises (Forbes and Rigobon, 2002). When volatility is not modeled in the correlation dynamics, changes in volatility are spuriously picked up by time-varying correlation estimates. As a result, market integration wrongly appears quite volatile.

[12] Please see Akbari and Ng (2020) for a recent survey on market integration measures and their plausible explanatory variables.

Table 1
The number of firms in the sample.

Country	Start	1989–2015	1989–1995	1996–2002	2003–2009	2010–2015
Panel A: All, developed (DEV), and emerging markets (EMG)						
Mean All		39202	11234	21867	22339	21764
Mean DEV		28411	9360	17673	16756	14152
Mean EMG		10791	1874	4194	5583	7612
Panel B: Developed markets (DEV)						
Australia	1989	1515	225	584	853	950
Austria	1989	139	72	96	71	54
Belgium	1989	187	61	131	129	107
Canada	1989	1959	286	666	1234	1231
Denmark	1989	227	134	183	114	84
Finland	1989	193	76	150	131	123
France	1989	1124	428	707	646	526
Germany	1989	1170	384	805	722	634
Hong Kong	1989	831	213	446	637	581
Ireland	1989	90	44	54	60	44
Italy	1989	436	165	260	309	236
Japan	1989	3890	1016	3011	2452	1888
Netherlands	1989	265	167	218	159	104
New Zealand	1989	149	45	83	91	86
Norway	1989	356	77	178	250	207
Singapore	1989	679	162	318	458	306
Spain	1989	231	131	167	158	121
Sweden	1989	519	135	319	291	329
Switzerland	1989	295	144	205	225	185
United Kingdom	1989	3346	1257	1840	1982	1489
United States	1989	10810	4138	7252	5784	4867
Panel C: Emerging markets (EMG)						
Argentina	1993	73	29	65	53	24
Brazil	1992	342	87	173	221	199
Chile	1992	126	65	94	59	66
China	1993	2375	11	53	1131	2279
Egypt	1999	84		14	59	74
Greece	1992	295	117	255	179	68
India	1993	1313	173	286	695	1119
Indonesia	1990	325	126	174	154	185
Israel	1995	110	4	46	56	82
Malaysia	1989	934	247	411	698	526
Mexico	1992	151	58	103	89	92
Pakistan	1993	106	33	56	49	71
Philippines	1989	146	71	101	76	81
Poland	1995	307	23	71	141	248
Portugal	1991	81	44	63	42	35
South Africa	1989	477	128	368	230	193
South Korea	1989	1446	248	772	485	935
Taiwan	1989	1255	182	576	562	906
Thailand	1989	559	190	295	360	282
Turkey	1991	286	38	218	244	147

This table shows the average number of firms available within the full sample period and four sub-periods, and the starting year of the available data (Start) by country. It provides the information across all 41 markets (All), 21 developing markets (DEV), and 20 emerging markets (EMG). The sample period is from January 1989 to December 2015.

b. *Information/Openness*

The information environment and economic openness have been identified in the literature as salient factors affecting international financial investments and market integration (e.g., Bae, Bailey, and Mao, 2006; Carrieri, Chaieb, and Errunza, 2013). Information and monitoring costs may make it difficult for foreign investors to assess financial risks and deter investments in capital markets. On the other hand, the availability of timely and reliable information in an economy may help investors to recognize risks and improve risk sharing. Such environments can be measured by a country's adoption of the International Financial Reporting Standards (IFRS), equity market openness (Equity Mkt Openness), and financial openness (Financial Openness). Another measure of information proxy is the number of internet users per 1000 people (Internet) that captures the general ease of information access in a particular country, and particularly, the extent of the country's communications technology. A country's level of economic openness measures the degree of free trade or capital movements, and the greater degree of openness promotes globalization. We use current account openness (Current Account Openness), capital account openness (Capital Account Openness), and trade openness (Trade Openness) as proxies for economic openness.

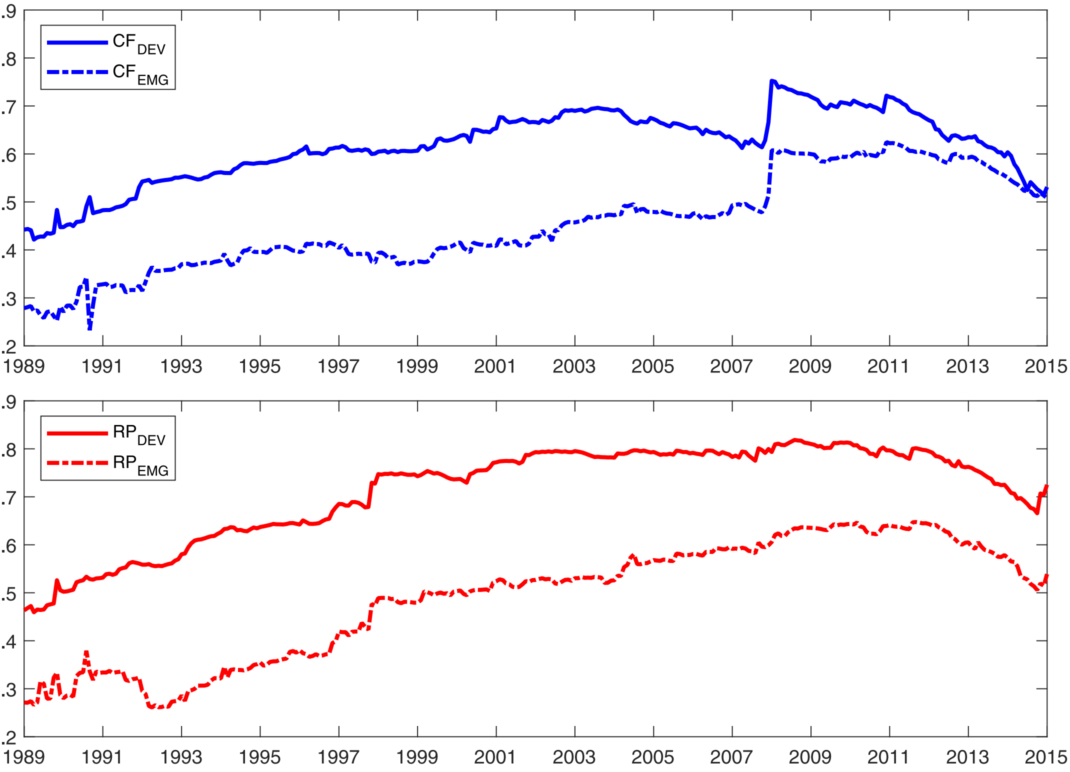

Fig. 3. Dynamic conditional correlations for cash flow news (CF) and risk price adjustments (RP) by market type: The top chart shows equal-weighted conditional correlations of world and country-level cash flow news for developed (DEV) and emerging markets (EMG). The bottom chart depicts equal-weighted conditional correlations of world and country-level risk pricing adjustments for developed (DEV) and emerging (EMG) countries. The dynamics of conditional correlations of CF depicts the time-variation in economic integration and the dynamics of conditional correlations of RP indicates the time-variation in financial integration.

To promote information transparency and openness, a country needs to have strong legal institutions, such as law and order (Law & Order), high investment profile (Investment Profile), and strong shareholder protection measured by anti-director index (Anti-director). Law & Order measures the strength and impartiality of the legal system and the extent of popular observance and enforcement of the law. Investment Profile reflects the risk of expropriation, contract viability, payment delays, and the ability to repatriate profits, and Anti-director accounts for the extent of shareholder protection. Hence, these three measures are included in this category of integration determinants.

c. *Financial development*

Our study gauges the extent of a country's financial development by the development of its stock market and banking sector. We use the ratio of stock market capitalization to GDP (Market Cap) as a proxy for stock market advancement and the ratio of private credit provided by financial institutions to GDP (Private Credit) as a measure of the banking sector development. As banks are dominant financing sources in many emerging and bank-based countries, poor banking sector development (represented by low levels of Private Credit) can significantly hamper integration (Levine and Zervos, 1998). Also, such development facilitates a more efficient allocation of capital (Wurgler, 2000) and can promote global integration.

d. *International trade*

To measure a country's intensity in domestic and foreign trade activities, the international tradeability of goods and services, as well as the free flow of capital are critical conditions for promoting integration (Edwards, 1993; Bekaert and Harvey, 1995). A country's total trade (Trade), merchandise trade (Merchandise Trade), as well as its exports (Exports) and imports (Imports) are employed as a means to gauge its extent in international trade.

e. *Foreign direct investment*

Besides financial market investments, foreign direct investment (FDI) should improve integration. FDI is measured using the amounts of FDI inflow (FDI Inflow), FDI outflow (FDI Outflow), and the sum of FDI inflow and outflow (FDI Total).

f. *Cyclical variables*

We add variables that are commonly linked to business and even financial cycles. Drawn from the growth literature (see, for e.g., Barro, 1996), we use both GDP growth rate (gGDP) and world GDP growth rate (gGDP$_w$) as measures of the overall growth of a country and the world, respectively. gGDP$_w$ captures the world (cyclical) business cycle and can affect global risk pricing changes as well as investors' cash flow expectations. Following Bekaert et al. (2007), we include the uncertainty of world growth rate (gUncertainty$_w$). As financial variables are often used as indicators of the business or financial cycle (e.g., BHLS, 2011),

Table 2
Correlation of determinants of economic and financial integration.

	Elec	Scho	gPop	Life	InPr	ADir	CaAO	CuAO	FiOp	Law	EqOp	IFRS	Inter	TrOp	MCap	PrCr
Economic Development Proxies																
GDPC (GDPC)	**0.70**	**0.78**	**−0.46**	**0.84**	**0.52**	−0.04	**0.76**	**0.70**	**0.74**	**0.67**	**0.74**	**0.45**	**0.63**	**0.52**	**0.30**	**0.55**
Elec (Electricity)		**0.58**	**−0.29**	**0.54**	**0.37**	**0.10**	**0.55**	**0.48**	**0.54**	**0.61**	**0.57**	**0.22**	**0.46**	**0.27**	**0.15**	**0.38**
Scho (School)			**−0.46**	**0.65**	**0.38**	**−0.09**	**0.59**	**0.52**	**0.56**	**0.55**	**0.55**	**0.34**	**0.46**	**0.43**	**0.10**	**0.31**
gPop (gPopulation)				**−0.47**	**−0.21**	**0.21**	**−0.38**	**−0.32**	**−0.35**	**−0.40**	**−0.31**	**−0.22**	**−0.28**	**−0.18**	0.01	**−0.25**
Life (Life)					**0.45**	**−0.07**	**0.69**	**0.71**	**0.69**	**0.63**	**0.63**	**0.35**	**0.58**	**0.40**	**0.25**	**0.43**
Information/Openness Proxies																
InPr (Investment Profile)						**0.10**	**0.46**	**0.44**	**0.45**	**0.29**	**0.36**	**0.40**	**0.66**	**0.27**	**0.36**	**0.44**
ADir (Anti-Director)							−0.06	**−0.15**	−0.06	**−0.07**	**−0.08**	0.04	**0.09**	**−0.18**	**0.21**	**0.22**
CaAO (Capital Account Openness)								**0.85**	**0.87**	**0.67**	**0.83**	**0.28**	**0.42**	**0.42**	**0.26**	**0.40**
CuAO (Current Account Openness)									**0.84**	**0.61**	**0.75**	**0.26**	**0.41**	**0.51**	**0.23**	**0.29**
FiOp (Financial Openness)										**0.64**	**0.79**	**0.29**	**0.44**	**0.45**	**0.28**	**0.46**
Law (Law & Order)											**0.65**	**0.13**	**0.28**	**0.23**	**0.14**	**0.41**
EqOp (Equity Mkt Openness)												**0.20**	**0.36**	**0.43**	**0.24**	**0.38**
IFRS (IFRS)													**0.71**	**0.16**	**0.20**	**0.32**
Inter (Internet)														**0.22**	**0.33**	**0.52**
TrOp (Trade Openness)															**0.13**	**0.15**
Financial Development Proxies																
MCap (Market Cap)																**0.47**
PrCr (Private Credit)																

	Expo	Impo	MeTr	Trd	FDII	FDIO	FDIT	gGDP	gUnc	R_{-1}	$gGDP_w$	CSprd	VIX
Economic Development Proxies													
GDPC (GDPC)	**0.25**	**0.23**	**0.20**	**0.24**	**0.23**	**0.34**	**0.31**	**−0.34**	**−0.12**	**−0.07**	−0.00	**0.10**	0.02
Elec (Electricity)	**0.09**	0.04	0.04	**0.06**	0.06	**0.17**	**0.13**	**−0.24**	**−0.06**	−0.03	0.01	0.04	0.02
Scho (School)	**0.07**	0.04	0.03	0.06	**0.15**	**0.23**	**0.22**	**−0.31**	**−0.15**	−0.05	0.01	0.03	0.01
gPop (gPopulation)	**0.13**	**0.16**	**0.13**	**0.14**	−0.02	**−0.09**	**−0.06**	**0.27**	**0.08**	0.03	−0.02	−0.01	0.002
Life (Life)	**0.27**	**0.25**	**0.23**	**0.26**	**0.24**	**0.30**	**0.29**	**−0.25**	**−0.14**	−0.06	−0.00	**0.08**	0.02
Information/Openness Proxies													
InPr (Investment Profile)	**0.24**	**0.23**	**0.21**	**0.24**	**0.24**	**0.31**	**0.30**	**−0.08**	**−0.38**	−0.01	**0.08**	**0.22**	**0.17**
ADir (Anti-Director)	0.00	0.01	−0.00	0.01	0.02	−0.03	−0.01	0.04	**−0.09**	−0.01	0.01	0.06	0.03
CaAO (Capital Account Openness)	**0.24**	**0.24**	**0.20**	**0.24**	**0.24**	**0.32**	**0.30**	**−0.28**	**−0.08**	−0.04	0.01	0.04	0.03
CuAO (Current Account Openness)	**0.27**	**0.26**	**0.23**	**0.26**	**0.22**	**0.28**	**0.27**	**−0.23**	**−0.08**	−0.02	0.002	0.04	0.02
FiOp (Financial Openness)	**0.27**	**0.26**	**0.23**	**0.27**	**0.24**	**0.31**	**0.30**	**−0.27**	**−0.10**	−0.04	0.02	0.04	0.02
Law (Law & Order)	**0.17**	**0.15**	**0.12**	**0.16**	**0.19**	**0.26**	**0.24**	**−0.13**	0.02	−0.04	0.01	**−0.08**	−0.06
EqOp (Equity Mkt Openness)	**0.24**	**0.23**	**0.19**	**0.23**	**0.22**	**0.31**	**0.28**	**−0.26**	−0.04	−0.03	0.03	−0.01	0.02
IFRS (IFRS)	**0.11**	**0.11**	**0.09**	**0.11**	**0.19**	**0.26**	**0.25**	**−0.27**	**−0.21**	**−0.07**	−0.05	**0.24**	−0.02
Inter (Internet)	**0.24**	**0.21**	**0.20**	**0.23**	**0.23**	**0.32**	**0.31**	**−0.26**	**−0.36**	−0.06	−0.01	**0.28**	**0.06**
TrOp (Trade Openness)	**0.16**	**0.16**	**0.15**	**0.16**	**0.09**	**0.12**	**0.11**	**−0.25**	**−0.06**	−0.002	−0.01	0.02	0.01
Financial Development Proxies													
MCap (Market Cap)	**0.64**	**0.66**	**0.65**	**0.65**	**0.49**	**0.52**	**0.53**	0.03	**−0.13**	**0.08**	0.06	−0.02	−0.03
PrCr (Private Credit)	**0.27**	**0.25**	**0.24**	**0.26**	**0.18**	**0.26**	**0.23**	**−0.21**	**−0.12**	**−0.08**	**−0.01**	**0.10**	0.03
International Trade													
Expo (Export)		**0.99**	**0.97**	**1.00**	**0.63**	**0.53**	**0.61**	**0.13**	**−0.08**	−0.003	0.03	0.05	0.02
Impo (Import)			**0.98**	**1.00**	**0.63**	**0.52**	**0.60**	**0.13**	**−0.07**	−0.02	0.02	0.06	0.02
MeTr (Merchandise Trade)				**0.98**	**0.62**	**0.51**	**0.60**	**0.13**	**−0.07**	−0.01	0.03	0.05	0.02
Trd (Trade)					**0.63**	**0.53**	**0.61**	**0.13**	**−0.08**	−0.01	0.02	0.05	0.02

(*continued on next page*)

we, therefore, include corporate credit spread (Credit Spread), the past year's return performance (R_{-1}), and the expectation of volatility of the U.S. stock market (VIX) to capture cyclical effects.

4. Determinants of economic and financial integration

In this section, we identify the mechanisms that can explain the varying degrees of economic and financial integration exhibited by different countries worldwide. More importantly, our analysis offers insights into the underlying drivers of the integration levels between DEV and EMG markets.

The RFR technique allows us to take an inclusive approach for the set of plausible determinants of market integration, as detailed in Section 3.2. However, in interpreting the results, we benefit from the evidence documented in ANS (2020) on the cross-sectional and time-series economic and financial integration patterns.

Table 2 (*continued*).

	Expo	Impo	MeTr	Trd	FDII	FDIO	FDIT	gGDP	gUnc	R_{-1}	$gGDP_w$	CSprd	VIX
Foreign Direct Investment													
FDII (FDI Inflow)						**0.86**	**0.96**	**0.14**	**−0.12**	**−0.02**	0.08	**0.06**	0.05
DUIO (FDI Outflow)							**0.96**	0.04	**−0.14**	−0.04	**0.09**	**0.08**	**0.06**
FDIT (FDI Total)								**0.09**	**−0.14**	**−0.03**	0.09	**0.07**	0.06
Cyclical Proxies													
gGDP (gGDP)									**−0.12**	**0.12**	**0.37**	**−0.14**	**−0.12**
gUnc ($gUncertainty_w$)										**0.14**	**−0.45**	**−0.10**	**−0.07**
R_{-1} (R_{-1})											**−0.09**	**−0.43**	**−0.39**
$gGDP_w$ ($gGDP_w$)												**−0.23**	**−0.15**
CSprd (Credit Spread)													**0.69**

This table reports pairwise correlations of the potential determinants of economic and financial integration in our sample. For ease of reference, we shorten some of the variable names and split the correlation matrix in two pages. The variables are *Economic Development Proxies*: Internet (Inter), GDPC (GDPC), Electricity (Elec), School (Scho), gPopulation (gPop), Life (Life); *Information/Openness Proxies*: Investment Profile (InPr), Anti-Director (ADir), Capital Account Openness (CaAO), Current Account Openness (CuAO), Financial Openness (FiOp), Law & Order (Law), Equity Mkt Openness (EqOp), IFRS (IFRS), Trade Openness (TrOp); *Financial Development Proxies*: Market Cap (MCap), Private Credit (PrCr); *International Trade*: Exports (Expo), Imports (Impo), Merchandise Trade (MeTr), Trade (Trd); *Foreign Direct Investment*: FDI Inflow (FDII), FDI Outflow (FDIO), FDI Total (FDIT); *Cyclical Proxies*: gGDP (gGDP), $gUncertainty_w$ (gUnc), R_{-1} (R_{-1}), $gGDP_w$ ($gGDP_w$), Credit Spread (CSprd), VIX (VIX). Statistically significant correlations at the 5% level are highlighted in bold.

ANS (2020) finds that the time trends associated with economic and financial integration are positive and statistically significant and remain qualitatively unchanged even after controlling market volatility. The time-trend coefficient is larger for financial than economic integration, especially in DEV markets, suggesting that the average speed of financial integration is faster than that of economic integration over the entire sample period. However, economic integration has been growing at a much quicker pace in EMG markets than in their DEV counterparts, whereas the reverse applies to financial integration. The opening of their economies to the world market since the start of the new millennium could contribute to the rapid increase in EMG markets' economic integration. International trade (e.g., merchandise trade) and foreign direct investment inflows (FDI) have increased for these markets relative to their DEV peers. To demonstrate, Fig. 4 plots the time-series of merchandise trade and FDI, measured in US dollars and scaled to their respective 1989 values. The plots show that the two variables are rising faster in EMG than in DEV markets, especially after the global crisis period.

To identify the drivers of economic and financial integration dynamics, we first estimate the RFR's model parameters, separately, for economic integration and financial integration measures. Then, we estimate the variable importance score, VI (hereafter RFR-assigned importance score), according to Eq. (5), for all plausible explanatory variables of market integration.

4.1. RFR results

Table 3 shows the VI for each of the 30 commonly-employed variables in explaining economic and financial integration, separately. We generate VI using the "Mean Decrease Accuracy" within the *RFR* approach, as described in Section 2. This technique identifies the information values of all input determinants relative to other determinants.[13] As a result, the sum of VIs of all 30 variables is normalized to equal 100%. Given the data availability of candidate determinant variables, the analysis of our integration measures derived in Section 3 are based on annual frequency from 1989 to 2015.

The table reveals several interesting results. First, the findings suggest that a country's economic development plays the most crucial role in explaining economic integration. Economic development accounts for 45.4% of the time-series and cross-country variation in economic integration. Among the various proxies for economic development, GDP per capita (GDPC) is the strongest determinant of economic integration (26.7%), followed by population growth rate (gPop) (7.2%). The statistically significant negative correlation between GDPC and gPop (−46%, Table 2) suggests that countries, especially emerging countries, that control their population growth achieve a greater degree of economic integration. Information/openness is the second most important category (20.5%), followed by international trade (16.1%).[14] Other determinant categories have a much lesser influence on economic integration.

Results for financial integration, however, differ markedly. The most important determinants are the proxies for financial development (42.0%) and information/openness (31.4%). The capital market development (Market Cap), i.e., the market size relative to GDP, plays a vital role in explaining financial integration variation.[15] Internet (15.1%) and Investment Profile (11.4%) also exhibit a significant effect, an implication that the development of information technology and the savvy of Internet investors influence the globalization of financial markets.[16] The investment profile proxy allows investors to evaluate the investment riskiness

[13] The "Mean Decrease Accuracy" and "Mean Decrease Impurity" produce similar rankings of importance for the candidate explanatory variables. Results are available from the authors upon request.

[14] Within the international trade category, the variables are highly correlated (typically above 0.95, Table 2); hence, it is hard to disentangle their individual importance measures.

[15] As we mentioned before, one cannot infer causality in such an analysis. For example, a large capital market could lead to higher financial integration, or higher financial integration could lead to a large capital market.

[16] Emery and Gulen (2018) find that countries with better internet access exhibit a lower geographic bias, and that demand for online financial information facilitates the channel between internet access and investment.

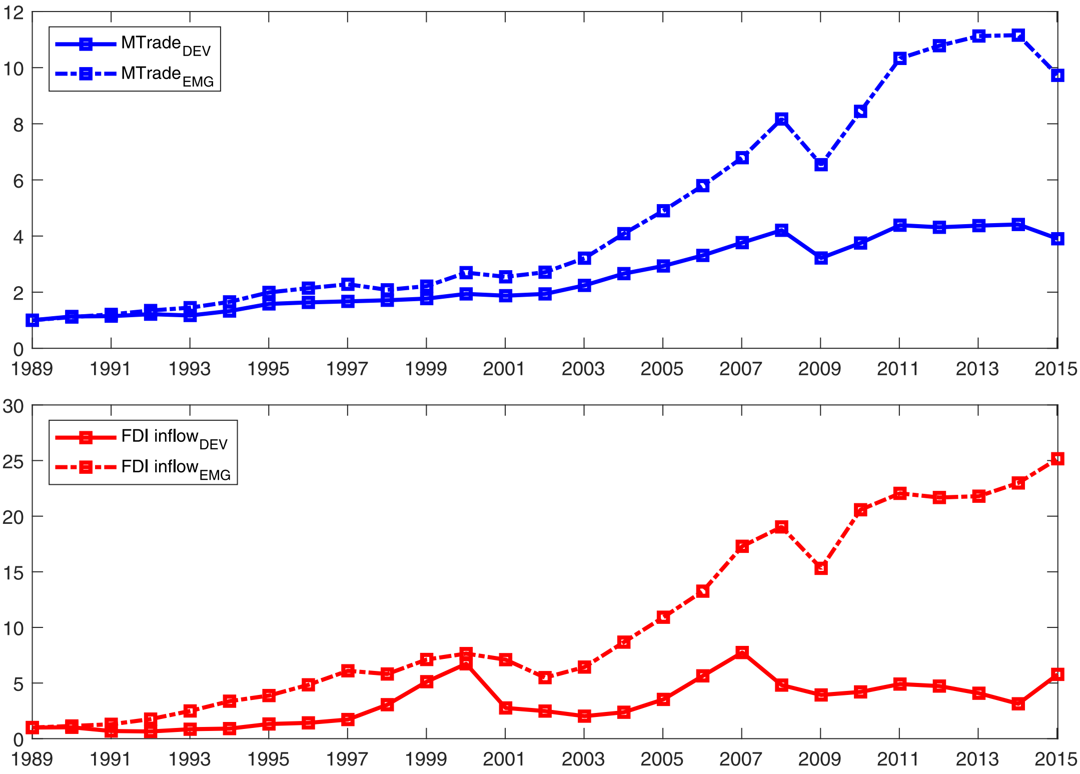

Fig. 4. Evolution of merchandise trade and foreign direct investment by market type: The top chart shows merchandise trade (MTrade) as measured by the sum of exports and imports for developed (DEV) and emerging markets (EMG). The bottom chart depicts foreign direct investment inflows (FDI inflow) for developed (DEV) and emerging markets (EMG). Both series are measured in US dollars and are scaled to their respective values reported in 1989.

of a country, specifically in areas of expropriation, profits repatriation, and payment delays. It is no surprise that countries with higher investment risk profiles are less financially integrated. Economic development explains some variation in financial integration (15.8%), as expected, but exhibits much less impact on financial integration than economic integration. The remaining determinant categories have little impact on financial integration.

Finally, none of the cyclical variables has an importance measure of above 2%, which indicates their limited role in explaining integration, especially financial integration. Perhaps short-term cyclical events tend not to have immediate and drastic effects on the global integration of economies and capital markets. This finding supports our earlier result that integration is a slow-moving process mostly affected by fundamental drivers, such as economic and financial development, or improvement in information and regulatory/political environment.

We now explain our earlier finding that economic integration in EMG countries has converged over time but much less in their financial integration. We have computed the time trend for each of the determinants in separate panel regressions for EMG and DEV countries. Our unreported results highlight the effects of the most important determinants. The average time trend in GDPC, the most important driver for economic integration, is 44% higher for EMG than for DEV countries, suggesting that economic integration has grown much faster for the former. The annual average GDPC growth rate is 0.69% for EMG markets, compared to 0.36% for DEV countries. The second most important economic development variable is population growth. Endemic population growth hampers economic development. Over the sample period, EMG markets have experienced a drastic drop in population growth. The time trend of population growth is negative for EMG countries but positive for DEV countries. The decreased population growth in EMG countries has helped advance their economic development and contributed to their growing economic integration with the world market. Other determinants, however, produce mixed signals and are of minor importance compared to economic development.

Financial integration is primarily driven by the extent of a country's financial development and information/openness. Market Cap, a key proxy for financial development, has a time trend of 240% higher for DEV than for EMG markets. Among the information proxies, the Internet and Investment Profile are, by far, the major contributors. The time trend of Investment Profile is 120% larger for DEV than for EMG counterparts, whereas that of the Internet is 70% larger. While the growing economic development of EMG countries would reduce the financial integration gap between EMG and DEV countries, all other vital determinants offset this effect. Thus, it is not surprising to find the financial integration gap between DEV and EMG countries remains wide.

Table 3
Variable importance measures of integration determinants.

	Economic integration	Financial integration
Economic Development Proxies		
GDPC	0.267	0.037
Electricity	0.046	0.042
School	0.041	0.022
gPop	0.072	0.016
Life	0.028	0.040
Total	**0.454**	**0.158**
Information/Openness Proxies		
IFRS	0.001	0.001
Equity Mkt Openness	0.007	0.003
Financial Openness	0.004	0.004
Internet	0.100	0.151
Current Account Openness	0.029	0.021
Capital Account Openness	0.013	0.003
Trade Openness	0.004	0.000
Law & Order	0.005	0.012
Investment Profile	0.028	0.114
Anti-Director	0.013	0.004
Total	**0.205**	**0.314**
Financial Development Proxies		
Market Cap	0.020	0.389
Private Credit	0.031	0.031
Total	**0.051**	**0.420**
	Economic integration	Financial integration
International Trade		
Trade	0.022	0.006
Merchandise Trade	0.056	0.007
Exports	0.042	0.009
Imports	0.040	0.008
Total	**0.161**	**0.030**
Foreign Direct Investment (FDI)		
FDI Inflow	0.027	0.020
FDI Outflow	0.016	0.008
FDI Total	0.015	0.016
Total	**0.057**	**0.043**
Cyclical Variables		
gGDP	0.016	0.012
gGDP$_w$	0.009	0.003
gUncertainty$_w$	0.018	0.008
Credit Spread	0.009	0.003
R$_{-1}$	0.011	0.005
VIX	0.010	0.004
Total	**0.072**	**0.034**
Aggregate total	**1.000**	**1.000**

This table shows the importance measure of a set of variables that can possibly explain cross-country and time-series variations in economic and financial integration measures. The variable importance measure is computed based on the prediction error of the random forests regression technique, in a context of a multitude of decision trees. The set of variables include six different categories, namely (i) economic development proxies, (ii) information/openness proxies, (iii) financial development proxies, (iv) international trade, (v) foreign direct investment, and (vi) cyclical variables. All variables are defined in Table A.1

4.2. Some economic intuition

Our set of predetermined explanatory variables has complex multivariate non-linear relationships with ANS's integration metrics. It is, therefore, imperative to show that these commonly-employed variables are major integration determinants and have sensible economic relations with the economic and financial integration metrics. To start, we investigate whether our most important variables would yield high explanatory power in a traditional linear regression setting. While we are critical of this simple linear approach as it is poorly adapted to handle a large number of highly correlated explanatory variables, we provide the results for

Table 4
Market integration and the most importance determinants.

Determinant	Economic integration	Financial integration
Economic Development Proxies		
GDPC	0.318*** (0.081)	
Information/Openness Proxies		
Internet	0.236*** (0.047)	0.234*** (0.052)
Investment Profile		0.466*** (0.058)
Financial Development Proxies		
Market Cap		0.300*** (0.050)
International Trade		
Merchandise Trade	0.317*** (0.056)	
Constant	−0.006 (0.006)	0.001 (0.002)
Observations	1016	1016
Adjusted R^2	0.387	0.524

This table reports estimates of the slope coefficients from panel regressions of economic and financial integration measures on their three most important determinants (identified in Table 3). Regressions are over the full time sample from 1989 to 2015 at the annual frequency. The variables are scaled by their respective standard deviations. P-values are estimated using standard errors clustered at year and country levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

comparison purposes. We choose the three most important variables with RFR-assigned importance scores greater than 10%,[17] as suggested in Table 3.

Table 4 reports the coefficient estimates of panel regressions with alternately economic and financial integration metrics as the dependent variable. All variables are scaled by their respective standard deviations to facilitate a comparison of their economic significance.[18] The robust standard errors reported in parenthesis are clustered by country and year. The coefficient estimates are positive and statistically significant, consistent with prior empirical evidence in the literature. The three determinants explain a large portion of the market integration variation — 38.7% for economic integration and 52.4% for financial integration. The magnitude of the coefficient estimates suggests the relative importance of the variables. For example, in terms of their economic significance, a one-standard-deviation increase in GDPC leads to a 0.318 standard deviation increase in economic integration. In comparison, Table 3 indicates that the relative importance of GDPC is 0.267 and of the Economic Development category is 0.454. These findings are not surprising, given the differences in the methodologies and variable inclusion. Our experiment is a simple illustration that our RFR results would not be unrelated to what would be found in more traditional approaches.

To provide more economic intuition on our baseline evidence, we also conduct a univariate analysis of each significant variable-integration metric relationship qualitatively and quantitatively. Table 5 tabulates the slope coefficient of each univariate regression and the correlation coefficient. Most variables have statistically significant relationships with the integration metrics, and the signs of their coefficients are broadly consistent with economic expectations. But Table 2 indicates that the pairwise correlations of these variables are high. For example, in the case of economic integration, based on RFR-assigned importance scores, GDPC is the most important proxy for economic development, followed by gPOP. The importance of gPOP with a negative sign (−0.072) probably attributes to the inclusion of EMG countries. As mentioned earlier, the latter have better control of their population explosion while growing in terms of GDP and thus are better positioned to be part of the global economy. The various Information/Openness proxies capture different aspects of information availability, reliability, and market openness, with the Internet having the most considerable RFR-assigned importance measure. This result suggests that information technology plays a significant role in the global economy. Proxies for trades, which are highly correlated, have a strong and positive influence on economic integration, consistent with our earlier finding that RFR assigns relatively high importance to international trade in economic integration. Lastly, the correlation between proxies for the other categories and economic integration is weak, which is in line with their low RFR-assigned importance measures.

Similarly, all proxies for economic development, information/openness, and financial development significantly correlate with the financial integration metric, and their slope coefficients bear the expected signs. RFR assigns considerable importance to these three categories and little significance to the remaining three types (i.e., International Trade, FDI, and Cyclical Variables). While

[17] While the International Trade category has a RFR-assigned importance value of 16.1% and all its proxies are highly correlated (over 0.98), our selected Merchandise Trade has the largest RFR-assigned importance value in the category.

[18] Note that the coefficients sum to 1.0 for financial integration and 0.9 for economic integration.

Table 5
Market integration and their determinants — univariate analysis.

	Economic integration		Financial integration	
Determinant	ρ	β	ρ	β
Economic Development Proxies				
GDPC	0.496	0.066***	0.638	0.109***
Electricity	0.366	0.012***	0.493	0.021***
School	0.341	0.002***	0.490	0.005***
gPOP	−0.326	−0.072***	−0.286	−0.081***
Life	0.372	0.763***	0.488	1.292***
Information/Openness Proxies				
IFRS	0.248	0.095***	0.375	0.184***
Equity Mkt Openness	0.291	0.136***	0.399	0.241***
Financial Openness	0.358	0.184***	0.518	0.343***
Internet	0.465	0.025***	0.620	0.042***
Current Account Openness	0.358	0.003***	0.478	0.005***
Capital Account Openness	0.419	0.003***	0.506	0.005***
Trade Openness	0.054	0.040	0.250	0.240***
Law & Order	0.288	0.038***	0.312	0.054***
Investment Profile	0.452	0.032***	0.628	0.058***
Anti-Director	0.053	0.008	0.165	0.031
Financial Development Proxies				
Market Cap	0.201	0.000	0.409	0.019***
Private Credit	0.342	0.001***	0.434	0.002***
International Trade				
Trade	0.461	0.056***	0.091	0.000
Merchandise Trade	0.441	0.064***	0.086	0.000
Export	0.464	0.117***	0.099	0.001
Import	0.453	0.104***	0.081	0.001
Foreign Direct Investment				
FDI Inflow	0.099	0.002	0.142	0.004*
FDI Outflow	0.221	0.005**	0.247	0.007***
FDI Total	0.165	0.002**	0.211	0.003**
Cyclical Proxies				
gGDP	−0.221	−0.011***	−0.270	−0.018***
gGDP$_w$	−0.069	−0.009	−0.008	−0.001
gUncertainty$_w$	−0.072	−0.057	−0.133	−0.136
Credit Spread	0.197	0.063***	0.158	0.065***
R$_{-1}$	−0.042	0.000	−0.015	0.000
VIX	0.100	0.003	0.106	0.004

This table reports the pairwise correlation coefficient (ρ) and slope coefficient (β) from univariate regressions of economic and financial integration measures on a set of predetermined variables that can explain their cross-country and time-series variations. P-values are estimated using standard errors clustered at the year and country levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

GDPC is highly correlated with financial integration (0.638), its RFR-assigned importance is rather low (0.037). Other aspects of development are likely more critical for financial integration than pure economic growth. For example, China and India have experienced tremendous production growth, but their financial markets remain relatively closed.

In summary, RFR results are consistent with quantitative and qualitative economic expectations, and it has the ability to circumvent the multicollinearity and nonlinearity issues and highlight the important variables (see Section 4.4).

4.3. Robustness tests

In the preceding section, while our correlation measures are fairly persistent, it is possible that RFR ignores such time series properties. To ensure that our results are robust, we include lagged explanatory variables in the model to adjust for the serial correlation.[19] We re-estimate the RFR model using one-year lagged values combined with their contemporaneous values of the predetermined variables (60 in total). Results are reported in Table 6, alongside the original results. When the model includes both contemporaneous and lagged variables, the RFR-assigned importance measure of each category does not change noticeably from that of Table 3. For example, the RFR-assigned importance measure of Economic Development proxies for economic integration is 0.454 when the RFR estimation includes no lagged values and is 0.434 when it contains both contemporaneous and lagged variables. A closer analysis of each category's sub-components (contemporaneous vs. lagged variables) suggests that lagged variables tend to be

[19] We thank a referee for making this excellent suggestion.

Table 6
Robustness test.

	Without lagged variables		With lagged variables							
			Economic integration			Financial integration				
Variable category	Economic integration	Financial integration	Total	Contemp.	Lagged	Total	Contemp.	Lagged	Excluding 2007–09	
Economic Development Proxies	0.454	0.158	0.434	0.208	0.226	0.127	0.062	0.065	0.487	0.135
Information/Openness Proxies	0.205	0.314	0.205	0.079	0.126	0.287	0.142	0.145	0.202	0.290
Financial Development Proxies	0.051	0.420	0.047	0.020	0.027	0.490	0.145	0.345	0.047	0.479
International Trade	0.161	0.030	0.168	0.083	0.085	0.025	0.011	0.014	0.152	0.026
Foreign Direct Investment	0.057	0.043	0.071	0.026	0.045	0.042	0.022	0.020	0.062	0.043
Cyclical Proxies	0.072	0.034	0.074	0.031	0.043	0.031	0.014	0.017	0.050	0.027
Aggregate	1.000	1.000	1.000	0.447	0.553	1.000	0.395	0.605	1.000	1.000

This table shows robustness results for the aggregate RFR-assigned importance measure of a set of variables that can possibly explain cross-country and time-series variations in measures of economic and financial integration. The first two columns report our baseline results on the RFR-assigned importance measures of each determinant category from Table 3. The next six columns report the RFR-assigned importance measures for determinant categories using information on both contemporaneous and one-year lagged values of the determinants. For each integration metric, the table reports the contribution of the sub-components under each category stemming from contemporaneous and one-year lagged variables. The last two columns report the RFR-assigned importance measures for each determinant category estimated on a subsample that excludes the global financial crisis period (the 2007–2009 years).

Table 7
Variable importance measures of integration determinants and the leaf node size.

	Table 3		Number of observations per leaf = 20	
Determinant category	Economic integration	Financial integration	Economic integration	Financial integration
Economic Development Proxies	0.454	0.158	0.520	0.141
Information/Openness Proxies	0.205	0.314	0.202	0.324
Financial Development Proxies	0.051	0.420	0.029	0.488
International Trade	0.161	0.030	0.182	0.011
Foreign Direct Investment	0.057	0.043	0.041	0.026
Cyclical Proxies	0.072	0.034	0.027	0.011
Aggregate total	1.000	1.000	1.000	1.000

This table shows the robustness results for the minimum leaf node size in estimating the RFR-assigned importance measure of a set of variables that can possibly explain cross-country and time-series variations in measures of economic and financial integration. It presents the RFR-assigned aggregate variable importance for each determinant category. The first two columns report the importance measures using a minimum of 5 observations per leaf node, which is the default case in our analysis (presented in Table 3). The last two columns show the importance measures using a minimum of 20 observations per leaf node.

slightly more important than their contemporaneous counterparts for economic integration (55.3% vs. 44.7%), but their difference becomes more pronounced for financial integration (60.5% vs 39.5%), suggesting a possible Granger-causality direction from our economic explanatory variables to integration metrics.

Alternatively, we can introduce dynamics in the model and investigate time variation in the explanatory variables that could reveal time variation in integration. But the relative importance of the variables could evolve as well. However, the availability of low-frequency annual economic data of 27 years does not allow us to split our full sample into sub-samples for time-variation analyses. Instead, we investigate whether the real and financial shock of the 2007–2009 global financial crisis is driving some of our results by replicating our baseline model using a sample that excludes the 2007–2009 crisis period. The results, shown in the last two columns of Table 6, suggest that the order of the most important variables does not change using this subsample. There exists only a marginal change in the RFR-assigned importance measures, compared to the main results.

Additionally, we explore whether increasing the number of minimum observations in the leaf nodes affects the results. We stop splitting when the MSE shows no improvement in our implementation, but we set a minimum of five observations to get a reasonable estimate of the fitted conditional mean. If we increase the minimum number to say 20, we might get a more accurate estimate of the conditional value at the cost of reducing the depth of a tree and its number of branches. Nevertheless, we re-estimate the model using 20 as the maximum number of observations and present the results in Table 7 by determinant category. For comparison purposes, we also show Table 3's results in the first two columns. While the importance of each category differs slightly, the main conclusion remains materially unaffected.

4.4. Multicollinearity, predictability, and nonlinearity issues

In the presence of highly correlated explanatory variables, the estimated variances of the coefficients are underestimated. As a result, the econometrician wrongfully assigns a lower *p*-value to the variables. Therefore, prior research that merely focuses on the *p*-value of regression coefficient estimates is prone to multicollinearity issues. In contrast, RFR takes a different approach to assign

Table 8
RFR comparison with linear models: out-of-sample fit.

Model	Economic integration	Financial integration
RFR	0.063	0.663
LASSO Regression		
$\lambda = 0.1$	−0.406	0.272
$\lambda = 1$	−0.124	0.277
$\lambda = 10$	−0.047	0.000
Ridge Regression		
$\lambda = 0.1$	−0.712	0.392
$\lambda = 1$	−0.696	0.389
$\lambda = 10$	−0.621	0.361

This table shows the out-of-sample R^2 score for the goodness of the fit for economic and financial integration measures, using Random Forest Regression (RFR), least absolute shrinkage and selection operator (LASSO Regression), and the Ridge regression. The training set is based on the 1989–2014 sample and the test set is based on the 2015 sample. The results of the LASSO and Ridge regressions are presented using three regularization values of $\lambda \in \{0.1, 1, 10\}$.

variable importance by emphasizing the precision of out-of-sample predictions rather than in-sample fit. Note that multicollinearity does not reduce the model's predictive power but only affects the estimation of model coefficients.

In this section, we study the comparative performance of RFR in predicting market integration measures in our setting. We compare the out-of-sample goodness of fit score (R^2) of the RFR model with two commonly used linear models in machine learning, namely LASSO and Ridge regressions, as follows.

- LASSO Regression: $argmin_\beta \ \sum_{t=1}^{T}(y_t - \sum_{i=0}^{K}\beta_i x_{i,t})^2 + \lambda \sum_{i=0}^{K}|\beta_i|$
- Ridge Regression: $argmin_\beta \ \sum_{t=1}^{T}(y_t - \sum_{i=0}^{K}\beta_i x_{i,t})^2 + \lambda \sum_{i=0}^{K}\beta_i^2$

where y_t is the vector of economic (or financial) measures of integration for 40 countries in our sample and $x_{i,t}$ denotes each of the 30 plausible determinants in our pool of explanatory variables for the countries in our sample. λ is the regularization parameter that penalizes in-sample accuracy for a better out-of-sample fit. Low values of λ result in a higher in-sample fit (also known as lower bias) that might come at the cost of poor out-of-sample performance (also known as higher variance). The two models are similar in several dimensions. They differ in how they penalize the size of the model coefficients, β_i. Generally, a LASSO regression tends to result in a model with fewer parameters (i.e., $\beta_i = 0$ for several determinants), which might control for the multicollinearity in our sample. On the other hand, a Ridge regression can result in a model with smaller sensitivity to the explanatory variables (i.e., smaller β_i), which might lead to a better out-of-sample performance.

We estimate the above models on a subsample from 1989 to 2014 (the training sample) and then calculate the out-of-sample R^2 of the fitted models using the observations of 2015 (our test sample). To be consistent across the models, we do not use a cross-validation sample needed to estimate the regularization parameter, λ. Instead, we report the score results for the LASSO and Ridge regressions for three values of $\lambda \in \{0.1, 1, 10\}$. Table 8 presents the results.

The table shows that RFR strongly outperforms the linear models in fitting out-of-sample observations for both economic and financial integration measures.[20] Linear models tend to perform relatively better for the financial integration, resulting in a positive out-of-sample R^2, consistent with the high autocorrelation observed in the financial integration measure in the recent years (see Fig. 3). However, considering that economic integration decreases following the 2008 financial crisis, the linear models fail to fit this downward trend in the 2015 subsample, resulting in the observed negative out-of-sample R^2 estimates for these models. By construction, these models impose the same slope coefficient throughout the training set and fail to capture shifts and non-linearity in the integration measure dynamics.

In our sample, after the RFR, the LASSO regression with $\lambda = 10$ provides the second-best score, albeit negative, for the economic integration. But it is also the worst for financial integration. For the financial integration, the RFR fits 0.663 of the variation in out-of-sample observations, and the Ridge regression with $\lambda = 0.1$ is the second-best model to fit the out-of-sample data with $R^2 = 0.392$. But it is the worst for economic integration.

It might be useful to illustrate how RFR help deals with nonlinearity as opposed to conventional regression models. Fig. 5 illustrates this point. We plot the relationship between GDPC and the measure of economic integration for China. Looking at the actual data, the relationship between GDPC and economic integration, as depicted by blue dots, is positive but not necessarily linear. The red line depicts the linear relationship, implied by the least square estimator, and we observe large and persistent deviations from this trend for several observations. The green line shows that the RFR's implied relationship does a better job of fitting the data through a series of piece-wise linear relationships in each tree's leaf node.

5. Conclusion

We study the drivers of economic and financial integration across countries and through time. Our study employs the random forests regression (RFR) technique to overcome the pitfalls of the regression-based variable selection and evaluation procedures. RFR,

[20] Gu et al. (2020), studying US stocks' predictability, also find that RFR outperforms LASSO and Ridge regressions out of sample.

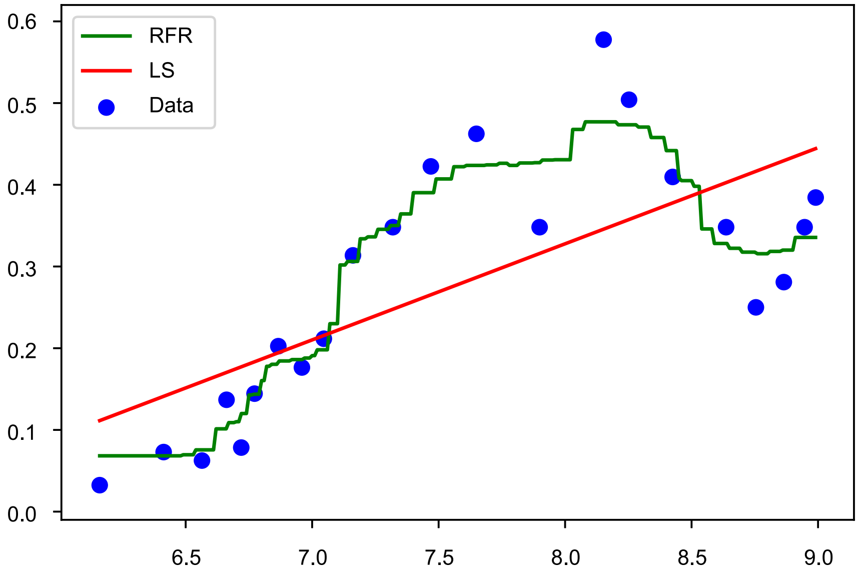

Fig. 5. RFR vs. linear regression approach: an example of the economic integration-GDPC relationship. The plot shows the performance of RFR (represented by the green line) vs. a linear regression method (represented by a red line) by using the information for China as an example. It provides their estimates of the relationship between the economic integration measure, plotted in the y-axis, and the gross domestic product per capita, plotted in the x-axis, as well as the Chinese data points (represented by blue dots). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

initially introduced in Breiman (2001), is an ensemble machine learning method in the context of a multitude of decision trees. RFR allows us to identify the importance of integration determinants in a large and highly correlated pool of potential explanatory variables. RFR accommodates a more general form of relationships, including nonlinear relationships, between dependent and independent variables. They also correct for over-fitting, which may result from a large set of explanatory variables often employed in determining the drivers of integration.

We follow Akbari et al. (2020) approach in measuring economic and financial integration dynamics. Our study offers new insights on the cross-country variation of economic and financial integration based on a large sample of firms from 41 different countries worldwide and a relatively exhaustive list of 30 explanatory variables. Our findings suggest that the dynamics of various country-specific characteristics have contributed to the relative pace of economic and financial integration between developed and emerging countries. Over the last three decades, general economic growth, increasing international trade, and progressive reduction in population growth among emerging economies have helped these countries attain a level of economic integration similar to their developed counterparts. However, the financial integration gap between developed and emerging markets still remains wide. Slow financial development and high investment riskiness have hindered the speed of emerging markets' financial integration with the world market. This phenomenon would possibly explain why these countries still have not caught up with their developed peers in terms of financial integration. These results also suggest that integration is a gradual process driven mainly by fundamental economic and financial variables rather than by cyclical or transitory events.

CRediT authorship contribution statement

Amir Akbari: Conceptualization, Data curation, Methodology, Investigation, Writing. **Lilian Ng:** Writing - review & editing. **Bruno Solnik:** Conceptualization, Methodology, Investigation, Writing - review & editing.

Appendix. Smooth-transitioning dynamic conditional correlations

We employ the smooth-transition dynamic conditional correlation specification to generate time-varying measures of a country's levels of economic and financial integration. The measure of a country's financial integration is the square of the correlation between its own risk pricing revision RP_c and the world risk pricing component RP_w, whereas its economic integration measure is given by the square of the correlation between its own cash flow revision CF_c and its world counterpart CF_w.

Assuming that $CF_{c,t}$ and $CF_{w,t}$ follow an AR(1) process with their residuals (innovations) denoted by $u_{c,t}$ and $u_{w,t}$, respectively. Let $\mathbf{u_t} = (u_{c,t} \quad u_{w,t})' = \mathbf{H_t^{1/2}} \mathbf{v_t}$, where $\mathbf{H_t}$ is the conditional covariance matrix at time t, and $\mathbf{v_t}$ is assumed to be independently identically normally distributed random variable with zero mean and variance of one. $\mathbf{H_t}$ is defined as $\mathbf{H_t} = \mathbf{D_t C_t D_t}$, where $\mathbf{D_t}$ is the conditional variance of each CF and assumes a GARCH(1, 1) process, and $\mathbf{C_t}$ is the time-varying conditional correlation. Assuming bivariate normality, $\mathbf{C_t}$ is modeled as follows.

$$\mathbf{C_t} = \text{diag}(q_{cc,t} \quad q_{ww,t})^{-1/2} \, \mathbf{Q_t} \, \text{diag}(q_{cc,t} \quad q_{ww,t})^{-1/2}, \tag{A.1}$$

$$\mathbf{Q_t} = (1 - a - b)\,\bar{\mathbf{Q}}_\mathbf{t} + b\,\mathbf{Q_{t-1}} + a\,\epsilon_\mathbf{t-1}\epsilon'_\mathbf{t-1}, \tag{A.2}$$

$$\bar{\mathbf{Q}}_\mathbf{t} = (1 - G(s_t;\gamma,d))\,\bar{\mathbf{Q}}^{(\mathbf{1})} + G(s_t;\gamma,d)\,\bar{\mathbf{Q}}^{(\mathbf{2})}, \tag{A.3}$$

where $q_{cc,t}$ and $q_{ww,t}$ are the diagonal elements of $\mathbf{Q_t}$, where $\mathbf{Q_t}$ is the 2×2 matrix driving the dynamics of $\mathbf{C_t}$, $\epsilon_\mathbf{t} = \mathbf{D}_\mathbf{t}^{-1}\mathbf{u_t}$ is a standardized error vector, $\bar{\mathbf{Q}}_t$ is the unconditional correlation matrix of the standardized error $\epsilon_\mathbf{t}$, changing smoothly from $\bar{\mathbf{Q}}^{(\mathbf{1})}$ to $\bar{\mathbf{Q}}^{(\mathbf{2})}$ through time, and G is a logistic transition function given by

$$G(s_t;\gamma,d) = \frac{1}{1 + exp(-\gamma(s_t - d))}, \gamma > 0. \tag{A.4}$$

In Eq. (A.4), s_t (i.e., $s_t = t/T$) is a time trend, employed as a transition variable capturing long-run trends in unconditional correlation, d is a location parameter specifying the center of the transition, and γ is a smoothness parameter specifying the speed of transition. The same procedure is applied when estimating the conditional correlation between RP_c and RP_w. All estimations are implemented using the maximum likelihood approach.

Appendix. Additional tables

See Table A.1.

Table A.1
Variable definition and data source.

Variable	Description	Data source
Returns, cash flow news, risk pricing revisions, and integration metrics		
R_c	Country returns — monthly value-weighted average of stocks' capital gain returns in a country.	DataStream
CF_c	Country cash flow news — monthly value-weighted average of stocks' cash flow news in a country.	I/B/E/S
RP_c	Country risk pricing adjustments — monthly value-weighted average of stocks' risk pricing adjustments in a country.	I/B/E/S
R_w	World returns — monthly value-weighted average of country capital gain returns.	DataStream
CF_w	World cash flow news — monthly value-weighted average of countries'' cash flow news.	I/B/E/S
RP_w	World risk pricing changes — monthly value-weighted average of countries' risk pricing changes.	I/B/E/S
R^2_{Econ}	Measure of economic integration — the square of correlations of country's monthly cash flow changes CF_c and CF_w, using the STDCC specification.	Akbari et al. (2020)
R^2_{Fin}	Measure of financial integration — the square of correlations of country's monthly risk pricing changes RP_c and that of the world RP_w, using the STDCC specification.	Akbari et al. (2020)
Economic Development Proxies		
GDPC	log of annual Gross Domestic Product (GDP) per capita	WDI
Electricity	Electric power consumption measures the production of power plants and combined heat and power plants less transmission, distribution, and transformation losses and own use by heat and power plants.	WDI
School	Ratio of total secondary school enrollment, regardless of age, to the population of the age group.	WDI
gPop	Population growth — a country's annual population growth.	WDI
Life	Log of life expectancy at birth.	WDI
Information/Openness Proxies		
IFRS	A dummy variable that is equal to one if the country adopts International Financial Reporting System in the year and zero otherwise.	iasplus.com
Equity Market Openness	Equity market openness is one minus the equity market restrictions from Fernández et al. (2016). The dataset covers from 1995 to 2013. Following Bekaert et al. (2016), we predict the values for 1989 to 1994 from Quinn and Toyoda (2008) and Chinn and Ito (2008).	Fernández et al. (2016)
Financial Openness	Financial Openness from Chinn and Ito (2008). The dataset coverage is up to 2014.	Chinn and Ito (2008)
Internet	The annual number of internet users per 1000 people.	WDI
Current Account Openness	Current Account Openness — annual publications of the IMF ends in 2011 Quinn and Toyoda (2008), and extend using Fernández et al. (2016) variables.	Various sources
Capital Account Openness	The index ranges from zero to four and is constructed from IMF annual publications which end in 2011 (Quinn and Toyoda, 2008). Following Bekaert et al. (2016), we predict the values for 2012 and 2013 from Fernández et al. (2016) variables with linear predictive regressions.	Various sources
Trade Openness	A dummy variable equals one if the trade of the country is liberalized in the year. The trade liberalization date is based on five criteria: average tariff rates of 40% or more; non-tariff barriers covering 40% or more of trade; a black market exchange rate that is depreciated by 20% or more; a state monopoly on major exports; and a socialist economic system.	Wacziarg and Welch (2008)
Law & Order	Measures the strength and impartiality of the legal system and popular observance of the law.	ICRG
Investment Profile	Investment profile index constructed to assess factors (i.e., country expropriation, profits repatriation, and payment delays) affecting the risk to investment.	ICRG
Anti-Director	The index covers years 1993 to 2002. Before 1993 and after 2002, we assume the anti-director index is constant over time.	Pagano and Volpin (2005)

(continued on next page)

Table A.1 (*continued*).

Variable	Description	Data source
Financial Development Proxies		
Market Cap	Ratio of stock market capitalization to GDP in a year.	WDI; DataStream
Private Credit	Financial resources available to the private sector, through loans, purchases of non-equity securities, and trade credits and other accounts receivable scaled by GDP	WDI
International Trade		
Trade	Sum of exports and imports of goods and services measured as a share of GDP.	WDI
Merchandise Trade	Sum of merchandise exports and imports divided by the value of GDP.	WDI
Exports	Ratio of the value of all goods and other market services provided to the rest of the world to GDP	WDI
Imports	Imports of goods and services representing the value of all goods and other market services received from the rest of the world scaled by GDP.	WDI
Foreign Direct Investment		
FDI Inflow	Ratio of the sum of absolute values of Foreign Direct Investment inflows to GDP.	WDI
FDI Outflow	Ratio of the sum of absolute values of Foreign Direct Investment outflows to GDP.	WDI
FDI Total	Ratio of the sum of absolute values of Foreign Direct Investment inflows and outflows to GDP.	WDI
Cyclical Proxies		
$gGDP_c$	Country GDP growth — annual GDP growth of a country	WDI
$gGDP_w$	World GDP growth — annual world GDP growth	WDI
$gUncertainty_w$	log of the standard deviation of real GDP growth across countries in the sample	IMF
Credit Spread	US corporate spread calculated as the difference between US Baa and Aaa bond yields.	Federal Reserve Bank of St. Louis
R_{-1}	Past-year local stock market annual returns.	Datastream
VIX	The 30-day implied volatility index derived from S&P500 option prices; the options are traded on Chicago Board of Options Exchange (CBOE).	CBOE

References

Akbari, A., Ng, L., 2020. International market integration: A survey. Asia-Pacific J. Financial Stud. 49, 161–185. http://dx.doi.org/10.1111/ajfs.12297.

Akbari, A., Ng, L.K., Solnik, B., 2020. Emerging markets are catching up: Economic or financial integration? J. Financ. Quant. Anal. 55, 2270–2303. http://dx.doi.org/10.1017/S0022109019000681.

Bae, K.-H., Bailey, W., Mao, C.X., 2006. Stock market liberalization and the information environment. J. Int. Money Finance 25, 404–428. http://dx.doi.org/10.1016/j.jimonfin.2006.01.004.

Barro, R.J., Determinants of Economic Growth: A Cross-Country Empirical Study, Working Paper, 1996.

Bekaert, G., Harvey, C.R., 1995. Time-varying world market integration. J. Finance 50, 403–444. http://dx.doi.org/10.2307/2329414.

Bekaert, G., Harvey, C.R., Kiguel, A., Wang, X., 2016. Globalization and asset returns. Ann. Rev. Financial Econ. 8, 221–288. http://dx.doi.org/10.1146/annurev-financial-121415-032905.

Bekaert, G., Harvey, C.R., Lundblad, C., 2001. Emerging equity markets and economic development. J. Dev. Econ. 66, 465–504.

Bekaert, G., Harvey, C.R., Lundblad, C., Siegel, S., 2007. Global growth opportunities and market integration. J. Finance 62, 1081–1137. http://dx.doi.org/10.1111/j.1540-6261.2007.01231.x.

Bekaert, G., Harvey, C.R., Lundblad, C.T., Siegel, S., 2011. What segments equity markets? Rev. Financ. Stud. 24, 3841–3890. http://dx.doi.org/10.1093/rfs/hhr082.

Bekaert, G., Hodrick, R.J., Zhang, X., 2009. International stock return comovements. J. Finance 64, 2591–2626. http://dx.doi.org/10.1111/j.1540-6261.2009.01512.x.

Biau, G., 2012. Analysis of a random forests model. J. Mach. Learn. Res. 13, 1063–1095.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Carrieri, F., Chaieb, I., Errunza, V., 2013. Do implicit barriers matter for globalization? Rev. Financ. Stud. 26, 1694–1739. http://dx.doi.org/10.1093/rfs/hht003.

Carrieri, F., Errunza, V., Hogan, K., 2007. Characterizing world market integration through time. J. Financ. Quant. Anal. 42, 915–940. http://dx.doi.org/10.1017/S0022109000003446.

Chambet, A., Gibson, R., 2008. Financial integration, economic instability and trade structure in emerging markets. J. Int. Money Finance 27, 654–675. http://dx.doi.org/10.1016/j.jimonfin.2008.02.007.

Chinn, M.D., Ito, H., 2008. A new measure of financial openness. J. Comp. Policy Anal. Res. Pract. 10, 309–322. http://dx.doi.org/10.1080/13876980802231123.

Edwards, S., 1993. Openness, trade liberalization, and growth in developing countries. J. Econ. Lit. 31, 1358–1393.

Eiling, E., Gerard, B., 2015. Emerging equity market comovements: Trends and macroeconomic fundamentals. Rev. Finance 19, 1543–1585. http://dx.doi.org/10.1093/rof/rfu036.

Emery, L.P., Gulen, H., Expanding Horizons: The Effect of Information Access on Geographically Biased Investing, Working Paper, 2018.

Engle, R., 2002. Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J. Bus. Econom. Statist. 20, 339–350.

Feng, G., Giglio, S., Xiu, D., 2020. Taming the factor zoo: A test of new factors. J. Finance 75, 1327–1370. http://dx.doi.org/10.1111/jofi.12883.

Fernández, A., Klein, M.W., Rebucci, A., Schindler, M., Uribe, M., 2016. Capital control measures: A new dataset. IMF Econ. Rev. 64, 548–574. http://dx.doi.org/10.1057/imfer.2016.11.

Forbes, K.J., Rigobon, R., 2002. No contagion, only interdependence: Measuring stock market comovements. J. Finance 57, 2223–2261. http://dx.doi.org/10.1111/0022-1082.00494.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. Rev. Financ. Stud. 33, 2326–2377. http://dx.doi.org/10.1093/rfs/hhz123.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63, 3–42. http://dx.doi.org/10.1007/s10994-006-6226-1.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33, 2223–2273. http://dx.doi.org/10.1093/rfs/hhaa009.

Ince, O.S., Porter, R.B., 2006. Individual equity return data from Thomson datastream: Handle with care! J. Financial Res. 29, 463–479. http://dx.doi.org/10.1111/j.1475-6803.2006.00189.x.

Kelly, B.T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. J. Financ. Econ. 134, 501–524. http://dx.doi.org/10.1016/j.jfineco.2019.05.001.

Khandani, A.E., Kim, A.J., Lo, A.W., 2010. Consumer credit-risk models via machine-learning algorithms. J. Bank. Financ. 34, 2767–2787. http://dx.doi.org/10.1016/j.jbankfin.2010.06.001.

Lehkonen, H., 2014. Stock market integration and the global financial crisis. Rev. Finance 2039–2094. http://dx.doi.org/10.1093/rof/rfu039.

Levine, R., Zervos, S., 1998. Stock markets, banks, and economic growth. Am. Econ. Rev. 88, 537–558.

Love, I., 2003. Financial development and financing constraints: International evidence from the structural investment model. Rev. Financ. Stud. 16, 765–791. http://dx.doi.org/10.1093/rfs/hhg013.

Mentch, L., Hooker, G., 2016. Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. J. Mach. Learn. Res. 17, 1–41.

Ohashi, K., Okimoto, T., 2016. Increasing trends in the excess comovement of commodity prices. J. Comm. Mark. 1, 48–64. http://dx.doi.org/10.1016/j.jcomm.2016.02.001.

Pagano, M., Volpin, P.F., 2005. The political economy of corporate governance. Amer. Econ. Rev. 95, 1005–1030. http://dx.doi.org/10.1257/0002828054825646.

Phylaktis, K., Ravazzolo, F., 2002. Measuring financial and economic integration with equity prices in emerging markets. J. Int. Money Finance 21, 879–903. http://dx.doi.org/10.1016/S0261-5606(02)00027-X.

Pukthuanthong, K., Roll, R., 2009. Global market integration: An alternative measure and its application. J. Financ. Econ. 94, 214–232. http://dx.doi.org/10.1016/j.jfineco.2008.12.004.

Quinn, D.P., Toyoda, A.M., 2008. Does capital account liberalization lead to growth? Rev. Financ. Stud. 21, 1403–1449. http://dx.doi.org/10.1093/rfs/hhn034.

Scornet, E., Biau, G., Vert, J.-P., 2015. Consistency of random forests. Ann. Statist. 43, 1716–1741. http://dx.doi.org/10.1214/15-AOS1321.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinformatics 9, 307. http://dx.doi.org/10.1186/1471-2105-9-307.

Wacziarg, R., Welch, K.H., 2008. Trade liberalization and growth: New evidence. World Bank Econ. Rev. 22, 187–231. http://dx.doi.org/10.1093/wber/lhn007.

Wager, S., Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. J. Amer. Statist. Assoc. 113, 1228–1242.

Wager, S., Hastie, T., Efron, B., 2014. Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. J. Mach. Learn. Res. 15, 1625–1651. http://dx.doi.org/10.1080/01621459.2017.1319839.

Wurgler, J., 2000. Financial markets and the allocation of capital. J. Financ. Econ. 58, 187–214. http://dx.doi.org/10.1016/S0304-405X(00)00070-2.