Nascent markets:
Understanding the success and failure of new stock markets

José Albuquerque de Sousa, Thorsten Beck, Peter A.G. van Bergeijk, and Mathijs van Dijk*

March 2023

Abstract

We study the success and failure of newly established stock markets, as measured by listings, market capitalization, and trading activity. Early success is a necessary but not sufficient condition for long-term success, while small population, high corruption, limited law and order, low globalization, and low GDP per capita are necessary but not sufficient conditions for failure. Out of many determinants identified in the literature, merely three indicators jointly predict success with high accuracy: ethnolinguistic fractionalization, democratic participation, and GDP. Our results highlight the importance of a good start and of "structural factors" (vs. "policy factors") in predicting nascent market success.

***** Albuquerque de Sousa is at the Norwegian School of Economics (NHH). Beck is at the Florence School of Banking and Finance at the European University Institute and CEPR. Van Bergeijk is at the International Institute of Social Studies, Erasmus University. Van Dijk is at Rotterdam School of Management, Erasmus University. E-mail addresses: Jose.Albuquerque.de.Sousa@nhh.no, Thorsten.Beck@eui.eu, bergeijk@iss.nl, and madijk@rsm.nl. We thank Stijn Claessens, Jan Dul, Erik Feyen, Andrei Kirilenko, Thomas Lambert, Ross Levine, Svein-Arne Persson, Vincent Somville, Henri Servaes, Roberto Steri (EFA discussant), and seminar participants at the European Finance Association Meetings, the International Institute of Social Studies, the Netherlands Institute for Advanced Study in the Humanities and Social Sciences, and the Rotterdam School of Management for helpful comments.

1. Introduction

Although theory is ambiguous, a large body of empirical research emphasizes the importance of well-developed and efficient financial markets for economic growth, at least in developing and emerging economies.[1] First, financial markets can stimulate the accumulation of capital in the economy (Bencivenga and Smith 1991; Jappelli and Pagano, 1993; O'Hara, 1995; Morck, Yavuz, and Yeung, 2011). Second, financial markets can foster more efficient allocation of capital (Schumpeter, 1934; Rajan and Zingales, 1998; Beck, Levine, and Loayza, 2000; Wurgler, 2000; Fisman and Love, 2004).

Many developing countries have both underdeveloped and concentrated financial systems, dominated by a few banks and lacking liquid capital markets. Some studies suggest that setting up a vibrant stock market in developing countries may be problematic because the stock market pricing process is inherently volatile and arbitrary due to monopolistic abuses and inadequate government regulations (Singh, 1997; Rioja and Valev, 2014). Such stock markets are thus bound to fail, and governments may be better off investing in the development of the banking sector. It is therefore crucial to understand how stock markets develop in the first years after establishment, whether early success is a precondition for long-term success and, more generally, what factors predict long-term nascent market success and failure.

This paper explores conditions for the successful establishment of public equity markets across a sample of developing countries that have opened a stock exchange since 1975. We find that long-term success (as measured by performance 15 to 20 years after establishment) is in part determined by early success: a high initial number of listings and trading activity are necessary, though not sufficient, conditions for long-term success. Furthermore, we find that a small population and low levels of control of corruption, law and order, globalization index, and GDP per capita are necessary but not sufficient conditions for failure. Three socio-economic and political indicators predict the success of 87.2% of the 47 nascent markets in our sample: ethnolinguistic fractionalization, democratic participation, and GDP.

[1] See, e.g., Grossman and Stiglitz (1980), Diamond and Verrecchia (1982), Laffont and Tirole (1988), Scharfstein (1988), Devereux and Smith (1994), Obstfeld (1994), Bencivenga, Smith, and Starr (1996), and Greenwood and Smith (1997) for theoretical arguments. Empirical studies include Levine (1991), Demirgüç-Kunt and Levine (1996, 2001), Levine and Zervos (1998), and Beck and Levine (2004). Levine (1997, 2005) surveys the literature.

Since 1975, the number of countries with at least one stock market has more than tripled, from 53 to 165. The vast majority of academic studies to date (even the "emerging markets" literature) focus on at most 50-60 of these 165 countries, and thus we know little about the (determinants of) development of many recently established ("nascent") stock markets. As of 2022, there were still 49 countries without a stock market, but several are planning to open a stock exchange, so determining the conditions for successful establishment remains an important policy concern.

In our study of the conditions for nascent market success and failure, we specifically gauge: (1) how nascent markets evolve in their initial stages of development, (2) whether early nascent market success is a necessary condition for long-term success, and (3) which factors predict long-term nascent market success. Given the host of potential determinants of stock market development identified in the financial development literature, pinpointing causality for each specific variable is challenging. To alleviate this issue, we use a variety of methods that are novel to this literature (cluster analysis, necessary condition analysis, and machine learning) to address these three questions. In particular, we aim to evaluate the relative importance of structural, economic, and policy determinants of financial sector development that have been studied extensively in the literature.

We analyze the development of nascent markets by using the three measures of stock market development most commonly used in the literature: number of listed domestic companies, aggregate market capitalization as a percentage of GDP, and aggregate turnover of stocks traded (a measure of trading activity). Pagano (1993) shows that market size is positively related to a market's ability to mobilize capital and induce risk-sharing. Turnover captures the amount of trading activity of the market and is often used as a basic indicator of transaction costs or liquidity. In liquid markets, investors can adjust their portfolios quickly and cheaply, which facilitates risk sharing and improves capital allocation (Devereux and Smith, 1994; Obstfeld, 1994). The literature also considers other proxies for financial market development, such as volatility, market concentration, and pricing errors. Demirgüç-Kunt and Levine (1996) show that these measures are positively correlated among relatively mature stock markets. In this paper, we limit our

analysis to three key success measures since they are widely-used and easy to interpret, and since we lack the data to compute the other measures for many of the nascent markets in the sample.

We find substantial variation in the success of different nascent markets, as illustrated in Figure 1, which shows the development of the number of listings on three nascent markets in our sample (Czech Republic, established 1993; Tanzania, 1998; Vietnam, 2003) over the first decade after their establishment. Some markets slowly but steadily come to fruition (Vietnam), others perish after thriving initially (Czech Republic), and yet others essentially remain dormant (Tanzania).

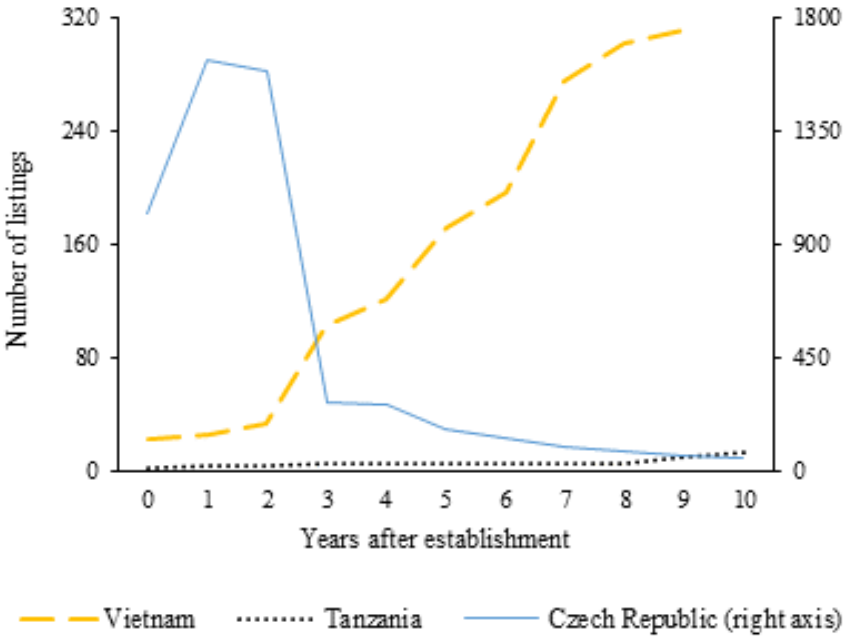


Figure 1. Number of listings in first ten years of select nascent markets

Unlike in previous work that focuses on more established markets (Demirgüç-Kunt and Levine, 1996), we find that correlations among the three success measures are low in the early stage of development of nascent markets, although they increase as markets mature. This suggests that nascent markets may initially thrive according to some measures but not others. It only becomes clear over time which markets succeed in attaining a high number of listings, large aggregate market cap, as well as high trading activity.

Using cluster analysis based on the three success measures simultaneously, we clearly identify two clusters that represent the least and the most successful markets after 20 years of trading. The most successful nascent markets on average fare significantly better according to each of the three success measures than the least successful markets. For example, the stock markets in Kuwait, Poland, and Thailand (in the cluster of most successful markets) have more listings, a greater market capitalization to GDP ratio, and higher turnover after 20 years than the markets in Kazakhstan, Panama, and Tanzania (in the cluster of least successful markets). These results are not materially affected when we scale the number of listings by population or GDP.

Long-term nascent market success is not unambiguously determined in the first years after establishment. Some markets that turn out to be successful after 20 years (such as Qatar) initially score relatively poorly on the success measures, while other markets that score relatively well initially (such as Slovak Republic) perish later.[2] Whether initial success is an important condition for long-term success is thus a relevant policy issue for the debate on whether opening a stock market is sensible when the interest from firms and investors may still be relatively limited. Liquid markets with substantial opportunities for investment and diversification are likely necessary to support investor demand, which may in turn stimulate demand from the corporate sector for exchange listings and raising public equity capital. Similarly, a large aggregate market cap might be a reputation signal necessary to attract investors and issuers alike. In other words, markets with a low initial number of listings, market cap, and turnover could run the risk of becoming dormant, possibly resulting in negative path dependence. Should these countries wait for such interest to develop or could opening a market in an early stage induce the necessary interest from firms and investors to generate an adequate number of listings, market cap, and trading activity in a later stage?

We investigate these issues using necessary condition analysis (NCA; Dul, 2016). In contrast to traditional sufficiency-based statistical methods such as regressions, NCA allows us to identify the conditions that are necessary (but may not be sufficient) for certain outcomes. It is relevant to note that these conditions are examined individually and not simultaneously. We find that a minimum number of

[2] Claessens, Djankov, and Klingebiel (2001) show that the number of listings on the (re)opened stock exchanges of transition economies was high in the early 1990s because of mass privatizations, but often dwindled subsequently.

listings and turnover in the first five years are necessary conditions for success along both of these dimensions after 20 years. Stock markets that start out with few listings and low trading activity fail to attract a considerable number of listings and to spur adequate trading activity in a later stage, and run the risk of quickly becoming dormant. However, there is little evidence that the initial market cap is a necessary condition for long-term success. These results suggest that only liquid markets with substantial opportunities for diversification from the outset are able to generate sufficient interest from firms and investors to thrive. There may indeed be a reputational cost to establishing an idle stock market. This may justify the choice of several countries that, given limited local demand, either postponed opening a stock exchange or decided to join forces and form a regional rather than a national exchange.

We proceed with a more comprehensive analysis of the factors explaining the success and failure of nascent markets. Broadly speaking, the main debate in the financial development literature focuses on the relative importance of "structural factors" – such as demographic and geographic structure (Beck and Feyen, 2013; De La Torre, Feyen, and Ize, 2013), legal origin (La Porta, Lopez-de-Silanes, Shleifer, and Vishny, 1997), social capital (Guiso, Sapienza, and Zingales, 2004), political system (Rajan and Zingales, 2003; Acemoğlu and Johnson, 2005), and other "inherited" characteristics – versus "policy factors" – such as contractual and informational frameworks (Djankov, McLiesh, and Shleifer, 2007), macroeconomic fundamentals (Boyd, Levine, and Smith, 2001), technological development, and regulatory and supervisory frameworks (Beck and Feyen, 2013) – in determining financial development.[3]

We collect data on over 60 variables that are commonly used as empirical proxies for such structural and policy factors (an overview of these variables can be found in Table A1 of the Appendix). Identifying the relation of each individual variable with nascent market success is challenging, since correlations among these variables are often substantial, since the number of observations in our analysis is limited (with annual data for up to around 60 markets), and since these relations may be non-linear.

[3] We acknowledge that the distinction between these two categories is not always clear-cut, since some structural factors could change over the medium- to long-term (e.g., political system), while some policy factors could take a long time to respond to changing policies (e.g., control of corruption).

To alleviate these concerns, we again use necessary condition analysis (NCA) – which is applied to explanatory variables individually – and also machine learning techniques (in particular, random forest analysis; Ho, 1995) – which focus on predicting outcomes rather than on understanding causality and which consider complex non-linear relations as well as interactions between predictive variables. Using these methods, we examine how nascent market success is related to these variables as measured at the time of establishment of the exchange ("initial conditions").

The NCA results uncover no necessary conditions for nascent market success. However, consistent with the literature, a small population, low control of corruption, low level of law and order, low globalization index, and low GDP per capita are each necessary (but not sufficient) conditions for failure. In other words, all of the nascent markets in our sample with low values for the three success measures after 15-20 years have low values for each of these five variables at the time of establishment of the exchange.

Our random forest analysis (using 500 trees) based on 47 nascent markets shows that merely three indicators correctly predict the success of 41 (87.2%) of these markets (where success is a binary indicator based on the cluster analysis): ethnolinguistic fractionalization, democratic participation, and GDP (each measured at the time of nascent market establishment). We note that each of these variables can be classified as a "structural" socio-economic and political factor, as opposed to a policy factor.

Interpreting the nature of the predictive relation of these three variables is not straightforward, since random forest analysis relates these variables to nascent market success in complex, non-linear ways that involve intricate interactions between the variables, and also since the nature of the relations varies across the 500 trees in the forest. That said, we can make various insightful observations based on our results. First, all nascent markets in our sample established in countries with a very low level of democratic participation (close to zero) belong to the cluster of "most successful" markets. The result that nascent markets flourish in autocratic states may seem surprising, but is consistent with Yang (2011), who empirically shows that democracy is not positively related to stock market development. Second, nascent markets established in countries with high levels of democratic participation and high levels of ethnolinguistic fractionalization always belong to the "least successful" cluster. This result is consistent

with the notion that ethnolinguistic fractionalization undermines trust and cooperation and results in less developed institutions and corruption, thereby hurting stock market development (Mauro, 1995; Yenkey, 2015). Third, higher GDP predicts success for most of the nascent markets in our sample, consistent with the notion that bigger economies are better able to sustain flourishing stock markets – although this pattern is not as clearcut for countries in the higher range of democratic participation and in the lower range of ethnolinguistic fractionalization. Fourth, we find that the most successful nascent markets have higher levels of private credit to GDP, in line with prior studies suggesting that banks and stock markets are complements, and that both play an important role in a country's development (Boyd and Smith, 1996; Beck and Levine, 2004; Demirgüç-Kunt, Feyen, and Levine, 2011),

In sum, our analysis shows that, despite the large variety of determinants of stock market development identified by the literature, a combination of merely three socio-economic and political indicators predict success of nascent markets with high accuracy. Moreover, our findings highlight the importance of "structural factors" (as opposed to "policy factors") in predicting nascent market success.

Our paper is related to several strands of the literature. First, our paper is related to the expansive finance and growth literature (see Beck, 2013, and Levine, 2005, for literature surveys). Although recent studies point to important non-linearities in the relation between financial and economic development (e.g., Arcand, Berkes and Panizza, 2015), there is increasing consensus on the critical role of financial sector development for economic growth in low- and middle-income countries (Claessens and Feyen, 2006). Several studies emphasize banking sector development rather than capital markets driving financial and economic development in developing countries (e.g., Demirgüç-Kunt, Feyen, and Levine, 2013), while other studies have pointed to the importance of diversified financial systems, with a variety of different institutions and markets. Specifically, Levine and Zervos (1998) and Beck and Levine (2004) find independent effects of both banking sector and equity market development on economic growth. While this literature typically focuses on already established stock exchanges, we contribute to this literature by gauging the success criteria for newly established stock markets in developing and emerging markets, where the impact of capital market development has been shown to be largest.

Second, our paper is related to the literature on the determinants of financial sector development. This literature examines the role of many structural factors (La Porta et al., 1997; Guiso, Sapienza, and Zingales, 2004; Rajan and Zingales, 2003; Acemoğlu and Johnson, 2005) and economic and policy factors (Boyd et al., 2001; Glaeser, Johnson, and Shleifer, 2001; Djankov et al., 2007; Beck and Feyen, 2013) as determinants of financial development. However, most of the studies in this strand of literature are based on cross-country correlations in a sample of relatively developed countries with stable financial structures. Levine's (1997, p.702) observation that "we do not have a sufficiently rigorous understanding of the emergence, development, and economic implications of different financial structures" still rings true today. Our paper contributes to this literature by analyzing the role of structural and policy factors in the context of developing new segments of the financial systems, notably public equity markets, in a large sample of less-developed countries that have received relatively little attention so far. Although some other studies explore the development of several newly established stock markets (Claessens, Djankov, and Klingebiel, 2001; Minier, 2009; Weber, Davis, and Lounsbury, 2009), we are not aware of prior work that provides a comprehensive analysis of whether initial success as well as structural and policy factors help to explain why some nascent markets succeed while others do not.

Before proceeding, we would like to point to several caveats. First, our analysis is largely based on (partial) correlations and does not imply causality. Specifically, we gauge the predictive power of initial conditions and different structural and policy variables for success indicators of nascent stock markets. Although reverse causation is not necessarily a concern for our analysis, we will refrain from causal interpretations. Nonetheless, we believe that our analyses and findings provide important and novel insights that help researchers and policy makers understand the drivers of stock market development. Second, we focus on nascent markets in general, some of which are actually "re-emerged" markets, i.e., they reopen after having been closed for several decades, mostly due to political constraints. While this distinction matters for computation of long-term returns and volatility measures (Goetzmann and Jorion, 1999), we see this distinction as less significant in terms of scale and liquidity of markets. Third, our work does not

speak to the discussion on the extent to which public equity markets contribute to economic development but rather what are the criteria for a successful development of such markets in the first place.

2. Data and methods

2.1. Data

We collect data on the year of establishment of stock markets around the world, indicators of stock market development, and a host of country-level variables that may help to explain stock market development. Our data sources include the World Development Indicators, the S&P Emerging Markets database, the World Governance Indicators, the Financial Development and Structure Dataset, websites of stock exchanges, academic papers, and several others. Variable definitions and data sources of all variables used in our analyses can be found in Table A1 of the Appendix.

We analyze the development of nascent stock markets in their first 40 years of activity. Our analysis begins in 1975 since that is the first year for which data on our stock market development indicators are available.[4] Table 1 presents an overview of the number and type (national or regional) of markets opened before and after 1975 and Figure 2 presents their geographic distribution. Before 1975, 53 countries had at least one stock exchange. This number has more than tripled in the past 40 years: as of 2021, 165 of the existing 214 countries[5] have at least one stock market, and some of these have more than one.[6] Hence, there are still 49 countries without a stock market as of 2022, but we find evidence that several of these countries have plans to open an exchange. Table 2 shows the list of all 85 nascent markets that are included in at least one of our analyses. We note that some of these 85 markets were established before 1975, but they have data post 1975 that are within the first 40 years of activity. However, most of our analyses include at most

[4] In some stock markets, trading does not immediately start at the official date of stock market establishment. Since we are interested in the stock markets' activity, and turnover is one of our success measures, we also collect information on the year trading started in each market and use this year as the first year in the life of the market.
[5] The number of existing countries is based on the World Bank list of countries, retrieved in September 2014.
[6] Not each of these 165 countries has its own stock market: 23 countries share a regional stock exchange. The largest regional stock exchanges are located in Africa (Bourse Régionale des Valeurs Mobilières, BRVM, in West Africa, and Bourse des Valeurs Mobilières d'Afrique Centrale, BVMAC, in Central Africa). A considerable number of countries (14) re-opened a stock market that had been closed due to the prevalence of a communist regime.

the 70 nascent markets that were opened in 1975 or later, and some analyses use an even smaller sample as we require data on all three success measures over a prolonged period after establishment.

2.1.1. Indicators of stock market development ("success measures")

We use three measures to assess the "success" of a stock market: number of listed domestic companies, aggregate market capitalization to GDP, and aggregate turnover of the stocks traded (total value of stocks traded to average market capitalization). We collect data on these measures from the World Development Indicators (WDI)[7], the S&P Emerging Markets database (EMDB)[8] and websites of the stock exchanges and stock exchange associations. Although the values of our three success measures taken from the different databases and websites are generally highly consistent, we find some slight differences in individual observations, in which cases we take the average values across the different databases. For countries that opened more than one exchange over our sample period (such as China's Shanghai and Shenzhen exchanges; both established 1989), we aggregate success measures across exchanges to obtain indicators of a country's overall stock market development.

2.1.2. Factors explaining nascent market success

We collect data on more than 60 potential factors explaining nascent market success, as shown in Table A1 of the Appendix. We categorize these variables into 6 main categories: geo-demographic indicators, socio-cultural indicators, political and legal indicators, financial indicators, economic indicators, and stock market design indicators. The list of potential factors explaining nascent market success above is not exhaustive. For example, other research suggests that informal institutions such as societal norms (Guiso, Sapienza and Zingales, 2004; Garretsen, Lensink, and Sterken, 2004) and stock market design characteristics such as trading mechanism and transaction taxes (Green, Maggioni, and Murinde, 2000; Kairys, Kruza, and Kumpins, 2000) may in part explain stock market development. However, comparative data on these and

[7] Before 2015, WDI stock market data were sourced from Standard & Poor's. Time series have been replaced in December 2015 with data from the World Federation of Exchanges. We collect both datasets.
[8] Over the period 1975-1995.

other variables for the markets in our sample are scarce. Therefore, we focus on the main groups of potential variables listed above (for which we can obtain data on most nascent markets in our sample), and present some suggestive evidence on several other variables in the Internet Appendix.

2.2. Methods

We use a variety of methods to answer the three main research questions in this paper: (1) how do nascent markets evolve in their initial stages of development?; (2) is early nascent market success a necessary condition for long-term success?; and (3) what factors predict long-term nascent market success? We use correlations, scatter plots, and cluster analysis to address question (1), necessary condition analysis to address question (2), and necessary condition analysis and random forest analysis to address question (3). Since cluster analysis, necessary condition analysis, and random forest analysis appear to be new to the financial development literature, we briefly discuss these methods here.

2.2.1. Cluster analysis

We use cluster analysis (Sneath and Sokal, 1973) to assess whether two or more clusters of relatively less and relatively more successful nascent markets can be distinguished after 20 years of trading based on the three success measures (what we refer to as long-term success). To make sure each of the three measures has equal weight in the clustering, we standardize each of the success measures to the interval [0,1] across the whole period for the cluster analysis. We use the *k*-means method (Hartigan and Wong, 1979) to identify clusters. This method minimizes the within-cluster sum of squared distances to the center of the cluster along the three success measures. We follow the approach of Charrad, Ghazzali, Boiteau, and Niknafs (2014) to determine the optimal number of clusters for our dataset. Their approach uses thirty different methods to determine the optimal number of clusters and selects the number of clusters indicated by the majority of those methods.[9]

[9] We implement this approach using the R-package "NbClust" (Charrad, Ghazzali, Boiteau, and Niknaf, 2014).

2.2.2. Necessary condition analysis

We use necessary condition analysis (NCA; Dul, 2016) to examine whether the early success of nascent markets as well as a host of other potential determinants are a necessary condition for long-term success. The approach of NCA is fundamentally different from the traditional sufficiency-based approach. Traditional multi-causality analysis presumes that each determinant is sufficient to increase the outcome, while none is necessary. In such type of analysis, causality can be expressed in additive models, such as multiple linear regression. In NCA, absence of the necessary determinant results in outcome failure, independently of the value of the other determinants. In other words, NCA is applied to individual variables rather than to multiple variables jointly. The necessary condition must be present for attaining an outcome, although its presence is not sufficient to guarantee that outcome. Traditional sufficiency-based approaches are not appropriate for testing such statements – they only test whether a determinant explains an outcome, not whether it is a necessary condition for the outcome to materialize.

To test whether high levels of each of the success measures in the early stage of a nascent market are necessary conditions for attaining high levels of those measures in the long-term, we follow the method proposed by Dul (2016). We first plot each of our dependent variables (long-term values of success measures) on the y-axis against each of the independent variables (initial values of success measures or other potential determinants) on the x-axis. We then evaluate whether an independent variable is a necessary condition for a dependent variable by examining whether there is an empty area (i.e., without observations) in the top left corner of the corresponding scatter plot, because such an empty area suggests that high values for the dependent variable cannot be attained in case of low values for the independent variable. The larger the empty area, the stronger the evidence for a necessary condition.

The "effect size" is defined as the ratio between the surface of this empty area to the surface of the "scope" of the analysis, which corresponds to total area of the scatter plot, where the borders of this area are defined by the minimum and maximum values of the dependent and independent variables. There are two ways to determine the area size of the empty space in the top left corner of the scatter plot, both of which are based on a "ceiling line" that defines the border of the empty area. First, the "ceiling envelopment

with free disposal hull" (CE-FDH) draws a ceiling line that connects the upper left observations in the scatter plot. In particular, this technique pulls an envelope (piece-wise linear function) along these observations using linear programming. Second, the "ceiling regression with free disposal hull" (CR-FDH) estimates a regression through the upper left observations. We compute effect sizes and their statistical significance (*p*-value) based on both techniques. Furthermore, NCA allows us to calculate a "bottleneck table" for each of the dependent variables. This table displays the minimum percentage of the range of the independent variable (across all observations in the sample) that is necessary to attain a given percentage of the range of the dependent variable.[10]

We use the same approach to determine whether a set of key variables related to the size and demographic structure of the country, as well as its economic development, legal environment, political environment, financial development, and supply of capital are necessary conditions for nascent market success. The initiative to open the stock exchange might also matter – it can either stem from the government, the private sector, or a combination of both.[11] In this analysis, the key variables are measured in the first five years of stock market establishment; the success measures are measured 16-20 years later.

The technique described above allows us to identify necessary conditions for success by analyzing the empty area in the top left corner of the scatter plot linking each variable to each success measure. However, the key variables mentioned above could also be necessary conditions for failure, leading to an empty area in the bottom left corner of the scatter plot. To identify such necessary conditions for failure, we redo the analysis by multiplying the success measures by -1. This allows us to test whether the key variables are a necessary condition for failure.

Similarly, low values of the key variables could be a necessary condition for either success or failure (empty areas in the top and bottom right corners of the scatter plot, respectively). For example, a

[10] We run NCA using the R-package available at https://cran.r-project.org/web/packages/NCA/index.html (downloaded December 2019). We note that just like in sufficiency-based approaches like regressions, endogeneity is a potential concern in NCA. Although reverse causation is not a problem in our application of NCA, the necessary conditions we identify could in part be driven by (unobserved) other factors.

[11] As private initiative may signal sufficient interest from companies (Minier, 2009), nascent markets established by private sector initiative might be more successful in the long-term.

low "law and order" score could be a necessary condition for failure. To include all these possibilities in our study, we also test whether the initial conditions multiplied by -1 are necessary conditions for either success or failure.

In sum, for each potential determinant, we run 4 (2×2) versions of NCA based on whether either low or high values of the determinant are a necessary condition for either success or failure. We note that, in each of these versions of the NCA (and in contrast to the random forest analysis below), success (and failure) are measured using the continuous values of the three success measures (listings, market cap, turnover) – as opposed to a binary success indicator based on the cluster analysis – since NCA cannot be carried out with a binary dependent variable. We will present average results for the four versions of the NCA across the three success measures.

2.2.3. Random forest analysis

NCA allows us to identify necessary conditions for success, but it does not allow us to identify predictors of nascent market success. Most variables that potentially could explain nascent market success or failure are endogenous, which makes causal interpretation problematic. Therefore, we steer away from regression analysis and causal interpretations when identifying predictors of success and failure. Instead, we expand the earlier analysis by using a commonly used machine learning method for classification: the random forest method, first proposed by Ho (1995). This method allows us to conduct a "horse race" with the potential factors predicting nascent market success while avoiding the challenges and limitations of regression analysis, such as the choice of variables to include in the regression model – we simply include all the variables for which we have data in the analysis. Further, we avoid having to use different regressions for each of the success measures: our classification variable is binary, based on the cluster analysis across all three success measures described in Section 2.2.1, and classifies each market as belonging to the "least successful" or "most successful" group. We use all the variables included in the NCA, measured in the first five years of stock market establishment, as potential predictors of success. The random forest analysis indicates which variables are more effective in predicting success, without revealing a coefficient for the

relation, as regression analysis does. To gauge the direction of that relation, we also conduct *t*-tests for the difference in means between the clusters of "least successful" and "most successful" markets.

3. Empirical results

In this section, we first provide a general analysis of the development of nascent markets (Section 3.1). We then present a cluster analysis to distinguish between less and more successful markets (Section 3.2) and a necessary condition analysis of the question whether long-term success requires early success (Section 3.3). In Section 3.4, we analyze the necessary conditions for long-term success or failure based on a broader set of variables. We conclude with the analysis of which of those variables are long-term success predictors in our random forest analysis in Section 3.5.

3.1. How do nascent markets evolve in their initial stage of development?

To obtain a first impression of how nascent markets develop in their first 40 years of activity and of whether the different success measures develop in a similar way, Figure 3 presents pairwise scatter plots of the three success measures in eight 5-year time intervals following market establishment. We first take logs of each of the three success measures to correct for skewness, and then take the average of the logs of the annual values of each measure within each 5-year interval to reduce noise. We keep a market in the sample for a specific 5-year interval if we have at least one annual observation for each success measure for that market in that interval. Panel A of Figure 3 shows the first five years after establishment and Panels B-H show the subsequent 5-year intervals.

Since we aim to exploit the full amount of information on nascent market success in their first 40 years in this analysis, we also include markets that opened before 1975 – as long as they have data for the success measures in at least one of the eight 5-year intervals after establishment. In the most extreme case, Zimbabwe (the first country in Table 2) opened a stock exchange in 1946, and appears in our sample in Panel H, as we have at least one observation (three in total) for all three success measures over the period

1981-1985, the eighth 5-year interval after establishment of the Zimbabwe Stock Exchange. The number of markets included in each of the panels of Figure 3 ranges from 18 to 56 and is reported in the panel title. The cross-country average of the success measure on the x-axis of the three scatter plots in each panel is depicted with a vertical line. We note that caution needs to be applied in direct comparisons of these averages across different panels, since they may be based on different samples. Nonetheless, the scatter plots in Panels A-H of Figure 3 indicate that the average number of listings and the market cap to GDP of the nascent markets in our sample generally increase as markets mature, while the average turnover barely changes over the first 40 years of trading. This could either be because most markets do not become more liquid in the first 40 years, or because the increase in turnover on some markets is counterweighted by a decrease on other markets. We will show in the next subsection that the latter is the case. These initial findings provide pointers that growing success of nascent markets cannot be taken for granted once they have been established. Markets do not necessarily spur higher trading activity over time, although on average they do become larger in terms of market cap.

We proceed by analyzing the correlations between the different success measures. The scatter plots in Panels A-H of Figure 3 also show regression lines and pairwise correlations (ρ) between the three success measures in each of the 5-year intervals. The correlation between the number of listings and turnover is high (between 0.31 and 0.77, see leftmost scatter plots) and increases throughout the first 40 years of activity. However, the correlation between market cap and both the number of listings and turnover (middle and rightmost scatter plots) is close to zero or even negative in the first three 5-year intervals. Some markets start out with very high turnover and relatively low market cap (e.g., Poland), while others start out with large market cap and low turnover (e.g., Jordan). After the first 15 years, these correlations increase and reach levels of around 0.5 towards the end of the 40 years period. As Poland develops, its market cap quickly increases to match its high turnover. As Jordan develops, on the other hand, its level of market cap decreases to match an only slightly improved turnover. In sum, these findings indicate that different success measures can lead to different conclusions about the success of a market in the first 15 years after establishment (consistent with, e.g., Feyen, 2010). One should therefore take care to evaluate whether

nascent markets have succeeded along all three dimensions of success only after this period, when correlations between the measures have become reliably positive.

3.2. Which nascent markets succeed and which ones fail?

The scatter plots in Figure 3 reveal some general patterns in the development of nascent markets, but they are not very informative about the large cross-sectional dispersion in nascent market success. In this section, we examine variation in success across markets. We proceed in two steps. First, we attempt to identify clusters of relatively less and more successful nascent markets based on the long-term values of the three success measures. Then, we examine the values of the success measures in the first five years for these clusters (formed based on long-term success) to gauge how the markets in the least and most successful clusters have developed over time. For these analyses, we use the common sample of 40 nascent markets for which values of the success measures are available in both the first (1-5 years) and fourth (16-20 years) interval. We assess long-term success based on the period of 16-20 years after establishment because Figure 3 shows that correlations between the success measures are all positive and relatively stable after this period and because the number of nascent markets for which we have data declines rapidly after 20 years.

Table 3 presents summary statistics of the three success measures for each of these 40 markets, sorted by the first year of trading. The oldest stock market in this sample is Thailand (established 1975), and the youngest is Guyana (2003). The averages across markets of each success measure for the initial and final 5-year intervals are presented at the bottom of the table. Consistent with the patterns in Figure 3, the average number of listings and market cap increase across these periods (from 60.5 listings and market cap of 10.6% of GDP in the first 5-year interval to 141.7 listings and market cap of 33.8% of GDP in the fourth 5-year interval). Average turnover actually decreases in this sample (from 48.8% in the first 5-year interval to 20.0% in the fourth interval). However, these changes over time in turnover across markets conceal the large differences in the development of individual markets. Turnover increases for 14 markets, decreases for 23 markets, and remains stable in one market (Ghana).

As a first step to analyze variation in success across markets, we apply cluster analyses to the three success measures for the 40 markets in this sample based on the period of 16-20 years after establishment. Following the approach of Charrad et al. (2014) to determine the number of clusters, we find that 12 methods propose two clusters, while other numbers of clusters are proposed by at most four methods. We conclude that the optimal number of clusters is two and do a robustness check with three clusters.

Figure 4 presents the results of the cluster analysis. In Panel A, we plot all 40 nascent markets in this sample along the three dimensions of success as measured after 16-20 years. The x-axis represents the number of listings, the y-axis represents turnover, and the diameter of the circles (indicating the individual markets) represents their market cap to GDP ratio. The plot shows a clear distinction between the two clusters of nascent markets that our analysis identifies (indicated in different colors): a cluster of markets with a relatively high number of listings, large market cap, and high turnover (green), and a cluster of markets with relatively low values for each of these measures (orange). China and Eswatini (formerly named Swaziland) are the two extremes along the three dimensions of success. China has the most successful stock market, with an average of 1,529 listed companies (Shanghai and Shenzhen combined), market cap representing 75.4% of GDP, and turnover of 149.4% over the period of 16-20 years.[12] Eswatini is the least successful market, with an average of 6 listed companies, market cap representing 6.7% of GDP, and almost no trading activity (0.5%) over the period of 16-20 years. Figure IA1 in the Appendix depicts the geographic distribution of the markets in the different clusters.[13]

The two vertical lines in the plot in Panel A of Figure 4 represent the average number of listings for each cluster. The two horizontal lines represent the average turnover for each cluster. The average market cap to GDP for each cluster is indicated in the bottom right corner. On average, the number of

[12] As a comparison, in 2015, the combined number of listings on all U.S. stock exchanges was 4,381, with an aggregate market cap to GDP of 140%, and an aggregate turnover of 165%.

[13] As a robustness check, we redo the cluster analysis using four different ways of scaling the number of listings. Scaling by log population or by log GDP does not alter the results. When scaling by population or by GDP, the difference in the average (scaled) number of listings after 16-20 years across the two clusters is less significant, which makes the classification of some countries as "successful" less clear-cut. However, the allocation of countries between the two clusters remains by-and-large the same. We present these results in Figure IA2 of the Internet Appendix. As a further robustness check, we also use three instead of two clusters. Figure IA3 of the Internet Appendix shows that the same markets are classified as successful, but the least successful markets with very low market cap to GDP are identified as a separate cluster. Overall, our conclusions are the same.

listings after 16-20 years is around 75% higher for the markets in the most successful cluster than for the markets in the least successful cluster (0.60 vs. 0.34).[14] The average market cap is around 50% larger for the most successful markets (0.77 vs. 0.50), while average turnover is almost 3 times higher (0.50 vs. 0.18). These numbers indicate that large differences arise in the success of nascent markets after 16-20 years, especially for number of listings and turnover. Such large differences can have considerable consequences for companies, investors, and economic development more generally, which underlines the importance of understanding the determinants of nascent market success.

These conclusions are supported by Panel B of Figure 4, which shows histograms of the three success measures for the clusters of least and most successful markets separately, and by Table 4, which tests for the statistical significance of the differences in the average success measures after 16-20 years across both clusters. The histograms show that, although there is also substantial variation in the success measures within each cluster, the cluster of most successful markets on average clearly scores better along all three dimensions of success than the cluster of least successful markets, especially for turnover. Table 4 shows that the difference in average values across the two clusters is highly statistically significant for all three success measures.

Next, to provide an initial analysis of how the least and most successful markets develop over time and of the extent to which long-term success is determined by early success, we study the initial success (first 5-year interval) of the markets included in the cluster analysis. Panel A of Figure 5 presents a similar three-dimensional plot of the success measures of the 40 markets as Panel A of Figure 4, but then based on the first five years after establishment. However, the colors of the circles representing the different countries still indicate whether the nascent markets in these countries were in the least or most successful clusters after 16-20 years. At first sight, it is hard to discern a clear pattern. Some markets that are part of the most successful cluster after 16-20 years (such as Qatar) have a comparatively low number of listings and turnover in the first five years, while some markets that start out with relatively high values for the success measures (such as Slovak Republic) end up in the cluster of least successful markets later.

[14] We note that, as discussed in Section 2.2.1, each of the success measures has been standardized over the interval [0,1] across the whole period, facilitating comparison across the measures and across first and fourth 5-year intervals.

One observation that does emerge from Panel A of Figure 5 is that markets with an insufficiently high initial number of listings and turnover in the first five years fail to make it into the cluster of most successful markets after 16-20 years: all the markets in the very bottom left of the graph are part of the least successful cluster. On the other hand, markets that start out small in terms of market cap to GDP, but with a relatively high number of listings and turnover from the outset (such as China) can still develop into markets that are successful along all three dimensions of success in the long-term. These conclusions based on visual inspection are buttressed by the histograms in Panel B of Figure 5 and the tests for the statistical significance of the differences in the average success measures in the first 5-year interval across both clusters in Table 5. Markets that turn out to be successful after 20 years on average do not have a significantly greater market cap to GDP in the first five years of trading (neither statistically nor economically) than markets that end up in the cluster of least successful markets after 20 years. In contrast, the most successful markets after 20 years on average do already have more listings (from both a statistical and economic perspective) and turnover (from an economic perspective;close to statistically significant) in the first five years of trading than the least successful markets. In the next subsection, we test more formally whether early success is necessary to attain long-term success.

3.3. Is early nascent market success a necessary condition for long-term success?

We use necessary condition analysis (NCA; Dul (2016) to assess the importance of initial success for long-term success. We start by plotting each of our dependent variables (number of listings, market cap, and turnover 16-20 years after establishment) on the y-axis against the independent variables (number of listings, market cap, and turnover after 1-5 years) on the x-axis in Panel A of Figures 6-8. Since NCA determines whether condition *x* is a necessary condition for outcome variable *y* in a univariate way, this yields nine (3×3) scatter plots. Each scatter plot shows the observation for each of the 40 nascent markets in this sample in small circles, two ceiling lines demarcating the empty area in the top left corner of the scatter plot (based on both the ceiling envelopment with free disposal hull, CE-FDH, and the ceiling regression with free disposal hull, CR-FDH, methods, see Section 2.2.2), and – for comparison purposes –

a simple OLS regression line. As described in Section 2.2.2, the "effect size" is defined as the ratio between the surface of the empty area and the surface of the total area of the scatter plot. Panel B of Figures 6-8 present the "bottleneck tables" corresponding to the necessary conditions for each of the dependent variables. These tables indicate the percentage of the range of values of each independent variable that is necessary to attain the corresponding percentage of the range of the dependent variables, and also show the effect sizes based on both the CE-FDH and the CR-FDH method.

Figure 6 shows that a high initial number of listings as well as high initial turnover are each a necessary condition for attaining a high number of listings in the long-term. The empty areas at the top left of the first and third scatter plots in Panel A are large relative to the total area (CE-FDH effect sizes of 0.32 and 0.31, respectively, and CR-FDH effect sizes very similar and all statistically significant; see Panel B). Panel B presents the "bottleneck table" of the necessary conditions for attaining a high number of listings. The first column shows the different percentages of the range of the number of listings measured in the 16-20 year interval after the establishment of the stock exchange. Each of the other columns represents the percentage of the range of values of each independent variable that is necessary to attain the corresponding percentage of the range of the number of listings. For example, for a market to attain a number of listings at the 50^{th} percentage of the range of values of number of listings after 16-20 years across markets, it must attain a number of listings at the 29^{th} percentage of the range of values of number of listings after 1-5 years across markets (first column of Panel B). Similarly, to attain a number of listings at the 80^{th} percentage of the range of number of listings after 16-20 years, the number of listings after 1-5 years needs to be at least at the 76^{th} percentage of the range of that variable in that 5-year interval across the 40 markets.

The bottleneck table shows that a considerable number of listings and turnover in the first five years are needed to attain a high number of listings after 16-20 years. Most numbers in the first and third column of Panel B of Figure 6 are similar, suggesting that, in general high initial turnover is as critical in attaining a high number of listings after 16-20 years as a high initial number of listings (third column of Panel B). In contrast, there is little evidence that large initial market cap is a necessary condition for a high number of listings in the long term. The surface of the empty area is relatively small (CE-FDH and CR-FDH effect

sizes only 0.06 and 0.03, respectively; see Panel B), and the bottleneck table in Panel B indicates that a certain initial market cap is only needed to attain a very high (close to the maximum) number of listings after 16-20 years.

Figure 7 presents the NCA results for market cap to GDP after 16-20 years as dependent variable. There is little evidence that high initial values for any of the three success measures are necessary conditions for attaining a large market cap after 16-20 years. The surface of the empty areas in the top left of each of the three scatter plots is relatively small (CE-FDH effects sizes of 0.07, 0.09, and 0.06, respectively; see Panel B) and the bottleneck table in Panel B indicates that relatively large market cap after 16-20 years can be attained without any condition on either initial number of listings, market cap, or turnover. There is some indication that attaining a very large market cap (close to the maximum) after 16-20 years requires some minimum level of market cap after 1-5 years, but, at 16.1, the necessary percentage is relatively small.

Similar to Figure 6, Figure 8 shows that a high initial number of listings as well as high initial turnover are each a necessary condition for attaining high turnover in the long-term. Both the scatter plots in Panel A and the bottleneck table in Panel B indicate that relatively high turnover after 16-20 years cannot be attained without considerable initial levels of number of listings and turnover. The effect sizes for initial number of listings and initial turnover as necessary conditions for long-term turnover are 0.28 and 0.26, respectively. Again, a high initial market cap is not a necessary condition for high turnover after 16-20 years, although very high levels of turnover (close to the maximum) after 16-20 years do not occur without a certain minimum initial market cap (35^{th} percentage of the range of initial market cap across markets, see second column of Panel B).

In sum, both cluster analysis and NCA indicate that a high initial number of listings and turnover are necessary conditions for a market to thrive in the long term. In contrast, the initial market cap is not a necessary condition for long-term success. These findings suggest that it is important for newly established

stock markets to ensure that the market is sufficiently liquid and offers sufficient opportunities for diversification early on to retain the opportunity for long-term success.[15]

3.4. What are further necessary conditions for long-term nascent market success and failure?

We are not only interested in the importance of initial success for long-term nascent market success. As discussed in Section 2.3, the academic literature examines a large number of other potential factors explaining stock market development. In this section, we discuss the results of the necessary condition analysis used to identify further necessary conditions for long-term nascent market success and failure. We focus on a set of key variables related to the size and demographic structure of the country, as well as its economic development, legal environment, political environment, financial development, supply of capital, and the initiative to open the market (government or private). We measure those variables in the first five years of establishment ("initial conditions"). We note that we can include some more countries in our sample than before since we do not require data on initial success.

Next, we consider which variables in Table A1 in the Appendix to include in the NCA. Since the calculation of effect sizes depends on the size of the empty areas in the scatter plots, and the size of such areas depends on the "scope" of the analysis (defined by the minimum and maximum values of the dependent and independent variables), we impose three criteria for the inclusion of a variable in the NCA to avoid generating conclusions for variables for which we have insufficient data or a limited "scope": (i) the number of countries with data available on each potential necessary condition is greater than 50% of the number of countries with data available on each success measure across the whole sample; (ii) the minimum standardized observation (linearly transformed to the [0-1] interval across the whole sample) of each success measure is below 0.2 (in other words, we require a sufficiently low minimum value for the success measures for the sample for which the variable is available); (iii) the maximum standardized observation of each success measure is greater than 0.8 (in other words, we require a sufficiently high

[15] This does not necessarily imply that countries should wait for enough demand from companies and investors to develop before opening a stock market. Taking part in a regional stock exchange could be an alternative. Some authors (Irving, 2005; Piesse and Hearn, 2005) show that certain African countries integrated several markets into a single regional stock exchange to ensure more liquidity than if they would have acted on their own.

maximum value for the success measures for the sample for which the variable is available). Table A3 in the Appendix assesses these criteria for each variable and Table A4 lists all variables that satisfy the criteria (Table IA1 in the Internet Appendix presents NCA results also for variables that do not satisfy the criteria).

Columns 1-4 of Table 6 summarize the results for the variables that satisfy the criteria. Variables are presented in several panels, following the categorization of variables presented in Table A1 of the Appendix. For each analysis, we present the effect sizes and *p*-values calculated using both the CE-FDH and CR-FDH methods. Initial conditions are measured as the average value of each potential determinant in the first five years after stock market establishment. Success is measured as the average number of listings scaled by population, market capitalization, and turnover in the 16-20 years after stock market establishment, expressed in logs (hence, not as a binary variable). Failure is the value of each success measure multiplied by -1. We also test whether the inverse of each variable is a necessary condition for either success or failure, by multiplying the variable by -1.

For each potential determinant (initial condition), we present the average *p*-value across the three measures of success. Columns 1 and 2 present the results of each initial condition as a necessary condition for failure and success, respectively. Columns 3 and 4 present the results of the analysis of the negative of each variable as a necessary condition for failure and success, respectively. *P*-values are indicated in bold when the *p*-value for at least one of the individual success measures is statistically significant. We also indicate the significance of the average *p*-value at the 1%, 5%, and 10% level with asterisks.

The first main result of this analysis is that Columns 1 and 2 show that essentially none of the initial conditions tested is necessary for either success or failure: the only marginally significant (at the 10% level) average *p*-value is the one for the black market premium as a necessary condition for failure. However, it is not necessary for success (as indicated by the insignificant *p*-value in Column 2 across all three measures of success).

That none of the initial conditions tested in Table 6 is a necessary condition for success may seem surprising, given that these variables largely cover the extensive body of literature on the determinants of stock market development. However, Columns 3 and 4 show that the *negative* of the following variables is

a statistically significant (based on the average *p*-value) necessary condition for failure (defined as the negative of the success measures): population, control of corruption, level of law and order, globalization, and GDP per capita. In other words, a small population, low control of corruption, low level of law and order, low globalization index, and low GDP per capita are each necessary (but not sufficient) conditions for failure. Rephrased yet again, all of the nascent markets in our sample with low values for the three success measures after 15-20 years have low values for each of these five variables at the time of nascent market establishment. These five necessary conditions for failure range from geo-demographic, political and legal, to economic characteristics. We note that the empty area on the bottom right of the scatter plots that relate these five variables to nascent market success (or, equivalently, the empty area in the top left of the scatter plots that relate the negative of these variables to failure) is consistent with (and may be responsible for) the positive relation between these variables and stock market development found in the literature.

The novel insight of our study, important for policy purposes, is that ensuring that a country has a high level of these five initial conditions before opening a stock market is no guarantee for success, but a high level of these conditions has not been associated with failure in the past.

3.5. Which initial conditions predict long-term nascent market success?

Necessary condition analysis allows us to identify necessary initial conditions for success, but it does not allow us to identify predictors of nascent market success. Given the identification challenges associated with the use of regression analysis, we abstain from using it. Instead, we conduct a classification random forest analysis to determine which initial conditions best predict long-term nascent market success. The binary classification of nascent markets used for this analysis into least successful and most successful is based on the cluster analysis. Since we are analyzing the factors explaining long-term success, we can include countries for which data on initial success are not available. This allows us to expand our sample from 40 to 58 countries. To verify that expanding the sample does not lead to a different classification of countries as least successful or more successful, we redo the cluster analysis of Figure 4 for this larger

sample of countries. Moreover, we scale listings by population to avoid the size of the country affecting our identification of initial conditions for success. The results of the extended cluster analysis can be found in Figure A1 of the Appendix. Only one country (Nepal) is classified differently from the original cluster analysis. Two countries (North Macedonia and Eswatini) are dropped, because of unavailable data on population. Table A2 in the Appendix shows the additional 20 countries included in the analysis, and their classification, compared to the classification presented in Figure 4. Ten of those additional countries are classified as least successful, and the other ten as most successful.

The variables included in the random forest analysis are the same as the ones included in the NCA (presented in Table A4). To conduct the analysis, we use Breiman and Cutler's Random Forest R-package[16], based on Breiman (2001). To maximize the number of observations in the analysis, we allow for data imputation[17], and conduct an analysis with 500 trees and 5 variables tried at each split. Panel A of Table 7 presents the random forest confusion matrix. This matrix depicts, for each cluster of least successful and most successful countries, the number of countries predicted as least successful and most successful, as well as the classification error (the percentage of incorrect predictions). The model correctly predicts nascent market success for 40 of the 55 countries, which implies an out-of-bag estimation error of 27.3%. Panel B of Table 7 shows how the prediction for each country compares to actual success. The classification error was higher among the most successful markets (33% vs 21% for least successful markets).

The random forest analysis allows us to identify the most important predictors of long-term nascent market success, by calculating the variable importance index. This index captures the total decrease in node impurities from splitting on each variable, averaged over all trees (node impurity is measured by the Gini index), and is therefore a measure for how important each variable is in determining the model's predictions (Breiman, 2001). Column 5 of Table 6 presents the variable importance index of each of the variables included in our random forest analysis. Panel A of Figure 9 presents the values of the importance index for the10 most important predictive variables (in declining order of importance).

[16] Retrieved from https://CRAN.R-project.org/package=randomForest

[17] According to the R-package used, "For continuous predictors, the imputed value is the weighted average of the non-missing obervations, where the weights are the proximities. For categorical predictors, the imputed value is the category with the largest average proximity",

The results indicate that the three most important variables in predicting nascent market success are ethnolinguistic fractionalization, GDP, and democratic participation. Three other indicators of a country's degree of democracy (democratic competition, democracy score from Polity, and the Van Hanen democracy index) also show up in the top 10 of variable importance. The size of the banking sector (measured by the percentage of private credit to GDP) is a further important predictor of success. The top 10 is complemented by globalization, the black market premium, and GDP growth.

A limitation of the variable importance index is that it does not identify the nature of the relation between the predictive variable and the outcome (nascent market success or failure). This relation could be either positive or negative, and is likely non-linear and/or dependent on interactions with other variables. To gauge the overall direction of the relation between each variable and long-term nascent market success in a univariate way, we conduct *t*-tests for the difference in the means of each predictive variable between the clusters of "least successful" and "most successful" markets.

Columns 6-8 of Table 6 present the results of these *t*-tests. The results suggest an overall negative association between ethnolinguistic fractionalization and nascent market success: the most successful nascent markets have on average a 0.21 points lower level of ethnolinguistic fractionalization than the least successful nascent markets, and this difference is significant at the 1% level. Remarkably, the level of democracy of a country also relates negatively to success – the average level of all "democracy" indicators included in the analysis is lower for the cluster of most successful nascent markets than for the cluster of least successful nascent markets. This difference is significant at the 5% level for both the "Polity" and van Hanen's "democratic competition" indicator. GDP and GDP per capita, on the other hand, are significantly higher for the most successful cluster. Furthermore, the *t*-tests show that countries with the least successful stock exchanges are more likely to be landlocked, have a higher percentage of Catholics, and have been established by private initiative. The countries with the most successful stock exchanges are more often exporters of fuels, more densely populated, have a higher percentage of Muslims, and higher levels of control of corruption, law and order, private credit, and globalization.

As discussed above, the relation between the initial conditions identified by the random forest analysis and nascent market success is not necessarily linear. Although the *t*-test results in Table 6 allow us to say something about the nature of the predictive relations, we also present a scatter plot to show how the three variables with the highest importance index (ethnolinguistic fractionalization, GDP, and democratic participation) relate to each other and to nascent market success in Panel B of Figure 9. The vertical axis depicts ethnolinguistic fractionalization, the horizontal axis depicts democratic participation, and the size of the circles is proportional to the country's GDP. Orange circles represent countries in the cluster of least successful markets, green circles represent countries in the cluster of most successful markets.

The figure shows a clear pattern in the relation between the level of democratic participation and nascent market success: all nascent markets in our sample established in countries with a very low level of democratic participation (close to zero) belong to the cluster of most successful markets. This result may be viewed as surprising, but is consistent with the findings of Yang (2011), who shows that democracy is not positively related to stock market development. Countries with higher levels of democratic participation *and* high levels of ethnolinguistic fractionalization mostly belong to the least successful cluster. This result is consistent with the notion that ethnolinguistic fractionalization harms trust and cooperation and is associated with corruption and less developed institutions, thereby hampering stock market development (Mauro, 1995; Yenkey, 2015). The pattern is not as clearcut for countries in the higher range of democratic participation *and* in the lower range of ethnolinguistic fractionalization, since there we observe a mix of green and orange circles. Furthermore, the observation that green circles tend to be bigger than orange circles indicates that the nascent stock markets of larger economies (countries with higher GDP) tend to be more successful.

A visual inspection of the scatter plot suggests that these three variables by themselves seem to predict nascent market success with considerable accuracy. To assess how the predictive power of our model changes when discarding all other predictors, we redo the classification random forest analysis using these three predictive variables only. Table 8 presents the results of the analysis based on imputed data, and 500 trees with 1 variable tried at each split. Panel A presents the random forest confusion matrix. With only

three variables, the same number of markets is correctly predicted. The classification error among the most successful markets decreased slightly (from 33% to 32%), whereas the classification error among the least successful markets slightly increased (from 21% to 22%). Panel B shows how the prediction for each country compares to actual success. All in all, the results in Table 8 indicate that these three variables combined are the most accurate predictors of the observed long-term success, since including all of the other variables adds virtually nothing to prediction success.

The analysis above is, as previously mentioned, partially based on imputed data. To verify that the predictions of the model hold without such imputation, we manually construct a simplified decision tree to predict nascent market success in Figure 10, following the analysis reported in Table 8. The decision tree is based on the observed values of each predictive variable in the first five years after the stock market is established. Therefore, it classifies a sub-sample of 47 countries with available data on ethnolinguistic fractionalization, GDP, and democratic participation in these first five years after market establishment. Column 4 of Table A2 presents an overview of these countries and the corresponding prediction. The colors indicate whether the country belongs to the cluster of least successful (orange) or most successful (green) countries.

The decision tree in Figure 10 allows us to correctly predict nascent market success in 41 out of the 47 nascent markets (87.2%). The figure also illustrates the types of interactions that random forest analysis considers. Low democratic participation (values <10.5%) always correctly predicts success. In other words, the 9 countries with a value for the democratic participation variable below 10.5% all belong to the cluster of most successful nascent markets. Second, for countries with higher democratic participation, high ethnolinguistic fractionalization (>50%) always correctly predicts failure. In other words, the 17 countries in our sample that have relatively high democratic participation *and* relatively high ethnolinguistic fractionalization are correctly classified as belonging to the cluster of least successful markets. Third, GDP helps to distinguish the other cases (countries with higher democratic participation but lower ethnolinguistic fractionalization), but not perfectly. In those cases, high GDP correctly predicts success in 66.7% of the cases (6/9 countries), whereas low GDP correctly predicts failure in 75% of the cases (9/12 countries).

To conclude, our random forest analysis shows that, despite the large variety of determinants of stock market development identified by the literature, a combination of merely three socio-economic and political indicators predicts nascent market success with high accuracy. Moreover, this analysis highlights the importance of "structural factors" (as opposed to "policy factors") in predicting nascent market success.

4. Conclusions

Given the importance of financial sector development for economic development, understanding the determinants of the successful establishment of public equity markets is relevant from both an academic and a policy perspective. In this paper, we analyze the development of newly established ("nascent") stock markets in their first years of activity, and examine the determinants of long-term nascent market success. We are particularly interested in the question whether various "structural factors" and "policy factors" identified by the financial development literature as important determinants of financial market development can help explain the success of this sample of relatively understudied stock markets.

We find that nascent markets show very different levels of success, and that the correlation between the three success measures used in this paper (number of listings, market cap to GDP, and turnover) steadily increases as markets become more mature. We show that a minimum initial number of listings and initial turnover are necessary but not sufficient conditions for long-term nascent market success along these dimensions. We further show that a small population, low control of corruption, low level of law and order, low globalization index, and low GDP per capita are each necessary but not sufficient conditions for failure.

Three socio-economic and political structural factors at the time of nascent market establishment correctly predict the success of 87.2% of the 47 nascent markets in our sample: ethnolinguistic fractionalization, GDP, and democratic participation. All nascent markets in our sample established in countries in which the level of democratic participation is close to zero belong to the cluster of "most successful" markets. Countries with higher levels of democratic participation and relatively high levels of ethnolinguistic fractionalization always belong to the "least successful" cluster. The pattern is not as clear

for countries in the higher range of democratic participation and in the lower range of ethnolinguistic fractionalization, but higher GDP does predict success for this group of nascent markets.

In sum, our finding that long-term success can be predicted quite accurately in the initial stage of a market's development suggests that stock markets thrive when they are established at the right stage of a country's development. The initial number of listings and liquidity are necessary conditions for success, while socio-economic and political structural factors help to predict long-term success with relatively greater accuracy than policy factors.

References

Acemoğlu, D., S. Johnson, J.A. Robinson, and Y. Thaicharoen, 2003, "Institutional causes, macroeconomic symptoms: Volatility, crises and growth," *Journal of Monetary Economics* 50, 49-123.

Acemoğlu, D. and S. Johnson, 2005, "Unbundling institutions," *Journal of Political Economy* 113, 949-995.

Aghion, P., Y. Algan, P. Cahuc, and A. Shleifer, 2010, "Regulation and distrust," *Quarterly Journal of Economics* 125, 1015-1049.

Arcand, J.L., E. Berkes, and U. Panizza, 2015, "Too much finance?," *Journal of Economic Growth* 20, 105-148.

Bhattacharya, U., and H. Daouk, 2002, "The world price of insider trading," *Journal of Finance* 57, 75-108.

Bekaert, G., C.R. Harvey, and C. Lundblad, 2005, "Does financial liberalization spur growth?," *Journal of Financial Economics* 77, 3-55.

Beck, T., 2013, "Finance, growth and fragility: The role of government," *International Journal of Banking, Accounting and Finance* 5, 49-77.

Beck, T., and E. Feyen, 2013, "Benchmarking financial systems: Introducing the financial possibility frontier," *World Bank Policy Research Working Paper* 6615.

Beck, T., A. Demirgüç-Kunt, and R. Levine, 2003, "Law, endowments and finance," *Journal of Financial Economics* 70, 137-181.

Beck, T., and R. Levine, 2004, "Stock markets, banks, and growth: Panel evidence," *Journal of Banking and Finance* 28, 423-442.

Beck, T., R. Levine, and N. Loayza, 2000, "Finance and the sources of growth," *Journal of Financial Economics* 58, 261-300.

Bencivenga, V.R., and B.D. Smith, 1991, "Financial intermediation and endogenous growth," *Review of Economic Studies* 58, 195-209.

Bencivenga, V.R., B.D. Smith, and R.M. Starr, 1996, "Equity markets, transaction costs, and capital accumulation: An illustration," *World Bank Economic Review* 10, 241-265.

Boyd, J., R. Levine, and B. Smith, 2001, "The impact of inflation on financial sector performance," *Journal of Monetary Economics* 47, 221-248.

Boyd, J., and B. Smith, 1996, "The coevolution of the real and financial sectors in the growth process," *World Bank Economic Review* 10, 371-396.

Breiman, L., 2001, "Random Forests," *Machine Learning* 45, 5 - 32

Charrad, M., N. Ghazzali, V. Boiteau, and A. Niknafs, 2014, "NbClust: An R package for determining the relevant number of clusters in a data set," *Journal of Statistical Software* 61, 1-36.

Claessens, S., S. Djankov, and D. Klingebiel, 2001, "Stock markets in transition economies," in Bokros, L., A. Fleming, and C. Votava, Eds, *Financial transition in Europe and Central Asia. Challenges of the new decade*, The World Bank, Washington D.C.

Claessens, S., and E. Feyen, 2006, "Financial sector development and the millennium development goals," World Bank Working Paper No. 89, World Bank, Washington, D.C.

De La Torre, A., E. Feyen, and A. Ize, 2013, "Financial development: Structure and dynamics," *The World Bank Economic Review*.

Demirgüç-Kunt, A., E. Feyen, and R. Levine, 2011, "Optimal financial structures and development: The evolving importance of banks and markets," *World Bank Strategy Research Working Paper* (5805). Demirgüç-Kunt, A., E. Feyen, and R. Levine, 2013, "The evolving importance of banks and securities markets," *World Bank Economic Review* 27, 476-490.

Demirgüç-Kunt, A., and R. Levine, 1996, "Stock market development and financial intermediaries: Stylized facts," *World Bank Economic Review* 10, 291-321.

Demirgüç-Kunt, A., and R. Levine, 2001, "Financial structure and economic growth: Overview," in Demirgüç-Kunt, A. and R. Levine, Eds., *Financial structure and economic growth: a cross-country comparison of banks, markets, and development*, MIT Press, Cambridge, MA.

Devereux, M.B., and G.W. Smith, 1994, "International risk sharing and economic growth," *International Economic Review* 35, 535-50.

Diamond, D.W., and R.E. Verrecchia, 1982, "Optimal managerial contracts and equilibrium security prices," *Journal of Finance* 37, 275-287.

Djankov, S., C. McLiesh and A. Shleifer, 2007, "Private credit in 129 countries," *Journal of Financial Economics* 84, 299-329.

Dul, J., 2016, "Necessary condition analysis (NCA): Logic and methodology of "necessary but not sufficient" causality," *Organizational Research Methods* 19, 10-52.

Feyen, E., 2010, "Finances of Egyptian listed firms and the performance of the Egyptian Stock Exchange," *World Bank Policy Research Working Paper* (5213).

Fisman, R. and I. Love, 2004, "Financial development and intersectoral allocation: A new approach," *Journal of Finance* 59, 2785-2807.

Garretsen, H., R. Lensink, and E. Sterken, 2004, "Growth, financial development, societal norms and legal institutions," *Journal of International Financial Markets, Institutions and Money* 14, 165-183.

Glaeser, E., S. Johnson, and A. Shleifer, 2001, "Coase versus the Coasians," *Quarterly Journal of Economics* 116, 853-899.

Goetzmann, W. and P. Jorion, 1999, "Re-emerging markets," *Journal of Financial and Quantitative Analysis* 34, 1-32.

Green, C.J., P. Maggioni, V. Murinde, 2000, "Regulatory lessons for emerging stock markets from a century of evidence on transactions costs and share price volatility in the London Stock Exchange," *Journal of Banking and Finance* 24, 577-601.

Greenwood, J., and B.D. Smith, 1997, "Financial markets in development, and the development of financial markets," *Journal of Economic Dynamics and Control* 21, 145-181.

Grossman, S.J., and J.E. Stiglitz, 1980, "On the impossibility of informationally efficient markets," *American Economic Review* 70, 393-408.

Guiso, L., P. Sapienza, and L. Zingales, 2004, "The role of social capital in financial development," *The American Economic Review* 94, 526-556.

Hartigan, J.A., and M.A. Wong, 1979, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society* 28, 100-108.

Ho, T. K., 1995, "Random Decision Forests", *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, pp. 278–282.

Hofstede, G., G.J. Hofstede, and M. Minkov, 2010, *Cultures and organizations: Software of the mind*,. McGraw-Hill, New York.

Irving, J., 2005, "Regional integration of stock exchanges in Eastern and Southern Africa: Progress and prospects," Working paper, IMF.

Jappelli, T., and M. Pagano, 1993, "Savings, growth, and liquidity constraints," *Quarterly Journal of Economics* 108, 93-109.

Kairys Jr., J.P., R. Kruza, R. Kumpins, 2000, "Winners and losers from the introduction of continuous variable price trading: Evidence from the Riga Stock Exchange," *Journal of Banking and Finance* 24, 603-624.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R.W. Vishny, 1997, "Legal determinants of external finance," *Journal of Finance* 52, 1131-1150.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R.W. Vishny. 1999. "The quality of government," *Journal of Law, Economics and Organization* 15, 222-279.

Laffont, J.-J., and J. Tirole, 1988, "Repeated auctions of incentive contracts, investment, and bidding parity with an application to takeovers," *RAND Journal of Economics* 19, 516-537.

Levine, R., 1991, "Stock markets, growth and tax policy," *Journal of Finance*, 46, 1445-1465.

Levine, R., 1997, "Financial development and economic growth: Views and agenda," *Journal of Economic Literature* 35, 688-726.

Levine, R., 2005, "Finance and growth: Theory and evidence," in Aghion, P., and S.N. Durlauf, Eds., *Handbook of Economic Growth*, Elsevier, Amsterdam.

Levine, R., and S. Zervos, 1998, "Stock markets, banks, and economic growth," *American Economic Review* 88, 537-558.

Mauro, P., 1995, "Corruption and growth," *Quarterly Journal of Economics* 110(3), 681-712.

Minier, J., 2009, "Opening a stock exchange," *Journal of Development Economics* 90, 135-143.

Morck, R., D. Yavuz, and B. Yeung, 2011, "Banking system control, capital allocation and economy performance," *Journal of Financial Economics* 100, 264-283.

Obstfeld, M., 1994, Risk-taking, global diversification and growth," *American Economic Review* 85, 1310-1329.

O'Hara, M., 1995, *Market microstructure theory*, Blackwell, Malden, MA.

Pagano, M., 1993, "The flotation of companies on the stock market: A coordination failure model," *European Economic Review* 37, 1101-1125.

Piesse, J., and B. Hearn, 2005, "Regional integration of equity markets in Sub-Saharan Africa," *South African Journal of Economics* 73, 36-53.

Rajan, R., and L. Zingales, 1998, "Financial dependence and growth," *American Economic Review* 88, 559-586.

Rajan, R., and L. Zingales, 2003, "The great reversals: The politics of financial development in the 20th century," *Journal of Financial Economics* 69, 5-50.

Rioja, F., and N. Valev, 2014, "Stock markets, banks and the sources of economic growth in low and high income countries," *Journal of Economics and Finance* 38, 302-320.

Scharfstein, D., 1988, "The disciplinary role of takeovers," *Review of Economic Studies* 55, 185-99.

Schumpeter, J., 1934, *The theory of economic development*, Harvard University Press, Cambridge, MA.

Singh, A., 1997, "Financial liberalization, stock markets, and economic development," *Economic Journal* 107, 771-782.

Sneath, P., and R. Sokal, 1973, *Principles of numerical taxonomy*, Freeman, San Francisco.

Weber, K., G.F. Davis, and M. Lounsbury, 2009, "Policy as myth and ceremony? The global spread of stock exchanges, 1980-2005," *Academy of Management Journal* 52, 1319-1347.

Wurgler, J., 2000, "Financial markets and the allocation of capital," *Journal of Financial Economics* 58, 187-214.

Yang, B., 2011, "Does democracy foster financial development? An empirical analysis," *Economics Letters* 112(3), 262-265.

Yenkey, C.B., 2015. "Mobilizing a market: Ethnic segmentation and investor recruitment into the Nairobi Securities Exchange." *Administrative Science Quarterly* 60(4), 561-595.

Table 1. Overview of number of existing stock exchanges

This table presents an overview of the number of countries that established a stock exchange before and after 1975 (the first years for which data on our success measures are available) as well as the number of countries that do not have a stock exchange yet. The second column shows the number of exchanges present in those countries. We refer to Section 3 for a description of the data sources.

	Number of exchanges	**Number of countries**
Established before 1975	1 or more exchanges	53 countries
Re-established after 1975	2 or more exchanges	4 countries
	1 exchange	10 countries
Established after 1975	2 or more exchanges	5 countries
	1 exchange	70 countries
	regional exchange	23 countries
Not established yet	–	49 countries (14 countries have plans to establish an exchange)

Table 2. Overview of all 85 nascent markets in the sample

This table presents all 85 nascent markets included in at least one of the empirical analyses in this paper. Columns present the name of the country, the year when trading started, and the name of the first stock exchange(s) established (or re-opened) in that country. Countries are ordered by the year in which trading started**.**

Country	Year	Stock exchange(s)
Zimbabwe	1946	Zimbabwe Stock Exchange
Venezuela	1947	Bolsa de Valores de Caracas
Israel	1953	Tel Aviv Stock Exchange
Bangladesh	1954	Dhaka Stock Exchange
Kenya	1954	Nairobi Securities Exchange
Korea, Rep.	1956	Korea Exchange
Nigeria	1961	Nigerian Stock Exchange
Malaysia	1964	Bursa Malaysia
Iran	1967	Tehran Stock Exchange
Jamaica	1968	Jamaica Stock Exchange
Ecuador	1969	Bolsa de Valores de Guayaquil
		Bolsa de Valores de Quito
Tunisia	1969	Bourse de Tunis
Bermuda	1971	Bermuda Stock Exchange
Singapore	1973	Singapore Exchange
New Zealand	1974	New Zealand Stock Exchange
Thailand	1975	Stock Exchange of Thailand
Costa Rica	1976	Bolsa Nacional de Valores
El Salvador	1976	Bolsa de Valores de El Salvador
Jordan	1978	Amman Stock Exchange
Fiji	1979	South Pacific Stock Exchange
Trin. & Tobago	1981	Trinidad and Tobago Stock Exchange
Iceland	1985	Iceland Stock Exchange
Kuwait	1985	Kuwait Stock Exchange
Saudi Arabia	1985	Tadawul
Barbados	1987	Barbados Stock Exchange
Guatemala	1987	Bolsa de Valores Nacional
Oman	1988	Muscat Securities Market
Bahrain	1989	Bahrain Stock Exchange
Bolivia	1989	Bolsa Boliviana de Valores
Botswana	1989	Botswana Stock Exchange
Mauritius	1989	Stock Exchange of Mauritius
China	1990	Shanghai Stock Exchange
		Shenzhen Stock Exchange
Ghana	1990	Ghana Stock Exchange
Honduras	1990	Bolsa Hondureña de Valores
Hungary	1990	Budapest Stock Exchange
Panama	1990	Bolsa de Valores de Panama
Slovenia	1990	Ljubljana Stock Exchange
Eswatini	1990	Eswatini Stock Exchange
Bulgaria	1991	Bulgarian Stock Exchange
Croatia	1991	Zagreb Stock Exchange
Mongolia	1991	Mongolian Stock Exchange
Poland	1991	Warsaw Stock Exchange
Russian Fed.	1991	Stock Exchange Saint-Pietersburg
Serbia	1991	Belgrade Stock Exchange
Malta	1992	Malta Stock Exchange
Namibia	1992	Namibian Stock Exchange
Ukraine	1992	PFTS Ukraine Stock Exchange
Czech Republic	1993	Prague Stock Exchange
Lithuania	1993	NASDAQ OMX Vilnius
Montenegro	1993	Montenegro Berza
Nepal	1993	Nepal Stock Exchange
Paraguay	1993	Bolsa de Valores de Paraguay
Slovak Republic	1993	Bratislava Stock Exchange
Kyrgyz Republic	1994	Kyrgyz Stock Exchange
Uzbekistan	1994	Tashkent Stock Exchange
Zambia	1994	Lusaka Stock Exchange
Latvia	1995	NASDAQ OMX Riga
Moldova	1995	Moldova Stock Exchange
Romania	1995	Bursa de Valori București
Sudan	1995	Khartoum Stock Exchange
W. Bank & Gaza	1995	Palestine Exchange
Cyprus	1996	Cyprus Stock Exchange
Estonia	1996	NASDAQ OMX Tallinn
Kazakhstan	1996	Kazakhstan Stock Exchange
Macedonia	1996	Macedonian Stock Exchange
Malawi	1996	Malawi Stock Exchange
Cayman Islands	1997	Cayman Islands Stock Exchange
Qatar	1997	Qatar Exchange
Channel Islands	1998	Channel Islands Stock Exchange
Côte d'Ivoire	1998	Bourse Régionale des Valeurs Mobilières
Tanzania	1998	Dar es Salaam Stock Exchange
Uganda	1998	Uganda Securities Exchange
Algeria	1999	Algiers Stock Exchange
Mozambique	1999	Bolsa de Valores de Mocambique
Papua N. Guinea	1999	Port Moresby Stock Exchange
Georgia	2000	Georgian Stock Exchange
U.A.E.	2000	Abu Dhabi Securities Exchange
		Dubai Financial Market
Vietnam	2000	Ho Chi Minh City Stock Exchange
Armenia	2001	NASDAQ OMX Armenia
Cameroon	2001	Douala Stock Exchange
St. Kitts & Nevis	2001	Eastern Caribbean Securities Exchange
Guyana	2003	Guyana Stock Exchange
Rwanda	2005	Rwanda Stock Exchange
Libya	2007	Libyan Stock Market
Sierra Leone	2009	Sierra Leone Stock Exchange

Table 3. Success measures for 40 nascent markets included in cluster analysis

(1-5 years and 16-20 years after establishment)

This table presents the year in which trading started, average number of listings, market capitalization (% GDP) and turnover ratio (%) in the first -year interval (1-5 years) and the fourth 5-year interval (16-20 years) after the year trading started, for all 40 nascent in the sample that have data on all three success measures in both intervals and are thus included in the cluster analysis. Countries are ordered by the year in which trading started. The bottom row presents averages for each measure and each interval. We refer to Table A1 of the Appendix for variable definitions and data sources.

Country	**1st year of trading**	**# of Listings**		**Market cap**		**Turnover**	
		1-5y	***16-20y***	***1-5y***	***16-20y***	***1-5y***	***16-20y***
Thailand	1975	42.0	389.5	3.9	58.8	126.8	92.2
Jordan	1978	73.4	111.3	43.0	73.2	9.6	14.8
Kuwait	1985	62.8	89.8	52.5	89.2	20.0	45.5
Barbados	1987	14.0	19.6	15.2	124.2	3.1	12.8
Oman	1988	56.3	142.3	8.9	42.3	12.9	24.8
Botswana	1989	10.1	23.0	6.6	33.1	6.3	2.4
Mauritius	1989	20.7	57.1	13.6	54.2	3.5	6.0
China	1990	135.0	1529.3	4.9	75.4	184.2	149.4
Ghana	1990	15.0	31.5	8.4	20.9	3.7	3.7
Hungary	1990	23.5	41.6	2.1	26.7	14.5	95.6
Panama	1990	25.0	24.6	5.5	31.9	6.5	2.7
Eswatini	1990	50.2	6.0	11.8	6.7	7.3	0.5
Bulgaria	1991	15.1	381.1	0.3	25.3	10.5	15.0
Croatia	1991	24.1	281.1	3.2	58.6	9.9	6.8
Poland	1991	30.7	394.4	1.7	36.3	128.6	44.5
Russian Federation	1991	91.5	437.3	0.7	75.4	174.0	67.9
Malta	1992	4.1	19.0	6.3	45.1	10.0	1.4
Namibia	1992	7.4	15.0	4.6	10.8	8.4	2.1
Czech Republic	1993	564.3	16.5	17.3	20.1	33.5	41.9
Lithuania	1993	357.8	37.2	7.3	10.7	21.5	5.9
Nepal	1993	85.8	180.2	5.3	31.4	3.9	3.3
Paraguay	1993	32.3	64.7	1.9	2.5	13.9	3.1
Slovak Republic	1993	179.7	101.0	9.1	5.1	105.1	4.1
Zambia	1994	5.9	19.8	8.1	30.8	1.2	2.4
Latvia	1995	47.5	32.0	3.3	4.3	21.2	3.0
Romania	1995	12.2	327.8	1.1	11.0	34.8	12.2
West Bank and Gaza	1995	20.8	45.5	15.2	25.7	15.8	11.4
Cyprus	1996	58.7	98.1	37.3	11.6	72.5	6.9
Estonia	1996	24.5	15.5	23.7	8.7	52.2	10.8
Kazakhstan	1996	16.3	71.0	7.4	13.8	2.5	4.1
Macedonia	1996	2.8	32.0	0.6	5.6	776.5	6.9
Malawi	1996	8.0	14.0	10.8	24.1	4.2	3.4
Qatar	1997	16.0	42.8	33.4	84.3	6.7	16.9
Côte d'Ivoire	1998	38.0	72.0	11.8	14.0	2.4	2.5
Uganda	1998	2.5	16.0	0.7	38.2	2.4	1.5
Algeria	1999	3.0	4.8	0.4	0.2	0.8	2.3
United Arab Emirates	2000	34.0	125.7	18.3	60.4	4.3	22.6
Vietnam	2000	24.0	323.7	0.4	21.4	31.2	36.0
Armenia	2001	174.2	20.3	0.6	2.5	6.7	8.4
Guyana	2003	11.0	13.0	15.9	37.3	1.1	0.5
Average		**60.5**	**141.7**	**10.6**	**33.8**	**48.8**	**20.0**

Table 4. Average success measures of least and most successful nascent markets

(16-20 years after establishment)

This table presents the results of *t*-tests of the significance of the difference in means of each of the three success measures (number of listings, market cap to GDP, and turnover) over the period 16-20 years after establishment between the two clusters of least and most successful nascent markets from Panel A of Figure 4. Success measures are expressed in logs and then standardized to the interval [0,1] across the whole period 1-20 years after establishment to facilitate comparison across measures and time periods. The clusters are formed based on the values of the three success measures over the period 16-20 years after establishment. We refer to Figure 4 for more information on the cluster analysis. This table reports the mean of each success measure for each cluster, the difference between the means over the period 16-20 years after establishment, and the *p*-value of the difference. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. We refer to Table A1 of the Appendix for variable definitions and data sources.

	Mean "most successful" cluster	**Mean "least successful" cluster**	**Difference (*p*-value)**
Number of listings	0.60	0.34	0.27*** (0.000)
Market cap	0.77	0.50	0.27*** (0.000)
Turnover	0.50	0.18	0.32*** (0.000)

Table 5. Average success measures of least and most successful nascent markets

(1-5 years after establishment)

This table presents the results of *t*-tests of the significance of the difference in means of each of the three success measures (number of listings, market cap to GDP, and turnover) over the period 1-5 years after establishment between the two clusters of least and most successful nascent markets from Panel A of Figure 4. Success measures are expressed in logs and then standardized to the interval [0,1] across the whole period 1-20 years after establishment to facilitate comparison across measures and time periods. The clusters are formed based on the values of the three success measures over the period 16-20 years after establishment. We refer to Figure 4 for more information on the cluster analysis. The table reports the mean of each success measure for each cluster, the difference between the means over the period 1-5 years after establishment, and the *p*-value of the difference. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. We refer to Table A1 of the Appendix for variable definitions and data sources.

	Mean "most successful" cluster	**Mean "least successful" cluster**	**Difference (*p*-value)**
Number of listings	0.39	0.27	0.12*** (0.026)
Market cap	0.39	0.35	0.04 (0.616)
Turnover	0.44	0.34	0.11 (0.143)

Table 6. Initial conditions for stock market success

This table summarizes the results of the different analyses aimed at identifying initial conditions for nascent market success. Initial conditions are measured as the average value of each variable in the first five years after stock market establishment. Success is measured as the average number of listings scaled by population, market capitalization, and turnover in the 16-20 years after stock market establishment. Columns 1-4 of this table present the results of the necessary condition analysis (NCA). For each initial condition, we present the average *p*-value across the three measures of success (reported individually in Table IA1 of the Internet Appendix). Columns 1 and 2 present the results of each initial condition as a necessary condition for failure (success measure multiplied by -1) and success. Columns 3 and 4 present the results of the analysis of the "negative" of each variable (variable multiplied by -1) as a necessary condition for failure and success. *P*-values are indicated in bold when the *p*-value for at least one of the individual success measures is statistically significant. Column 5 presents the variable importance index of the random forest analysis. Columns 6-8 present the results of the *t*-tests for difference in means between the cluster of least successful (L) and most successful (M) countries. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. We refer to Table A1 of the Appendix for variable definitions and data sources.

	NCA				**R. forest**	***t*-test**		
	Positive		**Negative**					
Initial conditions	**Failure** (1)	**Success** (2)	**Failure** (3)	**Success** (4)	**Variable imp.** (5)	**L** (6)	**M** (7)	**diff** (8)
Geo-demographic								
Landlocked	1.000	0.758	0.473	**0.571**	0.060	0.39	0.12	**-0.27****
Natural resources	**0.349**	0.464	**0.382**	0.673	0.646	0.65	0.79	0.15
Exporter of fuels	0.711	1.000	0.771	0.641	0.014	0.07	0.20	**0.13***
Population	**0.581**	**0.387**	**0.029****	0.829	0.696	15.38	15.43	0.05
Population density	0.538	**0.402**	**0.296**	0.521	0.881	3.63	4.30	**0.67***
Ethno-ling. fractional.	0.483	0.671	0.532	**0.377**	**1.965**	0.55	0.34	**-0.21*****
Catholic (%)	0.619	0.848	0.468	**0.224**	0.910	32.43	16.07	**-16.36****
Protestant (%)	0.812	0.714	0.626	0.716	0.884	13.73	8.30	-5.44
Muslim (%)	0.268	0.569	**0.709**	0.465	0.923	14.79	33.26	**18.46****
Other religions (%)	0.773	**0.225**	0.309	0.768	0.599	40.26	42.38	2.12
Political and legal								
Transition economy	0.770	1.000	**0.505**	**0.484**	0.056	0.32	0.29	0.03
Democracy (Polity)	0.476	0.775	0.487	**0.119**	**1.087**	4.41	-0.10	**-4.51****
Democratic participation	0.376	0.635	0.474	0.643	**1.476**	35.11	32.89	-2.21
Democratic competition	**0.299**	0.720	0.826	0.224	**1.350**	44.10	31.91	**-12.19****
Democ. index (Van Hanen)	0.301	0.661	**0.558**	0.507	**1.030**	16.31	15.98	-0.33
Control of corruption	0.952	**0.215**	**0.054***	0.672	0.723	2.93	3.56	**0.63***
Trade openness	0.777	0.603	**0.464**	**0.384**	0.518	4.37	4.37	0.00
Globalization	0.654	0.595	**0.097***	0.423	**1.124**	41.42	48.58	**7.15***
Government expenditure	0.793	**0.343**	**0.131**	0.567	0.512	2.81	2.94	0.13
Civil law	0.373	0.581	1.000	0.709	0.046	0.56	0.67	0.11
Law and order	0.904	**0.333**	**0.030****	0.576	0.771	3.57	4.54	**0.97*****
Others								
Uncertainty avoidance	0.335	**0.418**	0.571	0.798	0.970	62.44	69.21	6.77
Long-term orientation	0.822	**0.241**	**0.143**	0.828	0.836	46.12	53.50	7.37
Private credit	0.736	**0.215**	**0.142**	0.803	**1.338**	2.71	3.40	**0.69*****
Gross capital formation	0.497	0.356	0.244	0.670	0.485	19.07	19.39	0.31
National savings	0.702	**0.123**	**0.242**	0.775	0.757	6.56	10.07	3.51
GDP	0.672	**0.396**	**0.159**	0.801	**1.505**	22.53	23.60	**1.07*****
GDP per capita	0.700	**0.465**	**0.028****	0.652	0.957	7.16	8.10	**0.94*****
GDP growth	0.504	0.459	0.520	0.599	**1.057**	3.72	3.53	-0.19
Inflation	0.594	0.787	**0.136**	0.191	0.776	2.41	2.74	0.32
World GDP growth	**0.153**	**0.219**	0.680	0.662	0.808	2.92	2.81	-0.11
Black market premium	**0.055***	0.393	0.585	1.000	**1.110**	2.31	1.91	-0.40
Public initiative	0.542	**0.548**	**0.799**	0.779	0.028	0.57	0.67	0.10
Private initiative	1.000	1.000	0.485	**0.228**	0.088	0.25	0.11	**-0.14***
Both public and private	1.000	1.000	0.371	0.490	0.020	0.14	0.11	-0.03

Table 7. Random forest confusion matrix and country classification

This table presents the results of the classification random forest analysis used to predict whether a stock market belongs to the cluster of "least successful" or "most successful" markets, based on the set of key variables presented in Table A4 (measured in the first five years after the stock market establishment). The analysis is based on imputed data, and 500 trees with 5 variables tried at each split. Panel A presents the random forest confusion matrix. This matrix depicts, for each cluster of "least successful" and "most successful" countries, the number of countries predicted as "least successful" and "most successful", as well as the classification error (the percentage of incorrect predictions). Panel B presents, for each country included in the analysis, whether it belongs to the cluster of "least successful" (L) of "most successful" (M) markets, and the corresponding random forest prediction. Correct predictions are marked in bold.

Panel A. Random forest confusion matrix
Data imputed

Type of random forest: Classification (most/ least successful)
Number of trees: 500
No. of variables tried at each split: 5
OOB estimate of error rate: 27.27%

	Predicted		
	Least successful	**Most successful**	**Classification error**
Least successful	22	6	0.21
Most successful	9	18	0.33

Panel B. Country classification: Cluster analysis vs Random forest prediction

Country	**Cluster**	**Prediction**
Algeria	L	M
Armenia	**L**	**L**
Bahrain	**M**	**M**
Barbados	M	L
Bolivia	**L**	**L**
Botswana	**L**	**L**
Bulgaria	**M**	**M**
China	**M**	**M**
Costa Rica	**L**	**L**
Cote d'Ivoire	**L**	**L**
Croatia	**M**	**M**
Cyprus	L	M
Czech Republic	M	L
Estonia	**L**	**L**
Fiji	L	M
Ghana	**L**	**L**
Guyana	**L**	**L**
Hungary	**M**	**M**
Iceland	M	L
Jamaica	L	-
Jordan	**M**	**M**
Kazakhstan	**L**	**L**
Kuwait	**M**	**M**
Kyrgyz Republic	**L**	**L**
Latvia	**L**	**L**
Lithuania	**L**	**L**
Malawi	**L**	**L**
Malaysia	M	-
Malta	L	M

Country	**Cluster**	**Prediction**
Mauritius	M	L
Mongolia	M	L
Montenegro	M	L
Mozambique	**L**	**L**
Namibia	**L**	**L**
Nepal	**L**	**L**
New Zealand	M	L
Nigeria	L	-
Oman	**M**	**M**
Panama	**L**	**L**
Paraguay	**L**	**L**
Poland	**M**	**M**
Qatar	**M**	**M**
Romania	**M**	**M**
Russian Federation	**M**	**M**
Saudi Arabia	**M**	**M**
Serbia	M	L
Singapore	**M**	**M**
Slovak Republic	L	M
Slovenia	**M**	**M**
Sudan	**L**	**L**
Thailand	**M**	**M**
Trinidad and Tobago	**L**	**L**
Uganda	**L**	**L**
Ukraine	L	M
United Arab Emirates	**M**	**M**
Vietnam	**M**	**M**
West Bank and Gaza	M	L
Zambia	**L**	**L**

Table 8. Random forest confusion matrix and country classification based on three variables with highest importance

This table presents the results of the classification random forest analysis used to predict whether a stock market belongs to the cluster of "least successful" or "most successful" markets, based on the 3 variables with the highest importance in the analysis presented in Table 7 (and depicted in Figure 9). The analysis is based on imputed data, and 500 trees with 1 variable tried at each split. Panel A presents the random forest confusion matrix. This matrix depicts, for each cluster of "least successful" and "most successful" countries, the number of countries predicted as "least successful" and "most successful", as well as the classification error (the percentage of incorrect predictions). Panel B presents, for each country included in the analysis, whether it belongs to the cluster of "least successful" (L) of "most successful" (M) markets, and the corresponding random forest prediction.

Panel A. Random forest prediction top 3 variable importance

Data imputed
Type of random forest: Classification (most/ least successful)
Number of trees: 500
No. of variables tried at each split: 1
OOB estimate of error rate: 27.27%

	Predicted		
	Least successful	**Most successful**	**Classification error**
Least successful	19	9	0.32
Most successful	6	21	0.22

Panel B. Country classification: Cluster analysis vs Random forest prediction

Country	**Cluster**	**Prediction**
Algeria	L	M
Armenia	**L**	**L**
Bahrain	**M**	**M**
Barbados	**M**	**M**
Bolivia	**L**	**L**
Botswana	L	M
Bulgaria	**M**	**M**
China	**M**	**M**
Costa Rica	**L**	**L**
Cote d'Ivoire	L	M
Croatia	**M**	**M**
Cyprus	L	M
Czech Republic	M	L
Estonia	L	M
Fiji	**L**	**L**
Ghana	**L**	**L**
Guyana	**L**	**L**
Hungary	**M**	**M**
Iceland	**M**	**M**
Jamaica	L	-
Jordan	**M**	**M**
Kazakhstan	L	M
Kuwait	**M**	**M**
Kyrgyz Republic	**L**	**L**
Latvia	**L**	**L**
Lithuania	**L**	**L**
Malawi	**L**	**L**
Malaysia	M	-
Malta	L	M

Country	**Cluster**	**Prediction**
Mauritius	**M**	**M**
Mongolia	M	L
Montenegro	M	L
Mozambique	**L**	**L**
Namibia	**L**	**L**
Nepal	**L**	**L**
New Zealand	**M**	**M**
Nigeria	L	-
Oman	**M**	**M**
Panama	**L**	**L**
Paraguay	**L**	**L**
Poland	**M**	**M**
Qatar	**M**	**M**
Romania	**M**	**M**
Russian Federation	**M**	**M**
Saudi Arabia	**M**	**M**
Serbia	M	L
Singapore	M	L
Slovak Republic	L	M
Slovenia	**M**	**M**
Sudan	**L**	**L**
Thailand	**M**	**M**
Trinidad and Tobago	**L**	**L**
Uganda	**L**	**L**
Ukraine	L	M
United Arab Emirates	**M**	**M**
Vietnam	**M**	**M**
West Bank and Gaza	M	L
Zambia	**L**	**L**

Figure 2. Countries in which a first stock market was (re)established since 1975

This world map depicts stock markets (re)established before and after 1975. Countries depicted in white have not established a stock market as of 2022.

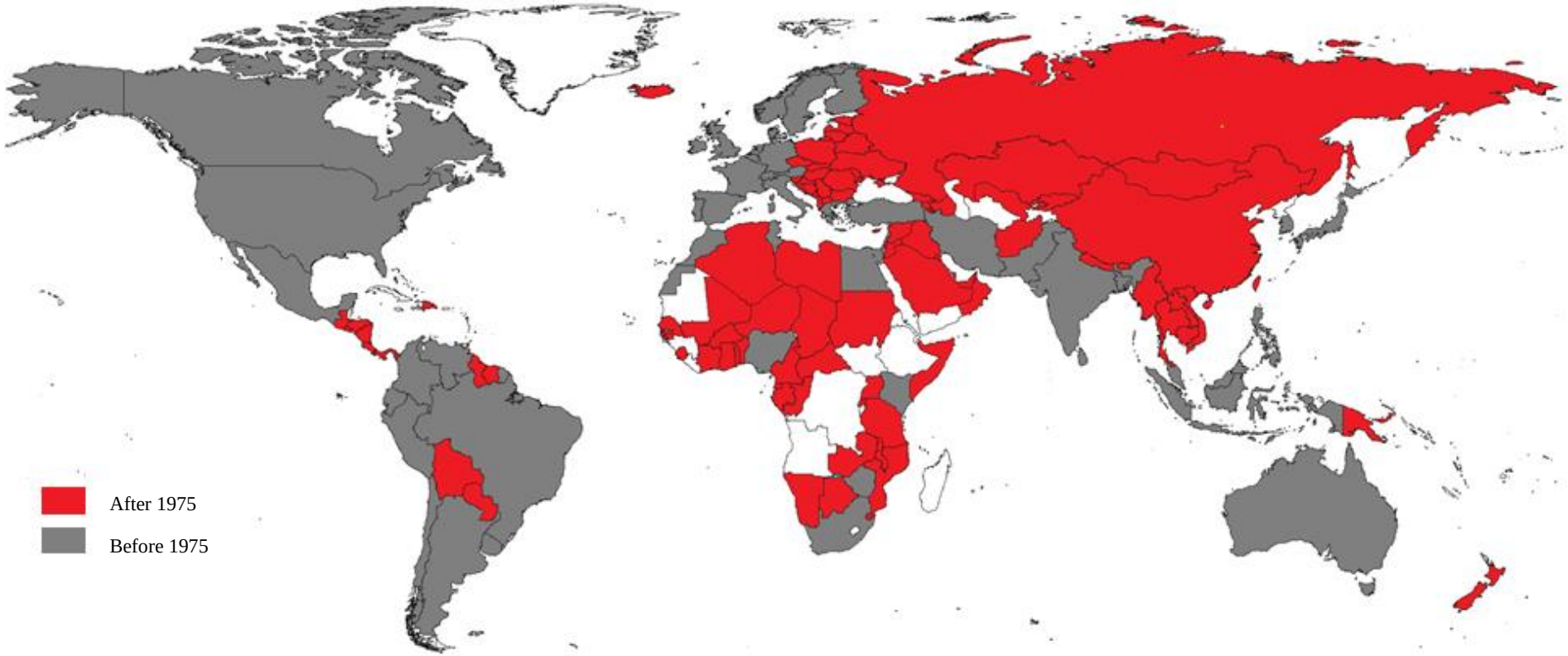
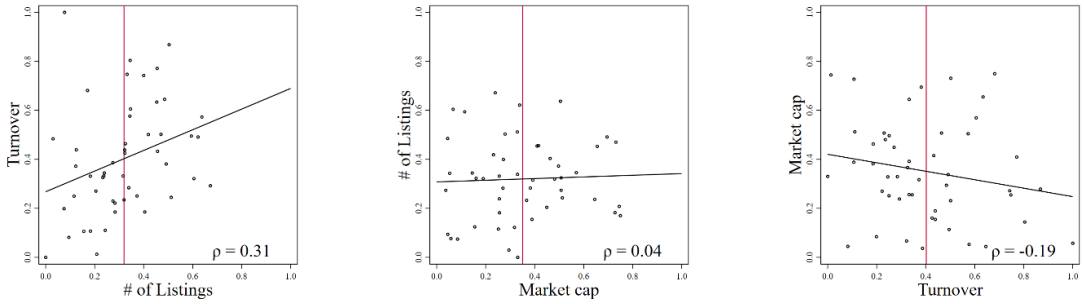


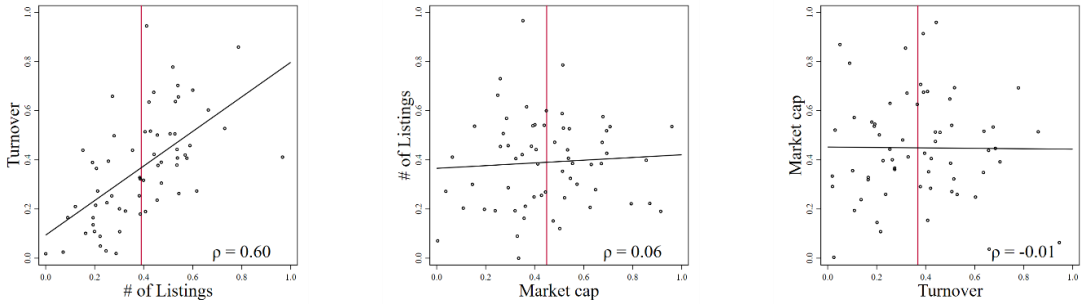
Figure 3. Scatter plots of nascent market success measures (5-year intervals after establishment)

This figure presents pairwise scatter plots of the three measures of nascent market success (number of listings, market cap to GDP, and turnover) in eight 5-year intervals after establishment in Panels A-G. Panel A presents the scatter plots of the success measures averaged 1-5 years after establishment (expressed in logs). Panels B-H present scatter plots for each of the subsequent seven 5-year intervals. Each scatter plot represents the relation between two success measures. Each point in the plot represents a different market. The average of the success measure on the x-axis is indicated by a vertical line. The scatter plots also show OLS regression lines. The correlation ρ between the measures is presented in the bottom right corner of the plot. Each panel is based on all markets for which data are available in that 5-year interval.

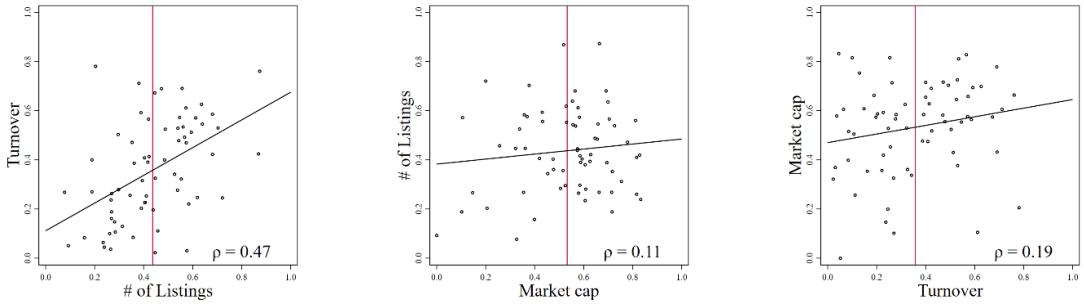
Panel A: 1-5 years after establishment (49 markets)



Panel B: 6-10 years after establishment (62 markets)



Panel C: 11-15 years after establishment (66 markets)



Panel D: 16-20 years after establishment (60 markets)

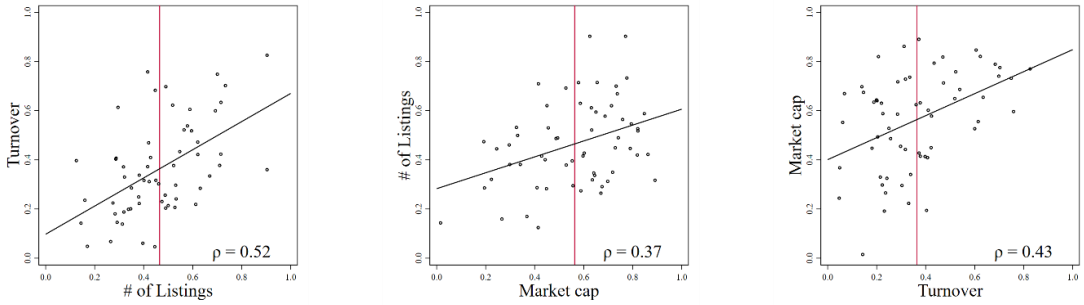
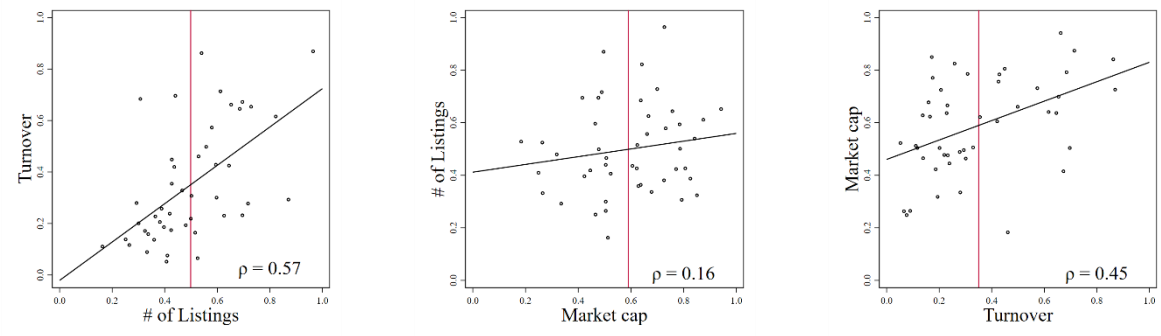
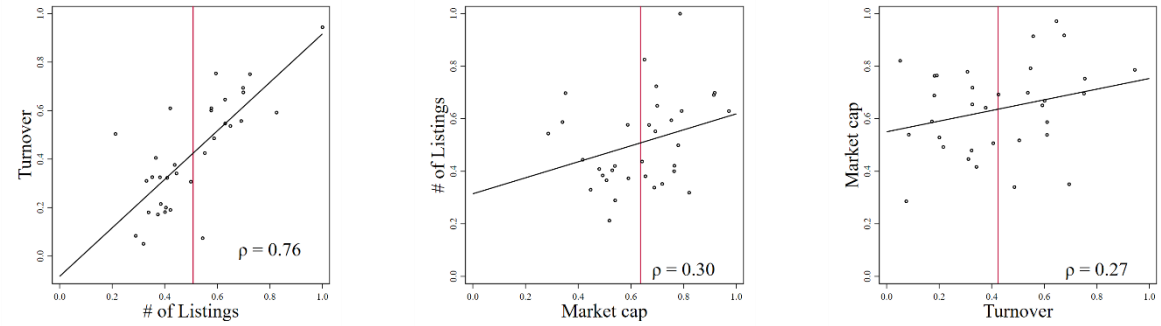


Figure 3 - Continued

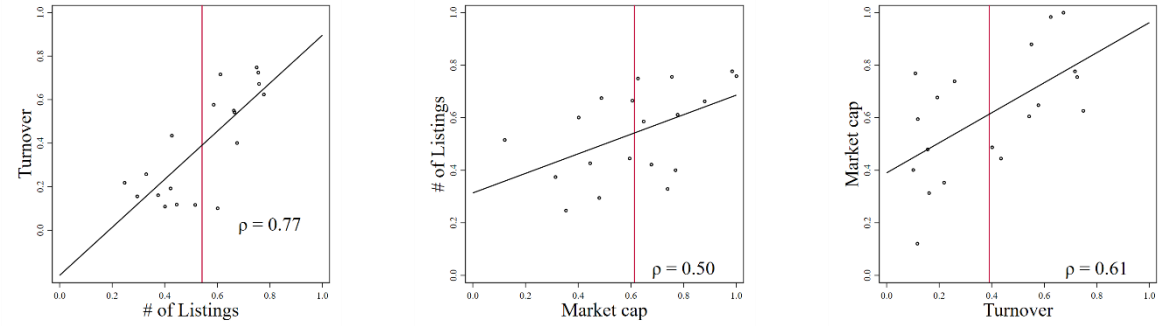
Panel E: 21-25 years after establishment (46 markets)



Panel F: 26-30 years after establishment (33 markets)



Panel G: 31-35 years after establishment (19 markets)



Panel H: 36-40 years after establishment (19 markets)

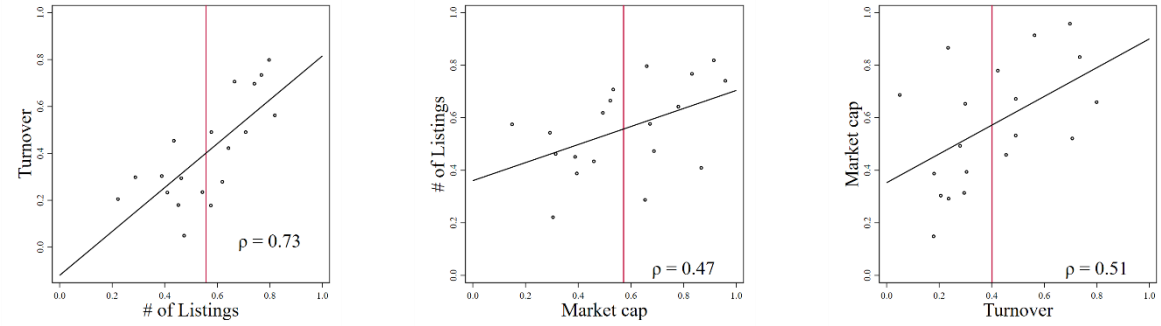
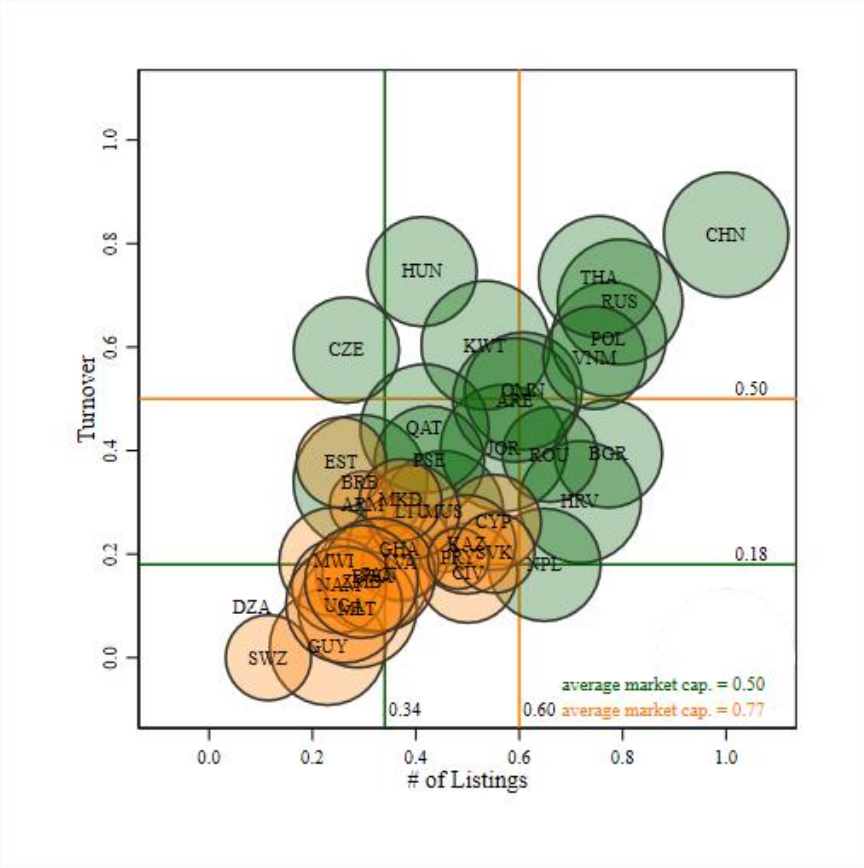


Figure 4. Cluster analysis of nascent market success (16-20 years after establishment)

This figure shows the cluster analysis results based on the three measures of nascent market success (number of listings, market cap to GDP, and turnover) over the period 16-20 years after establishment, yielding a cluster of "least successful" markets (orange) and a cluster of "most successful" markets (green) after 16-20 years. The sample includes 40 markets. Success measures are expressed in logs and then standardized to the interval [0,1] across the whole period 1-20 years after establishment to facilitate comparison across measures and time periods. The plot in Panel A presents depicts the position of each market along the three dimensions of success after 16-20 years: the x-axis represents number of listings, the y-axis represents turnover, and the diameter of the circle represents market cap. The horizontal lines indicate the average turnover of each cluster, the vertical lines represent the average number of listings of each cluster, and the average market capitalization of each cluster is indicated in the bottom right corner. Panel B shows histograms of the success measures for each cluster after 16-20 years. We refer to Table A1 of the Appendix for variable definitions and data sources.

Panel A: Plot of three success measures of least/most successful clusters after 16-20 years



Panel B: Histograms of three success measures of least/most successful clusters after 16-20 years

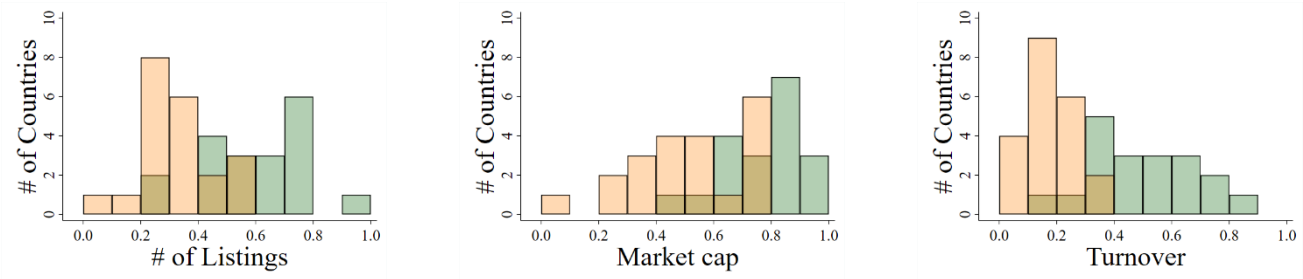
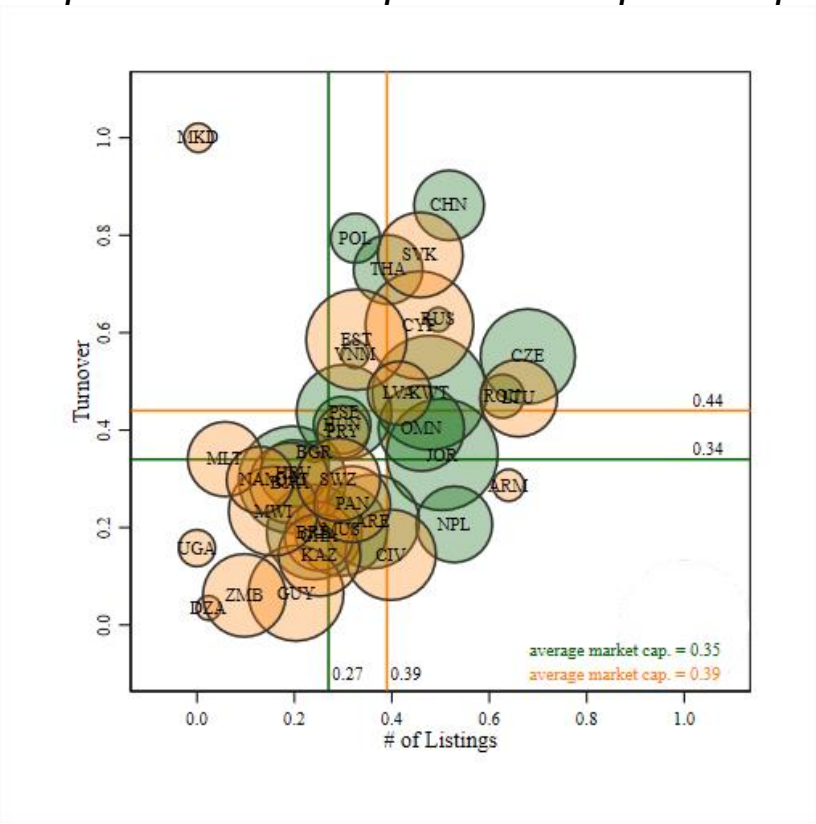


Figure 5. Success measures of least/most successful clusters (1-5 years after establishment)

This figure shows the three measures of nascent market success (number of listings, market cap to GDP, and turnover) over the period 1-5 years after establishment for the clusters of least (orange) and most successful (green) nascent markets from Panel A of Figure 4. The clusters are formed based on the values of the three success measures over the period 16-20 years after establishment. Success measures are expressed in logs and then standardized to the interval [0,1] across the whole period 1-20 years after establishment to facilitate comparison across measures and time periods. The plot in Panel A presents depicts the position of each market along the three dimensions of success after 1-5 years: the x-axis represents number of listings, the y-axis represents turnover, and the diameter of the circle represents market cap. The horizontal lines indicate the average turnover of each cluster, the vertical lines represent the average number of listings of each cluster, and the average market capitalization of each cluster is indicated in the bottom right corner. Panel B shows histograms of the success measures for each cluster after 1-5 years. We refer to Table A1 of the Appendix for variable definitions and data sources.

Panel A: Plot of three success measures of least/most successful clusters after 1-5 years



Panel B: Histograms of three success measures of least/most successful clusters after 1-5 years

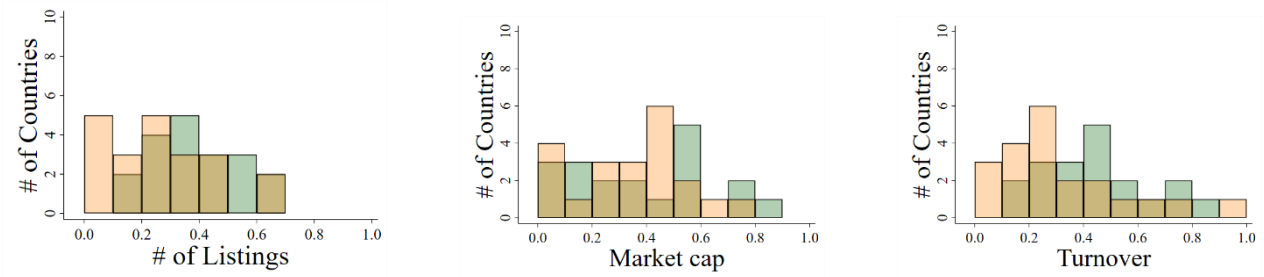
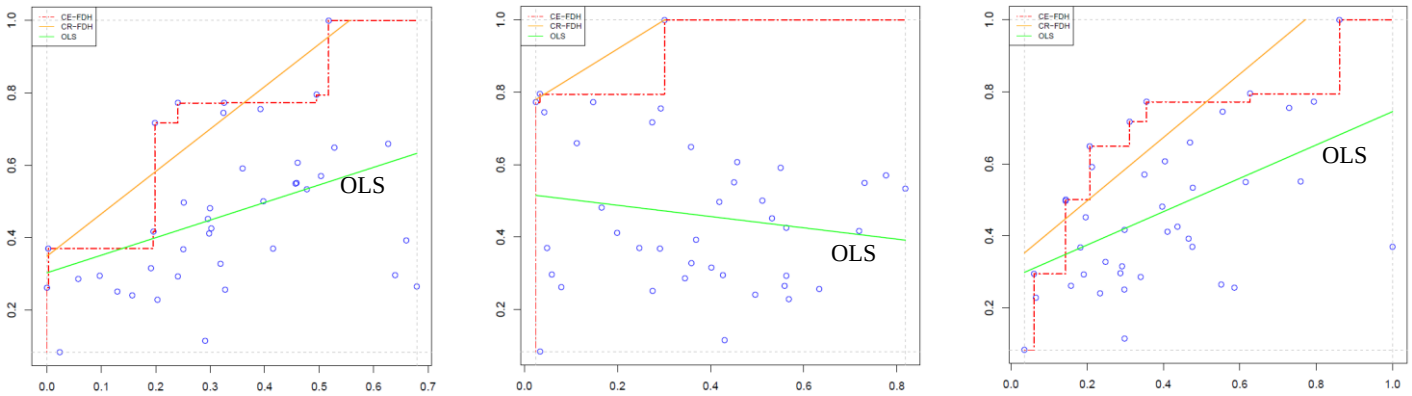


Figure 6. Necessary condition analysis of number of listings (16-20 years) as dependent variable

This figure presents the results of necessary condition analysis (NCA) for the number of listings. Panel A shows scatter plots of the number of listings over the period 16-20 after establishment as dependent variable (y-axis) and three different independent variables (x-axis, from left to right: number of listings over 1-5 years after establishment, market cap over 1-5 years, and turnover over 1-5 years) for the 40 markets included in the analysis. Success measures are expressed in logs. The dotted lines at the outer border of each plot indicate the "scope" of the analysis (defined by the minimum and maximum values of the dependent and independent variables). The dash-dot line (step function) represents the "ceiling line" based on the "ceiling envelopment with free disposal hull" (CE-FDH) method. The diagonal line in the upper left corner represents the ceiling line based on the "ceiling regression with free disposal hull" (CR-FDH) method. The plots also show OLS regression lines. Panel B presents the "bottleneck table" of the necessary conditions for attaining a high number of listings. The first column represents the different percentages of the range of the number of listings (16-20 years). Each of the other columns represents the percentage of the range of values of each independent variable that is necessary to attain the corresponding percentage of the range of the number of listings ("NN" stands for "Not necessary"). The bottom row shows effect sizes (based on the ceiling line using the CE-FDH and CR-FDH methods; ***, **, and * indicate significance at the 1%, 5%, and 10% level (Dul, 2016)). We refer to Section 3.2.2 for a discussion of NCA and to Table A.1 of the Appendix for variable definitions and data sources.

Panel A. Scatter plots of number of listings (16-20 years) vs. initial success measures (1-5 years)



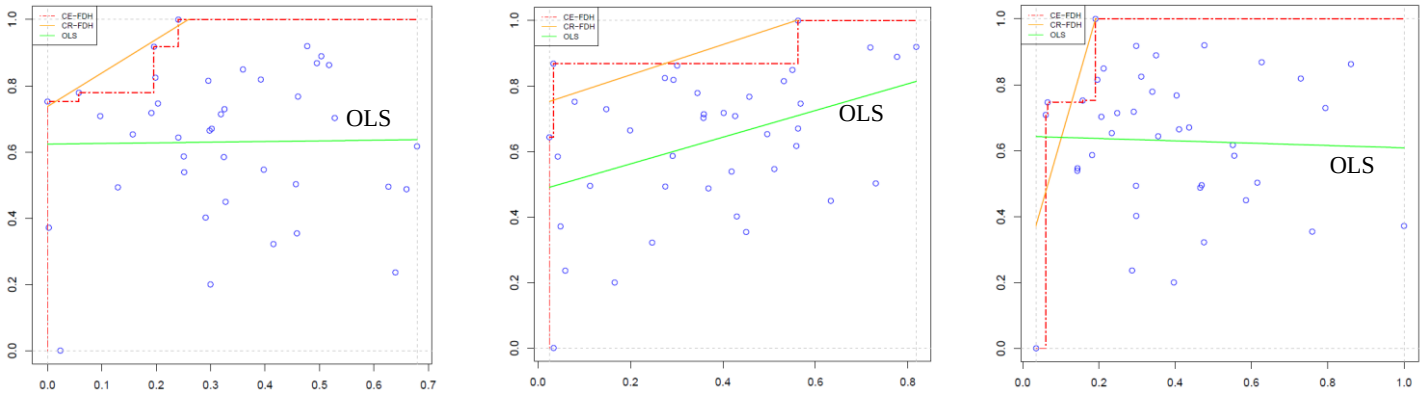
Panel B. Bottleneck table of necessary conditions for number of listings (16-20 years)

# of Listings (16-20y)	# of Listings (1-5y)	Market cap (1-5y)	Turnover (1-5y)
0	NN	NN	NN
10	NN	NN	2.7
20	0.4	NN	2.7
30	0.4	NN	11.2
40	29.2	NN	11.2
50	29.2	NN	17.8
60	29.2	NN	17.8
70	35.4	NN	33.2
80	76.2	34.9	85.6
90	76.2	34.9	85.6
100	76.2	34.9	85.6
Effect size:			
CE-FDH	0.32***	0.06	0.31***
(*p*-value)	(0.010)	(0.713)	(0.002)
CR-FDH	0.29***	0.03	0.27***
(*p*-value)	(0.008)	(0.770)	(0.005)

Figure 7. Necessary condition analysis of market cap (16-20 years) as dependent variable

This figure presents the results of necessary condition analysis (NCA) for market cap to GDP. Panel A shows scatter plots of market cap over the period 16-20 after establishment as dependent variable (y-axis) and three different independent variables (x-axis, from left to right: number of listings over 1-5 years after establishment, market cap over 1-5 years, and turnover over 1-5 years) for the 40 markets included in the analysis. Success measures are expressed in logs. The dotted lines at the outer border of each plot indicate the "scope" of the analysis (defined by the minimum and maximum values of the dependent and independent variables). The dash-dot line (step function) represents the "ceiling line" based on the "ceiling envelopment with free disposal hull" (CE-FDH) method. The diagonal line in the upper left corner represents the ceiling line based on the "ceiling regression with free disposal hull" (CR-FDH) method. The plots also show OLS regression lines. Panel B presents the "bottleneck table" of the necessary conditions for attaining a large market cap. The first column represents the different percentages of the range of market cap (16-20 years). Each of the other columns represents the percentage of the range of values of each independent variable that is necessary to attain the corresponding percentage of the range of market cap ("NN" stands for "Not necessary"). The bottom row shows effect sizes (based on the ceiling line using the CE-FDH and CR-FDH methods; ***, **, and * indicate significance at the 1%, 5%, and 10% level (Dul, 2016)). We refer to Section 3.2.2 for a discussion of NCA and to Table A.1 of the Appendix for variable definitions and data sources.

Panel A. Scatter plots of market cap (16-20 years) vs. initial success measures (1-5 years)



# of Listings (1-5y) | Market cap (1-5y) | Turnover (1-5y)

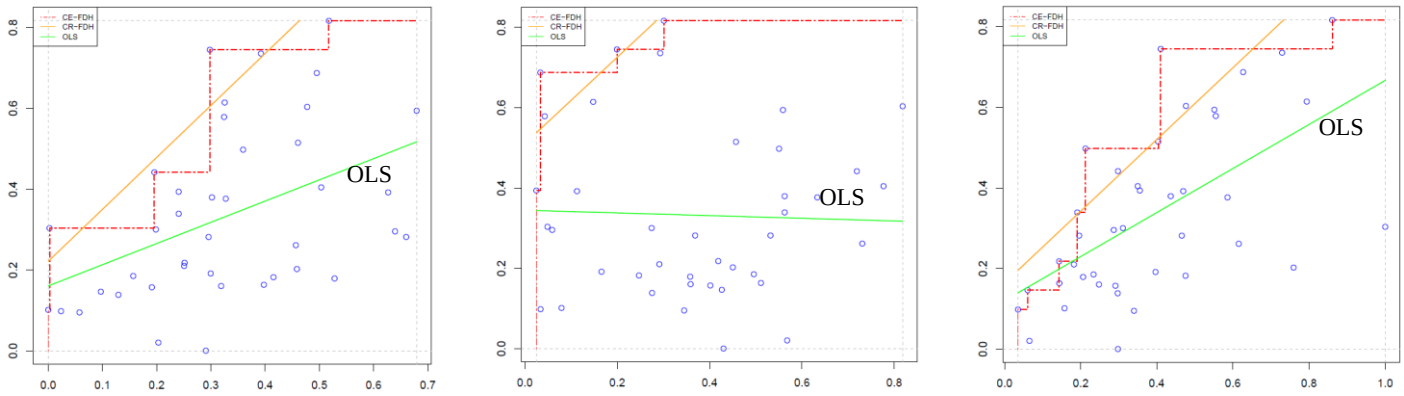
Panel B. Bottleneck table of necessary conditions for market capitalization (16-20 years)

Market cap (16-20y)	# of Listings (1-5y)	Market cap (1-5y)	Turnover (1-5y)
0	NN	NN	NN
10	NN	NN	2.7
20	NN	NN	2.7
30	NN	NN	2.7
40	NN	NN	2.7
50	NN	NN	2.7
60	NN	NN	2.7
70	NN	1.1	2.7
80	28.8	1.1	16.1
90	28.8	67.8	16.1
100	35.4	67.8	16.1
Effect size:			
CE-FDH	0.07	0.09*	0.06
(*p*-value)	(0.479)	(0.095)	(0.392)
CR-FDH	0.05	0.08	0.05
(*p*-value)	(0.614)	(0.190)	(0.420)

Figure 8. Necessary condition analysis of turnover (16-20 years) as dependent variable

This figure presents the results of necessary condition analysis (NCA) for turnover. Panel A shows scatter plots of turnover over the period 16-20 after establishment as dependent variable (y-axis) and three different independent variables (x-axis, from left to right: number of listings over 1-5 years after establishment, market cap over 1-5 years, and turnover over 1-5 years) for the 40 markets included in the analysis. Success measures are expressed in logs. The dotted lines at the outer border of each plot indicate the "scope" of the analysis (defined by the minimum and maximum values of the dependent and independent variables). The dash-dot line (step function) represents the "ceiling line" based on the "ceiling envelopment with free disposal hull" (CE-FDH) method. The diagonal line in the upper left corner represents the ceiling line based on the "ceiling regression with free disposal hull" (CR-FDH) method. The plots also show OLS regression lines. Panel B presents the "bottleneck table" of the necessary conditions for attaining a high turnover. The first column represents the different percentages of the range of turnover (16-20 years). Each of the other columns represents the percentage of the range of values of each independent variable that is necessary to attain the corresponding percentage of the range of turnover ("NN" stands for "Not necessary"). The bottom row shows effect sizes (based on the ceiling line using the CE-FDH and CR-FDH methods; ***, **, and * indicate significance at the 1%, 5%, and 10% level (Dul, 2016)). We refer to Section 3.2.2 for a discussion of NCA and to Table A.1 of the Appendix for variable definitions and data sources.

Panel A. Scatter plots of turnover (16-20 years) vs. initial success measures (1-5y)



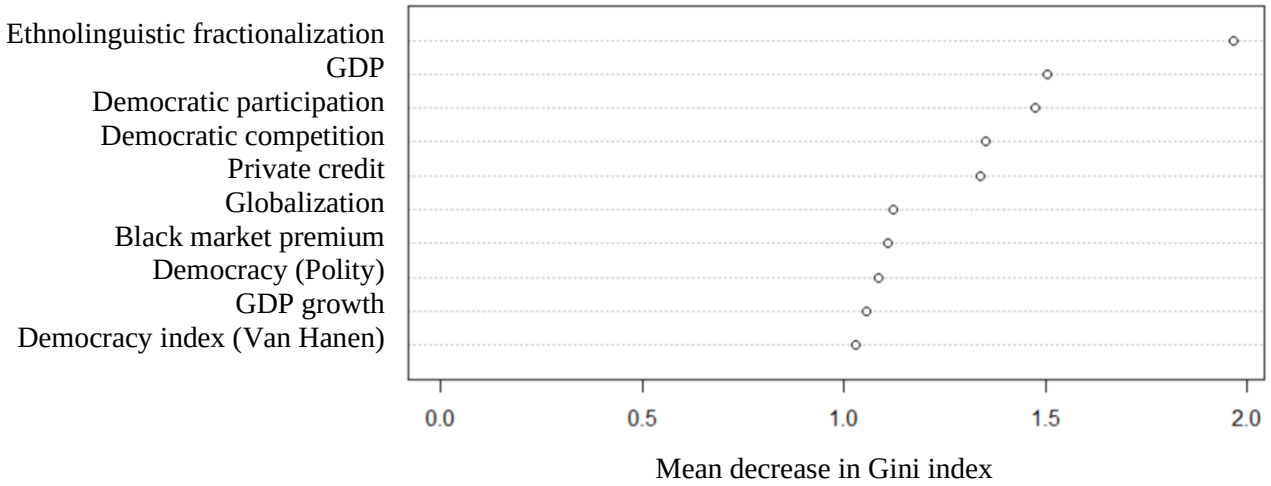
Panel B. Bottleneck table of necessary conditions for turnover (16-20 years)

Turnover (16-20y)	# of Listings (1-5y)	Market cap (1-5y)	Turnover (1-5y)
0	NN	NN	NN
10	NN	NN	NN
20	0.4	NN	11.2
30	0.4	NN	16.1
40	28.8	NN	16.1
50	28.8	1.1	18.4
60	43.9	1.1	18.4
70	43.9	1.1	38.9
80	43.9	1.1	38.9
90	43.9	22.0	38.9
100	76.2	34.9	85.6
Effect size:			
CE-FDH	0.28***	0.05	0.26***
(*p*-value)	(0.003)	(0.686)	(0.000)
CR-FDH	0.25***	0.06	0.28***
(*p*-value)	(0.009)	(0.661)	(0.001)

Figure 9. Random forest variable importance

This figure presents the results of the classification random forest analysis used to predict long-term nascent market success. Panel A depicts the 10 most important predictive variables and the total decrease in node impurities from splitting on each variable, averaged over all trees (node impurity is measured by the Gini index). Panel B presents a scatter plot showing how the 3 variables with the highest importance vary accross the countries in our sample. The vertical axis depicts ethnolinguistic fractionalization, the horizontal axis depicts democratic participation, and the size of the circles is proportional to the country's GDP. Orange circles represent countries in the cluster of "least successful" markets, green circles represent countries in the cluster of "most successful" markets.

Panel A. Random forest variable importance (Top 10)



Panel B. Top 3 variable importance

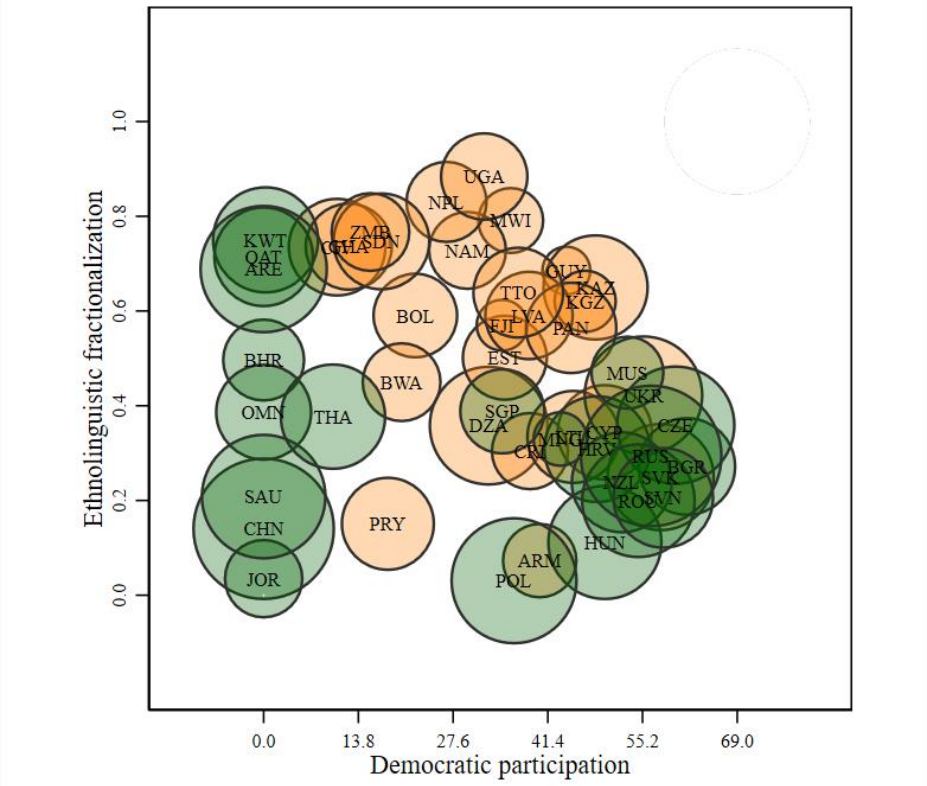
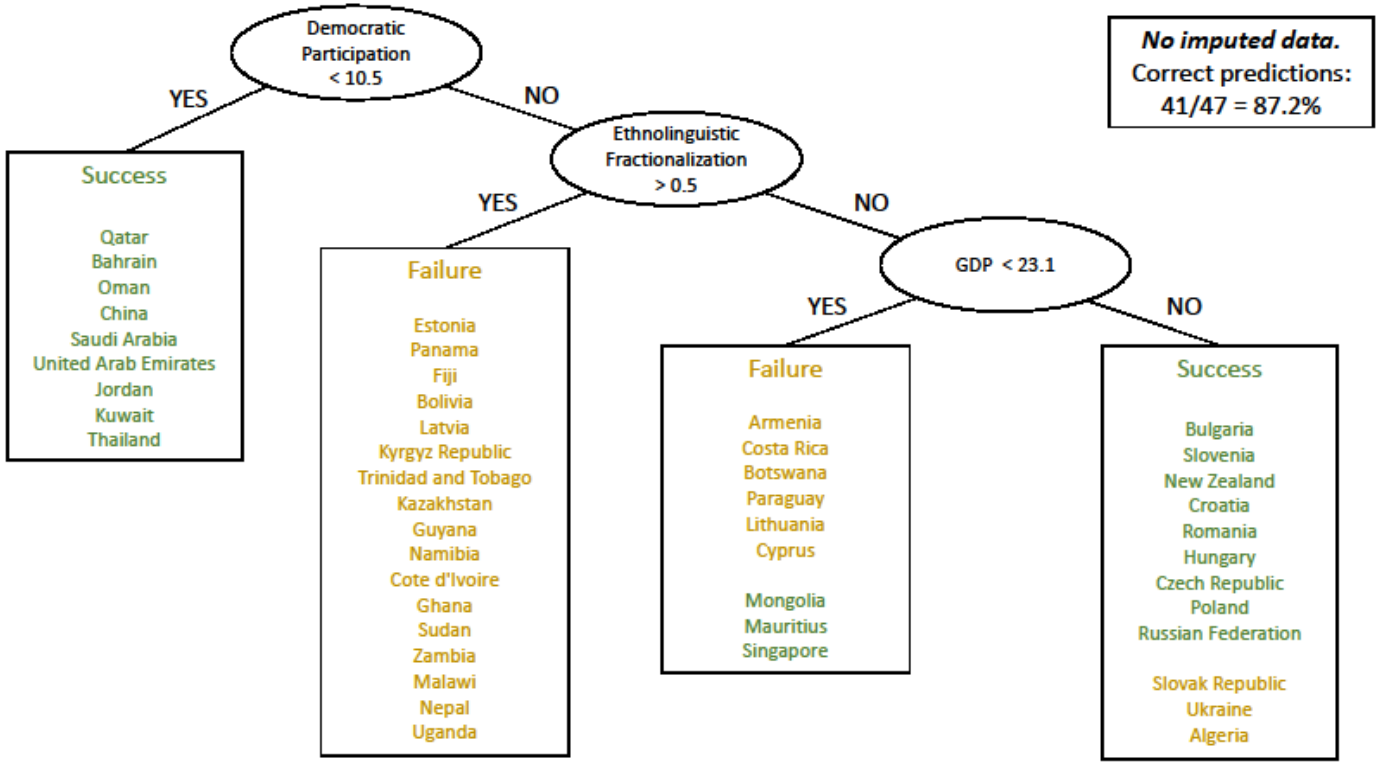


Figure 10. Decision tree to predict nascent market success based on 3 variables with highest importance in Random Forest

In this figure, we manually construct a simplified decision tree to predict nascent market success based on the 3 variables with the highest importance in the random forest analysis presented in Table 7 (Ethnolinguistic fractionalization, GDP, and democratic participation). The decision tree is based on the observed values of each variable in the first five years after the stock market is established. The colors indicate whether the country belongs to the cluster of "least successful" (orange) or "most successful" (green) countries.



APPENDIX

Table A1. Variable definitions and data sources

This table presents the variable definition and data sources for the success measures (Panel A), geo-demographic indicators (Panel B), social indicators (Panel C), political and legal indicators (Panel D), , financial indicators (Panel E), economic indicators (Panel F), and stock market design indicators (Panel G).

Panel A. Success measures

Variable	Source	Definition
Number of listings	World Development Indicators, stock exchange websites, stock exchange association websites	Number of listed domestic companies. Listed domestic companies are the domestically incorporated companies listed on the country's stock exchanges at the end of the year. This indicator does not include investment companies, mutual funds, or other collective investment vehicles.
Market capitalization	World Development Indicators, stock exchange websites, stock exchange association websites	Market capitalization (% GDP). Market capitalization (also known as market value) is the share price times the number of shares outstanding. Listed domestic companies are the domestically incorporated companies listed on the country's stock exchanges at the end of the year. Listed companies does not include investment companies, mutual funds, or other collective investment vehicles.
Turnover	World Development Indicators, stock exchange websites, stock exchange association websites	Stocks traded, turnover ratio (%). Turnover ratio is the total value of shares traded during the period divided by the average market capitalization for the period. Average market capitalization is calculated as the average of the end-of-period values for the current period and the previous period.

Table A1 – continued

Panel B: Geo-demographic indicators

Variable	Source	Definition
Region	World Development Indicators	Region dummies (Middle East & North Africa, Sub-Saharan Africa, Latin America and Caribbean, Asia, Europe & Central Asia)
Latitude	World Bank Social Indicators and Fixed Factors	Geographic coordinate in degrees.
Landlocked	World Bank Social Indicators and Fixed Factors	Dummy variable. Takes the value "1" if the country is landlocked, and "0"otherwise.
Natural resources	World Development Indicators	Total natural resources rents (% of GDP). Total natural resources rents are the sum of oil rents, natural gas rents, coal rents (hard and soft), mineral rents, and forest rents.
Exporters of fuels	World Bank Social Indicators and Fixed Factors	Exporters of fuels (mainly oil). Dummy variable. Takes the value "1"for exporters of fuel, and "0"otherwise. Major export category: Major exports are those that account for 50 percent or more of total exports of goods and services from one category, in the period 1988-92. The categories are: nonfuel primary (SITC 0,1,2,4, plus 68), fuels (SITC 3), manufactures (SITC 5 to 9, less 68), and services (factor and nonfactor service receipts plus workers' remittances). If no single category accounts for 50 percent or more of total exports, the economy is classified as diversified.
Population	World Development Indicators	Total population. Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship--except for refugees not permanently settled in the country of asylum, who are generally considered part of the population of their country of origin. The values shown are midyear estimates.
Population density	World Development Indicators	Population density is midyear population divided by land area in square kilometers. Population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship - except for refugees not permanently settled in the country of asylum, who are generally considered part of the population of their country of origin. Land area is a country's total area, excluding area under inland water bodies, national claims to continental shelf, and exclusive economic zones. In most cases the definition of inland water bodies includes major rivers and lakes.
Ethno-linguistic fractionalization	Historical Index of Ethnic Fractionalization	Historical Index of Ethnic Fractionalization (HIEF) dataset contains an ethnic fractionalization index for 165 countries across all continents. The dataset covers annually the period 1945-2013. The ethnic fractionalization index corresponds to the probability that two randomly drawn individuals within a country are not from the same ethnic group. The applications of HIEF pertain to the pattern of ethnic diversity across countries and over time.
Settler mortality	Acemoğlu et al. (2003)	Log of estimated mortality for European settlers during the early period of European colonization (before 1850). Settler mortality is calculated from the mortality rates of European-born soldiers, sailors and bishops when stationed in colonies.
Catholic ***Protestant*** ***Muslim*** ***Other religion***	La Porta et al. (1999)	Identifies the percentage of the population of each country that belonged to the three most widely spread religions in the world in 1980. For countries of recent formation, the data is available for 1990-1995. The numbers are in percent (scale from 0 to 100). The three religions identified here are (1) Roman Catholic; (2) Protestant; and (3) Muslim. The residual is called "other religions". Sources: Barrett, 1982, Worldmark Encyclopedia of Nations 1995, Statistical Abstract of the World 1995, United Nations, 1995, CIA 1996.

Table A1 – continued

Panel C: Socio-cultural indicators

Variable	Source	Definition
Power distance	Hofstede et al. (2010)	The extent to which the less powerful members of organizations and institutions (like the family) accept and expect that power is distributed unequally. This represents inequality (more versus less), but defined from below, not from above.
Individuality	Hofstede et al. (2010)	The degree to which individuals are integrated into groups. On the individualist side we find societies in which the ties between individuals are loose: everyone is expected to look after her/himself and her/his immediate family. On the collectivist side, we find societies in which people from birth onwards are integrated into strong, cohesive in-groups, often extended families (with uncles, aunts and grandparents) which continue protecting them in exchange for unquestioning loyalty.
Masculinity	Hofstede et al. (2010)	Refers to the distribution of emotional roles between the genders which is another fundamental issue for any society to which a range of solutions are found.
Uncertainty avoidance	Hofstede et al. (2010)	Society's tolerance for uncertainty and ambiguity. It indicates to what extent a culture programs its members to feel either uncomfortable or comfortable in unstructured situations. Unstructured situations are novel, unknown, surprising, different from usual.
Long-term orientation	Hofstede et al. (2010)	Research by Michael Bond and colleagues among students in 23 countries led him in 1991 to adding a fifth dimension called Long- versus Short-Term Orientation. In 2010, research by Michael Minkov allowed to extend the number of country scores for this dimension to 93, using recent World Values Survey data from representative samples of national populations. Long- term oriented societies foster pragmatic virtues oriented towards future rewards, in particular saving, persistence, and adapting to changing circumstances. Short-term oriented societies foster virtues related to the past and present such as national pride, respect for tradition, preservation of "face", and fulfilling social obligations.
Indulgence	Hofstede et al. (2010)	Indulgence versus restraint. Also based on Minkov's World Values Survey data analysis for 93 countries. Indulgence stands for a society that allows relatively free gratification of basic and natural human drives related to enjoying life and having fun. Restraint stands for a society that suppresses gratification of needs and regulates it by means of strict social norms.
Distrust	Aghion et al. (2010)	Share of people who answer "need to very careful in dealing with people" to the question: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?". Average country level of distrust over the four waves of the WVS [World Values Survey].

Table A1 – continued

Panel D: Political and legal indicators

Variable	Source	Definition
Transition economy	World Bank Social Indicators and Fixed Factors	Dummy variable. Takes the value "1" for economies in transition and "0" otherwise.
Democracy	Polity IV	Polity IV – Polity2 Score. Revised Combined Polity Score: The POLITY score is computed by subtracting the AUTOC (autocracy) score from the DEMOC (democracy) score; the resulting unified polity scale ranges from +10 (strongly democratic) to -10 (strongly autocratic).
Democracy index	Van Hanen	The Van Hanen's Index of Democracy is produced by combining data on political participation and electoral competition. Specifically: Index = (% of the adult population who voted in the elections) * (100 - % of the winning party's votes in a national election)/100
Democratic competition	Van Hanen	The smaller parties' share of the votes cast in parliamentary or presidential elections, or both, is used to indicate the degree of competition. It is calculated by subtracting the percentage of votes won by the largest party from 100. If the largest party gets, for example, 40 percent of the votes, the share of the smaller parties is 60 percent. If data on the distribution of votes are not available, the value of this variable is calculated on the basis of the distribution of seats in parliament. The distribution of seats is used also in cases in which it seems to indicate power relations more realistically than the distribution of votes.
Democratic participation	Van Hanen	The percentage of the population which actually voted in the same elections is used to measure the degree of participation (= Participation). This percentage is calculated from the total population, not from the adult or enfranchized population.
Government effectiveness	World Governance Indicators	Perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.
Political stability	World Governance Indicators	Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including politically-motivated violence and terrorism.
Control of corruption	World Governance Indicators	
Control of corruption (ICRG)	International Country Risk Guide	This is an assessment of corruption within the political system.
Freedom of the press	World Press Freedom – Freedom House	The press freedom index that Reporters Without Borders publishes every year measures the level of freedom of information in 180 countries. It reflects the degree of freedom that journalists, news organizations and netizens enjoy in each country, and the efforts made by the authorities to respect and ensure respect for this freedom.
War	UCDP/PRIO Armed Conflict Dataset v.4-2015, 1946 – 2014	Dummy variable that takes the value "1" if an armed conflict where at least one party is the government of a state is registered in that year, and "0" otherwise. Only conflicts with more than 1,000 deaths included.
Trade openness	World Development Indicators	Sum of exports and imports (% of GDP).
Globalization index	KOF Index of Globalization	Measures the three main dimensions of globalization: economic, social and political. Sub-indices: actual economic flows, economic restrictions, information flows, personal contact, cultural proximity.
Government expenditure	World Development Indicators	General government final consumption expenditure (% of GDP). General government final consumption expenditure (formerly general government consumption) includes all government current expenditures for purchases of goods and services (including compensation of employees). It also includes most expenditures on national defense and security, but excludes government military expenditures that are part of government capital formation.
Government debt	World Development Indicators	Central government debt, total (% of GDP). Debt is the entire stock of direct government fixed-term contractual obligations to others outstanding on a particular date. It includes domestic and foreign liabilities such as currency and money deposits, securities other than shares, and loans. It is the gross amount of government liabilities reduced by the amount of equity and financial derivatives held by the government. Because debt is a stock rather than a flow, it is measured as of a given date, usually the last day of the fiscal year.
Cash surplus/ deficit	World Development Indicators	Cash surplus or deficit is revenue (including grants) minus expense, minus net acquisition of nonfinancial assets.

Table A1 – continued

Panel D: Political and legal indicators (continued)		
Variable	**Source**	**Definition**
Legal origin	LaPorta, et al. (1999)	Dummy variable. Takes the value "1"if the country has a "civil law" origin and "0"if it has a "common law" origin.
Regulatory quality	World Governance Indicators	Perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.
Rule of law	World Governance Indicators	Perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.
Voice and accountability	World Governance Indicators	Perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.
Law and order	International Country Risk Guide	"Law and Order" form a single component, but its two elements are assessed separately, with each element being scored from zero to three points. To assess the "Law" element, the strength and impartiality of the legal system are considered, while the "Order" element is an assessment of popular observance of the law. Thus, a country can enjoy a high rating – 3 – in terms of its judicial system, but a low rating – 1 – if it suffers from a very high crime rate if the law is routinely ignored without effective sanction (for example, widespread illegal strikes).

Table A1 – continued

Panel E: Financial indicators

Variable	Source	Definition
Private credit	Financial Development and Structure Dataset	Private credit by deposit money banks and other financial institutions to GDP.
Bank concentration	Financial Development and Structure Dataset	Assets of three largest banks as a share of assets of all commercial banks.
Offshore deposits	Financial Development and Structure Dataset	Offshore bank deposits to domestic bank deposits (%). Offshore bank deposit data from October 2008 version of BIS Statistical Appendix Table 7B: External loans and deposits of reporting banks vis-à-vis the non-bank sector; bank deposits from IFS (IFS lines 24 and 25).
Gross capital formation	World Development Indicators	Gross capital formation (% of GDP). Gross capital formation (formerly gross domestic investment) consists of outlays on additions to the fixed assets of the economy plus net changes in the level of inventories. Fixed assets include land improvements (fences, ditches, drains, and so on); plant, machinery, and equipment purchases; and the construction of roads, railways, and the like, including schools, offices, hospitals, private residential dwellings, and commercial and industrial buildings. Inventories are stocks of goods held by firms to meet temporary or unexpected fluctuations in production or sales, and "work in progress." According to the 1993 SNA, net acquisitions of valuables are also considered capital formation.
Number of MSME's	MSME Country indicators	Number of MSME's per 1000 people. Where possible, MSMEs are defined as follows: micro enterprises: 1-9 employees; small: 10-49 employees; and medium: 50-249 employees. However, in the majority of countries, this definition did not match the local definition, in which cases the local definition took precedence. Only firms with at least one employee are included.
National savings	World Development Indicators	Net national savings (% of GNI). Net national savings are equal to gross national savings less the value of consumption of fixed capital.
Life insurance premium	Financial Development and Structure Dataset	Life insurance premium volume to GDP. Premium data is taken from various issues of Sigma reports (Swiss Re). Data on GDP in US dollars is from the electronic version of the World Development Indicators.

Table A1 – continued

Panel F: Economic indicators

Variable	Source	Definition
GDP	World Development Indicators	GDP, current USD. Sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. Dollar figures for GDP are converted from domestic currencies using single year official exchange rates.
GDP per capita	World Development Indicators	GDP per capita, current USD. GDP per capita is gross domestic product divided by midyear population.
GDP growth	World Development Indicators	GDP growth (annual %). Annual percentage growth rate of GDP at market prices based on constant local currency. Aggregates are based on constant 2005 U.S. dollars. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources.
Real interest rate (st.dev.)	World Development Indicators	Standard deviation of real interest rate, measured over the previous 15 years. Real interest rate is the lending interest rate adjusted for inflation as measured by the GDP deflator.
Inflation (st.dev.)	World Development Indicators	Standard deviation of inflation, measured over the previous 15 years. Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. The Laspereyes formula is generally used.
World GDP growth	World Development Indicators	World total value of GDP growth (annual %).
Shadow economy	Financial Development and Structure Dataset	Includes all market-based legal production of goods and services that are deliberately concealed from public authorities for any of the following reasons: (1) to avoid payment of income, value added or other taxes, (2) to avoid payment of social security contributions, (3) to avoid having to meet certain legal labor market standards, such as minimum wages, maximum working hours, safety standards, etc., and (4) to avoid complying with certain administrative procedures, such as completing statistical questionnaires or other administrative forms.
Black market premium	Global Development Network Growth Database	Black Market Premium (%). Levine and Renelt. World's Currency Yearbook (for 1985, 1990-93); Adrian Wood, Global trends in real exchange rates: 1960-84, WB Discussion paper no. 35. 1988 (filling in missing observations); Global Development Finance & World Development Indicators (for 1996-1997, calculated as (parallel Xrate/official Xrate-1)*100); values for industrial countries are added as 0).
GINI index	World Development Indicators	Gini index measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution. A Lorenz curve plots the cumulative percentages of total income received against the cumulative number of recipients, starting with the poorest individual or household. The Gini index measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line. Thus a Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality.
High technology exports	World Development Indicators	High Technology Exports (% of manufactured exports). High-technology exports are products with high R&D intensity, such as in aerospace, computers, pharmaceuticals, scientific instruments, and electrical machinery.
Scientific and technical articles	World Development Indicators	Scientific and technical journal articles. Scientific and technical journal articles refer to the number of scientific and engineering articles published in the following fields: physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences.
Secondary schooling	World Development Indicators	School enrollment, secondary (% net). Net enrollment rate is the ratio of children of official school age who are enrolled in school to the population of the corresponding official school age. Secondary education completes the provision of basic education that began at the primary level, and aims at laying the foundations for lifelong learning and human development, by offering more subject- or skill-oriented instruction using more specialized teachers.

Table A1 - continued

Panel G: Stock market design indicators		
Variable	**Source**	**Definition**
Government initiative ***Private initiative*** ***Both government and private initiative***	Minier (2009) + own collection	Initiative of opening the exchange. Dummy variables that indicate whether the exchange was opened by government initiative, private initiative or both.
Insider trading laws	Bhattacharya et al. (2002)	Dummy that takes the value "0" before insider trading laws are implement, and "1" thereafter. Dates come from Bhattacharya et al. (2002). Dates came from the answers given by all national regulators and officials of stock markets of the world in March 1999 to the question "When (mm/yy), if at all, were insider trading laws established in your exchange".
Insider trading law enforcement	Bhattacharya et al. (2002)	Dummy that takes the value "0" before insider trading laws are enforced, and "1" thereafter. Dates come from Bhattacharya et al. (2002). Dates came from the answers given by all national regulators and officials of stock markets of the world in March 1999 to the question "If answer to (1) above is YES, when (mm/yy), if at all, was the first prosecution under these laws?".
Equity market liberalization	Bekaert, Harvey, and Lundblad (2005).	Corresponding to a date of formal regulatory change after which foreign investors officially have the opportunity to invest in domestic equity securities. This chronology is based on over 50 different source materials.
Minority investor protection	Doing Business, World Bank	Doing Business measures the protection of minority investors from conflicts of interest through one set of indicators and shareholders' rights in corporate gover-nance through another. The data come from a question-naire administered to corporate and securities lawyers and are based on securities regulations, company laws, civil procedure codes and court rules of evidence. The ranking of economies on the strength of minority investor protec-tions is determined by sorting their scores for protecting minority investors. These scores are the sum of the extent of conflict of interest regulation index and the extent of shareholder governance index.

Table A2. Sample of countries and cluster in each analysis

This table presents an overview of the 85 countries included in the scatter plot analysis presented in Figure 3. For each of the countries, column (1) indicates whether the country belongs to the "least successful" or "most successful" cluster in the Cluster Analysis presented in Figure 4. Column (2) indicates the cluster of the analysis presented in Figure A1 (extended analysis to include countries that have data for the 16-20 year interval). Column (3) indicates the cluster of the analysis presented in Table 7 (cluster predicted by the random forest analysis). Column (4) indicates the cluster of the analysis presented in Figure 10 (Decision tree to predict nascent market success based on variables with highest importance in random forest).

	Figure 4	Figure A1	Table 7	Figure 10		Figure 4	Figure A1	Table 7	Figure 10
Country	(1)	(2)	(3)	(4)	Country	(1)	(2)	(3)	(4)
Algeria	L	L	M	M	Malaysia	-	M	-	L
Armenia	L	L	L	L	Malta	L	L	M	-
Bahrain	-	M	M	M	Mauritius	M	M	L	L
Bangladesh	-	-	-	-	Moldova	-	-	-	-
Barbados	M	M	L	-	Mongolia	-	M	L	L
Bermuda	-	-	-	-	Montenegro	-	M	L	-
Bolivia	-	L	L	L	Mozambique	-	L	L	-
Botswana	L	L	L	L	Namibia	L	L	L	L
Bulgaria	M	M	M	M	Nepal	M	L	L	L
Cameroon	-	-	-	-	New Zealand	-	M	L	M
Cayman Islands	-	-	-	-	Nigeria	-	L	-	-
Channel Islands	-	-	-	-	Oman	M	M	M	M
China	M	M	M	M	Panama	L	L	L	-
Costa Rica	-	L	L	L	Papua New Guinea	-	-	-	-
Cote d'Ivoire	L	L	L	L	Paraguay	L	L	L	L
Croatia	M	M	M	M	Poland	M	M	M	M
Cyprus	L	L	M	L	Qatar	M	M	M	M
Czech Republic	M	M	L	M	Romania	M	M	M	M
Ecuador	-	-	-	-	Russian Federation	M	M	M	M
El Salvador	-	-	-	-	Rwanda	-	-	-	-
Estonia	L	L	L	L	Saudi Arabia	-	M	M	M
Fiji	-	L	M	L	Serbia	-	M	L	-
Fiji	-	-	-	-	Sierra Leone	-	-	-	-
Georgia	-	-	-	-	Singapore	-	M	M	M
Ghana	L	L	L	L	Slovak Republic	L	L	M	M
Guyana	L	L	L	L	Slovenia	-	M	M	M
Honduras	-	-	-	-	St Kitts and Nevis	-	-	-	-
Hungary	M	M	M	M	Sudan	-	L	L	L
Iceland	-	M	L	-	Eswatini	L	-	-	-
Iran	-	-	-	-	Tanzania	-	-	-	-
Israel	-	-	-	-	Thailand	M	M	M	M
Jamaica	-	L	-	-	Trinidad and Tobago	-	L	L	L
Jordan	M	M	M	M	Tunisia	-	-	-	-
Kazakhstan	L	L	L	L	Uganda	L	L	L	L
Kenya	-	-	-	-	Ukraine	-	L	M	M
Korea, Rep.	-	-	-	-	United Arab Emirates	M	M	M	M
Kuwait	M	M	M	M	Uzbekistan	-	-	-	-
Kyrgyz Republic	-	L	L	L	Venezuela	-	-	-	-
Latvia	L	L	L	L	Vietnam	M	M	M	-
Libya	-	-	-	-	West Bank and Gaza	M	M	L	-
Lithuania	L	L	L	L	Zambia	L	L	L	L
North Macedonia	L	-	-	-	Zimbabwe	-	-	-	-
Malawi	L	L	L	-					
					# of countries	40	58	55	47

Table A3. Inclusion criteria for NCA – detailed

This table presents, for each variable identified as a potential determinant of nascent market success, the number of countries with data available, expressed as a percentage of the number of countries with data available on each success measure (listings/ population, market cap and turnover), as well as the corresponding minimum and maximum standardized observation of each success measure (linearly transformed to the [0-1] interval across the whole sample). The criteria for inclusion in the NCA are: i) the number of countries with data available is greater than 50% of the number of countries with data available on each success measure; ii) the minimum standardized observation of each success measure is below 0.2; iii) the maximum standardized observation of each success measure is greater than 0.8. Variables marked in red do not satisfy the criteria for inclusion in the NCA. The statistics are presented in several panels, following the categorization of variables presented in Table A1 of the Appendix.

	listings/pop			market cap			turnover		
	N	min	max	N	min	max	N	min	max
Panel B. Geo-demographic indicators									
landlocked	96.6 %	0.00	0.83	96.5 %	0.00	1.00	96.6 %	0.00	1.00
natural resources	83.1 %	0.00	0.83	82.5 %	0.00	1.00	81.0 %	0.00	1.00
exporter of fuels	96.6 %	0.00	0.83	96.5 %	0.00	1.00	96.6 %	0.00	1.00
population	100.0 %	0.00	1.00	96.5 %	0.00	1.00	96.6 %	0.00	1.00
population density	98.3 %	0.00	1.00	98.2 %	0.00	1.00	98.3 %	0.00	1.00
ethno-ling. fractional.	83.1 %	0.02	1.00	87.7 %	0.00	0.97	86.2 %	0.00	1.00
log settler mortality	33.9 %	0.02	0.69	33.3 %	0.00	1.00	32.8 %	0.08	1.00
catholic (%)	96.6 %	0.00	0.83	96.5 %	0.00	1.00	96.6 %	0.00	1.00
protestant (%)	94.9 %	0.00	0.83	94.7 %	0.00	1.00	94.8 %	0.00	1.00
muslim (%)	96.6 %	0.00	0.83	96.5 %	0.00	1.00	96.6 %	0.00	1.00
other religions (%)	94.9 %	0.00	0.83	94.7 %	0.00	1.00	94.8 %	0.00	1.00
Panel C. Socio-cultural indicators									
uncertainty avoidance	57.6 %	0.00	1.00	59.6 %	0.29	0.92	58.6 %	0.17	1.00
long-term orientation	54.2 %	0.00	1.00	57.9 %	0.00	0.92	56.9 %	0.00	1.00
distrust	25.4 %	0.02	0.76	26.3 %	0.00	0.89	25.9 %	0.17	1.00
Panel D. Political and legal indicators									
transition economy	96.6 %	0.00	0.83	96.5 %	0.00	1.00	96.6 %	0.00	1.00
democracy (Polity)	86.4 %	0.00	0.83	89.5 %	0.00	0.97	87.9 %	0.00	1.00
democratic participation	93.2 %	0.00	0.83	94.7 %	0.00	1.00	93.1 %	0.00	1.00
democratic competition	93.2 %	0.00	0.83	94.7 %	0.00	1.00	93.1 %	0.00	1.00
dem. index (vanHanen)	93.2 %	0.00	0.83	94.7 %	0.00	1.00	93.1 %	0.00	1.00
govern. effectiveness	35.6 %	0.00	0.64	33.3 %	0.00	0.92	34.5 %	0.00	0.73
political stability	33.9 %	0.00	0.64	33.3 %	0.00	0.92	34.5 %	0.00	0.73
control corruption (WGI)	35.6 %	0.00	0.64	33.3 %	0.00	0.92	34.5 %	0.00	0.73
contr. corruption (ICRG)	59.3 %	0.00	0.83	57.9 %	0.00	0.97	56.9 %	0.08	1.00
freedom of the press	50.8 %	0.00	0.64	49.1 %	0.00	0.92	48.3 %	0.00	0.74
trade openness	84.7 %	0.00	0.83	84.2 %	0.00	1.00	82.8 %	0.00	1.00
globalization	74.6 %	0.00	1.00	73.7 %	0.00	1.00	74.1 %	0.02	1.00
government expenditure	83.1 %	0.00	0.83	86.0 %	0.00	1.00	84.5 %	0.00	1.00
government debt	25.4 %	0.08	0.83	22.8 %	0.32	0.97	22.4 %	0.00	0.94
cash surplus/ deficit	23.7 %	0.08	0.83	21.1 %	0.20	0.75	20.7 %	0.00	0.74
civil law	86.4 %	0.00	0.83	86.0 %	0.00	1.00	84.5 %	0.00	1.00
regulatory quality	35.6 %	0.00	0.64	33.3 %	0.00	0.92	34.5 %	0.00	0.73
rule of law	35.6 %	0.00	0.64	33.3 %	0.00	0.92	34.5 %	0.00	0.73
voice and accountability	35.6 %	0.00	0.64	33.3 %	0.00	0.92	34.5 %	0.00	0.73
law and order	59.3 %	0.00	0.83	57.9 %	0.00	0.97	56.9 %	0.08	1.00
Panel E. Financial indicators									
private credit	88.1 %	0.00	0.83	91.2 %	0.00	1.00	89.7 %	0.00	1.00
bank concentration	30.5 %	0.00	0.64	29.8 %	0.00	0.92	31.0 %	0.00	0.73
offshore deposits	33.9 %	0.00	0.64	35.1 %	0.00	0.92	34.5 %	0.00	0.73
gross capital formation	79.7 %	0.00	1.00	80.7 %	0.00	0.97	81.0 %	0.00	1.00
number of MSMEs	15.3 %	0.28	0.69	17.5 %	0.24	0.86	17.2 %	0.00	1.00
national savings	61.0 %	0.08	0.83	61.4 %	0.20	1.00	60.3 %	0.00	1.00
life insurance premium	44.1 %	0.00	0.64	45.6 %	0.00	0.85	44.8 %	0.00	0.92

Table A3 - continued

	listings/pop			market cap			turnover		
	N	min	max	N	min	max	N	min	max
Panel F. Economic indicators									
GDP	91.5 %	0.00	0.83	89.5 %	0.00	1.00	87.9 %	0.00	1.00
GDPpc	91.5 %	0.00	0.83	89.5 %	0.00	1.00	87.9 %	0.00	1.00
GDP growth	88.1 %	0.00	0.83	86.0 %	0.00	1.00	84.5 %	0.00	1.00
real interest rate	45.8 %	0.00	0.76	45.6 %	0.00	1.00	44.8 %	0.06	1.00
inflation	79.7 %	0.02	0.76	82.5 %	0.00	1.00	81.0 %	0.00	1.00
world GDP growth	100.0 %	0.00	1.00	100.0 %	0.00	1.00	100.0 %	0.00	1.00
shadow economy	15.3 %	0.00	0.60	14.0 %	0.00	0.85	13.8 %	0.00	0.73
black market premium	55.9 %	0.00	0.69	57.9 %	0.00	1.00	58.6 %	0.02	1.00
GINI index	44.1 %	0.00	0.76	43.9 %	0.20	0.86	43.1 %	0.00	1.00
scientific articles	13.6 %	0.00	0.60	12.3 %	0.00	0.85	12.1 %	0.00	0.73
secondary schooling	25.4 %	0.00	0.64	26.3 %	0.00	1.00	25.9 %	0.00	0.91
Panel G. Stock market design indicators									
public	100.0 %	0.00	1.00	100.0 %	0.00	1.00	100.0 %	0.00	1.00
private	100.0 %	0.00	1.00	100.0 %	0.00	1.00	100.0 %	0.00	1.00
both	100.0 %	0.00	1.00	100.0 %	0.00	1.00	100.0 %	0.00	1.00

Table A4. Variables that satisfy conditions for inclusion in NCA and random forest

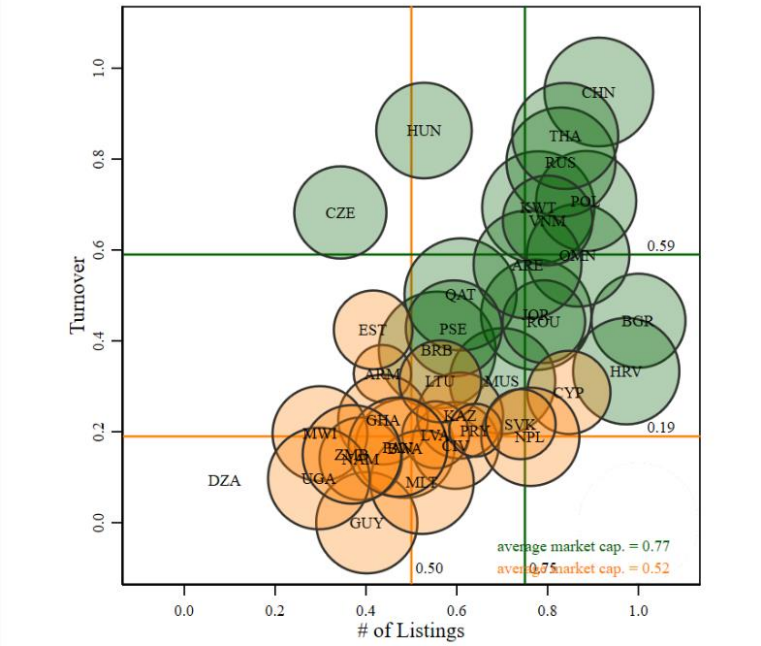
This table presents the variables that satisfy the conditions imposed for inclusion in the necessary condition analysis (NCA): i) the number of countries with data available is greater than 50% of the number of countries with data available on each success measure; ii) the minimum standardized observation of each success measure (linearly transformed to the [0-1] interval across the whole sample) is below 0.2; iii) the maximum standardized observation of each success measure is greater than 0.8. Each column represents a different type of variable, as indicated in the header of the table. We refer to Table A1 of the Appendix for variable definitions and data sources.

Panel B. Geo-demographic indicators	***Panel C. Socio-cultural indicators***	***Panel D. Political and legal indicators***	***Panel E. Financial indicators***	***Panel F. Economic indicators***	***Panel G. Stock market design indicators***
Landlocked	Uncertainty avoidance	Transition economy	Private credit	GDP	Public
Natural resources	Long-term orientation	Democracy (Polity)	Gross capital formation	GDP per capita	Private
Exporter of fuels		Democratic participation	National savings	GDP growth	Both
Population		Democratic competition		Inflation	
Population density		Democratic index (Van Hanen)		World GDP growth	
Ethno-linguistic fractionalization		Control of corruption (ICRG)		Black market premium	
Catholic (%)		Trade openness			
Protestant (%)		Globalization			
Muslim (%)		Government expenditure			
Other religions (%)		Civil law			
		Law and order			

Figure A1. Extended cluster analysis

This figure presents the results of the cluster analysis conducted as in Figure 4 when scaling the number of listings by population (Panel A), and extended to include countries for which data on the success measures in the first 5 years after establishment is not available (Panel B). The x-axis represents number of listings scaled by population, the y-axis represents turnover, and the diameter of the circle represents market cap. The horizontal lines indicate the average turnover of each cluster, the vertical lines represent the average number of listings of each cluster, and the average market capitalization of each cluster is indicated in the bottom-right corner. We refer to Table A1 of the Appendix for variable definitions and data sources.

Panel A. Success measures, countries with data available for 1-5y and 16-20y intervals



Panel B. Success measures, all countries with data available for 16-20y interval