

The political determinants of foreign direct investment: a firm-level analysis

Vincent Arel-Bundock

May 4, 2016

Keywords: Foreign direct investment, international political economy, domestic institutions

Abstract

Many large-N cross-national studies claim to show that political institutions and phenomena determine where foreign direct investment (FDI) flows. In this paper, I argue that these studies tend to overemphasize statistical significance, and often neglect to assess the explanatory or predictive power of their theories. To illustrate the problem, I estimate variations of a statistical model published in an influential article on *Political Risk, Institutions, and FDI*. I find that none of the political variables that the authors consider account for much of the variation in aggregate FDI inflows. To ensure that this underwhelming result is not driven by misspecification or measurement error, I leverage a large firm-level dataset on the investment location decisions of thousands of multinational firms. Using non-parametric machine-learning techniques and out-of-sample tests, I show that gravity variables can help us develop very accurate expectations about firm behavior, but that none of the 31 "political determinants" of FDI that I consider can do much to improve our priors. These findings have important implications, because they suggest that governments retain some room to move in the face of economic globalization.

^ovincent.arel-bundock@umontreal.ca. I thank William R. Clark, Robert J. Franzese Jr., Andrew Kerner, Walter R. Mebane Jr., Philip B.K. Potter, several anonymous reviewers, and the Editors of II.

Political-economists have produced a rich body of work on the political determinants of foreign direct investment (FDI). Much of their efforts are rooted in the “obsolescing bargain” tradition (Vernon 1971), which emphasises the shift in bargaining power that occurs when firms make capital investments abroad. Before settling in a new country, a company can credibly threaten to choose a different location, which increases its power vis-à-vis host governments. After fixed capital investments are made, however, exit threats evaporate, and the host government can seize assets. With that in mind, several authors have argued that international investors pay close attention to political conditions in host countries, because those conditions can affect business costs and the likelihood of expropriation. This, in turn, implies that governments can be punished by international capital markets if they do not adopt institutions that foster “good governance.”

Empirical support for this line of research is plentiful. Indeed, many large-N cross-national studies claim to show that political institutions and phenomena determine where foreign direct investment flows. However, as I explain below, these studies tend to overemphasize statistical significance, and often neglect to assess the fit between theory and data. As a result, we still know very little about the explanatory or predictive power of theories of politics and international investment.

The present paper is designed as a (partial) remedy to this problem. I begin by arguing that, contrary to common wisdom, there are good substantive reasons to expect that macro-political country-year variables will only be weakly related to international investment flows. Whether the political environment is a powerful explanator very much remains an open empirical question.

To answer this question, I consider several variations on an influential statistical model of *Political Risk, Institutions, and FDI*. In their article, Busse and Hefeker (2007) consider a large sample of developing countries and estimate the association between many political regressors¹

¹The variables considered by Busse and Hefeker (2007) are published by the PRS Group.

and an aggregate measure of net FDI inflows. As in the original analysis, I find that favorable political conditions tend to be positively correlated with FDI inflows. But although many of the regression coefficients of interest are statistically significant, I show that none of the variables in question can account for much of the variance in the dependent variable: adding a political regressor to the baseline economic model never increases the adjusted coefficient of determination by more than 0.013.

This result is underwhelming, but it could reflect measurement error or misspecification rather than a true lack of explanatory power. To alleviate these concerns, I introduce a large firm-level dataset on the investment location decisions of thousands of multinational corporations (Bureau van Dijk 2013). I then use machine-learning techniques to evaluate the predictive power of 31 “political determinants” of FDI.

This empirical exercise shows that standard gravity variables can help us develop very accurate expectations about firm behavior, but that none of the political measures I consider can do much to improve our priors. In fact, models that ignore politics altogether fare just as well as those that leverage all 31 variables in out-of-sample tests. Again, country-level measures of political phenomena and institutions are statistically related to investment location decisions, but that relationship does not hold much power to explain or predict real-world outcomes.

This finding has two important sets of implications for the field. First, the fact that macro-level political variables are weak predictors of the location and size of international investment flows suggests that states may retain some “room to move” in the face of economic globalization. Indeed, if markets punish governments for adopting the “wrong” modes of governance, the magnitude of this sanction must be proportional to the strength of association between political risk, institutions, and FDI. And if political factors only play a marginal role in explaining investment decisions, then market-based rationales for policy convergence across countries lose some of their appeal.

Second, this paper stands as a challenge for future researchers to demonstrate that their

theories materially improve prior expectations about investment decisions. Methodologically, political-economists would do well to draw on established best practices which include paying more attention to model fit and out-of-sample tests. Thematically, this paper shows that the field could benefit from shifting the focus away from country-year political indicators and aggregate FDI flow data, and toward micro-level mechanisms and other aspects of investment behavior. Or, flipping the script, many more valuable contributions could be made to the existing literature on the political effects of international investment.

1 POLITICS AND FOREIGN DIRECT INVESTMENT

The relationship between politics and foreign direct investment is the topic of a large and growing body of work. These efforts build on rapid developments in the study of political risk during the 1960-70s in the fields of business management and economics (see Kobrin 1979 for a contemporary survey). Recent contributions by political scientists (see Jensen et al. 2012) tend to emphasize the role of institutions in overcoming the obsolescing bargain problem identified in Vernon (1971).²

1.1 *Political variables should be strong predictors of investment behavior*

Many political factors are considered important in this literature. For example, some analysts explore the relationship between regime type and FDI flows (Jensen 2003; Li and Resnick 2003; O Neal 1994). Others have focused on partisanship (Pinto 2013; Pinto and Pinto 2008), the contractual environment and transaction costs (Henisz 2000; Henisz and Williamson 1999), or corruption (Caprio, Faccio, and McConnell 2013; Fredriksson, List, and Millimet 2003).

By and large, the articles cited above make *unconditional* arguments about the effect of politics on investment decisions, but several authors have also emphasized the *conditional* nature

²The expression “obsolescing bargain” refers to the idea that fixed capital assets make MNCs vulnerable to expropriation by host governments who renege on promises made at the time of initial investment.

of that relationship. Conditionality can operate at the host-country level: in Mosley (2003), for instance, the author shows that international investors do not get much actionable information by observing politics in developed countries, but that they could take cues from political developments in emerging economies when making investment decisions. Conditional arguments can also be deployed at the firm-level as in Wellhausen (2015), which points out that bondholders and direct investors have different preferences over property rights enforcement. Similarly, Kerner and Lawrence (2014) remind us that firms from different sectors, with different levels of capital intensity, may be more or less sensitive to expropriation; they show that measures of fixed capital expenditures are more closely related to political factors than broad measures of FDI flows or stocks.

Taken together, these works suggest that there are good theoretical reasons to expect that political conditions in host countries will affect MNCs' profitability and investment decisions. They also offer impressive empirical support for their claims. Pinto (2013, 124), for example, estimates that the semi-elasticity of FDI with respect to a *left-wing government* dummy variable could reach unity (in OECD countries, 1974-1996). In other words, governance from the left could be associated with a near-doubling of net FDI inflows. Jensen (2003, 62) also makes a strong case for the importance of political explanators when he concludes that a "move from an authoritarian regime to a democratic regime increases FDI inflows by 60 percent."

1.2 Political variables should be weak predictors of investment behavior

Yet, in some respects, such large point estimates should strike readers as surprising. Indeed, an emerging strand of research in economics points out that there are, in fact, very few robust determinants of FDI. These studies use a variety of statistical techniques, including sensitivity analysis (Chakrabarti 2001), extreme bounds analysis (Moosa and Cardak 2006), and Bayesian model averaging (Blonigen and Piger 2014; Eicher, Helfman, and Lenkoski 2012), and come to similar conclusions: traditional gravity variables exert a strong influence on investment be-

havior, but most other variables (infrastructure, business costs, credit markets) are not robustly associated with FDI flows.³ Li (2015, 4) comes to similar conclusions after reviewing empirical work on the relationship between tax incentives and FDI flows, concluding that an “extensive body of literature shows that the effects of tax incentives are dubious at best.”

These results have clear implications for our reading of political science works on the topic: if economic factors that are tightly linked to MNCs’ bottom line are fragile predictors of FDI flows, then we must be cautious when assessing the influence of (political) explanators that are further removed in the causal chain.

There are other reasons to be careful when interpreting the output of country-year FDI regressions. For instance, the obsolescing bargain theoretical frame can lead us to underappreciate the fact that firms can use several strategies to mitigate risk: they can purchase investment insurance, invest in joint-ventures with local firms, build surplus capacity, or appeal to the investor-state dispute settlement provisions of international investment treaties. To be sure, implementing those strategies can be costly, and the need to purchase insurance (in one form or another) could discourage some investors. Still, the availability of risk mitigation strategies for firms should weaken the relationship between investment decisions and country-level indicators of political risk.

Another underappreciated feature of the political-economy of FDI is that “poor governance” can have ambiguous effects on firms’ incentives. Corruption, for example, imposes costs and uncertainty, since it acts as a *de facto* tax, and because bribery agreements are not enforceable by law. However, corruption may also be attractive for foreign investors if it helps firms strike advantageous deals with local officials (e.g., to gain access to resources, suppress labor, or relax environmental regulation). Indeed, there is a long tradition of work in political science and economics which shows that the “grabbing hand” of corruption is sometimes counterbalanced

³The conclusion in Blonigen and Piger (2014, 777) is representative: “The covariates with consistently high inclusion probabilities include traditional gravity variables, cultural distance factors, relative labour endowments and trade agreements. Variables with little support for inclusion are multilateral trade openness, most host-country business costs, host-country infrastructure (including credit markets) and host-country institutions.”

by a “helping hand” (see Tanzi (1998) for a review). Lui (1985), for example, points out that bribery can often be used to speed-up bureaucratic processes in the host country; Huntington (1968, 61) writes that corruption may allow multinationals to gain access to the political system in countries where they have a weak domestic base of support; and Li and Resnick (2003) argue that in corrupt environments, MNEs may seek to bribe and collude with the host government to influence domestic politics in the host country. There is also empirical support for this alternative view: Egger and Winner (2005), for instance, use panel data and show that, after controlling for confounders, corruption is positively correlated with FDI inflows.

In sum, while there has been much theoretical and empirical work on the political determinants of FDI, there are also good reasons to expect that country-level features of the political environment will only be weakly related to international investment flows. Whether, on balance, political factors are important explanators/predictors of investment flows remains an open empirical question.

2 MODEL FIT AND PREDICTIVE POWER

Political methodologists have long recognized the important role that fit statistics must play as complements to parameter and uncertainty estimates. A case in point is *The Great R^2 /SEE Debate of 1990*, everyone agreed that prediction-based goodness-of-fit measures must inform the interpretation of quantitative data analyses, but simply disagreed over which statistic to deploy in practice (Achen 1990; King 1990; Lewis-Beck and Skalaban 1990).

Yet, despite the obvious importance of model fit, researchers in many fields of political science still know little about the explanatory or predictive power of their theories. In *The perils of policy by p-value*, for instance, Ward, Greenhill, and Bakke (2010) show how the traditional focus on statistical significance in conflict studies has led scholars to emphasize relationships that have little bearing on real-world outcomes.

Likewise, very little attention has been paid to the explanatory or predictive power of theories which link politics to FDI. In light of the above discussion, it seems important to assess whether macro-political variables can help us develop better expectations about international investment flows. In the next section, I take a step in that direction by estimating variations on a statistical model from an important study of *Political Risk, Institutions, and Foreign Direct Investment* (Busse and Hefeker 2007).

3 EXPLAINED VARIANCE IN A PUBLISHED STUDY

In their article, Busse and Hefeker (2007, 397) aim to identify the political “indicators that matter most for the activities of multinational corporations.” Using regression analysis on panel data, the authors find “that government stability, internal and external conflict, corruption and ethnic tensions, law and order, democratic accountability of government, and quality of bureaucracy are highly significant determinants of foreign investment inflows.”

The work of Busse and Hefeker is a good starting point because (a) their article has already had tremendous influence on the field, and it continues to be cited regularly⁴; (b) their sample covers the set of developing countries where political risk should matter most for investment decisions; and (c) their empirical strategy is representative of many similar efforts, with a baseline model which should be familiar to readers:⁵

⁴On 2016-05-04, Google Scholar recorded 698 citations for Busse and Hefeker (2007).

⁵The estimates that this model produces could be subject to endogeneity bias if FDI has a feedback effect on political conditions in the host country. For instance, pressure from foreign investors could push local government to improve property rights protection, or investment could spur economic growth and an attendant improvement in political stability. For the purposes of this study, these factors do not pose a major threat to inference: they suggest that β_5 may be biased *away* from zero (the direction of the omitted relationship). In other words, it seems unlikely that the weak explanatory power of political variables is an artefact of endogeneity.

$$\ln \text{FDI} = \beta_1 \ln \text{GDP} + \beta_2 \ln \text{Growth} + \beta_3 \ln \text{Trade} + \beta_4 \ln \text{Inflation} + \\ \beta_5 \text{Political} + \beta_6 \text{Year} + \beta_7 \text{Year}^2 + \Gamma \Omega + \varepsilon.$$

In a first series of tests, I consider data on net FDI inflows from UNCTAD's World Investment Reports.⁶ Data on GDP (constant 2005 US\$), GDP growth, ratio of trade to GDP, and inflation are from the World Bank's World Development Indicators (2015).⁷ I consider 10 political regressors, all published by the Political Risk Services Group. Ω represents a vector of country fixed effects. I also include the year and its square to account for time trends. To guard against the possibility that year-to-year swings in FDI skew the results, all variables are averaged over 4-year intervals. The sample covers 89 countries which the World Bank identifies as low or middle income.

To see if political variables materially improve our capacity to explain changes in investment behavior, I estimate a baseline model without any political variable. Then, I estimate a series of models with different political predictors and compare measures of fit. Tables 1 and 2 present the results from this exercise (column names identify the *Political* variable used in each model, and standard error are robust to heteroskedasticity).

The results are broadly consistent with the original published analysis: favorable political conditions (high values on the *Political* variables) tend to be associated with higher levels of FDI inflows, and several of the *Political* coefficients are statistically significant. Coefficient estimates and standard errors, however, only tell part of the story. Here, the main quantity of interest is the increase in explained variance which can be obtained by adding a political variable to the

⁶Busse and Hefeker (2007) consider the effect of political variables on FDI flows *per capita*. In similar exercises, several other authors have considered net FDI inflows instead, adjusting for the size of the economy by controlling for GDP. Tables 2 and 3 of Busse and Hefeker (2007) already make clear that political variables do not make much contribution to fit in models of FDI *per capita*. Since an exact replication would be superfluous, all the models I estimate here focus on net FDI inflows.

⁷transform

Table 1: The Political determinants of foreign direct investment flows. Linear models on four-Year intervals with country fixed effects. (Table 1 of 2)

	Baseline	Accountability	Bureaucracy	Corruption	Ethnic Conf.	External Conf.
Ln GDP	-0.365 (0.369)	-0.332 (0.374)	-0.354 (0.361)	-0.366 (0.368)	-0.460 (0.372)	-0.603 (0.370)
Ln GDP growth	0.068 (0.027)	0.063 (0.026)	0.068 (0.027)	0.069 (0.026)	0.064 (0.026)	0.063 (0.026)
Ln (X+M)/GDP	0.846 (0.234)	0.827 (0.235)	0.848 (0.234)	0.844 (0.236)	0.820 (0.238)	0.815 (0.224)
Ln Inflation	-0.151 (0.061)	-0.133 (0.061)	-0.153 (0.061)	-0.149 (0.061)	-0.149 (0.058)	-0.158 (0.059)
Year	9.184 (3.583)	8.495 (3.585)	9.254 (3.547)	8.645 (3.679)	5.181 (3.628)	3.285 (4.233)
Year ²	-0.002 (0.001)	-0.002 (0.001)	-0.002 (0.001)	-0.002 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Political		0.198 (0.081)	-0.020 (0.139)	0.147 (0.106)	0.416 (0.093)	0.153 (0.054)
Adj. R ²	0.708	0.711	0.708	0.709	0.716	0.712
Num. obs.	651	651	651	651	651	651
RMSE	1.658	1.651	1.659	1.656	1.634	1.647

Table 2: The Political determinants of foreign direct investment flows. Linear models on four-Year intervals with country fixed effects. (Table 2 of 2)

	Baseline	Internal Conf.	Investment Profile	Law and Order	Military	Religious
Ln GDP	-0.365 (0.369)	-0.504 (0.368)	-0.806 (0.379)	-0.645 (0.369)	-0.571 (0.381)	-0.384 (0.370)
Ln GDP growth	0.068 (0.027)	0.065 (0.026)	0.061 (0.024)	0.064 (0.026)	0.066 (0.026)	0.068 (0.027)
Ln (X+M)/GDP	0.846 (0.234)	0.781 (0.226)	0.850 (0.242)	0.870 (0.227)	0.834 (0.227)	0.859 (0.234)
Ln Inflation	-0.151 (0.061)	-0.117 (0.059)	-0.062 (0.061)	-0.106 (0.059)	-0.138 (0.060)	-0.148 (0.061)
Year	9.184 (3.583)	3.785 (4.130)	2.694 (4.183)	3.207 (4.101)	7.845 (3.597)	8.696 (3.625)
Year ²	-0.002 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.002 (0.001)	-0.002 (0.001)
Political		0.180 (0.053)	0.298 (0.064)	0.399 (0.110)	0.208 (0.086)	0.071 (0.109)
Adj. R ²	0.708	0.715	0.721	0.716	0.711	0.708
Num. obs.	651	651	651	651	651	651
RMSE	1.658	1.639	1.622	1.637	1.650	1.658

baseline apolitical” model. In that respect, the results in Tables 1 and 2 are striking: none of the 10 PRS variables yields more than a 0.013 increase in the value of adjusted R^2 .

One concern is that the models shown above may be underspecified, since they omit several host-country determinants of investment location (labor force characteristics, technological capacity, etc.). To account for such factors, the models in Table 8 of the online appendix add four proxy variables: GDP *per capita*, literacy rate, share of urban population, and number of telephones per 1000 people (World Bank, 2014).^{8,9} Another potential problem stems from the nature of the dependent variable. As Kerner and Lawrence (2014) point out, considering flows may not always be appropriate to test theories of politics and FDI. Tables 9 through 16 of the online appendix thus show results using four alternative dependent variables. The first is a measure of inward FDI stock from UNCTAD. The other three measures are from the US Bureau of Economic Analysis: the value of Plant, Property, and Equipment; total assets; and capital expenditures.¹⁰ Finally, I present other tests with regional dummy variables, and country-year data instead of 4-year averages (Tables 17 to 20). None of these alternatives reveal a large difference in model fit between apolitical and political models.

3.1 *What could explain the weak explanatory power of the country-level models?*

These underwhelming results seem to cast doubt on the idea that PRS variables could hold strong, independent explanatory power over flows of international investment. But before drawing conclusions from this reexamination, it is worth considering four reasons why the political variables considered in Busse and Hefeker (2007) could appear to be such weak explanators.

⁸These additional controls are not available for the full panel. Accordingly, I apply the multiple imputation routine implemented by Honaker and King (2010).

⁹In terms of omitted variable bias, it is important to keep in mind that the type of structural/economic conditions which favor inward FDI flows tend to be positively correlated with “good governance”. It thus seems reasonable to expect that omitting to control for these conditions will induce a positive bias in the *political* coefficient estimates. In other words, the expected direction of bias works *against* the main finding of this paper, which is that political variables are weak predictors of investment decisions.

¹⁰The BEA data was obtained in panel data format from the replication package of Kerner and Lawrence (2014).

A) *Measurement error.* The main data that allow cross-national and over time comparisons in FDI flows at the aggregate level are published in UNCTAD's World Investment Reports and in the International Monetary Fund's Balance of Payments Yearbooks. The breadth of coverage and convenience of these data have made them a primary source for research on international investment in both political science and economics.

It is well known, however, that drawing inference from these datasets is fraught with problems. Kerner and Lawrence (2014) and Kerner (2014) provide excellent overviews of the methodological problems involved in measuring aggregate FDI, with a focus on the use of these measures in studies of the political determinants of FDI. Three problems, in particular, seem to be worth considering. First, bilateral FDI data typically record information on the location of an investment and the location of its direct owner, rather than its global ultimate owner. Second, cross-national comparisons are seriously hampered by discrepancies in data-collection methods and by inconsistent reporting standards. This is particularly problematic in the accounting of reinvested earnings, a category which makes up a very large share of total FDI. Third, MNCs often control capital assets that were financed through debt issued in the host-country's financial market, and these funds are typically unaccounted for in the aggregate FDI data.

These problems pose a nontrivial threat to inference, because political explanators are likely to be correlated with the error in our measures of FDI. For example, if property rights protection in the host-country dictates a firm's decision to invest there in the first place, it should also influence its choice to reinvest or repatriate foreign-earned income. And while typical treatments of the linear model tend to conclude that measurement error in the dependent variable is rather unproblematic (Greene 2008, 326), this conclusion does not hold where, as is the case here, the error is correlated with regressors (Wooldridge 2010, 76–82), or when the quantity of interest is the coefficient of determination (Majeske, Lynch-Caris, and Brelin-Fornari 2010). Therefore, it seems plausible that the weak results found above are due to measurement error.

B) Wrong explanators. The authors I cited in my brief review of the literature must be commended for developing precise concepts which can be translated into measurable quantities and testable propositions. As a general matter, however, it is fair to say that the political determinants of FDI could work through a variety of distinct channels. For example, “institutional constraints on the Executive” could reduce the cost of doing business by making expropriation less likely (property rights protection), by preserving the status quo (veto players), by empowering a professional and independent judiciary (rule of law), or by limiting the discretionary power and rent extraction capabilities of the bureaucracy and the Executive (corruption). Each of these causal pathways, in turn, could be operationalized using a slew of alternative measures. Sifting through the hundreds of “quality of governance” indicators assembled by Teorell et al. (2013) should suffice to convince anyone that there exists a plethora of likely political determinants of investment behavior.

In that context, readers could easily challenge the results of Busse and Hefeker (2007) on theoretical or measurement grounds, by arguing that the authors did not consider the “right” measure of political risk or institutions. It is thus possible that the lack of improvement in explained variance that we observed above would be due to mismeasurement of the explanatory variables.

But if several stories can link politics to FDI, and if we can test any of those arguments using a large set of alternative regressors, we run the risk of finding some statistically significant coefficient estimates, even if only by chance. In other words, the availability of many country-year political variables makes “fishing” easier than it perhaps should be, a problem which is compounded by the tendency of academic journals to publish novel results and articles that reject null hypotheses (Gerber and Malhotra 2008; Sterling 1959).

This implies that some published results about politics and FDI may be spurious, and suggests the need for a unified statistical framework which allows us to test the explanatory or predictive power of a large number of political predictors simultaneously.

C) *Conditional association.* As I explained above, Busse and Hefeker (2007) are far from alone in making “unconditional” claims about the relationship between political phenomena and international investment flows. However, several authors have since refined the theoretical analysis by arguing that the relationship in question should be moderated by firm or host-country characteristics. If the relationship between politics and FDI is indeed conditional, then finding weak results in a pooled sample should not come as a surprise.

D) *Within vs. between variance.* One notable feature of the research design by Busse and Hefeker (2007) is that the authors use a fixed effects regression model to estimate their quantities of interest. This choice is easy to understand and motivate on methodological grounds, and it is not surprising to see that the approach has become standard in this field of research.¹¹

It is important to recognize, however, that the within model can be limiting, because it forces us to ignore much of the relevant variation in the phenomena of interest. Indeed, political risk variables and institutions tend to be very sticky over time, and most of the variation in politics is observed between countries rather than within them. For example, consider the PRS *Bureaucracy Quality* variable: in the sample of 89 developing countries considered above, the mean of within-country standard deviations is 0.37, whereas the standard deviation of within-country means is 1.06.

A replication of models from Busse and Hefeker (2007) can show that political variables do not explain much of the within-country variance in net FDI inflows, but this exercise does not carry much information about politics and FDI along the cross-sectional dimension. Since that is where most of the relevant variation occurs, it seems possible that the tests presented above understate the true strength of association.

¹¹See for example All2011; Biglaiser and DeRouen (2006); Bütthe and Milner (2014); Garriga (2015); Garland and Biglaiser (2008); Halvorsen and Jakobsen (2013); Jensen (2003); Kerner (2009); Kim et al. (2015); Staats and Biglaiser (2012).

4 A FIRM-LEVEL DATA ANALYSIS

In the previous section, I replicated the results from a well known study of political risk, institutions and FDI, and showed that macro-level political variables have little effect on the share of variance in FDI that can be accounted for by our statistical models. I also pointed to four methodological problems which could explain this weak result. In the rest of this paper, I leverage a big firm-level dataset with non-parametric machine learning techniques to address these concerns and reassess the relationship between politics and FDI.

To address the question of measurement error, I jettison the flawed aggregate FDI data that are used in much of the literature, and consider a large firm-level dataset on the investment location decisions of multinationals. These data are cross-sectional. As I suggested above, this feature should not hamper the investigation of the effect of political variables, because political/institutional variables are sticky over time, and because most of the variation in political explanators is cross-sectional; that variation across countries should be particularly salient for business managers who must choose where they will invest.

To leverage these data, I use *Random Forests*, an algorithm which is often deployed in the field of machine learning to assess the predictive power of multiple predictors at once. This approach is particularly well suited to the context at hand, because forests can take into account complex interactive data structures like those implied by our conditional theories of politics and FDI.

In sum, applying the random forest algorithm to firm-level data allows us to assess the strength of association between politics and FDI, while addressing the four concerns raised above.

4.1 *Dependent variable*

The Orbis database is maintained by Bureau Van Dijk. It includes information on over 120 million private companies in nearly every country. Most of these firms run purely domestic operations, but many are multinationals. In the analyses, I consider the subset of companies that Orbis links to a Global Ultimate Owner (GUO) (subsidiaries), and exclude those that reside in the same country as their GUO (domestic investments). This leaves a sample of about 193,000 foreign subsidiaries and 157,000 parent companies. Orbis offers the distinct advantage of tracking multinationals from many countries, unlike more regionally-focused databases like the U.S. Bureau of Economic Analysis' Survey of Direct Investment Abroad. For all the analyses in this paper, I use a binary dependent variable which indicates whether a particular multinational owns a subsidiary in a given country.¹²

Orbis does not appear to have been exploited much in political science, but it is often used in other disciplines, where it tends to be treated as representative of the worldwide population of large GUOs.¹³ It is important to recognize, however, that Bureau van Dijk (2013) does not claim to produce a random sample of firms, but rather emphasizes the comprehensiveness and accuracy of its publications. Unfortunately, due to the many problems that afflict other measures of FDI, we do not have a good external benchmark to definitively establish the representativeness of the Orbis data.

Nevertheless, descriptive statistics can provide a measure of reassurance: Table 3 shows that the share of firms recorded by Orbis in countries of each income category is closely related to FDI flows. The rank of individual countries in the worldwide distribution of both variables is also very similar (see online appendix). Given that the two variables measure different but strongly related concept (aggregate levels of FDI vs. location decisions), this high degree of similarity is

¹²The paucity of reliable firm-level data on other features such as revenues or number of employees prevents me from using alternative dependent variables in the context of this paper.

¹³Recent examples include Arulampalam, Devereux, and Maffini (2012); Bhaumik, Driffield, and Pal (2010); Bloom, Sadun, and Reenen (2012); Chen, Li, and Shapiro (2012); Lampel and Giachetti (2013); Maffini and Mokkas (2011); Martins and Yang (2014); Nunn and Trefler (2013); Temouri and Driffield (2009); Voget (2011).

encouraging.

Table 3: Share of net FDI flows and share of recorded foreign subsidiaries by country income level. Sources: World Bank and Orbis.

	Inward		Outward	
	Orbis	FDI	Orbis	FDI
High income	62.6	74.9	84.1	92.3
Upper middle income	33.3	21.1	12.7	6.6
Lower middle income	3.6	3.6	3.1	1.0
Low income	0.5	0.3	0.2	0.0

4.2 Sampling

The unit of analysis for all the tests I describe below is the GUO-country. Because there are many possible hosts, and given that firms typically invest only in a handful of places, the dependent variable is highly unbalanced: just under 2% of observations show positive values. This “unbalanced labels” situation is common in machine-learning applications, and it poses two minor challenges: considering all observations imposes an unnecessary computational burden, and it is easy to devise a highly accurate but trivial prediction model.¹⁴

One common solution, which I adopt here, is to produce a balanced sample by selecting on the dependent variable: choose all observations with a value of one, and draw a random sample of non-events of the same size. This choice-based approach can be used with any of the classification algorithms that I describe below. The strategy is analogous to the sampling scheme used by Prentice and Pyke (1979) and advocated by King and Zeng (2001) in the rare events logit case. It produces datasets that are less computationally costly to use, and allows us to calculate measures of predictive accuracy that value our ability to classify ones and zeros equally.

¹⁴For example, guessing “no investment” for all observations gives about 98% prediction accuracy.

Table 4: Determinants of investment location. Logistic regression.

	Apolitical	Political
Intercept	-17.4972 (0.0721)	-16.3773 (0.0720)
Ln GDP	0.6931 (0.0027)	0.6292 (0.0028)
Distance	-0.1239 (0.0012)	-0.1172 (0.0012)
Polity		0.0801 (0.0010)
Num. obs.	276477	276477
Pred.Acc. %	77.9	78.4

4.3 A first look

Before considering the full set of predictors, it is useful to look at a subset of variables in the familiar context of logistic regression. My goal is not to compute credible point estimates. Rather, this descriptive exercise aims to provide a sanity check for the firm-level data, and to illustrate the fact that in large datasets, political predictors can appear strongly related to the conditional probability of investment even if this relationship is of minor relevance to real-world outcomes.

I begin by randomly splitting the dataset in two equal parts. The first is used to estimate logit models, and the second to test their predictive accuracy in new data. The models I consider in Table 4 include at most three regressors: log *GDP* of the host country (World Bank 2014), geographic *distance* between the host and the firm’s home-country (Mayer and Zignago 2006), and the level of *democracy* of the host (Marshall, Jaggers, and Gurr 2012).

Three points are especially noteworthy. First, all the coefficients are statistically different from zero. Given the sample size, this is not surprising. Second, the Orbis data generally behave as expected: *distance* between home and host countries is negatively associated with the likelihood of investment, but democracy and GDP both show positive coefficients.

The third conclusion we can draw from Table 4 is most crucial for the arguments in this paper: including a measure of regime type to the baseline model yields very little improvement in the out-of-sample predictive accuracy of the logit model. Indeed, the difference in predictive accuracy for the two models is less than 0.5 percentage point, which suggests that the *Polity* variable does not appear to have much predictive power, above and beyond the information that is already conveyed by gravity variables.

To generalize these insights, I conduct several other tests using the same basic Logit framework. To begin, I substitute the *polity* index for each of the 10 political regressors produced by the PRS group. Then, I consider a set of “thicker” models which control for distance, log GDP, log GDP *per capita*, share of urban population, and for the number of telephone lines per 1000 people. As Table 24 of the appendix shows, the main conclusions are unaltered. In the rest of the paper, I push further by using an algorithm which allows us to consider the predictive power of a much broader set of covariates.

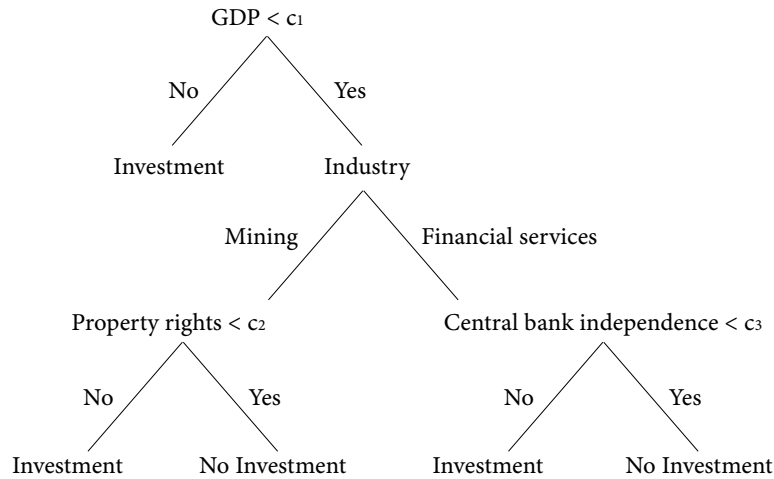
5 A NON-PARAMETRIC PREDICTION MODEL

Over the past years, many supervised learning algorithms have been developed to predict binary outcome variables using many predictors. Here, I consider one set of techniques in more detail: classification trees and random forests. As I argue below, this approach is particularly well-suited to the problem at hand, because it can be used to model unconditional as well as conditional relationships between politics and FDI.

5.1 *Trees*

Classification trees are a non-linear and non-parametric prediction technique developed by Breiman et al. (1984). Their accuracy, scalability, and ability to capture complex interactive data structures help explain why classification trees have had such an impact on statistical prac-

Figure 1: A hypothetical subtree for prediction of a binary outcome.



tice; the work of Breiman et al. (1984) has garnered over 28,000 citations¹⁵, and trees have been deployed in a wide variety of settings, from astronomy and genomics to sociology. In short, classification trees are not a novelty item. They are one of the most basic and useful tools available to researchers who work with high-dimensional data.

To understand how this technique works, it is useful to proceed in two steps. To begin, I illustrate how an existing tree can be used to make predictions on a binary variable. Then, I consider some of the computational aspects involved in growing a tree.

Figure 1 shows a decision rule designed to predict if a large multinational x will invest in country y . Starting from the top, we begin by checking if the host is a small economy (GDP smaller than some cutpoint c_1). If the host is large, we predict that the firm will have a subsidiary there; if the host is small, more conditions need to be considered. The top node of the tree refers to a property of the host-country (GDP), but the second level node splits the data based on firm characteristics instead (industry). If the multinational works in mining, a capital-intensive activity which renders it sensitive to expropriation, property rights enforcement could be an important factor in location decisions. If the firm comes from the financial sector instead, then price stability and central bank independence may be better predictors of investment behavior.

¹⁵Google Scholar, 2015-10-16.

This decision rule illustrates one of the main benefits of the approach: classification trees can capture the types of complex interactive data structures that are often of interest to political scientists (Brambor, Clark, and Golder 2006; Franzese and Kam 2009). In Figure 1, the investment decision was unconditionally related to the size of the host economy (larger markets attract international investors). The top node of that tree could also be interpreted as embodying the (conditional) broad vs. narrow distinction from Mosley (2003, 37), where the author argues that international investors get more useful information by monitoring politics in the developing world rather than in developed economies.¹⁶ Figure 1 also accounts for the possibility that capital-intensive industries like mining may be more sensitive to property rights protection, an argument which is consonant with the work of Kerner and Lawrence (2014) and several others.

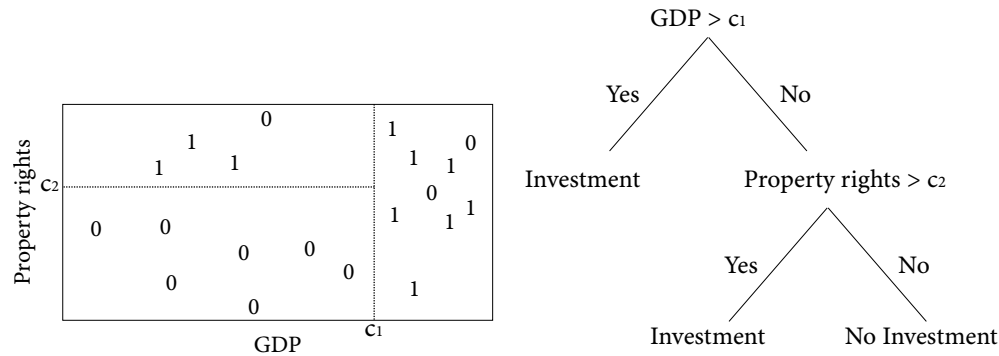
Classification trees like this one are easy to construct. To understand how, it is useful to think of trees not in static terms, but rather as a process of recursive partitioning of the predictor space. More concretely, we need to answer three main questions: (1) which variables will be used to split the data, (2) in what order, (3) based on which cut-points?

Figure 2 illustrates the process of growing a tree in a simple case with two predictors (GDP and property rights) and 20 observations on a binary outcome. In the left panel, each number represents the decision to invest in a host country with a particular combination of GDP and property rights (“1” means investment, and “0” no investment).

To begin, notice that if we naively predict 1 for all observations, we achieve 50% predictive accuracy (0s and 1s are represented in equal proportions in this dataset). We can improve on this by making different predictions for different subsets of the data. For example, if we expect more investment in rich countries, we could predict 1 for every observation that falls to the right of cut-point c_1 , and 0 for all others. This would allow us to classify 15 of the 20 observations accurately. We can further improve the model by recursive partitioning: after the c_1 split, we

¹⁶The comparison between Figure 1 and the argument in Mosley (2003) is admittedly imperfect, as the latter focuses on portfolio investors rather than direct investors. The broad vs. narrow distinction would also be better characterized as a function of GDP *per capita* rather than GDP.

Figure 2: Partitioning a predictor space with two variables and a binary outcome.



make a second cut at c_2 , and predict 1 everywhere but in the lower-left section. This raises prediction accuracy to 85%. The resulting decision rule is represented in tree-form in the right panel of Figure 2.¹⁷ Notice that, as in the previous example, conditional/interactive arguments will be accommodated by the model, *as long as they improve fit*.

In the simple case of Figure 2, it was easy to see that c_1 and c_2 were optimal cut-points. In practice, however, we often want to grow trees that include many more predictors. To deal with high-dimensional data, software implementations¹⁸ of the model use greedy algorithms¹⁹ that select appropriate partitions by minimizing a prediction-based loss function (Hastie et al. 2013, 309). Greedy procedures are computationally efficient, and they tend to perform well in most applications.

Unfortunately, greedy procedures also impart an undesirable property to the prediction model: when considering a set of correlated predictors, the algorithm will sometimes pick one variable and ignore its cousins. Since many of the political determinants of FDI listed in Table 7 are correlated, this means that we risk growing a tree which relies on a subset of predictors that was chosen somewhat arbitrarily. Classification trees are also burdened by a second important

¹⁷This procedure can also be used to predict continuous or categorical outcomes by predicting constants in each subsets of the data.

¹⁸I use the Python implementation of Pedregosa et al. (2011), but easy-to-use programs are available on most platforms.

¹⁹Greedy algorithms make a series of locally optimal decisions in hope of finding the global optimum. Readers will find a discussion of their properties in reference books on algorithms such as Cormen et al. (2009 Ch.16).

drawback: they tend to have low bias but high variance (Hastie et al. 2013). This, in turn, may reduce their ability to make accurate out-of-sample predictions.

5.2 Forests

Thankfully, both of these concerns can be assuaged by using an ensemble method like bagging: “The essential idea in bagging is to average many noisy but approximately unbiased models, and hence reduce the variance. Trees are ideal candidates for bagging, since they can capture complex interaction structures in the data, and if grown sufficiently deep, have relatively low bias” (Hastie et al. 2013, 587). In that spirit, Breiman (2001) developed a modification of bagging which he named “Random Forests”. Forests are grown as follows:

1. create many bootstrap samples by drawing observations with replacement from the original dataset;
2. for each bootstrap sample, randomly select a subset of predictor variables, and use this subset of predictors to grow a classification tree;
3. calculate the ensemble predictions by taking a majority vote amongst all the trees (for each observation, the ensemble prediction corresponds to the most common predicted value across the bootstrapped trees).

The most important characteristic of this procedure is that bootstrapped trees are grown using different predictors and that, as a result, they will yield a diverse set of predictions for the outcome variable. This matters because, as Dietterich (2000, 1) points out, a “necessary and sufficient condition for an ensemble of classifiers to be more accurate than any of its individual members is if the classifiers are accurate and diverse (Hansen and Salamon 1990).”²⁰ For this

²⁰“An accurate classifier is one that has an error rate of better than random guessing on new x values. Two classifiers are diverse if they make different errors on new data points (Dietterich 2000, 1–2).”

reason, random forests typically outperform single trees in terms of variance and out-of-sample predictive accuracy (Breiman 2001; Hastie et al. 2013).

An additional benefit of this random subsetting approach, is that it gives each predictor a chance to be considered in the absence of its close correlates. This reduces the likelihood that a variable will be completely ignored when other correlated predictors are greedily selected by the optimization algorithm. Thus, forests tend to handle multicollinearity much better than single trees.

6 THE POLITICAL DETERMINANTS OF INVESTMENT LOCATION DECISIONS

The main purpose of this paper is to determine if we can use political information to build more accurate models of investment behavior. In this section, I use split-sample tests to compare the predictive accuracy of two random forests. The first is a minimalist *apolitical* model which only relies on the following predictors: GDP of the host country, (World Bank 2014), and the geographic distance between home and host countries (Mayer and Zignago 2006). The second forest is a *political* model which adds 31 political variables to the baseline (Table 7 lists all variables and sources).

The selection criteria for country-level political variables are exceedingly liberal: they need to measure a macro feature of politics in the host country, to be plausibly related to the behavior of international investors, and to be available for a broad cross-section of countries. My goal is to include many variables in order to stack the odds in favor of a “politics matters” conclusion. The list of political predictors that I consider includes many of the measures that were used in prior research on the political determinants of FDI (e.g., regime type, property rights protection, left-wing government), as well as several other variables that do not seem to have been considered

so far (e.g., religion in politics).²¹

Readers will note that the list of variables in Table 7 is limited in (at least) one important respect: it includes political factors that could be characterized as “institutional” or “environmental”, but excludes more specific public policies such as corporate tax rates. This choice imposes strong constraints on the inference which can be drawn from the research design, but this is a compromise we must accept. To see why, compare how political scientists have operationalized the link between democratic institutions and FDI (Jensen 2003; Li and Resnick 2003) to the rich empirical literature on corporate taxation produced since the pioneering Hartman (1984) paper. While the democracy-FDI relationship can often be modelled using a linear additive structure or with simple multiplicative interactions, estimating the effect of tax policy on international investment requires a very different methodological apparatus. Indeed, public finance economists tend to employ complex response functions that account for a wide array of factors, including the wedge between statutory tax rates at home and abroad, whether the income of a multinational is subject to territorial or worldwide taxation, the tax rate that applies to repatriated earnings, the difference between expected after-tax returns in home and host countries, and the cost of relocation or income shifting (Hartman 1985; Hines and Rice 1994; Mooij and Ederveen 2008; Slemrod 1990). As I explained above, the prediction algorithms that I use can account for the linear and interactive data structures that are used in the “political determinants of FDI” literature, but it cannot act as substitute for the types of identification strategies that need to be deployed for policy evaluation. Restricting the set of variables that I consider clarifies the scope conditions of my claims, and links those claims more tightly to the political science literature to which this paper speaks.

In addition to the variables described above, the prediction models also leverage firm-level predictors. First, I include a measure of *internationalization*, that is, of the number of countries

²¹Because an MNC’s presence today depends on the political environment in the recent past, I take country averages for each variable over the ten years that preceded acquisition of the cross-sectional Orbis data (2002-2012). Unfortunately, the Orbis data does not reveal when subsidiaries were formed.

in which a company has operations. This is a useful predictor, which acts as a proxy for a firm's size and for the intensity of its international activities. Second, I introduce a series of binary variables that record whether firms have activities in each of the 25 top-level codes of the North American Industry Classification System (NAICS). These are important, because firms in certain sectors such as shipping may establish many more subsidiaries, simply due to the nature of their activities.²²

The NAICS dummies are also critical because they allow the random forests to accommodate the possibility that some firms are more sensitive to political conditions than others. For example, since there is a strong relation between industrial category and capital intensity (Table 5), a classification tree with NAICS dummies can account for (conditional) theoretical arguments which link capital-intensity, likelihood of expropriation, and investment location decisions. The dummies also allow us to represent other types of conditional arguments, such as those that inspired the tree in Figure 1.

6.1 *Random forest results*

As in the logistic regression exercise shown above, I split the dataset in two equal parts randomly. I use the “training set” to fit a random forest model, and the “test set” to measure its predictive accuracy in new data.²³

The first row of Table 6 reports the out-of-sample predictive accuracy for the baseline political and apolitical models. Two main conclusions emerge. First, the out-of-sample predictive

²²The Orbis dataset that I use is very limited in terms of firm-level information. For instance, while it includes variables for the number of employees and revenues of subsidiaries, these data display enough (non-random) missingness to be useless for inference purposes.

²³In applications where predictive power is paramount, a better approach would be to divide the dataset in three, using one part as a “validation set” to optimize the tuning parameters of the predictive model through grid search and cross-validation. For simplicity, I grow a forest of 50 trees using a reasonable set of default tuning parameters: the criterion used to measure the quality of splits is Gini impurity, the maximum number of predictors is fixed to the square root of the number of available predictors, and the depth of trees is not otherwise constrained. In Table 23 of the online appendix, I show that using different tuning parameters does not improve predictive accuracy substantially, and that it does not produce a larger gap between the predictive accuracies of apolitical and political models.

Table 5: Average capital shares by industry (1987-2005).

Industry	Capital share
Educational services	0.10
Management of companies and enterprises	0.20
Health care and social assistance	0.22
Durable goods	0.27
Administrative and waste management services	0.28
Construction	0.32
Other services, except government	0.33
Professional, scientific, and technical services	0.34
Transportation and warehousing	0.35
Accommodation and food services	0.36
Retail trade	0.42
Arts, entertainment, and recreation	0.42
Finance and insurance	0.45
Wholesale trade	0.46
Nondurable goods	0.47
Information	0.53
Mining	0.66
Utilities	0.77

Source: Acemoglu and Guerrieri (2008).

accuracy of the random forest model is high, reaching about 90% in a test set with equal proportions of 0s and 1s. Second, *considering politics does not improve our expectations with respect to investment behavior*. Even if we add a large array of political predictors to the baseline, the predictive accuracy of the political and apolitical models remains substantively indistinguishable.

So far, the main empirical strategy has been to compare the predictive accuracy of competing models, which I considered as monolithic blocks of predictors. An alternative way to assess the predictive power of *individual* variables is to compute their *importance*, or *mean decrease impurity*. For each node t in the tree T , we make a split s_t in the dataset using variable $\nu(s_t)$. Let $p(t)$ be the proportion of observations to reach node t in the decision tree, and $\Delta i(s_t, t)$ be the decrease in gini impurity that results from splitting at t . The importance of variable X_m is defined as the weighted impurity decrease at all nodes where ν is used, averaged over all N_T trees in the forest (Louppe et al. 2013, 2):

$$Imp(X_m) = \frac{1}{N_T} \sum_T \sum_{t \in T: \nu(s_t) = X_m} p(t) \Delta i(s_t, t),$$

Intuitively, the *importance* of X_m measures the reduction in misclassification that results from using X_m to partition the sample space, weighed by number of observations that were classified using that variable. Variables which allow accurate classification and/or which are located near the top of the tree will thus show a high level of importance. Figure 3 shows that, unsurprisingly, three of the variables hold much more importance than any other in producing accurate classification: *GDP*, *Distance*, and *Internationalization*.

Figure 3: Variable importance in the baseline random forest model.

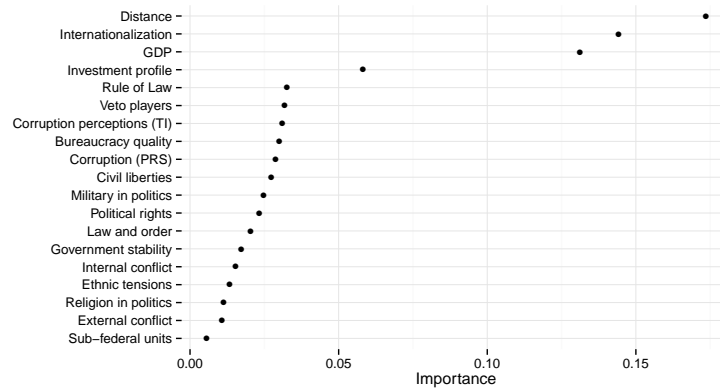


Table 6: Out-of-sample prediction accuracy (%) of political and apolitical models.

Sample	Apolitical	Political
Baseline	90.6	90.9
Many predictors	88.6	88.7
Low income	92.2	92.4
High income	87.0	87.2

6.2 Robustness

I take several steps to ensure that these findings are robust. First, because not all the political variables have broad country coverage, I re-estimate the prediction models using a subset of 16 variables that cover more host countries (Table 6, row 2). Second, to account for the possibility that political indicators carry less information in developed nations (Mosley 2003), I calculate predictive accuracy separately for high and low income countries (Table 6, rows 3-4).²⁴ Third, because the models described in Table 6 could be considered underspecified, I estimate “thicker” prediction models which also include the share of urban population, number of telephones per 1000 people, GDP *per capita*, and population size (World Bank 2014); as well as measures of economic, administrative and demographic distance between home and host countries (Berry, Guillén, and Zhou 2010). Fourth, to be certain that my results are not driven by firms in low

²⁴I refer to countries as “low income” when they fall below the high income threshold established by the World Bank (\$12,616 GNI *per capita* in 2012).

tax jurisdictions, I exclude all countries that belong to the list of tax havens assembled by Gravelle (2013). Fifth, as a further check on the representativeness of the Orbis data, I re-estimate the models using a stratified sample of GUOs, drawn in proportion to the GDP of their home countries. Sixth, to confirm that results are not driven by the choice of prediction algorithm, I consider four alternative procedures: logistic regression, Extremely Randomized Trees (Geurts, Ernst, and Wehenkel 2006), Bagging (Pedregosa et al. 2011), and AdaBoost (Freund, Schapire, and others 1996). Seventh, to guard against the possibility that the predictive power of (geographically clustered) political variables is captured by the *distance* variable, I re-estimate all models without considering geography. Finally, I re-estimate models while excluding outlier countries. To identify those outliers, I regress the share of subsidiaries (parents) against the share of FDI inflow (outflow) that goes to each country, and I inspect the studentized residuals from those regressions. I consider each country with a Bonferroni p-value smaller than 0.05 as an outlier, and I exclude them from the prediction model.²⁵

The results for robustness checks 3–8 are reported in the online appendix. None of these alternative approaches alter the main conclusion: political variables remain very weak predictors of investment location decisions.

7 CONCLUSION

In theory, many political phenomena could affect multinationals' propensity to invest in a country. The goal of this paper was to offer a systematic empirical assessment of the explanatory and predictive power of several frequently cited "political determinants of FDI". Using a replication exercise and an original data analysis for a large sample of multinationals, I showed that the information carried by country-level political indicators does *not* allow us to develop substantially better expectations about firm behavior.

²⁵Outliers: Cyprus, Germany, Italy, Netherlands, Romania.

This finding could have important consequences for how we understand policymaking in an integrated world. If macro-political and institutional features only play a marginal role in the location decisions of MNCs, then nations may enjoy latitude in the design of their governing institutions, in spite of growing pressure by international capital markets.

This paper's findings should also inform a re-evaluation of the state of knowledge in our field. To be clear, the results described herein must not lead to a wholesale rejection of the idea that politics matters for FDI. Rather, prior works on the topic must be assessed as a function of readers' substantive interest and purpose. If we are mainly interested in the mechanisms that link political variables to investment behavior, either because of their intrinsic importance or because their logic carries over to other contexts, then it makes little sense to adopt predictive accuracy or share of explained variance as a criteria of substantive importance. In contrast, if we motivate our inquiry by pointing to politics as a "driver" of FDI, then it is incumbent on the researcher to demonstrate that the proposed theory fits new data well enough to have practical significance. In such cases, I argued that computing out-of-sample measures of predictive accuracy imposes itself as a best practice for model assessment.

Finally, the present paper leaves open several promising avenues for future research. For instance, it is important to recognize that the location of subsidiaries and aggregate levels of FDI flows are not the only salient features of MNCs' behavior. Much could thus be learned by following the example of authors such as Malesky (2008) and Henisz (2000), who study how other aspects of international investment such as province-level investment stocks, mode of entry (joint-ventures vs. fully-owned subsidiaries), financing decisions (equity vs. debt, local vs. international borrowing), or the location of headquarters (corporate inversions). Alternatively, political-economists could develop new theory to explain why different types of FDI may be more sensitive to political conditions than others²⁶, or focus on political consequences of FDI

²⁶Note that the results of random forest models do account for industry-based heterogeneity in sensitivity to political risk.

flows.²⁷

²⁷There is already much excellent work in this line of research. A good example can be found in Manger (2005) and Manger (2009), where the author argues that FDI increases governments' incentives to conclude preferential trade agreements.

8 ONLINE APPENDIX

8.1 *Variations on Busse and Hefeker (2007)*

Table 7: Variable sources

Variable	Source
Alesina et al.	Ethnic fractionalization
Cingranelli & Richards	Physical integrity rights
Crowe & Meade	Central bank independence
Database of Political Institutions	Left-wing executive
Database of Political Institutions	Maximum difference of orientation
Database of Political Institutions	Proportional representation
Database of Political Institutions	Total fractionalization
Fraser Institute	Size of government
Freedom House	Civil liberties
Freedom House	Political rights
Henisz	Political constraints
Henisz	Sub-federal units
Heritage Foundation	Business freedom
Heritage Foundation	Labor freedom
Heritage Foundation	Property rights freedom
Heritage Foundation	Trade freedom
PRS Group	Bureaucracy quality
PRS Group	Corruption
PRS Group	Ethnic tensions
PRS Group	External conflict
PRS Group	Government stability
PRS Group	Internal conflict
PRS Group	Investment profile
PRS Group	Law and order
PRS Group	Military in politics
PRS Group	Religion in politics
Polity IV	Polity IV
Reporters Sans Frontieres	Press freedom
Transparency International	Corruption perceptions
World Bank	Rule of law
World Economic Forum	Judicial independence

Table 8: Coefficients of determination for regression models with host-country controls and with multiple imputation

Variable	R2
Baseline	0.68
Investment Profile	0.70
Internal Conf.	0.69
External Conf.	0.68
Corruption	0.68
Military	0.68
Religious	0.68
Law and Order	0.69
Ethnic Conf.	0.69
Accountability	0.69
Bureaucracy	0.68

Table 9: FDI stock. (Table 1 of 2)

	Baseline	Accountability	Bureaucracy	Corruption	Ethnic Conf.	External Conf.
Ln GDP	1.978 (0.126)	1.922 (0.120)	1.970 (0.121)	1.786 (0.131)	1.912 (0.122)	2.011 (0.122)
Ln GDP growth	0.049 (0.017)	0.045 (0.016)	0.049 (0.017)	0.045 (0.016)	0.045 (0.016)	0.050 (0.017)
Ln (X+M)/GDP	1.348 (0.220)	1.333 (0.221)	1.336 (0.224)	1.295 (0.205)	1.328 (0.211)	1.373 (0.224)
Ln Inflation	-0.164 (0.052)	-0.153 (0.052)	-0.161 (0.053)	-0.161 (0.050)	-0.164 (0.050)	-0.162 (0.051)
Political		0.101 (0.052)	0.052 (0.077)	-0.299 (0.071)	0.206 (0.074)	-0.027 (0.027)
Adj. R ²	0.834	0.835	0.834	0.842	0.838	0.834
Num. obs.	645	645	645	645	645	645
RMSE	0.962	0.959	0.962	0.941	0.952	0.962

Table 10: FDI stock. (Table 2 of 2)

	Baseline	Internal Conf.	Investment Profile	Law and Order	Military	Religious
Ln GDP	1.978 (0.126)	2.017 (0.123)	1.814 (0.148)	1.987 (0.125)	2.003 (0.128)	1.979 (0.126)
Ln GDP growth	0.049 (0.017)	0.050 (0.017)	0.045 (0.016)	0.049 (0.017)	0.049 (0.017)	0.048 (0.017)
Ln (X+M)/GDP	1.348 (0.220)	1.397 (0.234)	1.332 (0.217)	1.355 (0.222)	1.377 (0.219)	1.356 (0.223)
Ln Inflation	-0.164 (0.052)	-0.172 (0.052)	-0.141 (0.053)	-0.168 (0.053)	-0.170 (0.051)	-0.166 (0.052)
Political		-0.041 (0.025)	0.077 (0.036)	-0.045 (0.078)	-0.127 (0.033)	-0.049 (0.045)
Adj. R ²	0.834	0.835	0.836	0.834	0.836	0.834
Num. obs.	645	645	645	645	645	645
RMSE	0.962	0.961	0.958	0.962	0.957	0.962

Table 11: BEA Property, Plants, and Equipment. (Table 1 of 2)

	Baseline	Accountability	Bureaucracy	Corruption	Ethnic Conf.	External Conf.
Ln GDP	1.024 (0.159)	1.018 (0.165)	1.023 (0.159)	0.982 (0.164)	1.090 (0.158)	1.001 (0.162)
Ln GDP growth	-0.000 (0.013)	-0.001 (0.014)	-0.001 (0.013)	-0.002 (0.014)	0.001 (0.013)	0.001 (0.014)
Ln (X+M)/GDP	0.199 (0.244)	0.195 (0.246)	0.194 (0.243)	0.180 (0.236)	0.219 (0.243)	0.203 (0.243)
Ln Inflation	-0.142 (0.063)	-0.142 (0.063)	-0.141 (0.062)	-0.138 (0.061)	-0.143 (0.061)	-0.140 (0.063)
Political		0.010 (0.042)	0.112 (0.107)	-0.063 (0.059)	0.178 (0.116)	-0.034 (0.036)
Adj. R ²	0.959	0.959	0.959	0.959	0.960	0.959
Num. obs.	281	281	281	281	281	281
RMSE	0.584	0.586	0.585	0.585	0.577	0.585

Table 12: BEA Property, Plants, and Equipment. (Table 2 of 2)

	Baseline	Internal Conf.	Investment Profile	Law and Order	Military	Religious
Ln GDP	1.024 (0.159)	1.025 (0.161)	1.019 (0.169)	0.932 (0.167)	1.058 (0.166)	1.032 (0.159)
Ln GDP growth	-0.000 (0.013)	-0.000 (0.013)	-0.000 (0.013)	-0.002 (0.013)	-0.000 (0.013)	-0.000 (0.013)
Ln (X+M)/GDP	0.199 (0.244)	0.204 (0.238)	0.197 (0.243)	0.184 (0.243)	0.236 (0.246)	0.195 (0.240)
Ln Inflation	-0.142 (0.063)	-0.142 (0.063)	-0.141 (0.064)	-0.148 (0.062)	-0.144 (0.062)	-0.145 (0.065)
Political		-0.011 (0.035)	0.002 (0.025)	-0.144 (0.059)	-0.080 (0.040)	0.036 (0.060)
Adj. R ²	0.959	0.959	0.959	0.960	0.959	0.959
Num. obs.	281	281	281	281	281	281
RMSE	0.584	0.586	0.586	0.582	0.584	0.585

Table 13: BEA Capital Expenditures. (Table 1 of 2)

	Baseline	Accountability	Bureaucracy	Corruption	Ethnic Conf.	External Conf.
Ln GDP	0.906 (0.195)	0.855 (0.194)	0.905 (0.195)	0.976 (0.209)	0.943 (0.204)	0.921 (0.198)
Ln GDP growth	-0.034 (0.014)	-0.037 (0.015)	-0.034 (0.014)	-0.032 (0.015)	-0.034 (0.014)	-0.035 (0.015)
Ln (X+M)/GDP	-0.143 (0.302)	-0.180 (0.306)	-0.143 (0.303)	-0.114 (0.300)	-0.135 (0.301)	-0.152 (0.298)
Ln Inflation	0.041 (0.063)	0.046 (0.062)	0.043 (0.062)	0.033 (0.062)	0.042 (0.062)	0.041 (0.063)
Political		0.083 (0.058)	0.197 (0.162)	0.084 (0.080)	0.107 (0.117)	0.029 (0.050)
Adj. R ²	0.932	0.932	0.932	0.932	0.932	0.932
Num. obs.	302	302	302	302	302	302
RMSE	0.736	0.735	0.736	0.736	0.735	0.737

Table 14: BEA Capital Expenditures. (Table 2 of 2)

	Baseline	Internal Conf.	Investment Profile	Law and Order	Military	Religious
Ln GDP	0.906 (0.195)	0.905 (0.195)	0.807 (0.212)	0.940 (0.210)	0.928 (0.199)	0.970 (0.194)
Ln GDP growth	-0.034 (0.014)	-0.034 (0.014)	-0.035 (0.014)	-0.034 (0.014)	-0.034 (0.014)	-0.033 (0.014)
Ln (X+M)/GDP	-0.143 (0.302)	-0.145 (0.301)	-0.173 (0.300)	-0.138 (0.301)	-0.117 (0.307)	-0.157 (0.297)
Ln Inflation	0.041 (0.063)	0.042 (0.063)	0.056 (0.064)	0.044 (0.062)	0.040 (0.063)	0.036 (0.063)
Political		0.005 (0.039)	0.049 (0.031)	0.052 (0.083)	-0.064 (0.078)	0.197 (0.073)
Adj. R ²	0.932	0.932	0.932	0.932	0.932	0.934
Num. obs.	302	302	302	302	302	302
RMSE	0.736	0.738	0.736	0.737	0.737	0.727

Table 15: BEA Total Assets (Table 1 of 2)

	Baseline	Accountability	Bureaucracy	Corruption	Ethnic Conf.	External Conf.
Ln GDP	1.392 (0.199)	1.373 (0.203)	1.392 (0.199)	1.346 (0.199)	1.402 (0.195)	1.361 (0.202)
Ln GDP growth	0.008 (0.008)	0.007 (0.009)	0.008 (0.008)	0.006 (0.008)	0.008 (0.008)	0.009 (0.009)
Ln (X+M)/GDP	0.226 (0.194)	0.213 (0.197)	0.225 (0.194)	0.205 (0.194)	0.229 (0.196)	0.232 (0.195)
Ln Inflation	-0.157 (0.057)	-0.155 (0.057)	-0.157 (0.057)	-0.152 (0.057)	-0.157 (0.057)	-0.154 (0.056)
Political		0.035 (0.052)	0.001 (0.082)	-0.069 (0.042)	0.028 (0.062)	-0.045 (0.033)
Adj. R ²	0.966	0.966	0.966	0.966	0.966	0.966
Num. obs.	281	281	281	281	281	281
RMSE	0.566	0.567	0.567	0.566	0.567	0.566

Table 16: BEA Total Assets (Table 2 of 2)

	Baseline	Internal Conf.	Investment Profile	Law and Order	Military	Religious
Ln GDP	1.392 (0.199)	1.391 (0.200)	1.395 (0.191)	1.246 (0.211)	1.431 (0.205)	1.386 (0.199)
Ln GDP growth	0.008 (0.008)	0.008 (0.008)	0.008 (0.008)	0.006 (0.008)	0.008 (0.008)	0.008 (0.008)
Ln (X+M)/GDP	0.226 (0.194)	0.223 (0.190)	0.227 (0.196)	0.202 (0.190)	0.268 (0.189)	0.228 (0.193)
Ln Inflation	-0.157 (0.057)	-0.157 (0.057)	-0.158 (0.059)	-0.166 (0.056)	-0.159 (0.056)	-0.155 (0.058)
Political		0.006 (0.036)	-0.002 (0.025)	-0.227 (0.051)	-0.092 (0.041)	-0.024 (0.051)
Adj. R ²	0.966	0.966	0.966	0.967	0.966	0.966
Num. obs.	281	281	281	281	281	281
RMSE	0.566	0.567	0.567	0.557	0.565	0.567

Table 17: Region fixed effects. (Table 1 of 2)

	Baseline	Accountability	Bureaucracy	Corruption	Ethnic Conf.	External Conf.
Ln GDP	2.859 (0.238)	2.659 (0.230)	2.884 (0.239)	2.819 (0.249)	2.697 (0.234)	2.561 (0.229)
Ln GDP growth	0.113 (0.029)	0.100 (0.027)	0.112 (0.029)	0.112 (0.029)	0.104 (0.028)	0.101 (0.028)
Ln (X+M)/GDP	1.275 (0.367)	1.198 (0.364)	1.294 (0.374)	1.274 (0.364)	1.187 (0.365)	1.152 (0.334)
Ln Inflation	-0.205 (0.070)	-0.165 (0.070)	-0.212 (0.071)	-0.205 (0.069)	-0.204 (0.066)	-0.221 (0.067)
Political		0.375 (0.084)	-0.145 (0.156)	-0.057 (0.120)	0.530 (0.104)	0.214 (0.049)
Adj. R ²	0.647	0.658	0.647	0.647	0.662	0.657
Num. obs.	651	651	651	651	651	651
RMSE	1.823	1.796	1.823	1.824	1.785	1.797

Table 18: Region fixed effects. (Table 2 of 2)

	Baseline	Internal Conf.	Investment Profile	Law and Order	Military	Religious
Ln GDP	2.859 (0.238)	2.568 (0.223)	1.997 (0.247)	2.771 (0.230)	2.839 (0.237)	2.849 (0.235)
Ln GDP growth	0.113 (0.029)	0.102 (0.027)	0.094 (0.026)	0.107 (0.028)	0.112 (0.028)	0.113 (0.029)
Ln (X+M)/GDP	1.275 (0.367)	1.095 (0.319)	1.163 (0.366)	1.260 (0.350)	1.271 (0.362)	1.289 (0.358)
Ln Inflation	-0.205 (0.070)	-0.160 (0.068)	-0.085 (0.068)	-0.170 (0.068)	-0.201 (0.070)	-0.201 (0.070)
Political		0.256 (0.051)	0.409 (0.058)	0.372 (0.102)	0.084 (0.092)	0.098 (0.117)
Adj. R ²	0.647	0.663	0.675	0.654	0.647	0.647
Num. obs.	651	651	651	651	651	651
RMSE	1.823	1.781	1.751	1.804	1.823	1.824

Table 19: Country-Year dataset. BEA Property, Plants, and Equipment. (Table 1 of 2)

	Baseline	Accountability	Bureaucracy	Corruption	Ethnic Conf.	External Conf.
Ln GDP	1.366 (0.142)	1.390 (0.147)	1.362 (0.141)	1.317 (0.141)	1.394 (0.141)	1.339 (0.145)
Ln GDP growth	-0.010 (0.004)	-0.010 (0.004)	-0.010 (0.004)	-0.010 (0.004)	-0.010 (0.004)	-0.009 (0.004)
Ln (X+M)/GDP	0.079 (0.127)	0.080 (0.127)	0.077 (0.126)	0.069 (0.122)	0.097 (0.127)	0.095 (0.128)
Ln Inflation	-0.019 (0.010)	-0.019 (0.010)	-0.019 (0.010)	-0.020 (0.010)	-0.019 (0.010)	-0.021 (0.010)
Political		-0.031 (0.024)	0.089 (0.077)	-0.051 (0.030)	0.094 (0.050)	-0.041 (0.018)
Adj. R ²	0.969	0.969	0.969	0.969	0.969	0.969
Num. obs.	808	808	808	808	808	808
RMSE	0.508	0.508	0.508	0.507	0.506	0.507

Table 20: Country-Year dataset. BEA Property, Plants, and Equipment. (Table 2 of 2)

	Baseline	Internal Conf.	Investment Profile	Law and Order	Military	Religious
Ln GDP	1.366 (0.142)	1.383 (0.147)	1.353 (0.142)	1.307 (0.143)	1.410 (0.147)	1.360 (0.142)
Ln GDP growth	-0.010 (0.004)	-0.010 (0.004)	-0.010 (0.004)	-0.010 (0.004)	-0.010 (0.004)	-0.010 (0.004)
Ln (X+M)/GDP	0.079 (0.127)	0.085 (0.125)	0.079 (0.127)	0.061 (0.127)	0.116 (0.126)	0.087 (0.126)
Ln Inflation	-0.019 (0.010)	-0.020 (0.010)	-0.018 (0.010)	-0.024 (0.010)	-0.019 (0.010)	-0.018 (0.010)
Political		-0.020 (0.019)	0.006 (0.012)	-0.108 (0.027)	-0.091 (0.026)	-0.033 (0.030)
Adj. R ²	0.969	0.969	0.969	0.969	0.969	0.969
Num. obs.	808	808	808	808	808	808
RMSE	0.508	0.508	0.508	0.506	0.506	0.508

8.2 Variations on the firm-level data analysis

Tables 21 and 22 show the out-of-sample predictive accuracy of several machine-learning models. The results are as before: political models do not show marked improvement over apolitical models. Readers will also note that, in a few cases, prediction accuracy goes *down* when we increase the number of predictors. This counter-intuitive result should not be over-interpreted, since the difference between apolitical and political models is never substantively meaningful. Moreover, it is not uncommon for richer models to underperform parsimonious ones in out-of-sample tests (due to overfitting). This phenomenon is precisely what motivates the development of regularized models in the seminal work of Tibshirani (1996).

Table 21: Out-of-sample prediction accuracy (%) of political and apolitical models, using various samples and models (1 of 2).

Sample	Classifier	Apolitical	Political
Baseline	Ada Boost	87.0	88.4
	Bagging	90.3	90.5
	Extra Trees	89.7	90.2
	Logistic	83.0	84.5
	Random Forest	90.6	90.9
Alt. Distance	Ada Boost	87.5	87.9
	Bagging	90.0	90.1
	Extra Trees	90.0	90.3
	Logistic	82.9	84.1
	Random Forest	90.7	90.8
High income	Ada Boost	82.2	83.8
	Bagging	86.8	86.9
	Extra Trees	86.1	86.5
	Logistic	76.4	78.8
	Random Forest	87.0	87.2
Host characteristics	Ada Boost	87.9	88.3
	Bagging	90.5	90.5
	Extra Trees	90.3	90.2
	Logistic	80.0	84.9
	Random Forest	90.9	90.9
Low income	Ada Boost	90.3	90.9
	Bagging	91.8	91.9
	Extra Trees	91.4	91.7
	Logistic	85.6	88.4
	Random Forest	92.2	92.4

Table 22: Out-of-sample prediction accuracy (%) of political and apolitical models, using various samples and models (2 of 2).

Sample	Classifier	Apolitical	Political
Many predictors	Ada Boost	85.4	86.0
	Bagging	88.3	88.4
	Extra Trees	87.7	88.0
	Logistic	77.5	84.0
	Random Forest	88.6	88.7
No distance	Ada Boost	83.6	87.3
	Bagging	88.3	88.5
	Extra Trees	87.1	88.4
	Logistic	69.6	82.5
	Random Forest	88.1	88.7
Outliers	Ada Boost	74.3	74.9
	Bagging	78.5	78.5
	Extra Trees	78.3	78.3
	Logistic	66.6	73.0
	Random Forest	79.0	79.0
Stratified	Ada Boost	87.0	88.4
	Bagging	90.3	90.5
	Extra Trees	89.7	90.2
	Logistic	83.0	84.5
	Random Forest	90.6	90.9
Tax havens	Ada Boost	88.1	88.8
	Bagging	90.6	90.7
	Extra Trees	90.2	90.5
	Logistic	83.6	85.2
	Random Forest	90.9	91.1

Table 23: Random forest models using different tuning parameters.

Criterion	Max features	N estimators	Minimum samples per leaf	Apolitical	Political
gini	5	10	1	90.26	90.59
gini	5	10	50	88.58	89.26
gini	5	50	1	90.64	90.84
gini	5	50	50	88.67	89.55
gini	5	100	1	90.68	90.89
gini	5	100	50	88.66	89.54
gini	20	10	1	90.45	90.61
gini	20	10	50	90.39	90.44
gini	20	50	1	90.76	90.88
gini	20	50	50	90.47	90.55
gini	20	100	1	90.82	90.94
gini	20	100	50	90.48	90.56
gini	auto	10	1	90.26	90.55
gini	auto	10	50	88.58	89.61
gini	auto	50	1	90.64	90.86
gini	auto	50	50	88.67	89.74
gini	auto	100	1	90.68	90.90
gini	auto	100	50	88.66	89.73
entropy	5	10	1	90.30	90.57
entropy	5	10	50	88.43	89.29
entropy	5	50	1	90.61	90.84
entropy	5	50	50	88.74	89.46
entropy	5	100	1	90.68	90.89
entropy	5	100	50	88.65	89.54
entropy	20	10	1	90.58	90.73
entropy	20	10	50	90.37	90.55
entropy	20	50	1	90.88	91.01
entropy	20	50	50	90.49	90.60
entropy	20	100	1	90.93	91.03
entropy	20	100	50	90.47	90.59
entropy	auto	10	1	90.30	90.56
entropy	auto	10	50	88.43	89.67
entropy	auto	50	1	90.61	90.89
entropy	auto	50	50	88.74	89.75
entropy	auto	100	1	90.68	90.92
entropy	auto	100	50	88.65	89.75

Table 24: Out-of-sample prediction accuracy (%) of political and apolitical logit models. Thin models controls for distance and ln(GDP). Thick models control for distance, ln(GDP), ln(GDP per capita), share of urban population, and number of telephone lines per 1000 people. The samples for thick and thin models are slightly different due to data availability.

Model	Thin	Thick
Apolitical	77.1	78.2
Bureaucracy quality	77.1	78.2
Corruption (PRS)	78.1	78.2
Ethnic tensions	76.9	78.3
External conflict	77.5	78.6
Government stability	77.1	78.3
Internal conflict	78.0	78.6
Investment profile	77.6	77.4
Law and order	77.6	78.2
Military in politics	77.5	78.0
Religion in politics	77.8	78.8

Figure 4: Predictive accuracy of four algorithms. The dotted line indicates equal performance by the x-axis and the y-axis classifiers.

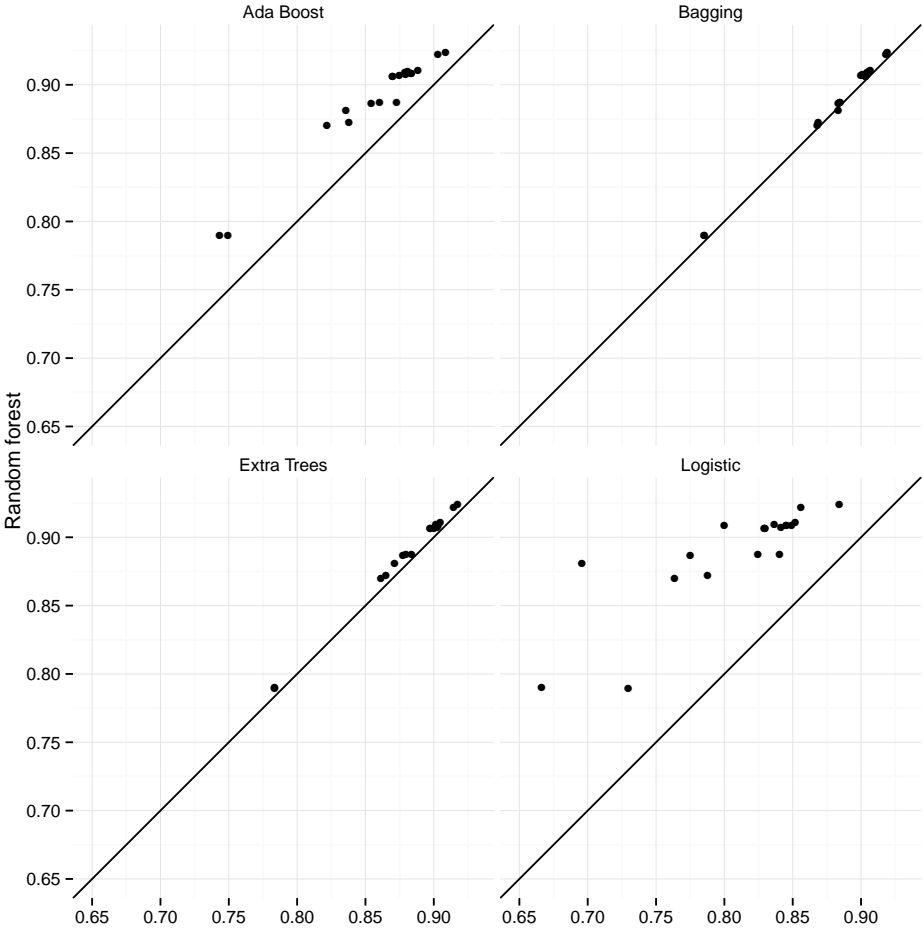
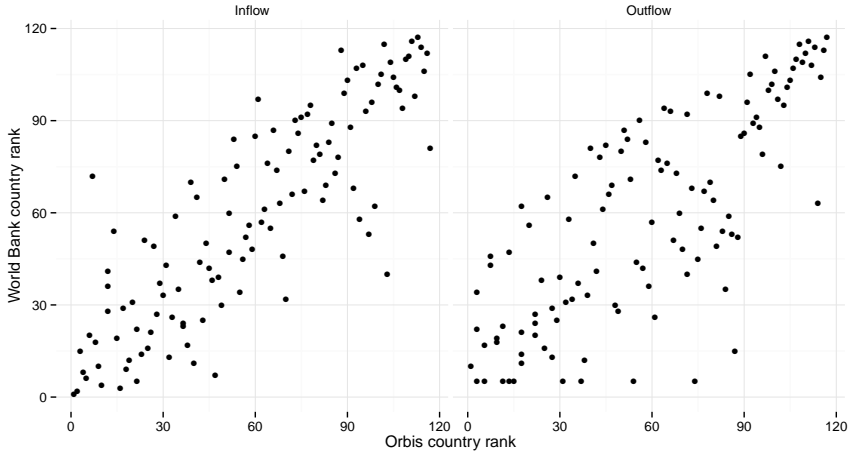


Figure 5: Ranks of countries in the Orbis and World Bank FDI datasets.



9 REFERENCES

- Achen, Christopher H. 1990. "What Does 'Explained Variance' Explain?: Reply." *Political Analysis* 2 (1): 173–84.
- Arulampalam, Wiji, Michael P. Devereux, and Giorgia Maffini. 2012. "The Direct Incidence of Corporate Income Tax on Wages." *European Economic Review* 56 (6): 1038–54.
- Berry, Heather, Mauro F Guillén, and Nan Zhou. 2010. "An Institutional Approach to Cross-National Distance." *Journal of International Business Studies* 41 (9): 1460–80.
- Bhaumik, Sumon Kumar, Nigel Driffield, and Sarmistha Pal. 2010. "Does Ownership Structure of Emerging-Market Firms Affect Their Outward FDI? The Case of the Indian Automotive and Pharmaceutical Sectors." *Journal of International Business Studies* 41 (3): 437–50.
- Biglaiser, Glen, and Karl R. DeRouen. 2006. "Economic Reforms and Inflows of Foreign Direct Investment in Latin America." *Latin American Research Review* 41 (1): 51–75.
- Blonigen, Bruce A., and Jeremy Piger. 2014. "Determinants of Foreign Direct Investment." *Canadian Journal of Economics* 47 (3).
- Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. "The Organization of Firms Across Countries*." *The Quarterly Journal of Economics* 127 (4): 1663–1705.
- Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (1): 63–82.
- Breiman, Leo. 2001. "Statistical Modeling: The Two Cultures." *Statistical Science* 16 (3): 199–215.
- Breiman, Leo, Jerome Friedman, Charles J. Stone, and Richard A. Olshen. 1984. *Classification and Regression Trees*. CRC press.
- Bureau van Dijk. 2013. "Orbis: Company Information Around the Globe."
- Busse, Matthias, and Carsten Hefeker. 2007. "Political Risk, Institutions and Foreign Direct Investment." *European Journal of Political Economy* 23 (2): 397–415.
- Büthe, Tim, and Helen V. Milner. 2014. "Foreign Direct Investment and Institutional Diversity in Trade Agreements: Credibility, Commitment, and Economic Flows in the Developing World, 1971–2007." *World Politics* 66 (01): 88–122.
- Caprio, Lorenzo, Mara Faccio, and John J. McConnell. 2013. "Sheltering Corporate Assets from Political Extraction." *Journal of Law, Economics, and Organization* 29 (2): 332–54.
- Chakrabarti, Avik. 2001. "The Determinants of Foreign Direct Investments: Sensitivity Analyses of Cross-Country Regressions." *Kyklos* 54 (1): 89–114.
- Chen, Victor Zitian, Jing Li, and Daniel M. Shapiro. 2012. "International Reverse Spillover Effects on Parent Firms: Evidences from Emerging-Market MNEs in Developed Markets." *European Management Journal, Multinational Companies from Emerging Markets and Strategic Fit* 30 (3): 204–18.
- Cormen, Thomas H, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. 2009. *Introduction to Algorithms*. Cambridge, Massachusetts; London: The MIT Press.
- Dietterich, Thomas G. 2000. "Ensemble Methods in Machine Learning." In 1–15. Lecture Notes in Computer Science. Springer Berlin Heidelberg.

- Egger, Peter, and Hannes Winner. 2005. "Evidence on Corruption as an Incentive for Foreign Direct Investment." *European Journal of Political Economy* 21 (4): 932–52.
- Eicher, Theo S., Lindy Helfman, and Alex Lenkoski. 2012. "Robust FDI Determinants: Bayesian Model Averaging in the Presence of Selection Bias." *Journal of Macroeconomics* 34 (3): 637–51.
- Franzese, Robert, and Cindy Kam. 2009. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. University of Michigan Press.
- Fredriksson, Per G., John A. List, and Daniel L. Millimet. 2003. "Bureaucratic Corruption, Environmental Policy and Inbound US FDI: theory and Evidence." *Journal of Public Economics* 87 (7–8): 1407–30.
- Freund, Yoav, Robert E Schapire, and others. 1996. "Experiments with a New Boosting Algorithm." In *ICML*, 96:148–56.
- Garland, Marshall W., and Glen Biglaiser. 2008. "Do Electoral Rules Matter? Political Institutions and Foreign Direct Investment in Latin America." *Comparative Political Studies*, November.
- Garriga, Ana Carolina. 2015. "Human Rights Regimes, Reputation, and Foreign Direct Investment." *International Studies Quarterly*.
- Gerber, Alan, and Neil Malhotra. 2008. "Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals." *Quarterly Journal of Political Science* 3 (3): 313–26.
- Geurts, Pierre, Damien Ernst, and Louis Wehenkel. 2006. "Extremely Randomized Trees." *Machine Learning* 63 (1): 3–42.
- Gravelle, J G. 2013. "Tax Havens: International Tax Avoidance and Evasion." *Congressional Research Service*.
- Greene, William H. 2008. *Econometric Analysis, 6th Edition*. Pearson Prentice-Hall.
- Halvorsen, Thomas, and Jo Jakobsen. 2013. "Democrats, Republicans—or Both? An Empirical Analysis of the Effects of the Composition of State Governments on FDI, 1977–2004." *International Interactions* 39 (2): 167–91.
- Hansen, Lars Kai, and Peter Salamon. 1990. "Neural Network Ensembles." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12: 993–1001.
- Hartman, David G. 1984. "Tax Policy and Foreign Direct Investment in the United States." *National Tax Journal* 37 (4): 475–87.
- . 1985. "Tax Policy and Foreign Direct Investment." *Journal of Public Economics* 26 (1): 107–21.
- Hastie, Trevor, Robert Tibshirani, Jerome Friedman, and James Franklin. 2013. *The Elements of Statistical Learning: data Mining, Inference and Prediction*. 11th print.
- Henisz, W. J. 2000. "The Institutional Environment for Multinational Investment." *Journal of Law, Economics, and Organization* 16 (2): 334–64.
- Henisz, W. J., and Oliver E. Williamson. 1999. "Comparative Economic Organization—Within and Between Countries." *Business & Politics* 1 (3): 261.
- Hines, James R., Jr., and Eric M. Rice. 1994. "Fiscal Paradise: Foreign Tax Havens and American Business." *The Quarterly Journal of Economics* 109 (1): 149–82.

Honaker, James, and Gary King. 2010. "What to Do About Missing Values in Time-Series Cross-Section Data." *American Journal of Political Science* 54 (2): 561–81.

Huntington, Samuel P. 1968. *Political Order in Changing Societies*. New Haven Connecticut: Yale University Press.

Jensen, Nathan M. 2003. "Democratic Governance and Multinational Corporations: Political Regimes and Inflows of Foreign Direct Investment." *International Organization* 57 (03): 587–616.

Jensen, Nathan M., Glen Biglaiser, Quan Li, Edmund Malesky, Pablo M. Pinto, Santiago M. Pinto, and Joseph L. Staats. 2012. *Politics and Foreign Direct Investment*. University of Michigan Press.

Kerner, Andrew. 2009. "Why Should I Believe You? The Costs and Consequences of Bilateral Investment Treaties." *International Studies Quarterly* 53 (1): 73–102.

———. 2014. "Can Foreign Stock Investors Influence Policymaking?" *Comparative Political Studies*, 0010414013519411.

Kerner, Andrew, and Jane Lawrence. 2014. "What's the Risk? Bilateral Investment Treaties, Political Risk and Fixed Capital Accumulation." *British Journal of Political Science* 44 (01): 107–21.

Kim, Moonhawk, Amy H. Liu, Kim-Lee Tuxhorn, David S. Brown, and David Leblang. 2015. "Lingua Mercatoria: Language and Foreign Direct Investment." *International Studies Quarterly* 59 (2): 330–43.

King, Gary. 1990. "Stochastic Variation: A Comment on Lewis-Beck and Skalaban's "The R-Squared." *Political Analysis* 2 (January): 185–200.

King, Gary, and Langche Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis* 9 (2): 137–63.

Kobrin, Stephen J. 1979. "Political Risk: A Review and Reconsideration." *Journal of International Business Studies* 10 (1): 67–80.

Lampel, Joseph, and Claudio Giachetti. 2013. "International Diversification of Manufacturing Operations: Performance Implications and Moderating Forces." *Journal of Operations Management* 31 (4): 213–27.

Lewis-Beck, Michael S., and Andrew Skalaban. 1990. "The R-Squared: Some Straight Talk." *Political Analysis* 2 (January): 153–71.

Li, Quan. 2015. "Fiscal Decentralization and Tax Incentives in the Developing World." *Review of International Political Economy* 0 (0): 1–32.

Li, Quan, and Adam Resnick. 2003. "Reversal of Fortunes: Democratic Institutions and Foreign Direct Investment Inflows to Developing Countries." *International Organization* 57 (1): 175–211.

Louppe, Gilles, Louis Wehenkel, Antonio Sutera, and Pierre Geurts. 2013. "Understanding Variable Importances in Forests of Randomized Trees." In *Advances in Neural Information Processing Systems*, 431–39. <http://papers.nips.cc/paper/4928-understanding-variable-importances-in-forests-of-randomized-trees>

Lui, Francis T. 1985. "An Equilibrium Queuing Model of Bribery." *Journal of Political Economy* 93 (4): 760–81.

Maffini, Giorgia, and Socrates Mokkalas. 2011. "Profit Shifting and Measured Productivity of Multinational Firms*." *Oxford Bulletin of Economics and Statistics* 73 (1): 1–20.

- Majeske, Karl D., Terri Lynch-Caris, and Janet Brelin-Fornari. 2010. "Quantifying R² Bias in the Presence of Measurement Error." *Journal of Applied Statistics* 37 (4): 667–77.
- Malesky, Edmund. 2008. "Straight Ahead on Red: How Foreign Direct Investment Empowers Subnational Leaders." *The Journal of Politics* 70 (01): 97–119.
- Manger, Mark. 2005. "Competition and Bilateralism in Trade Policy: the Case of Japan's Free Trade Agreements." *Review of International Political Economy* 12 (5): 804–28.
- . 2009. *Investing in Protection: The Politics of Preferential Trade Agreements Between North and South*. Cambridge University Press.
- Marshall, Monty G, Keith Jagers, and Ted Robert Gurr. 2012. *Polity IV Project*.
- Martins, Pedro S., and Yong Yang. 2014. "Globalized Labour Markets? International Rent Sharing Across 47 Countries." *British Journal of Industrial Relations*, February, n/a–/a.
- Mayer, Thierry, and Soledad Zignago. 2006. "GeoDist: the CEPII Distances and Geographical Database."
- Mooij, Ruud A. de, and Sjeff Ederveen. 2008. "Corporate Tax Elasticities: a Reader's Guide to Empirical Findings." *Oxford Review of Economic Policy* 24 (4): 680–97.
- Moosa, Imad A., and Buly A. Cardak. 2006. "The Determinants of Foreign Direct Investment: An Extreme Bounds Analysis." *Journal of Multinational Financial Management* 16 (2): 199–211.
- Mosley, Layna. 2003. *Global Capital and National Governments*. Cambridge University Press.
- Nunn, Nathan, and Daniel Trefler. 2013. "Incomplete Contracts and the Boundaries of the Multinational Firm." *Journal of Economic Behavior & Organization* 94 (October): 330–44.
- Oneal, John R. 1994. "The Affinity of Foreign Investors for Authoritarian Regimes." *Political Research Quarterly* 47 (3): 565–88.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30.
- Pinto, Pablo M. 2013. *Partisan Investment in the Global Economy: why the Left Loves Foreign Direct Investment and FDI Loves the Left*. Cambridge; New York: Cambridge University Press.
- Pinto, Pablo M., and Santiago M. Pinto. 2008. "The Politics of Investment Partisanship: And the Sectoral Allocation of Foreign Direct Investment." *Economics & Politics* 20 (2): 216–54.
- Prentice, R. L., and R. Pyke. 1979. "Logistic Disease Incidence Models and Case-Control Studies." *Biometrika* 66 (3): 403–11.
- Slemrod, Joel B. 1990. "Tax Effects on Foreign Direct Investment in the United States: Evidence from a Cross-Country Comparison." In 79–122. University of Chicago Press, 1990.
- Staats, Joseph L., and Glen Biglaiser. 2012. "Foreign Direct Investment in Latin America: The Importance of Judicial Strength and Rule of Law¹." *International Studies Quarterly* 56 (1): 193–202.
- Sterling, Theodore D. 1959. "Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance—Or Vice Versa." *Journal of the American Statistical Association* 54 (285): 30–34.
- Tanzi, Vito. 1998. "Corruption Around the World: Causes, Consequences, Scope, and Cures." *Staff Papers (International Monetary Fund)* 45 (4): 559–94.

- Temouri, Yama, and Nigel L Driffield. 2009. "Does German Foreign Direct Investment Lead to Job Losses at Home?" *Applied Economics Quarterly* 55 (3): 243–63.
- Teorell, Jan, Nicholas Charron, Stefan Dahlberg, Sören Homberg, Bo Rothstein, Petrus Sundin, and Richard Svensson. 2013. "The Quality of Governance Dataset." University of Gothenburg: The Quality of Governance Institute.
- Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–88.
- Vernon, Raymond. 1971. *Sovereignty at Bay*. Basic Books.
- Voget, Johannes. 2011. "Relocation of Headquarters and International Taxation." *Journal of Public Economics* 95 (9–10): 1067–81.
- Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. 2010. "The Perils of Policy by P-Value: Predicting Civil Conflicts." *Journal of Peace Research* 47 (4): 363–75.
- Wellhausen, Rachel L. 2015. "Bondholders Vs. Direct Investors? Competing Responses to Expropriation." *International Studies Quarterly*, n/a–/a.
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2 edition. The MIT Press.
- World Bank. 2014. *World Development Indicators*.