WP/20/278

IMF Working Paper

Identifying Reform Priorities: The Role of Non-linearities

by Klaus-Peter Hellwig

INTERNATIONAL MONETARY FUND

 WP/20/278

IMF Working Paper

Asia Pacific Department

Identifying Reform Priorities: The Role of Non-linearities

Prepared by Klaus-Peter Hellwig

Authorized for distribution by Rahul Anand

December 2020

Abstract

Can countries improve their business climate through reforms in specific policy areas? Kraay and Tawara (2013) find that the answer depends on how we measure the business climate. When regressing seven different business climate indices on 38 policy indicators, they find little agreement among the seven models as to which of those policy indicators matter most. I revisit this puzzle using the same data but replacing their linear models with a Random Forest algorithm. I find a strong consensus across models on the importance ranking of policy indicators: No matter which business climate index is considered, the top ten contributors to a better business climate always include high recovery rates in insolvency proceedings (i.e., cents on the dollar for creditors), shorter border formalities for both importers and exporters, and low costs for starting a business. I show that the marginal effect of reforms is heterogeneous across countries and document how reform priorities depend on country specific circumstances.

JEL Classification Numbers: O43

Keywords: Structural reforms; Random Forest; Business climate; Shapley values

Author's E-Mail Address: khellwig@imf.org

Identifying Reform Priorities: The Role of Non-linearities

Klaus-Peter Hellwig*†

November 16, 2020

Abstract

Can countries improve their business climate through reforms in specific policy areas? Kraay and Tawara (2013) find that the answer depends on how we measure the business climate. When regressing seven different business climate indices on 38 policy indicators, they find little agreement among the seven models as to which of those policy indicators matter most. I revisit this puzzle using the same data but replacing their linear models with a Random Forest algorithm. I find a strong consensus across models on the importance ranking of policy indicators: No matter which business climate index is considered, the top ten contributors to a better business climate always include high recovery rates in insolvency proceedings (i.e., cents on the dollar for creditors), shorter border formalities for both importers and exporters, and low costs for starting a business. I show that the marginal effect of reforms is heterogeneous across countries and document how reform priorities depend on country specific circumstances.

JEL codes: O43

Keywords: Structural reforms; Random forest; Business climate; Shapley values

*khellwig@imf.org.

†**The views expressed in this paper are those of the author and do not necessarily represent the views of the IMF, the IMF Executive Board, or IMF Management.**

1 Introduction[1]

The idea that governments can foster long-term growth by improving the business climate is widely accepted among economists and policymakers. Hence, the general advice to become more business friendly is a staple of many country reports and has been taken up in many development plans.[2] However, in most countries, the list of potential areas for improvement is long – too long to do everything at once. For example, cutting red tape, simplifying the tax code, facilitating the enforcement of contracts are prescriptions that few economists would disagree with. But the literature has yet to reach consensus on which of these areas, if any, would deliver the largest improvements in the business environment and should therefore be at the top of a government's reform agenda.

The empiricist's approach to this question is to take an index of business friendliness and search for policy variables that can best explain variations in the index. Kraay and Tawara (2013) – henceforth KT – construct seven such business climate indices using expert assessments and survey responses from seven different sources. Remarkably, even though all seven indices try to measure broadly similar concepts and are highly correlated with each other, they offer very diverging conclusions about which reform areas should be prioritized. Specifically, when regressing each of the seven indices on the 38 components of the World Bank's Doing Business (DB) ranking, KT find that the relative importance of each DB indicator for explaining variations in the business climate is highly sensitive to the choice of left-hand side variable – too sensitive to offer robust guidance for policy.

In this paper, I revisit KT's findings using the same data but exploring alternative functional forms for the regression model. Their analysis assumes that perceptions of the business climate are a linear function of the individual DB indicators. This could be a strong assumption, since it implies that reform priorities should be the same across countries, irrespective of country specific circumstances. Therefore, I replace KT's linear Bayesian Model Averaging (BMA) algorithm with a Random Forest (Breiman, 2001) algorithm. I thereby retain the idea of averaging over many different models but

[1]I thank Paolo Mauro for introducing me to the Kraay and Tawara (2013) puzzle, Aart Kraay for sharing his data, and Romain Duval, Giulio Lisi, Kevin Wiseman, and seminar participants at the IMF for helpful comments.

[2]For example, recent reports advising country authorities to improve the business environment include the World Bank's June 2020 Global Economic Prospects or the IMF's October 2020 Regional Economic Outlooks for Sub-saharan Africa and for the Middle East and Central Asia Region.

impose a regression tree structure rather than a linear structure on each constituent model. Regression tree algorithms are designed to search for threshold effects and can also capture complex interactions between variables. If interaction terms matter, then policy makers need to pay increased attention to the proper sequencing of reforms. Of course, Random Forest (RF) is not the only tree-based method, let alone the only way to explore interactions and other non-linearities.[3] The choice of RF over other methods is motivated by its similarity to BMA in that it combines model averaging with shrinkage to obtain robust model estimates.

My main result is that, when using RF instead of BMA, the importance ranking of DB indicators becomes less sensitive to the choice of the left-hand side variable. Four indicators are among the top ten irrespective of the left-hand side variable. By contrast, when using BMA, not a single variable is in the top 10 for all left-hand side variables. The broader consensus under RF also applies at the lower end of the ranking. 21 DB variables are never in the top 10 (as opposed to 12 under BMA).

The paper is organized as follows: Section 2 estimates the BMA and RF regressions and compares the stability of variable importance across outcome variables for the two methods. Section 3 explores the nature of the relevant non-linearities in the RF and discusses policy implications for identifying reform priorities. Section 4 concludes.

2 Stability of reform priorities

The basic assumption underlying KT is that a country's perceived overall quality of the business environment can be explained by the country's performance in a number of specific and more disaggregated aspects of the business environment. Specifically, they assume that, in a cross-section of countries, country c's business climate $y_{c,i}$, as perceived by some observer i, can be explained as a function of a vector $\mathbf{x_c}$ of that country's 38 indicators in the World Bank's DB study and some unobservable disturbance $\epsilon_{c,i}$:

$$y_{c,i} = f_i(\mathbf{x_c}) + \epsilon_{c,i}$$

The DB indicators quantify the experience of a hypothetical business in eleven different broader areas. Rather than asking survey participants to assign subjective scores, performance is measured in observable quantities (e.g., the number of tax payments

[3] For example, IMF (2019) uses Bayesian Hierarchical Analysis to explore the role of interactions in the relationship between structural reforms and growth.

the hypothetical business needs to make or the number of documents it needs to fill out to export its products). These quantities can be, directly or indirectly, influenced by policy makers. Hence, knowledge of the function f_i would allow policymakers to identify reforms that would improve observer i's perception of the business climate. It is important to note that the mapping f_i is not an economic relationship or process. It merely translates the situation faced by businesses on the ground into a one-dimensional score. Hence, as pointed out by KT, establishing causality is less challenging than in other contexts where the dependent variable is an economic outcome.

KT estimate the function f_i seven times, each time using a different outcome variable y_i.[4] They construct the seven outcome variables by using information from seven different sources: Economist Intelligence Unit, Political Risk Services, Global Insight Global Risk Service, Global Insight Business Risk Conditions, Cerberus Corporate Intelligence Gray Area Dynamics, Global Competitiveness Report, and World Bank Country Policy and Institutional Assessment.[5]

Bayesian Model Averaging (BMA)

KT assume that each observer i applies a linear function: $y_{c,i} = \mathbf{x_c}'\beta_\mathbf{i} + \epsilon_{c,i}$. They use a BMA approach which allows them to be agnostic ex ante about which of the 38 DB variables should have a non-zero slope coefficient. Instead of estimating a single model, BMA considers the fit of all candidate models (i.e., all possible combinations of regressors) and, based on model fit, computes posterior probabilities that a certain variable is included in observer i's true model.[6] Loosely speaking, if models that include variable x^j tend to have a better fit than models without x^j, then x^j is likely to be part of the true model from which the data were generated. KT use this posterior inclusion probability (PIP) as their measure of a variable's importance.

Table 1 reports the PIPs for each variable in a series of BMA regressions, replicating KT's results. Each column represents a different outcome variable. Note that results are only reported for six outcome variables, due to the confidentiality of the CPIA data set. In each column, the cells with the top 10 DB indicators in terms of PIP are shaded green. KT's main result is that there is no DB indicator that is among the top 10 in

[4]The estimations are done on a cross-section of countries using data for 2009. Country coverage varies across outcome variables (See Annex A).

[5]See Kraay and Tawara (2013) for details on the construction of outcome variables.

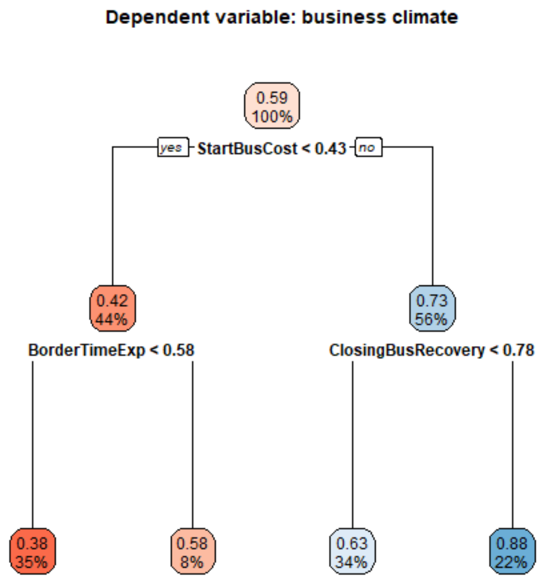
[6]KT impose a prior that the number of variables in the true model is 10, so that parsimonious models are favored ex ante.

Table 1: Bayesian Model Averaging: Posterior Inclusion Probabilities

	DRI	EIU	GAD	GCS	PRS	WMO
StartBusProc	4.0	26.6	37.2	42.7	15.5	4.3
StartBusTime	8.3	4.2	35.9	27.5	4.5	7.1
StartBusCost	9.2	13.2	11.1	5.6	50.6	83.7
StartBusCapital	10.2	5.7	3.9	6.0	3.7	4.5
ConstPermProc	24.1	5.6	6.9	4.2	5.2	4.1
ConstPermTime	63.9	5.7	4.4	66.5	15.2	3.6
ConstPermCost	4.0	33.4	14.0	21.6	23.4	97.6
EmployHiringDiff	15.3	4.5	4.6	10.2	4.4	7.1
EmployHoursRig	4.5	6.9	7.6	19.0	4.6	9.2
EmployFiringDiff	12.9	24.3	3.6	3.6	4.8	5.0
EmployFiringCost	84.0	29.9	12.1	11.0	70.2	47.0
RegPropertyProc	3.9	5.3	13.0	77.1	3.3	10.8
RegPropertyTime	4.5	3.7	12.7	6.3	4.6	6.8
RegPropertyCost	4.3	4.8	3.6	4.7	5.0	4.7
CreditLegal	93.1	9.3	100.0	31.1	11.1	97.9
CreditInfo	4.0	11.8	6.0	20.8	10.5	3.7
CreditPubReg	4.3	3.6	5.0	3.6	4.0	3.2
CreditPrivBureau	4.6	15.3	40.9	5.6	3.9	10.0
InvestorDisclosure	4.2	3.3	4.2	6.5	3.5	4.1
InvestorLiability	4.9	3.8	4.0	4.7	7.1	3.0
InvestorShareholder	5.2	4.5	4.0	11.2	28.8	4.3
TaxPayment	3.8	59.1	96.6	7.8	53.2	95.4
TaxTime	4.6	20.3	4.6	98.8	20.7	15.3
TaxProfit	67.7	5.0	4.5	6.9	5.1	4.6
TaxLabor	9.0	4.0	44.7	5.1	4.4	10.9
TaxOther	55.0	8.5	4.0	3.9	29.0	5.3
BorderDocExp	13.5	43.3	16.5	3.6	9.8	92.8
BorderTimeExp	16.5	9.3	47.6	11.3	11.6	93.3
BorderCostExp	27.5	10.7	6.0	73.4	10.0	21.0
BorderDocImp	5.6	5.5	33.7	4.1	4.2	19.5
BorderTimeImp	71.9	73.9	18.3	6.1	52.2	9.3
BorderCostImp	66.1	50.7	4.7	29.6	17.7	20.0
ContractProc	8.3	40.1	17.0	11.3	5.6	4.3
ContractTime	90.0	4.6	6.0	6.7	9.9	4.2
ContractCost	3.7	13.0	3.8	3.8	5.2	7.7
ClosingBusTime	3.8	31.1	4.4	3.9	13.5	12.4
ClosingBusCost	4.3	10.8	7.0	6.1	4.0	6.2
ClosingBusRecovery	3.3	62.0	99.8	99.2	62.6	90.1

Notes: Numbers indicate variable importance in terms of Posterior Inclusion Probabilities in percent. Shaded cells indicate top 10 DB indicators. Columns represent BMA regressions with different outcome variables, each constructed from a different source: DRI=Global Insight Global Risk Service, EIU=Economist Intelligence Unit, GAD=Cerberus Corporate Intelligence Gray Area Dynamics, GCS=Global Competitiveness Report, PRS=Political Risk Services, WMO=Global Insight Business Risk Conditions. See Kraay and Tawara (2013) for BMA parameter settings. See Annex C for variable descriptions.

Figure 1: Regression tree example



Notes: Dependent variable is the business climate index constructed from Global Insight Business Risk Conditions (WMO); N=174 observations; Explanatory variables: 38 DB indicators, rescaled to $[0, 1]$; higher DB scores indicate better performance. See Annex C for variable descriptions.

all columns.

Random Forest (RF)

I now consider an alternative functional form. Instead of assuming that the subject expert i who assesses a country's business climate $y_{c,i}$ follows a linear decision rule, I assume that they follow a tree-shaped decision rule. Decision trees sequentially split the sample into buckets with high and low business climate scores, depending on whether some variable is above or below its critical threshold value.

To give an example, Figure 1 estimates a short decision tree using the Breiman et al. (1984) regression tree algorithm. This algorithm begins with the full sample and searches for the variable and threshold that deliver the best separation between high and low business climate scores. In this example, the winning variable is the DB score measuring the cost of starting a business, and the optimal threshold is 0.43. The algorithm then takes the two sub-samples and looks for the next best split. For observations with a low score for startup cost (i.e., the left branch), the sample is split using the time to clear export formalities; and for observations with a high score at the first split, the sample is split by the recovery rates (cents on the dollar for creditors) in insolvency procedures.

From the example it is easy to see how regression trees naturally capture interactions between variables: Creditors' recovery rates are used only when startup costs are low (in the right branch), while the length of border formalities matters only when startup costs are high. Moreover, it allows variables to enter the decision rule in a non-linear way. For example, once a country is performing well enough in a certain area (e.g., time to clear customs), additional improvements in that area may deliver only relatively small gains in the overall perception of the business climate. Similarly, an improvement from extremely poor performance to very poor performance in a particular area may not help the overall perception, if the expert applies some minimum performance standard rather than a linear rule.

A main drawback of the decision tree algorithm is that uncovering the "right" tree is extremely difficult. Finding the globally optimal tree – one that best explains the data with a minimum number of splits and branches – requires exploring a potentially infinite number of combinations and is therefore de facto impossible. Therefore, decision tree algorithms used in practice look for locally, instead of globally, optimal paths. The resulting trees can be highly sensitive to small changes in the estimation sample, since changes in the variable or threshold applied at the top of the tree tend to affect the entire structure of the tree.

To overcome the lack of robustness of individual trees, Breiman (2001) proposes the RF approach. Rather than relying on a single tree, RF estimates a large number of trees, each estimated on a different sample and using a different set of variables. The heterogeneity in samples is achieved by taking random draws with replacement (i.e., bootstrapping) from the original sample. The heterogeneity in variables is achieved by limiting the set of variables available to the algorithm to a small number m_{try} of variables drawn randomly from the full set of variables. This random subset is redrawn at each split. On the one hand, this randomization implies that individual trees are poor representations of the data generating process. However, the diversity among the trees implies that model errors tend to offset each other. Hence, if the number of trees is sufficiently large, RF models can achieve predictive performance that often outperforms linear models.

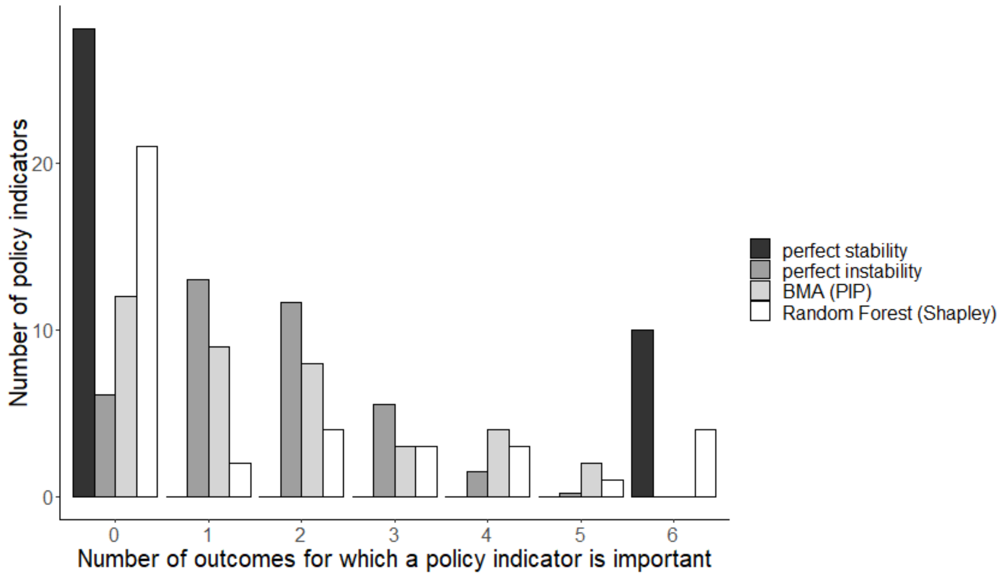
Table 2 reports the importance of each variable in the RF models estimated for the six outcome variables of interest. The number of trees is set to 5000 and the pa-

Table 2: Random Forest: Shapley Values

	DRI	EIU	GAD	GCS	PRS	WMO
StartBusProc	0.1	0.5	0.5	0.3	0.5	0.3
StartBusTime	0.2	0.4	0.8	0.6	0.6	0.5
StartBusCost	0.5	1.2	1.7	0.7	1.8	3.6
StartBusCapital	0.1	0.1	0.4	0.2	0.2	0.1
ConstPermProc	0.1	0.1	0.1	0.1	0.2	0.1
ConstPermTime	0.4	0.3	0.1	0.4	0.8	0.2
ConstPermCost	0.2	2.0	1.3	0.7	1.8	2.7
EmployHiringDiff	0.0	0.0	0.1	0.2	0.3	0.0
EmployHoursRig	0.1	0.0	0.1	0.1	0.1	0.1
EmployFiringDiff	0.1	0.4	0.1	0.1	0.2	0.1
EmployFiringCost	0.4	0.3	0.2	0.1	0.6	0.4
RegPropertyProc	0.1	0.1	0.2	0.3	0.2	0.1
RegPropertyTime	0.1	0.2	0.4	0.2	0.2	0.2
RegPropertyCost	0.1	0.4	0.2	0.1	0.3	0.2
CreditLegal	0.3	0.2	1.3	0.4	0.4	0.7
CreditInfo	0.3	0.5	0.7	0.2	0.7	0.1
CreditPubReg	0.1	0.1	0.1	0.1	0.1	0.0
CreditPrivBureau	0.2	1.0	0.8	0.2	0.7	0.8
InvestorDisclosure	0.1	0.2	0.2	0.1	0.2	0.2
InvestorLiability	0.2	0.1	0.2	0.2	0.1	0.2
InvestorShareholder	0.1	0.1	0.3	0.1	0.3	0.2
TaxPayment	0.1	0.7	0.8	0.5	1.4	1.0
TaxTime	0.2	0.4	0.1	0.4	0.7	0.4
TaxProfit	0.2	0.1	0.1	0.1	0.1	0.1
TaxLabor	0.1	0.2	0.2	0.1	0.1	0.1
TaxOther	0.6	0.8	0.3	0.2	0.3	0.3
BorderDocExp	0.7	0.7	0.8	0.4	0.7	1.2
BorderTimeExp	1.0	1.1	1.3	0.7	0.9	3.0
BorderCostExp	0.7	0.7	0.2	0.6	0.3	0.1
BorderDocImp	0.2	0.3	0.6	0.3	0.3	0.4
BorderTimeImp	1.3	2.2	1.5	0.8	1.2	1.8
BorderCostImp	1.0	1.3	0.3	0.5	0.4	0.2
ContractProc	0.0	0.6	0.6	0.4	0.2	0.4
ContractTime	0.1	0.4	0.1	0.1	0.1	0.1
ContractCost	0.0	0.2	0.3	0.2	0.2	0.4
ClosingBusTime	0.1	0.9	1.1	0.5	0.5	0.6
ClosingBusCost	0.0	0.3	0.4	0.2	0.2	0.4
ClosingBusRecovery	0.4	2.6	1.9	0.9	1.9	3.1

Notes: Numbers indicate variable importance in terms of absolute Shapley values (averaged over all observations). Shaded cells indicate top 10 DB indicators. Columns represent RF regressions with different outcome variables, each constructed from a different source: DRI=Global Insight Global Risk Service, EIU=Economist Intelligence Unit, GAD=Cerberus Corporate Intelligence Gray Area Dynamics, GCS=Global Competitiveness Report, PRS=Political Risk Services, WMO=Global Insight Business Risk Conditions. See Annex A for RF hyper-parameter settings. See Annex C for variable descriptions.

Figure 2: Stability of variable importance across outcome variables



rameter m_{try} is obtained through cross-validation (see Annex A).[7] Variable importance is evaluated in terms of Shapley values.[8] In cooperative game theory, Shapley values are obtained by averaging the marginal values of a player in all possible combinations of other players. Shapley (1953) shows that these average marginal values yield the unique payoff distribution among players such that four axioms are satisfied. Strumbelj and Kononeko (2010) show that the same concept can be applied to prediction models, where an individual prediction takes the role of aggregate payoff and individual predictors take the role of players. For each observation this method yields a vector of Shapley values indicating the contributions of individual predictors in explaining why the model prediction for that observation deviates from the sample average. From these local variable importance measures, I compute a general importance measure (reported in Table 2) by averaging the absolute Shapley values over all countries.

A quick look at Table 2 reveals that the six RF models are in agreement on the relative importance of many of the DB indicators, irrespective of the dependent variable. Figure 2 applies the graphical method used in KT to summarize the stability of variable importance across outcome variables. For each DB indicator, the chart counts the number of dependent variables for which it is among the ten most important determinants. The resulting distribution is plotted as follows: Bucket n counts the number

[7]Each tree is grown exhaustively until no more splits are possible.
[8]An alternative importance measure is used in Annex B, with fairly similar results.

of DB indicators that are in the top 10 for n outcomes and outside the top 10 for 6-n outcomes. Perfect stability would imply that all six models agree on the top 10. There would be 10 indicators in the n=6 bucket, zero indicators in the n=5 bucket, and so forth. 28 indicators would be in the n=0 bucket. Perfect instability would be a situation in which variable importance is random, so that the expected distribution is that of a binomial distribution with success probability 10/38 and six trials. As in KT, the distribution derived from the BMA results is remarkably close to the "perfect instability" benchmark, with not a single DB indicator in the top 10 for all outcome variables. By comparison, the distribution of variable importance for RF is a lot closer to the "perfect stability" benchmark, with four DB indicators in the top 10.

The four indicators deemed important among all six RF models are (i) the recovery rates for creditors in insolvency cases, (ii) the time it takes exporters to complete border formalities, (iii) the time it takes importers to complete border formalities, and (iv) the cost of starting a business. The stability of RF models across outcome variables also manifests itself in the variables that are not deemed important. 21 variables are never among the top 10. These include labor and profit tax rates, employment regulations, protections for minority shareholders, and mechanisms to register property and enforce contracts.

However, it is important to keep in mind that Table 2 shows each variable's *average* importance across all observations. To identify reform priorities, we need to keep in mind that, due to non-linearities, the importance of a variable can vary substantially across observations in the sample. Moreover, when non-linearities are at work, a variable's high importance as a determinant of country *c*'s business climate score does not imply that the variable should be a reform priority for country *c*. This is because reform priorities should reflect *marginal* effects from changing a variable, not the average effect of that variable. Hence, unlike in linear models, identifying reform priorities requires looking at each country's unique circumstances. The next section will explore in more detail how priorities depend on country specific circumstances.

3 The role of non-linearities

The previous section has shown that relative importance of DB indicators is less sensitive to the choice of dependent variable if we use non-linear models. I now explore the nature of the relevant non-linearities in more detail. But, instead of looking at each business climate index y_i separately, I construct a new business climate index $\tilde{y}$ that combines the six measures from the previous section:[9]

$$\tilde{y}_c = \sum_i \frac{y_{c,i} - mean(y_i)}{std.dev.(y_i)}$$

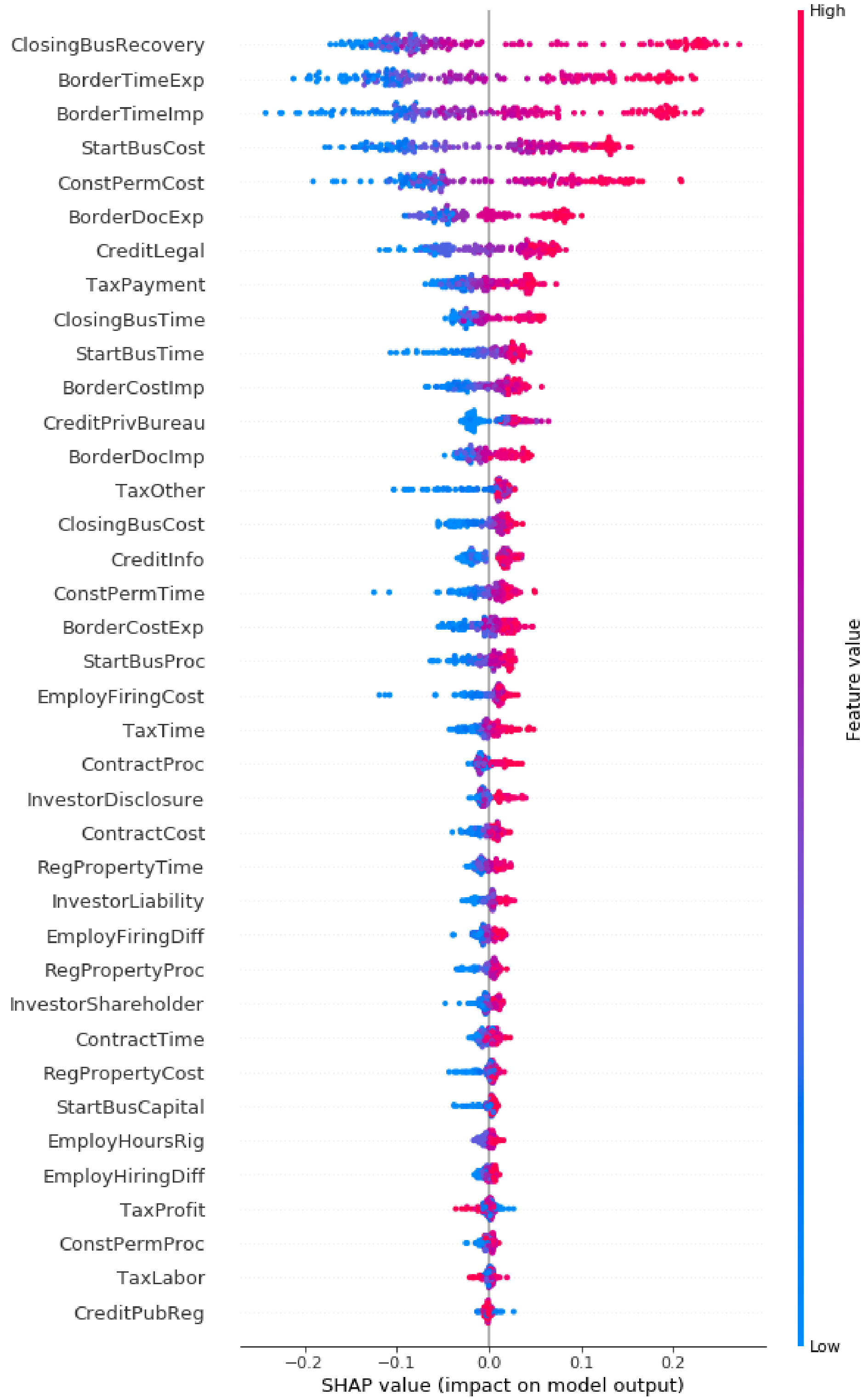
I re-estimate the RF with this new business climate index $\tilde{y}$ as the left-hand side variable. Figure 3 shows the variable importance for each DB indicator and each observation.[10] The Shapley value $s_{c,j}$ for country c and indicator j measures the contribution of indicator j in explaining the deviation of $\tilde{y}_c$ from the sample average. As in the previous section, the recovery rate for creditors in bankruptcy procedures, and the time to export or import are the top three indicators. Countries with high scores (plotted in red) in these areas tend to have a better overall score $\tilde{y}_c$ than countries with low scores (plotted in blue). But the relationship is more complex than a simple linear one, as points are often bunched together, indicating similar Shapley values, even though they have different colors (i.e., different levels of the DB indicator). Moreover, while the ranking gives a sense of general variable importance, it also highlights that, in some circumstances, indicators that are otherwise relatively unimportant can obtain a larger role. A case in point is the cost of firing employees which, ranked 19th, is generally less important, except for a handful of cases where it substantially reduces the business climate perception $\tilde{y}$. When using BMA, these outliers are influential enough to assign a high PIP to the cost of firing employees.

Figure 4 takes a closer look at the relationship between the most important DB indicators and $\tilde{y}$. Recovery rates in insolvency resolution, shown in Figure 4a, are a potent example for the role of non-linearities. While, on average, higher recovery rates are associated with a better business climate, the relationship appears to be far from linear. For values below 0.6, the slope is only slightly positive. As in Figure 1, there is a threshold somewhere between 0.6 and 0.8 at which the perceived business climate improves drastically with the DB indicator. But for values above 0.8, further improvements in recovery rates have no impact, presumably because at that point the bankruptcy resolution framework is deemed "good enough". This highlights a crucial difference between RF and BMA: the marginal returns to reform efforts depend on

[9] whenever $y_{c,i}$ is missing, it is replaced with $mean(y_i)$, as long as at least one outcome value is available for that country. As a result, the sample is the union of the six samples used in the previous section.

[10] Charts in this section are produced by the SHAP package by Lundberg and Lee (2017).

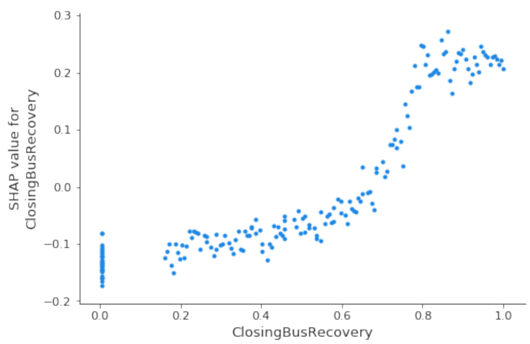
Figure 3: Shapley values: Summary



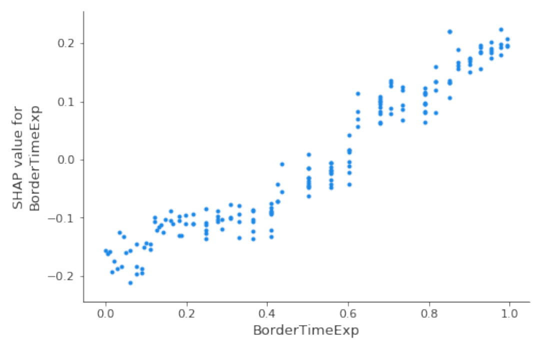
Notes: Higher feature (i.e., DB indicator) values indicate better performance. See Annex C for variable descriptions.

Figure 4: Shapley values by indicator

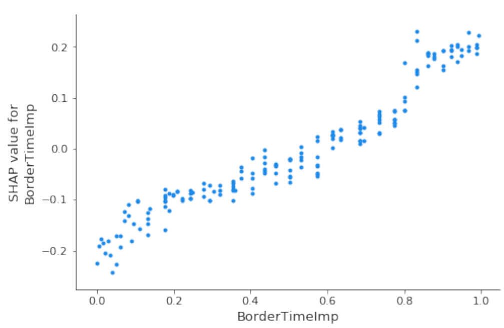
(a) Resolving insolvency: recovery rate



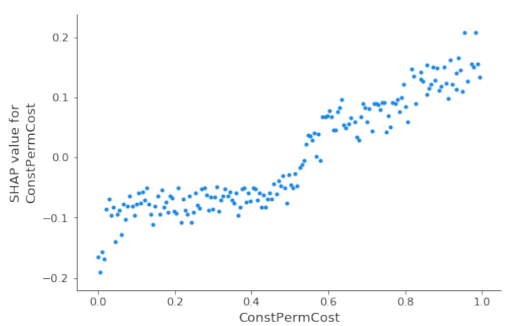
(b) Trading across borders: time to export



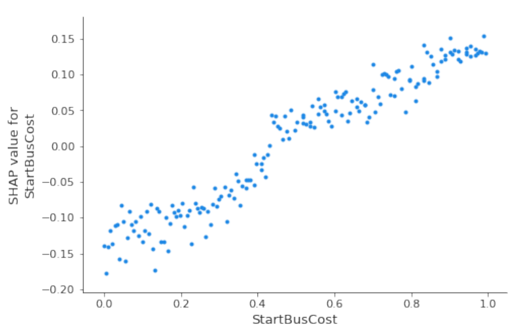
(c) Trading across borders: time to import



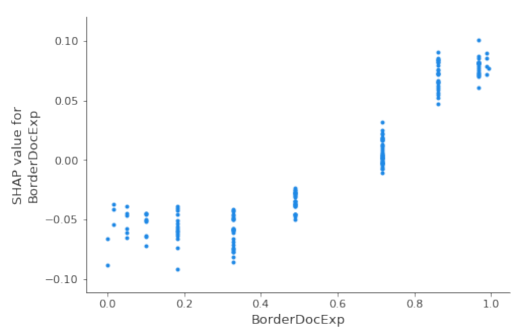
(d) Cost of construction permit



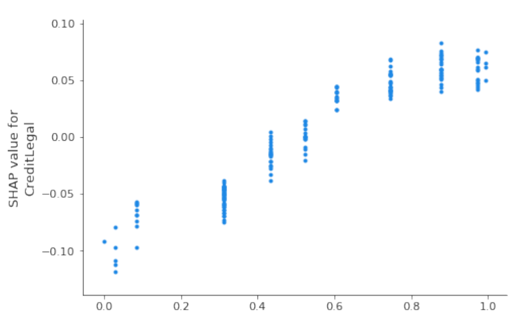
(e) Cost of starting a business



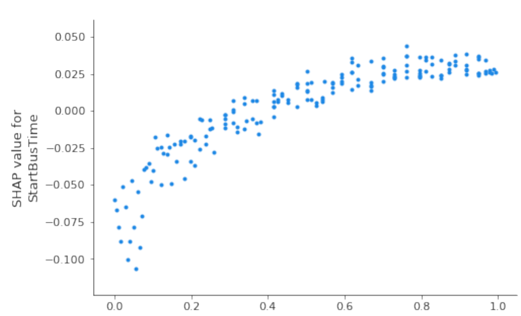
(f) Documents required to export



(g) Legal rights of borrowers and lenders



(h) Time to start a business



Notes: Higher feature (i.e., DB indicator) values indicate better performance. All DB indicators are scaled to $[0, 1]$.

country specific circumstances: For countries with very low values of the DB indicator, the low recovery rate explains the poor perception of the business environment. But that does not make reforms in this area a low-hanging fruit, as the efforts required to cross the threshold are potentially large. Note also that, if we had imposed a linear functional form, the estimated slope would depend on the share of observations with values below 0.6. Small changes to country coverage of the dependent variable can therefore lead to big changes in the estimated variable importance.

The form of the relationship between the DB indicators and $\tilde{y}$ appears to vary substantially across indicators. Figures 4.b and 4.c suggest that $\tilde{y}$ is close to linear in the time it takes to clear the border formalities when exporting or importing, whereas improvements in the cost of obtaining a construction permit (Figure 4.d) matter a lot more once that indicator is above 0.5. Similar thresholds seem to play a role for the cost of starting a business, the documentation requirements for exporters, and the strength of legal rights for borrowers and lenders.

Do these patterns uncovered from the data represent causal relationships? Given the similarity in approach, the RF inherits any concerns about causality from BMA, and switching from one method to the other does not increase or reduce those concerns. As noted in Section 2, in this specific context, the concerns about endogeneity should be limited.[11] The remaining concerns are twofold: First, the 38 DB variables may not be comprehensive enough to cover all important aspects of the business climate, which would lead to omitted variable bias.[12] And second, both BMA and RF induce parameter bias to avoid overfitting. In the case of RF, this bias means that the measured importance of the top variables is more likely to be diluted than exaggerated.[13] With these caveats about the magnitude of the estimated effects, the results do allow for a causal interpretation.

It also important to note that absence of evidence of an effect does not imply evidence of absence of an effect, particularly in such small samples. The fact that, in this particular sample, some variables appear to be of very limited importance could be due to an absence of sufficient variation in those variables.

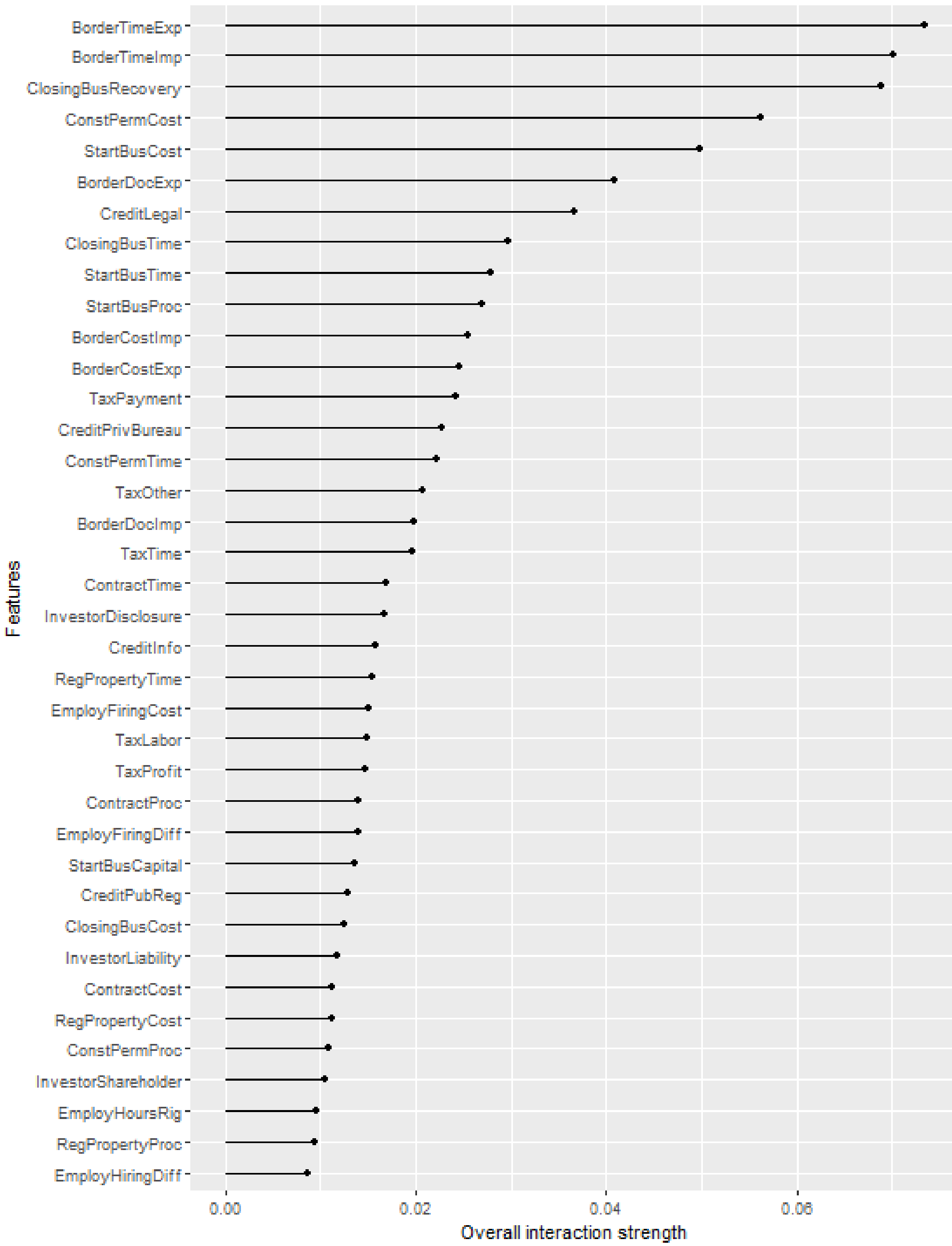
The final question is to what extent the importance of one indicator depends on the strength of other indicators: are the DB indicators complements or substitutes in

[11] That is, the dependent variable (perceived business climate) does not capture an economic outcome but merely aggregates the more granular information on the right-hand side.

[12] E.g., unlike later editions, the 2009 DB indicators do not cover the ease of getting electricity.

[13] Note that most important variables are used less often than warranted in an unbiased model, because they are not always part of the random set of m_{try} available variables.

Figure 5: Interaction strength: Summary



Note: Chart reports interaction strength measured using H-statistics (see Friedman and Popescu, 2008). See Annex C for variable descriptions.

determining the overall perception of the business environment? To assess this question, I compute the H-statistics for each pair of variables, as proposed by Friedman and Popescu (2008). The H-statistic for indicator j measures the variation in the prediction function that is explained by interactions between indicator j and all other indicators (excluding j), as a share of the total variation in the prediction function explained by indicator j. Figure 5 reports the H-statistics and finds that interactions are very weak: for any DB indicator, less than 10 percent of explanatory power is obtained from interactions. For policymakers, this is an encouraging result. The low degree of complementarity means that sequencing of reforms may not be a first-order issue, as the payoff from reforms in one area does not depend on the status of reforms in other areas. It should be noted that RF does not explore all possible variable interactions. It only considers an interaction if at least one of the two variables matters in itself (i.e., without the other one). But in most cases that seems to be a fairly mild and plausible restriction.

4 Conclusion

This paper has shown that perceptions of a country's business environment can be tied to specific indicators that are under the control of policymakers. Generally speaking, efficient insolvency procedures, speedy border formalities, and low startup costs are found to be of first-order importance. However, the exact order of priorities depends on country specific circumstances. Identifying these country-specific priorities is made possible by the use of non-linear modeling techniques like Random Forest. Moreover, when using Random Forest instead of linear techniques, explanations for cross-country differences in the business environment are not sensitive to how the overall business environment is measured.

References

[1] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[2] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.

[3] Jerome H Friedman, Bogdan E Popescu, et al. Predictive learning via rule ensembles. *The Annals of Applied Statistics*, 2(3):916–954, 2008.

[4] IMF. Reigniting growth in low-income and emerging market economies: What role can structural reforms play? *World Economic Outlook, October, Chapter 3*, 2019.

[5] Aart Kraay and Norikazu Tawara. Can specific policy indicators identify reform priorities? *Journal of Economic Growth*, 18(3):253–283, 2013.

[6] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *CoRR*, abs/1705.07874, 2017.

[7] Lloyd S Shapley. A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317, 1953.

[8] Erik Strumbelj and Igor Kononenko. An efficient explanation of individual classifications using game theory. *The Journal of Machine Learning Research*, 11:1–18, 2010.

Annex

A RF hyper-parameters

To find the optimal hyper parameter m_{try}, I use the following repeated cross-validation procedure:

1. The sample is divided into 10 random subsamples.
2. I guess a value for m_{try}
3. For each subsample I estimate the RF model using all observations except that subsample and then use the held-out subsample to make out-of-sample predictions. Performance is measured as the average RMSE over all 10 subsamples.
4. Steps 2-3 is repeated for values of m_{try} between 2 and 20.
5. Steps 1-4 are repeated 10 times
6. The value for m_{try} that delivers the best average RMSE over the 100 held-out samples is used to then estimate the RF model using the full sample.

Table 3 reports the selected hyper-parameters.

Table 3: Hyper-parameters

Outcome variable	DRI	EIU	GAD	GCS	PRS	WMO	$\tilde{y}$
#observations	137	144	158	130	133	178	178
m_{try}	5	10	5	4	6	12	7

Notes: DRI=Global Insight Global Risk Service, EIU=Economist Intelligence Unit, GAD=Cerberus Corporate Intelligence Gray Area Dynamics, GCS=Global Competitiveness Report, PRS=Political Risk Services, WMO=Global Insight Business Risk Conditions.

B Alternative variable importance measure for RF

Table 4: Random Forest: OOB permuted predictor importance

	DRI	EIU	GAD	GCS	PRS	WMO
StartBusProc	1.6	14.9	12.6	15.1	8.5	10.5
StartBusTime	4.8	7.0	22.2	17.5	6.5	13.8
StartBusCost	17.5	27.4	35.1	21.1	29.1	54.9
StartBusCapital	5.8	-0.2	16.4	3.6	2.9	2.8
ConstPermProc	0.0	2.7	2.7	4.7	1.7	4.3
ConstPermTime	6.9	-0.3	5.1	22.4	16.2	4.4
ConstPermCost	10.0	36.6	28.1	27.0	28.0	50.3
EmployHiringDiff	-2.0	-2.1	2.5	6.8	6.4	-0.2
EmployHoursRig	4.9	1.8	3.2	6.3	0.2	5.7
EmployFiringDiff	4.6	13.8	3.4	-4.4	3.6	3.5
EmployFiringCost	17.8	7.1	8.0	9.1	18.1	15.2
RegPropertyProc	0.8	-1.4	1.8	19.4	4.1	2.0
RegPropertyTime	1.7	4.1	7.2	7.5	-4.1	1.9
RegPropertyCost	5.3	12.2	4.3	-1.7	5.0	7.7
CreditLegal	12.9	7.8	31.7	8.5	1.7	23.3
CreditInfo	6.9	12.1	15.8	1.6	14.7	6.4
CreditPubReg	-4.1	-1.9	2.3	3.0	1.0	0.7
CreditPrivBureau	5.1	20.5	18.6	3.8	12.3	22.9
InvestorDisclosure	3.4	7.9	7.6	-0.3	3.1	7.1
InvestorLiability	4.1	0.3	7.8	7.6	1.1	8.6
InvestorShareholder	0.1	-0.6	8.8	1.2	7.9	4.9
TaxPayment	2.9	16.8	21.1	18.7	31.2	28.6
TaxTime	1.4	5.4	-0.2	27.2	22.9	15.2
TaxProfit	6.8	6.3	2.5	-1.7	0.4	1.5
TaxLabor	0.3	3.9	5.9	3.6	-2.9	2.5
TaxOther	24.0	21.4	9.6	1.2	5.6	8.6
BorderDocExp	22.5	18.9	22.9	17.3	16.8	33.4
BorderTimeExp	29.3	30.4	31.1	31.1	22.2	52.3
BorderCostExp	22.7	21.7	9.5	28.1	6.2	10.5
BorderDocImp	7.3	11.6	18.9	18.3	8.7	17.7
BorderTimeImp	34.6	39.3	34.6	31.2	24.3	39.3
BorderCostImp	26.5	28.0	10.4	26.1	6.2	10.1
ContractProc	2.7	14.4	15.3	16.9	1.6	10.6
ContractTime	1.4	12.8	6.3	4.5	3.7	4.8
ContractCost	-0.3	10.1	10.0	9.6	1.3	18.6
ClosingBusTime	4.7	21.3	28.6	20.9	8.5	21.9
ClosingBusCost	-0.4	15.5	17.4	10.7	5.4	16.7
ClosingBusRecovery	14.4	42.8	40.9	30.5	30.3	53.6

Notes: Numbers indicate variable importance in terms of out-of-bag permuted predictor importance. Shaded cells indicate top 10 DB indicators. Columns represent RF regressions with different outcome variables, each constructed from a different source: DRI=Global Insight Global Risk Service, EIU=Economist Intelligence Unit, GAD=Cerberus Corporate Intelligence Gray Area Dynamics, GCS=Global Competitiveness Report, PRS=Political Risk Services, WMO=Global Insight Business Risk Conditions. See Annex A for RF hyper-parameter settings. See Annex C for variable descriptions.

C Doing Business variable descriptions

Table 5: List of Doing Business variables

Abbreviation	Description
Starting a business	
StartBusProc	Number of procedures
StartBusTime	Time
StartBusCost	Cost (% of income per capita)
StartBusCapital	Paid-in minimum capital (% of income per capita)
Dealing with construction permits	
ConstPermProc	Number of procedures
ConstPermTime	Time
ConstPermCost	Cost (% of warehouse value)
Employing workers	
EmployHiringDiff	Difficulty of hiring (index)
EmployHoursRig	Rigidity of hours (index)
EmployFiringDiff	Difficulty of firing (index)
EmployFiringCost	Cost of firing (weeks of salary)
Registering property	
RegPropertyProc	Number of procedures
RegPropertyTime	Time
RegPropertyCost	Cost (% of property value)
Getting credit	
CreditLegal	Strength of legal rights (index)
CreditInfo	Depth of credit information (index)
CreditPubReg	Credit registry coverage (% of adults)
CreditPrivBureau	Credit bureau coverage (% of adults)
Protecting minority investors	
InvestorDisclosure	Extent of disclosure (index)
InvestorLiability	Extent of director liability (index)
InvestorShareholder	Ease of shareholder suits (index)
Paying taxes	
TaxPayment	Number of payments
TaxTime	Time (hours per year)
TaxProfit	Profit tax (% of profits)
TaxLabor	Labor tax and contributions (% of profits)
TaxOther	Other taxes (% of profits)
Trading across borders	
BorderDocExp	Documents to export (number)
BorderTimeExp	Time to export
BorderCostExp	Cost to export (US$ per container)
BorderDocImp	Documents to import (number)
BorderTimeImp	Time to import
BorderCostImp	Cost to import (US$ per container)
Enforcing contracts	
ContractProc	Number of procedures
ContractTime	Time
ContractCost	Cost (% of claim)
Resolving insolvency	
ClosingBusTime	Time
ClosingBusCost	Cost (% of estate)
ClosingBusRecovery	Recovery rate (cents on the dollar)