

FDI and Trade Outcomes at the Industry Level

A Data-Driven Approach

Jean-Christophe Maur

Milan Nedeljkovic

Erik von Uexkull



WORLD BANK GROUP

Finance, Competitiveness and Innovation Global Practice

&

Macroeconomics, Trade and Investment Global Practice

January 2022

Abstract

This paper proposes a novel empirical methodology to reveal factors associated with foreign direct investment decisions and export success at the industry level. Faced with large amounts of policy and economic indicators, as well as significant product and sectoral diversity, the motivation of this research is twofold. First, it selects among the vast number of indicators and potential factors relevant to investment and trade outcomes a manageable subset of variables that

can usefully guide the understanding of investment and trade outcomes. Second, it provides robust estimates of how these variables affect the probability at the margin of investment and trade outcomes at the sector level. Finally, the paper uses these estimates to produce new metrics (“scores”) of trade and investment climate performance at the country and sectoral levels.

This paper is a product of the Finance, Competitiveness and Innovation Global Practice and the Macroeconomics, Trade and Investment Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at jmaur@worldbank.org and jvonuexkull@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

FDI and Trade Outcomes at the Industry Level – A Data-Driven Approach*

Jean-Christophe Maur[†], Milan Nedeljkovic[‡] and Erik von Uexkull[§]

Keywords: trade, investment, export competitiveness, dimensionality reduction, machine learning

JEL classification codes: F14, F21, C55

* The authors would like to thank Mary Hallward-Driemeier, Jean-François Arvis, Yan Liu, Samuel Rosenow, and Victor Steenbergen for very useful comments and suggestions on earlier versions of this paper, as well as Samuel Fraiberger and Melise Jaud for productive discussions. We are deeply grateful to Miodrag Petkovic and Maria Reinholdt Andersen for their assistance in preparing respectively the policy dataset and the benchmarking excel tool. This research benefitted from support provided by the Umbrella Facility for Trade trust fund that receives contributions from the governments of the Netherlands, Norway, Sweden, Switzerland, and the United Kingdom. All errors remain the authors' sole.

[†] World Bank, jmaur@worldbank.org

[‡] FEFA, Metropolitan University, Belgrade, mnedeljkovic@fefu.edu.rs

[§] World Bank, jvonuexkull@worldbank.org

Motivation

Endeavors by multilateral agencies, donors, non-governmental organizations and academia to create a better understanding how institutions and policies determine economic and social outcomes have led to the considerable development of databases seeking to present meaningful global indicators of policies and policy outcomes across countries.

One of the most influential works that led to this boom has been the academic work on institutional economics (starting with Djankov, La Porta, Lopez de Silanes and Schleifer, 2002), itself inspired by the work of Hernan De Soto (1989). The research used cross-country variations in institutional quality (measured by indicators of performance) to demonstrate their importance for economic outcomes. This later led to the production of annual Doing Business indicators in 2003 and an earlier iteration of this in the World Development Report 2002 (Djankov, 2016). Doing Business has been hugely influential in informing development economics. Other influential indicators include the World Economic Forum's competitiveness report, and the Logistics Performance Index. The business of indicators is now thriving: the World Bank's TCdata360 on trade and competitiveness, an aggregator site, offers access to over 2,400 indicators.

The statistical relevance of many of these policy indicators has been amply demonstrated in the academic literature, and their relevance to economic policy through the widespread adoption by policy makers. However, the exact nature of what these indicators measure remains often subject to interpretation. They vary in degree of quality, and approaches to measurement: from ordinal measures such as perception of given issues, expert judgment of policy quality, to objective measurements of economic variables.⁵ Amidst this variety, the sheer volume of indicators has become challenging to navigate, and thus to incorporate into coherent frameworks. For instance, we know that what affects trade or investment success is due to a very wide range of causal factors, but where are the best measures of these factors? Most analyses using indicators only look at selected subsets of indicators, and one can question how these indicators have been selected.

Aware that different sets of factors matter in different country contexts, policy advisors have developed broader frameworks (starting with Doing Business) that would offer a more comprehensive picture, guiding policy makers in their reform efforts, and helping them identify and prioritize where action is needed. There is a plethora of tools and a large industry serving prospective investors and policy makers to assist them in identifying priority markets and factors impacting them, illustrating the complexity of the environment facing them and therefore the difficulty to reduce it to the key factors that would majorly impact their decisions. The exercise is not without challenges, as noted by Kraay and Tawara (2013). Their research finds significant shortcomings when attempting to identify sets of policy indicators that could be particularly correlated with specific (desired) policy outcomes. They find that sets of policies identified for one outcome do not correlate well with closely related and similar outcomes, thus suggesting that any attempt to identify partial effects of specific policies and to pick a subset of variables could be spurious.

A recent framework was developed for a new strategic exercise in the World Bank Group, the Country Private Sector Diagnostics (CPSD), designed to inform support efforts to mobilize private sector initiative for development. To guide the understanding about the extent of factors constraining private

⁵ For a discussion, see Hausmann and Klinger (2008).

investments -- and for which industries these constraints bind or not – a benchmarking tool using 163 indicators (including Doing Business and WEF indicators) for 140 countries was developed. The selection of the indicators, however, was only based on expert judgment.

Building on these earlier efforts, we propose in this paper a novel empirical methodology to identify factors associated with respectively foreign investment decisions and international trade success, using publicly available data measuring the environment in which trade and investment happen, with the following aims:

- Reduce the dimensionality of the universe of indicators to a subset of indicators that are significantly associated with trade and investment outcomes (“success”).
- Robustly identify the subset of indicators.
- For the selected subset of indicators, benchmark country and sector performance against the rest of the world and take initial steps toward the building of a predictive model of trade and investment “success”.

This empirical strategy is ultimately designed to help policy makers and economic actors improve their strategic focus on the variables that have the greatest impact on their economic objective, and possibly on sector specific interventions with the greatest value for money.

[A review of commonly used investment and trade climate indicators and identification of the strategy environment for trade and investment](#)

There is a vast empirical literature investigating the drivers of trade and investment outcomes. We review here briefly the main empirical strategies and analytical frameworks that are the most commonly used in the policy literature, reflecting on the considerable growth of available information on indicators tracking factors of trade and investment climate, while also noting some of the limitations of currently used approaches.

- [Drivers of investment climate and business environment](#)

Derived from the new comparative economics literature, investment climate and business environment indicators rankings aim to exhaustively benchmark the private sector or investment climate environment at the country level. For each country, the performance on individual attributes relative to the rest of the world and best performers is being recorded. The Doing Business Indicator series, which was prepared annually by the World Bank Group between 2003 and 2020 measures 11 different areas using 41 indicators for 180 economies.⁶ For each area, each country’s performance is measured by a score indicating the country’s level relative to best practice (Distance to Frontier indicator).⁷ Another widely used set of indicators are the Global Competitiveness Indicator (GCI) series prepared annually by the World Economic Forum since 2007. The WEF strategy is to gather existing indicators (including Doing Business ones) alongside own perception surveys of executives. Its 2017-18 edition covers 12 areas of

⁶ The areas are: starting a business, dealing with construction permits, getting electricity, registering property, getting credit, protecting minority investors, paying taxes, trading across borders, enforcing contracts, resolving insolvency and labor market regulations.

⁷ The methodology of the Doing Business indicators was updated in 2015 to adopt the DTF measure, which measures progress over time over the previous ordinal ranking measure.

competitiveness using 114 indicators in 137 economies.⁸ It provides a score of between 0-100 for aggregate indicators ranking countries as well as for each of the areas of competitiveness. Another tool using a similar approach is the *Export Competitiveness Diagnostic Toolkit* from the World Bank (Reis and Farole, 2012) which uses quantitative analysis to assess constraints to export competitiveness in 13 broad policy areas. It assesses a country's overall basket of exports and/or specific traded sectors and identifies the main constraints to improved trade competitiveness and related policy responses.

Investment Climate Surveys are conducted at the firm level and thus focus primarily on investigating firm characteristics and how different firms relate to specific policy issues.⁹ Conducted by the World Bank since the 1990s, *Enterprise Surveys* provide firm-level surveys of a representative sample of an economy's private sector.¹⁰ While methodologically different to benchmarking approaches, they offer the same possibility of cross-country panel study as surveys share identical questions across the various countries in which they have been implemented. The surveys cover a broad range of business environment topics including access to finance, corruption, infrastructure, crime, competition, and performance measures. The available information now includes information on 139 countries gathered from 131,000 firms. Constraints facing firms can be ranked and the data can be parsed by size of firm and to a lesser extent by sectors.¹¹ The data consists of a mix of perception/recall data and actual revenue and cost information provided by firms. Enterprise surveys are periodically repeated allowing for a review of how constraints evolve over time and selectively for the use of panels. Based on the Enterprise Surveys, Investment Climate Assessment reports have been often produced which in addition to an in-depth analysis of the survey data also undertake a deeper analysis of investment climate constraints and potential solutions.

- **Identification Frameworks**

Already implicitly present in tools such as Doing Business and World Economic Forum approaches, is the idea of providing a comprehensive and organized framework through sets of policy variables (the "pillars") that aim to reflect all the policies that impact respectively private sector activity and competitiveness. However, these frameworks come short of providing a true analytical structure to relate the various policies to the desired outcome and identify which matter more particularly given a specific country context.

The *Growth Diagnostics* proposed by Hausmann and Rodrik¹² focuses on formulating growth strategies (to maximize social welfare) by identifying key reform priorities through an analysis of where market imperfections and distortions, and therefore failure of public interventions result in a difference between private and social rates of return. Since wholesale reform is not realistic, the framework uses a methodology to identify binding constraints focusing on the supply side of investment (i.e., inadequate

⁸ The areas are: basic requirements (i.e., institutions, infrastructure, macroeconomic environment and health and primary education); efficiency enhancers (i.e., higher education and training, goods market efficiency, labor market efficiency, financial market development, technological readiness and market size); and innovation and sophistication factors (i.e. business sophistication and innovation).

⁹ <http://www.enterprisesurveys.org/Methodology/Enterprise-Surveys-versus-Doing-Business>

¹⁰ The surveys are administered by a central unit in the World Bank though there is region specific collaboration with other institutions (e.g., with EBRD in Eastern Europe and Central Asia where the surveys are known as Business Environment and Enterprise Performance Surveys).

¹¹ The main limitation for sector analysis is that the surveys focus on manufacturing oriented formal firms.

¹² Hausmann, Rodrik and Velasco (2005) and Hausmann, Klinger and Wagner (2008).

access to finance) and the demand side (i.e., inadequate returns to investment and/or inadequate private appropriability of returns).

Another approach proposed more recently is the *Growth Identification and Facilitation Framework* (GIFF) which focuses on exports and on first the identification of sectors of latent comparative advantage and second on feasible and focused policies to achieve structural transformation toward the production in these sectors (Lin and Monga, 2010). Like the product space analysis proposed by Hausmann and Hidalgo,¹³ this approach commences with the identification of a list of tradeable goods and services. However, the focus is not on related products but on products exported in the previous 20 years in dynamically growing countries with similar endowment structures and per capita income. The methodology then focuses on specific measures to develop priority sectors including: identifying binding constraints; attracting global investors and/or incubating domestic firms; highlighting unique country endowments and new technologies/industries; developing targeted interventions to address poor infrastructure and bad business environments; and providing time bound incentives to pioneer firms.

- [Empirical work on determinants of trade and investment](#)

The availability of indicators has led to advances in the empirical understanding of determinants of trade and investment outcomes at a more granular level and a considerable literature. Offering firm level data and a large array of firm level indicators, including direct and indirect measurement of policy induced costs, Enterprise Survey data is a key contributor. Among the seminal works, Dollar, Hallward-Driemeier and Mengistae (2005) show using enterprise survey data that the investment climate affects various metrics of firms' performance. Dollar, Hallward-Driemeier and Mengistae (2006) use surveys from 8 countries to assess the impact of policies affecting investment climate on firm international trade integration. Low customs clearance times, reliable infrastructure and good financial services are associated with the likelihood to export. Eifert, Gelb and Rachamandran (2005) and Eifert, Gelb and Rachamandran (2008) find the importance of high indirect and business environment cost impact negatively the productivity of African firms relative to those in other countries.

Likewise, Doing Business data has led to a vast literature on structural and policy factors linked with participation in international trade and investments flows. Early works include Djankov, Freund and Pham (2006), who measure how border policies impact the ability of firms to trade, showing notably that time lost at the border is especially costly. Sekkat and Varoudakis (2007) examine the impact of investment climate on FDI.

In the empirical literature on the effects on trade of institutional and structural variables, the predictive power of the gravity model introduced by Tinbergen (1962) has made it the most popular tool. Gravity modeling helped establish firmly the empirical truth about the role of distance and economic size as key factors of bilateral international trade flows. The model has been refined over the years, as well as grounded to the main theories of international trade (see for a comprehensive review Head and Mayer, 2013). From its inception, the gravity model has been used to test the existence of policy dimensions linked to trade, and it has been essential in furthering the understanding of the drivers of international trade and transaction costs (transport costs, and trade policies) directing trade flows. The gravity approach has also been applied with success to other bilateral economic flows: services (Kimura and

¹³ Hidalgo, Klinger, Barabasi and Hausmann (2007).

Lee, 2006), cross-border portfolio investments (Portes and Rey, 2005), and FDIs (Bergstrand and Egger, 2007; Head and Ries, 2008).

A challenge of the gravity approach is its focus on dyadic variables -- elements that bind pairs of countries (Head and Mayer, 2013), which requires adding country (importer and exporter) fixed effects, thus making the treatment of country specific policy variables more complicated. As noted by Head and Mayer (2013) there are some workarounds to retain country specific variables, as well as the possibility of using panel approaches.

A rich literature on gravity modeling has also led to reviewing the evidence collected by these studies through meta-analysis. Head and Mayer (2013) expand the dataset used by Disdier and Head (2008) to 159 papers to review 2500 coefficient estimates of policy variables. However, their focus is only to investigate the impact a few trade policy variables of regional integration and currency union on bilateral trade outcomes.

In a similar spirit to this research, and the meta-analysis approach (albeit on a much more limited sample of 8 previous papers), Blonigen and Piger (2014) use Bayesian Model Averaging of 56 potential covariates with FDIs, which include measures related to GDP, labor endowments, capital, land and natural resources endowments, trade openness, FDI and investment climate, tax policy, communication infrastructure, financial infrastructure, policy environment, as well as dyadic variables commonly included into gravity models: distance, cultural proximity, and geographic proximity measures. In line with the findings of our research, their results find a narrower set of variables (between 7 and 16) with high inclusion probability as predictor of FDI, thus suggesting that a parsimonious model could explain FDI outcomes. They find little support for policy variables controlled by the host country (such as multilateral trade costs, business costs, infrastructure, or political institutions), influence FDI, with an exception for bilateral trade and investment policies.

- [Shortcomings of existing empirical approaches](#)

The methodology proposed in this paper is in part motivated by some of the limitations of existing empirical analyses of factors associated with investments and trade outcomes, namely the tendency to focus on measuring the relevance of individual policies rather than packages of reforms, how to select from a very large number of indicators and dimensions, and the relative absence of focus on sector specific dimensions.

Focus on individual policies rather than package of reforms. The focus of most tools has often been on measuring individual policies and either assess whether such policies matter (gravity models) or whether they lag best practice (Doing Business and WEF). Thus, where these tools are the most effective is at providing better knowledge on the quality of individual policies that are *assumed* to matter for economic success (growth, trade or investment performance). As to the question of how these individual policies matter relatively to each other, tools like Doing Business or WEF do not offer much in the answer except to highlight by benchmarking which policies are lagging the most compared with other countries and best practice. Identification frameworks attempt to provide a heuristic way to sort out between policies that matter more and others, but not an empirically testable one.

Proliferation of indicators. Beyond that, the proliferation and multiplicity of indicators, while offering more wealth of information than ever, does not really help further in identifying issues of prime importance. On the contrary, by adding to the overall dimensionality, and in providing added policy

granularity as well as focus on new policy issues (including sectoral ones) they make the problem in a way more complex, albeit in a positive way richer in information content. There is also the question of varieties of methodologies used in producing these indicators, which can be subject to limitations, thus creating uncertainty as to what are best-practice indicators.

Absence of sector dimensions. A third limitation is that most tools fail to account for sector or product specific dimensions. In some cases, the sector dimension is nearly fully absent of the discussion and the focus is almost exclusively on country-level cross-cutting issues such as in Doing Business, the WEF indicators or Growth Diagnostics. Arguably, even policies that affect all sectors have different degrees of relevance for different sectors and thus priority reform order will be sector/product specific to some extent. Tools that use product and sector specific information such as gravity modeling and the GIFF are also not really designed to evaluate for sector specific policy dimensions. An early attempt to find the relevance and priority of policies at the sector level can be found in Pathikonda and Farole (2016). Another relevant development has been the use of sectoral gravity models to acknowledge the vastly different nature of trade for different products and industries (see Yotov et al., 2015, for a review).

In conclusion none of the current tools provides clear guidance on the choice of the relevant set of public policy reforms that could lead to improved outcomes. Because of their partial nature, existing tools while helping identify relevant policies offer little guarantee that reform will lead to change in outcomes, because other binding constraints may be missing from the diagnostic. Second because of the multiplicity of indicators, identification tends to be biased by how these indicators are constructed and it might be difficult also to distinguish between closely related indicators and which convey better information. Also, by failing to account for the sector specific responses existing tools lack focus. Arguably, the identification of constraints at the country level will reflect the constraints that affect sectors that are core to an economy because of averaging effect, but by the same token average effects might hide constraints that are very important for a core sector but not so much for the rest of the economy.

Ideally, one would want to achieve a level of analysis that cuts across both dimensions i.e., those which affect the economy as whole (cross-cutting enabling conditions), and those aspects which affect only particular economic sectors or firms. Moreover, the analysis should be detailed enough to offer concrete guidance for policy makers in view of implementation and focus on results.

Constraints benchmarking

In this section, we briefly describe the investment constraints benchmarking tool (CBT) from which this empirical work derives, and which was developed to inform the World Bank Group's Country Private Sector Diagnostics. The idea behind the tool is two-fold. In a spirit similar to the WEF, the CBT brings together and organizes in one place a substantial set of indicators. The tool builds on international surveys which contain information regarding factor prices and quantities (such as the World Development Indicators), and metrics of policies (such as Doing Business and Investment Climate surveys). It also builds on previous attempts to create composite measures of these dimensions such as the World Economic Forum's competitiveness index. Second, and in a development from existing tools, the CBT proposes a way to measure each country's performance at the sector level (instead of just at the economy-level) by observing the revealed preference of individual foreign direct investments.

The goal is to measure constraints for each sector in the country of study and compare these measurements with benchmarks of performance in countries across the world where investments have been successfully made in this same sector. This is expected to provide a revealed measure of whether a constraint is absolutely binding or not for investors in a sector. For example, if a given indicator (e.g. electricity supply) is essential for a sector (e.g. manufacturing), only countries with good electricity supply will witness successful private investment in manufacturing. For other sectors that depend less on electricity supply, such as agriculture, countries with poor electricity supply may still attract successful private sector investment. This benchmarking of constraints thus enables singling out the set of constraints for each sector where the country's performance is below the minimum level required for successful investment. If constraints are frequently below investment level across different sectors, then this would also reveal constraints that may be binding for the entire economy.

The analysis is based on country-specific performance indicators (163 in total sourced from the World Bank TCdata360)¹⁴ as sub-indicators of 12 categories representing: Demand, Factors (labor, capital), Key Inputs (infrastructures), and Institutions (macroeconomy, governance). The binding nature of the constraints measured by these indicators is then assessed by benchmarking them with a metric of third countries and sector specific score of success. In its current form, this metric is the existence of actual FDI in a sector and country¹⁵ as recorded by the Financial Times' *fDi Markets* database.

The CBT is designed to guide analysis through a large amount of information and provide an initial indication of where the biggest issues may be. However, the approach comes with limitations. First, it does not provide the user with the information on whether the variables included are associated in a statistically significant way with the success outcome of interest (i.e., attracted FDI or a revealed comparative advantage in a sector). Yet, the ability to understand which variables, among many variables, are the statistically significant (key) binding constraints is of high importance for efficient decision making (both from the investor and policy maker perspective). Second, the benchmarking process (i.e., assigning scores from 1-5 to each indicator for every country/sector pair) considers each indicator in isolation and is based only on the past information about the values of the indicator and the sector-level success performance. The process does not account for the effects of other relevant indicators or of potential associated factors, thereby potentially producing biased estimates.

Identification of factors associated with trade and investment in a high-dimensional framework

Objectives

Our approach is based on complementing the work discussed above by addressing some of its limitations. In particular, the objectives of the analysis are the following:

- Reduce the complexity of the universe of indicators associated with trade and investment outcomes success at the sector level.

¹⁴ Accessible at <https://tcdata360.worldbank.org/>

¹⁵In an earlier version of the tool, IFC's DOTS database has been used. The DOTS database records IFC investments (so not only foreign investments but also domestic ones to which IFC is part of) and their performance. Thus, the tool enabled to benchmark performance against that of profitable investments, offering a more robust benchmark.

- Develop a robust empirical method to identify the subset of indicators (from a large set) which are significantly associated with measures of trade and investment outcomes at the sector level (trade and investment sector “success”).
- Use the model estimates to construct robust benchmark scores of country/sector performance against the rest of the world for the identified subset of indicators.

In this way, econometric analysis complements the existing ordinal-ranking based CPSD methodology by: i) providing statistical assessment of the strength of association between the individual indicators and the sector specific performance and by ii) explicitly taking into account indicator’s forecasting performance as well as the effects of other indicators and potential confounding factors in the process of constraints benchmarking. The produced benchmarking heatmap empowered by the empirical model estimates provides actionable and refreshable signals of the identified sector level constraints which can be used to facilitate efficient data-driven economic policy decision making.

In the remaining parts of this section we first discuss data which is used in the empirical analysis before discussing in detail the empirical design and econometric approach. We then present the baseline results together with conducted specification checks.

Data

We use three general types of data in the analysis:

I. Foreign direct investment data

We use the Financial Times’s *fDi Markets* database for FDI data. The database provides a comprehensive source of information on the firm-level greenfield investment, containing reported deals/projects world-wide over the 2003-2019 period for 206 countries. The data set contains information on the recipient and the source country, principal industry sector to which the foreign direct investment type can be mapped, the year/month of the transaction, the name of the investing firm as well as various other information. The *fDi Markets* database focuses on the *new* FDI projects and provides cleaner information on the FDI sector level patterns relative to the balance of payments (BOP) based FDI data. The latter contains sector level information on FDI flows only for a limited number of countries, does not distinguish between existing (reinvested earnings) and new FDI activities and may also be distorted by “round-tripping” behavior of firms. A potential disadvantage of the *fDi Markets* database is that the reported values of FDI projects are estimates which may differ from the actual recorded values.¹⁶ Moreover, while the data set coverage is extensive after 2010, in earlier years the coverage may not be fully comprehensive. We circumvent these disadvantages by focusing the analysis on the last decade (2010-19 period) and using the available information to construct binary FDI success variable – which takes value 1 if the country/sector pair received FDI (from any country) over the specified period and zero otherwise. The use of a binary indicator is imperfect as it does not distinguish by size and number of investments, and thus the relative share/frequency of inward FDI that a country may command relative to the rest of the world and to the size of its economy. However, a discrete indicator is possibly less inadequate than potential alternatives. The project value in *fDi Markets* database is a rough

¹⁶ The *fDi Markets* database uses public information sources such as media sources to inform its database. See: <https://www.fdimarkets.com/faqs/>

approximation in most cases,¹⁷ and the author's view is that a value indicator would introduce significant additional noise. For the same reason, introducing a measure relative to GDP or population size would introduce the same source of distortion. An alternative could be a count variable of the number of FDI projects, but it would heavily bias the measure against small countries. The constructed binary variable is our dependent variable in the empirical analysis.

II. Trade data

We use the BACI data set prepared by CEPII to construct a measure of the revealed comparative advantage at the disaggregated 4-digit sector level.¹⁸ The BACI data set provides disaggregated data on bilateral trade flows for more than 5,000 products and 200 countries. The BACI database is built from data directly reported by each country to the United Nations' COMTRADE database, where additional procedures are applied by CEPII to reconcile the declarations of the exporter and the importer and provide more accurate measurement of trade flows relative to COMTRADE. The data set is available at annual frequency. The 4-digit level of disaggregation was chosen because it offers a good compromise between data set size and use of results for the construction of the benchmarking values.¹⁹

We use trade data to calculate the measure of the revealed comparative advantage (RCA) index (Belassa, 1967) for each country/sector pair. Based on the calculated RCA for each sector/country pair over the specified period, we construct the binary trade success variable which takes value 1 if the estimated RCA for country/sector pair is above 1 and 0 otherwise. We consider the use of a binary RCA variable as the most appropriate (though imperfect) approximation of whether a country has successfully exported a given product, relative to the rest of the world and export basket composition. The definition of export success in this case does not distinguish between instances of artificial competitiveness (for instance subsidized exports) and genuine ones, another limitation of the RCA and other trade outcome measures. Another downside of the indicator of success used in this analysis is that it is discrete despite being constructed from a data set with continuous and richer information. Using trade volumes or market shares would not be appropriate because it reflects the size of the exporting economy rather than specialization. Taking into account the RCA value or specific RCA levels in the variable would introduce information from this key variable that escapes easy interpretation: as it would pick up significant variations of RCA levels that can just as well be a reflection of the lack of economic diversification in other products rather than a reflection of success in the product in question on global markets. The constructed binary variable is our dependent variable in the empirical analysis.

III. Country indicators data

The indicators of the characteristics of each economy are collected from World Bank TCdata360, complemented by IMF WEO data for several macroeconomic indicators. Indicators are grouped in 4 broad categories: demand, production factors, key inputs and institutions. The latter three categories are further decomposed into subcategories summarized in Table 1. These indicators were selected to correspond to the country characteristics a potential investor or trader might consider when targeting a

¹⁷ Authors observed that this data seems often informed by public announcement of investments. Realized investments often differ quite considerably in scope. This was confirmed in discussions with World Bank colleagues working with this data.

¹⁸ http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=37

¹⁹ The empirical investigation could be conducted at a more disaggregated level, e.g. HS-6 digit.

destination country. They include policy variables that are amenable to change in the short to medium term, as well as country fundamentals such as per capita GDP or population size. While these indicators may not be under the control of policy makers, their inclusion nevertheless has an important policy dimension to insert a reality check into sectoral FDI strategies. For instance, countries may find that policy initiatives intended to promote FDI in a certain industry are doomed to fail if country fundamentals are detrimental to the development of such an industry in the country. In other cases, there might be short-term initiatives that policy makers can consider to compensate for shortcomings on particular fundamentals, such as a more open immigration regime to make up for skills shortages or spatial solutions to reduce transport costs and disadvantages related to remote location.

Table 1. Indicators categories

Category	Sub-category
DEMAND	Demand
PRODUCTION FACTORS	Labor and skills
PRODUCTION FACTORS	Geography and natural resources
PRODUCTION FACTORS	Existing capabilities
INPUTS	Energy
INPUTS	Transport
INPUTS	Finance
INSTITUTIONS	Regulatory barriers and taxation
INSTITUTIONS	Rule of law and property rights
INSTITUTIONS	Market contestability
INSTITUTIONS	Macroeconomic and political stability

Empirical Design and Approach

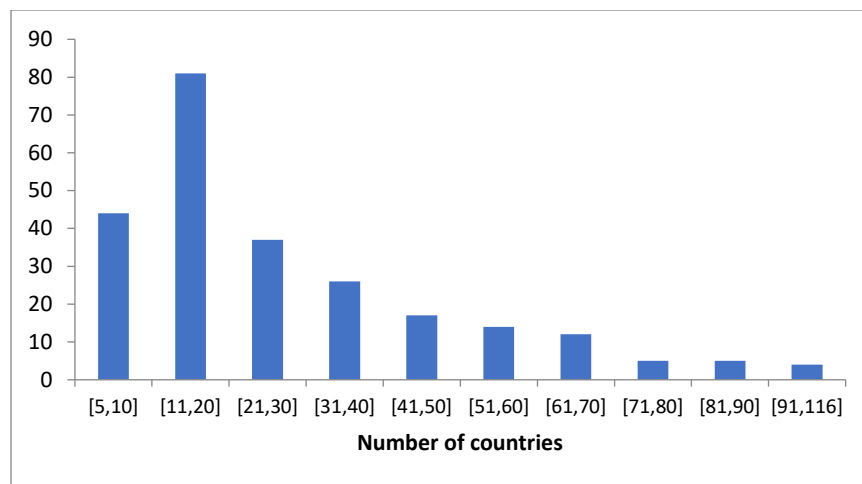
Data preparation

The initial country indicators data list included 594 potential indicators for 217 countries, collected over the 1960-2019 period. Not surprisingly, the country coverage is limited until the most recent periods. Due to large number of missing data points per individual indicators, we try to maximize the trade-off between including the maximum number of indicators while maintaining adequate country data coverage. After thorough data examination and cleaning, the final balanced data set includes 190 potential indicators which are available for 116 countries over the 2010-19 period, the period which matches the availability of FDI data while being representative in terms of the number of countries selected (Annex 3 provides the full list).

We average the data for indicators over two non-overlapping five-year periods (2010-2014 and 2015-19). The first period (2010-2014) is used for computation of the indicators to avoid any potential reverse causality issues (the lagged indicators are predetermined relative to the FDI and trade flows which are computed over the 2015-19 period). We do not use the information on indicators from the second period in estimations, but only for calculation of the thresholds for sector benchmarking exercise (see section below). Since we are interested in analyzing the key structural and macroeconomic factors associated with FDI and trade success from the perspective of the recipient country, we only use country level indicators and do not include any global or "push" factors.

We follow the international finance literature (e.g., Chinn and Prasad, 2003; Gruber and Kamin, 2007; Chinn, Eichengreen, and Ito, 2014) in using a 5-year period to smooth out the effects of business cycles and transitory (and global) factors, as well as to allow for the fact that the effect of the changes in structural drivers on FDI and RCA tend to materialize over time (adjustment lag).²⁰ We merge the indicators data set with FDI and trade data sets. The raw FDI data contained 702 sectors (sector/activity pairs in the *fDi Markets'* nomenclature) for 206 countries with large number of zeros (no FDI activity over the entire five-year period) for individual sectors. We first excluded countries for which the country indicators are not available over the period of analysis (2010-19). We then excluded those sectors which have fewer than 5 non-zero observations for the remaining countries, arriving to the final number of 245 sectors. Using the information on FDI flows, we construct a variable $Y_{i,j}$ which takes value 1 if sector i of country j was the recipient of any FDI flow over the five-year period 2015-19, and 0 otherwise. Analogously, we create a variable $Z_{i,j}$ based on the information on country/sector pair FDI "success" over the 2010-14 period. Figure 1 shows the histogram of the empirical distribution of the number countries which received FDI flows in any of the 245 sectors over the 2015-19 period. We see that the distribution is largely skewed toward sectors in which FDI flows went to only a subset of countries (the left-hand side of Figure 1). In the median sector (by the number of recipient countries) only 20 countries are the recipients of FDI inflows.

Figure 1. Country recipient distribution of sector level FDI



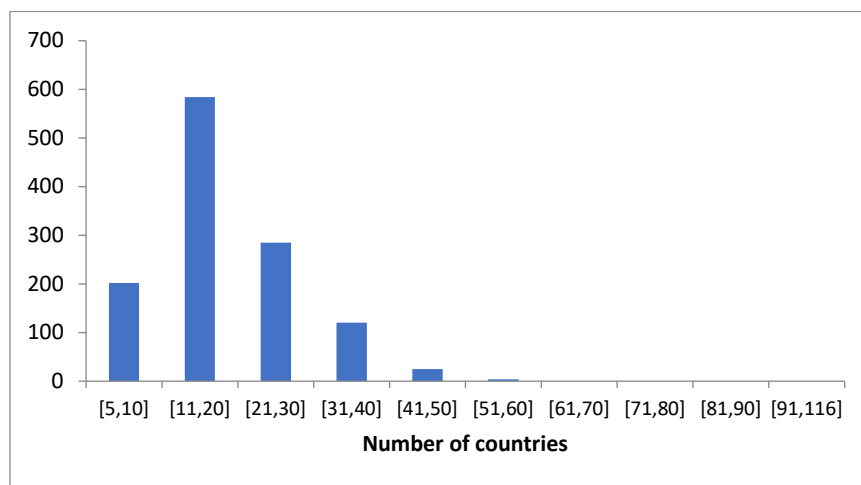
Note: the histogram reports the number of countries for each of the 245 sector/activity pairs which received FDI inflows over the 2015-19 period.

The same steps are used with the trade data. The initial data set included information on RCA for 1,245 4-digit sectors. We excluded sectors which have fewer than 5 non-zero observations for the sample countries, leading to a final number of 1,220 4-digit sectors. We transform the data set and construct a dummy variable $Y_{i,j}$ which takes value 1 if the country j has RCA in sector i above 1 and 0 otherwise. The RCA is calculated based on the sum of export and import flows over the 2015-19 period. Analogously, we create a variable $Z_{i,j}$ based on the information on country/sector pair RCA "success" over the 2010-14 period. Like Figure 1, Figure 2 shows the histogram of the empirical distribution of number of

²⁰ There is the possibility of reverse flows (i.e., an outflow investment is made and then disinvested) during the period, but if any at all, these are likely to be negligible relative to the overall sum of investment or trade flows.

countries that exhibit RCA above one (“success”) in any of the 1,220 sectors. Even more so than with FDI, the distribution is heavily skewed toward sectors in which only a subset of countries (the left-hand side of the Figure 2) exhibit revealed comparative advantage. In contrast to FDI flows and in line with economic intuition, the maximum number of countries with RCA in a sector over the 2015-19 period is only 58. In the median sector (by the number of countries with RCA) roughly 17 countries exhibit a revealed comparative advantage.

Figure 2. Country distribution of sector level RCA

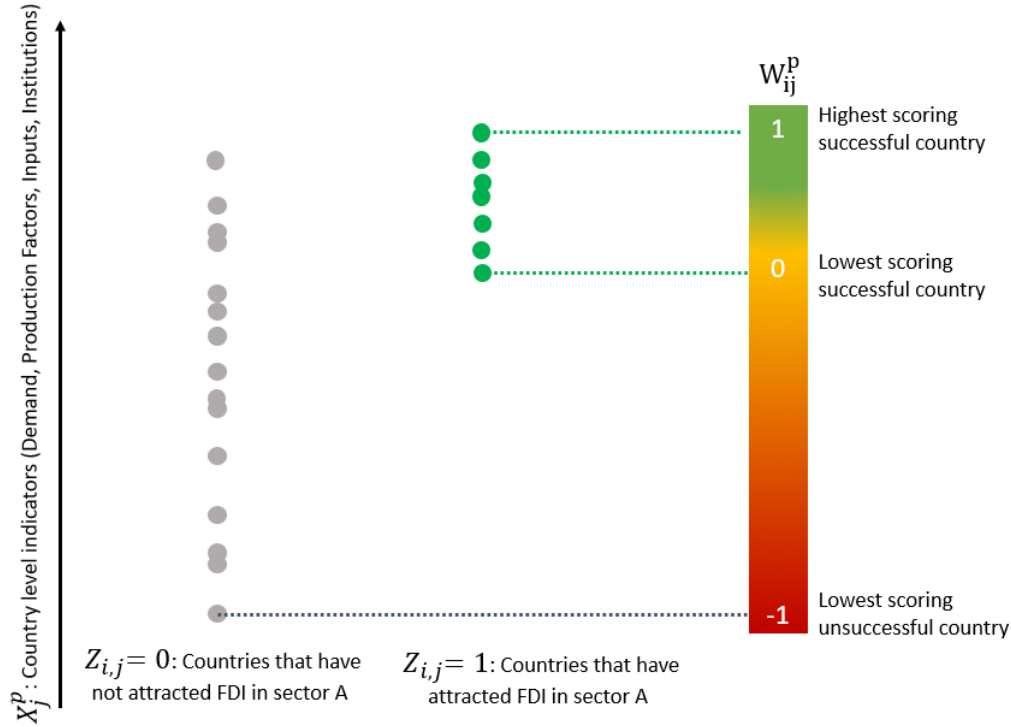


Note: the histogram reports the number of countries with RCA>1 in each of 1220 hs4 sectors over the 2015-19 period.

Finally, to build a sector-level benchmark indicator, we construct the sector level values of the indicators. The sector-level value of the p -th indicator, $W_{i,j}^p$ is obtained as a nonlinear transformation of the original country level indicator, X_j^p , using the information on the country and sector level FDI (or trade) performance over the corresponding 2010-14 period, $Z_{i,j}$: $W_{i,j}^p = h(X_j^p, Z_{i,j})$ where function $h(*)$ applies a max-min transformation.

Specifically, W_{ij}^p is a synthetic variable constructed by exploiting the full cross-sectional information on 1) the values of the variable p for all countries in the sample, and 2) success (in FDI or trade outcomes) of all countries in the given sector i . It should be noted that potential endogeneity or overfitting effects from the inclusion of the FDI outcome indicator in the construction of this variable should be limited to the extent that any individual country’s position influences the distribution of FDI (trade) success across all countries, which can be expected to be minimal. The frontier is defined as the minimum indicator threshold where success (trade or investment) is observed. We applied the max-min transformation, such that W_{ij}^p is defined in the range $[-1,1]$: -1 is assigned if the country j was not successful in sector i over the 2010-14 period ($Z_{i,j} = 0$) and has the lowest value of X_j^p in the cross-section; 1 is assigned if the country j was successful over this period ($Z_{i,j} = 1$) and has the highest value of X_j^p in the cross-section. The realized value of W_{ij}^p between these extremes depends on the relative value of X_j^p with respect to the lowest value of the indicator among the successful countries (“the minimum threshold”, where higher value of the indicator is associated with higher likelihood of success), scaled by the distance between the highest value of the indicator and the minimum threshold. Figure 3 illustrates the construction of these sector-level indicators.

Figure 3: Illustration of sector-level benchmark indicator



In this way, we generate sectoral variation in the indicators and transform data to the comparable scale for all potential indicators. The transformation profiles the value of each indicator for each country and sector pair with respect to the relative performance of that country *vis-à-vis* all other countries in FDI (or trade) success in a given sector.

The intuition behind this approach is two-fold: First, what matters in determining investment or export success is a country's *relative* endowment in factors and policies compared to potential competitors, rather than their level. Second, depending on their specific requirements and production functions, industries may respond differently to these factors so their impact on investment and export success is sector specific.

This, presumably, could offer more information for the estimation of the parameters of interest. The sector specific values of explanatory variables are also required for the production of the benchmarking heatmap, which was an integral part of this project.

Econometric approach

We use a panel binary response single index model with additional unobserved sector and country fixed effects for estimating which factors are associated with trade and investment outcomes:

$$Y_{i,j} = g\left\{\alpha_i + \gamma_j + \sum_{p=1}^K \beta_p W_{i,j}^p \geq \varepsilon_{i,j}\right\}, \varepsilon_{i,j} \mid W_{i,j,\alpha,\gamma} \sim F_{\varepsilon}; i = 1, 2, \dots, N; j = 1, 2, \dots, M \quad (1)$$

where $Y_{i,j}$ is the success variable (FDI or trade outcomes over the 2015-19 period), $W_{i,j}^p$ is the policy indicator (average value over the 2010-14 period), K is the number of potential indicators (190), function $g(*)$ is the zero/one indicator function and we assume that error distribution F_{ε} is a logistic cumulative

distribution function (CDF).²¹ In addition, we allow for the unobserved country and sector characteristics which can have a confounding effect on the outcome (FDI or trade success), captured by parameters α_i and γ_j . The country and sector fixed effects proxy for individual heterogeneity across both dimensions (the ubiquity of sectors across countries and countries diversity across sectors), which if not controlled for can potentially bias the estimates.

We propose the following algorithm to estimate the unknown parameters and conduct inference on the estimated parameters which, in essence, extends recent work in two separate strands of the literature – estimation and inference in high dimensional generalized linear models (Fei and Li, 2021) and estimation and inference in nonlinear panel data models with large dimensions (Fernandez-Val and Weinder, 2016):

- 1) Randomly split the data (with no replacement) in two groups N_1 and N_2 . Obey the panel structure during the splitting, such that the split is made only along one of the panel dimensions dimension (say, N).
- 2) Apply a machine learning model (with fixed effects as the nuisance parameters) to predict $Y_{i,j}$ on N_1 part of the sample. Based on the obtained results, select the vector W^Q of the Q best predictors among $W_{i,j}^p$ (a Q -dimensional subset of the K -dimensional set of all indicators).
- 3) Estimate coefficient for each variable $W_{i,j}^p$ by estimating separate nonlinear panel models defined in (1) on N_2 part of the sample for each variable and using W^Q from the step 2 as the covariates in these estimations. Apply analytic bias correction from Fernandez-Val and Weidner (2016) to the coefficient estimate. Save the results.
- 4) Repeat steps 1-3 B times.
- 5) Estimate $\widehat{\beta}_p$ as the mean from all B random splits:
$$\widehat{\beta}_p = \sum_{b=1}^B \widehat{\beta}_{p,b} \quad (2)$$
- 6) Estimate standard errors and the corresponding confidence intervals using the distribution of the estimated coefficients and nonparametric delta method (Efron, 2014).

The outlined algorithm controls for three types of bias in the estimates of the parameters of interest, which may be present in large-dimensional nonlinear panel data with individual effects framework. By selecting a relatively large Q in step 2, the "under-fitting" bias (the probability that the machine learning model may not encompass the true model) is minimized and asymptotically negligible. However, selecting a too large number of variables in the model selection phase leads to the "over-fitting" bias (inclusion of non-relevant variables) which can be minimized using repeating sample splitting in step 4 (Wang et al, 2020, apply similar idea in a different context). By repeating the steps multiple times, the arbitrariness in the sample splitting is avoided and the potential loss of efficiency from using only the half of the sample for the estimation of the parameters of interest is minimized. Finally, nonlinear panel data models with fixed effects suffer from the "incidental bias" problem which is addressed by using the analytic bias correction from Fernandez-Val and Weidner (2016) in step 3. The proposed algorithm is very flexible and allows the researcher to use any popular machine learning algorithm in step 2. In estimations we use

²¹ Please note that we use the logistic CDF in relation to binary character of our dependent variable. The econometric approach outlined here can be straightforwardly extended to include count type of dependent variable (Poisson) or continuous variable.

random forest and gradient boosting algorithms in step 2, with data driven (cross-validation with 5 folds) choice of the hyper-parameters. We set $Q \sim K/2$ and repeat the splits $B=1000$ times.

The empirical approach is tailored to our key interests: i) implementing a methodology that allows feature selection based on the specification (1) while producing the statistical inference (confidence intervals) for *all* parameters of the selected variables; ii) verifying economic consistency of the estimated parameters by examining the sign of the estimated parameters.

Modern machine learning tools, as an alternative to our approach, can be efficiently used to capture hard-to-observe patterns in the data and generate more accurate predictions of the variables of interest. However, the existing machine learning literature dominantly focuses on prediction improvements and is just starting to provide measures of inherent uncertainty around the variables included in the machine learning models (see Wager and Athey, 2018; Farrell et al, 2021), which is the key required component of our approach.

This problem of high-dimensional variable selection has been an active area of research in statistics and econometrics over the past 25 years, dating back to original proposal of Lasso (Tibshirani, 1996) and the number of other sparse (regularization) estimators that followed (Zhou and Hastie, 2005, and the subsequent literature). These estimators produce the first step of model/variable selection by identifying a subset of variables, while setting the value of coefficients to zero for the remaining variables. Nevertheless, performing statistical inference on the (non-zero) parameters estimated by lasso and/or the related estimators (the second step of the model/variable selection) has been challenging given the irregular character of the estimators. The recent literature has taken four general approaches to inference in high dimensional framework. One stream of the literature focuses on a particular type of the high-dimensional (or machine learning) estimator and provides methods for valid inference conditional on application of the chosen estimator (Lasso has been the most popular choice). Early examples are de-biased or de-sparsified lasso estimators in the linear regression context (Zhang and Zhang, 2014, Javanmard and Montanari, 2014; Zhang and Cheng, 2017) and in the generalized linear model context (van de Geer et al., 2014). However, to the best of our knowledge no existing method of inference is available in the nonlinear panel data framework like ours. The second stream of the literature centers on inference on the parameters of the selected model, conditional on particular model's selection. Examples in this strand of the literature are Lockhart et al. (2014), Fithian et al. (2014), Lee et al. (2016), Tibshirani et al. (2016). The third strand of the literature focuses on providing valid inference uniformly over the methods used for model selection (Berk et al., 2013, Rinaldo et al., 2019). In these papers the inference is not provided for the individual parameters, but rather for some transformation of the underlying parameters (the best linear projection of the dependent variable or the leave-out-covariates projection parameters). If the researcher is interested only in specific parameter related to the intervention variable of interest (rather than identifying the subset of key variables and conducting inference on them), the post-double selection estimator of Belloni et al. (2014) and related works provides alternative inference approach. None of these approaches hence would enable recovering the key interests of our project.

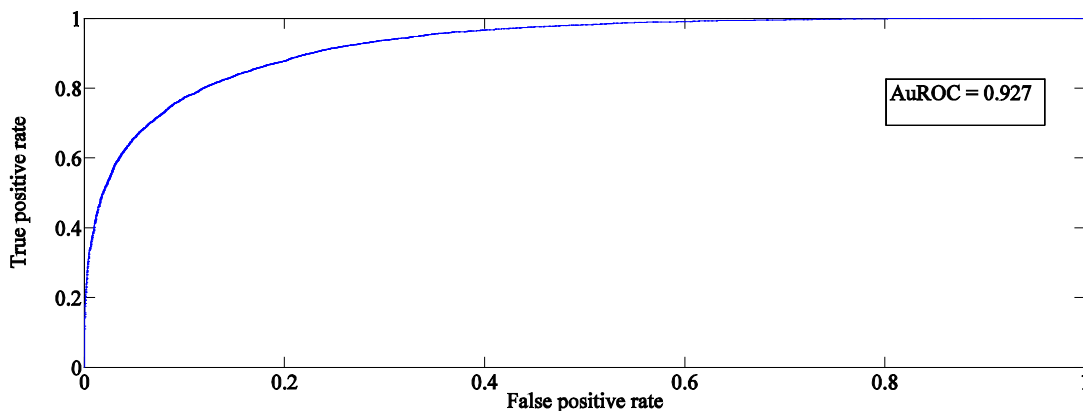
Empirical Results

Foreign direct investment

Before proceeding to the estimates, we evaluate the overall classification performance of the empirical model using the receiver operating characteristic (ROC) curve. The ROC curve plots the share of model's correct classification of positive FDI flows in total number of sectors with FDI flows (true positive rate) versus the percentage of incorrect signals of positive FDI flows out of all sectors without FDI flows (false positive signals), along varying probability cutoff levels. The relative quality of the model's empirical performance can be observed (and estimated) by looking at the area under the curve. If the model unveils associations with the FDI performance, the ROC curve should lie closer to the top left corner. The corresponding area under the ROC curve (AuROC) for an informative signal is then strictly above 0.5, which is the value for an uninformative signal.

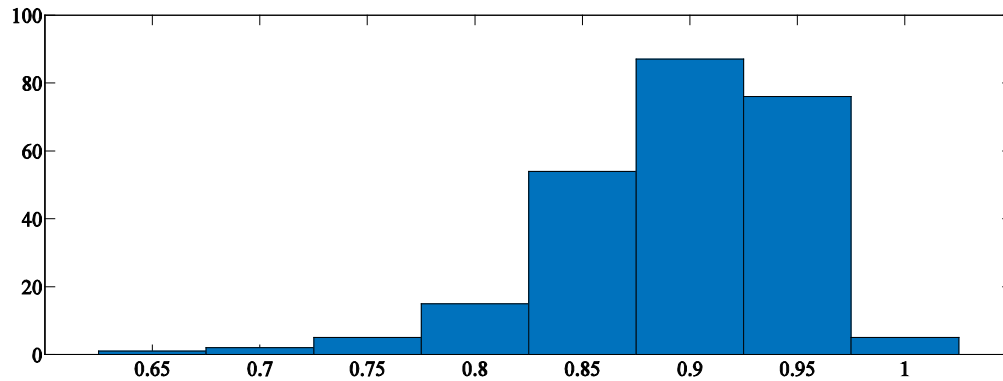
Figure 4 shows very good empirical performance of the estimated model. The ROC curve is close to the top left corner and AuROC's value of 0.93 is close to the maximum attainable level. To understand better the model's performance, we also construct sector level values of AuROC. Figure 5 shows the distribution of the sector level AuROC values. Even though the model is not estimated at the individual sector level, the model is highly informative about the sector's empirical performance, as all AuROC values are strongly above 0.5. As expected, we observe some heterogeneity in sector level performance, nevertheless for majority of sectors the estimated value of AuROC is above 0.8.

Figure 4. ROC curve: FDI data



Note: The figure reports estimated receiver operating characteristic (ROC) curve. Total positive rate is the share of model's correct classification of positive FDI flows in total number of sectors with FDI flows. False positive rate is the share of model's incorrect signals of positive FDI flows in total number of sectors without FDI flows. The curve is plotted for varying values of threshold probability (threshold probability determines whether the observation is classified as a signal of positive or zero FDI flows). Estimation and classification are performed over the in-sample 2015-19 period. Estimated area under the ROC curve (AuROC) equals 0.927 and significantly differs from 0.5.

Figure 5. Distribution of sector level values of AuROC: FDI data



Note: The figure reports histogram of estimated AuROC values for each sector. Estimation and classification are performed over the in-sample 2015-19 period.

Econometric results meet expectations: the model picks a comprehensive set of non-sector specific variables that can be understood as impacting various costs, as well as demand drivers that will impact future investment returns. This confirms the initial intuition behind the empirical set up: the dataset was built by identifying, building on literature and expert knowledge, factors associated with firm decision to invest that could be linked to economy-wide indicators of demand, as well as the costs of factors, essential infrastructure, the efficiency of institutions in providing product market rules, a stable macroeconomic environment and a political stability. Heuristically, this approach is akin to an expected country-risk-adjusted profit function.

Because results only capture economy-wide effects, they cannot be construed as identifying all the variables (let alone the most important ones) associated with investment decisions since sector-specific dimensions which are controlled for in the regression may very well be the paramount criteria. For instance, the availability of arable land for investments into industrial agricultural production, or the existence of specific skills in many industries, etc. Therefore, a correct reading of the results is that the model identifies factors primarily associated at the economy-wide level with investment decisions or trade success, considering available sector-level information.

Table 2 reports the results for variables with statistically significant estimated coefficients (at the 10% significance level). For parsimony we do not report the estimated coefficients for the full set of potential indicators (190). Following the discussion in the data preparation section, note that we construct the sector level values of each indicator as a nonlinear transformation of the original country level indicator using the information on the country and sector level FDI performance over the corresponding 2010-14 period. Following the CBT methodology, the transformation profiles the value of each indicator for each country and sector pair with respect to the relative performance of that country vis-à-vis all other countries that have recorded success in attracting FDI in a sector.

Before the transformation, we profile each variable such that higher values of the variable imply positive effect on the likelihood of receiving the FDI in a given sector. Indeed, not all variables reflect conditions that create positive incentives for investments: while some variable scores (e.g., GDP growth or

availability of key inputs) indicate the existence of favorable dimensions for successful investments, others convey the higher costs of an investment climate. To ease the interpretation of the estimated coefficients, therefore we assign a positive or negative sign to each value depending on the assumed theoretical direction of influence of a given variable on the probability of successful investment, such that higher values of all included variables imply positive effect on the likelihood of receiving the FDI.

Table 2. Results for FDI data

Group	Variable	Beta	Standard deviation	Transformation sign
1	GDP growth - 5Y forecast	0.931	0.319	1
1	Final consumption expenditure (% of GDP)	2.058	0.375	1
2	Labor force, total	0.697	0.382	1
2	Tertiary education enrollment, gross %	0.849	0.498	1
2	Labor tax and contributions (% of profit)	0.928	0.341	-1
3	Urban population growth (annual %)	0.662	0.436	1
4	Industry (including construction), value added (annual % growth)	1.908	0.733	1
4	Scientific and technical journal articles	0.784	0.269	1
4	Availability of scientists and engineers, 1-7 (best)	0.766	0.473	1
5	Total natural resources rents (% of GDP)	0.697	0.242	-1
5	Forest rents (% of GDP)	0.604	0.179	-1
5	Mineral rents (% of GDP)	0.590	0.202	-1
5	Crude Oil Proved Reserves (Billion Barrels)	0.495	0.219	1
6	Logistics performance index: Ability to track and trace consignments (1=low to 5=high)	0.998	0.617	1
7	Soundness of banks, 1-7 (best)	1.006	0.223	1
8	Profit tax (% of commercial profits)	0.719	0.458	-1
8	Total tax and contribution rate (% of commercial profits)	0.818	0.489	-1
8	Time to pay taxes (hrs/year)	0.616	0.284	-1
8	Registering property: Cost, % of property value	0.881	0.284	-1
9	Commencement of proceedings to resolve insolvency index (0-3)	0.799	0.342	1
9	Judicial independence, 1-7 (best)	0.942	0.609	1
10	Cost to Import: Documentary Compliance (USD)	0.423	0.173	-1
10	Cost to Export: Border Compliance (USD)	0.936	0.298	-1
10	Cost to Export: Documentary Compliance (USD)	0.725	0.302	-1
10	Prevalence of trade barriers, 1-7 (best)	1.033	0.639	1
10	Tariff rate, most favored nation, simple mean, all products (%)	1.031	0.519	-1
10	General government final consumption expenditure (% of GDP)	2.423	0.853	-1
11	Inflation, consumer prices (annual %)	1.281	0.503	-1
11	General government net lending/borrowing, Percent of GDP	1.697	0.317	1
11	General government gross debt, Percent of GDP	0.656	0.352	-1
11	Political Stability No Violence	1.613	0.416	1
11	Business costs of crime and violence, 1-7 (best)	0.672	0.345	1
12	Foreign Direct Investment: Inward stock (USD per capita)	0.543	0.316	1

Note: The table reports the results for variables with statistically significant estimated coefficients (at the 10% significance level). The first column profiles the variables with respect to groups outlined in Table 1. The results are obtained using the outlined methodology with country/sector panel data. The dependent variable is the binary indicator of FDI performance over 2015-19 period. The regressors are the policy indicators (average value over the 2010-14 period).

The model picks at least one variable in each of the key pre-identified categories (table 2). This suggests that there is empirical validity that country-level measures of performance corresponding to these categories have a relevance to assessing the investment climate. This broadly validates approaches like the Doing Business and Competitiveness indicators.

One surprise is the relative absence of indicators linked with infrastructure (only one indicator of transport connectivity) and financial services. In the specific context of the data used for the estimation, this may be explained by the fact that large investments (FDI captured in the database are presumably the decision of large investors) may be able to internalize these types of constraints. Another point of note is that the FDI data sample (unlike the trade data one) includes services and goods sectors and that we cannot exclude that results would vary in a specification differentiating goods and services sectors.

Natural resources rents pose a specific conundrum as they may influence market readiness for investments in opposite directions. Resource rents are obviously specific to resource sectors and play positively on FDI prospects in these sectors. They also impact other sectors, with a mix of potentially positive and negative effects. Resource windfall through induced demand effect, potential Dutch disease and resource curse effects that should favor non-trade sectors and thus favor market seeking FDI. However, these effects also come at the expense of efficiency in other productive sectors thus detracting potential investment in these. Resource-curse effects may negatively affect the overall attractiveness of investing in a country.

Because of the importance of investments that are linked to extractive industries (in poorer countries in particular), we ran a specification of the model omitting investments in the resource-based sectors (mining and oil sectors). We find that that results are robust to the omission maintaining an overall good fit for the model, and with most variables being stable to the specification. The model also assigns the same significance to rent variables in both specifications. Taken together these results suggest that the drivers of investments are not determined by investment behavior in the extractive sectors and that the presence of extractive resources do have a significant impact on investments in the rest of economic sectors.²²

Turning to the variables that are associated with extractive rents, results find that resource rents as share of the overall economy tend to have a depressing effect on the likelihood of investment. However, this is not the case for oil reserves which are on the contrary linked with higher likelihood of investment.

As would be expected, demand size dimensions are linked to investment decisions, and have higher magnitude of the coefficients (positive for Final consumption expenditure as % of GDP) as is the involvement of government in markets (negative for General government final consumption expenditure as % of GDP) which suggests that this factor having more relevance than others. Interestingly, the coefficient for industry value added coefficient is also higher. We take this variable as an indication of technological development capacity, which thus would suggest that more sophisticated markets offer more possibilities for economic returns than less sophisticated ones.

Finally, there is generally a positive signaling effect created by other investments (measured by the stock of FDI). While it is intuitive to think of realized investments as indeed signaling the availability of economic opportunities, and in line with the literature findings (see e.g., Walsh and Yu, 2010). The

²² However, this does not imply that dimensions that matter for the extractive sector are identical to those associated with non-extractive sectors.

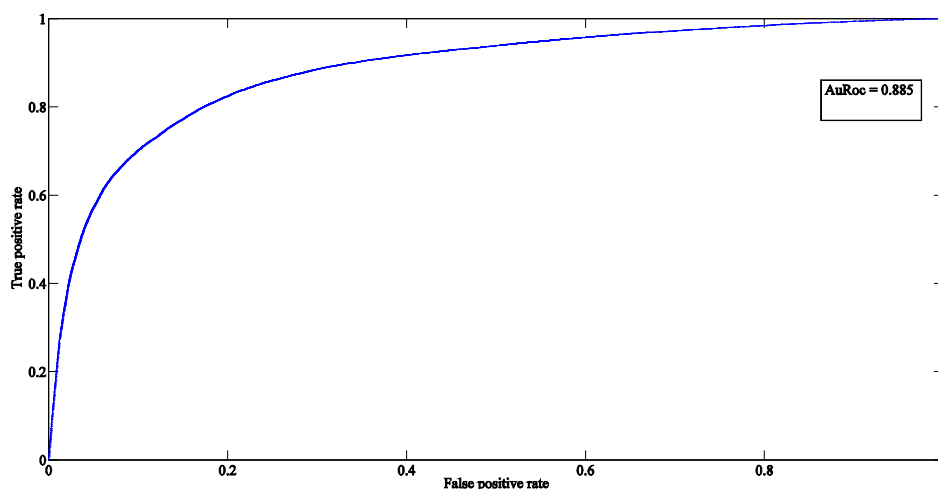
coefficient is positive in both regression (with and without rent sectors) but interestingly much higher in the regression without the rent sectors, confirming the value of the signaling effect when the nature and size of economic rents is harder to figure out.

Since the model allows for the computation of an estimate of the impact on probability of success of investment and constructing a metric of this impact at any given measurement for each significant variable, it becomes possible to construct an aggregate metric of the probability of successfully attracting FDI for each sector *at the level of performance observed* for each country. We derive from these results a simple metric that produces an overall score for countries (Annex 1), reflecting their overall performance across the significant variables and how these relate to the likelihood of attracting investment (a high score representing higher likelihood).

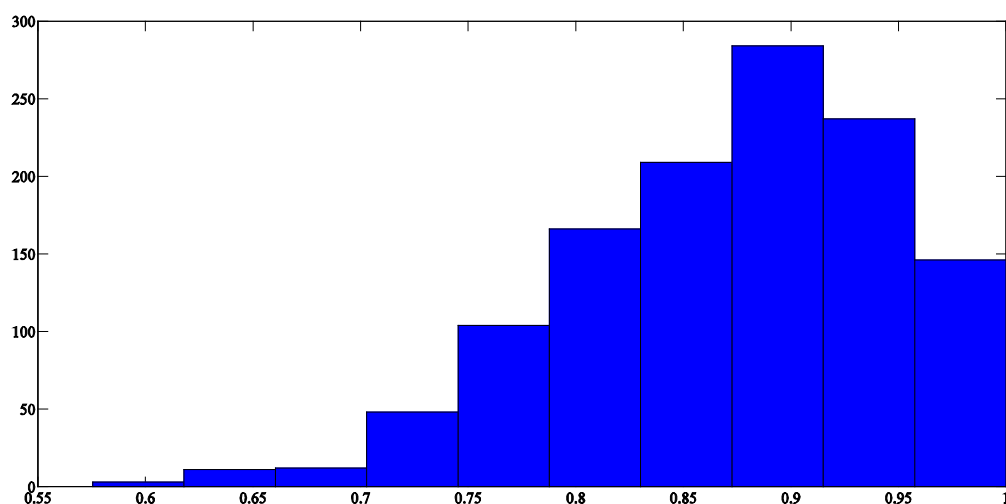
Revealed comparative advantage

Analogous to the FDI analysis, we first examine the overall classification performance of the empirical model using the receiver operating characteristic (ROC) curve. Figure 6 shows very good empirical performance of the estimated model for the trade flows too. The ROC curve is close to the top left corner and AuROC's value of 0.88 is close to the maximum attainable level. Figure 7 confirms that the model is highly informative about sector's empirical performance, as all AuROC values are strongly above 0.5.

Figure 6. ROC curve: RCA data



Note: The figure reports estimated receiver operating characteristic (ROC) curve. Total positive rate is the share of model's correct classification of RCA in total number of sectors with RCA. False positive rate is the share of model's incorrect signals of RCA in total number of sectors without RCA. The curve is plotted for varying values of threshold probability (threshold probability determines whether the observation is classified as a signal of positive or zero RCA). Estimation and classification are performed over the in-sample 2015-19 period. Estimated area under the ROC curve (AuROC) equals 0.885 and significantly differs from 0.5.



Note: The figure reports histogram of estimated AuROC values for each sector. Estimation and classification are performed over the in-sample 2015-19 period.

The specification using trade RCA data selects a larger number of significant variables than in the case of FDI. Table 3 reports the results for variables with statistically significant estimated coefficients (at the 10% significance level). We find a very strong overlap between dimensions associated with attractiveness to FDI and successful trade performance: 26 of 33 variables are common to both specifications. This is expected given the strong complementarities between investment and trade, noting also that FDI and trade can move in opposite direction and be substitute (the market-seeking argument). The overlap suggests strong complementarities in policies dominate over policies that may for instance favor infant industry type approaches (limiting trade to attract more domestic based production).

Differences between the RCA specification and earlier results for FDI are that network infrastructure (ICT, road transport) and financing variables become significant. There may be several reasons for this result. First it may be more complicated to internalize infrastructure services on an international basis (as opposed to more domestic focused in the case of FDI), and infrastructure services are key to external connectivity. Second, trade success englobes not only large players but also smaller firms who may be more reliant on external supply of key inputs.

Another expected result is that trade policies take more importance. The model picks an additional trade policy related variable in that belonging to an FTA is significantly associated with export success.

The sign of total natural resource rents changes direction when compared to FDI (but not mineral, coal or natural gas rents). In the context of trade it is generally admitted that natural resources abundance can lead to Dutch Disease type effect, an anti-export bias. At the same time resource windfall may grow the size of the domestic market and thus play a positive role in attracting market seeking FDIs. This would explain the change in sign between the two specifications.

Table 3. Results for Export Revealed Comparative Advantage (4-digit product level)

Group	Variable	Beta	Standard deviation	Transformation sign
1	GDP growth - 5Y forecast	4.103	0.619	1
1	Final consumption expenditure (% of GDP)	3.333	0.453	1
2	Employment to population ratio, 15+, total (%) (modeled ILO estimate)	2.150	0.698	1
2	Labor force participation rate, total (% of total population ages 15-64)	2.672	0.679	1
2	Labor force, total	2.148	0.610	1
2	Cooperation in labor-employer relations, 1-7 (best)	2.411	0.607	1
2	Country capacity to retain talent, 1-7 (best)	1.090	0.564	1
2	Flexibility of wage determination, 1-7 (best)	2.645	0.281	1
2	Hiring and firing practices, 1-7 (best)	2.126	0.681	1
2	Primary education enrollment, net %	0.609	0.356	1
2	Labor tax and contributions (% of profit)	3.532	0.507	-1
2	Pay and productivity, 1-7 (best)	2.260	0.409	1
3	Urban population (% of total)	1.213	0.556	1
3	Urban population growth (annual %)	5.803	0.960	1
3	Arable land (% of land area)	3.335	0.497	1
3	Water resources: total internal renewable water resources	1.154	0.451	1
3	Total renewable water resources per inhabitant (billion cubic meters)	1.605	0.499	1
4	Industry (including construction), value added per worker (constant 2010 US\$)	1.755	0.532	1
4	Industry (including construction), value added (annual % growth)	1.802	0.554	1
4	Gross fixed capital formation (% of GDP)	2.757	0.648	1
4	Scientific and technical journal articles	1.128	0.490	1
4	PCT patents, applications/million pop.	2.378	0.436	1
4	Availability of scientists and engineers, 1-7 (best)	2.049	0.822	1
4	Company spending on R&D, 1-7 (best)	1.871	0.560	1
4	Number of persons engaged (in millions)	1.960	0.604	1
5	Getting electricity: Cost, % of income per capita	-0.835	0.255	-1
5	Total natural resources rents (% of GDP)	-0.556	0.288	-1
5	Coal rents (% of GDP)	0.314	0.167	-1
5	Mineral rents (% of GDP)	0.379	0.219	-1
5	Natural gas rents (% of GDP)	0.701	0.179	-1
5	Electricity Production (kWh/capita)	1.647	0.528	1
5	Crude Oil Proved Reserves (Billion Barrels)	2.750	0.651	1
6	Quality of roads, 1-7 (best)	2.094	0.532	1
6	% of population covered by a mobile-cellular network	0.460	0.274	1
6	% of population covered by at least a 4G mobile broadband network	1.854	0.471	1
6	Mobile network performance score	0.737	0.354	1
7	Domestic credit to private sector by banks (% of GDP)	1.737	0.806	1
7	Private credit bureau coverage (% of adults)	3.114	0.649	1
7	Getting credit - Distance to Frontier (DB15-19 methodology)	2.659	0.667	1
7	Financing through local equity market, 1-7 (best)	1.546	0.632	1
7	Soundness of banks, 1-7 (best)	1.803	0.967	1

Group	Variable	Beta	Standard deviation	Transformation sign
8	Profit tax (% of commercial profits)	3.817	0.523	-1
8	Total tax and contribution rate (% of commercial profits)	1.088	0.589	-1
8	Time to pay taxes (hrs/year)	1.030	0.314	-1
8	Registering property: Cost, % of property value	1.687	0.384	-1
8	Burden of government regulation, 1-7 (best)	2.154	0.417	1
9	Commencement of proceedings to resolve insolvency index (0-3)	1.369	0.402	1
9	Enforcing contracts: Cost, % of claim	1.310	0.349	-1
9	Resolving insolvency: Recovery rate, cents on the dollar	0.714	0.358	1
9	Strength of investor protection, 0-10 (best)	1.993	0.423	1
10	Cost to Export: Border Compliance (USD)	1.252	0.448	-1
10	Cost to Export: Documentary Compliance (USD)	0.753	0.274	-1
10	Extent of market dominance, 1-7 (best)	1.935	0.926	1
10	Tariff rate, most favored nation, weighted mean, manufactured products (%)	1.308	0.662	-1
10	Tariff rate, most favored nation, simple mean, all products (%)	2.258	0.693	-1
10	General government final consumption expenditure (% of GDP)	2.056	0.510	-1
11	Inflation, consumer prices (annual %)	-1.058	0.597	-1
11	General government net lending/borrowing, Percent of GDP	3.299	0.653	1
11	General government gross debt, Percent of GDP	1.723	0.433	-1
11	Voice and Accountability	2.181	0.540	1
11	Business costs of crime and violence, 1-7 (best)	2.161	0.389	1
11	Public trust in politicians, 1-7 (best)	2.243	0.470	1
12	Foreign Direct Investment: Inward stock (USD per capita)	1.862	0.701	1
12	FTA in % of world gdp	2.749	0.744	1

Note: The table reports the results for variables with statistically significant estimated coefficients (at the 10% significance level). The first column profiles the variables with respect to groups outlined in Table 1. The results are obtained using the outlined methodology with country/sector panel data. The dependent variable is the binary indicator of RCA over 2015-19 period. The regressors are the policy indicators (average value over the 2010-14 period).

Lastly, one variable is significant but does not behave as expected: the level of inflation. As noted earlier, we assigned the model how to interpret each independent variable expected influence on the probability of success. In the case of inflation, this assumed that higher levels of inflation are a sign of macroeconomic imbalances which would tend to discourage investments (see for instance Schneider and Frey, 1985). It is possible that inflation relation to RCA success is non-linear. Moderately high level of inflation may be a sign of strong growth while higher levels betray deeper macroeconomic issues.

Sector benchmarking: revealed constraints

A shortcoming of most measures of the trade and investment climate determinants is that they are available at the economy-wide level rather than at the industry or firm level. The reasons for this range from the nature of variables which are not necessarily sector specific to measurement limitations and data availability.

We build here on the approach of the constraints benchmarking tool to derive a country and sector specific performance metric for the significant variables identified by the model. Each variable for any

country/sector (for investments) or country/product (for trade) pair is scored against the panel of country/sector or product pairs where investment or trade success is observed. The intuition behind this being that investment and trade success reveal the conditions under which they become possible, and therefore a good policy strategy would be to try to meet *at minima* similar conditions to hope for a similar outcome. The results are presented on a simple Excel spreadsheet that allows to read an assessment of the country performance on each of the variables selected by the model.

There are different ways to build a benchmark. A simple way, as was done for the CBT, would be for each variable and each sector (product) to observe where a country ranks relative to successful peers. One way to do this would be to use quantiles defined according to performance (from worst to best) and score each country's performance in a variable/sector (product) according to which quantile it would fall into. Another method would be to use an indexing measure such as the Distance to Frontier score (a normalized measure of the percentage of the best score) such as used in the Doing Business rankings to see how far a country is from the best practice. Thus, it would become possible to assess for each sector how a country fares along each variable relative to countries with successful outcomes.

These methods, however, would only use the model for the selection of the significant variables but leave aside the information that the model produces on the effect of each variable score on the likelihood of investment and RCA success. We therefore take an approach which utilizes the model insights to calculate benchmarks. Our goal is to produce categorical scores (from 1: worse to 5: best) for each indicator previously selected by the model and relative to which each country's performance on this indicator can be benchmarked. To construct scores, we must define four thresholds which map the entire universe of values for each indicator into the five defined categories (bins). Once the thresholds W^p are defined, they are transformed back to the original units of measurements for each variable.

The scores are obtained using the concept of accumulated local effects (ALE, Apley and Zhu, 2019). Based on the estimates from the econometric model, ALE provides a measure of the impact of the variable (evaluated at any point) on the probability of success (receiving FDI or having RCA) once the effects of all other variables are integrated out. For each variable, we simulate a large number of realizations (2,000) and for each of them compute the respected accumulated local effect. We then need to map the set of simulated effects on the success probability (ALE) into five scores (bins). Akin to credit risk scores, the goal is to produce the scores which comprehensively capture the similar profile of the probability of success within the score (within score homogeneity) and, at the same time, differ substantially between each other (between score heterogeneity). Hence, the mapping should provide that all points (e.g., values of W^p) in the first bin imply very low (negligible) effect on the success probability, while all points (e.g., values of W^p) in the last bin imply the maximum effect on the success probability (to the extent possible which is determined by the estimated coefficient for this variable). The values in the three bins in between the first and the fifth category imply gradually increasing probability of success. A question then is how to perform this mapping.

A simple approach is to form bins based on the quantiles of the distribution of estimated ALE (say 20th, 40th, 60th, 80th). Potential limitations with such approach are that splitting is made arbitrary (at pre-defined quantiles) and there is no guarantee that bins (scores) formed in this way will maintain the within score homogeneity and between the score heterogeneity with respect to the success probability, especially if the distribution of ALE is skewed. To alleviate these shortcomings, we use a k-means clustering approach which partitions the estimated ALE into five bins, such that within bin dissimilarity in

success probability is minimized (i.e., the within score homogeneity is achieved) while the difference between the bins is maintained (i.e., the between score heterogeneity is achieved).

The benchmarking provides a convenient way to rapidly read results at the country and sector level and help focus on where the model predicts shortcomings (or success) in terms of performance across indicators. It becomes then straightforward to present results in a simple spreadsheet format that allows for the rapid selection of results for the country, sector or product desired. The spreadsheet can also be easily updated with new estimates when new data becomes available.

The benchmarking tool could provide an information-rich decision tool to assist policy makers in targeting policy areas that would contribute to improve the country's performance in attracting FDI and participating in international trade. Similarly, the benchmarking score could help investors score country performances for the variable or the sector of interest to them.

Conclusions and suggestions for further research

This paper has proposed two contributions. First, a machine-learning based approach for the empirical use of large amounts of existing information on the environment surrounding foreign investment and trade flows. This approach offers a way to handle the large quantity of existing indicators in a numerically robust fashion and select from these a narrow and manageable subset that are shown to be associated with export success and ability to attract foreign investments. This is a first step toward a desirable objective of reducing information complexity to focus on the most relevant dimensions contributing to successful economic outcomes. Incidentally, the empirical exercise also confirms that there are large areas of commonality between predictors associated with investment and export performance, which is strongly suggestive of complementarity between investment and export flows. Second, we propose an approach to derive sector-relevant information. Economic sectors differ in terms of the nature of conditions that lead to their respective success, but data generally does not reflect these sector specificities. We find indeed measurable differences across sectors.

Finally, the research undertaken for this paper suggests a large potential agenda of future research, including alternative approaches to complexity reduction (for instance other ML algorithm techniques could help pick a subset number of variables), use of sub-sector clusters or heterogeneous coefficient panels to capture sector variations, and nested approaches for classes of explanatory variables, or use of different levels of aggregation.

References

- Apley, D., and J. Zhu, 2019. "Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models", arXiv:1612.08468 [stat.ME].
- Belassa, B., 1967. " 'Revealed' Comparative Advantage Revisited: An Analysis of Relative Export Shares of the Industrial Countries", 1953-1971, *Manchester School* 45, pp. 327-44.
- Belloni, A., V. Chernozhukov, and C. Hansen, 2014. "Inference on Treatment Effects After Selection Among High-Dimensional Controls, *The Review of Economic Studies*, 81(2), 608-650.
- Bergstrand J., and P. Egger, 2007. "A Knowledge and Physical Capital Model of International Trade Flows, Foreign Direct Investment, and Multinational Enterprises", *Journal of International Economics*, 2007, vol. 73, issue 2, 278-308.
- Berk, R., L. Brown, A. Buja, K. Zhang, and L. Zhao, 2013. "Valid Post-Selection Inference", *The Annals of Statistics*, 41(2), 802-837.
- Blonigen, B., and J. Piger, 2014. "Determinants of Foreign Direct Investment", *Canadian Journal of Economics*, Volume 47:3, pp. 775-812.
- Chinn, M., and E. Prasad, 2003. "Medium-Term Determinants of Current Accounts in Industrial and Developing Countries: An Empirical Exploration", *Journal of International Economics*, 59, pp. 47-76.
- Chinn, M., B. Eichengreen, and H. Ito, 2014. "A Forensic Analysis of Global Imbalances", *Oxford Economic Papers*, 66(2), pp. 465-490.
- Disdier, A.-C., and K. Head, 2008. "The Puzzling Persistence of the Distance Effect on Bilateral Trade". *The Review of Economics and Statistics* 90 (1), pp. 37-48.
- Djankov, S., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer, 2002. "The Regulation of Entry", *The Quarterly Journal of Economics*, Volume 117, Issue 1, pp. 1-37.
- Djankov, S., 2016. "The Doing Business Project: How it Started", *Journal of Economic Perspectives*, Volume 30:1, Correspondence, Pages 247-248.
- Djankov, S., C. Freund, and C. S. Pham, 2010. "Trading on Time", *The Review of Economics and Statistics*, 92(1), pp. 166-173.
- de Soto, H., 1989. *The Other Path*. New York: Harper and Row.
- Dollar, D., M. Hallward-Driemeier, and T. Mengistae, 2005. "Investment Climate and Firm Performance in Developing Economies", *Economic Development and Cultural Change*, 54(1), pp. 1-31
- Dollar, D., M. Hallward-Driemeier, and T. Mengistae, 2006. "Investment climate and international integration", *World Development*, 39(9), pp. 1498-1516.
- Eifert B., A. Gelb and V. Rachamandran, 2005. "Business Environment and Comparative Advantage in Africa: Evidence from the Investment Climate Data", *Center for Global Development Working Paper* Nb 56.

- Eifert, B., A. Gelb and V. Ramachandran, 2008. "The Cost of Doing Business in Africa: Evidence from Enterprise Survey Data", *World Development*, 36(9), pp. 1531-1546.
- Efron, B., 2014. "Estimation and Accuracy After Model Selection", *Journal of the American Statistical Association*, 109:507, pp. 991-1007.
- Farrell, M. H., Liang, T., & Misra, S, 2021. "Deep Neural Networks for Estimation and Inference", *Econometrica*, 89(1), pp. 181-213.
- Fithian, W., D.L. Sun, and J.E. Taylor, 2014. "Optimal Inference After Model Selection", arXiv:1410.2597 [math.ST]
- Fei, Z., and Y. Li, 2021. "Estimation and Inference for High Dimensional Generalized Linear Models: A Splitting and Smoothing Approach", *Journal of Machine Learning Research*, 22, pp. 1-32
- Fernandez-Val, I., and M. Weinder, 2016. "Individual and time effects in nonlinear panel models with large N, T", *Journal of Econometrics*, 192(1), pp. 291-312.
- Gruber J., and S. Kamin, 2007. "Explaining the Global Pattern of Current Account Imbalances", *Journal of International Money and Finance*, 26(4), pp. 500-522.
- Hausmann, R., B. Klinger, and R. Wagner, 2008. "Doing Growth Diagnostics in Practice: A 'Mindbook'", *CID Working Paper No. 177*.
- Hausmann, R., D. Rodrik, and A. Velasco, 2005. "Growth Diagnostics", Working Paper, Harvard University.
- Head, K., and T. Mayer, 2013. "Gravity Equations: Workhorse, Toolkit, and Cookbook", CEPII Working Paper.
- Head, K. and J. Ries, 2008. "FDI as an outcome of the market for corporate control: Theory and evidence", *Journal of International Economics*, 74(1), pp. 2-20.
- Hidalgo, C., B. Klinger, A.-L. Barabasi, and R. Hausmann, 2007. "The Product Space Conditions the Development of Nations", *Science*, Vol. 317: 5837, pp. 482-487.
- Javanmard, A. and A. Montanari, 2014. "Confidence Intervals and Hypothesis Testing for High-Dimensional Regression", *The Journal of Machine Learning Research*, 15(1), 2869-2909.
- Kimura, F., and HH. Lee, 2006. "The Gravity Equation in International Trade in Services". *Review of World Economics*, 142, pp. 92-121.
- Kraay, A., and N. Tawara, 2013. "Can Specific Policy Indicators Identify Reform Priorities?", *Journal of Economic Growth*, Vol. 18, pp. 253-283.
- Lee, J. D., D. L. Sun, Y. Sun, and J.E. Taylor, 2016. "Exact Post-Selection Inference, With Application to the Lasso", *The Annals of Statistics*, (44)(3), 907-927.
- Lockhart, R., J. Taylor, R., Tibshirani, R. J., and R. Tibshirani, 2014. "A Significance Test for the Lasso", *The Annals of Statistics*, 42(2), 413-468.

- Portes, R. and H. Rey, 2005. "The Determinants of Cross-Border Equity Flows", *Journal of International Economics*, 65(2), pp. 269-296.
- Pathikonda, V., and T. Farole, 2016. "The Capabilities Driving Participation in Global Value Chains". Policy Research Working Paper, No. 7804. World Bank.
- Reis, J.-G., and T. Farole, 2012. *Trade Competitiveness Diagnostic Toolkit*, World Bank.
- Rinaldo, A., L. Wasserman, and M. G'Sell, 2019. "Bootstrapping and Sample Splitting for High-Dimensional, Assumption-Lean Inference", *The Annals of Statistics*, 47(6), 3438-3469.
- Schneider, F., and B. Frey, 1985. "Economic and Political Determinants of Foreign Direct Investment", *World Development*, 13(2), pp. 161-175.
- Sekkat, K., and M-A. Venganzones-Varoudakis, 2007. "Openness, Investment Climate, and FDI in Developing Countries", *Review of Development Economics*, 11(4), pp. 607-620.
- Tibshirani, R., 1996. "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society Series B (Methodological)*, 58(1), 267-288.
- Tibshirani, R. J.E., Taylor, J., Lockhart, R. and R. Tibshirani, 2016. "Exact Post-Selection Inference for Sequential Regression Procedures", *Journal of the American Statistical Association*, 111(514), 600-620.
- van de Geer, S., P. Bühlmann, Y., Ritov, and R. Dezeure, 2014. "On Asymptotically Optimal Confidence Regions and Tests for High Dimensional Models. *The Annals of Statistics*, 42(3), 1166- 1202.
- Wager, S. and S. Athey, 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests", *Journal of the American Statistical Association*, Vol. 113, Issue 523, 1228-1242.
- Walsh, J. P., and J. Yu, 2010. "Determinants of Foreign Direct Investments: A Sectoral and Institutional Approach", *IMF Working Paper*, WP/10/187.
- Wang, J., He, X., and Xu, G., 2020. "Debiased Inference on Treatment Effect in a High-dimensional Model", *Journal of the American Statistical Association*, 115(529), pp.442-454.
- Yotov, Y., R. Piermartini, J.A. Monteiro, and M. Larch, 2015. *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model Online Revised Edition*, WTO and UNCTAD.
- Zhang, X. and G. Cheng, 2017. "Simultaneous Inference for High-Dimensional Linear Models. *Journal of the American Statistical Association*, 112(518), 757-768.
- Zhang, C. H. and S.S. Zhang, 2014. "Confidence Intervals for Low-Dimensional Parameters with High-Dimensional Data", *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 76(1), 217-242.
- Zhou, H., and T. Hastie, 2005. "Regularization and Variable Selection Via the Elastic Net", *Journal of the Royal Statistical Society Series B (Methodological)*, 67(2), 310-320.

Annex 1: Receiving FDI: country scores

Country	Score				
		Ghana	17.4%	Netherlands	66.4%
Albania	9.0%	Greece	35.0%	New Zealand	28.0%
Algeria	15.9%	Guatemala	11.7%	Nicaragua	6.0%
Angola	8.1%	Guinea	9.4%	Nigeria	25.1%
Argentina	27.1%	Honduras	12.0%	North Macedonia	6.5%
Australia	51.9%	Hungary	31.8%	Norway	14.8%
Austria	31.1%	Iceland	9.6%	Oman	4.0%
Azerbaijan	8.2%	India	79.4%	Pakistan	29.9%
Bahrain	7.1%	Indonesia	47.8%	Panama	19.4%
Bangladesh	23.0%	Ireland	73.6%	Paraguay	11.5%
Belgium	46.6%	Israel	26.2%	Peru	17.0%
Benin	4.5%	Italy	56.7%	Philippines	43.1%
Bolivia	4.0%	Jamaica	21.9%	Poland	46.1%
Bosnia & Herzegovina	16.1%	Japan	49.6%	Portugal	44.3%
Botswana	4.5%	Jordan	14.5%	Romania	46.4%
Brazil	36.6%	Kazakhstan	17.9%	Russian Federation	44.2%
Bulgaria	23.8%	Kenya	38.4%	Rwanda	13.1%
Burkina Faso	1.8%	Korea, Rep.	32.6%	Saudi Arabia	12.1%
Burundi	2.6%	Kyrgyz Republic	9.9%	Senegal	8.7%
Cambodia	25.9%	Lao PDR	12.8%	Sierra Leone	0.5%
Cameroon	3.7%	Latvia	13.5%	Singapore	49.4%
Chile	17.2%	Lebanon	4.0%	Slovak Republic	17.7%
China	66.2%	Lesotho	1.7%	Slovenia	17.2%
Colombia	37.5%	Lithuania	30.9%	South Africa	33.9%
Costa Rica	29.2%	Luxembourg	16.1%	Spain	80.1%
Côte d'Ivoire	21.8%	Madagascar	3.6%	Sri Lanka	22.6%
Croatia	22.4%	Malawi	2.2%	Sweden	25.8%
Cyprus	36.5%	Malaysia	52.9%	Tanzania	21.9%
Denmark	45.8%	Mali	1.5%	Thailand	48.2%
Dominican Republic	15.3%	Mauritania	0.8%	Tunisia	8.3%
Ecuador	10.5%	Mauritius	15.7%	Turkey	28.1%
Egypt, Arab Rep.	36.7%	Mexico	57.6%	Uganda	12.5%
El Salvador	6.8%	Moldova	3.0%	Ukraine	18.3%
Estonia	11.2%	Mongolia	3.4%	United Kingdom	89.2%
Finland	46.1%	Morocco	31.6%	United States	95.4%
France	66.8%	Mozambique	11.4%	Uruguay	7.1%
Gambia, The	1.8%	Myanmar	20.1%	Venezuela, RB	1.7%
Georgia	13.1%	Namibia	5.3%	Vietnam	58.8%
Germany	69.4%	Nepal	11.1%	Zambia	7.9%

Annex 2: Constraints benchmarking for FDI – Example for Mauritius, plastics sector

Relative Performance of Mauritius to All Successful Countries for activities in Plastics			Design, Development & Testing	Headquarters	Logistics, Distribution & Transportation	Manufacturing	Recycling	Research & Development	Sales, Marketing & Support
Mauritius' performance									
GDP growth	5y forecast	3.92	3	3	3	3	3	3	3
Final consumption expenditure	% of GDP	89.89	5	5	5	5	4	5	5
Labor force	Total	604127.60	1	1	1	1	1	1	1
Flexibility of wage determination	1-7 (best)	4.64	2	2	2	2	2	2	2
Tertiary education enrollment	gross %	38.85	3	3	3	3	3	3	3
Labor tax and contributions	% of profit	7.86	5	5	5	5	5	5	5
Urban population growth	annual %	-0.11	1	1	1	2	1	1	1
Total renewable water resources per inhabitant	billion cubic meters	2175.00	2	2	2	2	2	2	2
Industry (including construction), value added	annual % growth	1.31	2	2	2	2	1	1	2
Scientific and technical journal articles	number of articles	144.24	1	1	1	2	1	1	1
Availability of scientists and engineers	1-7 (best)	3.78	2	1	2	4	2	1	2
Local supplier quantity	1-7 (best)	4.85	2	2	3	5	2	3	3
Total natural resources rents	% of GDP	0.00	5	5	5	5	5	5	5
Forest rents	% of GDP	0.00	5	5	5	5	5	5	5
Mineral rents	% of GDP	0.00	5	5	5	5	5	5	5
Crude Oil Proved Reserves	billion barrels	14.00	5	5	5	5	5	5	5
Quality of port infrastructure	1-7 (best)	4.39	3	2	3	4	3	2	3
Ability to track and trace consignments	(1=low to 5=high)	3.00	2	2	3	4	3	3	3
Mobile tariff	0-100 (best)	50.30	3	3	3	4	3	3	3
Soundness of banks	1-7 (best)	5.48	3	5	4	5	3	4	4

Relative Performance of Mauritius to All Successful Countries for activities in Plastics			Design, Development & Testing	Headquarters	Logistics, Distribution & Transportation	Manufacturing	Recycling	Research & Development	Sales, Marketing & Support
Mauritius' performance									
Profit tax	% commercial profits	10.36	4	4	4	4	4	4	4
Total tax and contribution rate	% commercial profits	21.90	5	5	5	5	5	5	5
Time to pay taxes	hrs/year	149.60	5	3	5	5	5	3	5
Registering property: Cost	% of property value	4.60	5	5	5	5	5	5	5
Commencement of proceedings to resolve insolvency index	(0-3)	3.00	5	5	5	5	5	5	5
Judicial independence	1-7 (best)	5.03	4	4	4	5	4	4	5
Cost to Import: Documentary Compliance	USD	165.60	1	1	2	4	1	2	3
Cost to Export: Border Compliance	USD	302.60	3	3	3	4	3	3	4
Cost to Export: Documentary Compliance	USD	128.10	2	3	3	3	2	1	3
Prevalence of trade barriers	1-7 (best)	4.62	3	2	4	5	3	4	5
Tariff rate, most favored nation, simple mean, all products	%	0.88	5	5	5	5	5	5	5
General government final consumption expenditure	% of GDP	15.20	4	3	3	4	3	2	4
Inflation, consumer prices	annual %	2.01	4	2	4	5	3	4	4
General government net lending/borrowing	% of GDP	-2.59	3	4	3	4	3	3	4
General government gross debt	% of GDP	65.83	2	2	2	2	1	1	2
Political Stability No Violence	-2.5 to 2.5 (best)	0.96	5	5	5	5	5	5	5
Business costs of crime and violence	1-7 (best)	5.15	4	2	4	5	4	4	4
Foreign Direct Investment: Inward stock	% of GDP	37.64	4	3	4	4	3	3	4

Annex 3: List of variables used

Sign	Indicator name	Pillar	Source
1	Domestic market size index, 1-7 (best)	Demand	TCdata360
1	GDP growth (annual %)	Demand	WDI
1	GDP per capita (constant 2010 US\$)	Demand	WDI
1	GDP per capita growth (annual %)	Demand	WDI
1	GDP (constant 2010 US\$)	Demand	WDI
1	GDP, PPP (constant 2011 international \$)	Demand	WDI
1	GDP growth - 5Y forecast	Demand	IMF WEO
1	Final consumption expenditure (% of GDP)	Demand	WDI
1	Households and NPISHs final consumption expenditure (% of GDP)	Demand	WDI
1	Foreign market size index, 1-7 (best)	Demand	TCData360
1	Employment to population ratio, 15+, female (%) (modeled ILO estimate)	Production Factors	WDI
1	Employment to population ratio, 15+, male (%) (modeled ILO estimate)	Production Factors	WDI
1	Employment to population ratio, 15+, total (%) (modeled ILO estimate)	Production Factors	WDI
1	Wage and salaried workers, female (% of female employment) (modeled ILO estimate)	Production Factors	WDI
1	Wage and salaried workers, male (% of male employment) (modeled ILO estimate)	Production Factors	WDI
1	Wage and salaried workers, total (% of total employment) (modeled ILO estimate)	Production Factors	WDI
1	Employment in industry (% of total employment) (modeled ILO estimate)	Production Factors	WDI
1	Labor force participation rate, female (% of female population ages 15-64) (modeled ILO estimate)	Production Factors	WDI
1	Labor force participation rate, male (% of male population ages 15-64) (modeled ILO estimate)	Production Factors	WDI
1	Labor force participation rate, total (% of total population ages 15-64) (modeled ILO estimate)	Production Factors	WDI
1	Labor Market Efficiency	Production Factors	TCData360
1	Labor force, total	Production Factors	WDI
1	Cooperation in labor-employer relations, 1-7 (best)	Production Factors	TCdata360
1	Country capacity to attract talent, 1-7 (best)	Production Factors	TCdata360
1	Country capacity to retain talent, 1-7 (best)	Production Factors	TCdata360
1	Flexibility of wage determination, 1-7 (best)	Production Factors	TCdata360
1	Hiring and firing practices, 1-7 (best)	Production Factors	TCdata360
1	Primary education enrollment, net %	Production Factors	TCdata360
1	Secondary education enrollment, gross %	Production Factors	TCdata360
1	Tertiary education enrollment, gross %	Production Factors	TCdata360
-1	Labor tax and contributions (% of profit)	Production Factors	WB DB
1	Quality of educational system, 1-7 (best)	Production Factors	TCData360
1	Quality of management schools, 1-7 (best)	Production Factors	TCData360
1	Quality of math & science education, 1-7 (best)	Production Factors	TCData360
1	Pay and productivity, 1-7 (best)	Production Factors	TCdata360
1	Extent of staff training, 1-7 (best)	Production Factors	TCData360
1	Life expectancy at birth, female (years)	Production Factors	WDI
1	Life expectancy at birth, male (years)	Production Factors	WDI
1	Life expectancy at birth, total (years)	Production Factors	WDI
-1	Mortality rate, under-5 (per 1,000 live births)	Production Factors	WDI

Sign	Indicator name	Pillar	Source
1	Population, total	Production Factors	WDI
1	Population growth (annual %)	Production Factors	WDI
1	Urban population (% of total)	Production Factors	WDI
1	Urban population growth (annual %)	Production Factors	WDI
1	Arable land (% of land area)	Production Factors	WDI
1	Urban population	Production Factors	WDI
1	Total internal renewable water resources per capita (m3/inhab/year)	Production Factors	FAO
1	Water resources: total internal renewable water resources	Production Factors	FAO
1	Total renewable water resources per inhabitant (billion cubic meters)	Production Factors	FAO
1	Industry (including construction), value added per worker (constant 2010 US\$)	Production Factors	WDI
1	Industry (including construction), value added (constant 2010 US\$)	Production Factors	WDI
1	Industry (including construction), value added (annual % growth)	Production Factors	WDI
1	Industry (including construction), value added (% of GDP)	Production Factors	WDI
1	Gross fixed capital formation (% of GDP)	Production Factors	WDI
1	Gross capital formation (% of GDP)	Production Factors	WDI
1	Scientific and technical journal articles	Production Factors	WDI
1	PCT patents, applications/million pop.	Production Factors	TCData360
1	Availability of latest technologies, 1-7 (best)	Production Factors	TCData360
1	Capacity for innovation, 1-7 (best)	Production Factors	TCData360
1	Availability of scientists and engineers, 1-7 (best)	Production Factors	TCData360
1	Business Sophistication	Production Factors	TCData360
1	Company spending on R&D, 1-7 (best)	Production Factors	TCData360
1	Firm-level technology absorption, 1-7 (best)	Production Factors	TCData360
1	Local supplier quality, 1-7 (best)	Production Factors	TCData360
1	Local supplier quantity, 1-7 (best)	Production Factors	TCData360
1	Prevalence of foreign ownership, 1-7 (best)	Production Factors	TCData360
1	Production process sophistication, 1-7 (best)	Production Factors	TCData360
1	State of cluster development, 1-7 (best)	Production Factors	TCData360
1	Capital per worker (in 2011US\$)	Production Factors	Penn WT
1	Capital stock at current PPPs (in mil. 2011US\$)	Production Factors	Penn WT
1	Number of persons engaged (in millions)	Production Factors	Penn WT
1	Access to electricity (% of population)	Inputs	WDI
1	Access to electricity, urban (% of urban population)	Inputs	WDI
1	Quality of electricity supply, 1-7 (best)	Inputs	TCData360
-1	Getting electricity: Cost, % of income per capita	Inputs	TCdata360
1	Environmental Performance Index	Inputs	Yale
-1	Total natural resources rents (% of GDP)	Inputs	WDI
-1	Coal rents (% of GDP)	Inputs	WDI
-1	Forest rents (% of GDP)	Inputs	WDI
-1	Mineral rents (% of GDP)	Inputs	WDI
-1	Natural gas rents (% of GDP)	Inputs	WDI
-1	Oil rents (% of GDP)	Inputs	WDI

Sign	Indicator name	Pillar	Source
1	Electricity Production (kWh/capita)	Inputs	TCdata360
1	Crude Oil Proved Reserves (Billion Barrels)	Inputs	U.S. EIA
1	Quality of air transport infrastructure, 1-7 (best)	Inputs	TCData360
1	Quality of port infrastructure	Inputs	TCData360
1	Quality of roads, 1-7 (best)	Inputs	TCData360
1	Logistics performance index: Overall (1=low to 5=high)	Inputs	WDI
1	Logistics performance index: Efficiency of customs clearance process (1=low to 5=high)	Inputs	WDI
1	Logistics performance index: Ability to track and trace consignments (1=low to 5=high)	Inputs	WDI
1	% of population covered by a mobile-cellular network	Inputs	GSMA
1	% of population covered by at least a 3G mobile broadband network	Inputs	GSMA
1	% of population covered by at least a 4G mobile broadband network	Inputs	GSMA
1	Mobile cellular subscriptions (per 100 people)	Inputs	WDI
1	Individuals using the Internet (% of population)	Inputs	WDI
1	Mobile network performance score	Inputs	GSMA
1	Mobile download speed	Inputs	GSMA
1	Mobile tariff (based on Mobile Connectivity Index)	Inputs	GSMA
1	Mobile apps penetration: Mobile Connectivity Index	Inputs	GSMA
1	Domestic credit to private sector by banks (% of GDP)	Inputs	WDI
1	Domestic credit to private sector (% of GDP)	Inputs	WDI
1	Private credit bureau coverage (% of adults)	Inputs	WDI
1	Financial Market Development	Inputs	TCData360
1	Getting credit - Composite Distance to Frontier	Inputs	TCData360
1	Getting credit - Distance to Frontier (DB15-19 methodology)	Inputs	TCData360
1	Affordability of financial services, 1-7 (best)	Inputs	TCData360
1	Availability of financial services, 1-7 (best)	Inputs	TCData360
1	Ease of access to loans, 1-7 (best)	Inputs	TCData360
1	Financing through local equity market, 1-7 (best)	Inputs	TCData360
1	Soundness of banks, 1-7 (best)	Inputs	TCData360
1	Domestic credit provided by financial sector (% of GDP)	Inputs	WDI
1	Depth of credit information index (0=low to 8=high)	Inputs	WDI
-1	Labor tax and contributions (% of commercial profits)	Institutions	WDI
-1	Other taxes payable by businesses (% of commercial profits)	Institutions	WDI
-1	Profit tax (% of commercial profits)	Institutions	WDI
-1	Total tax and contribution rate (% of commercial profits)	Institutions	WDI
-1	Profit tax (% of profit)	Institutions	TCData360
-1	Time to pay taxes (hrs/year)	Institutions	TCData360
-1	Total tax rate (% of profit)	Institutions	TCData360
1	Ease of doing business - Composite Distance to Frontier	Institutions	TCData360
1	Ease of doing business - Distance to Frontier (DB16 methodology)	Institutions	TCData360
-1	Dealing with construction permits: Cost, % of Warehouse value	Institutions	TCData360
-1	Registering property: Cost, % of property value	Institutions	TCData360
-1	Starting a business: Cost - Men, % of income per capita	Institutions	TCData360

Sign	Indicator name	Pillar	Source
-1	Starting a business: Cost - Women, % of income per capita	Institutions	TCData360
-1	Trading across borders : Time to export: Border compliance, hours (DB16-19 methodology)	Institutions	TCData360
-1	Trading across borders : Time to import: Border compliance, hours (DB16-19 methodology)	Institutions	TCData360
1	Burden of government regulation, 1-7 (best)	Institutions	TCData360
1	Regulatory Quality	Institutions	TCData360
1	E-government development index	Institutions	UNDESA
1	2. Ethics and corruption	Institutions	TCData360
1	Wastefulness of government spending, 1-7 (best)	Institutions	TCData360
1	1st pillar Institutions	Institutions	TCData360
1	Control of Corruption	Institutions	TCData360
1	Government Effectiveness	Institutions	TCData360
1	Rule of Law	Institutions	TCData360
1	Commencement of proceedings to resolve insolvency index (0-3)	Institutions	WB DB
-1	Enforcing contracts: Cost, % of claim	Institutions	TCdata360
1	Resolving insolvency: Recovery rate, cents on the dollar	Institutions	TCdata360
1	Strength of insolvency framework index (0-16)	Institutions	WB DB
1	Strength of legal credit rights index (0-12)	Institutions	WB DB
1	Efficiency of legal system in challenging regs, 1-7 (best)	Institutions	TCData360
1	Efficiency of legal system in settling disputes, 1-7 (best)	Institutions	TCData360
1	Intellectual property protection, 1-7 (best)	Institutions	TCData360
1	Judicial independence, 1-7 (best)	Institutions	TCData360
1	Strength of investor protection, 0-10 (best)	Institutions	TCData360
-1	Cost to Import: Border Compliance (USD)	Institutions	WB DB
-1	Cost to Import: Documentary Compliance (USD)	Institutions	WB DB
-1	Cost to Export: Border Compliance (USD)	Institutions	WB DB
-1	Cost to Export: Documentary Compliance (USD)	Institutions	WB DB
1	Prevalence of trade barriers, 1-7 (best)	Institutions	TCdata360
1	Business impact of rules on FDI, 1-7 (best)	Institutions	TCdata360
1	Effectiveness of antimonopoly policy	Institutions	TCdata360
1	Effectiveness of anti-monopoly policy, 1-7 (best)	Institutions	TCdata360
1	Extent of market dominance, 1-7 (best)	Institutions	TCdata360
1	Intensity of local competition, 1-7 (best)	Institutions	TCdata360
-1	Tariff rate, applied, simple mean, manufactured products (%)	Institutions	WDI
-1	Tariff rate, most favored nation, simple mean, manufactured products (%)	Institutions	WDI
-1	Tariff rate, applied, weighted mean, manufactured products (%)	Institutions	WDI
-1	Tariff rate, most favored nation, weighted mean, manufactured products (%)	Institutions	WDI
-1	Tariff rate, most favored nation, weighted mean, all products (%)	Institutions	WDI
-1	Tariff rate, applied, weighted mean, all products (%)	Institutions	WDI
-1	Tariff rate, most favored nation, simple mean, all products (%)	Institutions	WDI
-1	Tariff rate, applied, simple mean, all products (%)	Institutions	WDI
-1	General government final consumption expenditure (% of GDP)	Institutions	WDI
1	Current account balance (% of GDP)	Institutions	WDI

Sign	Indicator name	Pillar	Source
-1	Consumer price index (2010 = 100)	Institutions	WDI
-1	Inflation, consumer prices (annual %)	Institutions	WDI
1	General government net lending/borrowing, Percent of GDP	Institutions	WDI
-1	General government gross debt, Percent of GDP	Institutions	WDI
1	Voice and Accountability	Institutions	TCData360
1	Political Stability No Violence	Institutions	TCData360
1	Business costs of crime and violence, 1-7 (best)	Institutions	TCData360
1	Public trust in politicians, 1-7 (best)	Institutions	TCData360
1	Reliability of police services, 1-7 (best)	Institutions	TCData360
1	Foreign direct investment, net inflows (% of GDP)	Other	WDI
1	Foreign Direct Investment: Inward stock (% of world)	Other	TCdata360
1	Foreign Direct Investment: Inward stock (% of GDP)	Other	TCdata360
1	Foreign Direct Investment: Inward stock (USD per capita)	Other	TCdata360
1	Exports of goods and services (% of GDP)	Other	WDI
1	Export value index (2000 = 100)	Other	WDI
1	Export volume index (2000 = 100)	Other	WDI
1	Fuel exports (% of merchandise exports)	Other	WDI
1	Food exports (% of merchandise exports)	Other	WDI
1	Agricultural raw materials exports (% of merchandise exports)	Other	WDI
1	Ores and metals exports (% of merchandise exports)	Other	WDI
1	Manufactures exports (% of merchandise exports)	Other	WDI
-1	Merchandise: Concentration and diversification indices of exports by country (Concentration Index)	Other	TCdata360
-1	Merchandise: Concentration and diversification indices of exports by country (Diversification Index)	Other	TCdata360
1	FTA in % of world GDP	Other	CEPII

Notes: IMF WEO = IMF World Economic Outlook; Penn WT = Penn World Tables; TCdata360 = World Bank TCdata360; UNDESA = UN Department of Economic and Social Affairs; US EIA = US Energy Information Administration; WB DB = World Bank Doing Business; WDI = World Bank World Development Indicators.