How to Foster Innovation: a Data-Driven Approach to Measuring Economic Competitiveness

C. Kuhlman, K. Natesan Ramamurthy, P. Sattigeri, A.C. Lozano, L. Cao, C. Reddy, A. Mojsilovic, K.R. Varshney

Abstract

Innovation is a key factor driving economic growth in countries worldwide. However, innovation is hard to define, and therefore even harder to measure. To help policy makers and business leaders better understand how to foster innovation, we need robust ways to quantify innovation at local and global scales. In this work, we take a data-driven, machine learning approach to measuring innovation. Analyzing a large number of country-level metrics, we aim to discover actionable "levers" of innovation automatically. Using unsupervised learning methods we determine groups of related world development indicators among a collection compiled by the World Bank. We then train a Group Lasso predictive model using data from the World Economic Forum (WEF) that captures the perceived level of innovation in 150 countries. Aside from providing high predictive accuracy, the Group Lasso also provides a model that is easily interpretable. The result is the Open Innovation Index (OII), an automatic global model for measuring innovation using machine learning algorithms and open data. We predict the OII scores for countries that only have WDI data and no existing WEF innovation scores. Furthermore, we also present case studies where the innovation levers of a few representative countries are uncovered automatically by the proposed model.

Introduction

In this work we address the problem of using data to evaluate and study an important, yet ill-defined driver of economic and social growth – innovation. Through job creation and the development of new products and services, innovation increases competitiveness in local and global markets, and has the potential to advance economies in countries around the world. The United Nations recognized innovation as key to economic development when they presented their Sustainable Development Goals in 2015 [1], including a goal to "Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation". Over 190 world leaders committed to these goals to help end extreme poverty and fight inequality and injustice. Innovation is also recognized by the World Economic Forum (WEF) as one of the twelve pillars of economic competitiveness evaluated in their annual Global Competitiveness Report [2]. The 2016 report singles out innovation as an important element for competitiveness as well as openness and economic integration. The report posits that in 2016 the world was recovering from the Great Recession (the general economic decline observed in world markets during the late 2000s) and entering the fourth industrial revolution. In this context, creating an environment in which innovation flourishes is identified as critical for "economic diversification to reignite growth" [2].

Finding ways to foster innovation in countries worldwide is an important step toward a more prosperous, equitable, and sustainable world. But innovation is hard to define, and therefore even harder to measure. The World Economic Forum's Evaluation of Leading Indicators of Innovation study [3] shows that while many previous reports have attempted to quantify innovation, there has been little consensus on the appropriate measures. For instance, indices may focus on only particular outputs of a country, such as the number of patents produced [3], or use curated collections of metrics that have been determined by domain experts to be measures of innovation [4][5][6]. While manually constructed indices contain valuable expert domain knowledge, they may introduce selection bias, since the importance of different datasets is determined by expert opinion. This approach may exclude patterns in the data which automatic machine learning methods could identify. To help policy makers and business leaders better understand how to promote and foster innovation, we need more robust and comprehensive ways to quantify innovation at local and global scales, and evaluate progress made over time.

The need for better measures of innovation has been recognized by economic experts [7], who cite access to open data as a driver for robust innovation research and the development of national and international innovation indices. The goal of this work is to better understand innovation using a data-driven approach. Working with the World Economic Forum, we aim to discover actionable "levers" of innovation automatically from among a large number of country-level metrics, as opposed to manually constructing a measure from a predefined set of hand-selected indicators. Our model is trained using only open data - a vast collection of publicly available socioeconomic datasets. We design a clearly defined data analysis pipeline for evaluating innovation in a highly interpretable and reproducible manner. The result is an automatic Open Innovation Index (OII).

Challenges

Quantifying a high level notion such as innovation is a difficult task, with statistical as well as conceptual challenges. Innovation may have many different meanings in different contexts. There is no absolute “ground truth”, or directly measurable examples, with which to evaluate a model for predicting innovation since it is an intangible concept. Therefore, a proxy for innovation must be used, which introduces some inevitable error and subjectivity in the results. In addition, innovation may be perceived differently in different places due to variation in economies, cultures, and societies worldwide. A simple global model may fail to account for differences between countries, and a very complicated model may overfit to the training data and be a poor predictor for evaluating new countries.

In addition, although there is an enormous amount of public global development data available, it suffers from many data quality issues such as missing values and collection errors. It is important to ensure that high-fidelity data sets are used in the analysis, and predictions are based on actual data values, not on the availability of data. However, the reasons behind missing and noisy data, while difficult to determine, may be relevant to the question at hand. Therefore, data quality should be taken into account as a possible input to analysis.

Along with missing values, the data poses statistical challenges not only due to the data quality, but also because of the interrelated nature of the metrics contained within. A broad number of topics are covered by international development data, as shown in **Figure 1**. Information about latent variables not directly measured is captured by the data as well. Statistical models may not perform well when presented with highly correlated and interdependent data. In addition, there are relatively few example instances of countries (less than 150) compared to the number of potential explanatory metrics (thousands). A major challenge in this analysis is to find a predictive model which can handle the correlations among the explanatory variables, while retaining all the metrics relevant to the outcome.

The potentially large impact of using data science to shape policy is also a concern. The goal of the OII is not just to rank countries and measure progress over time, but also to provide actionable insights and offer evidence-based guidance on improving innovation outcomes at global and country-level scales. The negative impacts of an inaccurate model could substantially affect the lives of many people, and therefore a robust solution is required which can accurately pinpoint the significant factors that impact innovation in each country. There is a high penalty for errors, and careful consideration is required to ensure fair and equitable allocation of resources. A key strategy to address the high-stakes nature of this task is to use a highly interpretable model that decision makers can understand and trust. A clearly defined model using publicly available data to measure innovation can be easily inspected and reproduced. This is in contrast to other proprietary measures for innovation, and in general applications of machine learning to social issues which are not explicit in their formulation. So called "black box" models are not interpretable in this sense and may not be trusted by decision makers. Worse yet, they may be biased in ways that are not discoverable. This can unfairly impact the outcomes of applying these measures to real-world scenarios.

Methodology

Keeping interpretability and robustness in mind, the development of the OII model focuses on discovering relationships between country-level metrics and a measurement of innovation. As a means of understanding the factors which impact the level of innovation in each individual country, a comprehensive collection of World Development Indicators (WDI) compiled by the World Bank [8] is considered. A model is trained on this data to predict innovation scores for each country. The target innovation scores are provided by the World Economic Forum's Global Competitiveness Report (GCR). This report contains survey data which captures the perceived level of innovation in 150 countries over a 10-year timespan.

To develop our model we first employ unsupervised learning to transform the input feature space into a more compact and interpretable representation. Natural correlations between data sets are used to perform an automatic clustering [9]. This organizes the input feature space into a coarser granularity, allowing for more data to be incorporated into the model without overfitting. We then use Group Lasso regression [10] to construct a model which predicts innovation levels with high accuracy, while further reducing the feature space dimensionality. The resulting OII data analysis pipeline is described in detail. Results are compared to an alternative strategy of stability selection [11], and the consistency of the grouping strategy is evaluated for the data as it changes over time.

The rest of this article discusses the data itself, exploratory analysis and preprocessing, the development of the OII predictive model, and resulting innovation scores and individual country profiles. The predictions of the OII are presented, along with a discussion of the metrics chosen by the model. A few case studies of individual countries show that the OII provides customized innovation profiles at the country level.

Data

World Development Indicators

Our input data is the World Development Indicators (WDI) [8] data set published by the World Bank. The WDI data includes national, regional and global estimates measuring development.

Compiled from officially recognized international sources, this open data set represents the most current and accurate global development data available. This statistical reference includes over 1500 metrics covering more than 200 economies of countries and regions, spanning 56 years. The annual publication is released in April of each year, and the online database is updated three times a year. The World Bank's Open Data site provides access to the WDI database free of charge to all users. A selection of the WDI data is featured at data.worldbank.org. The WDI indicators measuring different aspects of an economy are grouped under the high-level topics listed in Figure 1. The statistics provided by these indicators are used as inputs to our analyses.

Global Competitiveness Report

The Global Competitiveness Report (GCR) [2], published annually by the World Economic Forum, includes data which captures the perceived level of innovation in 150 countries. This provides a target for the predictive model in this work. Since 2004 the report has ranked countries based on the Global Competitiveness Index, (GCI), which "assesses the ability of countries to provide high levels of prosperity to their citizens. This in turn depends on how productively a country uses available resources. Therefore, the Global Competitiveness Index measures the set of institutions, policies, and factors that set the sustainable current and medium-term levels of economic prosperity" [12]. Over 110 variables contribute to the index; two thirds of which come from the Executive Opinion Survey, and one third which comes from publicly available sources, such as the United Nations. The survey contains the responses of roughly 14,000 business leaders from 150 economies.

The GCI variables are organized into twelve pillars (see Figure 1), with each pillar representing an area considered as an important determinant of competitiveness. An overall score for each pillar is computed from several sub-components, which help measure that pillar. Of particular interest to this analysis is the 12th pillar: Innovation. The innovation score is the arithmetic mean of 6 metrics (listed in Figure 1 in the rightmost column). The first 5 metrics are responses to survey questions in the Executive Opinion Survey. In addition, one empirical measure, the number of patent applications per million people in the population, also contributes to this competitiveness pillar. We take the overall 12th pillar innovation scores from the GCR as our ground truth data upon which we train our index.

The innovation pillar provides a measure of the perceived level of innovation in each country. Although innovation is a complex concept which is inherently subjective and hence hard to measure, the survey data on which the score is based can be seen as uniquely providing a characterization of the innovation level of each country. The measures included in the WDI dataset, such as GDP for example, are simpler and easier to quantify compared to innovation, although some of them can still be based on survey data. We would like to create a replacement for perceived innovation levels using a model whose inputs are easier to quantify. By training a model to predict these scores, the OII evaluates which empirical measures best align with the perception of innovation, and thereby determines how well these metrics reflect and describe the level of innovation. We place a reasonable assumption here that *bonafide* surveys on metrics that can be easily quantifiable can be trusted.

Data Pre-processing

Although the WDI data represents the most comprehensive global development data available, the data set contains many missing values and requires preprocessing for meaningful analysis.

Due to the noise in the data we first take a number of carefully considered cleaning steps to ensure our analysis is statistically meaningful. These preprocessing steps are standardly adopted in statistical modeling while working with imperfect real-world data sets and they have been shown to be practically useful. The final cleaned data set used in our work consists of a 7-year time series from 2009 - 2015. 142 countries for which there is substantial data in the WDI data set as well as GCI innovation scores are considered for training the model.

Smoothing

For predictive modeling we use WDI data for each country averaged over 3 years. This smoothing step serves to fill some missing values. For instance, there will be a value for a metric even if even if there is only data available for one out of three years. In addition, it may account for noise in the data and better reflect the general state of a country over a short time. We expect that developments such as investments in infrastructure, health, and education do not have an instantaneous effect. We apply the same smoothing step to the GCI data by averaging the scores for every 3 years, resulting in a 7-year time series for 2009 through 2015.

Feature Selection

Even after averaging is performed, there are a significant number of missing values in the smoothed WDI data. We only include metrics which are consistently available for all countries. For each year, data are dropped if there are missing values for more than 50% of countries. This results in an input data set of about 700 metrics per year, reducing our original metric feature space by almost half.

Standardization

After the feature selection step the WDI data is standardized by setting each metric to have a mean of 0 and standard deviation of 1. This allows for a fair comparison of metrics with different scales.

Impute Missing Values

Both WDI metrics and GCI innovation scores may have some missing values after the above pre-processing steps. The remaining WDI missing values are replaced with the mean of the data**.** The original GCI Innovation score dataset contains scores for 150 countries over a 10-year period from 2007 through 2015. For each year, some countries may be omitted from the GCI report if the survey was not successfully conducted [12]. After the smoothing step, if there are years for which a country has missing values, we replace them using the mean of the scores for other years that are available for the country. The intuition for this strategy comes from the observation that the GCI innovation scores do not vary dramatically over time for each country. The scores for each year are then standardized across countries as required by our model.

Predictive Model for Innovation

We seek to develop an accurate model of innovation which aligns with the perceived levels of innovation described by the GCI reports, but one which is also highly interpretable. The metrics included in the model should help to illuminate the measurable factors which reflect and impact innovation. We hope that insights gained through this analysis will be accessible to decision makers and help them identify the most important factors impacting the growth of innovation in their own country. Toward this end, we choose linear regression analysis as our predictive model. This technique is familiar to analysts, economists and government, and has been shown to be effective in econometric analysis. Interpretation is facilitated through coefficient ranking. That is, the coefficients of each term indicate the impact of the corresponding feature on prediction.

Using the publicly available WDI socioeconomic country level metrics as explanatory variables, we built a series of predictive models for each year. The methodology for each model is consistent, while the underlying input data used may vary, depending on the quality of available data for that year. In fact, we want to be able to accommodate as many input data sets as possible in a general way, and have a flexible algorithm that can incorporate new data that may become available.

In order to understand which factors impact this measured level of innovation, the model should capture all relevant metrics from the WDI data set that contain information about innovation and have predictive power, despite possible correlation with other metrics. This will allow decision makers not only to use the index to chart their progress over time, but also to interpret the model, and use their own expertise to gain insight from the factors automatically chosen by the model. We want to provide all information that can shed light on the forces driving innovation or preventing it, without introducing our own bias and preconceived ideas into the process.

Therefore, the first step in our analysis is to transform this feature space into a representation which captures the relationships among the input data. We approach this using a clustering method based on correlation. Data points assigned to a cluster may be related due to some underlying latent factor, because they describe related topics, or occasionally due to some spurious correlation. In this analysis we do not attempt to explain these relations, but rather exploit them in order to reduce our feature space while retaining all relevant information.

Clustering

To group the metrics, we adapt an agglomerative hierarchical clustering method from a previous study on grouping features by correlation for predictive models [9]. Hierarchical clustering produces a hierarchy of groupings among features. The fact that finer clusters are nested within coarser ones is more natural and easier to interpret than clusterings obtained with other methods (e.g. k-means). Moreover, this allows for the exploration of clustering at various granularities and does not require a predetermined number of clusters. The Pearson correlation coefficient of each pair of metrics, $\rho_{XY} = \mathrm{cov}(X,Y)/\sigma_X\sigma_Y$, is adopted as the similarity measure for our clustering. This is a value between positive 1 and negative 1 inclusive, where a high positive value indicates strong positive correlation between the metrics, 0 no correlation, and a low value a negative correlation. A dissimilarity matrix is built using distance values $d(X,Y) = 1 - |\rho_{XY}|$ for each pair of metrics. We further refine this distance measure to only consider features which are strongly and significantly correlated with high confidence. Only distances for very highly correlated metrics ($\rho_{XY} > 0.75$) and with high statistical significance (p value < 0.05) are used; all other pairs of metrics are assigned a distance of 1.

Clustering is performed via an iterative process using this dissimilarity matrix. At each step, the two closest clusters are combined. Clustering proceeds in an agglomerative fashion until there is only one cluster. Average-linkage is used to determine the distance between clusters. That is, given clusters A and B, the distance between them is the mean distance between all pairs of metrics within $\text{avglink}(A, B) = \frac{1}{|A| \cdot |B|} \sum_{X \in A} \sum_{Y \in B} d(X, Y)$. Once all metrics have been combined, a cut point is found at which to halt the clustering process. The step with the greatest difference in avglink between the joined clusters in consecutive iterations is chosen. In this way the number of clusters is determined automatically.

To evaluate the goodness of our clustering method, we examined whether it produced similar results over time. After pre-processing, the WDI data is clustered for each year, and the clusterings are compared in a pairwise fashion. On average 266 groups were found for each year. The Adjusted Mutual Information (AMI) measure [13] is used to compare two clusterings while adjusting for possible correlations due to chance. An AMI score of 0 indicates that the clusterings could have been chosen by chance, and a 1 indicates identical clusterings. For the WDI input data, the average pairwise AMI score for the 7-year timespan ranged from 0.81 - 0.89, indicating a consistency among the groups across years.

Including this clustering step in our data analysis pipeline allows the naturally occurring patterns in the data to drive the analysis. Indeed, other possible groupings such as those defined by the World Bank or WEF could alternatively be used in the model to incorporate domain knowledge instead. However, by using correlation in this manner, we allow for the relationships between metrics to be discovered automatically, and to reflect the actual outcomes of each metric, country and year. After this grouping step, further analysis can be completed at the cluster level.

Group Lasso

Though our feature space has now become compressed, we still want to remove groups from the space which do not have predictive input for our task. Lasso [14] is a regression technique which performs both feature selection to reduce the feature space, and regularization to avoid an overly complicated model. Lasso has been shown to improve the predictive ability and interpretability of regression models [14]. By adding a constraint that the coefficients in the model sum to be less than a certain amount, this technique forces some of the coefficients to zero, thereby effectively performing feature selection. This regularization step also simplifies the model to avoid overfitting.

However, the Lasso model does not determine a unique solution to the prediction problem. When there are highly correlated metrics in the data set, Lasso will simply choose one of them as a predictor and set the coefficients for the other variables to zero [9]. This is problematic for the purposes of our analysis. If a feature is very closely correlated with another that has high predictive power, then both features are relevant to our question and should be included in the model. The question of why one is a better predictor than the other, or why they are so closely related are not for our model to answer. We simply want to choose both as interesting and present them to the analyst for further investigation.

Therefore, our solution is found using the correlation-based clustering described above in conjunction with the Group Lasso [10]. With G, a set of distinct groups of metrics, the Group Lasso will select either all or no features from each group to be included in the model. The Group Lasso is described by Equation 1.

$$(1) \quad \underset{\beta \in \mathbb{R}^p}{\arg\min} \left\{ \|\mathbf{Y} - \mathbf{X}\beta\|_2^2 + \lambda \sum_{g=1}^{G} \left\| \beta_{I_g} \right\|_2 \right\}$$

Given dependent variable vector $\mathbf{Y} \in \mathbb{R}^n$, an $n \times p$ design matrix $\mathbf{X}$, and vector of independent variables $\beta \in \mathbb{R}^p$, the estimator learns the model subject to the regularization term $\lambda \sum_{g=1}^{G} \left\| \beta_{I_g} \right\|_2$ which is applied for each group of metrics $g \in G$. The contributions of the groups to the model is spread across the features, with different coefficients assigned to each. In this way, the selection strategy avoids the case in which many highly correlated features are given and Lasso chooses only one almost arbitrarily. Group Lasso has been shown to be effective in other scenarios where the data are highly interdependent, and the number of features is much larger than the number of training examples [9].

Model Evaluation

Performance

To evaluate our model we use the coefficient of determination, or R^2 score. This measure is commonly used to evaluate statistical models. It measures the amount of variance explained by the model, and not due to noise. This is an appropriate test for the goodness of fit of the model. An R^2 score of 1 indicates a perfect fit, and a score of 0 indicates the model only predicts as well as using the mean of the data. Scores can also be negative, since the model may be an arbitrarily bad predictor. 5-fold cross validation is employed to evaluate the model. The average of 10 cross validation trials is taken, where each time the entire dataset is randomly shuffled to ensure that different subsets of the data are selected for the training and test sets. This approach was chosen due to the small number of training examples which makes the use of a holdout validation set impractical.

Our evaluation of the Group Lasso model results in an average R^2 score of 0.75. Any score above zero means that the model has some predictive power, however a low score indicates that the outcome may be due to features missing from the input data set, or that a linear approximation is not best for the problem. In our scenario, it is safe to assume that we do not have a complete picture of the factors which impact innovation in our model. However, we do have an accurate picture of what factors are measurable and actionable through policy and funding intervention. We consider this R^2 score to be high enough for our intended application.

The innovation score is not only valuable as a numeric value but also as a global ranking, and a basis for comparing different countries. The Spearman rank correlation coefficient measures the monotonic relationship of two rankings. A score of 1 indicates identical orderings, while a -1 means one rank is the inverse of the other. Comparing the resulting rankings of countries using the OII scores and the original GCI innovation scores yields an average correlation of 0.85, meaning the relative orderings of the two evaluations are quite similar.

Metric Evaluation

The Group Lasso model selects on average 26 groups containing 257 metrics. These metrics fall into the following categories, with many subcategories:

- *Economic Policy & Debt*
- *Education*

- *Environment*
- *Financial Sector*
- *Health*
- *Infrastructure*
- *Private Sector & Trade*
- *Public Sector*
- *Social Protection & Labor*

Model Comparison

To evaluate the predictive power and interpretability of our grouping strategy, we compare it to another data-driven feature selection methodology using a technique known as Stability Selection [11]. This method also addresses the problem of correlated input features to a Lasso regression model. If two features are highly correlated, then the Lasso model will choose one at random. For our purposes this is not desirable, since we do not know in advance which of the correlated features may be truly aligned with innovation levels, and which could be the result of spurious correlations, or correlated with innovation due to some other underlying factor.

To resolve this, stability selection employs an approach similar to bootstrapping. The model is trained many times, and each time a randomly selected subset of the input features is used. The data is also slightly perturbed, giving a different weighting to the individual metrics. At each iteration, the metrics that are assigned non-zero coefficients are recorded. After many runs, the metrics that have been chosen by the model the highest number of times are determined to have the most predictive power for the model. In this way the most important metrics to sift to the top naturally after many perturbations of the data

For our analysis we used the ElasticNet algorithm [15], a regression method which uses a combination of the L_1 Lasso and L_2 ridge regularizations and has been shown to improve performance in cases where the input feature space is larger than the number of examples available for training [15]. This model is able to achieve an average R^2 score of 0.72. Our empirical evaluation found that the technique required a very high number of iterations to provide consistent results after which the model chose 26 features as stable across at least 5 out of the 7 years. These features fall into only a few subcategories of the WDI designations:

- *Economic Policy & Debt*— GNI and GDP per capita, national accounts, purchasing power parity, adjusted net national income per capita.
- *Financial Sector*— Assets.
- *Health* — Health expenditure per capita.
- *Infrastructure for technology* — Communications, technology, transportation.
- *Private Sector & Trade*—Logistics performance indices, business environment.

Interpretability

Stability selection yields a simple model with an R^2 score similar to, but slightly lower than the Group Lasso, and it captures many fewer features. We observe that the majority of metrics picked in the stability selection experiments belong to a single cluster in the Group Lasso experiments. These metrics are shown by both models to be most strongly correlated with the innovation scores. It seems apparent that while the association of strong economies, technological infrastructure and private industry to innovation makes intuitive sense, it is not very informative for decision makers.

More interesting perhaps are the smaller groups and single metrics which are not correlated with the dominant group, but also present as indicators when enforcing the group structure. However, these deeper insights into our explanatory data come at the price of a more complicated model. When we consider which model is more interpretable, it can be difficult to evaluate, especially since interpretability may have different meanings [16].

For the purposes of better measuring innovation, the goal is to provide decision support. In the process of stability selection, given two very highly correlated metrics one may be a slightly better predictor of innovation than the other, and therefore it will be picked more often. In reality both features have important information to contribute to our understanding of how to measure innovation. An analyst may intuitively understand the importance of one metric over the other, or be interested to discover the underlying relationship between them. The benefit of the grouping strategy is that since these variables are highly correlated, they will both be included in a single group, and since at least one of them is an effective predictor for the model, the entire group will be chosen. The coefficients of the variables within the group will indicate that one is a slightly better predictor than the other. Therefore, using the Group Lasso model allows for a more meaningful result.

Open Innovation Index Results

The OII model provides an automatic way to evaluate innovation levels in a country using open data, and provides insights into the contributing factors for each score. **Figure 2** shows a choropleth map of predicted global OII scores using the most recent WEF data. The possible score ranges from 1 to 7, with actual predicted minimum score 2.50, maximum score 5.66, and standard deviation 0.63. Using our automatic, data-driven approach we are able to evaluate 219 countries and regions worldwide. This includes 77 countries that have WDI socioeconomic metrics but no GCI survey data, and were previously not considered for evaluation. The histograms below the map show the distributions of scores binned either by equal intervals of scores, as is given on the map, or by equal numbers of countries, which shows that 2/3 of country scores fall below the mean score of 3.41.

Use Cases for Individual Country Profiles

To evaluate the impact of different factors on the innovation scores for individual countries we can look at the coefficients used by our model to determine the innovation score. **Figure 3** shows 3 charts which visualize the formulas for Chile, Argentina and Norway. The visualizations show the most impactful metrics on the final score for each country. Metrics with a high value to the right are contributing to the score, while negative values to the left are detracting. To improve readability and the high-level descriptive quality of these visual profiles, we only include metrics which contribute to or detract from the overall score by at least 0.005. Metrics are colored according to the WEF categories given in Figure 1. At a glance we can see which sectors dominate the model for each country, and observe similarities and differences between countries.

Norway is ranked number 1 by the OII model, and this snapshot view shows strengths in almost all categories. Especially dominant are the extremely strong economy indicated by the block of red bars, and many significant contributing factors in the private sector and trade, infrastructure, and health categories. By contrast, Chile and Argentina are ranked 44th and 170th respectively. We can observe weaker scores across most categories than for the higher ranked Norway. Also of note is that although neighboring geographically, these two countries are assigned very different OII scores and have varying contributions to OII from different categories. To understand why this is, a detail view of each country can be inspected.

Figure 4 shows the detail view for the first two categories in Chile's country profile. A number of metrics from the private sector and trade and infrastructure categories are contributing positive values to the innovation score. In particular, the logistics index, efficient customs procedures and quality of port infrastructure metrics dominate the model. This indicates that Chile's strong trade relationships seem to facilitate a favorable environment for innovation. Argentina on the other hand is lacking infrastructure contributions, the private sector and trade category is not as strong, and the financial sector metrics are detracting from the overall OII score. To see where possible improvements could be made, a decision maker could dig deeper into the profiles of these countries to see the impacts of individual metrics. Such actionable insights provided in an automated way can help policy makers identify the correct areas for further focus and investigation. The Group Lasso model retains many useful and interesting metrics as shown in the detail view in Figure 4.

Discussions and Future Work

Traditional economic analysis begins with expert knowledge. Starting with a hypothesis, statistical methods are applied to confirm or rule out this assumption. Data science and machine learning can offer a different perspective on this type of analysis by flipping the role of expert knowledge. The data-driven approach starts with the data itself, ignoring or down-weighting prior assumptions and conventional ideas about the problem. A model that best describes the data is inferred which exposes naturally occurring patterns in the data that may have been previously overlooked. These results are then presented to experts for interpretation, and to lead their further investigation into the topic.

In this work, we do not claim to have designed a definitive measure of innovation – rather the OII provides a baseline global model for measuring innovation using empirical methods. The OII model unveils features which are predictors of innovation, and are clearly related to the question of how to grow and foster innovation. However, we know that correlation does not necessarily imply causation. Further study is needed to evaluate and then act on these observations.

In addition, taking a closer look at the errors for each country, we can gain insight into the limitations of a single global model for innovation. Surprisingly, this analysis shows that the countries for which innovation is hardest to measure are exactly those that are judged by the WEF survey data to be the most innovative. The OII model consistently underestimates this highly innovative group. Perhaps the characteristics of highly innovative countries are not captured by our data, or we simply do not have enough training examples because they are particularly unique.

Other natural groupings of countries share similar discriminative characteristics, beyond the group of very highly innovative countries. Countries vary considerable in terms of their size, development level, region, culture and many other factors. Surely, the manner in which innovation is manifested within countries also varies. The factors which are important for one group of countries may not be the same everywhere. Identifying similar countries in our feature space could improve the model by allowing for customization based on inherent characteristics of countries. In this way the global innovation model could perhaps be improved. Future work could investigate the use of more sophisticated modeling techniques such as other variations of group sparse models or tree-based models. Boosting or ensembling could also prove beneficial, using country or group adaptive weightings based on various characteristics such as region, income level or data quality, or neighborhood-based regularization to improve the model.

Finally, the input data used in the predictive modeling may not necessarily be accurate and can have uncertainty associated with it. This, along with a lack of sufficient data samples contributes to uncertainty in the predictions as well. Both of these uncertainties can be quantified using principled statistical methods, and this is a rich area for future research.

Conclusion

This work has presented the Open Innovation Index, a data-driven measure of the level of innovation in countries worldwide. Using high-fidelity, publically available data, the OII automatically reveals the most important global factors which impact innovation. Individual profiles reveal the role each factor plays in determining innovation in each individual country. We have shown that the OII model correlates with historical measures of the perception of innovation in each country, and provides a comprehensive, easily understandable, and replicable model of innovation. As more work is done to use open data to better facilitate economic growth and worldwide prosperity, the OII can easily be extended to incorporate new sources of information as they become available.

We hope this work can provide policy makers, business leaders, and individuals with deeper insights into the factors impacting innovation in their countries, and help inform efforts to grow and foster innovation worldwide. This predictive model of innovation based on regularly collected open metrics could replace manually conducted innovation surveys with automated “measurements”. In addition, it can provide an analysis of the innovation level in countries for which surveys were not conducted.

Acknowledgments

This work was conducted as part of the inaugural IBM Data Science for Social Good fellowship program Summer 2016, with input and guidance from the World Economic Forum.

References

1. United Nations, Sustainable Development. [Online]. Available: http://www.un.org/sustainabledevelopment/sustainable-development-goals/

2. K. Schwab and X Sala-i-Martin, World Economic Forum’s Global Competitiveness Report, 2016-2017. [Online]. Available: https://www.weforum.org/reports/the-global-chttps://www.weforum.org/reports/the-global-competitiveness-report-2016-2017-1ompetitiveness-report-2016-2017-1. Switzerland: World Economic Forum. 2016.

3. Global Agenda Council on the Economics of Innovation Evaluation of Leading Indicators of Innovation, World Economic Forum Technical Report. [Online]. Available: http://www3.weforum.org/docs/WEF_GAC_on_Economics_Innovation.pdf

4. European Innovation Scoreboard. [Online]. Available: http://ec.europa.eu/growth/industry/innovation/facts-figures/scoreboards/index_en.htm

5. The Global Innovation Index. [Online]. Available: https://www.globalinnovationindex.org/

6. Z. J. Acs, L. Szerb, and E. Autio. "The global entrepreneurship and development index." [Online]. Available: https://thegedi.org/global-entrepreneurship-and-development-index/

7. C. J. Schramm, A. Arora, R. K. Chandy, K. Cooper, D. W. Jorgenson, D. S. Siegel, D. L. Bernd et al. "Innovation measurement: Tracking the state of innovation in the American economy. A report to the Secretary of Commerce by The Advisory Committee on Measuring Innovation in the 21st Century Economy" (2008). [Online] Available: http://users.nber.org/~sewp/SEWPdigestFeb08/InnovationMeasurement2001_08.pdf

8. The World Bank, World Development Indicators (2014). [Online]. Available: http://databank.worldbank.org/data/download/archive/WDI_excel_2014_04.zip

9. P. Bühlmann, P. Rütimann, S. van de Geer, and C. H. Zhang, "Correlated variables in regression: clustering and sparse estimation," in *Journal of Statistical Planning and Inference*, vol. 143, no. 11, pp. 1835-1858, 2013.

10. M. Yuan, and Y. Lin. "Model selection and estimation in regression with grouped variables," in *Journal of the Royal Statistical Society: Series B (Statistical Methodolog*y), vol. 68, no.1, pp. 49-67, 2006.

11. N. Meinshausen and P. Bühlmann,"Stability selection" in *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 72, no. 4, pp. 417-473, 2010.

12. K. Schwab and X Sala-i-Martin. "The Global Competitiveness Report 2014-2015." [Online]. Available: http://reports.weforum.org/global-competitiveness-report-2014-2015/methodology/

13. N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *Journal of Machine Learning Research,* vol. 11, pp. 2837-2854, 2010.

14. R. Tibshirani, "Regression shrinkage and selection via the lasso," in *Journal of the Royal Statistical Society. Series B (Methodological),* pp. 267-288, 1996.

15. H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," in *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301-320, 2005.

16. Z. C. Lipton, "The Mythos of Model Interpretability." in *ICML Workshop on Human Interpretability in Machine Learning,* New York, 2016.

Bios

Caitlin Kuhlman *Computer Science Department, Worcester Polytechnic Institute, Worcester, MA 01609 USA (cakuhlman@wpi.edu).* Ms. Kuhlman is a Ph.D. student in the Department of Computer Science at Worcester Polytechnic Institute. She is a member of the Database Systems Research Group. Her research interests include machine learning and data mining, with a particular interest in using public data to solve social problems. In the summer of 2016 she was a member of the inaugural class of Data Science for Social Good Fellows at IBM Research.

Karthikeyan Natesan Ramamurthy *IBM Research Division, IBM Thomas J Watson Research Center, Yorktown Heights NY 10598 USA (knatesa@us.ibm.com).* Dr. Ramamurthy is a Research Staff Member in the Data Science Group at the IBM Thomas J Watson Research Center. He received his Ph.D. in Electrical Engineering from Arizona State University in 2013. His broad research interests are in understanding the geometry of high-dimensional data, developing theory and methods for efficiently modeling the data, and the interplay between humans, machines, and data. His papers have won best paper awards at the 2015 IEEE International Conference on Data Science and Advanced Analytics, and the 2015 SIAM (Society for Industrial and Applied Mathematics) International Conference on Data Mining. He is a member of IEEE.

Prasanna Sattigeri *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (psattig@us.ibm.com).* Dr. Sattigeri is a research staff member in the Data Science Group at the IBM Thomas J Watson Research Center in Yorktown Heights, NY. He received his Ph.D. in Electrical Engineering from Arizona State University. His broad research interests include developing theory and algorithms for representation learning of high dimensional data and building systems for semantic inferences. He is also interested in developing scalable solutions and has experience shipping machine learning products to consumers.

Aurélie C. Lozano *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (aclozano@us.ibm.com).* Dr. Lozano is a Research Staff Member at the IBM T.J. Watson Research Center. She received the M.S./Dipl.Ing. degree in Communication Systems from the Swiss Federal Institute of Technology Lausanne (EPFL) in 2001, and the M.A. and Ph.D. degrees in Electrical Engineering from Princeton University respectively in 2004 and 2007. She was an Adjunct Associate Professor in the Computer Science Department and the Industrial Engineering and Operations Research Department at Columbia University from 2014 to 2016. Her research interests include machine learning, statistics and optimization. Her current focus is on high dimensional data analysis and predictive modeling, with applications including biology, environmental sciences, business and infrastructure analytics, and social media analytics. She was a recipient of the best paper award at the conference on Uncertainty in Artificial Intelligence (UAI) 2013 and of the IBM Research Pat Goldberg Memorial best paper award, 2013.

Lei Cao *CSAIL, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (lcao@csail.mit.edu).* Dr. Cao is a postdoctoral researcher in the CSAIL (Computer Science and Artificial Intelligence Laboratory) at the Massachusetts Institute of Technology. He was previously a Research Staff Member in the Commerce Research department at the IBM T.J. Watson Research Center, where he worked on big data analytics for an e-commerce product. He received a Ph.D. degree in Computer Science from Worcester Polytechnic Institute in 2016. He is author or coauthor of a dozen technical papers at top data management and data mining venues including KDD (Knowledge Discovery and Data Mining), SIGMOD (Association for Computing Machinery's Special Interest Group on Management of Data), VLDB (Very Large Data Bases), ICDE (International Conference on Data Engineering), etc.

Chandra Reddy *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, New York 10598 (creddy@us.ibm.com).* Dr. Reddy is a researcher in the Mathematical Sciences Department at IBM Research. Dr. Reddy joined IBM in 1998 after receiving a Ph.D. degree in computer science from Oregon State University. Since then, he has been developing solutions to various analytics and optimization problems arising in industries such as steel, semiconductors, energy and construction. Several of these solutions have been successfully deployed and are in production use at various industrial facilities worldwide. He is a "Best of IBM 2015" honoree and has received an IBM Corporate Award and an IBM Outstanding Technical Accomplishment Award. His research, prior to joining IBM, focused on both theoretical and practical aspects of machine learning of compact hierarchical control knowledge for planning and problem solving. Dr. Reddy's research interests include machine learning, optimization, and artificial intelligence.

Aleksandra Mojsilović *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (aleksand@us.ibm.com).* Dr. Mojsilović is an IBM Fellow and scientist at the IBM T. J. Watson Research Center. She received her Ph.D. degree in electrical engineering from the University of Belgrade, Belgrade, Serbia in 1997. She was a Member of Technical Staff at the Bell Laboratories, Murray Hill, New Jersey (1998-2000), and then joined IBM Research, where she currently leads the Data Science group. Dr. Mojsilović is a founder and co-director of the IBM Social Good Fellowship program. Her research interests include multidimensional signal processing, predictive modeling and pattern recognition. She has applied her skills to problems in computer vision, healthcare, multimedia, finance, human resources, public affairs and economics. She is one of the pioneers of business analytics at IBM and in the industry; throughout her career she championed innovative uses of analytics for business decision support. For her technical contributions and the business impact of her work Dr. Mojsilović was appointed an IBM Fellow, the company's highest technical honor. She is the author of over 100 publications and holds 16 patents. Her work has been recognized with several awards including IEEE Signal Processing Society Young Author Best Paper Award, Institute for Operations Research and the Management Sciences (INFORMS) Wagner Prize, IBM Extraordinary Accomplishment Award, IBM Gerstner Prize, and Best Paper awards at the European Conference on Computer Vision (ECCV) and the Conference on Service Operations and Logistics, and Informatics (SOLI). She is an IEEE Fellow and a member of INFORMS and Society of Women Engineers (SWE).

Kush R. Varshney *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (krvarshn@us.ibm.com).* Dr. Varshney is a research staff member and manager in the Data Science Group at the IBM T. J. Watson Research Center in Yorktown Heights, NY. He received the Ph. D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 2010. He applies data science and predictive analytics to human capital management, healthcare, olfaction, public affairs, and international development. He conducts academic research on the theory and methods of statistical signal processing and machine learning. His work has been recognized through best paper awards at the Fusion 2009, SOLI (Service Operations and Logistics, and Informatics) 2013, KDD (Knowledge Discovery and Data Mining) 2014, and SDM (SIAM International Conference on Data Mining) 2015 conferences. Dr. Varshney is codirector of the IBM Social Good Fellowship program.

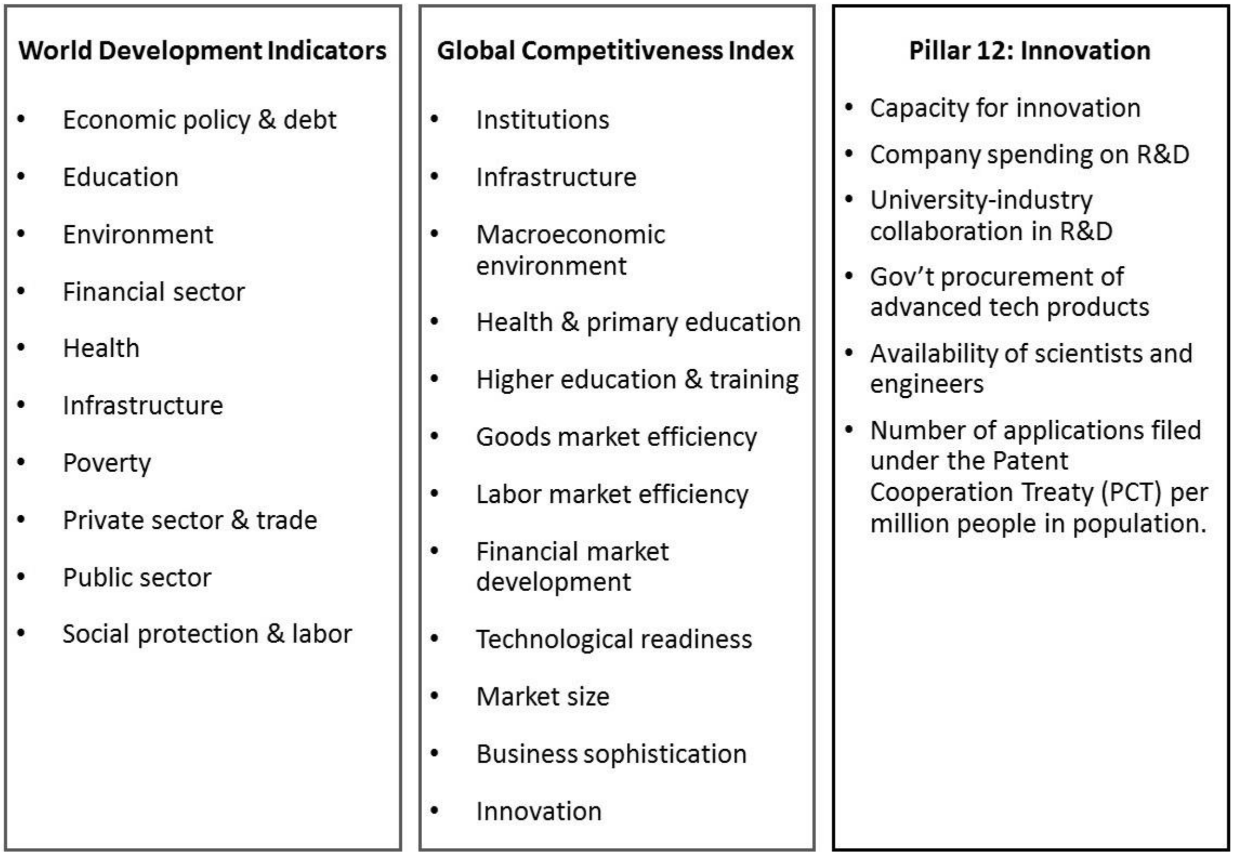

Figure 1 Datasets used in the construction of the Open Innovation Index.

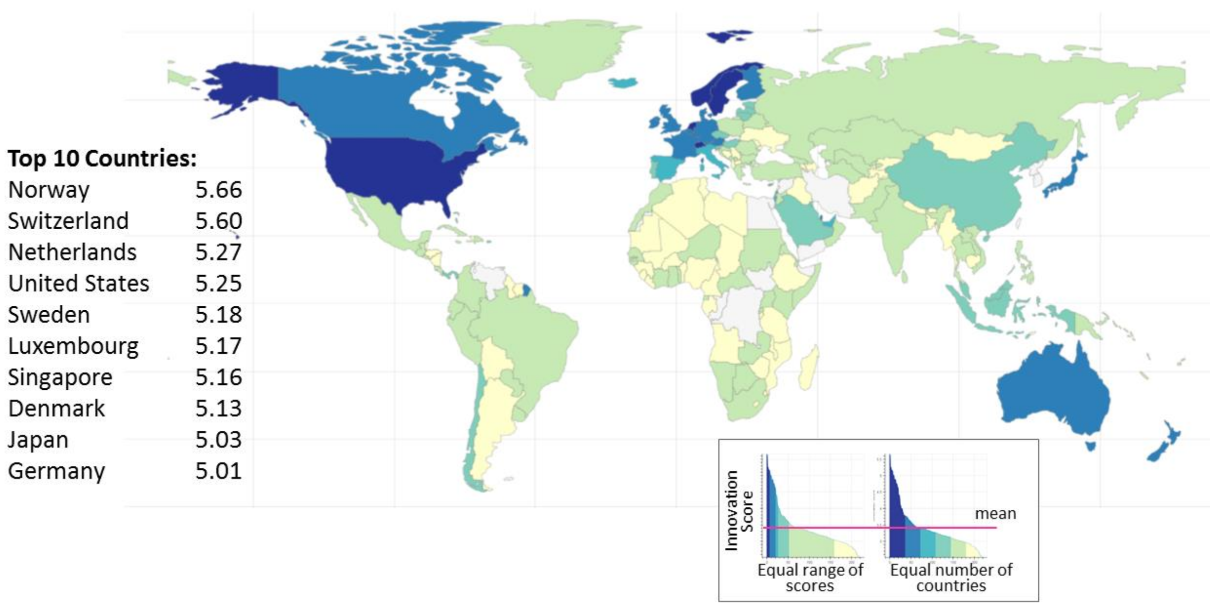

Figure 2 Predicted Open Innovation Index scores based on 2013-2015 WDI data.

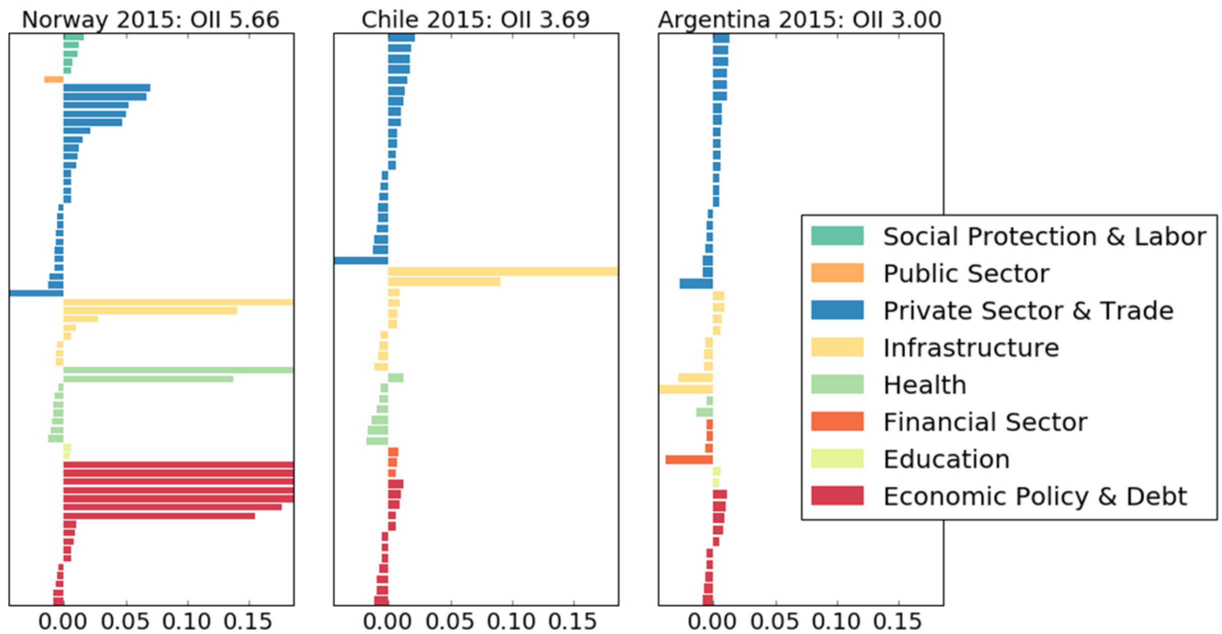

Figure 3 Condensed view of individual country profiles. Each bar along the y-axis represents a single metric. The numbers 0.00 to 0.15 on the x-axis indicate the amount each metric contributes to the overall OII score.

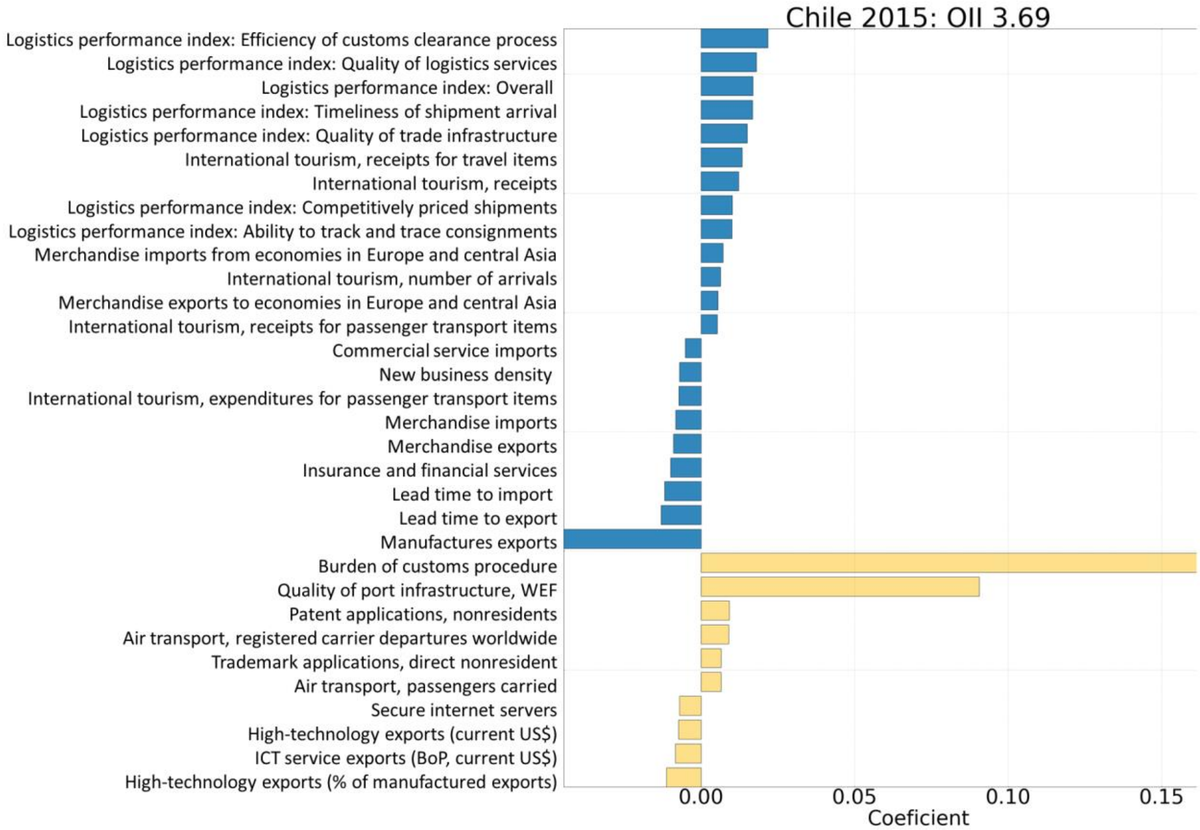

Figure 4 Detail view of the Private Sector & Trade and Infrastructure categories of the country profile for Chile. The numbers 0.00 to 0.15 on the x-axis indicate the amount each metric contributes to the overall OII score. (ICT: information and communications technology; BoP: balance of payments.)