# Decision Support for Policymaking: Causal Inference Algorithm and Case Study

Bhanukiran Vinzamuri
IBM Research
1101 Kitchawan Rd
Yorktown Heights, New York 10598
bhanu.vinzamuri@ibm.com

Aleksandra Mojsilović
IBM Research
1101 Kitchawan Rd
Yorktown Heights, New York 10598
aleksand@us.ibm.com

Kush R. Varshney
IBM Research
1101 Kitchawan Rd
Yorktown Heights, New York 10598
krvarshn@us.ibm.com

## ABSTRACT

Economic and public policy choices by country-level decision makers have direct influence on sustainable development in their countries. Innovativeness is one key component of sustainable development that is difficult to define and measure directly. Nevertheless, in recent years, several innovation indices such as the Global Innovation Index (GII) and Global Competitiveness Index (GCI) have been developed. It is increasingly important to standardize the way of defining and measuring innovation to track progress towards sustainable development, but more importantly, it is important to develop an understanding of the policy choices that lead a country to be more innovative. At its core, this is a problem of causal inference from observational time series data made challenging due to missingness, noise and high correlation patterns. In this paper, we study this inference problem on the World Development Indicators (WDI) dataset acquired from the World Bank using a novel framework that can handle the aforementioned challenges (not only for the innovation domain, but for any policy domain). We address sparsity and correlation using concave regularization, and error control in the presence of noise using stability selection. We conduct a case study on a developing country by comparing inferred causal levers in different domains (e.g., infrastructure, financial sector and public sector) with levers obtained from similar and dissimilar countries. We validate some of our inferences using results published by subject matter experts through independent studies. This analysis provides actionable insights to policymakers of the developing country for improving their innovativeness.

## KEYWORDS

Social good, innovation, policymaking, causal inference, sparse regression

**ACM Reference format:**
Bhanukiran Vinzamuri, Aleksandra Mojsilović, and Kush R. Varshney. 2016. Decision Support for Policymaking: Causal Inference Algorithm and Case Study . In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 6 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

*Conference'17, Washington, DC, USA*
© 2016 ACM. 978-x-xxxx-xxxx-x/YY/MM...$15.00
DOI: 10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

In September 2015, 193 members of the United Nations agreed upon a set of 17 sustainable development goals (SDGs) to transform our world by 2030 [7]. The main goal of SDG 9 (Industry, Innovation and Infrastructure) is *to build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation.*[1] This SDG is one of the six SDGs discussed at the Multi-Stakeholder Forum on Science, Technology and Innovation for SDGs in 2017.

As a result, public organizations and policymakers have worked in cohesion to define economic indices which quantify the extent of innovation, political growth and sustainable development observed for each country. The Global Competitiveness Index (GCI) is one among these indices produced by the World Economic Forum which captures metrics corresponding to twelve diverse pillars [12]. The GCI for *Bangladesh* in the year 2015 is illustrated in Figure 1. The horizontal axis represents the pillar score in the range 1-7. (We choose Bangladesh since our later case study focuses on this country.)

The GCI is constructed in a labor-intensive fashion primarily based on surveys of experts around the world and does not provide any insight into which levers a country should be pulling, i.e. which developmental investments a country should be making, to improve on these pillars. In this work, we propose to address these issues by taking a data-driven approach analyzing a large set of time series of country-level development indicators, specifically the set of World Development Indicators (WDI) produced by the World Bank. Our approach leads to a causal modeling exercise with WDI values as covariates and GCI scores as response variables. In this paper, we primarily focus on the innovation pillar of GCI.

The data-driven inference problem has several challenges:

1. *Temporality*: GCI scores vary with time on a yearly basis depending on various socioeconomic and political factors which affect the country during the year,
2. *Sparsity*: The high-dimensional nature of predictors makes it important to infer the true sparse set of causative levers for GCI scores,
3. *Correlation*: WDI predictors are highly correlated which can mislead inference algorithms by detecting false positives instead of the true ground-truth variables. This affects model stability and interpretability,

[1] http://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-9-industry-innovation-and-infrastructure/targets/

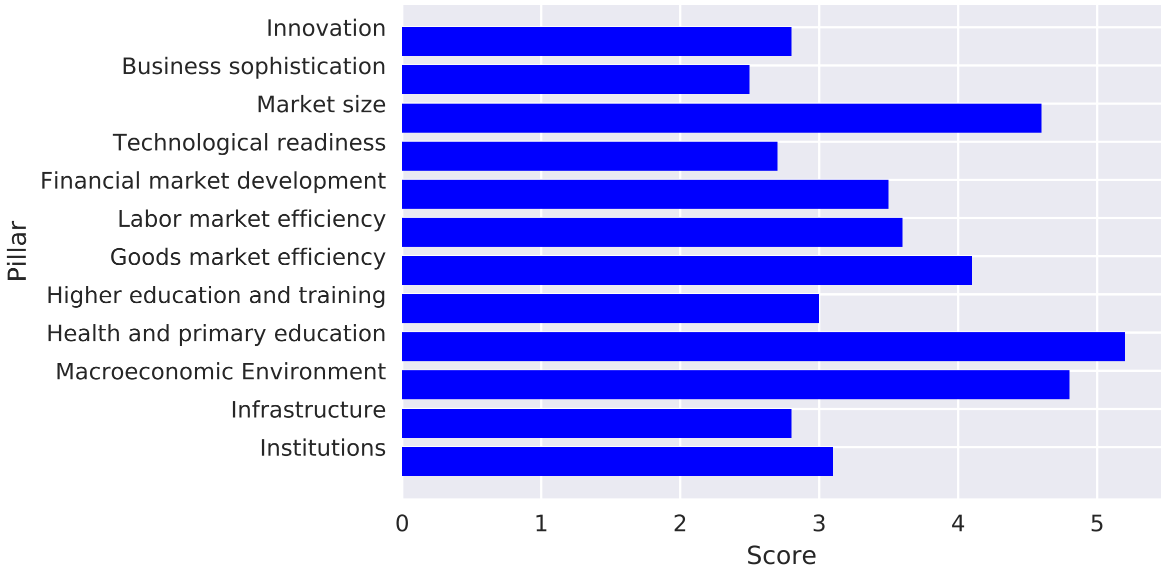


**Figure 1: Global Competitiveness Index for Bangladesh in 2015.**

(4) *Noise*: This stems from the inherent noise in the real-world WDI and GCI datasets which are collected from the World Bank and the World Economic Forum, respectively, and
(5) *Missing Values*: This phenomenon occurs due to the high-dimensional nature of the predictors and the lack of their availability across all countries for different years.

The problem of identifying a sparse set of causal levers from a time series of WDI predictors and GCI scores can be set up as a Granger causal inference problem [3]. Granger causality is an inference method where a time series $A$ is said to be causative to another time series $B$ if the past values of $A$ can help predict the current values of $B$. An important component of this inference method is the lag which determines the time period to go back in the past.

Our previous work has also looked at solving this problem using causal inference methodologies [5, 6, 11]. However, these methods use the knowledge of feature groups (clusters of features) determined by a semi-automated clustering approach which adds an additional parameter to the inference method. These inference methods also do not address all the problem-specific challenges outlined above. In contrast, in this paper, we propose a data-driven inference framework for identifying causal levers which uses concave regularization and randomization-based methods to address all the challenges. Concave regularizers are very effective at inferring the true sparse structure and perform better than convex regularizers under variable levels of noise and correlation [13]. Randomization-based methods such as stability selection reduce the variability of results with unknown noise levels which improves interpretability [9]. One of the key technical contributions of our proposed algorithm is that we can obtain theoretical guarantees on the expected number of false levers selected.[2] This can be very useful for a policymaker who wants to limit the number of false levers selected by the inference algorithm.

## 2 STATISTICAL METHODOLOGY

### 2.1 Background

In this section, we provide preliminary background on the statistical and machine learning techniques we employ for developing an effective causal inference algorithm for policymaking. The first technique described here is the linear regression problem which has the following form:

$$y = X\beta + \varepsilon, \tag{1}$$

where $y \in \mathbb{R}^n$ is a response variable, $X \in \mathbb{R}^{n\times p}$ is a feature matrix, $\beta \in \mathbb{R}^p$ is a coefficient vector, and $\varepsilon \in \mathbb{R}^n$ is a noise vector which has zero mean and sub-Gaussian noise such that $\varepsilon \sim N(0, \sigma^2 I_{n\times n})$.

In Table 1, we review some of the main notations used in this paper. Regularization techniques are used heavily among researchers from both the social science and machine learning communities. These techniques improve the generalizability of the solution obtained by reducing overfitting. The following class of regularized linear regression problems is studied in this paper:

$$\hat{\beta} = \underset{\beta\in\mathbb{R}^p}{\text{argmin}} L(\beta; \lambda; \gamma), \tag{2}$$

where

$$L(\beta; \lambda; \gamma) = \frac{1}{2n}\|y - X\beta\|_2^2 + \sum_{j=1}^{p} h(\beta_j; \lambda; \gamma), \tag{3}$$

$\beta = (\beta_1, \ldots, \beta_p)$, and $h(\beta_j; \lambda; \gamma)$ is a concave penalty function consisting of parameters $\lambda$ and $\gamma$. This formulation will be referred to as sparse regression in the later parts of this paper for simplicity. Our intuition for employing concave penalties here is due to their improved performance under high correlation and noise settings

[2] Technical details are in another paper currently under triple-blind peer review.

**Table 1: Notation**

| Name | Description |
|---|---|
| $n$ | number of instances |
| $p$ | number of features |
| $X$ | $\mathbb{R}^{n \times p}$ feature vector matrix |
| $y$ | $\mathbb{R}^{n}$ response variable |
| $\beta$ | $\mathbb{R}^{p}$ regression coefficient vector |
| $\lambda$ | scalar regularization parameter |
| $\Lambda$ | vector of regularization parameters |
| $\hat{\Pi}^{\Lambda}$ | $[0,1]^{p \times \lvert\Lambda\rvert}$ selection probability matrix |
| $\pi_{thr}$ | cut-off parameter |
| $\gamma$ | scalar concavity parameter |
| $h(\beta_j; \lambda; \gamma)$ | family of concave penalty functions |

which is generally observed in real-world policymaking datasets. They can also be implemented efficiently which makes them a good choice for this problem.

We now explain the other algorithmic paradigm used in this paper which is called stability selection [9]. This method samples several fixed set of instances repeatedly from the training data and fits sparse regression models on them to identify features which are selected consistently. Despite its simple formulation, this procedure has good theoretical guarantees on the number of false positives selected. We use this procedure for false positive control with the base concave regularized linear regression method described earlier. We mention some definitions used here in the context of stability selection.

A regularization path is given by the coefficient value of each variable over all regularization parameters $\{\hat{\beta}^{\lambda}_{k} : \lambda \in \Lambda, k = 1, \ldots, p\}$. For any regularization parameter $\lambda \in \Lambda$, the selected set $\hat{S}^{\lambda}$ is implicitly the function of the samples $I = \{1, \ldots, n\}$. We write $\hat{S}^{\lambda} = \hat{S}^{\lambda}(I)$ where necessary to express this dependence.

Let $I$ be a random subsample of $\{1, 2, \ldots, n\}$ of size $\frac{n}{2}$ drawn without replacement. For every set $K \subseteq \{1, 2, \ldots, p\}$, the probability of being in the selected set $\hat{S}^{\lambda}(I)$ is defined as

$$\hat{\Pi}^{\lambda}_{K} = \mathbb{P}^{*}\{K \subseteq \hat{S}^{\lambda}(I)\} \tag{4}$$

where $\mathbb{P}^{*}$ represents the probability estimate.

For every variable $k = \{1, 2, \ldots, p\}$, the stability path is given by the selection probabilities $\hat{\Pi}^{\lambda}_{K}, \lambda \in \Lambda$. It is a compliment to the usual path plots that show the coefficients of all variables $k = \{1, 2, \ldots, p\}$ as a function of the regularization parameter.

For a cut-off $\pi_{thr}$ with $0 < \pi_{thr} < 1$ and a set of regularization parameters $\Lambda$, the set of stable variables is defined as given below

$$\hat{S}^{\Lambda} = \{k : \max_{\lambda \in \Lambda}(\hat{\Pi}^{\lambda}_{k}) \geq \pi_{thr}\} \tag{5}$$

Variables with high selection probabilities are retained and those with low selection probabilities are disregarded. The exact cut-off $\pi_{thr}$ is a tuning parameter but the results vary surprisingly little for sensible choices in a range of the cut-off. An advantage of the stability selection procedure is that the choice of the initial set of regularization parameters $\Lambda$ typically has not a very strong influence on the results, as long as $\Lambda$ is varied within reason. These features make this procedure less parameter reliant and easier to use.

**Algorithm 1:** Causal Inference Algorithm with Stability Selection

```
Input: Time-series of descriptive metrics ($TS$); Response
       scores ($RS$); time-lag ($lag$);
Output: Causal levers $\hat{S}^{\Lambda}$
Initialize: Pre-process $TS$ and $RS$;
Stratify $TS, RS$ using $lag$ to create $Xtrain, ytrain, Xtest, ytest$;
Fit a sparse regression model using $Xtrain, ytrain$ and tune
  for $\Lambda$ using validation set. Apply model on $ytest$ to report
  prediction metrics.;
for $j = 1, \cdots, |\Lambda|$ do
    for $i = 1, \cdots, 100$ do
        Re-sample set of size $\frac{nrows(Xtest)}{2}$;
        Fit a sparse regression model using parameter $\Lambda_j$ and
          compute set of selected variables;
    end
end
$\hat{S}^{\Lambda} = \{\max_{\lambda \in \Lambda} \hat{\Pi}^{\lambda}_{k} \geq \pi_{thr}\}$;
```

## 2.2 Proposed Algorithm

In Algorithm 1, we propose a new procedure for inference of causal levers from time series data using Granger causal regression. The time-series ($TS$) consists of the pre-processed World Development Indicators and the GCI scores ($RS$). $TS$ is split according to the lag to create training (validation) and test data. Using the training data, we learn a sparse regression model and tune the regularization parameter over a validation set to identify a stable range of parameters which can be used for evaluation on the test set.

Subsequently, a selection probability matrix is initialized for storing the averaged feature selection probabilities while varying the regularization parameter in the specified range. The number of re-samplings is set to 100. The selection probabilities are calculated after re-sampling from the data iteratively after running the sparse regression algorithm for each parameter. These values are averaged and only those features (causal levers) are selected whose maximum selection probability exceeds the user defined threshold ($\pi_{thr} \in [0.6, 1)$).

# 3 INNOVATIVENESS CASE STUDY

## 3.1 Dataset Description

We consider the WDI dataset published by the World Bank which represents the most updated and accurate global development data available.[3] The indicators therein measure the progress of countries in different sectors such as achieving sustainable development, helping vulnerable populations, and reducing gender disparities.[4] The seven prominent categories among WDI are:

- *World View*: indicators pertaining to the size of the economy, different global goals, and women in development,

[3] http://databank.worldbank.org/data/download/archive/WDI_excel_2014_04.zip

[4] http://wdi.worldbank.org/tables

- *Poverty and Shared Prosperity*: poverty at national and international lines and distribution of income and prosperity indicators,
- *People*: a broad range of indicators such as (i) education, (ii) unemployment, (iii) health and diseases, and (iv) labor force structure,
- *Environment*: indicators in this category are (i) agriculture, (ii) electricity production and sources, and (iii) sustainable Energy,
- *Economy*: indicators related to (i) merchandise exports and imports, (ii) national income and other monetary indicators,
- *States and Markets*: indicators related to (i) private sector, (ii) public policy, and (iii) science, technology and statistical capacity, and
- *Global Links*: indicators related to (i) travel and tourism, (ii) debt and financial flows, and (iii) direction of trade of low and middle economies.

The following steps were followed for preprocessing the WDI data. We use smoothing to deal with missing values present in the the covariates. Pairwise year differences are calculated for the WDI indicators across years and they are smoothed by taking the mean over three years. The data is validated by converting country names to lowercase and renaming some countries. We deleted 45 countries which did not represent proper countries. We computed the set of commonly available metrics across 7 years by intersecting the WDI metrics.

The response variable for our analysis is the GCI score obtained from the World Economic Forum.[5] GCI data was preprocessed by removing redundant columns. The data considered was from years 2007 to 2015. The time series of scores was constructed for each individual year from 2007 to 2015. The non-innovation features (11 pillars) and innovation scores were separated. Innovation scores were averaged for every 3 years and stored. We use this for filling missing innovation scores for each country. We also computed nearest neighbors using WDI metrics for all countries which have both WDI and GCI scores. 5 nearest neighbors are computed using the cosine distance. We computed the set of countries present across all 7 years in both WDI and GCI data.

After preprocessing the WDI indicators and the GCI scores, we created a time-series of the WDI metrics and GCI scores from the years 2009–2015. We set the time-lag to 4 years, where we used 2009 WDI features and 2011 GCI sores to learn a prediction model. This was tested using 2013 WDI features. The predictions obtained were compared to 2015 GCI scores. The final dataset considered has 146 rows (countries) and 1500 columns (WDI metrics).

## 3.2 Experimental Setup

The code for preprocessing the WDI and GCI datasets was written in Python 2.7 and we used the *pandas* [8] library and their associated functions for performing all the steps mentioned earlier. The concave regularizer we used for sparse regression is the Minimax Concave Penalty [13]. This has a closed form solution in the unidimensional case and it can be solved using coordinate-descent-based optimization methods. We implemented the sparse regression component in Python 2.7 and invoked the R packages and optimization solvers through the *rpy2* interface [2]. This was done in order to leverage state-of-the-art optimization solvers which were readily available as R packages. The $\gamma$ value for this minimax concave penalty was fixed in a suitable range to obtain solution paths which are not very unstable. The rationale for this setting is based on understanding the properties of the regularizer which we have studied rigorously in an independent machine learning submission and these details are beyond the scope of this paper. We used matplotlib to generate the barplots [4]. The radial plots were generated using the *plotrix* library in R [10].

## 3.3 Causal Levers Inferred

In Figure 2, we visualize the contribution of different levers for predicting the GCI score as learned by our approach. We consider levers from eight sectors which are labelled in the diagram where each color corresponds to a unique sector. These colors can be mapped to the labels on the right side of the radial plot (such as Economic Policy and Debt) in a counter-clockwise manner. For each sector, we considered one or multiple levers which were representative of that sector. The radial plots demonstrate which sectors were responsible for fostering innovation in that country.

The countries selected for our study here are (i) *Bangladesh* — a developing country in Asia which would benefit heavily through effective policymaking, (ii) *Pakistan* — a developing country in Asia which is one of Bangladesh's nearest neighbors according to their WDI metrics, and (iii) *Singapore* — a well-developed country which contrasts *Pakistan* and *Bangladesh* on studying the impact of potential levers on innovation.

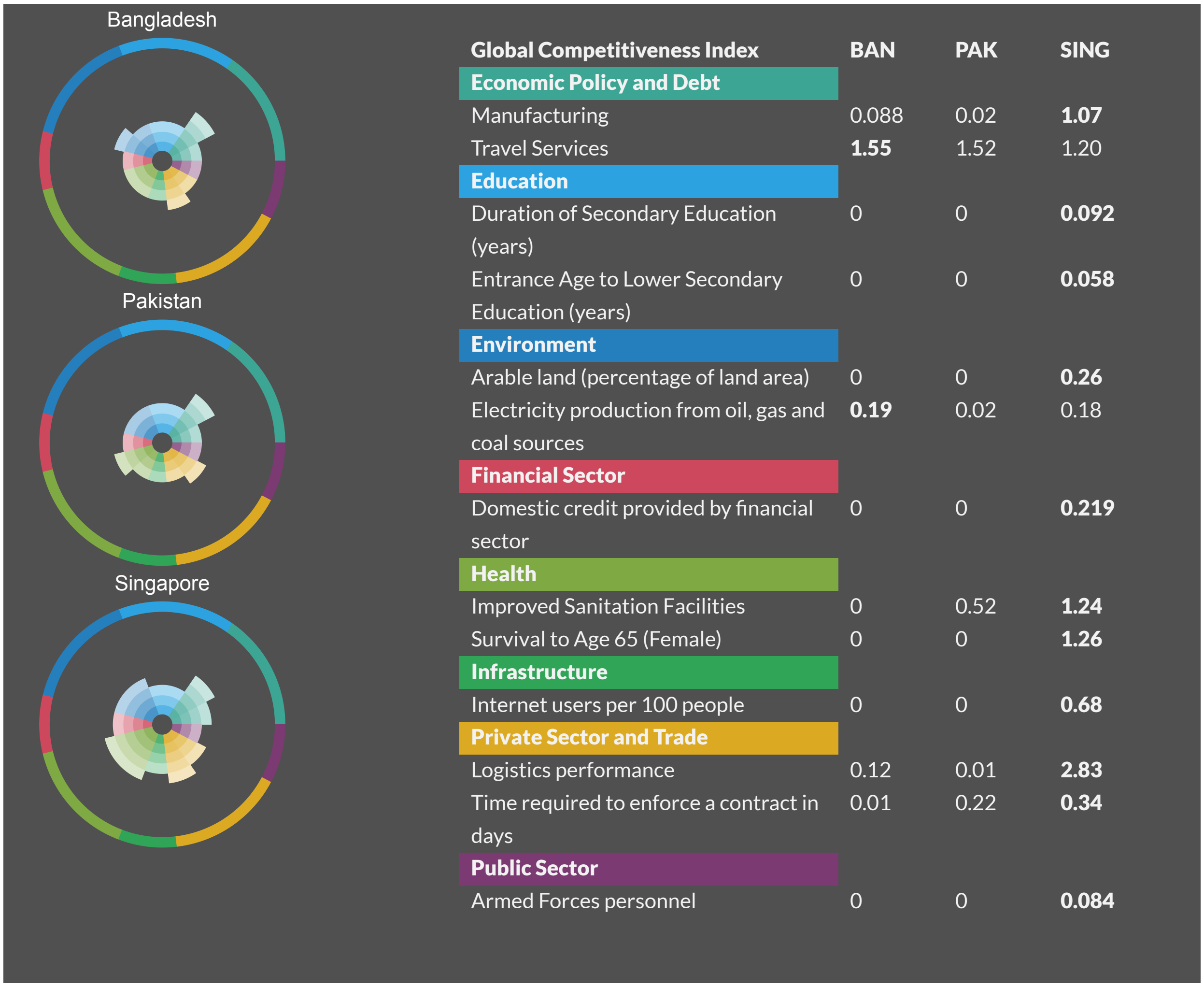
The contribution for each lever is calculated by computing the product of the normalized features for that lever and their corresponding learned regression model coefficients. These model coefficients are the causal strength of these levers. The contribution for each sector is the sum of the contributions of individual levers in that sector. Figure 2 shows a table of the causal strengths along with a visualization via radial plots (logarithmically transformed for better visual presentation).

From this figure, one can observe that innovativeness of both *Bangladesh* and *Pakistan* is caused to a large extent by the Economic Policy and Debt sector through Travel Services. In the Environment sector, electricity production from oil, gas and coal sources makes a good contribution to innovation for *Bangladesh* compared to *Pakistan* and *Singapore.* We found independent evidence for the significance of this lever for *Bangladesh* from the results provided by the Macro Economic Meter.[6] *Pakistan* has some contribution to innovation from the Health sector through improved sanitation facilities which is missing for Bangladesh. Both countries seem to be trailing *Singapore* significantly in the Infrastructure and Financial sectors. The Education sector is also not contributing to the GCI score for *Bangladesh* and *Pakistan.*

We can validate some of the results for *Bangladesh* obtained by our approach by comparing to a survey conducted by the World

[5] https://www.weforum.org/reports/the-global-competitiveness-report-2016-2017-1

[6] http://mecometer.com/whats/bangladesh/electricity-production-by-source/



| **Global Competitiveness Index** | **BAN** | **PAK** | **SING** |
|---|---|---|---|
| **Economic Policy and Debt** | | | |
| Manufacturing | 0.088 | 0.02 | **1.07** |
| Travel Services | **1.55** | 1.52 | 1.20 |
| **Education** | | | |
| Duration of Secondary Education (years) | 0 | 0 | **0.092** |
| Entrance Age to Lower Secondary Education (years) | 0 | 0 | **0.058** |
| **Environment** | | | |
| Arable land (percentage of land area) | 0 | 0 | **0.26** |
| Electricity production from oil, gas and coal sources | **0.19** | 0.02 | 0.18 |
| **Financial Sector** | | | |
| Domestic credit provided by financial sector | 0 | 0 | **0.219** |
| **Health** | | | |
| Improved Sanitation Facilities | 0 | 0.52 | **1.24** |
| Survival to Age 65 (Female) | 0 | 0 | **1.26** |
| **Infrastructure** | | | |
| Internet users per 100 people | 0 | 0 | **0.68** |
| **Private Sector and Trade** | | | |
| Logistics performance | 0.12 | 0.01 | **2.83** |
| Time required to enforce a contract in days | 0.01 | 0.22 | **0.34** |
| **Public Sector** | | | |
| Armed Forces personnel | 0 | 0 | **0.084** |

**Figure 2: Contribution of Levers from different sectors for predicting the GCI score as determined by CLEVER algorithm for Bangladesh, Pakistan and Singapore in the year 2015.**

Economic Forum to identify problematic factors for doing business in Bangladesh. In this survey, participants were asked to select the five most problematic factors from a list of factors (levers) with rankings ranging from 1 (most problematic) to 5. The levers were weighted by their ranking to compute an aggregated score which is shown in Figure 3. This survey ranks inadequate supply of infrastructure and access to financing as important issues to address for policymakers in *Bangladesh*. As illustrated in Figure 2, our data-driven approach assigns zero score to infrastructure and financial sector-based levers for *Bangladesh* compared to Singapore, therefore, suggesting that policymakers need to act on these levers for improving the innovativeness of *Bangladesh*. Similarly, poor work ethic in national labor force and poor public health in Figure 3 were also captured by the proposed algorithm by assigning zero scores for public sector and health sector-based levers. Therefore, the levers captured by our data-driven causal inference approach are similar to those identified through an independent study.

Apart from the visible benefits of using this approach for developing countries like *Bangladesh* and *Pakistan*, there are some takeaways for a developed country such as *Singapore* also. For example, the Education sector in Singapore can improve significantly to further improve the country’s innovativeness.

## 4 CONCLUSION

In this paper, we presented a data-driven approach for causal inference from a real-world dataset consisting of World Development Indicators for predicting the response innovation. Our approach was able to address several challenges associated with this task, namely, temporality, high-dimensional correlation and sparsity, and noise. The proposed algorithm unified two different algorithmic paradigms: (i) concave regularizers-based sparse regression and (ii) randomization-based stability selection. The former resulted in unbiased estimation with improved performance under

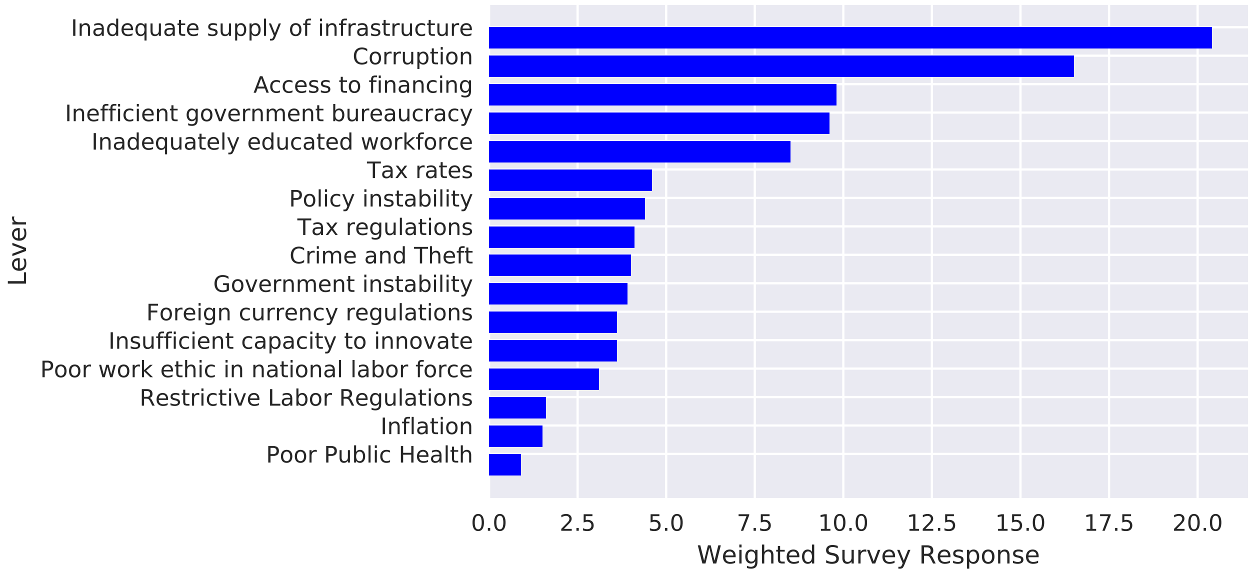


**Figure 3: Most problematic factors for doing business in Bangladesh as calculated through a survey conducted by the World Economic Forum.**

high levels of correlation and noise, and the latter reduced the variability of results by providing effective false positive control. We used this framework to identify the levers for a developing country (*Bangladesh*) and compared these levers to a similar developing country (*Pakistan*) and a well-developed country (*Singapore*). We validated some of our inferences using results published by subject matter experts and we identified some degree of similarity between these results.

Overall, the work herein represents an effective approach for the general problem of advising policymakers in a data-driven way when causal inferences are not easily obtained due to difficult data characteristics. This work can be extended by incorporating false negative control which would guarantee selection consistency. Alternative regularizers such as Ordered Weighted $\ell_1$ norm (OWL) [1] can also be incorporated to handle correlation and noise. We can also further validate our findings by collaborating with subject matter experts who can independently assess each lever for a country.

## ACKNOWLEDGMENTS

We acknowledge the contributions of Karthikeyan Natesan Ramamurthy and Prasanna Sattigeri for their comments on this work and providing their code.

## REFERENCES

[1] Mario Figueiredo and Robert Nowak. 2016. Ordered weighted l1 regularized regression with strongly correlated covariates: Theoretical aspects. In *Artificial Intelligence and Statistics*. 930–938.

[2] L Gautier. 2008. rpy2: A Simple and Efficient Access to R from Python. *URL http://rpy. sourceforge. net/rpy2. html* (2008).

[3] Clive WJ Granger. 1988. Causality, cointegration, and control. *Journal of Economic Dynamics and Control* 12, 2-3 (1988), 551–559.

[4] John D Hunter. 2007. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering* 9, 3 (2007), 90–95.

[5] Caitlin Kuhlman, Karthikeyan Natesan Ramamurthy, Prasanna Sattigeri, Aurélie Lozano, Lei Cao, Chandra Reddy, Aleksandra Mojsilović, and Kush R Varshney. 2017. How to Foster Innovation: a Data-Driven Approach to Measuring Economic Competitiveness. *IBM Journal of Research and Development* (2017), 1–19.

[6] Aurélie C Lozano, Prasanna Sattigeri, Aleksandra Mojsilović, and Kush R Varshney. 2016. Stable estimation of Granger-causal factors of country-level innovation. In *Signal and Information Processing (GlobalSIP), 2016 IEEE Global Conference on*. IEEE, 1290–1294.

[7] Yonglong Lu, Nebojsa Nakicenovic, Martin Visbeck, A Stevance, et al. 2015. Five priorities for the UN sustainable development goals. *Nature* 520, 7548 (2015), 432–433.

[8] Wes McKinney. 2012. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython.* " O'Reilly Media, Inc.".

[9] Nicolai Meinshausen and Peter Bühlmann. 2010. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72, 4 (2010), 417–473.

[10] Thomas Rahlf. 2017. *Data Visualization with R*. Springer.

[11] Prasanna Sattigeri, Aurélie Lozano, Aleksandra Mojsilović, Kush R Varshney, and Mahmoud Naghshineh. 2016. Understanding Innovation to Drive Sustainable Development. *ICML Workshop on Data4Good: Machine Learning in Social Good Applications* (2016), 21–25.

[12] Klaus Schwab and Xavier Sala-i Martin. 2015. World Economic Forum's Global Competitiveness Report, 2014-2015. *Retrived from* (2015).

[13] Cun-Hui Zhang et al. 2010. Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics* 38, 2 (2010), 894–942.