USING R TO UNCOVER INNOVATION LESSONS FOR EMERGING MARKETS

Source Code (available at https://github.com/AlecE-123/Repository-1), Output and Report by Alec E. *December 2019*

EXECUTIVE SUMMARY

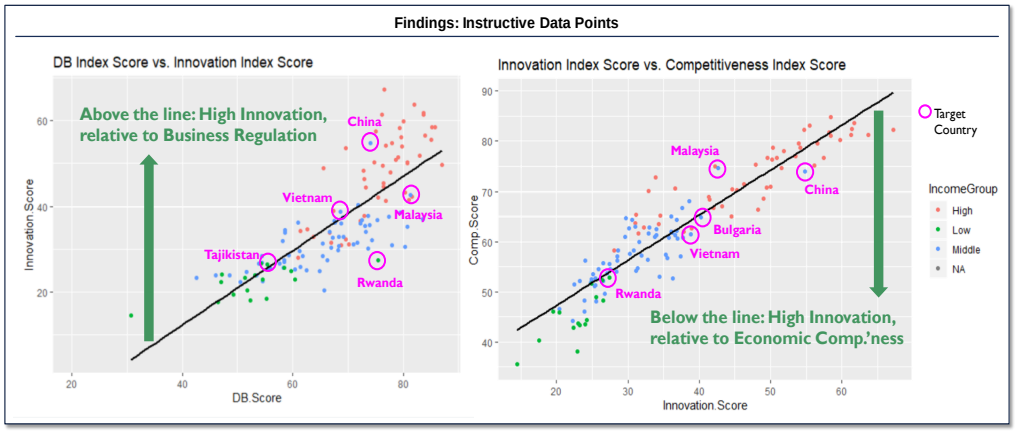
The purpose of this project is to use R to analyze the relationships between three country-level development indices to see if they can reveal any lessons or policy implications about how governments in emerging markets can promote innovation within their own borders.

- **Hypothesis:** A regression analysis of the three datasets can identify specific countries that are outperforming with regard to innovation, despite their comparatively lagging developmental context.
- **Rationale:** A novel theoretical approach to "innovation" suggests that there are specific lessons to be learned from countries that are leapfrogging their way to innovation success, sidestepping the traditionally-emphasized impediments of weak business regulation and poor economic development.

This analysis is an extension of work and recommendations I was involved in during an "Action Learning" project, as part of my current MBA program at the Asia School of Business in collaboration with MIT Sloan Management. The project was hosted by a real estate developer in Bangkok, Thailand and the goal was to identify high-impact policies, strategies and initiatives that governments around the world are using to successfully promote innovation, particularly in emerging market economies.

Though our Action Learning project had good results and we identified various governments that had made meaningful progress on innovation, there was surely more to learn. I undertook this R coding project in order to identify countries that might offer additional lessons and insights on how to achieve innovation. I used regression of country-level data points to establish the relationships between the development indices, and then looked for outliers that were "more innovative" than the overall relationship suggested, based on corresponding strength of regulatory environment and economic competitiveness. The regressions I coded were:

1. *Innovation vs. Business Regulatory Environment*
2. *Innovation vs. Economic Competitiveness*



The table below summarizes my findings. There were four countries that stood out as "successful outliers" when compared to the overall trend in both regressions: China, Malaysia, Vietnam and Rwanda. There were two additional countries that appeared as outliers in one of the regressions: Tajikistan and Bulgaria. This list, however, is partial and represents my assessment of the best candidate countries – there of course are others.

	Innovation Advantage Relative to:	
Study Priority	Business Regulatory Env.	Economic Competitiveness
HIGH	China	China
HIGH	Malaysia	Malaysia
HIGH	Vietnam	Vietnam
HIGH	Rwanda	Rwanda
MEDIUM	Tajikistan	Bulgaria

These findings also somewhat validated the results of our Action Learning Project, as we had spent a lot of time studying China and Malaysia as interesting emerging market case studies. The remaining names either received much less attention or none at all. The concluding argument of this R project is that those names remain low-key but high-potential subjects of study and may offer valuable insights about the specific dynamics, resources and interactions that can be created and leveraged in order to achieve innovation in emerging markets.

USING R TO UNCOVER INNOVATION LESSONS FOR EMERGING MARKETS

Source Code (available at https://github.com/AlecE-123/Repository-1), Output and Report by Alec E. *December 2019*

BACKGROUND

The purpose of this project is to use R to analyze the relationships between three comprehensive, country-level development indices (see Figure 1) to see if they can reveal any lessons or policy implications about how governments in emerging markets can promote innovation within their own borders. These reports are generally reviewed and used independently, but I wanted to see if integrating the data could uncover anything interesting.

Figure 1. Report Indices Used

Index 1: Doing Business 2019	*Index 2: Global Competitiveness Report 2019*	*Index 3: Global Innovation Index 2019*
		
THE WORLD BANK	WORLD ECONOMIC FORUM	JOHNSON Cornell University INSEAD WIPO WORLD INTELLECTUAL PROPERTY ORGANIZATION
"Measures complex regulatory processes by zeroing in on their quantifiable components"	*"Provides a detailed map of the factors and attributes that drive productivity, growth and human development"*	*"GII will continue to be a laboratory for measuring and analyzing emerging innovation data"*

Given the complexities of measuring country-level development, there is an inherent level of uncertainty in these indices and their underlying data. However, these reports are annually produced by institutions that are generally considered to be authorities in their respective fields of research, and offer the most accurate available data.

The first two indices – *Doing Business 2019* and *The Global Competitiveness Report 2019* – are developmental indicators that measure the efficacy of business regulation, and the general competitiveness of economies, respectively. They offer macro assessments of the general business environment and economic resources available in a country.

The third index – *Global Innovation Index 2019* – measures how innovative economies are. The purpose of this analysis is to understand how countries can achieve innovation; I'm therefore considering this third index as the dependent variable or output, while I've treated the other two indices as independent variables or inputs.

However, I wasn't looking for countries just with high innovation scores (the output) to see how they measured up on the other indices (the inputs). This likely would have led to the predictable conclusion that if a country wants to achieve a high innovation score, it must improve its scores on the other indices, because they usually progress together. This is not a particularly useful observation for the governments of middle-income countries.

Rather, I wanted to identify countries that have high innovation scores *relative* to low scores on the other indices, and / or relative to their income group (i.e. low, middle, or high based on GNI per capita). The reason I used this approach is because the findings of a prior project I was involved in suggested revisions to the "traditional" theory and practice of innovation in emerging markets. The next section provides further details but in short: *innovation is not just the inevitable output of a robust, high-income economy – innovation potential can be realized in weaker environments and with fewer resources if certain initiatives, structures and interactions are supported, regardless of per capita income levels.*

I wanted to find countries with poor regulatory environments and / or low economic competitiveness but that still had strong innovation levels. This would suggest that, relative to its otherwise weak position, that a particular economy possessed some inherent dynamic or resource that enabled innovation to flourish. I also wanted to find countries with innovation scores that were higher than those in the next income group (e.g. a middle-income country outscoring a high-income one). Again, this would indicate that the former was doing something right in terms of innovation that was enabling it to outpace its higher-income peers.

Identifying "innovative" economies in this way, and then evaluating them closely, might then uncover useful insights or policy lessons for governments in emerging markets looking to replicate such innovation success.

USING R TO UNCOVER INNOVATION LESSONS FOR EMERGING MARKETS

Source Code (available at https://github.com/AlecE-123/Repository-1), Output and Report by Alec E. *December 2019*

PROJECT CONTEXT AND MOTIVATION

This project was inspired by, and builds on, my work during an Action Learning project conducted during Fall 2019.[1] The goal of the project, located at a real estate developer in Bangkok, was to dig into the problem of how emerging market governments can directly support and promote innovation within their own borders. Our host company wanted us to identify global best practices and identify specific policies, initiatives, and mechanisms that could be applied to still-developing economies. Though numerous definitions exist, we used the following for our purposes:

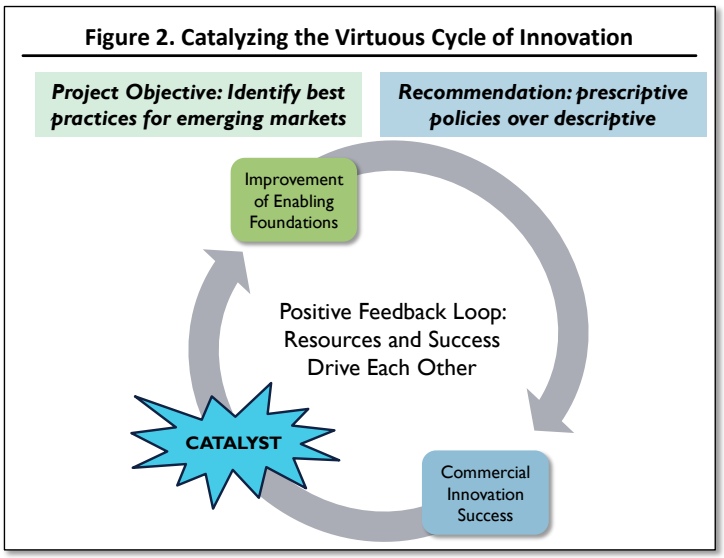
*Innovation = Invention * Commercialization*

By this definition, innovation requires not only a product, service or process breakthrough, but also some creation and capture of value via commercialization. Because the ability to sustainably produce innovation of this kind is key to economic "upgrading" (for example, escaping the well-documented middle income trap), the concept has become a global development priority.

During the project, we reviewed a variety of traditional theories of innovation. These academic approaches were predominantly measurement-driven: they identified the countries that were the best at innovation, measured a wide variety of "input" variables within those countries, and then concluded that in order to achieve a similar level of innovation, other countries had to achieve similar values for those input variables. The implication being, that innovation is the inevitable output.

This "descriptive" approach was too broad, too long-term and too passive to recommend to our client. We wanted to dig into something more precise, more targeted, and more actionable. We rethought the concept of how innovation fundamentally transpires and with further primary and secondary research, we realized that it has much more to do with the tactical, individual-level interactions and opportunities that a community can foster. Of course, it helps to have highly educated people and massive resources concentrated in locations like Cambridge, MA or Silicon Valley, CA. But it is also extremely difficult and costly to replicate those unique environments. The point is that for any given community, the right structures and support can help realize whatever innovation potential already exists there, regardless of how small it may seem on global scale.

Figure 2. Catalyzing the Virtuous Cycle of Innovation



Our team therefore recommended "prescriptive" policies to our host, which, rather than focusing on long-term pillars of development, were a set of specific initiatives that could be implemented in the near-term and that could be scaled up or down, depending on whatever level of resources was available. We had conceptualized development as a virtuous cycle (see Figure 2: commercial success will improve the surrounding economic environment, which will drive further success). But we also saw that such a cycle only builds momentum once initiated – it requires a catalyst. The "prescriptive" policies we recommended were precisely intended to serve as such catalysts, achieving innovation success with whatever people and resources already existed, kickstarting the cycle of development and leapfrogging the "traditional" long-term development models.

[1] Action Learning ("AL") is a feature of my current MBA program at the Asia School of Business in collaboration with MIT Sloan Management. AL projects function as compressed management consulting engagements, or "field" case studies. A team of 3-4 students is assigned to a host company located in the ASEAN region and travels onsite to address a live business challenge.

USING R TO UNCOVER INNOVATION LESSONS FOR EMERGING MARKETS

Source Code (available at https://github.com/AlecE-123/Repository-1), Output and Report by Alec E. *December 2019*

EXTENDING THE AL PROJECT WITH R

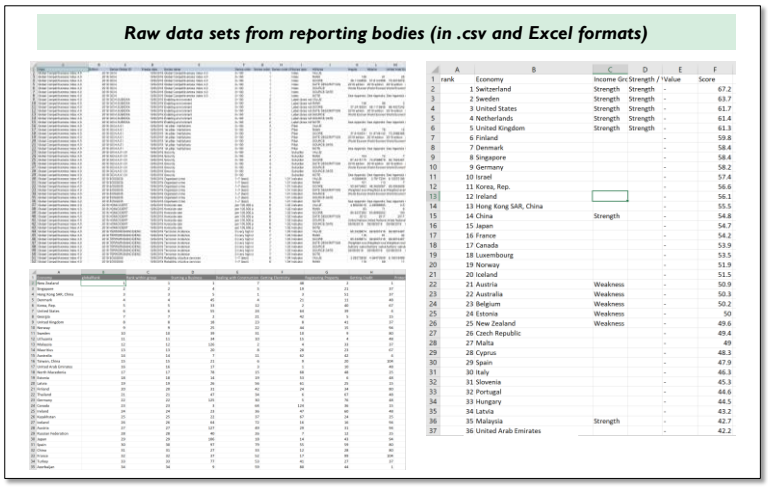
The recommendations we made to our AL host company were the result of much labor-intensive primary and secondary research including visits to Singapore, China, Thailand and Malaysia. We conducted interviews with professors and experts in the United States and Europe and attended a variety of conferences and meetings with businesses and government agencies. We combed through the academic literature on innovation and the middle-income trap, as well as case studies, articles, white papers and development reports (including those used for this R analysis). Our task was then to document, synthesize and make sense of all this information.

The reason I describe this process is to highlight the potential value of the output of this R analysis. During the AL project, we did not always know where to look for the most instructive case studies and were often surprised by where we did, or did not, find them. We canvassed as widely as we could, but what is certain is only that we have not looked everywhere – there are sure to be other best practices and new ideas being implemented out there. While our efforts ultimately yielded a valuable contribution, the challenge of solving innovation remains open-ended and unsolved.

This R project aims to make further progress on this challenge. The final output is simply a list of countries that are high-priority candidates for further research on innovation in emerging markets. The list provides partial validation of our prior findings, and a starting point for research into new countries. If they are somehow tapping into little-known economic, political, social, or cultural dynamics that benefit innovation, perhaps they can be studied and replicated in other contexts.

DATA SOURCES

I used three datasets for this project – the publicly-available data for each of the three index reports in question. I obtained the raw data directly from the reporting bodies.[2] For purposes of replicability, below I've described some minor data cleaning and tidying I completed in Excel and then in R.



The Excel work involved matching up country names across data sets. For example, the United States was sometimes listed as "United States" and sometimes as "United States of America." I manually eliminated these discrepancies so the data tables could be properly joined for analysis.

Data cleaning exercises in R included:

- Filtering to include only the desired metrics, as many sub-metrics are included in the datasets.
- Joining the datasets into a single table so it could be used for integrated analysis. This also required the removal of certain NA values, as not all datasets included the same list of countries, likely due to availability of data.
- Coding each observation / country for income level, so I could create the income group-based subsets.

[2] The full reports and data sets can be accessed at the following links:

Report	URL
Doing Business 2019	https://www.doingbusiness.org/en/reports/global-reports/doing-business-2019
Global Competitiveness Report 2019	https://www.weforum.org/reports/global-competitiveness-report-2019
Global Innovation Index 2019	https://www.wipo.int/publications/en/details.jsp?id=4434

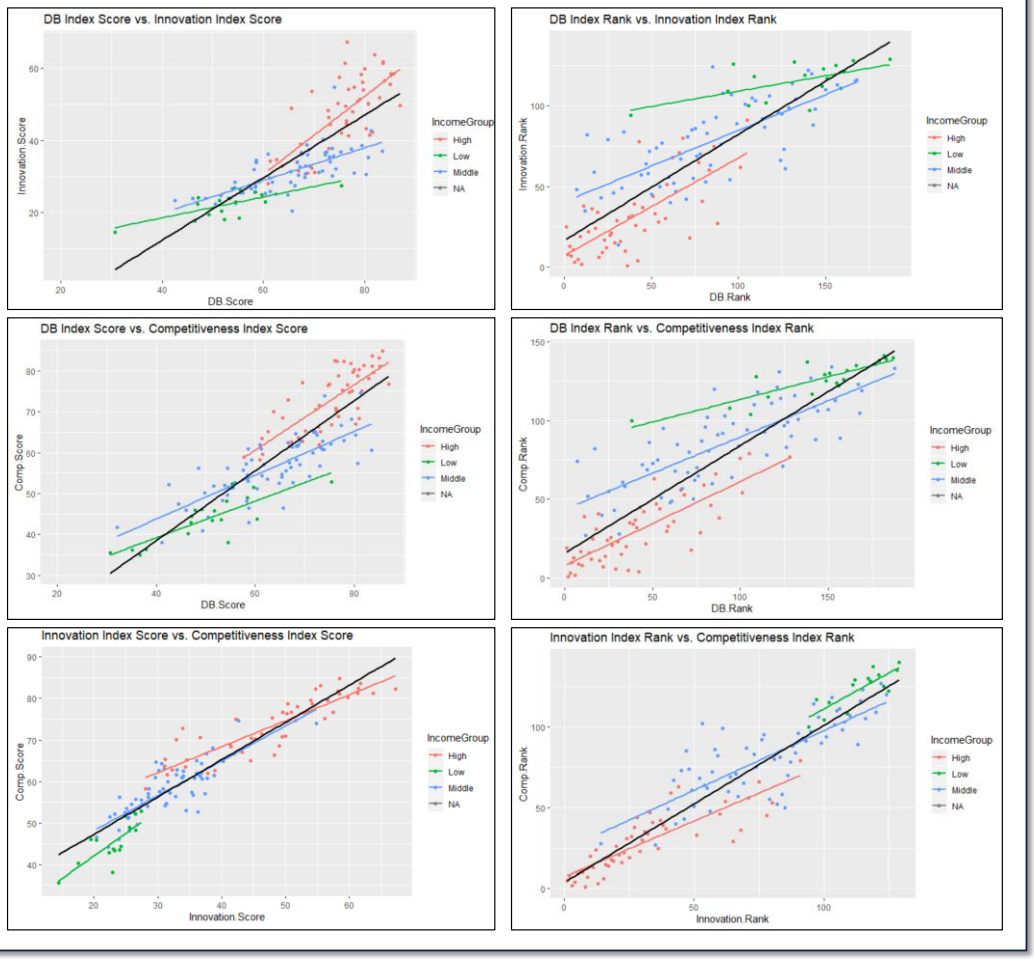
USING R TO UNCOVER INNOVATION LESSONS FOR EMERGING MARKETS

Source Code (available at https://github.com/AlecE-123/Repository-1), Output and Report by Alec E. *December 2019*

ANALYSIS AND INITIAL OBSERVATIONS

After cleaning the data, I had four datasets to work with. The first contained all countries (n=192); the second, high-income countries (n=59); the third, middle-income (n=102); and the fourth, low-income (n=30). The three indices were included in each dataset: Doing Business ("DB", regulatory quality), Competitiveness (economic strength) and Innovation.

Figure 3. Regression Output – Segmented by Income Group

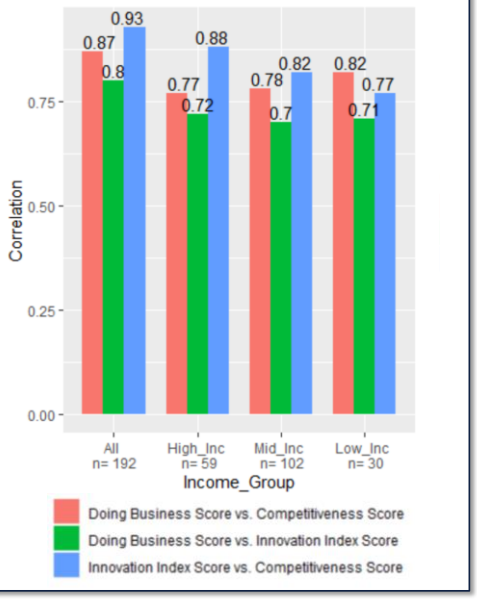


The datasets contained two values for each of the three indices: index SCORE and index RANK. The "score" value was the output of the respective reports, a blended metric that reflected measurements across numerous variables as set forth by the report's methodology. The "rank" value was simply how the country ranked globally, based on that score. Each observation row tied back to a country, so every country had six values, (e.g. for Singapore, there was both a SCORE and RANK value for each of the three indices).

In order to identify the "successful outliers" I produced multiple regression variations. I could then evaluate the full dataset, as well as the income group segments (see Figure 3). This also allowed me to evaluate the relative strength of the correlations between the indices (see Figure 4). I noted the following initial observations:

- Innovation and Competitiveness Scores are most tightly correlated.
- Doing Business and Innovation Scores are least tightly correlated.
- Regressions are strongest in the full sample (i.e. they weaken when segmented by any income group).
- There is no clear relationship between the strength of correlations across income groups (e.g. sometimes strengthens, sometimes weakens, when moving from High to Low income and vice versa).

Figure 4. Summary of Regression Coefficients



USING R TO UNCOVER INNOVATION LESSONS FOR EMERGING MARKETS

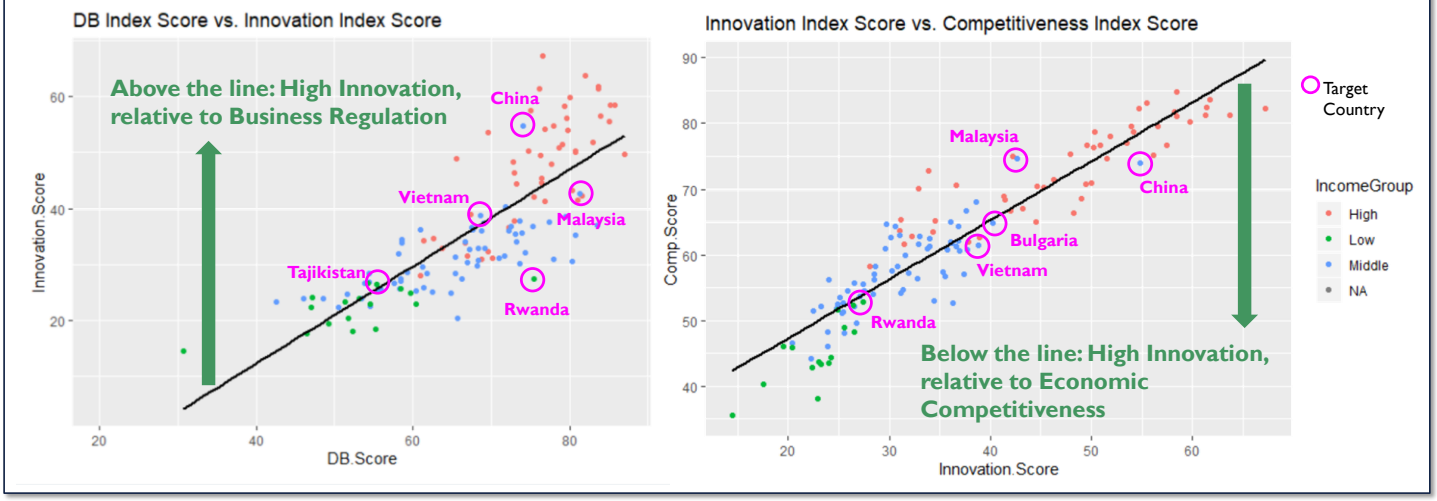
Source Code (available at https://github.com/AlecE-123/Repository-1), Output and Report by Alec E. *December 2019*

FINDINGS AND RECOMMENDATIONS

Not all regressions were equally useful. I was interested in the scores because they represented objective data for each country (ranks were less helpful). I selected two regressions for closer interpretation (see Figure 5):

1. Innovation Score vs. Doing Business Score
2. Innovation Score vs. Competitiveness Score

Figure 5. Instructive Data Points



I identified target countries based on their variance from the trend of the overall regression or income group relationships. I used these criteria to determine if a country was an interesting data point:

- If a country was scoring poorly on either Doing Business or Competitiveness, but still scoring relatively highly on Innovation, using the overall regression line as a predictive benchmark. (China, Vietnam and Tajikistan in the first regression; China, Bulgaria, Vietnam and Rwanda in the second).
- If a country's innovation score was better than those of countries in a higher income group. (Malaysia, China and Rwanda in the first regression; Malaysia ,China and Rwanda in the second regression.
- Low- or middle-income countries only, since we are looking at emerging markets.

Figure 6. Summary of Findings

Study Priority	Innovation Advantage Relative to: Business Regulatory Env.	Economic Competitiveness
HIGH	China	China
HIGH	Malaysia	Malaysia
HIGH	Vietnam	Vietnam
HIGH	Rwanda	Rwanda
MEDIUM	Tajikistan	Bulgaria

The list I compiled (see Figure 6) is only partial. I aimed to compile a list of countries that best satisfies all three of the above criteria. I used this criteria because the regression coefficients indicate that there is a strong relationship between each of these three indices. That is, Doing Business scores and Competitiveness scores are reasonably accurate predictors of a country's innovation level. Therefore, the "successful outliers" (i.e. countries that are successful at innovation, relative to where the general trend would predict them to be) were the most instructive data points.

I would therefore recommend to policymakers that they conduct in-depth reviews of the policy environments and economic dynamics of these countries. This analysis suggests that they are in some way defying their macroeconomic developmental context to produce disproportionately high levels of innovation.

USING R TO UNCOVER INNOVATION LESSONS FOR EMERGING MARKETS

Source Code (available at https://github.com/AlecE-123/Repository-1), Output and Report by Alec E. *December 2019*

APPENDIX A – OVERALL REGRESSION OUTPUTS (NOT SEGMENTED BY INCOME GROUP)