3-2022

Indicators of Political Instability in the Presence of Rapid Urban and Youth Population Growth

Douglas W. Hubbard

INDICATORS OF POLITICAL INSTABILITY IN THE PRESENCE OF RAPID URBAN AND YOUTH POPULATION GROWTH

THESIS

Douglas W. Hubbard, GS-11, USAF

AFIT-ENS-MS-22-M-139

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-22-M-139

INDICATORS OF POLITICAL INSTABILITY IN THE PRESENCE OF RAPID URBAN AND YOUTH POPULATION GROWTH

THESIS

Presented to the Faculty

Department of Operations Research

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Douglas W. Hubbard, BS

GS-11, USAF

March 2022

AFIT-ENS-MS-22-M-139

INDICATORS OF POLITICAL INSTABILITY IN THE PRESENCE OF RAPID URBAN AND YOUTH POPULATION GROWTH

Douglas W. Hubbard, BS

GS-11, USAF

Committee Membership:

Dr. Richard F. Deckro, DBA
Advisor

LTC Phillip M. LaCasse, PhD
Reader

AFIT-ENS-MS-22-M-139

Abstract

Large and rapidly growing cities and other urban agglomerations have the potential to become incubators of political instability. This is especially true of rapidly growing cities which are located in countries that are also experiencing high rates of growth in their youth population. Rapid growth rates put stress on urban infrastructure and other institutions, and these stresses can cause major problems for both city and national governments. Knowing when these cities and countries may be trending toward their tipping points regarding political instability will help governments and international organizations develop and implement effective strategies to mitigate the risk of instability.

Acknowledgments

I would like to express my sincere appreciation to my faculty advisor, Dr. Richard Deckro, and my reader, LTC Phillip LaCasse, for their guidance and support throughout the course of this thesis effort. In addition, I would like to thank my supervisors who have made this opportunity possible. Finally, and most importantly, I would also thank my family for their support during the course of this program.

Douglas W. Hubbard

Table of Contents

Page

List of Figures

Page

List of Tables

Page

INDICATORS OF POLITICAL INSTABILITY IN THE PRESENCE OF RAPID URBAN AND YOUTH POPULATION GROWTH

I. Introduction

Background

The human population is rapidly increasing throughout the world. The United Nations estimates for world population was 4.458 billion in 1980 and had risen to 7.795 billion by 2020 (United Nations, accessed 2020). That is an average 1.41% growth annually or an estimated 74.9% total increase in just four decades.

Recently, the largest rates of growth are occurring in Asia and Africa, with most of that growth happening in urban areas (Liotta and Miskel, 2009). The United Nations (UN) defines a megacity as an urban area with a population of ten million people or more (Liotta and Miskel, 2009). In this study, rapid urban growth countries will be the term used because some of the largest rates of urban growth are currently happening in countries with cities that are below the UN threshold.

While some very large urban areas may not currently be problems, the rate of growth, regardless of an urban area's size, is a key indicator in potential instability (Bettencourt, 2020). Often co-occurring with this urban population boom, the relative size and growth of a country's youth cohort can put additional strain on a country's economy, institutions, and infrastructure. For an up-and-coming nation needing stability for economic and societal growth, such periods of transition are challenging.

Economically speaking, having sufficient jobs providing sufficient income as the youth population enters the workforce is a noble goal. Unfortunately, the youth unemployment rate in Sub-Saharan Africa averaged 7.3% in 2019, while, in Africa, the

overall highest rate recorded was 15.4% in Tunisia (Ndjie, Ondoa, and Tabi, 2019). The associated pressures weigh heavily on various forms of infrastructure, health services, educational institutions, the job market, and the local and national economies. Internal migration, mainly from rural areas to these rapidly growing urban areas, generally exacerbates these issues.

When a nation's institutions and infrastructure are having difficulty keeping pace with the growth of their nation's urban population, decent living conditions, proper security services, and adequate infrastructure may not be available. As the local social order and traditional customs unravel, such places can form lucrative recruiting grounds for criminal gangs, unregulated militias, and anti-government insurgents.

For the reasons discussed above; policy makers, both those in the struggling countries and those in foreign countries wanting to provide aid, will need to be able to recognize the warning signs with enough lead time to put necessary plans in place.

Focus and Goals

This research is focused on rapidly growing urban populations, their country's youth cohort, and the myriad pressures that rapid growth rates can induce. Ultimately, such concerns in rapid growing urban areas place considerable pressure on political institutions.

To aid in determining which countries are either nearing a tipping point for instability, or are already there, this research's goal is to build an analytical procedure that will calculate a country fragility rating and a categorical stability grouping. This procedure will be verified using currently available data, the results be discussed, and the

procedure will be documented. The final product will be the outline of a process that can be quickly executed using readily available data and commonly available software. This process will work with a small group of six independent variables, use commonly available software, generating ratings for countries, both globally and regionally. These ratings can be compared to a research-based tipping point (see Table 1). Analysts and decision-makers will then be able to use these fragility ratings and their associated fragility categories to support sound decisions regarding policy and interventions.

Outline

Chapter II, Literature Review, will give an overview and summarize previous, relevant research. Chapter III, Research Definitions, will discuss terminology, and what that terminology means in the context of this study. Various multivariate analysis methods will be discussed. There will also be discussion of how the data set was collected and organized, as well as highlight how the model will be verified. Chapter IV, Methodology, will outline the reasons for choosing the combination of techniques that the analysis procedure will be using. Chapter V, The Analytical Procedure, is an outline of the analysis process used in this study. Chapter VI, Results, discusses the outcomes, both in summary and in detail, as the analysis steps were executed. This chapter will also include the steps taken for model verification. Chapter VII, Recommendations and Conclusions, will recommend items of further research and bring this study to its conclusion.

II. Literature Review

Chapter Overview

Relevant literature will be discussed and summarized in this chapter. This will show what has been done, how this work will be different, and why it matters.

Relevant Research

The main issue in research on this topic, as highlighted by Atin Basuchoudhary, James T. Bang, Tinni Sen, and John David, is to separate works that explain causal relationships well, and those that predict future conflict well. (Atin, Bang, Sen, and David, 2018)

One work that attempts to do both is 2018's, *Predicting Hotspots: Using Machine Learning to Understand Civil Conflict* by Atin Basuchoudhary, James Bang, Tinni Sen, and John David. The authors compiled data from seven databases. The data collected was then used to build three different classes of variables: civil conflict variables, quality of political institutions variables, and economic development variables. Within each of these classes, there were 9 to 24 variables (Atin, *et al* 2018). Some of these variables were purges, demonstrations, democratic accountability, corruption, changes in GDP per capita, and drought index. Utilizing a machine learning process, the data was analyzed to compute their target variable, which was Lagconflict. Lagconflict, as defined in their book, was the likelihood of conflict in a certain area 1 to 5 years after the appearance of the leading factors (Atin, *et al*, 2018).

Rapid Urban Growth

The ease with which a national government can lose control of an urban area undergoing rapid growth is illustrated in “The Mega-Eights: Urban Leviathans and International Instability” (Liotta and Miskel, 2009). This article looks at conditions in eight very large cities that are experiencing rapid growth. All eight of these cities lie within what’s known as the “10/40 Window,” a region between 10 deg N and 40 deg S latitude. This region stretches from West Africa to Oriental Asia (Liotta and Miskel, 2009.) While highlighting the problems present in all these locations, the authors discussed three cities in particular detail: Lagos, Nigeria; Karachi, Pakistan; and Cairo, Egypt (Liotta and Miskel, 2009).

According to the Indian Census bureau, the city of Mumbai had grown from around six million in 1971 to nearly 12.5 million in 2011, a growth rate of 27.9%. It is forecast to be near 27 million by 2025, for a growth rate of 5.6% between 2011 and 2025. Along the way, the city is facing many of the same problems as other rapid urban growth areas in the less developed world. A 2013 study, titled “Spatiotemporal urbanization processes in the megacity of Mumbai, India: A Markov chains-cellular automata urban growth model,” (Moghadam and Helbich, 2013) highlights these points. The aim of Moghadam and Helbich’s study was to forecast trends in land use and population growth for the 2020-2030 time period, and to provide insight and data to government leaders managing the city’s future.

In the article, “The Urbanization of Everything: Governance Challenges in Southeast Asia,” Giok Ling Ooi takes an in-depth look at rapid growth cities in Southeast Asia (Ooi, 2007). He notes that this region is not only one of the most populated areas of

the world, but also one of the fastest growing. This rapid growth had led to vast slum areas in the inner-city districts, pushing new developments to these cities' periphery. He also highlights that the goals of attracting economic investment often conflict with building necessary infrastructure (Ooi, 2007).

Another article, "Assessing the Relative Contribution of Economic, Political, and Environmental Factors on Past Conflicts and the Displacement of People in East Africa" looks at refugees and other displaced populations (Owain and Maslin, 2018). Consideration is given to both what causes the displacement, and the effects of the presence of displaced people. Both internal and external conflicts, climactic changes, population level changes, and the rates of change of these variables are explored.

New York City, for example, shows why large First World urban areas are not having the issues that some growing Third World urban areas are having. First, First World urban areas have the fortune of growing at a much more modest rate. For example, from 1950 to 2015, New York City grew by 30%, meaning that the city's infrastructure only needed to grow that same 30%. Second, the respective national governments are already established, both internally and internationally, which allows these urban areas to grow with national and regional support (Liotta and Miskel, 2009).

Urban areas in the developing world have not been so fortunate. These include, but are not limited to, Dhaka (Bangladesh), Lagos (Nigeria), and Karachi (Pakistan). From 1950 to 2015, the rate of population growth in the listed cities has been astronomical. Dhaka grew by 5400%, Lagos grew by 2400%, and Karachi grew by 2000% (Liotta and Miskel, 2009).

Adding additional pressure, many of the countries containing these rapidly growing urban areas have only been independent since the end of colonialism in the 1950's and 1960's. Establishing a government while having to defend their borders from invading neighbors or having to handle influxes of refugees from unstable neighboring countries has often absorbed or monopolized resources needed to properly govern their own rapidly growing urban areas (Liotta and Miskel, 2009).

Youth Bulge

An additional major factor, and one that is drawing increased attention in academic circles, is Africa's "youth bulge." For this thesis, the youth cohort is defined as the part of the population aged 15-24. The reasons for choosing this age grouping are given in Chapter III.

Although discussion and study of the effect of youth cohorts in the developing world has been ongoing at least since the 1970's, Froneberg maintains rigorous examination of this topic has only occurred since 2000 (2019).

Sub-Saharan Africa has proportionately the largest youth population in the world (Sommers, 2011). Alongside this, youth unemployment is also the highest of any region of the world. Marc Sommers, in his article "Governance, Security, and Culture: Assessing Africa's Youth Bulge" (2011), highlights the need to connect with, and understand the life of these youths, instead of simply making assumptions, as governmental organizations, internal and external, have all too often done. In their article, "Governance and Youth Unemployment in Africa," Abe Ndjie, Atangana Ondoa,

and Ngoa Tabi, point out that corruption, in both the public and private sector, stands as one of the largest obstacles to resolving the issue of high youth unemployment (2019).

In addition to the size of the youth cohort, the level of youth unemployment in Africa is very high. In 2017, the average youth unemployment rate in North Africa, including the Maghreb region, was 11.7% with Tunisia's rate, the highest for the first half of 2017, at 15.4%. While Sub-Saharan Africa's rates have been lower, 7.3% in 2019, those countries are experiencing an upward trend (Ndjie, Ondoa, and Tabi, 2019). That same study also concluded that good governance indicators, such as control of corruption and political stability, can reduce youth unemployment.

Internal Migration

A phenomenon where rapid urban growth areas and youth bulges merge is internal migration. Internal migration refers to movement of a part of the population from one region to another region within a country. This could be for economic advantage in search of jobs. It may also be due to drought or flooding, conflict, corruption or a number of other factors. There are differing forms of migration, and their effects on stability differ from region to region and country to country. The primary concern for this study is migration from a predominantly rural area to a predominantly urban area such as a rapidly growing city. With both the youth bulge variables increasing and the rapid growth city variables increasing, the pressure such migration puts on existing infrastructure and services can be immense. A sustained strain on city services often brings about increasing instability.

Discussion of Analysis Methods

Principal Component Analysis is an analysis technique that is useful for reducing the number of variables in a large data set into a set of common, underlying factors. If done extracting orthogonal factors, it also eliminates multicollinearity in the data. The technique preserves as much of the total variance of the original data set as possible (Hair, Black, Babin, and Anderson, 2019). The weakness of this technique is that interpretability can become difficult. In this application, this makes it difficult to trace findings back to the original data elements to identify which of the original variables are driving the dependent value.

Linear Discriminant Analysis is a technique for separating a data set into distinct groupings. The separation technique maximizes the distance between the group means and minimizing the variation between groupings (Hair, *et al*, 2019). While one can see which data points fall into which grouping, it can be difficult to derive why a particular grouping has its particular mix of elements in it.

K Nearest Neighbors is a grouping technique. Similarity between groupings is defined as "the shortest distance from any object in one cluster to any object in the other." (Hair, *et al*, 2019) The logic of this technique maintains interpretability. However, the inherent weakness is that the parameter K must be chosen before processing begins. Since the goal is to discover which data elements are important and how many groupings exist, such a requirement could potentially make this technique less tractable.

K-Means Clustering is another technique that divides the data set into K groups, which is pre-determined. Each group has its own mean value, and each data point is placed into the group whose mean value is closest (Hair, *et al*, 2019). While the logic can maintain a sense of interpretability, the process is still dependent on the choice of a parameter, K, and this must be done before processing begins. Again, and for the same reasons as K Nearest Neighbors, such a requirement could potentially make this technique less tractable.

Cluster Analysis is also a grouping technique. It groups the data points so that each point is more like its own group that it is to any other groups (Hair, *et al*, 2019). This can have good trace-back capability. However, it only shows the most influential data item in each cluster without showing other component variables or to what extent those component variables might influence the grouping. Thus, it would be possible for a variable that is second or even third most influential in its grouping to be missed when evaluating the whole model.

Linear Regression involves the assumption of linear relationships between the independent variables and the dependent variable (Hair, *et al*, 2019). Because there are three variables each for urban population and youth bulge population, the risk of multicollinearity is much higher than typical with this dataset.

Factor Analysis, using the JMP software package, is a dimension reduction method that extracts the common variance and places them into factors. This offers a way to trace back to match up to the original variables (Hair, *et al*, 2019). Using the Orthomax rotation method can arrange the generated factors into an orthogonal configuration. This arrangement provides a valid way to minimize multicollinearity in the model, as well as assuring a valid way to trace back to the original independent variables.

Problem Statement

Past studies have used large groups of variables that could take time to assemble into the form necessary for the analysis to be performed. This study will be researching to see if a small group of readily available population variables and readily available software can be used to assess and classify the level of instability quickly and accurately. Such information, timely acquired, could be helpful for policymakers when considering interventions. It may also help with identifying aspects of rapid growth that could lead to a city or an urban area becoming a nexus for political instability and conflict.

III. Research Definitions

Chapter Overview

This chapter will define the research metrics that will be used in this study. This will include how population growth will be measured, what is meant by rapid growth, and what is the age band for a youth bulge. The processes for building the data set, stating the independent and dependent variables, determining the stability categories, and how the model will be verified will be presented. The research goal for this study will also be outlined.

Different Measures for Population Growth

This study measures population growth in two different ways. One is raw growth, and the other is percentage growth. Raw growth is the actual difference in the relevant population from the start of the measurement period and the conclusion of the period. Percentage growth is calculated by taking the raw growth and dividing it by the relevant population at the beginning of the measurement period. Both urban population data (United Nations, 2022) and youth bulge population data (United Nations, 2020) are by country and for the five-year period from 2010-2015. This population data is taken from the UN Population Division.

Defining Rapid Growth

To place the scale of growth in perspective, between 2010 and 2015, it was estimated that the urban population in Bangladesh grew by 3.13% annually to 27.2 million. In that period, in Nigeria, the urban population grew by 3.49 % to 48.2 million, and in Pakistan, the urban population grew by 3.00% to 45.6 million. In contrast, in that same period, the United States was estimated to have grown by 1.33% to 198.4 million (United Nations, 2022). In a 2009 paper, not only had New York City's population been holding reasonable steady, but it has also had in place both an effective local government providing infrastructure and other services, and a national government with sufficient resources and standing to back the local government up (Liotta and Miskel, 2009).

Urban population data is drawn from the UN Population Division, which lists the population by greater metropolitan area. This data file contains 150 countries with cities that are above 300,000 (United Nations, 2022). The highest rate of country urban growth from 2010 to 2015 is 10.83% for Oman. Out of 150 countries, 14 were growing at 5% or higher, and 63 were growing at 2% or higher. Currently, over half of the world's population live in urban areas (Urdal and Hoelscher, 2012). The forecast is that that proportion will increase to two-thirds living in urban areas (Urdal and Hoelscher, 2012). Approximately 60 percent of this urban growth is expected to be in Asia (Ooi, 2007).

Defining Youth Bulge

For the purpose of this analysis, the youth cohort is defined as the part of the population aged 15-24. This period of a person's life varies from culture to culture, and it can sometimes vary by socioeconomic status as well. Calculating from the UN World Population Prospects database, in 2020, the estimated average share of young adults was 15.4% in Asia and 19.3% in Africa. This was near to or higher than that the world average of 15.5% (United Nations, 2020). Coupled with youth unemployment rates in Africa around 10% (Ndjie, Ondoa, and Tabi, 2019), a country's economy can be stressed to its limits and possibly beyond while trying to grow fast enough to keep pace with the growth of its workforce.

Building the Data Set

The master data set for this study was drawn from three sources. The urban population data is drawn from the UN Population Division's World Urbanization Prospects (United Nations, 2022). The column "Urban Raw Growth" is calculated as being the difference between the 2010 urban population and the 2015 urban population. The column "Urban Population Average Annual Growth" is the percentage calculated by dividing the urban raw growth by the 2010 urban population. This source was chosen as it was the most comprehensive collection of current city-level population data available for the study time period, containing countries both developed and developing.

The youth bulge data is from the UN Population Division's Population Prospects 2019, which does contain specific data for the 15-24 age bracket (United Nations, 2020). The column "Youth Bulge Raw Growth" is calculated as being the difference between the 2010 and 2015 youth population. The column "Youth Bulge Average Annual Growth" being the percentage calculated by dividing the youth bulge raw growth by 2010 population.

The state fragility indexes were compiled by Monty G. Marshall and Gabrielle Elzinga-Marshall for the Center for Systemic Peace (Marshal & Elzinga-Marshal, accessed 2020). In their table, titled "State Fragility Index and Matrix 2018", ratings were given for 167 countries, all of which were over 500,000 in total country population. Countries were rated in four categories: security, political, economic, and social. For each of these four categories, there are two subcategories: effectiveness and legitimacy. The definition of these categories and subcategories are listed in Appendix A. For each of these eight items, a rating from 0 to 3 (4 for economic effectiveness) was assigned by the index developers. These ratings were then added together giving a total rating of 0 to 25. Lower scores indicate higher stability. Higher scores indicate higher fragility.

Variable Definitions

The independent variables for this study are the different aspects of the population data. For the urban population data, there is a variable each for the city's population, the raw growth of that population during the study period, and the percentage growth during the study period. Similarly, for the youth bulge population data, there are three variables: one for population, one for raw growth, and one for percentage growth.

As shown in Appendix C, there is evidence for multicollinearity. For this reason, the factor analysis method with orthogonal rotation was applied to the original six independent variables. This results in three factors with significantly reduced multicollinearity. The new variables will be Raw Population Group, Youth Raw Growth, and Annual Percent Growth. The details of this reduction process are described in Chapter IV.

The dependent variable will be the state fragility index. This is a categorical rating based on a model-calculated numerical rating with a range of 0 to 25. This rating will be grouped into three categories, ranging from A to C. A rating of A will be for highly stable countries (*e.g.*, Canada or Switzerland) to a rating of C, meaning countries near or at total collapse (*e.g.,* Iraq or Afghanistan). These categories are listed and explained in Table 1. The outputs of the model presented in this thesis will be a list of countries, their predicted rating, and the associated stability category; and these ratings will be for the end of the research period.

Table 1 - Fragility Rating Categories

Category	Rating Range	Meaning
A	0 – 8	Stable: This country is reasonably stable, and its government is effective. This country would need minimal assistance and surveillance
B	9 – 12	Borderline: This country would need more thorough research before determining a course of action
C	13 – 25	Fragile: This country is either on or over the edge. This country requires close observation and assistance with security as well as with their economic needs.

Delineating State Fragility Categories

Having a numerical rating is good. Knowing where a tipping point along the numerical scale makes that rating useful.

The break points in Table 1 were found using linear discriminant analysis (LDA). First, factor analysis was run, generating a set of three orthogonal factors. Second, regression analysis was run until only significant factors remained in the model. Then, using Excel, a threshold level for stability was entered, and the worksheet calculated a binary variable of either “S” for stable or “F” for fragile. These values were entered into the JMP data table. The LDA process was executed using the Annual Percent Growth factor (as the only remaining factor, see Chapter VI) and the binary variables as the independent variables.

The number of misclassifications shown in the generated confusion matrices (See Appendix B) were used to determine intervals of overlapping between stable and fragile. In examining the misclassified countries, percentage population growth was determined to be the influencing factor between stable and fragile. Urban growth of 1.9% or higher was typically misclassified stable-to-fragile, with Guatemala being an exception at

1.35%. Guatemala has been have having significant difficulties with crime and corruption. Urban growth of 2.5% or lower was typically misclassified fragile-to-stable, Eritrea being an exception at 3.69%. Eritrea had been in a border conflict with Ethiopia, which had been settled by 2018. There is still much tension between the two countries.

It was determined that fragile-to-stable misclassification was to be minimized, because that form of misclassification could lead to serious unanticipated consequences in the future.

It was decided to use an overlap interval from nine to twelve for delineating the fragility rating categories. For those countries in category B, *i.e.,* the overlap area, an annual growth rate of 3% or more generally indicated more fragile, while a rate less than 3% generally indicated more stability.

Model Verification

This verification process was used to check that the predicted output was within a reasonable range. Two comparison checks will be made.

The first comparison will be to see if the mean of the given ratings from the state fragility table is within the 95% confidence interval of the mean of the model predictions. If this is so, then the model should be in close agreement with the known circumstances worldwide, based on the State Fragility Index.

The second comparison will be on the city level. For each city, the given ratings for the country from the stability table will be compared with 95% prediction interval of the predicted values from the model. If this is true for a large proportion of the cities, then the model should be in close agreement with observed trends worldwide.

If the model passes both tests, the model will have been considered verified, and can be used by decision makers and other officials to formulate policy and take proper and timely action.

Research Goal

The research goal is two-fold. The first is to show that a small number of population variables processed using a statistical analysis procedure can yield reliable results for rating potential political instability. The second goal is to outline this process so that future analysts and decision makers can make use of this in their work.

The population data can be found at the United Nations Population Division, while the fragility ratings can be found at the Center for Systemic Peace. The software used in the analysis process is the JMP software package and Microsoft Excel (although other appropriate analysis software could be used). Both the data elements and the processing software are readily available.

IV. Methodology

Chapter Overview

In this chapter, the process for analyzing the dataset will be explained.

Selection and Summary

Referencing the discussion on various multivariate analyses provided in Chapter II, a two-step analysis procedure was chosen. First, factor analysis with Orthomax rotation is applied to the original dataset to extract a set of orthogonal factors that minimizes possible multicollinearity. Second, using these generated factors as the independent regressor variables, a standard least squares regression model is generated. Once the model is generated, testing for the assumptions of multiple linear regression can be made (Hair, *et al*, 2019). The results of this testing are shown in Chapter VI. Next, if any of the regressor variables is found to be insignificant (p-value > 0.05), the regressor variable with the highest p-value is eliminated from the model. The model is run again. These elimination steps are repeated until only significant (p-value <= 0.05) remain. This mathematical model is then verified using the process outlined in Chapter III. Using the output from this analysis, the number of countries in each stability category is then counted and reported. This analysis process provides the desired trace-back capability to the original data elements, while also providing sufficient descriptive statistics to evaluate the model's fit and its predictive capabilities.

V. The Analysis Process

Here are the steps used in this analysis:

I. Collect, calculate, and format the data set as needed.

II. Factor Analysis

  a. Using JMP, run Factor Analysis using the independent variables from the data set.

  b. Using JMP, apply the Orthomax rotation to generate the factors

  c. Using JMP, save the generated factors.

III. Generate a Standard Least Squares Regression Model Eliminate Insignificant Regressor Variables

  a. Eliminate these variables one at a time, regenerating the model each time, until no insignificant variables remaining

IV. Test the Assumptions for Linear Regression

V. Perform Tests for Goodness of Fit using Model Statistics

VI. Using Table 1 and the model-calculated ratings, assign an instability category.

VII. Use the assigned categories for further evaluation and analysis as desired.

VIII. Tests for Verification (Optional and only when one has values for the dependent variable to check against)

  a. Using Excel, perform a comparison test of the overall means.

  b. Using Excel, compare each country's dataset ratings with the model's prediction interval.

VI. Results

Chapter Overview

In this chapter, the results of the analysis are reported and interpreted. The initial factor analysis is discussed, followed by the regression process, then the tests for the assumptions of linear regression, then a discussion of model adequacy, and finally, the generated results from the regression model. There will also be reporting for and discussion of two verification tests, a prediction interval test of the instability rating for each country in the dataset and a confidence interval test of the mean of the instability ratings for the dataset.

The Factor Analysis Step

The first step is to apply the factor analysis method to the dataset. One requirement of factor analysis is to standardize and normalize the data set before any processing. The JMP software performs this step in the background. This was verified by first running the process with the data set unchanged, and then running with manually standardized and normalized data. The results were identical in both instances.

Table 2- Eigenvalues from the Initial Factor Analysis

Factor Analysis (Using JMP software)			
Eigenvalues			
Number	Eigenvalue	Percent	Cumulative Percent
1	3.332	55.525	55.525
2	1.576	26.269	81.794
3	0.621	10.355	92.149
4	0.405	6.744	98.892
5	0.053	0.884	99.776
6	0.013	0.224	100.000

Continuing with the JMP software package, factor analysis is run on the six independent variables. The goal at this stage was to generate a set of orthogonal factors where scaling and multicollinearity issues are minimized in the model. Table 2 shows the eigenvalues for each factor. The factors are shown from the highest eigenvalue to the lowest. Factors with higher eigenvalues account for a greater percentage of the model variance. Typically, only those factors with an eigenvalue greater than or equal to one would be considered for retaining them in the model as this explains more variation than they contribute. Of the six factors that were generated, only two had an eigenvalue greater than one. Factor number three had an eigenvalue of 0.621, and the remaining factors were even lower. Typically, this would indicate that the first two factors should be sufficient for continued processing, and these two factors would account for approximately 82% of the variability in the model.

However, when the factor rotation was performed on the two highest ranking factors using the Orthomax rotation process. Only five of the six variables loaded significantly onto the two generated factors. Therefore, the rotation was applied for the three highest ranking factors, and all six variables loaded onto the three generated factors. For completeness, the factor rotation was applied for the four highest factors. However, the six variables only loaded significantly onto the first three factors with no significant loading on the fourth factor. With these considerations, three factors will be used in this study, and these three factors account for 92% of the variability in the model.

Table 3- Rotated Factor Loadings

Rotated Factor Loadings (from JMP)			
	Raw Population Group	Youth Raw Growth	Annual Growth
Country Youth Bulge 2015	0.966	-------------	------------
Urban Population 2015	0.876	-0.419	------------
Urban Population Raw Growth 2010-15	0.873	-0.487	------------
Youth Bulge Raw Growth 2010-15	-0.363	0.921	------------
Urban Population Average Annual Growth 2010-15	-------------	-------------	0.786
Youth Bulge Average Annual Growth 2010-15	-------------	-------------	0.715

Table 3 shows how the Factor Loadings were used to group and name the three factors. Table 4 shows which of the six data elements were assigned to which of the three factors, based on the rotated factor loadings. This indicates which of the original variables were represented to their greatest extent in which factor.

As stated by Hair, *et al*., "Factor loadings are the level of correlation between the variable and the factor" (p. 123, 2019). For positive factor loadings, these indicate that changes in the independent variable result in a corresponding change in the value of that factor. For negative factor loadings, these indicate that changes in the independent variables result in a corresponding, yet opposite, change in the value of that factor (Hair, *et al*., 2019)

Table 4 – Data Element Factor Assignment following Orthomax Rotation Process

Factor Name	Factor	Data Elements
Raw Population Group	Factor 1	Youth Bulge 2015 Urban Population 2015 Urban Raw Growth 2010-15
Youth Raw Growth	Factor 2	Youth Bulge Raw Growth 2010-15
Annual Growth	Factor 3	Urban Pop Avg Annual Growth Youth Bulge Avg Annual Growth

Table 5 – Significance Test for Sufficient Factors

Chi-squared Significance Test for Sufficient Factors				
Test	Degrees of Freedom	Criterion	Chi-square statistic	P-value for Chi-square
H0: 3 factors are sufficient HA: more factors are needed	1	0.036	5.258	0.022

The factor analysis provides information on how well the newly generated factors capture the fit of the original model. This will be done using the chi-square test for sufficiency. For this test, the null hypothesis, H_o, is that three factors are sufficient, while the alternate hypothesis is that more factors are needed. Table 5 shows the information for this test. The resulting p-value is 0.021, which is less than the level of significance, $\alpha = 0.05$. This would generally mean that the null hypothesis is rejected, and that these three factors would not be sufficient. However, when the analysis was performed for four factors, the input variables loaded onto the same three factors, with

only minimal loading onto the fourth factor. For this process, only the three factors shown in Table 3 will be used.

Table 6 – Variance Explained by each Factor

Variance explained by each Factor (from JMP)			
Factor	Raw Population Group	Youth Raw Growth	Annual Growth
Variance	2.596	1.298	1.155
Percent	43.259	21.628	19.255
Cumulative Percent	43.259	64.886	84.141

Table 6 shows that 84.141% of the variance in the model is accounted for by the three generated factors. These three factors were saved, and these factors became the data elements for the regression process outlined next.

It is interesting to note that, when the rotation process was applied, the average annual growth-related data items loaded onto a single factor, the 2015 population-related items loaded onto a single factor, but the raw growth-related data items loaded onto separate factors. This suggests that the annual growth data and the raw population data may affect the models in different ways or to different degrees. This will assist with interpretability later in the process.

Initial Regression Analysis

At this time, a standard least squares process was executed, using the three factors from the factor analysis as the independent variables and the state fragility score as the dependent variable.

Testing for the significance of the parameter estimates involved using a t-test for each parameter estimate (henceforth β_j). For each j = {1,2,3}, with each j corresponding to an independent variable (see Table 8), the null hypothesis, $H_{0,}$ is that $\beta_j = 0$. The alternative hypothesis, H_a, is that $\beta_j \neq 0$. The *p*-values for all 3 β_j's are shown in Table 7. Only the Intercept and Annual Percent Growth factor are below the $\alpha = 0.05$ level of significance set for this analysis, which means that null hypothesis is rejected this parameter estimate. For the other two parameter estimates, the null hypothesis is not rejected, which means that one may be able to eliminate them from the model.

Table 7 - Regression Coefficient Estimates for Three Model Factors

Parameter Estimates						
J	Independent Variable	Estimate β_j	Std Error	*t*-ratio	*p*-value	VIF
---	Intercept	8.307	0.433	19.17	< .0001	
1	Raw Pop Group	0.199	0.437	0.46	0.649	1.002
2	Youth Raw Growth	0.250	0.438	0.57	0.569	1.004
3	Annual Growth	3.746	0.499	7.50	< .0001	1.006

Reduction to only Significant Factors

Using the largest p-value (0.649), the Raw Population Group variable was eliminated from the model. Again, testing for the significance of the parameter estimates involved using a t-test for each parameter estimate (henceforth β_j). For each j = {1,2}, with each j corresponding to an independent variable (see Table 8), the null hypothesis, H_0, is that $\beta_j = 0$. The alternative hypothesis, H_a, is that $\beta_j \neq 0$. The p-values for both β_j's are shown in Table 8. Only the Annual Growth factor is below the $\alpha = 0.05$ level of significance set for this analysis, which means that null hypothesis is rejected this parameter estimate. For the Youth Raw Growth factor, the null hypothesis is not rejected, which means that one may be able to eliminate this factor from the model.

Table 8 - Regression Coefficient Estimates for Two Model Factors

Parameter Estimates						
J	Independent Variable	Estimate β_j	Std Error	t-ratio	p-value	VIF
---	Intercept	8.307	0.432	19.22	< .0001	
1	Youth Raw Growth	0.248	0.437	0.57	0.571	1.004
2	Annual Growth	3.756	0.497	7.55	< .0001	1.004

Table 9 - Regression Coefficient Estimates for One Model Factor

Parameter Estimates						
J	Independent Variable	Estimate β_j	Std Error	t-ratio	p-value	VIF
---	Intercept	8.307	0.431	19.27	< .0001	
1	Annual Growth	3.773	0.495	7.62	< .0001	1

As the sole remaining insignificant factor, the Youth Raw Growth variable was eliminated from the model. Testing for the significance of the remaining parameter estimate involved using a *t*-test. For this parameter estimate (henceforth β), which corresponds to the Annual Percent Growth variable (see Table 9), the null hypothesis, H_0, is that $\beta_j = 0$. The alternative hypothesis, H_a, is that $\beta_j \neq 0$. The *p*-values for β are shown in Table 9. The Annual Growth factor is below the $\alpha = 0.05$ level of significance set for this analysis, which means that null hypothesis is rejected this parameter estimate.

The regression model is as follows:

$$State\ Fragility = 8.307 + 3.773 * Annual\ Growth \qquad \text{(1) Equation}$$

Table 10 - Variance Inflation Factors

Variance Inflation Factors (VIFs) from JMP	
Annual Growth	1

Multicollinearity

As stated in Chapters II and IV, multicollinearity is a probable issue in this data set. The data set has three city-related population variables and three youth bulge-related variables. The factor analysis step was intended to minimize this to the maximum extent possible.

To check that this has been occurred, the variable inflation factors (VIF) can be examined. Table 10 shows the VIF's for the Annual Growth factor. A VIF of one is ideal (Hair, *et al*, 2019). If multicollinearity is still present, at least one variable's VIF will be substantially high. In this model, the only variable has a VIF of one, indicating that there is no or minimal multicollinearity in the model.

Testing the Assumptions of Simple Linear Regression

Here are the four basic assumptions of Linear Regression (Hair, *et al*, 2019).

1) A linear relationship exists between each independent variable and the dependent variable
2) The residuals have constant variance for every level of the independent variables
3) The residuals have a normal distribution
4) The residuals are independent

The results from performing these four tests are displayed here.

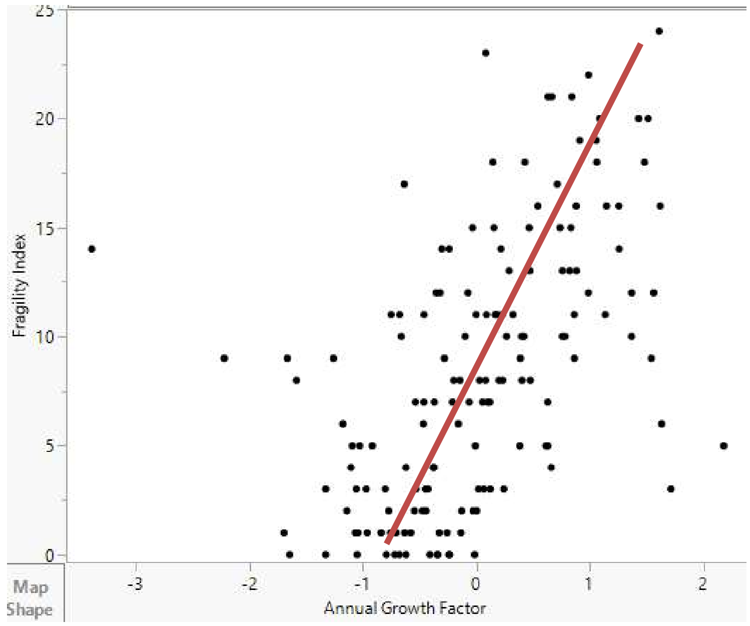


Figure 1 – Scatterplot of the Independent Variable vs. Actual Fragility Scores

The first test is for a linear relationship. Figure 1 shows a scatterplot for the Annual Growth Factor, the independent variable, and the actual value for the Fragility Score, the dependent variable. Ideally, one should be able to draw a rough line thru the data points (Hair, *et al*, 2019). While there appears to be a few outliers, one could fit a line to the data points. Another consideration is scale. If the horizontal and vertical axes were in proper scale to each other, the plots would be less vertically oriented.

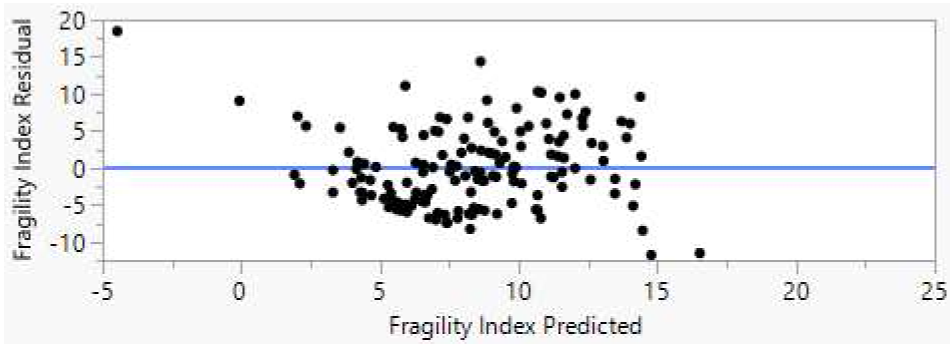


Figure 2 – Residuals by Predicted Values

The second test is the test for constant variance of the residuals, which can be done with a predicted vs. residual plot. Ideally, there is an even spread across the length of the plot and around the zero line (Hair, *et al*, 2019). There is no clear funnel, arrow, or diamond shape to the plot in Figure 2, suggesting that there is no material impact of heteroscedasticity in the model.

The third test is for the test for the residuals being normally distributed. This is done by examination of the normal quantile plot of the residuals. Ideally, the data points should lie along a 45° degree and within the 95% confidence interval limits (Hair, et al, 2019). These two conditions are met as shown in Figure 3.

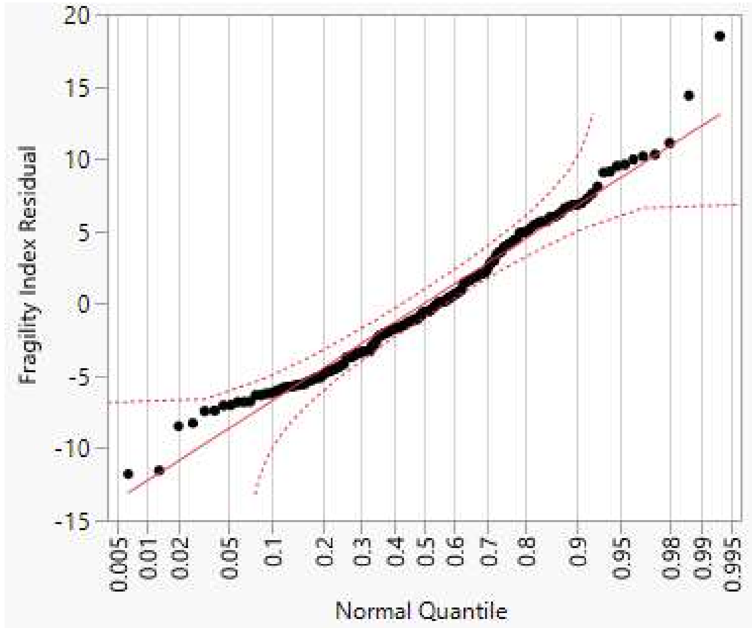


Figure 3 – Normal Quantile Plot of the Residuals

The fourth and final test is for independence of the residuals (Hair, *et al*, 2019). One can do this using the externally studentized residual plot, Figure 4, shows that most of the plot points are well scattered and within the 95% individual confidence interval limits (green lines). Most of the remaining points are within the 95% simultaneous confidence interval limits (red lines). This shows that the independence assumption is met.

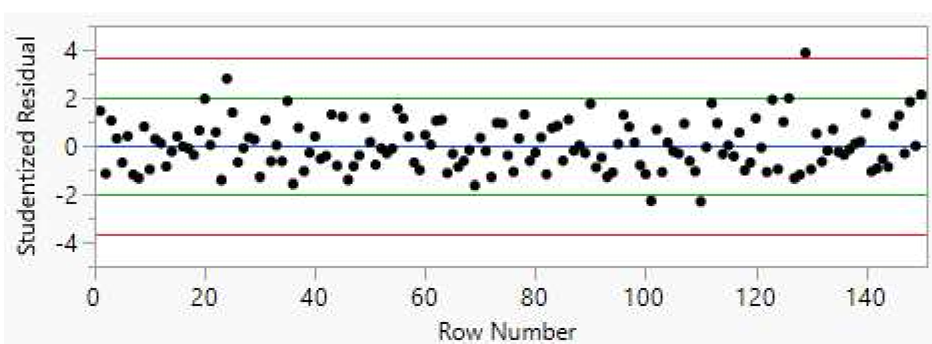


Figure 4- Externally Studentized Residuals with 95% confidence limits

All four assumptions for multivariate linear regression are being met. A linear regression model is deemed a valid approach for further processing.

Testing for Significance of Regression

With insignificant factors eliminated from the model, and after testing for the assumptions of linear regression, the process continues with testing for the significance of regression

A significance of regression test is conducted, using an *F*-test. The null hypothesis (H_0) is the regression equation is not significant. The alternative hypothesis (H_a) is that the regression equation is significant. The information for the *F*-test is shown in Table 11. The *p*-value is well below the $\alpha = 0.05$ level of significance, meaning that the null hypothesis is rejected. This suggests that the overall model a good fit for the data.

.

Table 11– F-test for Significance of Regression

F-test for Significance of Regression				
Source	Degrees of Freedom	Sum of Squares	Mean Square	*F*-Test statistic
Model	1	1617.959	1617.96	58.037
Error	148	4125.935	27.88	*P*-Value for F
C. Total	149	5743.893		< .0001

Taken together with tests for significance of the parameter estimates, these statistical tests show that this regression model is a good fit for the data, and that the predictions based on this model will be generally reliable.

Summary of Fit Statistics

From Table 12, the coefficient of determination, R^2, is quite low, from an engineering point of view, at 0.282, and the adjusted R^2 was 0.277. Low values would indicate that a large portion of the variance is explained by other variables other than the independent variables used in this model. However, it is not uncommon for R^2 values to be less than 0.50 when human behavior is involved (*Regression analysis: How do I interpret R-squared and assess the goodness-of-fit?* 2013). In this case, the human behavior in question is population growth, demographic changes, and politics.

Table 12 – Model Summary of Fit Statistics

Summary of Fit	
R^2	0.282
Adjusted R^2	0.277
RMSE – Standard Deviaton	5.280
Mean of the Respone Variable	8.307
Number of Observations	150

In a mathematical model, residuals are the difference between the actual value and the predicted value. These are often used to test the quality of the model's predictions. The normal quantile plot, Figure 3, shows the plot points in a rough 45° line and within the 95% confidence interval limits. This is the desired pattern for this graph.

In addition, the externally studentized residual plot, Figure 4, shows that most of the plot points are well scattered and within the individual confidence interval limits (green lines). Most of the remaining points are within the simultaneous confidence interval limits (red lines). Taken together, these plots indicate that the values of the residuals are keeping within the bounds of the assumptions of regression.

The difficulty shows itself when examining the mean of the fragility index value and the sample standard deviation. The Summary of Fit table, Table 12, shows a mean of 8.307 (on a range of 0 to 25), while the standard deviation is 5.280. With a standard deviation value that is this large relative to the related mean, this may indicate that there is considerable variability in the model.

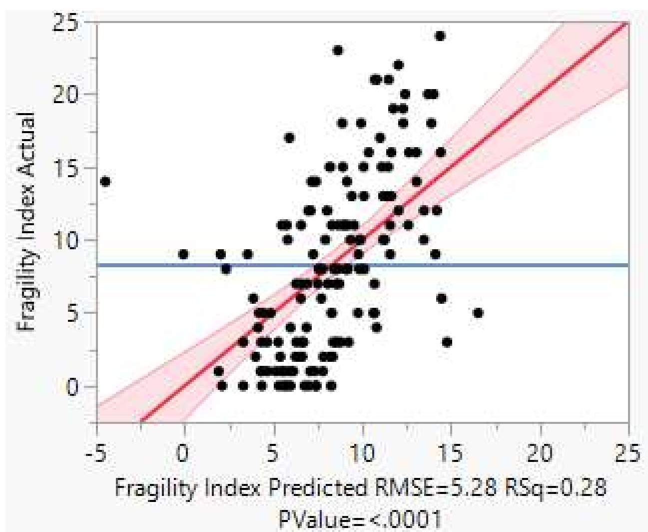


Figure 5- Actual by Predicted Plot for Dependent Variable

This is confirmed by the actual by predicted plot, Figure 5, where there is considerable width horizontally and height vertically in the graph. On such a plot, the desired look is a roughly 45° line, and within 95% confidence limits (Montgomery, Peck, and Vining; 2012). Although, there is a general upward and rightward trend in the plot, it is clearly not the desired 45°. The fact that one could still draw a trend line, and the R^2 is still within a reasonable range means that there is a rough linear fit to the data.

Oily Rag Theory

One explanation for the sizable variability shown in Figure 5 is what is called "Oily Rag Theory." This was first put forth by Sean O'Brien in his article, "Anticipating the Good, the Bad, and the Ugly: An Early Warning Approach to Conflict and Instability Analysis" (p. 791-811, 2002). In that article, O'Brien made the analogy of national instability to an oily rag in a garage. He described a situation where two garages were very similar in all material ways, including an oily rag left on a workbench in each case. In one case, the oily rag caught fire, and the garage burned down. In the other, nothing occurred, and the garage was intact and undamaged. Why did one garage catch fire, while the other did not? That could prove very difficult to answer, and there might not be enough evidence remaining to form an explanation. The same uncertainty exists when attempting to predict if a nation will destabilize. (O'Brien, 2002)

Table 13 - City Count by Fragility Category

Category	Range	Model Prediction	From the Data Set
A	0 – 8	80 countries	81 countries
B	9 – 12	54 countries	32 countries
C	13 – 25	16 countries	37 countries

Country Count by Categories

The model output from JMP was downloaded into an Excel spreadsheet. Excel was then used to count the number of countries in each category. The same procedure was applied to the ratings in the data set. The results are given in Table 13. Comparing the predicted ratings from the model to the assigned ratings in the data set, the model tends to overpredict membership in category B, but it underpredicts membership in category C.

Investigating Differences of Categorization

Using Excel, a confusion matrix was generated for classifying and investigating the instances that the data and model do not agree. This confusion matrix is shown in Table 14. Some patterns do emerge from this investigation.

A to B misclassification tends to be driven by an urban growth rate of 2% per year or higher. A to C misclassification is also driven by high city growth rates for both youth and urban populations. This growth rate is generally 5% per year or higher.

Table 14 – Confusion Matrix for Fragility Categories

Confusion Matrix for Fragility Categories			
Data Set \ Model	A	B	C
A	62	14	3
B	13	14	5
C	5	24	8

B to A misclassification is driven by urban growth rates of less than 2.5%. In fact, in many of these countries, the youth population is actually shrinking. B to C misclassification tends to happen when urban growth is 4.5% per year or more, and the youth population is increasing at a 3% or higher annual rate.

C to A misclassification tends to happen when the youth population is shrinking. C to B misclassification may happen when the urban growth rate is 2.5% per year or higher and the youth bulge growth rate is 2% per year or higher. The rate of growth for its urban population and for its youth population is important consideration for a country. Managing that growth properly is a key ingredient for a country's stability.

Verification Step

This step was performed to check if the model output was reasonable for the intended application.

In this step, two verification comparisons will be made. The first will be the mean of the fragility indices from the data set with the mean of the model-generated

indices. The second comparison will be a city-by-city comparison between that given by the dataset versus the model predictions.

These numbers were derived using JMP and Excel. From the given data, the mean of the fragility indices is 8.307 and the corresponding sample standard deviation is 6.209. From the model predictions, the mean is also 8.307, and the corresponding sample standard deviation is 5.280. The given mean from the data set and the predicted mean from the model are nearly identical, and the standard deviation from the model is lower than standard deviation from the data set. This indicates that the model should be predicting correctly.

Next, JMP was used to obtain the upper and lower bounds for 95% prediction interval values. Excel was used to calculate whether each city's fragility index from the data set lay within the corresponding prediction interval. Out of 150 countries, five of them fell outside their respective prediction interval bounds. This is a 96.67% prediction rate. Since the prediction rate is so high, the model appears to be predicting correctly.

Since the model passed both proposed verification checks, predictions made using this model should be reliable enough to base policy-making decisions and operational planning decisions upon with reasonable effectiveness.

Investigating Predictions outside the Intervals

In the verification section, five countries were mentioned. These were countries whose data-assigned fragility rating was outside the 95% prediction interval for the model-calculated fragility.

Two of these countries, Qatar and Oman had data-assigned ratings that were lower than their 95% prediction intervals from the model. Both countries showed very high levels of urban growth, and Qatar also shows very high levels of youth bulge growth.

The other three countries, Zimbabwe, Syria, and the Central African Republic had data-assigned rankings that were higher than their 95% prediction intervals. These countries had either modest growth or a decline in the youth and/or urban demographics. In the case of Syria, there is still an ongoing civil war, which might not be accounted for in a population-based model.

Effects of Migration

Internal migration may affect the fragility score, and therefore the instability category assigned to a country. Migration from one rural area to another rural area within a country will have little effect on the countrywide values for the youth and urban populations. Similarly, migration from one urban area to another urban area within a country will have little effect on the countrywide values for youth and urban population. In both cases, the effect of internal migration on the fragility score would likely to be low, and probably unpredictable.

Migration from a rural area to an urban area within a country, which is generally more common, would result in increases in the urban population values for that country. Migration from one country to another, whether for economic or security reasons, will affect the countrywide population values of both countries. If urban areas find themselves on or near a national border, the impact could be high. If a country is experiencing such migration, and it receives a rating that places it near the boundary between instability categories, thesesmigration factors need to be considered before assigning a category

Summary

Both country-level urban population growth and country-level youth population growth are concerning to political leaders, policy planners, and decision makers. Such statistics, handled properly, can be reliable predictors. There is a general trend that higher values for these parameters could suggest increased political instability in the coming years. However, the 95% prediction intervals for state fragility index can be, on average, around 20-21 points wide. The state fragility scores have a range of 0 to 25. These are very wide intervals in relation to the possible range of values. Therefore, while these types of population data parameters are readily available or readily calculable, and they can function as reasonable predictors of instability, one should look at other indicators before a final instability category is assigned.

VII. Recommendations and Conclusion

Recommendations for Future Research

The results show that a country with a rapidly growing urban population and a sizable and growing youth bulge is a potential candidate for future instability. However, the coefficient of determination is only approximately 0.282, and the root mean square error (RMSE) (5.280) is high compared to the mean (8.307). This would indicate that other items may also be important, or that intervening variables may be driving the results. Intrastate and foreign migration are possible variables. Other variables worthy of investigation would fall under conflict status, corruption in government, and economic status, among others.

Conflict status would entail considering whether internal instability is already present, or an external war is already ongoing. This may also be referring to a conflict that has recently ended. Political stability coupled with economic stability has already been cited as a factor that lowers youth unemployment.

Corruption in government can, and often does, exacerbate youth unemployment. In fact, it can often inhibit economic growth in general. Whether by raising the cost of doing business for a company in a particular region or country or by diverting funds away from developing businesses or from the general population, the result is the same. It means either a lower profit for the company or lower wages for workers. Often, this leads to both. In addition, the uncertainty fostered by corruption is an added business risk

Economic factors include the degree of diversification of a country's economy, level of unemployment, and relative pace of economic growth. Countries with high levels of corruption and have an abundance of natural resources often offer little or no reduction in their youth unemployment level (Ndije, *et al*, 2019). The economies of such countries are also sensitive to shocks in world trade of such commodities. Youth unemployment is widespread in Africa and Asia, and this could serve as a recruiting group for a rebellion and other forms of violence. The pace of economic growth relative to the growth of a youth bulge would indicate how readily an economy can absorb new persons into the work force. Even in a growing economy, an excess of unemployed can surpress wages

Lastly, it would likely pay great dividends to gain information about the youth cohort's standing in a particular country by interviewing or surveying the youth themselves. It can be easy to theorize about the obstacles that youth face from an academic setting. However, the youth cohort's members have first-hand knowledge about their status and grievances, and one will always be better hearing about that from them. The literature about youth bulges lacks first-hand information about youth in general (Sommers, 2011). Much has been theorized about African youth, but interviews with the youth themselves are conspicuously lacking (Sommers, 2011).

Conclusion

This analysis suggests that rate of growth is more important than whether or not the country's urban population is already significantly large. A common reference point for whether a country that sits on the boundary lines of fragility scale is a growth rate of 2%. Where the urban population is growing faster than that, it generally follows that the country is more fragile that it might seem. Similarly, where urban populations are growing at a more modest pace, those countries are generally more stable.

Potential youth bulges should be straight-forward to forecast with a simple inspection of birth records. However, in some countries, record-keeping may not be as rigorous as in others. Coupled with inadequate death records, this can also make knowing who survived into adulthood can be particularly difficult. The circumstances within a country can make ascertaining the size of the youth population beyond estimates tricky.

One of the goals of the analysis was to ascertain whether a large and growing youth bulge in a country with a large and rapidly growing city could lead to political instability. What has been shown is that rapid urban growth and rapidly growing youth bulges can lead to instability a few years into the future. However, the level of correlation is such that, while these are reasonable indicators, they are individual indicators on a larger list of several such indicators.

Significant indicators beyond the population data include strong adherence to traditional and cultural values, or the willingness to overturn and replace such values. Additionally, while established democracies and established authoritarian governments tend to be stable, instability often tends to accompany changes in style and form of government (Marshall, 2001). Migration, whether internal or external, to an urban area or away from an urban area, should be included in future formulation of predictive of mathematical models. The Oily Rag Theory may also be a factor in why one country is stable while another is falling into chaos (O'Brien, 2002).

The other goal was to provide an approach that, using a few readily available or readily calculable items of population data, can provide a quick measure of instability for a city, country, or region. An analyst or decision-maker can use this approach in some circumstances as a situational update when it has been some time since the last analysis had been done. It might also be useful when it is known that key factors have changed, and policymakers want to know the effects of these changes.

While this approach can provide an estimate of instability, in the long-term, it would be to a researcher's or policy maker's advantage to consider the wider list of indicators rather than focusing on just the population factors presented here.

Appendix A

Fragility Index Items (Marshall and Elzinga-Marshall, 2018)

Security effectiveness – a measure of general security and vulnerability to political violence, 1994-2018 (25 years), includes periods of direct involvement in wars, periods in between wars, and periods of peace, where there has not been a war since the most recent one ended

Security legitimacy – a measure of state regression and political repression, 2003-2018

Political effectiveness – a measure of regime/governance stability, including regime durability, current leader's years in office, and number of coup events 2003-2018

Political legitimacy – a measure of regime/governance inclusion, 2018, including factionalism, ethnic group political discrimination, political salience of elite ethnicity, and exclusionary ideology of ruling elite.

Economic effectiveness – a rating on a five point scale (0 to 4) measuring GDP per capita, 2010-2018.

Economic legitimacy – a measure of a country's share of export trade in manufactured goods, 2003-2018, including both manufactured goods and primary commodities.

Social effectiveness – a measure of human capital development, 2018, based on the UNDP Human Development Index.

Social legitimacy – a measure of human capital care, 2018, including infant mortality rates.

Appendix B

Listed here are the confusion matrices for various threshold fragility levels. For this part of the analysis, the Fragility Index from the data set was used. If the fragility index was less than or equal to this threshold value, the country was set to “S” for “Stable.” If the fragility index was higher than this threshold value the, the country was set to “F” for “Fragile.” These “S” or “F” values, along with the three vectors from the factor analysis, were input into a Linear Discriminant Analysis process. That process generated confusion matrices with the input values and predicted values.

Threshold 6
Misclassify 39

Matrix	F	S
F	63	24
S	15	48

Threshold 10
Misclassify 38

Matrix	F	S
F	40	14
S	24	72

Threshold 7
Misclassify 38

Matrix	F	S
F	58	20
S	18	54

Threshold 11
Misclassify 38

Matrix	F	S
F	32	11
S	27	80

Threshold 8
Misclassify 37

Matrix	F	S
F	51	18
S	19	62

Threshold 12
Misclassify 36

Matrix	F	S
F	29	8
S	28	85

Threshold 9
Misclassify 34

Matrix	F	S
F	48	14
S	20	68

Threshold 13
Misclassify 42

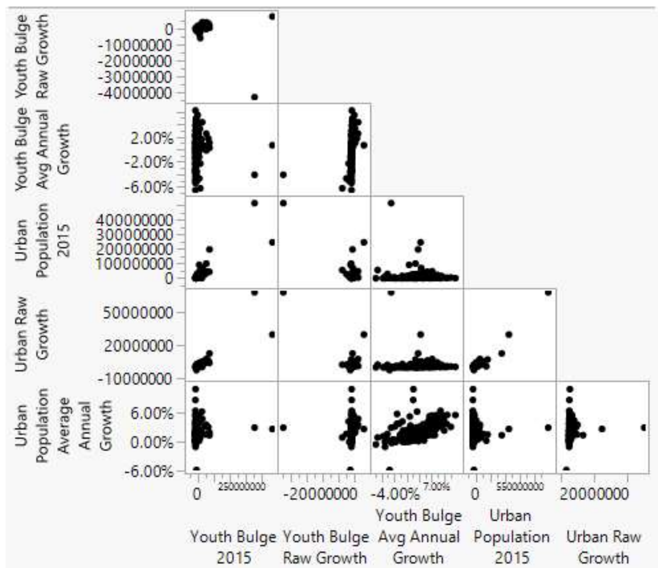
Matrix	F	S
F	23	9
S	33	85

Appendix C

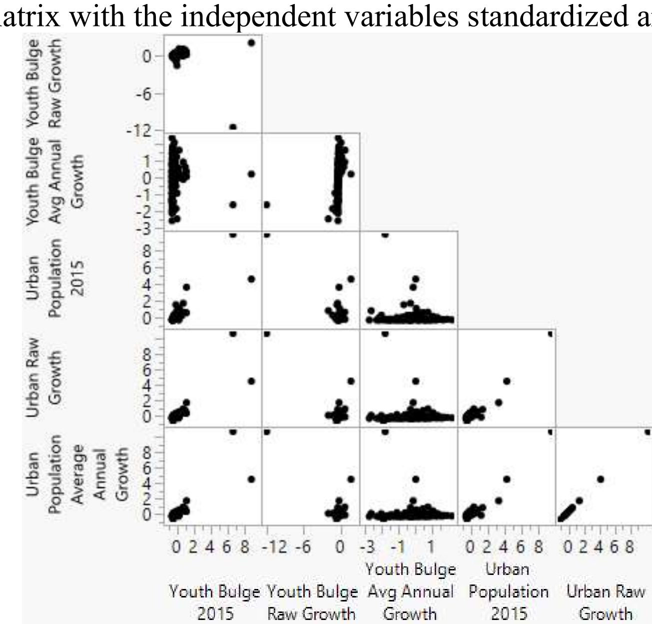
Correllation Matrices

1) Ideally, the individual charts should be a 45° line from lower left to upper right. Many of the individual charts are showing either a vertical pattern or a horizontal pattern. This means that there is evidence for possible multicollinearity.

2) This is the matrix with the independent variables not standardized.



3) This is the matrix with the independent variables standardized and normalized.



Appendix D

Table Notes

1) Population estimates are from the United Nations Population Division whose data was in thousands.
2) Growth rates are calculated from the population data, and they are an average annual percentage rate from 2010-2015.
3) CF 2018: Country Fragility (Score) from the Center for Systemic Peace for the year 2018

Country	CF 2018	Youth Bulge 2010	Youth Bulge 2015	Youth Bulge Raw Growth	Youth Bulge Avg Annual Growth	Urban Population 2010	Urban Population 2015	Urban Raw Growth	Urban Pop. Avg Annual Growth
Dem. Rep. of Congo	24	12,241,000	14,401,000	2,160,000	3.30%	18,694,000	23,187,000	4,493,000	4.40%
Central African Rep.	23	889,000	930,000	41,000	0.91%	717,000	798,000	81,000	2.16%
South Sudan	22	1,892,000	2,169,000	277,000	2.77%	255,000	321,000	66,000	4.71%
Yemen	21	5,311,000	5,844,000	533,000	1.93%	4,640,000	5,693,000	1,053,000	4.18%
Burundi	21	20,553,000	20,650,000	97,000	0.09%	556,000	751,000	195,000	6.20%
Sudan (North)	21	6,672,000	7,810,000	1,138,000	3.20%	6,740,000	7,696,000	956,000	2.69%
Somalia	20	2,276,000	2,821,000	545,000	4.39%	2,970,000	3,854,000	884,000	5.35%
Ethiopia	20	17,492,000	21,630,000	4,138,000	4.34%	4,162,000	5,240,000	1,078,000	4.71%
Afghanistan	20	5,570,000	7,278,000	1,708,000	5.49%	4,315,000	5,016,000	701,000	3.06%
Niger	19	2,978,000	3,765,000	787,000	4.80%	1,236,000	1,481,000	245,000	3.68%

Chad	19	2,355,000	2,852,000	497,000	3.90%	991,000	1,187,000	196,000	3.68%
Nigeria	18	30,300,000	34,305,000	4,005,000	2.51%	40,618,000	48,212,000	7,594,000	3.49%
Iraq	18	5,930,000	7,203,000	1,273,000	3.97%	14,058,000	16,401,000	2,343,000	3.13%
Guinea	18	2,107,000	2,414,000	307,000	2.76%	1,518,000	1,714,000	196,000	2.46%
Myanmar	18	9,161,000	9,574,000	413,000	0.89%	5,723,000	6,495,000	772,000	2.56%
Guinea-Bissau	17	317,000	351,000	34,000	2.06%	409,000	497,000	88,000	3.97%
Zimbabwe	17	2,873,000	2,828,000	(45,000)	-0.32%	2,482,000	2,514,000	32,000	0.26%
Angola	16	4,600,000	5,384,000	784,000	3.20%	8,460,000	11,036,000	2,576,000	5.46%
Cameroon	16	4,195,000	4,646,000	451,000	2.06%	6,091,000	7,859,000	1,768,000	5.23%
Burkina Faso	16	3,112,000	3,610,000	498,000	3.01%	2,329,000	2,956,000	627,000	4.88%
Mauritania	16	703,000	787,000	84,000	2.28%	851,000	1,058,000	207,000	4.45%
Côte d'Ivoire	16	4,106,000	4,771,000	665,000	3.05%	4,470,000	5,076,000	606,000	2.58%
Pakistan	15	37,234,000	40,597,000	3,363,000	1.74%	39,316,000	45,588,000	6,272,000	3.00%
Mali	15	2,890,000	3,334,000	444,000	2.90%	2,126,000	2,524,000	398,000	3.49%
Rwanda	15	2,133,000	2,249,000	116,000	1.06%	797,000	951,000	154,000	3.60%
Gambia	15	374,000	432,000	58,000	2.93%	393,000	420,000	27,000	1.34%

Eritrea	15	712,000	618,000	(94,000)	-2.79%	670,000	803,000	133,000	3.69%
Uganda	14	6,580,000	7,837,000	1,257,000	3.56%	2,016,000	2,577,000	561,000	5.03%
Philippines	14	18,339,000	19,864,000	1,525,000	1.61%	25,232,000	27,910,000	2,678,000	2.04%
Venezuela	14	5,318,000	5,315,000	(3,000)	-0.01%	13,207,000	14,226,000	1,019,000	1.50%
Libya	14	1,235,000	1,106,000	(129,000)	-2.18%	2,287,000	2,563,000	276,000	2.30%
Syria	14	4,391,000	3,521,000	(870,000)	-4.32%	10,069,000	7,451,000	(2,618,000)	-5.84%
Congo-Brazzaville	13	805,000	892,000	87,000	2.07%	2,392,000	2,962,000	570,000	4.37%
Liberia	13	734,000	856,000	122,000	3.12%	1,056,000	1,270,000	214,000	3.76%
Malawi	13	2,997,000	3,453,000	456,000	2.87%	1,432,000	1,713,000	281,000	3.65%
Sierra Leone	13	1,294,000	1,451,000	157,000	2.32%	905,000	1,043,000	138,000	2.88%
Togo	13	1,257,000	1,404,000	147,000	2.24%	1,466,000	1,635,000	169,000	2.21%
Equatorial Guinea	12	174,000	210,000	36,000	3.83%	230,000	313,000	83,000	6.36%
Madagascar	12	4,161,000	4,953,000	792,000	3.55%	2,482,000	3,223,000	741,000	5.36%
Zambia	12	2,751,000	3,236,000	485,000	3.30%	2,668,000	3,263,000	595,000	4.11%
Djibouti	12	175,000	178,000	3,000	0.34%	485,000	531,000	46,000	1.83%
Laos	12	1,412,000	1,410,000	(2,000)	-0.03%	606,000	642,000	36,000	1.16%

Kyrgyzstan	12	1,212,000	1,095,000	(117,000)	-2.01%	844,000	936,000	92,000	2.09%
Mozambique	11	4,559,000	5,382,000	823,000	3.37%	3,652,000	4,546,000	894,000	4.48%
Ghana	11	5,009,000	5,499,000	490,000	1.88%	4,968,000	6,111,000	1,143,000	4.23%
Bangladesh	11	29,514,000	30,351,000	837,000	0.56%	23,341,000	27,235,000	3,894,000	3.13%
Colombia	11	8,379,000	8,563,000	184,000	0.44%	23,162,000	25,974,000	2,812,000	2.32%
Haiti	11	2,067,000	2,139,000	72,000	0.69%	2,141,000	2,439,000	298,000	2.64%
Bolivia	11	1,942,000	2,088,000	146,000	1.46%	4,237,000	4,696,000	459,000	2.08%
India	11	234,647,000	242,350,000	7,703,000	0.65%	217,305,000	247,298,000	29,993,000	2.62%
Tajikistan	11	1,696,000	1,692,000	(4,000)	-0.05%	721,000	812,000	91,000	2.41%
Uzbekistan	11	6,289,000	5,984,000	(305,000)	-0.99%	3,798,000	4,094,000	296,000	1.51%
Sri Lanka	11	3,220,000	3,126,000	(94,000)	-0.59%	574,000	581,000	7,000	0.24%
Algeria	11	7,469,000	6,615,000	(854,000)	-2.40%	5,239,000	5,619,000	380,000	1.41%
Tanzania	10	8,825,000	10,003,000	1,178,000	2.54%	6,145,000	8,000,000	1,855,000	5.42%
Benin	10	1,813,000	2,099,000	286,000	2.97%	1,682,000	2,011,000	329,000	3.64%
Kenya	10	8,649,000	9,707,000	1,058,000	2.33%	5,241,000	6,287,000	1,046,000	3.71%
Senegal	10	2,593,000	2,885,000	292,000	2.16%	2,958,000	3,378,000	420,000	2.69%

Cambodia	10	2,989,000	3,198,000	209,000	1.36%	1,523,000	1,779,000	256,000	3.16%
Papua New Guinea	10	1,390,000	1,573,000	183,000	2.50%	312,000	345,000	33,000	2.03%
Egypt	10	16,589,000	16,603,000	14,000	0.02%	25,697,000	28,442,000	2,745,000	2.05%
Azerbaijan	10	1,840,000	1,613,000	(227,000)	-2.60%	2,690,000	2,867,000	177,000	1.28%
Bahrain	9	186,000	181,000	(5,000)	-0.54%	292,000	441,000	149,000	8.60%
Nepal	9	5,269,000	5,934,000	665,000	2.41%	1,216,000	1,507,000	291,000	4.38%
Turkey	9	12,539,000	13,008,000	469,000	0.74%	34,625,000	38,894,000	4,269,000	2.35%
Kazakhstan	9	3,127,000	2,619,000	(508,000)	-3.48%	4,815,000	5,605,000	790,000	3.09%
Iran	9	15,857,000	12,493,000	(3,364,000)	-4.66%	27,722,000	30,005,000	2,283,000	1.60%
Ukraine	9	6,362,000	4,905,000	(1,457,000)	-5.07%	12,240,000	12,202,000	(38,000)	-0.06%
Russia	9	21,392,000	15,505,000	(5,887,000)	-6.23%	54,291,000	56,704,000	2,413,000	0.87%
Gabon	8	325,000	366,000	41,000	2.40%	649,000	747,000	98,000	2.85%
Saudi Arabia	8	5,076,000	4,832,000	(244,000)	-0.98%	16,814,000	19,805,000	2,991,000	3.33%
Paraguay	8	1,303,000	1,350,000	47,000	0.71%	2,851,000	3,262,000	411,000	2.73%
Thailand	8	9,620,000	9,649,000	29,000	0.06%	21,662,000	24,498,000	2,836,000	2.49%
Guatemala	8	3,070,000	3,465,000	395,000	2.45%	2,560,000	2,738,000	178,000	1.35%

South Africa	8	10,449,000	10,014,000	(435,000)	-0.85%	21,171,000	23,984,000	2,813,000	2.53%
Indonesia	8	41,782,000	44,022,000	2,240,000	1.05%	40,345,000	44,368,000	4,023,000	1.92%
Turkmenistan	8	1,103,000	1,021,000	(82,000)	-1.53%	668,000	753,000	85,000	2.42%
Moldova	8	745,000	575,000	(170,000)	-5.05%	555,000	526,000	(29,000)	-1.07%
Honduras	7	1,726,000	1,933,000	207,000	2.29%	1,689,000	2,000,000	311,000	3.44%
Israel	7	1,107,000	1,182,000	75,000	1.32%	5,693,000	6,259,000	566,000	1.91%
Vietnam	7	17,543,000	15,640,000	(1,903,000)	-2.27%	13,596,000	16,677,000	3,081,000	4.17%
Ecuador	7	2,836,000	2,991,000	155,000	1.07%	4,684,000	5,153,000	469,000	1.93%
Mongolia	7	591,000	504,000	(87,000)	-3.13%	1,138,000	1,365,000	227,000	3.70%
China	7	229,260,000	186,270,000	(42,990,000)	-4.07%	448,571,000	517,241,000	68,670,000	2.89%
Morocco	7	6,179,000	6,035,000	(144,000)	-0.47%	10,674,000	11,454,000	780,000	1.42%
Nicaragua	7	1,223,000	1,210,000	(13,000)	-0.21%	992,000	1,027,000	35,000	0.70%
North Korea	7	3,912,000	3,958,000	46,000	0.23%	4,741,000	4,844,000	103,000	0.43%
Jordan	6	1,409,000	1,791,000	382,000	4.91%	2,982,000	3,900,000	918,000	5.51%
Peru	6	5,555,000	5,446,000	(109,000)	-0.40%	13,077,000	14,357,000	1,280,000	1.89%
Brazil	6	34,615,000	34,326,000	(289,000)	-0.17%	94,219,000	101,345,000	7,126,000	1.47%

Armenia	6	523,000	420,000	(103,000)	-4.29%	1,066,000	1,071,000	5,000	0.09%
Oman	5	642,000	626,000	(16,000)	-0.50%	893,000	1,493,000	600,000	10.83%
Lebanon	5	966,000	1,200,000	234,000	4.43%	1,990,000	2,229,000	239,000	2.29%
United Arab Emirates	5	1,158,000	988,000	(170,000)	-3.13%	4,314,000	5,698,000	1,384,000	5.72%
Namibia	5	462,000	488,000	26,000	1.10%	314,000	368,000	54,000	3.22%
Mexico	5	21,176,000	21,970,000	794,000	0.74%	65,251,000	70,652,000	5,401,000	1.60%
Cuba	5	1,589,000	1,452,000	(137,000)	-1.79%	2,869,000	2,859,000	(10,000)	-0.07%
Belarus	5	1,370,000	1,061,000	(309,000)	-4.98%	3,692,000	3,906,000	214,000	1.13%
Georgia	5	620,000	520,000	(100,000)	-3.46%	1,079,000	1,078,000	(1,000)	-0.02%
Kuwait	4	455,000	431,000	(24,000)	-1.08%	1,994,000	2,568,000	574,000	5.19%
United States	4	43,938,000	44,498,000	560,000	0.25%	185,753,000	198,426,000	12,673,000	1.33%
El Salvador	4	1,223,000	1,299,000	76,000	1.21%	1,086,000	1,099,000	13,000	0.24%
Tunisia	4	1,984,000	1,737,000	(247,000)	-2.62%	2,590,000	2,785,000	195,000	1.46%
Bosnia	4	530,000	440,000	(90,000)	-3.65%	342,000	342,000	-	0.00%
Qatar	3	267,000	361,000	94,000	6.22%	1,188,000	1,552,000	364,000	5.49%
Malaysia	3	5,703,000	5,693,000	(10,000)	-0.04%	10,148,000	11,723,000	1,575,000	2.93%

Panama	3	631,000	667,000	36,000	1.12%	1,504,000	1,673,000	169,000	2.15%
Dominican Republic	3	1,883,000	1,927,000	44,000	0.46%	3,190,000	3,571,000	381,000	2.28%
Singapore	3	747,000	784,000	37,000	0.97%	5,074,000	5,535,000	461,000	1.75%
Argentina	3	6,890,000	7,023,000	133,000	0.38%	24,185,000	25,327,000	1,142,000	0.93%
Chile	3	2,885,000	2,812,000	(73,000)	-0.51%	8,955,000	9,427,000	472,000	1.03%
Jamaica	3	532,000	540,000	8,000	0.30%	584,000	587,000	3,000	0.10%
Serbia	3	1,228,000	1,098,000	(130,000)	-2.21%	1,336,000	1,369,000	33,000	0.49%
Greece	3	1,215,000	1,086,000	(129,000)	-2.22%	3,977,000	3,970,000	(7,000)	-0.04%
Trinidad and Tobago	3	220,000	186,000	(34,000)	-3.30%	546,000	545,000	(1,000)	-0.04%
Romania	3	2,719,000	2,218,000	(501,000)	-3.99%	2,817,000	2,779,000	(38,000)	-0.27%
Australia	2	3,133,000	3,172,000	39,000	0.25%	15,290,000	16,786,000	1,496,000	1.88%
Norway	2	634,000	675,000	41,000	1.26%	898,000	969,000	71,000	1.53%
New Zealand	2	637,000	658,000	21,000	0.65%	2,097,000	2,254,000	157,000	1.45%
Uruguay	2	516,000	518,000	2,000	0.08%	1,659,000	1,707,000	48,000	0.57%
Belgium	2	1,323,000	1,325,000	2,000	0.03%	4,416,000	4,546,000	130,000	0.58%
North Macedonia	2	321,000	285,000	(36,000)	-2.35%	521,000	560,000	39,000	1.45%

Croatia	2	517,000	487,000	(30,000)	-1.19%	688,000	687,000	(1,000)	-0.03%
Bulgaria	2	894,000	701,000	(193,000)	-4.75%	1,883,000	1,933,000	50,000	0.53%
Costa Rica	1	851,000	822,000	(29,000)	-0.69%	1,712,000	1,903,000	191,000	2.14%
Switzerland	1	929,000	942,000	13,000	0.28%	3,070,000	3,255,000	185,000	1.18%
Albania	1	538,000	498,000	(40,000)	-1.53%	409,000	449,000	40,000	1.88%
Taiwan	1	3,199,000	3,111,000	(88,000)	-0.56%	12,896,000	13,382,000	486,000	0.74%
Spain	1	4,912,000	4,554,000	(358,000)	-1.50%	16,231,000	17,226,000	995,000	1.20%
United Kingdom	1	8,317,000	8,079,000	(238,000)	-0.58%	27,552,000	28,822,000	1,270,000	0.91%
Italy	1	5,914,000	5,857,000	(57,000)	-0.19%	24,836,000	25,477,000	641,000	0.51%
France	1	7,715,000	7,547,000	(168,000)	-0.44%	22,969,000	23,614,000	645,000	0.56%
Germany	1	8,999,000	8,665,000	(334,000)	-0.75%	16,000,000	16,384,000	384,000	0.48%
Slovakia	1	772,000	654,000	(118,000)	-3.26%	414,000	422,000	8,000	0.38%
Czech Republic	1	1,298,000	1,070,000	(228,000)	-3.79%	1,611,000	1,647,000	36,000	0.44%
Lithuania	1	455,000	376,000	(79,000)	-3.74%	526,000	531,000	5,000	0.19%
Japan	1	12,871,000	12,119,000	(752,000)	-1.20%	91,593,000	92,490,000	897,000	0.20%
Denmark	0	685,000	749,000	64,000	1.80%	1,192,000	1,271,000	79,000	1.29%

Canada	0	4,606,000	4,649,000	43,000	0.19%	20,411,000	21,809,000	1,398,000	1.33%
Finland	0	661,000	650,000	(11,000)	-0.34%	1,434,000	1,549,000	115,000	1.55%
Sweden	0	1,270,000	1,204,000	(66,000)	-1.06%	2,184,000	2,366,000	182,000	1.61%
Austria	0	1,023,000	1,008,000	(15,000)	-0.29%	1,731,000	1,835,000	104,000	1.17%
Netherlands	0	2,045,000	2,073,000	28,000	0.27%	3,499,000	3,626,000	127,000	0.72%
Ireland	0	599,000	535,000	(64,000)	-2.23%	1,100,000	1,163,000	63,000	1.12%
South Korea	0	6,562,000	6,704,000	142,000	0.43%	33,247,000	34,036,000	789,000	0.47%
Portugal	0	1,165,000	1,087,000	(78,000)	-1.38%	4,097,000	4,182,000	85,000	0.41%
Hungary	0	1,230,000	1,142,000	(88,000)	-1.47%	1,734,000	1,746,000	12,000	0.14%
Estonia	0	178,000	135,000	(43,000)	-5.38%	394,000	416,000	22,000	1.09%
Poland	0	5,376,000	4,479,000	(897,000)	-3.59%	5,972,000	5,957,000	(15,000)	-0.05%
Latvia	0	297,000	212,000	(85,000)	-6.52%	665,000	646,000	(19,000)	-0.58%

Appendix E

A report of the output from the analysis process

Country	Fragility Index	Youth Bulge Avg Annual Growth	Urban Population Average Annual Growth	Predicted Fragility Index
Afghanistan	20	5.49%	3.06%	12.41
Albania	1	-1.53%	1.88%	7.08
Algeria	11	-2.40%	1.41%	5.48
Angola	16	3.20%	5.46%	14.42
Argentina	3	0.38%	0.93%	6.70
Armenia	6	-4.29%	0.09%	3.88
Australia	2	0.25%	1.88%	8.33
Austria	0	-0.29%	1.17%	7.01
Azerbaijan	10	-2.60%	1.28%	5.81
Bahrain	9	-0.54%	8.60%	14.13
Bangladesh	11	0.56%	3.13%	9.53
Belarus	5	-4.98%	1.13%	4.44

Belgium	2	0.03%	0.58%	6.52
Benin	10	2.97%	3.64%	11.26
Bolivia	11	1.46%	2.08%	8.93
Bosnia	4	-3.65%	0.00%	4.15
Brazil	6	-0.17%	1.47%	6.55
Bulgaria	2	-4.75%	0.53%	4.01
Burkina Faso	16	3.01%	4.88%	12.63
Burundi	21	0.09%	6.20%	10.82
Cambodia	10	1.36%	3.16%	9.82
Cameroon	16	2.06%	5.23%	13.05
Canada	0	0.19%	1.33%	7.42
Central African Rep.	23	0.91%	2.16%	8.62
Chad	19	3.90%	3.68%	11.75
Chile	3	-0.51%	1.03%	6.61
China	7	-4.07%	2.89%	7.51
Colombia	11	0.44%	2.32%	9.17

Congo-Brazzaville	13	2.07%	4.37%	11.63
Costa Rica	1	-0.69%	2.14%	7.80
Côte d'Ivoire	16	3.05%	2.58%	10.36
Croatia	2	-1.19%	-0.03%	5.39
Cuba	5	-1.79%	-0.07%	4.85
Czech Republic	1	-3.79%	0.44%	4.37
Dem. Rep. of Congo	24	3.30%	4.40%	14.38
Denmark	0	1.80%	1.29%	8.25
Djibouti	12	0.34%	1.83%	8.03
Dominican Republic	3	0.46%	2.28%	8.56
Ecuador	7	1.07%	1.93%	8.52
Egypt	10	0.02%	2.05%	7.93
El Salvador	4	1.21%	0.24%	6.87
Equatorial Guinea	12	3.83%	6.36%	14.21
Eritrea	15	-2.79%	3.69%	8.18
Estonia	0	-5.38%	1.09%	4.35

Ethiopia	20	4.34%	4.71%	13.70
Finland	0	-0.34%	1.55%	7.39
France	1	-0.44%	0.56%	5.46
Gabon	8	2.40%	2.85%	10.10
Gambia	15	2.93%	1.34%	8.90
Georgia	5	-3.46%	-0.02%	4.19
Germany	1	-0.75%	0.48%	5.14
Ghana	11	1.88%	4.23%	11.56
Greece	3	-2.22%	-0.04%	4.65
Guatemala	8	2.45%	1.35%	8.62
Guinea	18	2.76%	2.46%	9.92
Guinea-Bissau	17	2.06%	3.97%	11.00
Haiti	11	0.69%	2.64%	9.00
Honduras	7	2.29%	3.44%	10.68
Hungary	0	-1.47%	0.14%	5.31
India	11	0.65%	2.62%	8.65

Indonesia	8	1.05%	1.92%	7.77
Iran	9	-4.66%	1.60%	3.56
Iraq	18	3.97%	3.13%	12.31
Ireland	0	-2.23%	1.12%	5.97
Israel	7	1.32%	1.91%	8.75
Italy	1	-0.19%	0.51%	5.64
Jamaica	3	0.30%	0.10%	6.30
Japan	1	-1.20%	0.20%	1.92
Jordan	6	4.91%	5.51%	14.47
Kazakhstan	9	-3.48%	3.09%	7.25
Kenya	10	2.33%	3.71%	11.17
Kuwait	4	-1.08%	5.19%	10.80
Kyrgyzstan	12	-2.01%	2.09%	6.98
Laos	12	-0.03%	1.16%	7.11
Latvia	0	-6.52%	-0.58%	2.11
Lebanon	5	4.43%	2.29%	10.69

Liberia	13	3.12%	3.76%	11.41
Libya	14	-2.18%	2.30%	7.16
Lithuania	1	-3.74%	0.19%	4.29
Madagascar	12	3.55%	5.36%	13.48
Malawi	13	2.87%	3.65%	11.17
Malaysia	3	-0.04%	2.93%	9.22
Mali	15	2.90%	3.49%	11.09
Mauritania	16	2.28%	4.45%	11.63
Mexico	5	0.74%	1.60%	8.28
Moldova	8	-5.05%	-1.07%	2.33
Mongolia	7	-3.13%	3.70%	8.08
Morocco	7	-0.47%	1.42%	6.91
Mozambique	11	3.37%	4.48%	12.59
Myanmar	18	0.89%	2.56%	8.86
Namibia	5	1.10%	3.22%	9.75
Nepal	9	2.41%	4.38%	11.57

Netherlands	0	0.27%	0.72%	6.77
New Zealand	2	0.65%	1.45%	7.83
Nicaragua	7	-0.21%	0.70%	6.57
Niger	19	4.80%	3.68%	12.29
Nigeria	18	2.51%	3.49%	13.89
North Korea	7	0.23%	0.43%	6.29
North Macedonia	2	-2.35%	1.45%	6.26
Norway	2	1.26%	1.53%	8.21
Oman	5	-0.50%	10.83%	16.53
Pakistan	15	1.74%	3.00%	11.46
Panama	3	1.12%	2.15%	8.77
Papua New Guinea	10	2.50%	2.03%	9.31
Paraguay	8	0.71%	2.73%	9.19
Peru	6	-0.40%	1.89%	7.71
Philippines	14	1.61%	2.04%	9.13
Poland	0	-3.59%	-0.05%	3.30

Portugal	0	-1.38%	0.41%	5.59
Qatar	3	6.22%	5.49%	14.78
Romania	3	-3.99%	-0.27%	3.30
Russia	9	-6.23%	0.87%	(0.07)
Rwanda	15	1.06%	3.60%	10.07
Saudi Arabia	8	-0.98%	3.33%	9.83
Senegal	10	2.16%	2.69%	9.90
Serbia	3	-2.21%	0.49%	5.28
Sierra Leone	13	2.32%	2.88%	10.08
Singapore	3	0.97%	1.75%	8.38
Slovakia	1	-3.26%	0.38%	4.69
Somalia	20	4.39%	5.35%	14.02
South Africa	8	-0.85%	2.53%	8.41
South Korea	0	0.43%	0.47%	5.75
South Sudan	22	2.77%	4.71%	12.04
Spain	1	-1.50%	1.20%	6.13

Sri Lanka	11	-0.59%	0.24%	5.76
Sudan (North)	21	3.20%	2.69%	10.69
Sweden	0	-1.06%	1.61%	7.04
Switzerland	1	0.28%	1.18%	7.34
Syria	14	-4.32%	-5.84%	(4.48)
Taiwan	1	-0.56%	0.74%	6.16
Tajikistan	11	-0.05%	2.41%	8.30
Tanzania	10	2.54%	5.42%	13.47
Thailand	8	0.06%	2.49%	9.06
Togo	13	2.24%	2.21%	9.40
Trinidad and Tobago	3	-3.30%	-0.04%	4.32
Tunisia	4	-2.62%	1.46%	5.97
Turkey	9	0.74%	2.35%	9.77
Turkmenistan	8	-1.53%	2.42%	7.56
Uganda	14	3.56%	5.03%	13.06
Ukraine	9	-5.07%	-0.06%	2.03

United Arab Emirates	5	-3.13%	5.72%	10.63
United Kingdom	1	-0.58%	0.91%	5.93
United States	4	0.25%	1.33%	6.90
Uruguay	2	0.08%	0.57%	6.63
Uzbekistan	11	-0.99%	1.51%	6.57
Venezuela	14	-0.01%	1.50%	7.41
Vietnam	7	-2.27%	4.17%	8.70
Yemen	21	1.93%	4.18%	11.48
Zambia	12	3.30%	4.11%	12.03
Zimbabwe	17	-0.32%	0.26%	5.91

Bibliography

Abé Ndjié, A., Atangana Ondoa, H., & Ngoa Tabi, H. (2019). Governance and youth unemployment in Africa. *Labor History*, *60*(6), 869–882. https://doi.org/10.1080/0023656x.2019.1645320 Accessed 22 June 2020.

Atin Basuchoudhary, Bang, J. T., Tinni Sen, & David, J. (2018). *Predicting hotspots : using machine learning to understand civil conflict*. Lexington Books.

Bettencourt, L. M. A. (2020). Urban growth and the emergent statistics of cities. Science Advances, 6(34). https://doi.org/10.1126/sciadv.aat8812

Froneberg, E. (2019). Youth Cohorts and the Risk of Conflict Recurrence: A Global Quantitative Analysis (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-384563 Accessed 9-18-2020

Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2022). *Multivariate Data Analysis, 8th edition*. CENGAGE INDIA.

Liotta, P. H., & Miskel, J. F. (2009). The 'Mega-Eights': Urban Leviathans and International Instability. *Orbis*, *53*(4), 647–663. https://doi.org/10.1016/j.orbis.2009.07.003 Accessed 9 June 2020.

Marshal, M. G., & Elzinga-Marshal, G. (n.d.). Table 1: State Fragility Index and Matrix 2018. N. Vienna, VA; Center for Systemic Peace. http://www.systemicpeace.org/inscr/SFImatrix2016c.pdf Accessed 13 November 2020.

Marshall, Monty. 2001. "Regime Authority, Opportunity and Outbreaks of State Failure Events." University of Maryland, Center for International Development and Conflict Management. [Google Scholar].

Moghadam, H.S., & Helbich, M. (2013). Spatiotemporal urbanization processes in the megacity of Mumbai, India: A Markov chains-cellular automata urban growth model. Applied Geography, 40, 140–149. https://doi.org/10.1016/j.apgeog.2013.01.009

Ndjie, A. A., Ondoa, H. A., & Tabi, H. N. (2019, July 26). *Governance and youth unemployment in Africa*. Taylor & Francis. Retrieved June 22, 2020, from https://www.tandfonline.com/doi/full/10.1080/0023656X.2019.1645320

O'Brien, S.P. (2002). Anticipating the Good, the Bad, and the Ugly: An Early Warning Approach to Conflict and Instability Analysis. *The Journal of Conflict Resolution*, 46(6), 791–811.

Ooi, G. L. (2007). The Urbanization of Everything: Governance Challenges in Southeast Asia. *Georgetown Journal of International Affairs*, *8*(2), 39–45. https://www.jstor.org/stable/43133730 Accessed 9 June 2020.

Owain, E. L., & Maslin, M. A. (2018). Assessing the relative contribution of economic, political and environmental factors on past conflict and the displacement of people in East Africa. *Palgrave Communications*, *4*(1). https://doi.org/10.1057/s41599-018-0096-6 Accessed 22 June 2020.

Sommers, M. (2011). Governance, Security and Culture: Assessing Africa's Youth Bulge. *International Journal of Conflict and Violence (IJCV)*, *5*(2), 292–303. https://doi.org/10.4119/ijcv-2874 Accessed 22 June 2020.

United Nations. (n.d.). *World population prospects - population division*. United Nations. Retrieved December 30, 2020, from https://population.un.org/wpp/Download/Standard/Population/

United Nations. (n.d.). *World urbanization prospects - population division*. United Nations. Retrieved February 4, 2022, from https://population.un.org/wup/Download/

Urdal, H. & Hoelscher, K. (2012) Explaining Urban Social Disorder and Violence: An Empirical Study of Event Data from Asian and Sub-Saharan African Cities. *International Interactions*, 38(4), 512-528. https://doi.org.afit.idm.oclc.org/10.1080/03050629.2012.697427

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	**2. REPORT TYPE**	**3. DATES COVERED** (From - To)
24-03-2022	Master's Thesis	June 2020 - March 2022

4. TITLE AND SUBTITLE
Indicators of Political Instability in the Presence of Rapid Urban and Youth Population Growth

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Hubbard, Douglas W, GS-11

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute Of Technology
Graduate School of Engineering and Management (AFIT/ENS)
2950 Hobson Way
Wright-Patterson AFB, OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENS-MS-22-M-193

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement A.
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES This work is declared a work of the U.S.Government and is not subject to copyright protection in the United States

14. ABSTRACT
Large and rapidly growing cities and other urban agglomerations have the potential to become incubators of political instability. This is especially true of rapidly growing cities which are located in countries that are also experiencing high rates of growth in their youth population. Rapid growth rates put stress on urban infrastructure and other institutions, and these stresses can cause major problems for both city and national governments. Knowing when these cities and countries may be trending toward their tipping points regarding political instability will help governments and international organizations develop and implement effective strategies to mitigate the risk of instability.

15. SUBJECT TERMS
Indicators of Political Instability; Rapid Urban Growth

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. Richard Deckro
					19b. TELEPHONE NUMBER (Include area code)
U	U	U	UU	82	(312) 785-3636 ext 4325