2006

Forecasting Stability Levels for the Countries of the Former Soviet Union

Margaret Erin Keck
Virginia Commonwealth University

The opinions and interpretation expressed herein are my own and are not necessarily shared by the U.S. Army, the Department of Defense any other agency of the U.S. government, or Virginia Commonwealth University

Forecasting Stability Levels for the Countries of the Former Soviet Union

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematical Sciences (Operations Research) at Virginia Commonwealth University.

by

Margaret Erin Keck
BA (Mathematics), Sweet Briar College, 2003

Director: Dr. Jason Merrick,
Associate Professor,
Department of Statistical Sciences & Operations Research

Virginia Commonwealth University
Richmond, Virginia
May, 2006

Acknowledgment

The author wishes to thank several people. First, I would like to thank my family and my fiancé, Terry, for their love, support and patience. I would like to thank Dr. Merrick for his time and effort in helping me to complete this project. And, finally, I would like to thank the National Ground Intelligence Center for giving me the opportunity to return to school and complete my degree.

Table of Contents

List of Tables

List of Figures

ABSTRACT

FORECASTING STABILITY LEVELS FOR THE COUNTRIES OF THE FORMER SOVIET UNION

By Margaret Erin Keck, B.A.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematical Sciences (Operations Research) at Virginia Commonwealth University.

Virginia Commonwealth University, 2006

Director: Dr. Jason Merrick,
Associate Professor, Department of Statistical Sciences & Operations Research

United States intelligence officers and policymakers need reliable forecasts of country, regional, and global stability or instability. Such forecasts require a methodology for identifying and analyzing factors that contribute to stability. The anticipation of this stability level can facilitate crisis warning and diplomatic strategies for various timelines, including five, ten, and twenty year forecasts. While the problem of forecasting can be tackled in various ways, in the interest of time and space, I will only go into a few of them. The approach I will use is multiple linear regression to generate a short-term forecast for the stability levels of the countries of the Former Soviet Union (FSU). This model could ultimately be used to help formulate policies that enhance stability in developing or transitioning countries.

1. Introduction

United States intelligence officers and policymakers need reliable forecasts of country, regional, and global stability or instability. Such forecasts require a methodology for identifying and analyzing factors that contribute to stability. The anticipation of this stability level can facilitate crisis warning and diplomatic strategies for various timelines, including five, ten, and twenty year forecasts. The approach I will use is multiple linear regression to generate a short-term forecast [short-term in this context refers to a 5 year forecast] for the stability levels of the countries of the Former Soviet Union (FSU). These countries, Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Russia, Tajikistan, Turkmenistan, Ukraine, and Uzbekistan, hereafter referred to as the countries of the FSU, came into existence in December, 1991 when the Soviet Union was dissolved, and each declared their respective independence. Since the Soviet Union was such a different entity than any of the present countries, both politically and financially (ethnically, religiously…etc.), using the factors that contributed to the stability of the Soviet Union would not be beneficial to our forecasts.

Information pertaining to a country's instability is useful in anticipating crisis (O'Brien, 2002). "*Those who make foreign and international policy seek more than explanation, they want better 'early warning' of impending conflicts so that preventative diplomacy and other conflict management tools can be brought in to play*" (Gurr and Moore, 1997). There are various crisis warning and diplomatic strategies that the U.S. government has at its disposal, and all depend to some degree on reliable forecasts of state stability. In this paper, I will discuss aspects of this problem. First, I will focus on multiple linear regression and how it is used in economic forecasting. Then, I will discuss some solutions to the forecasting problem, and give an overview of the O'Brien world-wide forecasting study discussed in O'Brien (2002). Finally, I will develop a strategy for forecasting stability in the countries of the FSU.

2. Literature Review

2.1 Regression Analysis and Forecasting

Multiple linear regression has been used in various scenarios to determine future values, often in economics by fitting a linear equation to observed data. Multiple linear regression is most applicable when there is a belief that there is a strong linear relationship between the independent variables and the variable being predicted. The fitted line is then a description of the mean response. Therefore, this method should only be used when there is accurate data and reliable predictions of the independent variables (National Center for Education Statistics, 2003). The use of multiple linear regression aims to produce a realistic forecast, not one that simply extends historical growth rates into the future, as this would be assuming that all of your variables will behave the same at any given point in the future. A more realistic approach is to specify the major factors and any inter-relationships between them (Wagle, 1965). It is true that a number of models could fit our data and, therefore, we cannot rely solely on a mathematical formula. Forecasting methodology can be used to support other non-quantitative approaches. "*The aim of mathematical models is not to supplant judgment and experience, but to enable the managers to get a greater insight into the nature of complex problems and to reduce a portion of uncertainty present in their problems*" (Wagle, 1965).

Forecasting is best developed and most widely used in economics, therefore an examination of economic forecasting can provide insight into some other types of forecasting. One of the main issues facing forecasters concerns the amount of data required. That is, how much historical data is needed in order to get an accurate forecast? Obviously, short-term forecasting can be the most accurate, but there are times when long-term forecasts are needed [short term is defined in this context as months, a year, or up to 5 years. Long-term is measured in multiple years for our purposes]. In the long-term context, is looking back twenty years enough to get a twenty year forecast? James Clark, in his article 'Long-Term forecasts of inflation rates using quantitative tools,' addresses this question. One approach is point estimation. Point estimation takes an average value from different time period lengths and with their standard deviation, uses these averages as a forecast of the future. The approach has a major downfall in the fact that it is missing variability. Another aspect covered in Clark is a computer program called Decisioneering (2001) (www.decisioneering.com). This program uses moving averages (exponential smoothing) to conclude that a forecast should be made for ½ to ¾ of the period of historical data used. Using this methodology, we would need 30-40 years of historical data to get a twenty year forecast. Clark stresses that you have to know your data set (2001). In forecasting inflation rates, for example, looking back farther than twenty years seems inappropriate. The state of the world, and the United States, prior to 1980 was dramatically different economically from today. A simple histogram of the data shows an inconsistency prior to 1982. The conclusion was that using the time period 1982-1999 to forecast 2001-2020 was going to be the best approach. Obviously, other tools can be used to validate the conclusion after making this decision (Clark, 2001).

The next question that arises after determining your historical data is: which forecasting method is the best to use? Armstrong (1984) examines whether more complex models actually provide better results. Table 1 from Armstrong (1984) demonstrates that simple methods seem to be as adequate as the complex models with most comparisons showing a negligible difference. On this chart, in the results column, a "+" means that the complex methods were more accurate, a "0" means that there was a negligible difference, and a "–" means that the simple methods were more accurate. Armstrong concludes saying, "*I suggest starting with the least expensive method (1984).*"

Table 1 - Model Complexity Comparison (Armstrong, 1984, pp. 52-66)

Study	Major Comparisons	Results
Winters [1960]	Exponential Smoothing versus moving averages	+
Frank [1969]	Exponential Smoothing versus moving averages	+
Eltom and Gruber [1972]	Exponential Smoothing versus moving averages	+
Chow [1965]	Adaptive versus constant parameters	+
Whybark [1972]	Adaptive versus constant parameters	+
Smith [1974]	Adaptive versus constant parameters	+
Dennis [1978]	Adaptive versus constant parameters	+
Brown and Rozeff [1978]	Box-Jenkins versus simple trend	+
Newbold and Granger [1974]	Box-Jenkins versus exponential smoothing	+
Reid [1975]	Box-Jenkins versus exponential smoothing	+
Dalrymple [1978]	Box-Jenkins versus regression	+
Kirby [1966]	Exponential Smoothing versus moving averages	0
Adam [1973]	Exponential Smoothing versus moving averages	0
Raine [1971]	Adaptive versus constant parameters	0
Dancer and Grey [1977]	Adaptive versus constant parameters	0
Chatfield and Prothro [1973]	Box-Jenkins versus no-change	0
Albercht et al. [1977]	Box-Jenkins versus no-change	0
Bates and Granger [1969]	Box-Jenkins versus exponential smoothing	0
Groff [1973]	Box-Jenkins versus exponential smoothing	0
Geurts and Ibrahim [1975]	Box-Jenkins versus exponential smoothing	0
Mabert [1976]	Box-Jenkins versus exponential smoothing	0
Chatfield [1978]	Box-Jenkins versus exponential smoothing	0
Kenny and Durbin [1982]	Box-Jenkins versus exponential smoothing	0
Torfin and Hoffman [1968]	6 models of varying complexity	0
Markland [1970]	4 models of varying complexity	0
Johnson and Schmitt [1974]	10 models of varying complexity	0
Carey [1978]	21 models of varying complexity	0
Hagerman and Ruland [1979]	3 models of varying complexity	0
Makridakis and Hibon [1979]	22 models of varying complexity	0
Kuland [1980]	8 models of varying complexity	0
Makridakis et al. [1982]	21 models of varying complexity	0
Armstong [1975]	Complex curve versus rule of thumb	0
Mabery [1978]	Adaptive versus constant parameters	-
Gardner [1979]	25 models of varying complexity	-
Gardner and Dannenbring [1980]	Adaptive versus constant parameters	-
McLeavey, Lee and Adam [1981]	Adaptive versus constant parameters	-
Ledolter and Abraham [1981]	Simple versus complex models	-
Coggin and Hunter [1982-3]	Simple versus complex models	-
Brandond, Jarrett and Khumawala [1983]	Box-Jenkins versus 8 simple models	-

2.2 Forecasting Political Instability

Various government agencies have tried to tackle the problem of forecasting country instability. In 1994 then Vice-President Al Gore, in the aftermath of the genocide in Rwanda, commissioned the State Failure Task Force (SFTF), now known as the Political Instability Task Force, to "*Identify and examine key factors associated with serious state crisis and to develop a methodology that could identify 'critical thresholds' in these factors so that they might provide early warning of state failures up to two years in advance*" (Esty et al., 1995). This task force established a classification of unstable events divided into four categories: genocides and politicides, ethnic wars, revolutionary wars, and adverse or disruptive regime transactions. Genocides and politicides are defined as sustained policies by states or their agents and, in civil wars, by contending authorities that result in the deaths of a substantial portion of members of communal or political groups. Ethnic wars are secessionist civil wars, rebellions, protracted communal warfare, and sustained episodes of mass protest by politically organized communal groups. Revolutionary wars are sustained military conflicts between insurgents and central governments, aimed at displacing the regime. Finally, adverse or disruptive regime transitions are defined as major, abrupt shifts in patterns or governance, including state collapse, periods of severe instability, and shifts towards authoritarian rule (Esty et al., 1995). The goals of the task force were to identify factors consistently associated with the onset of political stability or instability and to develop models that can accurately assess the relative vulnerability of countries world-wide to the onset of instability (Goldstone et al., 2005). This task force used logistic regression, neural networks and genetic algorithms to identify patterns between hundreds of explanatory variables and different types of state failures. All task force data was drawn from open sources or developed by the task force members themselves. The task force does not use any classified data (Goldstone et al., 2005). They derived a 'best' overall global model consisting of only three variables: measure of a country's level of democracy, trade openness, and infant mortality rate; these variables will be defined in more detail later. This model proved to be successful in predicting conflict 2/3 of the time but was received with much criticism (Esty et al., 1995). The task force did find that relatively simple models can identify the factors associated with a broad range of political violence and instability events around the world (Goldstone et al., 2005).

O'Brien (2002) uses pattern classification algorithm and fuzzy analysis of statistical evidence (FASE) to draw his conclusions, for an overview of FASE see Chen, 1995 or O'Brien, 2006. O'Brien analyses the relationships between country macro-structural factors and historical instances of country instability (2002). In addition to the occurrence of conflict, it also forecasts the level of intensity of the instability with 80% accuracy. The rating system used is similar to the SFTF system. The goal of the O'Brien study is to identify the probability of country instability as well as the certain conditions conducive to instability over the next 15 years (2002). It identifies factors, that, when combined with unknown triggers such as assassination, riots, and natural disasters, have historically been associated with certain levels of instability. In doing so, it attempts to forecast the certain conditions conducive to conflict, not the specific events that set off a particular conflict, noting that "*...regardless of our ability to predict the spark that will set them ablaze, we can forecast the oiliness of the oily rags with a reasonable degree of expected accuracy* (O'Brien, 2002)." The different factors contributing to instability serve as "oily rags" – the oilier the rag, the more likely a single spark such as a riot or assassination might produce an explosive situation. The reverse also holds, since the more stable a country is, the better equipped it is to handle such a spark (O'Brien, 2002).

Conflict data for this project was drawn from the KOSIMO database (German acronym for: Conflict-Simulation-Model), which was developed by the Heidelberg Institute of International Conflict Research and covers the time period of 1975-1999. The attached chart (Appendix A.1) is a summary of the KOSIMO conflicts (Pfetch and Rohloff, 2000). Conflict is defined here as:

- The "*clashing of overlapping interests (positional differences) around national values and issues (independence, self-determination, borders and territory, and access to or distribution of domestic or international power)*
- *The conflict has to be of some duration and magnitude of at least two parties (states, groups of states, organizations, or organized groups) that are determined to pursue their interests and win their case*
- *At least one party is the organized state*" (Heidelberg Institute of International Conflict Research (HIIK)).

Twelve different variables were found to be useful in this study. As shown in Table 2, these variables reflect a country's tendency to engage in conflict, it's ability and commitment to achieve economic performance, it's commitment to public health and the welfare of it's people, commitment to political rights and civil liberties and to the rules of global trade, and, finally, demographic indictors such as ethnic and religious diversity as well as the age distribution in each country.

Table 2 - Variable List

1. % of history spent in state of conflict	**7**. Political Rights index (1-7)
2. Infant Mortality Rate	**8**. Democracy (-10 - 10)
3. Trade Openness	**9**. Religious Diversity
4. Youth Bulge	**10**. Caloric Intake
5. Civil Liberties index (1-7)	**11**. GDP per capita
6. Life Expectancy	**12**.Ethnic Diversity

The definitions of each of these independent variables are included in an attached table, Appendix A.2 (O'Brien, 2002). It can be seen that the variables used in the SFTF full model (Democracy, Trade Openness, and Infant Mortality Rate) are included above. Next, O'Brien, 2002 developed an instability rating system to find the probability of each conflict type occurring in each country in a year and then find the combined probability for each relative instability level: high, moderate, or none/low intensity.

Figure 1 - Instability Levels (O'Brien, 2005b)

Once the combined probability is found, any country/year with an instability level probability ≥ 67% is labeled as such; for example, (1 - 2), (2 - 3), or (3 - 4). If more than one of these choices is ≥ 67% then the higher of the two options is chosen. If none apply (nothing ≥ 67%) the country/year is then labeled as "uncertain." Uncertain does not mean either correct or incorrect, only that there is not enough information to draw a conclusion. Historically, about 50% of these "uncertain" cases will result in the country experiencing at least some level of instability that year (O'Brien, 2002). In these cases the conflict type is rounded up. For example, the conflict in Northern Ireland may have received a score between 2 and 3, and, therefore, will be labeled as a violent crisis with moderate to high levels of instability.

In order to determine the model's performance, three ratios were established to measure the ability to predict instability. 'Overall Accuracy' is the proportion of a country-year observations that were classified correctly over any given forecast period.

'Recall' is the ability to correctly forecast the conflict type. And 'Precision' is the ability not to produce too many false positives (precision = 0.8 means 20% of the results were false positives or over-predictions). Ideally, recall and precision should be as close to 1 as possible, and all three of these measures should be at least 60% to be deemed "good" (O'Brien, 2005a).

Table 3 - Model Performance

Overall Accuracy	Recall	Precision
# of correct predictions / # of predictions made	# of correctly predicted conflicts / # of conflicts that occurred	# of correctly predicted conflicts / # of conflicts predicted to occur

As mentioned previously, Fuzzy Analysis of Statistical Evidence (FASE) was used as the methodology for this project. FASE was developed by Yuan Yan Chen, who is also employed by the Center for Army Analysis. It is a hybrid method combining elements from statistics, possibility theory and fuzzy logic. It utilizes the principle of inverse inference so it has properties similar to Bayesian classifiers. The principal difference is that classification is performed on the possibility rather than the probability measure (O'Brien, 2002). With FASE, the conditional probabilities from each attribute are normalized into possibilities and combined with a t-norm from fuzzy set theory (Chen, 1995). Once the attribute possibilities have been combined into overall possibilities for the class label, the class label possibilities can be normalized into probabilities for a more familiar interpretation. To find these class labels the Frank Rule (Frank, 1979) is applied to the possibility scores to produce overall likelihood ratios for each of the independent variables (O'Brien, 2002).

Frank Rule: $$T(x_1, x_2, \ldots, x_k) = \log\left(\frac{s(1+(sx_1-1)(sx_2-1)\cdots(sx_k-1))}{(s-1)(k-1)}\right),$$

where k is the number of attributes and s is an adjustment parameter that is set close to 0 if our independent variables are highly correlated, and close to (but not equal to) 1 if they are independent (O'Brien, 2002).

In addition O'Brien developed the Analyzing Complex Threats for Operations and Readiness (ACTOR) model, which takes the probability of conflict for 167 different countries from 2005-2015 using historical data from 1975-2003 and puts the conclusions into a user-friendly form (2002). This spreadsheet makes it possible to examine 'what-if' scenarios for each country year by changing the values of any of the variables to reflect different changes and then compute the 'what-if' conclusions easily. This allows anyone using the interface to change variables to see what kind of effect a change in environment will have on country stability (O'Brien, 2005a).

There are some inherent problems with the O'Brien, 2002 study, some of which are acknowledged in the conclusion. For instance, some variables cannot be measured or measurements are not available. Examples of these are water scarcity, deforestation, other environmental players as well as some religious data. Also, it is very difficult to predict what spark will set off unrest in any country, it is only possible to forecast how susceptible a country is to conflict. Finally, the farther out you try to forecast, the more unreliable the forecast becomes. The 5 year forecasts are 85% accurate but that percentage declines as the out-year forecast increases (O'Brien, 2002).

O'Brien did not deal specifically with the countries of the FSU (2002). Only the most recently available observations or averaged set of observations were available and those were used to develop a straight-line forecast through the forecast period (O'Brien, 2002). I have already established that this lacks variability as described in Clark (2001).

As I stated before, my project will focus solely on Russia and the fourteen other countries of the Former Soviet Union, therefore, I will use the FSU countries to determine variables and develop a specific FSU model using demographic, social, and cultural data. I can then use that model to forecast stability levels of a country out 5 years. This differs from O'Brien in two ways (2002). First is the simple realization that the history of the FSU is much different than the other countries of the world that have a much longer history. Within this area of the world, it is quite obvious that something other than a straight-line forecast is needed to gain an accurate forecast. Second, I will use historic information to identify the independent variables that contribute (negatively or positively) to the *stability* of a country. I will then develop a model that will forecast the level of stability in a given country utilizing these independent variables. I will begin by analyzing all of the same independent variables used in O'Brien, 2002 and narrowing from there. I will use the same response variable of Maximum KOSIMO level to evaluate stability. Countries with a stability level of 2 or below (described in depth later) will be given a label of "stable;" however, anything above a 2 will be considered "unstable." Working to minimize our response variable will maximize stability. With the final model we can see which variables have an impact on stability. Working to increase (or decrease) these significant variables in a certain country should help decision makers to implement policies to enhance stability in transitioning or emerging nations. As the SFTF noted, paraphrasing Tolstoy, "*All stable nations resemble one another; each unstable nation is unstable in its own way*[1]"(Goldstone et al., 2005). I will use multiple linear regression to determine my model, and in order to maximize accuracy, I will developing a 5 year forecast, as my data

---

[1] The original is from Anna Karennina: "All happy families resemble one another; each unhappy family is unhappy in it's own way."

set only contains seven years (1992-1999). In an attempt to determine the accuracy of my results I will use data from 1992-1999 to formulate my model and collect data from December 1, 2005 to compare against my results. Data from 1991 is not being used in this project as the countries in question declared their independence from the Soviet Union by December, 1991 (ranging from March, 1990 to December, 1991), so the overall starting point will be January 1, 1992. By the same logic, the variable called % history spent in state of conflict will be removed from my project. There has been no separation between inter- and intra-state conflicts in this project.

3. Methodology

3.1 The Data

To begin the analysis of this problem, I began with the same variables that O'Brien found significant, with the exception of percent of history spent in state of conflict (2002). The reasoning for not using this variable is that the timeframe that we have to work with is too short and this variable may, in turn, seem exaggerated or disproportionate. The eleven variables remaining are Infant Mortality Rate, Trade Openness, Political Rights Index, Democratic Level, Religious Diversity, Youth Bulge, Civil Liberties Index, Life Expectancy, Caloric Intake, GDP (gross domestic product) per capita, and Ethnic Diversity. The definitions for each variable are the same as in the O'Brien, 2002 (see Appendix A).

In several instances there are missing data points. Some cases (Case 1) a data point for one year for a certain variable is missing (i.e. 1992 caloric intake for Moldova). In other cases (Case 2) the entire data set for a certain variable for a country was missing (i.e. no data on religious diversity for Ukraine 1992-1999). For Case 1, a smoothing technique was used to extrapolate the missing year(s) from the other data available using simple linear regression techniques [all missing points fall at the beginning of the data - 1992, 1993, and 1994]. In Case 2, an average was found for each missing country/variable combination, and this average (mean) was used for all seven years. Since multiple linear regression develops a line based on mean response, this is feasible and will not have an undesired impact on the results.

Replacing these missing values is crucial to the outcome of this model because I need all of the observations to be used (again, since this data set is not very large). If one data point is missing in an observation, the entire observation is thrown out. In this case 70 observations had instances of either Case 1 or Case 2 missing data points which would leave only 50 observations from which to derive a model.

3.2 Finding the Best Model

Using SAS software, the first step taken in choosing a "best" model for this problem was to use stepwise, backward, and forward regression against the full model (Table B.1). The results of these tests were then compared with the adjusted R-squared and Cp statistics to determine a first and second choice model. In choosing my model, improvement is gauged by the value of R-squared and adjusted R-squared. These two values demonstrate the level of variation explained by that certain model. In addition to the R-squared, a significance level of alpha = .25 for each of the independent variables will be deemed as "good," anything lower than .25 is better (meaning Pr > |t| should be less than .25) (Mays, 2005).

It was determined by separate software that interactions between the independent variables are not significant. It was also tested to see if introducing lag variables would prove useful. This would take the place of the Percent of History in Conflict variable, which would duplicate the effects of the lag variables. Both would be measuring the effects of the previous year(s) conflict level on the current year. This also proved not to be significant or helpful to the model.

From there, the possibility of log transformations on both the independent and dependent variables will be explored. Influence diagnostics will be run using the "best" model and any necessary observations removed from the data set. We will test for multicollinearity between our variables and finally cross-validate our chosen model. If all of these steps prove successful we will have a final model (Mays, 2005).

The model chosen by both stepwise and forward regression was comprised of only caloric intake, trade openness, and ethnic diversity, also, as seen in Table B.2, it only had an R-squared of .26 compared with the model chosen by backwards selection which produced a significant model with and R-squared of .48 (Table B.3). This is not as high as the full model's R-squared of .49; however, all of the independent variables are significant at an alpha level of .25 in the backwards selection reduced model. This model contains all of the same variables as the full model except Big Religion. A second "best" model was chosen as a back-up by Cp. This model had an R-squared of .48 and removed the caloric intake variable as well as big religion. Two models are chosen in case the cross-validation of the first model fails and we cannot use it. This way we already have a second model to work with without beginning again.

Using the reduced model, tests were run to see whether Box-Tidwell and/or Box-Cox transformations would be useful. Box-Tidwell was run first to see if a log transformation on the independent variables would improve the model (Box and Cox, 1964). This was only tested on the variables caloric intake, infant mortality rate, life expectancy, youth bulge, gross domestic product, and openness as these are the only continuous variables in our model. Nominal and ordinal variables are coded as indicators, so it wouldn't change anything to transform them. In the following list, alpha represents the power transformations for each of the continuous independent variables.

Caloric intake: alpha = 15 Infant Mortality Rate: alpha = .51

Life Expectancy: alpha = -37 | Youth Bulge: alpha = -.10

GDP: alpha = .33 | Openness: alpha = -1.6

This proved to be successful as the R-squared increased from .4877 to .5720 and all of the independent variables were now significant at alpha = .20 (Table B.4). Next, a Box-Cox transformation was done on the dependent variable, MaxKOSIMO level (Box and Tidwell, 1962). A lambda value of -1.925, denoting the maximum of the log-likelihood function, was chosen with a 95% confidence interval of [-1.425, -2.50] (The Mathworks, 2006). This improved the model, only when combined with the transformed independent variables (Table B.5). This transformed MaxKOSIMO level is now labeled as YTRANS. The transformed independent variables are XTCAL, XTIMR, XTLIFE, XTYOUTH, XTGDP, and XTOPEN for caloric intake, infant mortality rate, life expectancy, youth bulge, GDP, and trade openness, respectively.

Our newest model, with both dependent and independent variable transformations, has the highest R-squared thus far of .6552. This model currently has YTRANS as the dependent variable with independent variables as shown in the following ANOVA table.

Table 4 - Analysis of Variance - Transformations

Source	**DF**	**Sum of Sq**	**Mean Sq**	**F value**	**Pr > F**
Model	10	13.51807	1.3181	20.71	<.0001
Error	109	7.11527	0.06528		
Corr. Total	119	20.63334			

Root MSE	0.225549	R-square	0.6552
Dependent Mean	0.68662	Adj R-sq	0.6235
Coeff Car	37.21067		

Parameter Estimates

Variable	**DF**	**Par. Est.**	**St. Error**	**t value**	**Pr > |t|**
Intercept	1	17.14101	2.90233	5.91	<.0001
XTCAL	1	1.87774e-54	1.65566e-54	1.13	.02592
XTIMR	1	-0.28985	0.04253	-6.82	<.0001
XTLIFE	1	1.494051e66	2.263945e65	6.60	<.0001

XTYOUTH	1	-12.46873	2.47959	-5.03	<.0001
PolRights	1	0.06488	0.03766	1.72	0.0877
CivilLib	1	-0.08518	0.03753	-2.27	0.0252
XTGDP	1	-0.04572	0.02338	-1.96	0.0531
BigEthnic	1	-0.01720	0.00216	-7.95	<.0001
Democracy	1	-0.00549	0.01017	-0.54	0.5905
XTOPEN	1	-0.07952	0.02499	-3.18	0.0019

As you can see, not all of the independent variables are significant in this model. Backwards and stepwise selection will be run again to determine what the best model is with the "new" variables. Backwards and stepwise selection both chose the same "best" model (Table B.6). This gives us a new model containing XTIMR, XTLIFE, XTYOUTH, PolRights, CivilLib, XTGDP, BigEthnic, and XTOPEN.

Table 5 - Analysis of Variance - Selection

Source	DF	Sum of Sq	Mean Sq	F value	Pr > F
Model	8	13.40260	1.6753	25.72	<.0001
Error	111	7.23074	0.06514		
Corr. Total	119	20.63334			

Root MSE	0.25523	R-square	0.6496
Dependent Mean	0.68662	Adj R-sq	0.6243
Coeff Car	37.17191		

Parameter Estimates

Variable	DF	Par. Est.	St. Error	t value	Pr > |t|
Intercept	1	16.14804	2.80169	5.75	<.0001
XTIMR	1	-0.28249	0.04103	-6.88	<.0001
XTLIFE	1	1.477215e66	2.232029e65	6.62	<.0001
XTYOUTH	1	-11.74774	2.39922	-4.90	<.0001
PolRights	1	0.08475	0.03039	2.79	0.0062
CivilLib	1	-0.07947	0.03658	-2.17	0.0319
XTGDP	1	-0.03552	0.02078	-1.71	0.0902
BigEthnic	1	-0.01713	0.00201	-8.50	<.0001
XTOPEN	1	-0.08549	0.02382	-3.59	0.0005

YTRANS is still the response variable. This final model has a R-squared of .6496 and an adjusted R-squared of .6243. The R-squared is a little lower than the last model, but the adjusted R-squared has increased. Adjusted R-squared is a better judge of the quality of the variables and goodness of fit as it is not affected by the number of variables in a model as R-squared is. R-squared will increase as additional variables are added to the model, while adjusted R-squared may or may not increase, depending on the quality of the additional variables. This explains the results with this model. Another improvement is that now all of the independent variables are significant at an alpha level of 0.20 in the newly selected model. At this time, the chosen model is:

YTRANS = 16.14804 - 0.28249XTIMR + 1.477215e66XTLIFE - 11.74774XTYOUTH + 0.08475PolRights - 0.07947CivilLib - 0.03552XTGDP - 0.01713Big Ethnic - 0.08549XTOPEN

3.3 Analyzing the Final Model

Influence Diagnostics were run to check for outliers, high leverage points, and high influence points (Table B.7). The R-student statistics is used to gauge outliers with a measure of $|t_i| \geq 2$. The observations with an R-student higher than |2| in this data set are (89, 105, 107, 113, and 120). In the same way, HAT diagonals are used to judge high leverage points with a measure of 2p/n (p = number of parameters, n = the number of observations), which in this case equals .15. The observations with HAT diagonals higher than .15 are (16, 33, 49, 81, 102, 103, 104, and 105). High influence points are determined by an overlap between outliers and high leverage points. In this case, observation 105 is a high influence point. This observation corresponds to the data point from Ukraine in 1992. Outliers are defined as an observation within the range of the x-variables, but with a noticeable shift in the y-direction. High leverage points are in the pattern of the y-data, but are shifted in the x-direction. A high influence point, as both an outlier and a high leverage point, is out of the pattern in both the x- and y-directions. High influence points should be removed from the data set as they may skew the results.

The removal of the Ukraine, 1992 observation should improve the results of our model. A possible reason why the Ukraine in 1992 is labeled as a high influence point may be a result of the missing data. The observation of Ukraine in 1992 was one of the few that had both a Case 1 and Case 2 missing value.

The next influence diagnostic to be run was to measure the extent of any influence using DFFITS and Cook's D. DFFITS is a scaled measure of the change in the predicted value for a single observation, while Cook's D measures the change in the parameter estimates caused by deleting each observation (SAS/Insight Users Guide, 1999). Any observation with a DFFITS ≥ 2 or a Cook's D ≥ 1 would be labeled as influencing the fit; however, this data set had no occurrences of either. After running all influence diagnostics, it was determined that only observation 105, Ukraine 1992, needs to be removed.

The next test was to check for multicollinearity between the variables (Table B.8). This measures the linear dependency among the independent variables as well as testing for a high correlation between the variables. This is done by analyzing the Variance Inflation Factors (VIF) to see if multicollinearity exists. A measure of 10 or higher reflects a severe problem, while a VIF > 4 could be a possible problem. If all VIFs are lower than 4 the process can stop because there is no evidence of multicollinearity. If it does exist, the eigenvalues and condition indexes are analyzed to determine how many dependencies there are. The number of eigenvalues 'near 0' estimate the number of serious dependencies. Using SAS, the condition index should not be greater than 30 (SAS outputs the square root of the condition index), otherwise, this means there is/are serious dependencies. Finally, proportions of variance can be checked for large values (close to 1) to see which variables are involved in the collinearity. Once this is determined, the affected variable(s) can be removed to see if this improves the model. In the tests run on the chosen model there were no VIFs greater than 10 and only 2 VIFs between 4 and 10 (5.75724, 5.47167). This shows that there might be multicollinearity. Continuing to the next step of checking eigenvalues, the smallest eigenvalue is 0.09951 with the square root of the condition index being 5.83909. This leaves no reason to examine the proportions of variation. The conclusion can be drawn that multicollinearity is not affecting this model and therefore no variables need to be removed.

The final model has the same variables as before; however, the analysis of variance results have improved as a result of removing the high influence point of Ukraine in 1992.

Table 6 - Analysis of Variance - Final Model

Source	**DF**	**Sum of Sq**	**Mean Sq**	**F value**	**Pr > F**
Model	8	13.56680	1.6958	27.09	<.0001
Error	110	6.88587	0.06260		
Corr. Total	118	20.45267			

Root MSE	0.25020	R-square	0.6633
Dependent Mean	0.69017	Adj R-sq	0.6388
Coeff Car	36.25134		

Parameter Estimates

Variable	**DF**	**Par. Est.**	**St. Error**	**t value**	**Pr > |t|**
Intercept	1	18.76518	2.96417	6.33	<.0001
XTIMR	1	-0.30837	0.04171	-7.39	<.0001
XTLIFE	1	1.526533e66	2.198095e65	6.94	<.0001
XTYOUTH	1	-14.06101	2.55008	-5.51	<.0001
PolRights	1	0.08061	0.02984	2.70	0.0080
CivilLib	1	-0.07855	0.03586	-2.19	0.0306
XTGDP	1	-0.03972	0.02045	-1.94	0.0547
BigEthnic	1	-0.01728	0.00198	-8.75	<.0001
XTOPEN	1	-0.09715	0.02387	-34.07	<.0001

Notice, a higher R- and adjusted R-squared of 0.6633 and 0.6388 showing that 66.33% of variation is explained by this model, a significant difference from the original (Table B.9). This results in a model of:

YTRANS = 18.76518 - 0.30837XTIMR + 1.526533e66XTLIFE

- 14.06101XTYOUTH + 0.08061PolRights - 0.07855CivilLib - 0.03972XTGDP

- 0.01728BigEthnic - 0.09715XTOPEN

3.4 Validation of the Final Model

The next step in evaluating this model is cross-validation. This consists of splitting the data set randomly into two parts and deleting the response variable points of one part. A prediction is then run with both parts and SAS gives a predicted value for the missing points. Since the response variable is transformed, the data that was removed (MaxKOSIMO level) needed to be transformed in order to compare it to the predicted values generated by SAS. A correlation test was then run against the original versus the predicted values to determine if correlation exists, and if cross-validation was successful (Table B.10). The statistic of r(YTRANS, YPRED) will be used to test Ho: $\rho = 0$ versus Ha: $\rho > 0$ where ρ is the population correlation coefficient. The test of the hypotheses is a t-test where

$$t = r(YTRANS, YPRED)\sqrt{\frac{n-2}{1-r(YTRANS, YPRED)^2}} \sim t_n\text{-}2$$

with n being the number of observations in the "deleted" part of the data set. In this case, t = 59. If the null hypothesis (Ho) is rejected, then the cross validation is successful and the model is useful.

For this model the cross-validation output is:

Table 7 - Cross Validation

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
YTRANS	59	0.64065	0.42585	37.79850	0.06930	1.00000
YPRED	59	0.62205	0.35180	36.70100	-0.02990	1.38290

Pearson Correlation Coefficients, N = 59
Prob > |r| under H0: Rho=0

	YTRANS	YPRED
YTRANS	1.00000	0.77563 <.0001
YPRED	0.77563	1.00000 <.0001

From this output, r (YTRANS, YPRED) = 0.77563, so

$$t = (.77563)\sqrt{\frac{59-2}{1-0.77563^2}} = 9.3585$$

The p-value is then: $p = P(t \geq 9.3585)$; SAS gives us this value as a two-tailed p-value. Dividing this by 2 will give us an accurate value. For this model, the p-value was <.0001, and <.001/2 = <.00005, which is small enough to be successful. This means a rejection of null hypothesis and approval of this model. We can now conclude that the model has been validated as useful for prediction.

4. Results

In either a regular or transformed state, the variables that proved to be significant in judging or predicting stability were infant mortality rate (transformed), life expectancy (transformed), youth bulge (transformed), Political Rights Index, Civil Liberties Index, Gross Domestic Product (transformed), Ethnic Diversity, and Trade Openness (transformed). The response variable is a transformation on the levels 1 through 4 of the KOSIMO database.

For the analysis of stability, we are going to interpret the response variable as a continuous scale of 1-4 with 1 being stable and 4 being unstable. For reference, in 2005, Iraq was the only country, worldwide to receive a 4 rating (Heidelberg Institute of International Conflict Research (HIIK), 2006). Table 8 shows the breakdown of analysis for how the absolute value of the YTRANS response correlates to the "stable" versus "unstable" result. The stability levels shown correspond to the KOSIMO rating levels.

Table 8 - YTRANS (-1.925)

Stability Level	Response	Absolute Value	Result
1	1	≥ 1	Stable
2	0.263340259	≥ 0.26	Stable
3	0.120653961	< 0.26	Unstable
4	0.069348092	< 0.06	Unstable

With our model minimizing all of our variables except life expectancy (transformed) and Political Rights will result in the most stability. Interpretation of most of the variables seems to be fairly logical, i.e. decreasing the infant mortality rate.

In order to determine the accuracy of this model in predicting future stability, I have collected data for the countries of the FSU in 2005. Using our model and entering in the pertinent values we can see how closely our prediction is to the actual 2005 rating for that specific country. With our results we will round decimals up to the next whole number/level for interpretation purposes. When this is either 1 or 2 we will label the country as "stable" with the acknowledgement that 1 is "maximum stability" and 2 is "mostly stable." Any value above a 2 will be labeled as "unstable."

For example, our model is:

YTRANS = 18.76518 - 0.30837XTIMR + 1.526533e66XTLIFE - 14.06101XTYOUTH + 0.08061PolRights - 0.07855CivilLib - 0.03972XTGDP - 0.01728BigEthnic - 0.09715XTOPEN

For the country of Kazakhstan in 2005[2] we will use:

XTIMR: $29^{.51} = 5.6595$

XTLIFE: $66.6^{-37} = 3.3998\text{E-}68$

XTYOUTH: $.8599^{-.1} = 1.0152$

PolRights: 6

CivilLib: 5

XTGDP: $1520^{0.33} = 11.2204$

BigEthnic: 53

XTOPEN: $1.0944^{-1.6} = 0.8656$

This gives us a model prediction of 0.50276 for Kazakhstan, which once the response is transformed with a lambda of -1.925 is a 2, mostly stable. The actual KOSIMO level for

[2] We have to use the 1999 data for openness at this time, all other data points are from 2005.

Kazakhstan in 2005 is 2, mostly stable. We can do the same thing for every country in the FSU as shown in Table 9, for prediction data on each country see Table A.3.

Table 9 – Level Results

	Armenia	Azerbaijan	Belarus	Estonia	Georgia	Kazakhstan
Model Prediction	0.097043	-1.001873	-0.00035	1.27991	0.452045	0.50276074
|Rounded|	0.09	1	0.0003	1.2	0.45	0.5
Transformed Result	3	1	4	1	2	2
KOSIMO 2005	2	2	2	1	2	2

	Kyrgyzstan	Latvia	Lithuania	Moldova	Russia	Tajikistan
Model Prediction	0.35416848	1.165864	0.810825	0.368696	-0.210	-0.625225
|Rounded|	0.4	1	0.8	0.35	0.21	0.62
Transformed Result	2	1	2	2	3	2
KOSIMO 2005	2	1	1	2	3	2

	Turkmenistan	Ukraine	Uzbekistan
Model Prediction	-0.2550771163	0.679291	-.05887058
|Rounded|	0.25	0.68	0.58
Transformed Result	3	2	2
KOSIMO 2005	1	2	2

However, we are looking to predict either stable or unstable, as we have established that anything higher than a 2 is unstable, and this may be for differing reasons. Therefore, we can label each of our predictions and determine accuracy from that. In Table 10, each country is given a label of "stable" for values of 1 or 2 and "unstable" for anything higher than 2. The KOSIMO levels are translated in the same way, 1 and 2 are "stable" and 3 and 4 are labeled as "unstable." The accuracy is shown in the table as a "1" for correct and a "0" for an incorrect prediction of stability. Following this methodology we can determine that our model is 80% accurate. With an accurate reading for trade openness for 2005 these accuracies may increase.

Table 10 – Stability Results

	Armenia	**Azerbaijan**	**Belarus**	**Estonia**	**Georgia**	**Kazakhstan**	**Kyrgyzstan**
Result	Unstable	Stable	Unstable	Stable	Stable	Stable	Stable
KOSIMO	Stable	Stable	Stable	Stable	Stable	Stable	Stable
Correct	0	1	0	1	1	1	1

	Latvia	**Lithuania**	**Moldova**	**Russia**	**Tajikistan**	**Turkmenistan**	**Ukraine**	**Uzbekistan**
Result	Stable	Stable	Stable	Unstable	Stable	Unstable	Stable	Stable
KOSIMO	Stable	Stable	Stable	Unstable	Stable	Stable	Stable	Stable
Correct	1	1	1	1	1	0	1	1
Accuracy	**0.8**			**80% Accuracy**				

The three countries out of fifteen that were incorrect in our prediction were Armenia, Belarus and Turkmenistan. These countries may be anomalies for various reasons, and therefore, would not fit into our model. Or, it may be possible that there is another variable that could be included in this model that was not examined which would explain the stability in those three countries. On a positive note, we have overestimated all three cases as opposed to predicting stability when the country may in fact be unstable.

5. Conclusions

Using multiple linear regression to generate a short-term forecast of the stability levels for the countries of the former Soviet Union (FSU), I developed the following model:

YTRANS = 18.76518 - 0.30837XTIMR + 1.526533e66XTLIFE - 14.06101XTYOUTH + 0.08061PolRights - 0.07855CivilLib - 0.03972XTGDP - 0.01728BigEthnic - 0.09715XTOPEN.

This shows that in either a regular or transformed state, infant mortality rate (transformed), life expectancy (transformed), youth bulge (transformed), Political Rights Index, Civil Liberties Index, Gross Domestic Product (transformed), Ethnic Diversity, and Trade Openness (transformed) are all relevant factors in predicting the future stability of a country in the FSU. The response variable is a transformation on the levels 1 through 4 of the KOSIMO database. Using this model to perform a test of a five-year projection using 2005 data, I determined that this model is at least 80% accurate for the short-term.

This model covers many of the main factors affecting stability; however, there may be other variables not considered in the beginning of this project that have a significant effect on stability. Also, as stated in O'Brien (2002) there are the "other" factors that are difficult or impossible to measure or collect (water scarcity, deforestation…).

The general methodology of this project can be useful in predicting stability for other areas of the world that do not necessarily fit into a "global" model such as parts of Africa, southeast Asia, or possibly to compare and contrast developed versus developing world countries. Data collected from 2000-2005 could be added to this current data set, and using the same methodology, a model could be derived that is capable of forecasting farther than five years. As stated previously, the limitation on our timeline stemmed from the small data set.

Forecasting stability and instability allows policymakers to focus resources on vulnerable states and develop sound operational plans, both diplomatic and military, before conflict erupts. Now that we can identify certain factors that increase stability, these issues can be a focus for policymakers in transitioning countries or in aiding a developing nation.

Appendix A - Tables and Figures

Table A.1 - KOSIMO Data Projects (www.hiik.de)

Standard deviations are in parentheses

		Total Conflict		
Conflict Type	*External*	*Internal*	*Total*	*% External*
Crisis	31	43	74	42%
Violent Crisis	30	91	121	25%
War	17	44	61	28%
Total	78	178	256	30%
		Total Country-Year Conflicts		
Conflict Type	*External*	*Internal*	*Total*	*% External*
Crisis	27	283	310	9%
Violent Crisis	49	494	543	9%
War	47	384	431	11%
Total	123	1161	1284	10%
		Average Conflict Duration (in years)		
Conflict Type		*External*		*Internal*
Crisis		5.26		5.75
		(4.67)		(4.26)
Violent Crisis		2.58		4.98
		(4.15)		(4.64)
War		3.5		5
		(5.15)		(5.30)

Table A.2 - Independent Variables (O'Brien, 2002)

% of history spent in state of conflict: % of time (in years) spent in a state of conflict as defined by KOSIMO *Source: KOSIMO data project*

Infant Mortality Rate: Number of deaths of children under 1 year of age per 1,000 live births *Source: U.S. Census Bureau, International database (www.census/gov.ipc.www)*

Trade Openness: value of a country's total imports and exports as a % of GDP per capita *Source: PENN World Tables (1975-1992) datacentre2.chass.utoronto.ca/pwt; 1999 World Bank Development Indicators (1993-1998)*

Youth Bulge: Ratio of population aged 15-29 to those aged 30-54 *Source: U.S. Census Bureau, International database (www.census/gov.ipc.www)*

Civil Liberties Index: Measure of the freedom of country's people "to develop views, institution, and personal autonomy apart from the state." Seven point scale 1=free, 7=not free. *Source: Freedom House (www.freedomhouse.org)*

Life Expectancy: Average life expectancy (males and females combined) *Source: U.S. Census Bureau, International database (www.census/gov.ipc.www)*

Political Rights Index: Measure of rights to participate meaningfully in the political process (using the same scale as the civil liberties index) *Source: Freedom House (www.freedomhouse.org)*

Democracy: Measure of degree of democracy: ranges from -10 (least democratic) to 10 (most democratic) *Source: Polity98 project (kleditsch.socsci.gla.ac.uk/Polity.html)*

Religious Diversity: Largest religious group in a country as a % of total population *Source: CIA World Fact Book; Country Indicators of Foreign Policy Project (CIFP); Elligensen 1996; Handbook of the Nations; Britannica Book of the Year; Demographic Yearbook*

Caloric Intake: Estimates of the average number of calories consumed per person, per day. *Source: Food and Agriculture Organization of the United Nations (http://apps.fao.org)*(Tafi, Jana. Poverty and Welfare over the 1990's in the Republic of Moldova: Indicators of Status and Trends. www.unicef-icdc.org/research/ESP/CountryReports2000_01/Moldova00.pdf)

GDP per capita: Annual gross domestic product per person measured in constant 1998 dollars *Source: World Bank (1999); World Development Indicators*

Ethnic Diversity: Largest ethnic group in a country as a % of total population *Source: same sources used to measure* ***religious diversity*** *above.*

Table A.3 2005 Data Analysis

	Armenia	Azerbaijan	Belarus	Estonia
XTIMR	4.948586846	9.463353531	3.699228085	2.887858391
XTLIFE	2.33463E-69	2.10231E-67	1.07796E-68	2.1057E-69
XTYOUTH	1.01819833	1.020380639	1.045112758	1.042419919
PolRights	5	6	7	1
CivilLib	4	5	6	6
XTGDP	9.041009803	8.7280104	10.81602892	15.67944289
BigEthnic	93	90	81.2	65
XTOPEN (99)	1.496744024	0.451835031	0.798730438	0.36358722

	Georgia	Kazakhstan	Kyrgyzstan	Latvia
XTIMR	4.489152268	5.569586829	6.218910052	3.235936569
XTLIFE	2.69808E-70	3.3998E-68	1.141247E-68	3.18743E-69
XTYOUTH	1.045287813	1.015204973	0.994993122	1.047153528
PolRights	3	6	6	1
CivilLib	4	5	5	2
XTGDP	8.477375252	11.22040115	6.495182216	14.74759324
BigEthnic	70	53	64.9	57
XTOPEN (99)	1.328587781	0.865602027	2.762420112	0.760011759

	<u>Lithuania</u>	<u>Moldova</u>	<u>Russia</u>	<u>Tajikistan</u>
XTIMR	2.697739435	6.562217121	3.979298719	11.04370302
XTLIFE	6.89329E-70	7.46172E-68	2.5779E-68	1.0505E-67
XTYOUTH	1.04885248	1.027383882	1.046049074	0.980184364
PolRights	2	3	6	6
CivilLib	2	4	5	5
XTGDP	15.0085868	7.563278359	12.54195589	5.549322015
BigEthnic	81	64.5	82	65
XTOPEN (99)	0.632414679	0.717858494	2.964744452	1.47

	Turkmenistan	**Ukraine**	**Uzbekistan**
XTIMR	8.915885363	3.235936569	8.79309485
XTLIFE	6.88288E-67	6.3152E-69	1.32192E-67
XTYOUTH	0.995659386	1.045277047	0.993167501
PolRights	8	4	7
CivilLib	7	3	6
XTGDP	10.37838581	9.003082273	6.639713787
BigEthnic	85	78	80
XTOPEN (99)	2.477547813	1.404837271	2.033108048

Appendix B – SAS Output

B.1 Original Model ANOVA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	70.85405	6.44128	9.45	<.0001
Error	108	73.64595	0.68191		
Corrected Total	119	144.50000			

Root MSE	0.82578	R-Square	0.4903
Dependent Mean	1.75000	Adj R-Sq	0.4384
Coeff Var	47.18723		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-16.13950	4.15648	-3.88	0.0002
Calories	1	-0.00035999	0.00029482	-1.22	0.2247
IMR	1	0.06641	0.01066	6.23	<.0001
LifeExp	1	0.24978	0.06448	3.87	0.0002
YouthBulge	1	-4.49468	1.00968	-4.45	<.0001
PolRights	1	-0.27984	0.12998	-2.15	0.0335
CivilLib	1	0.45511	0.11934	3.81	0.0002
GDP	1	0.00043009	0.00013321	3.23	0.0016
Big Religion	1	0.00302	0.00402	0.75	0.4541
BigEthnic	1	0.02904	0.00799	3.63	0.0004
Democracy	1	0.03889	0.03491	1.11	0.2678
openness	1	-0.73474	0.22848	-3.22	0.0017

B.2 Stepwise and Forward Selected Model ANOVA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	37.78521	12.59507	13.69	<.0001
Error	116	106.71479	0.91996		
Corrected Total	119	144.50000			

Root MSE	0.95914	R-Square	0.2615
Dependent Mean	1.75000	Adj R-Sq	0.2424
Coeff Var	54.80817		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.05588	0.86943	2.36	0.0197
Calories	1	-0.00073424	0.00022633	-3.24	0.0015
BigEthnic	1	0.02952	0.00678	4.35	<.0001
openness	1	-0.45644	0.21548	-2.12	0.0363

B.3 Backwards Selection Model ANOVA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	70.46910	7.04691	10.38	<.0001
Error	109	74.03090	0.67918		
Corrected Total	119	144.50000			

Root MSE	0.82413	R-Square	0.4877
Dependent Mean	1.75000	Adj R-Sq	0.4407
Coeff Var	47.09288		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-15.87241	4.13297	-3.84	0.0002
Calories	1	-0.00036738	0.00029407	-1.25	0.2142
IMR	1	0.06644	0.01064	6.25	<.0001
LifeExp	1	0.24429	0.06394	3.82	0.0002
YouthBulge	1	-4.41701	1.00236	-4.41	<.0001
PolRights	1	-0.27067	0.12915	-2.10	0.0384
CivilLib	1	0.46362	0.11857	3.91	0.0002
GDP	1	0.00042077	0.00013236	3.23	0.0019
BigEthnic	1	0.03180	0.00708	4.49	<.0001
Democracy	1	0.04364	0.03427	1.27	0.2056
openness	1	-0.73141	0.22798	-3.21	0.0018

B.4 Box-Tidwell Transformation ANOVA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	82.65336	8.26534	14.57	<.0001
Error	109	61.84664	0.56740		
Corrected Total	119	144.50000			

Root MSE	0.75326	R-Square	0.5720
Dependent Mean	1.75000	Adj R-Sq	0.5327
Coeff Var	43.04342		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-37.11540	8.55676	-4.34	<.0001
XTCAL	1	-6.7239E-54	4.88128E-54	-1.38	0.1712
XTIMR	1	0.76873	0.12538	6.13	<.0001
XTLIFE	1	-3.93293E66	6.674649E65	-5.89	<.0001
XTYOUTH	1	26.55009	7.31043	3.63	0.0004
PolRights	1	-0.21297	0.11102	-1.92	0.0577
CivilLib	1	0.42131	0.11065	3.81	0.0002
XTGDP	1	0.23859	0.06892	3.46	0.0008
BigEthnic	1	0.04408	0.00638	6.91	<.0001
Democracy	1	0.05431	0.02999	1.81	0.0729
XTOPEN	1	0.22130	0.07369	3.00	0.0033

B.5 Box-Cox Transformation with Transformed Independent Variables ANOVA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	13.51807	1.35181	20.71	<.0001
Error	109	7.11527	0.06528		
Corrected Total	119	20.63334			

Root MSE	0.25549	R-Square	0.6552
Dependent Mean	0.68662	Adj R-Sq	0.6235
Coeff Var	37.21067		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	17.14101	2.90233	5.91	<.0001
XTCAL	1	1.87774E-54	1.65566E-54	1.13	0.2592
XTIMR	1	-0.28985	0.04253	-6.82	<.0001
XTLIFE	1	1.494051E66	2.263945E65	6.60	<.0001
XTYOUTH	1	-12.46873	2.47959	-5.03	<.0001
PolRights	1	0.06488	0.03766	1.72	0.0877
CivilLib	1	-0.08518	0.03753	-2.27	0.0252
XTGDP	1	-0.04572	0.02338	-1.96	0.0531
BigEthnic	1	-0.01720	0.00216	-7.95	<.0001
Democracy	1	-0.00549	0.01017	-0.54	0.5905
XTOPEN	1	-0.07952	0.02499	-3.18	0.0019

B.6 Backwards and Stepwise Selected Model ANOVA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	13.40260	1.67533	25.72	<.0001
Error	111	7.23074	0.06514		
Corrected Total	119	20.63334			

Root MSE	0.25523	R-Square	0.6496
Dependent Mean	0.68662	Adj R-Sq	0.6243
Coeff Var	37.17191		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	16.14804	2.80169	5.76	<.0001
XTIMR	1	-0.28249	0.04103	-6.88	<.0001
XTLIFE	1	1.477215E66	2.232029E65	6.62	<.0001
XTYOUTH	1	-11.74774	2.39922	-4.90	<.0001
PolRights	1	0.08475	0.03039	2.79	0.0062
CivilLib	1	-0.07947	0.03658	-2.17	0.0319
XTGDP	1	-0.03552	0.02078	-1.71	0.0902
BigEthnic	1	-0.01713	0.00201	-8.50	<.0001
XTOPEN	1	-0.08549	0.02382	-3.59	0.0005

B.7 Influence Diagnostics

```
                              Output Statistics

                          Cook's          Hat Diag     Cov           -DFBETAS-
Obs   -2-1 0 1 2               D  RStudent       H    Ratio   DFFITS Intercept

  1 |    ***|      |       0.025   -1.5462  0.0857   0.9778  -0.4733   -0.0888
  2 |    ***|      |       0.018   -1.5312  0.0661   0.9608  -0.4074   -0.1968
  3 |     **|      |       0.008   -1.1406  0.0538   1.0314  -0.2720   -0.0495
  4 |       |      |       0.000   -0.0481  0.0425   1.1328  -0.0101   -0.0009
  5 |       |      |       0.000   -0.2375  0.0617   1.1508  -0.0609   -0.0023
  6 |       |      |       0.000   -0.2553  0.0607   1.1488  -0.0649   -0.0003
  7 |       |      |       0.000    0.2992  0.0423   1.1245   0.0628   -0.0075
  8 |       |      |       0.001    0.3256  0.0427   1.1235   0.0688   -0.0121
  9 |       |      |       0.000   -0.1130  0.0740   1.1703  -0.0319   -0.0112
 10 |       |      |       0.000   -0.2975  0.0399   1.1218  -0.0607   -0.0032
 11 |       |      |       0.001   -0.3683  0.0356   1.1125  -0.0708    0.0115
 12 |       |*     |       0.004    0.7981  0.0519   1.0863   0.1867   -0.0777
 13 |       |      |       0.002    0.4454  0.0689   1.1464   0.1212   -0.0450
 14 |       |      |       0.000    0.0919  0.1310   1.2476   0.0357   -0.0107
 15 |       |      |       0.001    0.1730  0.1427   1.2623   0.0706   -0.0273
 16 |       |      |       0.001    0.2469  0.1589   1.2834   0.1073   -0.0492
 17 |       |      |       0.001   -0.3248  0.0723   1.1593  -0.0907   -0.0203
 18 |       |      |       0.001   -0.3027  0.0802   1.1706  -0.0894   -0.0161
 19 |       |      |       0.000    0.0995  0.0447   1.1347   0.0215    0.0042
 20 |       |      |       0.000    0.0919  0.0685   1.1638   0.0249    0.0043
 21 |       |      |       0.000    0.1381  0.1044   1.2094   0.0471   -0.0017
 22 |       |      |       0.000    0.1435  0.1059   1.2113   0.0494   -0.0047
 23 |       |      |       0.001    0.2242  0.1057   1.2081   0.0771   -0.0097
 24 |       |      |       0.001    0.2746  0.1043   1.2038   0.0937   -0.0113
 25 |       |      |       0.000   -0.2070  0.0465   1.1338  -0.0457    0.0104
 26 |      *|      |       0.003   -0.6703  0.0634   1.1165  -0.1743   -0.0003
 27 |      *|      |       0.007   -0.9801  0.0656   1.0736  -0.2597   -0.0257
 28 |       |      |       0.001   -0.4244  0.0492   1.1243  -0.0966    0.0020
 29 |       |      |       0.000   -0.1435  0.0635   1.1565  -0.0374   -0.0011
 30 |       |      |       0.000   -0.1808  0.0799   1.1759  -0.0533    0.0003
 31 |       |      |       0.000   -0.1194  0.0899   1.1906  -0.0375    0.0010
 32 |       |      |       0.000   -0.0901  0.0886   1.1895  -0.0281    0.0010
 33 |       |      |       0.011   -0.4737  0.2994   1.5203  -0.3096   -0.1318
 34 |     **|      |       0.024   -1.2094  0.1287   1.1055  -0.4647   -0.1659
 35 |     **|      |       0.013   -1.2624  0.0691   1.0239  -0.3439   -0.0457
 36 |     **|      |       0.006   -1.0277  0.0477   1.0453  -0.2299    0.0689
 37 |     **|      |       0.006   -1.2268  0.0347   0.9945  -0.2325    0.0705
 38 |     **|      |       0.006   -1.0376  0.0462   1.0420  -0.2284    0.1122
 39 |     **|      |       0.006   -1.0158  0.0515   1.0516  -0.2367    0.1313
 40 |      *|      |       0.005   -0.8439  0.0549   1.0831  -0.2034    0.1219
 41 |       |      |       0.002    0.4315  0.0759   1.1563   0.1237   -0.0471
 42 |       |      |       0.000   -0.0325  0.0927   1.1956  -0.0104    0.0018
 43 |       |      |       0.000   -0.0276  0.0692   1.1654  -0.0075    0.0026
 44 |       |      |       0.000    0.2619  0.0552   1.1419   0.0633   -0.0185
 45 |       |      |       0.000    0.0521  0.0575   1.1508   0.0129   -0.0047
 46 |       |      |       0.000   -0.1266  0.0623   1.1555  -0.0326    0.0135
 47 |       |      |       0.001   -0.2856  0.0643   1.1518  -0.0749    0.0303
 48 |       |      |       0.001   -0.3507  0.0711   1.1562  -0.0970    0.0434
 49 |       |      |       0.000    0.0967  0.1636   1.2960   0.0427    0.0228
 50 |       |      |       0.000   -0.1328  0.1042   1.2093  -0.0453   -0.0217
```

```
 51 |      |      |  0.001   0.2712  0.0796  1.1716   0.0798   0.0417
 52 |      |*     |  0.003   0.7260  0.0561  1.1010   0.1770   0.0885
 53 |      |*     |  0.004   0.7301  0.0562  1.1005   0.1781   0.0807
 54 |      |*     |  0.006   0.8633  0.0636  1.0903   0.2250   0.0951
 55 |      |*     |  0.004   0.8626  0.0499  1.0746   0.1977   0.0731
 56 |      |**    |  0.007   1.0098  0.0593  1.0614   0.2536   0.0904
 57 |      |      |  0.001  -0.4732  0.0392  1.1087  -0.0956   0.0123
 58 |     *|      |  0.004  -0.8308  0.0527  1.0826  -0.1960  -0.0464
 59 |      |      |  0.000   0.0626  0.1316  1.2489   0.0244   0.0083
 60 |     *|      |  0.003  -0.6795  0.0487  1.0982  -0.1537  -0.0213
 61 |     *|      |  0.003  -0.7742  0.0478  1.0849  -0.1735  -0.0160
 62 |      |      |  0.001  -0.4244  0.0513  1.1268  -0.0987  -0.0014
 63 |      |      |  0.000  -0.2505  0.0450  1.1301  -0.0544  -0.0024
 64 |      |      |  0.000  -0.2511  0.0447  1.1298  -0.0543  -0.0050
 65 |      |**    |  0.005   1.0884  0.0391  1.0252   0.2196   0.0781
 66 |      |*     |  0.007   0.9695  0.0647  1.0744   0.2550   0.0980
 67 |      |*     |  0.007   0.9379  0.0637  1.0785   0.2446   0.0964
 68 |      |*     |  0.003   0.6562  0.0557  1.1091   0.1593   0.0810
 69 |      |*     |  0.003   0.6660  0.0528  1.1045   0.1572   0.0685
 70 |      |*     |  0.003   0.6952  0.0512  1.0992   0.1615   0.0579
 71 |      |*     |  0.004   0.8088  0.0522  1.0851   0.1898   0.0590
 72 |      |*     |  0.004   0.8426  0.0512  1.0791   0.1958   0.0691
 73 |      |**    |  0.005   1.0514  0.0408  1.0337   0.2169  -0.0840
 74 |      |***   |  0.010   1.6041  0.0355  0.9135   0.3078  -0.1205
 75 |      |**    |  0.008   1.2530  0.0416  0.9964   0.2611  -0.0931
 76 |      |**    |  0.008   1.2003  0.0463  1.0118   0.2646  -0.1197
 77 |      |**    |  0.013   1.3683  0.0588  0.9902   0.3419  -0.1443
 78 |      |**    |  0.012   1.2834  0.0599  1.0095   0.3240  -0.1364
 79 |      |***   |  0.025   1.5093  0.0896  0.9909   0.4735  -0.1194
 80 |      |***   |  0.024   1.5728  0.0823  0.9677   0.4709  -0.0764
 81 |      |*     |  0.006   0.5184  0.1732  1.2836   0.2373  -0.0800
 82 |      |      |  0.000   0.0625  0.1375  1.2574   0.0249  -0.0105
 83 |    **|      |  0.012  -1.1090  0.0825  1.0699  -0.3327   0.1157
 84 |     *|      |  0.008  -0.9743  0.0695  1.0792  -0.2664   0.1473
 85 |    **|      |  0.010  -1.2475  0.0551  1.0118  -0.3014   0.1688
 86 |    **|      |  0.011  -1.3028  0.0536  0.9987  -0.3100   0.1726
 87 |   ***|      |  0.012  -1.6678  0.0384  0.9011  -0.3333   0.1555
 88 |    **|      |  0.009  -1.0683  0.0661  1.0586  -0.2843   0.1200
 89 |  ****|      |  0.045  -2.4593  0.0650  0.7167  -0.6485  -0.3338
 90 |   ***|      |  0.028  -1.8865  0.0685  0.8745  -0.5117  -0.1151
 91 |   ***|      |  0.019  -1.5697  0.0657  0.9512  -0.4162  -0.0433
 92 |    **|      |  0.012  -1.2263  0.0668  1.0288  -0.3281   0.0298
 93 |    **|      |  0.010  -1.0839  0.0686  1.0586  -0.2943   0.0495
 94 |     *|      |  0.007  -0.9373  0.0635  1.0784  -0.2441   0.0417
 95 |     *|      |  0.005  -0.7879  0.0697  1.1085  -0.2156   0.0589
 96 |     *|      |  0.004  -0.6557  0.0771  1.1349  -0.1895   0.0677
 97 |      |**    |  0.011   1.0847  0.0779  1.0691   0.3154   0.0094
 98 |      |**    |  0.016   1.3512  0.0717  1.0078   0.3757   0.0337
 99 |      |*     |  0.004   0.7170  0.0599  1.1066   0.1810   0.0244
100 |      |      |  0.000   0.0593  0.0793  1.1780   0.0174   0.0032
101 |      |      |  0.000  -0.0939  0.1128  1.2219  -0.0335  -0.0067
102 |     *|      |  0.007  -0.5270  0.1736  1.2834  -0.2416  -0.0591
103 |     *|      |  0.023  -0.8439  0.2276  1.3254  -0.4581  -0.0979
104 |   ***|      |  0.106  -1.5307  0.2916  1.2668  -0.9821  -0.1225
105 |      |****  |  0.225   2.3472  0.2763  0.9656   1.4503  -0.9529
106 |    **|      |  0.013  -1.4317  0.0529  0.9701  -0.3383  -0.0938
107 |   ***|      |  0.017  -2.0221  0.0377  0.8117  -0.4001  -0.0374
```

108		|* |	0.003	0.7820	0.0479	1.0840	0.1754	0.0343
109		|* |	0.004	0.7436	0.0537	1.0958	0.1771	0.0304
110		|* |	0.003	0.7180	0.0561	1.1020	0.1751	0.0318
111		|* |	0.004	0.7885	0.0563	1.0927	0.1926	0.0417
112		|* |	0.005	0.8919	0.0560	1.0770	0.2172	0.0590
113	****|	|	0.041	-2.5314	0.0575	0.6915	-0.6255	-0.2821
114		|** |	0.011	1.1711	0.0671	1.0401	0.3140	0.0775
115		|*** |	0.019	1.6871	0.0585	0.9156	0.4205	0.1542
116		|*** |	0.016	1.6631	0.0514	0.9145	0.3870	0.1274
117		|*** |	0.016	1.7917	0.0430	0.8750	0.3798	0.1104
118		|*** |	0.014	1.7434	0.0402	0.8844	0.3568	0.1014
119		|*** |	0.018	1.9900	0.0394	0.8215	0.4031	0.1161
120		|**** |	0.021	2.1571	0.0403	0.7786	0.4419	0.1112

B.8 Multicollinearity

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	10.14483	1.44926	15.61	<.0001
Error	111	10.30784	0.09286		
Corrected Total	118	20.45267			

Root MSE	0.30473	R-Square	0.4960
Dependent Mean	0.69017	Adj R-Sq	0.4642
Coeff Var	44.15332		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	0.46764	1.98712	0.24	0.8144	0
XTLIFE	1	9.763046E65	2.519122E65	3.88	0.0002	1.46543
XTYOUTH	1	0.38250	1.99639	0.19	0.8484	3.42297
PolRights	1	0.01847	0.03488	0.53	0.5975	5.75724
CivilLib	1	-0.01921	0.04257	-0.45	0.6527	5.47167
XTGDP	1	0.07342	0.01652	4.44	<.0001	2.01233
BigEthnic	1	-0.01232	0.00226	-5.44	<.0001	1.14709
XTOPEN	1	-0.13453	0.02841	-4.73	<.0001	1.12342

Collinearity Diagnostics (intercept adjusted)

Number	Eigenvalue	Condition Index	Proportion of Variation XTLIFE	Proportion of Variation XTYOUTH	Proportion of Variation PolRights
1	3.39273	1.00000	0.02343	0.01924	0.01259
2	1.12556	1.73616	0.05931	0.01808	0.00009581
3	1.03056	1.81442	0.03847	0.00201	0.00027745
4	0.65405	2.27756	0.54775	0.00134	0.00002659
5	0.47935	2.66041	0.25808	0.01211	0.08370
6	0.21824	3.94281	0.06105	0.87178	0.00473
7	0.09951	5.83909	0.01189	0.07545	0.89858

Collinearity Diagnostics (intercept adjusted)

Number	Proportion of Variation CivilLib	Proportion of Variation XTGDP	Proportion of Variation BigEthnic	Proportion of Variation XTOPEN
1	0.01322	0.02313	0.00053293	0.00229
2	0.00052387	0.02236	0.10422	0.50888
3	0.00035701	0.03350	0.66712	0.06310
4	0.00001087	0.14622	0.03813	0.34135
5	0.07495	0.37786	0.00123	0.01133
6	0.08734	0.35812	0.17551	0.07253
7	0.82360	0.03882	0.01325	0.00052705

B.9 Final Model ANOVA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	13.56680	1.69585	27.09	<.0001
Error	110	6.88587	0.06260		
Corrected Total	118	20.45267			

Root MSE	0.25020	R-Square	0.6633
Dependent Mean	0.69017	Adj R-Sq	0.6388
Coeff Var	36.25134		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	18.76518	2.96417	6.33	<.0001
XTIMR	1	-0.30837	0.04171	-7.39	<.0001
XTLIFE	1	1.526533E66	2.198095E65	6.94	<.0001
XTYOUTH	1	-14.06101	2.55008	-5.51	<.0001
PolRights	1	0.08061	0.02984	2.70	0.0080
CivilLib	1	-0.07855	0.03586	-2.19	0.0306
XTGDP	1	-0.03972	0.02045	-1.94	0.0547
BigEthnic	1	-0.01728	0.00198	-8.75	<.0001
XTOPEN	1	-0.09715	0.02387	-4.07	<.0001

B.10 Cross- Validation

The CORR Procedure
2 Variables: YTRANS YPRED

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
YTRANS	59	0.64065	0.42585	37.79850	0.06930	1.00000
YPRED	59	0.62205	0.35180	36.70100	-0.02990	1.38290

Pearson Correlation Coefficients, N = 59
Prob > |r| under H0: Rho=0

	YTRANS	YPRED
YTRANS	1.00000	0.77563 <.0001
YPRED	0.77563 <.0001	1.00000

References

1. Armstrong, J. Scott, 1984. Forecasting by Extrapolation: Conclusions from 25 Years of Research. Interfaces. 14: 52-66.

2. Box, G.E.P. and D.R. Cox. 1964. An Analysis of Transformations (with discussion). Journal of the Royal Statistical Society. Series B 26: 211-246.

3. Box, G.E.P. and P.W. Tidwell. 1962. Transformations of the Independent Variables. Technometrics. 4: 531-550.

4. Chen, Yuan yan, 1995. Statistical Inference based on the possibility and belief measures. Transactions of the American Mathematical Society. 347: 1855-63.

5. Clark, James M., 2001. Long-Term Forecasts of Inflation Rates Using Quantitative Tools. Financial Engineering News. www.fenews.com/fen20/longtermforecasts.html. Last Accessed 10 April 2006.

6. Esty, Daniel, Jack Gladstone, Ted Robert Gurr, Pamela Surko, and Alan Unger, 1995. Working Papers: State Failures Task Force Report. McLean, VA: Science Applications International Corportion (SAIC).

7. Frank, M.J. 1979. On the Simultaneous Associativity of F(x,y) and X + y – F(x,y). Aequationes Mathematicae. 19:194-226.

8. Goldstone, Jack A., Ted Robert Gurr, Monty G. Marshall, Robert H. Bates, Michael Lustik, John Ulfelder, and Mark Woodward. 2005. A Global Forecasting Model of Political Instability. American Political Science Association. Science Applications Internal Corporation (SAIC).

9. Gurr, Ted Robert, and Will H. Moore, 1997. Ethno-political rebellion: A Cross-sectional analysis of the 1980s with risk assessments from the 1990s. American Journal of Political Science. 41: 1079-1104

10. Heidelberg Institute of International Conflict Research. 2006. www.hiik.de/en/index_e.htm. Last Accessed 10 April 2006.

11. Mays, D'arcy. 2005. Virginia Commonwealth University. STAT 643 - Applied Linear Regression. Unpublished course notes.

12. National Center for Education Statistics. 2003. Projection of Education Statistics to 2013. www.nces.ed.gov/programs/projections. Last Accessed 10 April 2006.

13. O'Brien, Sean P., 2002. Anticipating the Good, the Bad, and the Ugly: An Early Warning Approach for Conflict and Instability Analysis. The Journal of Conflict Resolution. 46:791-811

14. O'Brien, Sean P., 2005a. CAA's Crisis and Instability Forecasting Capbilities. Center for Army Analysis presentation.

15. O'Brien, Sean P., 2005b. An Analytical Approach for Monitoring, Shaping, and Evaluating Responses to Country Stability. Center for Army Analysis presentation.

16. O'Brien, Sean P., 2006. FASE Pattern Classification Methodology. www.geocities.com/seanob888/JRC/good2.htm#FASE%20PATTERN%20CLASSIFICATION%20METHODOLOGY. Last Accessed 10 April 2006.

17. Pfetsch, Frank and Christoph Rohloff. 200. National and International Conflicts 1945-1995. New Empirical and Theoretical Approaches. Heidelberg Institute on International Conflict Research.

18. SAS/Insight Users Guide. 1999. www.csc.fi/cschelp/sovellukset/stat/sas/sasdoc/sashtml/insight/index.htm Last Accessed 17 April 2006.

19. The Mathworks. 2006. www.mathworks.com/access/helpdesk/help/toolbox/finance/boxcox.html. Last Accessed 17 April 2006.

20. Wagle, B.V., 1965. A Review of Two Statistical Aids in Forecasting. The Statistician. Vol 15, No 2, 191-196

Vita

Margaret Erin Elizabeth Keck was born on August 6, 1981 in Richmond, Virginia, and is an American citizen. She graduated from Trinity Episcopal High School, Richmond, Virginia in 1999. She received her Bachelor of Arts in Mathematics from Sweet Briar College in 2003. Since graduation she has been employed by the U.S. Army, National Ground Intelligence Center as an intelligence analyst in Charlottesville, Virginia.