Sovereign Credit Rating Prediction

PH125.9x Data Science: Capstone

George Holt (Riskfree)

30 December 2020

Contents

1 Introduction

Can sovereign credit ratings be predicted using current and historical economic and other data about the sovereign? What are the economic and financial variables that best predict sovereign credit ratings? How reliable are the sovereign credit rating predictions given the available economic and financial data? These are some of the questions addressed in the research reviewed in this report.

Credit ratings are ordered categorical assessments of the credit quality of an entity, such as a corporation, that indicate the ability of the entity to satisfy its bond and other debt obligations in a timely fashion. Credit ratings can be assigned to entities by any observer, but the most widely published and respected credit ratings are assigned by three organizations: Moody's Investors Service ("Moodys"), Standard and Poors Corporation ("S&P"), and Fitch Ratings ("Fitch"). These three rating agencies are recognized by securities and banking regulatory authorities as Nationally Recognized Statistical Rating Organizations ("NRSROs") or External Credit Assessment Institutions ("ECAIs") as authorities on credit quality, typically as part of the authorities credit risk regulations.

Credit ratings are useful to investors in fixed income debt instruments, such as bonds issued by a rated entity or loans made to a rated entity. Credit ratings are useful to investors as indicators of the relative credit quality of different debt instruments held in their portfolios, which assists investors in diversifying the credit risk of the portfolios. Credit ratings of rated entities also affect market yield spreads between debt instruments issued by different entities, e.g. between corporate bonds and U.S. Treasury securities, and therefore affect the pricing of those instruments. See for example [Grothe (2013)}. As noted above, credit ratings are also used by banking organizations to calculate risk capital requirements imposed by bank regulatory authorities.

Each of the rating agencies has established a methodology for determining the appropriate credit rating category of an entity at a particular point in time. Generally, these methodologies are based on objective measures of an entity's ability to satisfy its bond and other debt obligations, such as its income and assets, but the selection of a particular rating category is also based upon human judgment, which means that the rating assignment is partly subjective. While the labels for rating categories differ somewhat across the agencies, the categories for all agencies can be grouped into a consistent grading hierarchy, and the rating categories for each agency can be rank ordered according to increasing credit risk.

The agencies generally provide credit ratings for four types of debt obligations: long-term foreign currency, short-term foreign currency, long-term local currency, and short-term local currency, where long-term means over one year and short-term under one year, and foreign currency means obligations denominated in the currency of a country other than the sovereign obligor and local currency means obligations denominated in the currency of the sovereign obligor. In this report, only long-term foreign currency credit ratings are considered, since they are the most commonly used sovereign credit ratings. The grading hierarchy and rank ordering of credit rating categories for the three rating agencies are presented in Table 1.

This table shows that Moody's has 22 rating categories, Standard & Poors has 23 rating categories, and Fitch has 25 rating categories.

One class of entities that is covered by the rating agencies is sovereign governments, which are generally the national governments of various countries. The ability to satisfy debt obligations differs among corporations, sovereigns, and other entities, so the rating agencies methodologies for sovereigns differ from the methodologies for other types of obligors. Considering these differences, the rating agencies list numerous economic, social, and political factors that underlie their sovereign credit rating assessments in their statements on rating criteria.

In the first systematic analysis of the determinants and impact of the sovereign credit ratings assigned by the credit ratings agencies, Richard Cantor and Frank Packer of the Federal Reserve Bank of New York identified eight variables that are repeatedly cited in rating agency reports as determinants of sovereign ratings. These variables are (see Cantor and Packer (1996)):

- *Per capita income.* Sovereigns with a higher per capita income have a larger potential tax base a greater ability to repay debt, so they should have a higher credit rating. This variable can also serve as a proxy for the level of political stability and other important factors.

Table 1: Long-Term Credit Rating Categories

Grade	Moodys	SandP	Fitch
Prime	Aaa	AAA	AAA
High grade	Aa1	AA+	AA+
	Aa2	AA	AA
	Aa3	AA-	AA-
Upper medium grade	A1	A+	A+
	A2	A	A
	A3	A-	A-
Lower medium grade	Baa1	BBB+	BBB+
	Baa2	BBB	BBB
	Baa3	BBB-	BBB-
Non-investment grade speculative	Ba1	BB+	BB+
	Ba2	BB	BB
	Ba3	BB-	BB-
Highly speculative	B1	B+	B+
	B2	B	B
	B3	B-	B-
Substantial risks	Caa1	CCC+	CCC+
	Caa2	CCC	CCC
	Caa3	CCC-	CCC-
Extremely speculative	Ca	CC	CC
			C
In default with little prospect for recovery		SD	RD
In default	C	D	D
			DD
			DDD
Not rated	WR	NR	

- *GDP growth.* A relatively high rate of economic growth suggests that a sovereign's existing debt burden will become easier to service over time.
- *Inflation.* A high rate of inflation points to structural problems in the government's finances. When a government appears unable or unwilling to pay for current budgetary expenses through taxes or debt issuance, it must resort to inflationary money finance. Public dissatisfaction with inflation may in turn lead to political instability.
- *Fiscal balance.* A large government deficit absorbs private domestic savings and suggests that a government lacks the ability or will to tax its citizenry to cover current expenses or to service its debt, while a government surplus indicates the ability to tax or reduce expenses to service debt.
- *External balance.* A large current account deficit indicates that the public and private sectors together rely heavily on funds from abroad. Current account deficits that persist result in growth in foreign indebtedness, which may become unsustainable over time.
- *External debt.* A higher debt burden should correspond to a higher risk of default. The weight of the burden increases as a country's foreign currency debt rises relative to its foreign currency earnings (exports).
- *Economic development.* Countries that are classified as economically developed, are expected to have a higher credit rating. They are perceived to have attained a certain minimum threshold of economic development for which default is very unlikely. In addition, these countries are often strongly integrated with the world economy, such that a default is less likely, as foreign creditors can more easily disrupt trade or seize assets abroad in case of default. A proxy for this minimum income or development level is a simple indicator variable noting whether or not a country is classified as industrialized by the International Monetary Fund.
- *Default history.* Other things being equal, a country that has defaulted on debt in the recent past is widely perceived as a high credit risk. Both theoretical considerations of the role of reputation in sovereign debt (see Eaton (1996)) and related empirical evidence indicate that defaulting sovereigns suffer a severe decline in their standing with creditors (see Ozler (1991)). Credit reputation can be incorporated by using an indicator variable that notes whether or not a country has defaulted on its international bank debt since 1970.

Several studies of sovereign credit rating determinants have been conducted subsequent to the research performed by Cantor and Packer. (See, e.g., Valle and Marín (2005)). These studies typically show that many of the economic and financial indicators listed above are important determinants of the credit ratings assigned to sovereigns. This report investigates whether the credit rating assessments of the credit rating agencies for sovereigns can be predicted using surrogates for the variables outlined by Cantor and Packer using machine learning models.

The discussion above provides some background on sovereign credit ratings and the observable variables that credit rating agencies use to assign credit ratings of sovereign obligors at various points in time. In the next section, the methodology for obtaining information about sovereign credit ratings and the economic and financial variables suggested as determining sovereign credit ratings is discussed. This section discusses both current and historical credit ratings as well as the credit rating actions upon which the historical credit ratings are based. Web scraping technology is employed to obtain the current rating and credit rating action data, which is subsequently analyzed and converted to historical time series of sovereign credit ratings.

The next section also discusses the sources for the economic and financial data suggested as determining sovereign credit ratings. Most of this data is obtained from databases maintained by the World Bank (IBRD) and the International Monetary Fund (IMF). This economic and financial data is downloaded from these databases as *.csv* files that are retained in a *GitHub* repository, and these files are accessed using the R language *read_csv()* function to create tibbles (data frames).

The following section develops modeling approaches used to estimate relationships between sovereign credit ratings and economic and financial indicators for sovereigns. These modeling approaches are applied to random training subsets of the historical observations for each agency's ratings to estimate model parameters, and the parameter estimates and other statistics for each fitted model are examined. The resulting fitted model for each agency is then used to predict credit ratings given a random test subset of associated indicator values that is a complement to the training observation subset. The resulting predicted credit ratings are displayed and compared to the actual historical credit ratings for each model.

In the next section, the predictive performance of each of the agency credit rating models is evaluated. The predictive performance of a model is measured in alternative ways: (i) whether the model predicts the same credit rating as was assigned by the rating agency, and (ii) whether the model predicts a credit rating that is the same or close to the credit rating assigned by the rating agency. In the latter case, alternative approaches to measuring the distance between different credit rating categories are applied. These performance measures provide a means of evaluating the ability of a model to predict credit ratings, and if so, how well each model predicts ratings similar to actual ratings.

The final section of this report summarizes the results of the model development and evaluation process and also offers conclusions and provides suggestions for further research on sovereign credit rating prediction.

2 Methodology and Analysis

In this section, sources of sovereign credit rating information are identified, and approaches to extracting this information from the sources are discussed. Descriptive statistics and charts for current sovereign credit ratings are produced. The results show that sovereign credit rating assessments are not uniformly distributed across each agency's credit rating categories. Instead, a small group of sovereigns is assigned each agency's highest rating categories, and there are a substantial number of sovereigns that are assigned to each agency's lower rating categories.

Current sovereign credit ratings alone are insufficient to develop predictive models, since only about 140 sovereigns have current credit ratings. Consequently, historical credit ratings for sovereigns must also be acquired to provide sufficient observations to develop predictive models. However, historical credit rating series for sovereigns are not readily available, and the historical series must be constructed from information about historical credit rating actions taken by each agency. (A credit rating action is an announcement by a rating agency that is the result of a review of a sovereign indicating the current status of the sovereign's rating and whether it changes the credit rating category for the sovereign.) Thus, this report describes the process of obtaining the histories of credit rating actions for each sovereign and for converting these actions into a historical series of credit ratings. This process results in an annual time series of historical credit rating category assessments by each rating agency for each sovereign.

This section also considers sources for economic and financial indicators that are considered determinants of sovereign credit ratings. Several of the indicators that were identified by Cantor and Packer are discussed, and the sources of these indicators are explored. Since predictive models for both current and historical credit ratings are developed, a history of each indicator is also required to estimate the relationship between indicator values and ratings. Consequently, the report describes the process for obtaining the time series of each indicator over a time interval consistent with the credit rating histories.

The next part of this section develops models for predicting sovereign credit ratings. Earlier research on sovereign credit ratings used traditional linear regression models for this purpose, but there are several theoretical and practical issues with this approach. To address these issues, this report develops two predictive models for sovereign credit ratings that employ machine learning techniques. The first modeling approach is called *Ordinal Forest Classification*, which is similar to the random forest models used for regression, and the second modeling approach is *Linear Discriminant Analysis*. Each of these modeling approaches is used to train a sovereign credit rating prediction model for each of the rating agencies. The training process uses a randomly-selected subset of observations of historical credit ratings for each rated sovereign and associated economic and financial indicator values. Each model that results from this process is then applied to a complementary randomly-selected subset of observations of historical economic and financial indicator values associated with the sovereigns to produce predicted credit ratings for each sovereign during each year. The predicted ratings for each sovereign, rating agency, and year are then compared to actual ratings, and these results are summarized in confusion matrices and displayed graphically in mosaic charts.

The final part of this section considers some metrics for measuring the predictive performance of ordinal classification models such as the sovereign credit rating prediction models developed in this report. In addition to traditional *Percentage Observed Agreement*, these metrics include three variants of Cohen's *kappa*, Kendall's *Coefficient of Concordance*, and Spearman's *Average Rank Correlation*. Each of these metrics is applied to the predictions from the models for each rating agency to obtain the desired statistics, and the performance results are presented and summarized.

2.1 Sovereign Credit Ratings

Analysis of sovereign credit ratings depends upon observations of current credit ratings as well as historical credit rating behavior, which consists of credit rating actions for sovereign entities over some historical time period. Each credit rating agency publishes current credit ratings and maintains a history of its credit rating actions, but accessing this data directly from each agency is not necessarily the most convenient method of collecting the data.

After investigating the potential sources of credit rating information, the internet website *countryeconomy*.com was identified as a more convenient source, since sovereign credit ratings and historical rating actions for all three rating agencies are available there. The *countryeconomy* website also offers a variety of additional economic and financial data for each country, so it is also convenient source for that information. The website claims to collect data from a variety of reputable sources, such as the credit rating agencies, the International Monetary Fund, The World Bank, and the central banks of various countries, but there may be differences between the website data and the original source data that have not been investigated while performing this analysis. Investigation of any differences is the subject of future research, so this analysis will assume that the data is consistent with the original source data.

To acquire the current credit ratings and historical credit rating action data from the *countryeconomy.com* website, a so-called web scraping process was developed that takes the *html* content for webpages and extracts the credit ratings and credit rating actions from those webpages. For purposes of this analysis, a main webpage containing a list of countries and current rating information and links to country specific webpages and a collection of webpages containing the historical credit rating actions for each country were the primary sources of the credit rating data. The process employed to perform web scraping involves downloading the *html* text, applying various functions in the *rvest* web scraping package to access nodes in the *html* text, using regular expression string functions to extract the rating information from the nodes, and assembling the rating information into data structures used in analysis.

After performing the web scraping process, there are 144 sovereigns currently rated by one or more rating agencies. These sovereigns are shown in Table 2.

Table 2: Rated Sovereigns

Sovereigns	Sovereigns.	Sovereigns..
United States	Dominican Republic	Malawi
United Kingdom	Ecuador	Mexico
Germany	Estonia	Malaysia
France	Egypt	Mozambique
Japan	Ethiopia	Namibia
Spain	Finland	Nigeria
Italy	Fiji	Nicaragua
Portugal	Gabon	Netherlands
Greece	Grenada	Norway
Ireland	Georgia	New Zealand
Andorra	Ghana	Oman
United Arab Emirates	The Gambia	Panama
Albania	Guatemala	Peru
Armenia	Hong Kong	Papua New Guinea
Angola	Honduras	Philippines
Argentina	Croatia	Pakistan
Austria	Hungary	Poland
Australia	Indonesia	Paraguay
Azerbaijan	Israel	Qatar
Bosnia and Herzegovina	India	Romania
Barbados	Iraq	Serbia
Bangladesh	Iran	Russia
Belgium	Iceland	Rwanda
Burkina Faso	Jamaica	Saudi Arabia
Bulgaria	Jordan	Seychelles
Bahrain	Kenya	Sweden
Benin	Cambodia	Singapore
Bolivia	South Korea	Slovenia
Brazil	Kuwait	Slovakia
Bahamas	Kazakhstan	San Marino
Botswana	Laos	Senegal
Belarus	Lebanon	Suriname
Belize	Liechtenstein	El Salvador
Canada	Sri Lanka	Thailand
Democratic Republic of the Congo	Lesotho	Turkmenistan
Republic of the Congo	Lithuania	Tunisia
Switzerland	Luxembourg	Turkey
Ivory Coast	Latvia	Trinidad and Tobago
Chile	Libya	Taiwan
Cameroon	Morocco	Ukraine
China	Moldova	Uganda
Colombia	Montenegro	Uruguay
Costa Rica	North Macedonia	Uzbekistan
Cuba	Mali	Saint Vincent and the Grenadines
Cape Verde	Mongolia	Venezuela
Cyprus	Malta	Vietnam
Czech Republic	Mauritius	South Africa
Denmark	Maldives	Zambia

2.1.1 Current Sovereign Credit Ratings

The current sovereign credit ratings are the cumulative result of credit rating actions taken by the rating agencies over the time period covered by the credit rating actions. Moody's currently has credit ratings for 110 countries, Standard & Poors currently has credit ratings for 121 countries, and Fitch currently has credit ratings for 124 countries. Credit rating assessments are not uniformly distributed across each agency's rating categories. This is demonstrated by histograms showing the number of countries in each Moody's rating category in Figure 1, the number of countries in each Standard & Poors rating category in Figure 2, and the number of countries in each Fitch rating category in Figure 3.

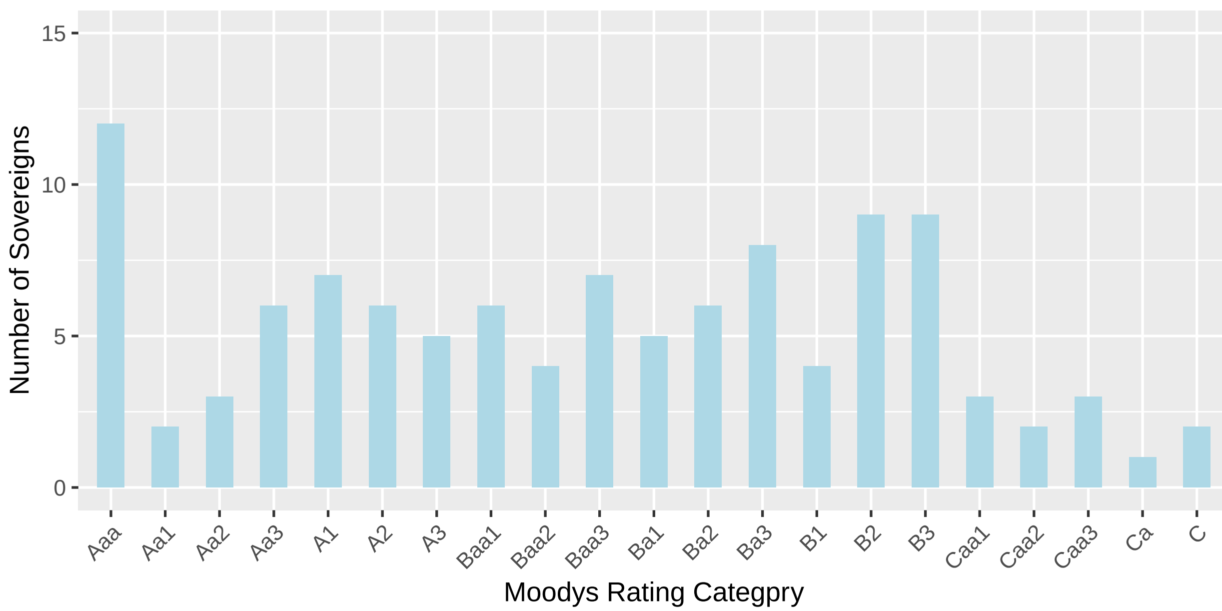

Figure 1: **Number of Sovereigns by Moody's Rating Category**

These figures show that there are about a dozen sovereigns that the rating agencies have assigned their highest rating category, e.g. "AAA", while slightly less than half of the sovereigns have credit ratings with credit grades that represent speculative investments, i.e. credit ratings below "BBB-".

2.1.2 Historical Sovereign Credit Rating Actions

Each credit rating agency performs reviews of a sovereign's creditworthiness periodically or as events affecting the sovereign dictate with the objective of providing investors with relatively current information about the riskiness of the sovereign's debt instruments. These reviews may result in announcements of credit rating actions, which are statements about what credit rating is appropriate for a sovereign at the time of the announcement. Announcements can indicate that a different rating category is appropriate than the category most recently indicated by the agency, or it may confirm the most recently announced rating category. If an agency has not previously performed a credit review of a sovereign and initiates ratings for a sovereign, the rating action would typically indicate the initial rating category for a sovereign.

Credit rating actions are sometimes accompanied by a qualifier, such as Stable, Negative, Positive, or Under Review, that provide further information about the agency's views on the direction of a sovereign's credit rating. While the qualifiers provide additional information to investors, they do not directly affect a sovereign's credit rating at a given point in time, and they do not affect the history of a sovereign's credit ratings over historical time periods. Since the emphasis in this report is on credit rating history, the significance of rating qualifiers is not considered in the analysis that follows, although analysis of qualifier significance may be an appropriate subject for further research.

When these qualifier events are excluded from the credit rating action data, the data used for analysis covers 141 countries and consists of 4421 observations. To better understand the process of taking credit rating actions at the credit rating agencies, it is useful to examine the number of rating actions where a rating is either confirmed or changed each year. The set of credit rating action observations used in this report

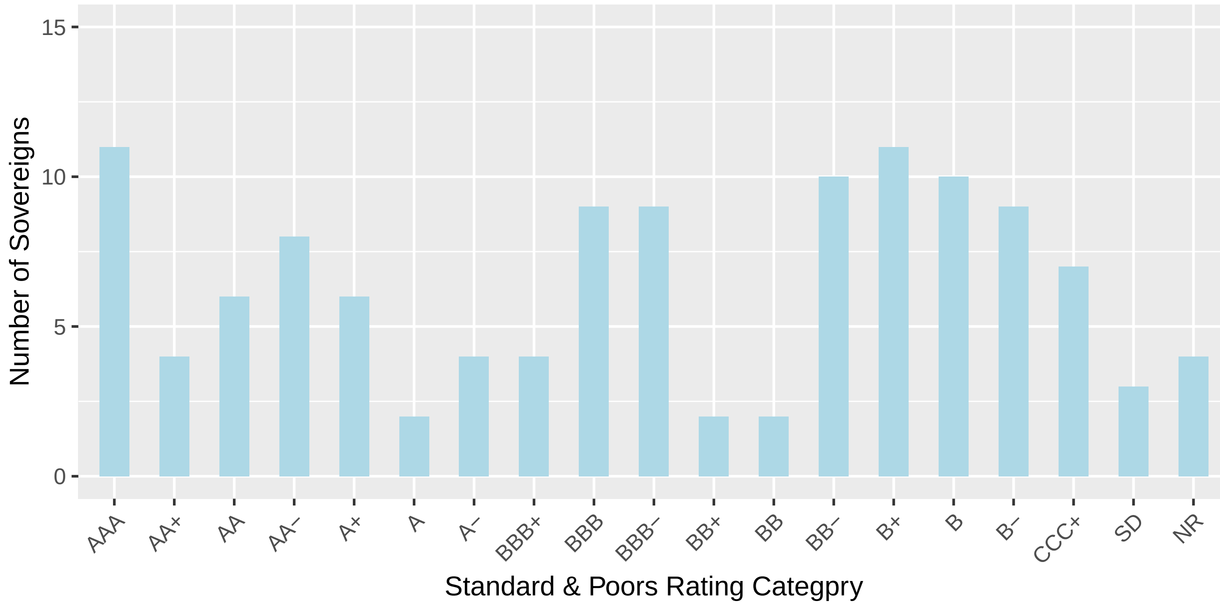

Figure 2: **Number of Sovereigns by Standard & Poors Rating Category**

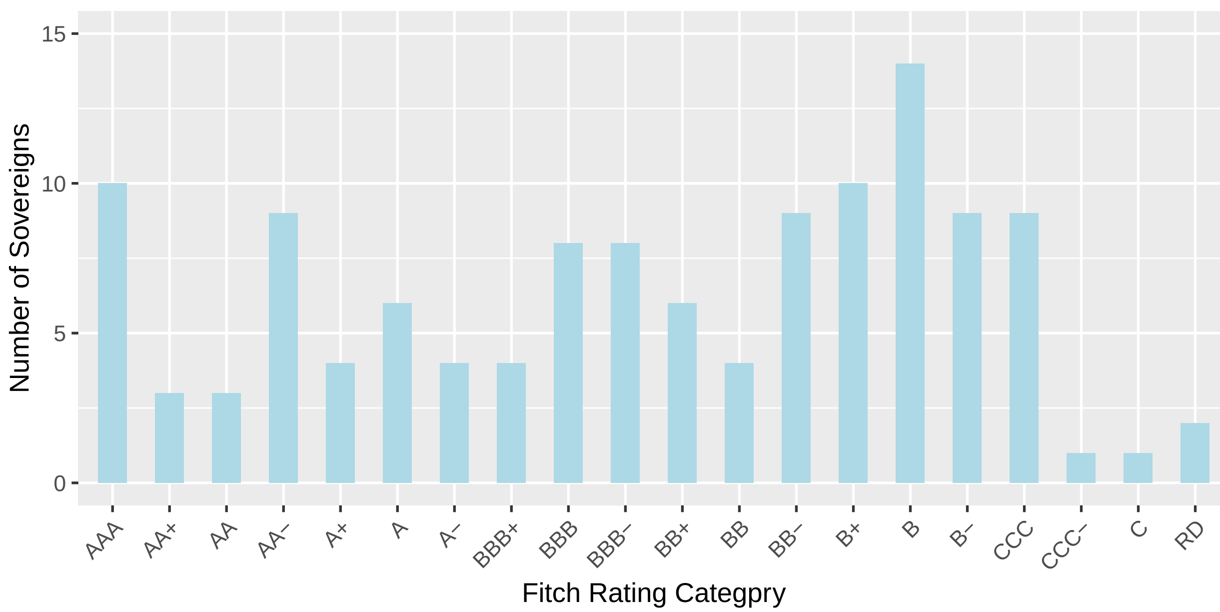

Figure 3: **Number of Sovereigns by Fitch Rating Category**

contains observations from 1994 through November 2020, excluding rating qualifier only events. The number of sovereign credit rating actions excluding rating qualifier events each year is shown in Figure 4.

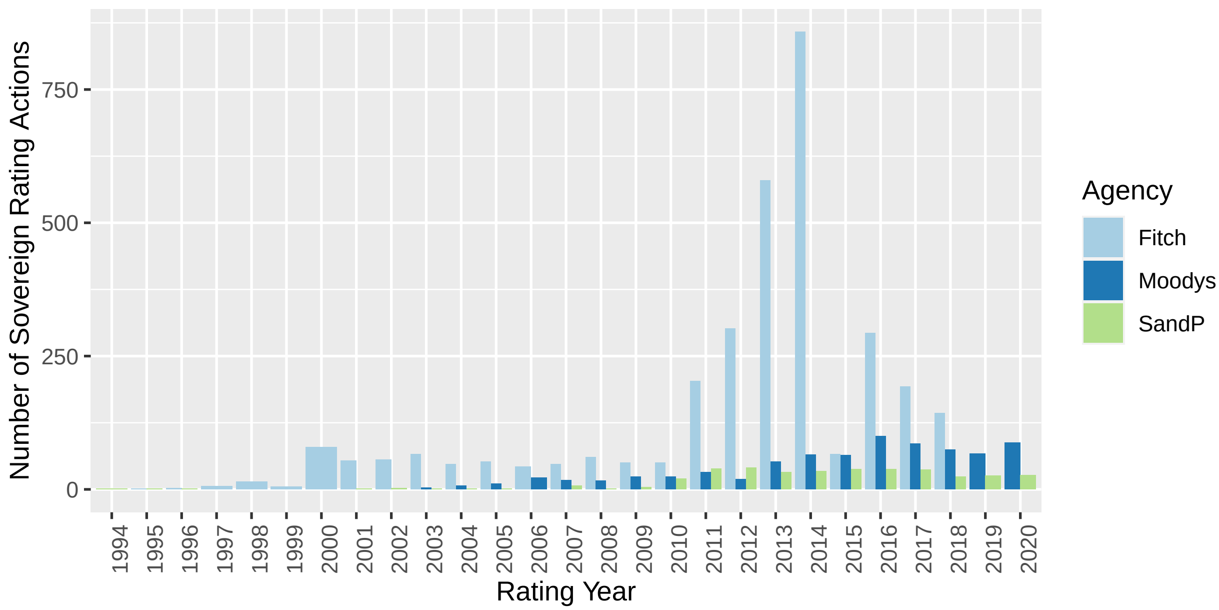

Figure 4: **Credit Rating Actions by Year**

This bar chart shows that there were very few credit rating actions by each rating agency prior to the year 2000 and overall less than 100 actions per year between 2000 and 2020, with the exception of 7 different years when Fitch produced between 130 and 820 rating actions. Fitch appears to be much more active than Moody's or Standard & Poors, but this may be because Fitch reconfirms many sovereign ratings each year without changing the rating category. Recently, Moody's has apparently been the most active agency in issuing rating actions in the past few years.

2.1.3 Historical Sovereign Credit Ratings

The information provided by the credit rating actions of each rating agency for each sovereign provides a means of constructing a history of the credit rating for each sovereign by each rating agency over historical time periods. As noted above, most of the credit rating actions for sovereigns have occurred over the dozen or so years prior to 2020. This information is sufficient to provide a time series of credit rating categories assigned by each rating agency for each sovereign for an interval of about 20 years, and it allows some insight into the dynamics of sovereign creditworthiness over this interval.

The process for constructing the annual credit rating time series for each sovereign and rating agency uses the average credit rating for each year where one or more credit rating actions occurred. When there is only one credit rating action by an agency for a sovereign during a given year, the average credit rating is the same as the credit rating category for the action. For example, if Moody's assigns a credit rating category of "A1" to Spain once during a given year, this credit rating action results in the average credit rating of "A1" for the Moody's credit rating for the year. When there are two or more rating actions by an agency for a sovereign during a given year, the average credit rating is the credit rating category obtained by calculating the average ranking of credit categories, rounding the resulting average to the nearest integer ranking, and using the rounded average ranking to obtain the corresponding rating category for the agency. For example, if Moody's assigns a credit rating category of "A1" to Spain on one day during a year and a rating category of "Aa2" on another day during the same year, these credit rating actions result in the average credit rating of "Aa3" as the Moody's credit rating for the year. (This result is obtained by averaging the ranking of 5 for the “A1” rating and the ranking of 3 for the “Aa2” rating, resulting in an average ranking of 4, corresponding to a “Aa3” rating.)

The rating agencies do not take credit rating actions for each sovereign during every year, so the averaging process outlined above does not necessarily result in an annual credit rating of a sovereign by a rating agency for every year during a given time interval. For years where no credit rating actions for a sovereign occurs, the appropriate credit rating category for the year is the credit rating category for the sovereign assigned by the rating agency by its the most recent credit rating action. For example, if Moodys did not issue a credit rating action for Spain during 2019, but Moodys most recent credit rating action for Spain occurred on April 18, 2018, the credit rating category for that action (“Baa1”) is the appropriate credit rating category for 2019.

2.2 Sovereign Economic and Financial Indicators

As discussed earlier, Cantor and Packer identified several economic and financial indicator variables that are important factors in determining the creditworthiness of sovereign debt issuers. Cantor and Packer also indicated the sources of the data they used in their study, with the World Bank being the primary source at the time of their research. Today, the World Bank remains the most important source of this data, so it is the primary source for the economic and financial data used in this report. Some data is also obtained from the International Monetary Fund (IMF). The following discussion considers each of the economic and financial indicators identified by Cantor and Packer and briefly explores the data available for these indicators.

All of the data obtained from the World Bank is retained in databases maintained by the bank. While the content of these databases is directly available through a web-page based interface, and the data from these databases can be downloaded using a web-based API, it appeared more convenient to separately download the data as *.csv* files and retain these files in a GitHub repository. These files can be downloaded from GitHib subsequently using the *read_csv()* function provided in the R language, which results in a data frame containing the data that can be easily used for further exploration and analysis. The IMF databases are also directly available through a web-page interface, but in this case downloading *.csv* files was the most convenient approach to obtaining the data.

To illustrate the comparative historical behavior of the economic and financial indicators across sovereigns, time series charts of each indicator for multiple sovereigns are included in the following discussion. However, since 127 sovereigns are included in the overall analysis, time series charts for only a limited set of sovereigns are presented as illustrations.

2.2.1 GNI Per Capita

Cantor and Parker identified per capita income as an important factor in determining sovereign credit ratings. In this report, the per capita value for gross national income (GNI. Formerly GNP) expressed in current international dollars and converted by a purchasing power parity (PPP) conversion factor is used as the indicator for per capita income. GNI is the sum of value added by all resident producers plus any product taxes (less subsidies) not included in the valuation of output plus net receipts of primary income (compensation of employees and property income) from abroad. The PPP conversion factor is a spatial price deflator and currency converter that eliminates the effects of the differences in price levels between countries. The data for GNI per capita for each sovereign is downloaded from a World Bank database.

The time series for GNI per capita are presented in Figure 5. This indicator appears to exhibit roughly linear growth over time.

2.2.2 GNI Growth

The second variable determining credit ratings identified by Cantor and Packer is sovereign GDP growth per year. GDP growth is now called GNI growth. In this report, the measure of sovereign economy’s GDP growth is measured by the change in the volume of its output or in the real incomes of its residents at market prices based on constant local currency. Aggregates are based on constant 2010 U.S. dollars. GNI is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for

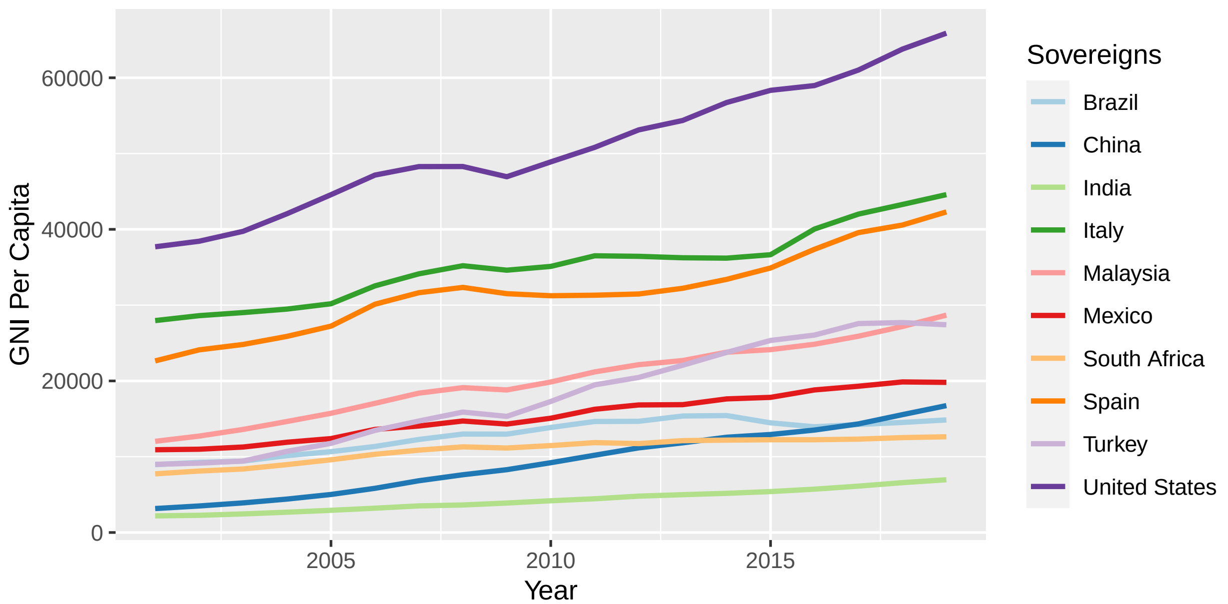

Figure 5: **GNI Per Capita by Year**

depreciation of fabricated assets or for depletion and degradation of natural resources. GDP accounts for all domestic production, regardless of whether the income accrues to domestic or foreign institutions. The data for GNI growth for each sovereign is downloaded from a World Bank database.

The time series for GNI growth are shown in Figure 6. This indicator appears to fluctuate substantially over the 20-year interval covered by the chart, with growth rates as high as 14 percent and as low as -6 percent. All of the selected set of sovereigns were impacted by events in 2009, although China was less affected than the other sovereigns shown.

2.2.3 CPI Inflation

Cantor and Packer identified consumer price index (CPI) inflation as a third determinant of sovereign credit ratings. In this report, inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. The Laspeyres formula is generally used. (The Laspeyres Index is calculated by working out the cost of a group of commodities at current prices, dividing this by the cost of the same group of commodities at base period prices, and then multiplying by 100. The data for CPI inflation for each sovereign is downloaded from a World Bank database.

The time series for CPI inflation rate are shown in Figure 7. Most of the selected sovereigns have inflation rates under 5 percent over the 20-year interval covered by the chart, but Turkey had extreme inflation early in this period and more recently had higher than typical inflation.

2.2.4 Fiscal Balance

The fourth determinant of sovereign credit ratings identified by Cantor and Packer is a sovereign's fiscal balance. Fiscal balance is defined as the average annual central government budget surplus relative to GDP. After investigating potential sources for this data, including the World Bank and the International Monetary Fund, sufficient data to perform the analyses in this report did not appear to be available. The U.S. Central Intelligence Agency (CIA) does occasionally publish fiscal balance information for many countries, but this informaion is typically only available for select years. Cantor and Packer indicated that there were data

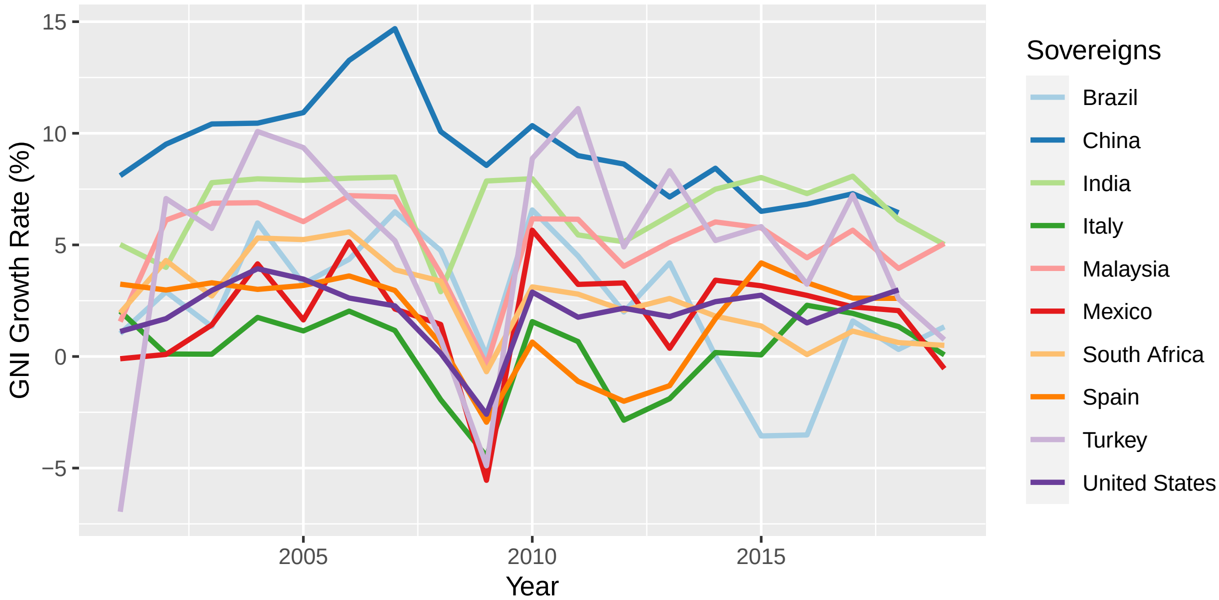

Figure 6: **GNI Growth by Year**

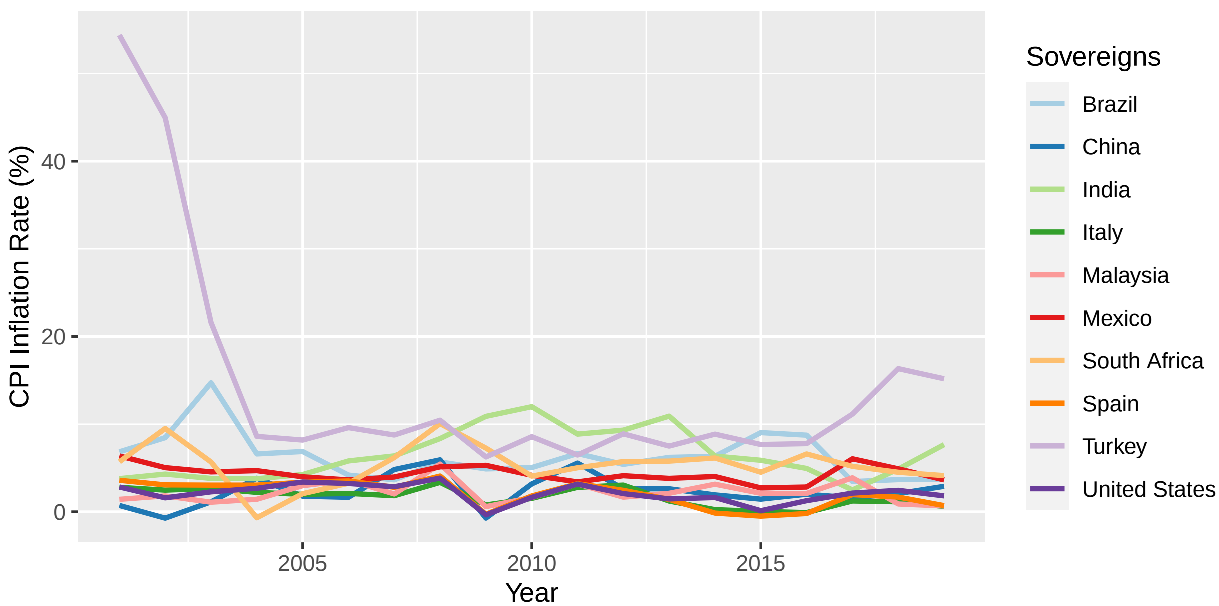

Figure 7: **CPI Inflation Rate by Year**

limitations that affected their use of this indicator. These data limitations may be more important today than in the 1990s, since there are many more sovereigns with ratings today. The linear regression analysis performed by Cantor and Packer also indicated that fiscal balance was relatively unimportant as an indicator for predicting sovereign credit ratings. Because of these limitations, fiscal balance is not included in the analysis in this report, and its usefulness in predicting sovereign credit ratings will be subject to further research.

2.2.5 Current Account Balance

Cantor and Packer identified a sovereign's external balance as a fifth determinant of sovereign credit ratings. They define external balance as a sovereign's average annual current account surplus relative to GDP. In this report, current account balance as a percent of GNI is used as the corresponding indicator of a sovereign's external balance. Current account balance is defined as the sum of net exports of goods and services, net primary income, and net secondary income. This indicator is often referred to as the balance of payments for a sovereign. The data for the current account balance as a percent of GDP for each sovereign is downloaded from a World Bank database.

The time series for current account balance are presented in Figure 8. The chart shows that until about 2013, the selected sovereigns had diverging current account balances as a percent of their GDP, with Malaysia having the largest surplus and Spain and Turkey having the largest deficits. Since 2013, the selected sovereigns appear to have current account balances that are less in absolute value than 5 percent.

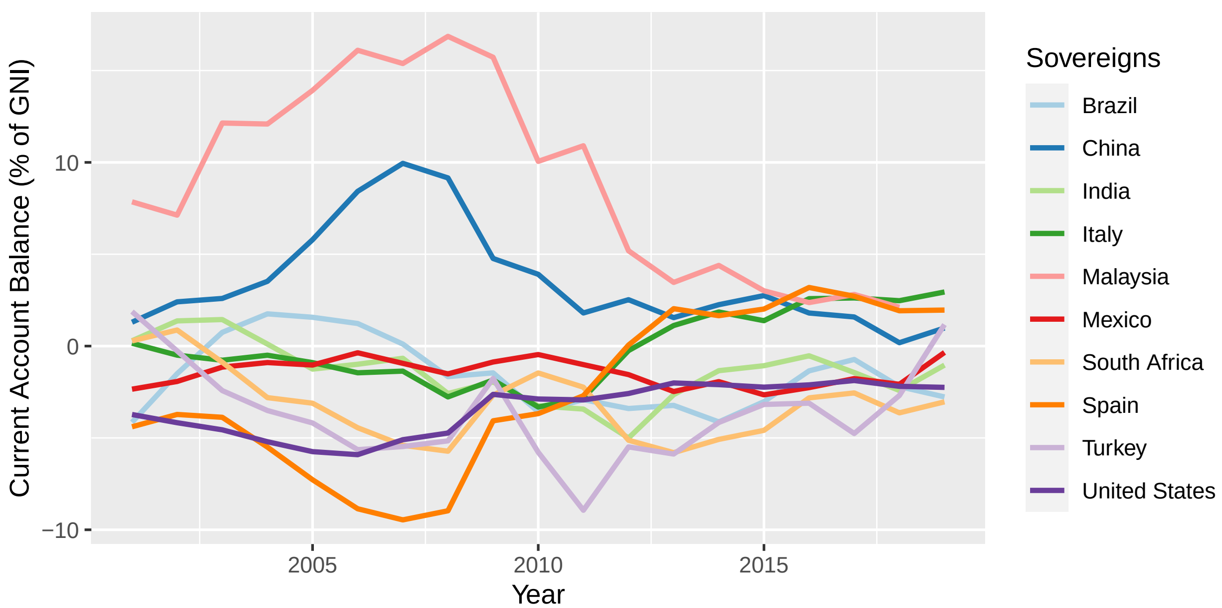

Figure 8: **Current Account Balance by Year**

2.2.6 External Debt

External debt was identified as a sixth determinant of sovereign credit ratings by Cantor and Packer. They define external debt as foreign currency debt as a percentage of exports. External debt is that part of the total debt in a country that is owed to creditors outside the country. The debtors can be the government, corporations or private households. The debt includes money owed to private commercial banks, other governments, or international financial institutions.

External indebtedness affects a sovereign's creditworthiness and investor perceptions. Non-reporting sovereigns might have outstanding debt with the World Bank, other international financial institutions, or private creditors. Total debt service is contrasted with sovereigns' ability to obtain foreign exchange through exports of goods, services, primary income, and workers' remittances. Debt ratios are used to assess the sustainability of a sovereign's debt service obligations, but no absolute rules determine what values are too high. Empirical analysis of developing countries' experience and debt service performance shows that debt service difficulties become increasingly likely when the present value of debt reaches 200 percent of exports. Still, what constitutes a sustainable debt burden varies by sovereign. Sovereigns with fast-growing economies and exports are likely to be able to sustain higher debt levels.

Information on total external debt is not readily available for many sovereigns, so central government debt as a percentage of GNI is used as an alternative measure of sovereign indebtedness in this report. Data for central government debt is available from the IMF for some sovereigns. This data was downloaded in the form of a *.csv* file from the IMF web site and retained in a *github* repository, and this data is extracted from the *.csv* file in the repository using the *read.csv()* function.

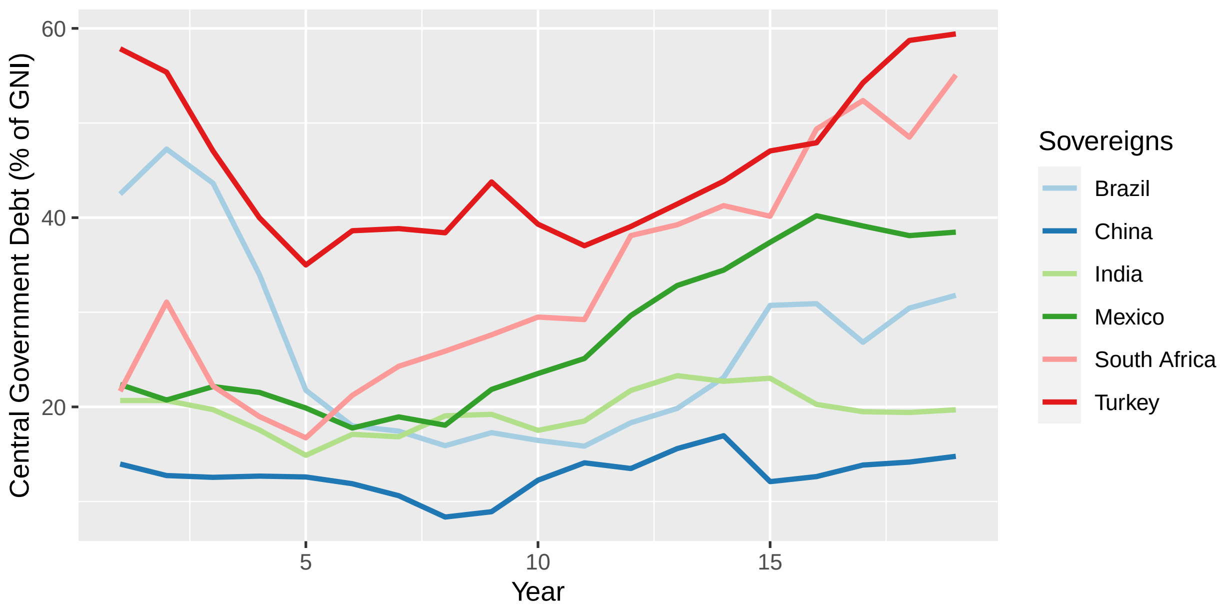

Figure 9: **External Debt Ratio by Year**

The chart in Figure 9 shows *external_debt_ratio* for some of the selected sovereigns. This chart shows that the ratio has been increasing in recent years.

Unfortunately, this dataset does not include information for many of the most highly rated-sovereigns, e.g., the United States. So its inclusion in subsequent analyses produces indicator observations that do not contain observations for these highly-rated sovereigns, which means that any models trained using indicator observations that include the *external_debt_ratio* will produce biased results that do not cover higher credit rating categories. Consequently, development of sovereign credit rating prediction models in the remainder of this report will not include *external_debt_ratio* as an indicator variabe.

2.2.7 Developing Economy Indicator

Cantor and Packer included an indicator for economic development as a determinant of sovereign credit ratings. According to the IMF, developing countries are those countries whose standard of living, income, economic and industrial development remain more or less below average. In this report, a list of 152 sovereigns provided by the IMF is used to identify the sovereigns with credit ratings that are considered to be developing countries. This data is contained in a *.csv* file retained in a *github* repository, and it is extracted from the repository using the *read_csv* function.

2.2.8 Sovereign Indicators

The indicators discussed above provide the economic and financial metrics required to develop, calibrate, and test credit rating prediction models. These indicators are assembled into a set of observations for each sovereign and year considered in the modeling process. A small number of sovereigns do not have sufficient data to support the modeling process, so these sovereigns are dropped from consideration. The modeling process also requires that the indicators for each sovereign, year, and indicator have values, i.e., are not NAs, so the indicators for each sovereign and year with missing values are also dropped from the analysis.

2.3 Sovereign Credit Rating Prediction

Credit rating prediction is an ordinal classification process from a data science viewpoint. Credit ratings for each rating agency are an ordered collection of rating categories, and the ordering of the rating categories for each agency provides additional ranking information that supplements the assignment of a particular rating category. From a theoretical viewpoint, sovereign credit rating prediction should account for ordinal classification across a set of discrete credit rating categories and for the information provided by the ranking of the credit rating categories.

Cantor and Packer and other earlier analysts of sovereign credit ratings used a linear regression approach to describe the relationship between credit rating categories and explanatory variables. The linear regression approach assigns a numerical value to each credit rating category that provides a ranking of the rating categories that effectively assumes that the distance between successive rating categories in the ranking is the same for all rating categories. There is nothing inherent in the rating agencies methodologies that supports this assumption, so the distances between successive ratings are effectively unknown. Thus the linear regression approach makes assumptions about rating categories and the distances between successive ratings that is not consistent with the ordinal classification assumptions.

This report offers an alternative perspective on predicting sovereign credit ratings using modeling approaches that are consistent with the ordinal classification process outlined above. In particular, two different modeling approaches based on machine learning models are developed and applied to the sovereign credit rating and indicator observations discussed in the previous sections of this report. These approaches are:

- Ordinal Forest Models
- Linear Discriminant Analysis Models

An important issue in developing sovereign credit rating prediction models is the relatively small number of credit rating observations for an individual rating agency each year. At best the number of observations is limited by the number of sovereigns rated by each agency, which is currently slightly over 120 sovereigns for each of the rating agencies, but has been less in previous years. Using a single years observations provides an insufficient basis for training (calibrating) a sovereign credit rating model, because all of the credit rating categories in a rating agency's rating hierarchy typically do not appear in a single year, and because the lack of one or more credit rating categories for an agency in a given year makes it impossible to incorporate the unobserved ratings in the predictive model.

To address this issue, this report uses all of the years for which credit rating and indicator observations are available as the basis for developing credit rating prediction models. Under this assumption, every year covered is treated the same as all other years in the overall observation set. This assumption leads to an observation set with more than 2,000 observations for each rating agency that includes observations of all of the credit rating categories in each of the rating agencies rating hierarchies, which appears sufficient to train the desired credit rating models. At the same time, this assumption effectively assumes that the importance of the indicator variables in predicting an agency's credit ratings does not change over time. Some research (see Reusens and Croux (2017)) has indicated that the rating agencies may change their weighting of some indicator variables over an extended period of time, so this assumption deserves some additional analysis, which will be deferred for future research.

Natural logarithms of the values of *cpi_inflation* are used as observations, consistent with the analysis of Cantor and Packer.

2.3.1 Relationship Between Ratings and Indicators

In this section, the relationships between each rating agency's credit rating categories and the economic and financial indicators are explored. This is accomplished by plotting the distribution of the values of each indicator for each credit rating category separately for each rating agency. The distributions are presented as horizontal boxplots.

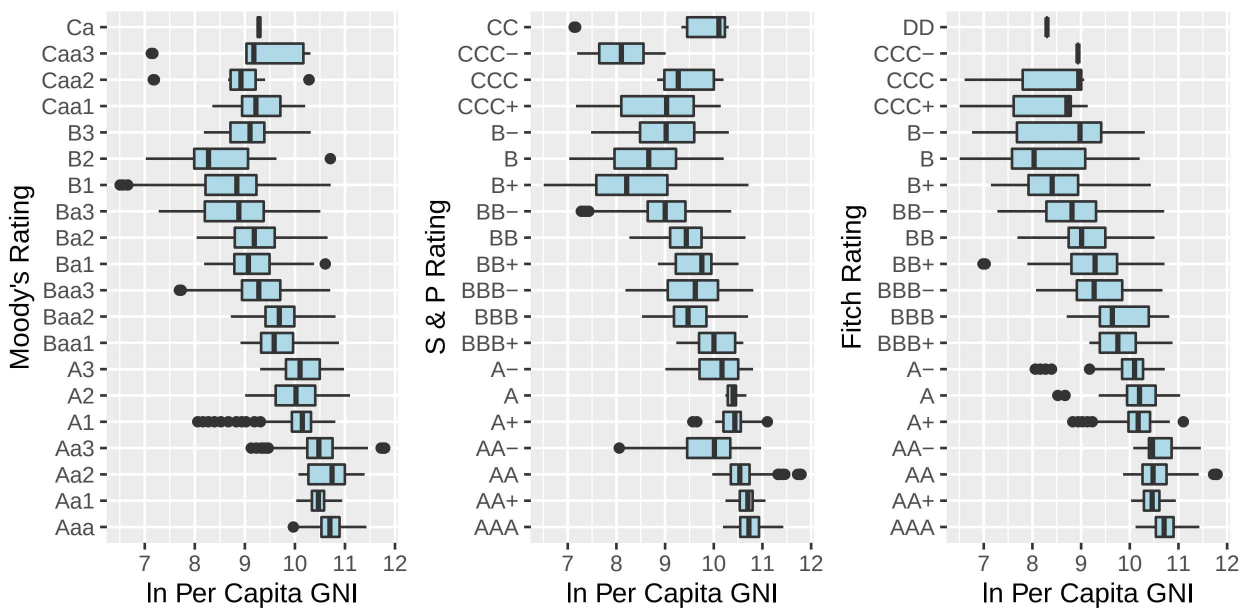

Figure 10: **Per Capita GNI vs Agency Rating Category**

The relationships between each rating agency's credit rating categories and *per_capita_gni* is presented in Figure 10. These plots appear to show a positive exponential relationship between higher ratings and *per_capita_gni*.

The relationships between each rating agency's credit rating categories and *gni_growth* is presented in Figure 11. *gni_growth* appears to have a modest negative impact on credit ratings.

The relationships between each rating agency's credit rating categories and *cpi_inflation* is presented in Figure 12. *cpi_inflation* appears to have a modest negative impact on credit ratings.

The relationships between each rating agency's credit rating categories and *current_accountt_ratio* is presented in Figure 13. *current_account_ratio* appears to have a modest positive impact on credit ratings.

Figure 14 shows separate distributions of credit rating categories for developing and developed economies for each rating agency. The distributions indicate that the development status for a sovereign's economy has a significant impact on credit rating assessments.

2.3.2 Ordinal Forest Models

Ordinal forest models (see Hornung (2020)) are a relatively new approach to describing the relationship between an ordinal response variable, such as a credit rating, and a set of explanatory variables. These models are similar to random forest regression models, where a continuous outcome is related to some explanatory variables, but in this case the outcomes are an ordered set of classes, such as credit ratings. A straightforward forest-based prediction method for ordinal response variables consists of simply considering a regression forest using the class values $1, ..., J$ of the response variable for the corresponding classes. However, this procedure

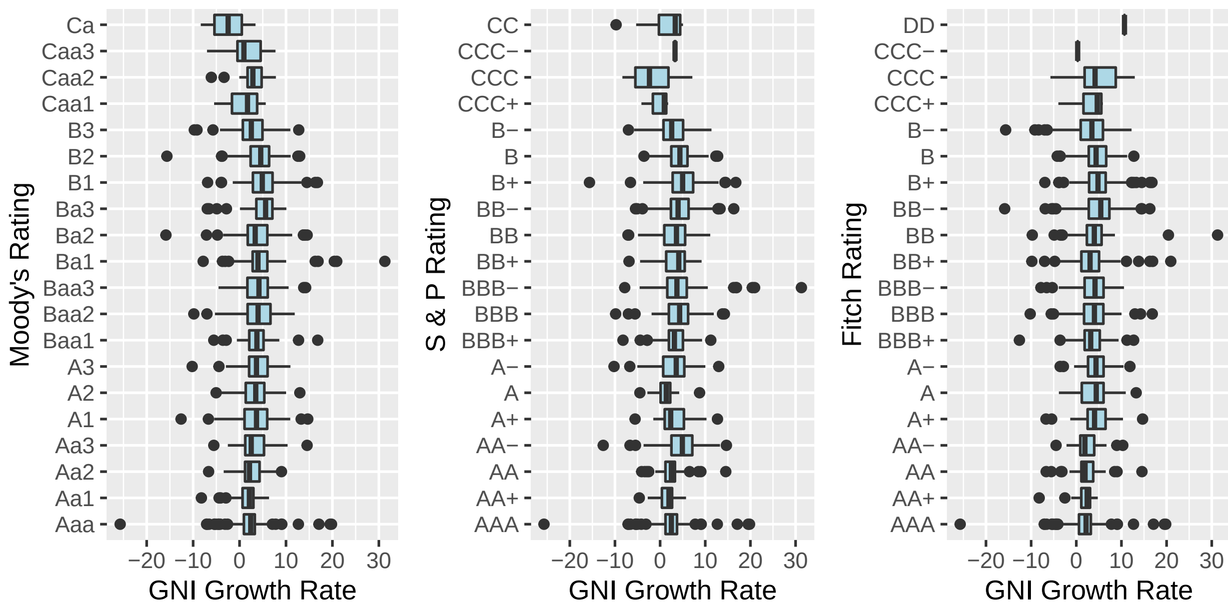

Figure 11: **GNI Growth Rate vs Agency Rating Category**

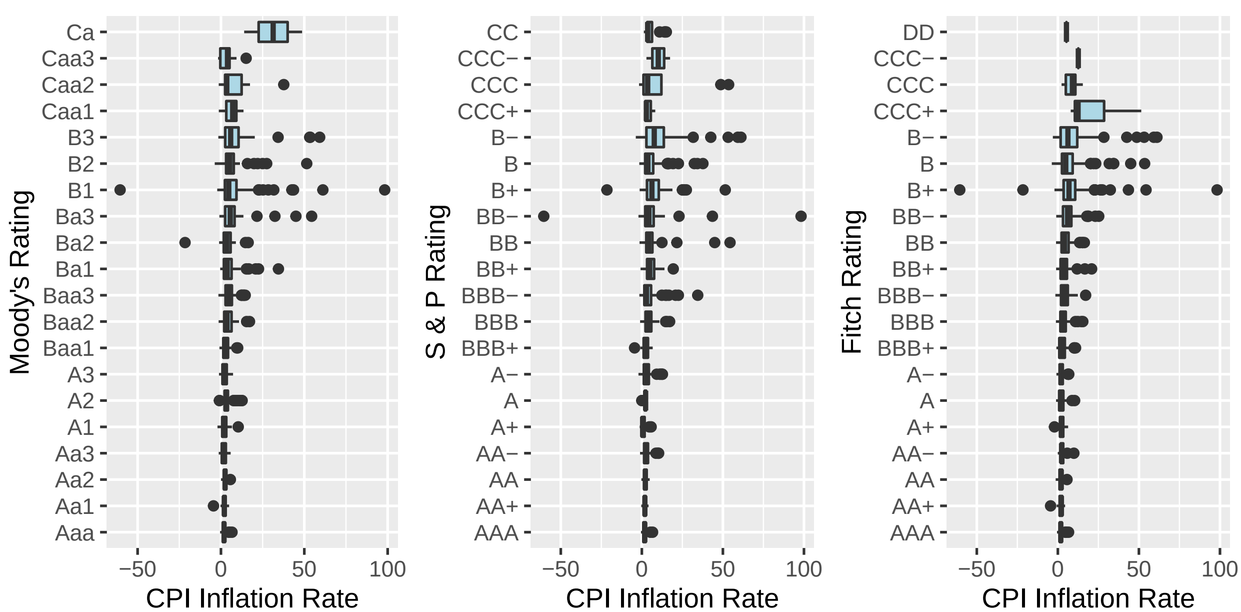

Figure 12: **CPI Inflation Rate vs Agency Rating Category**

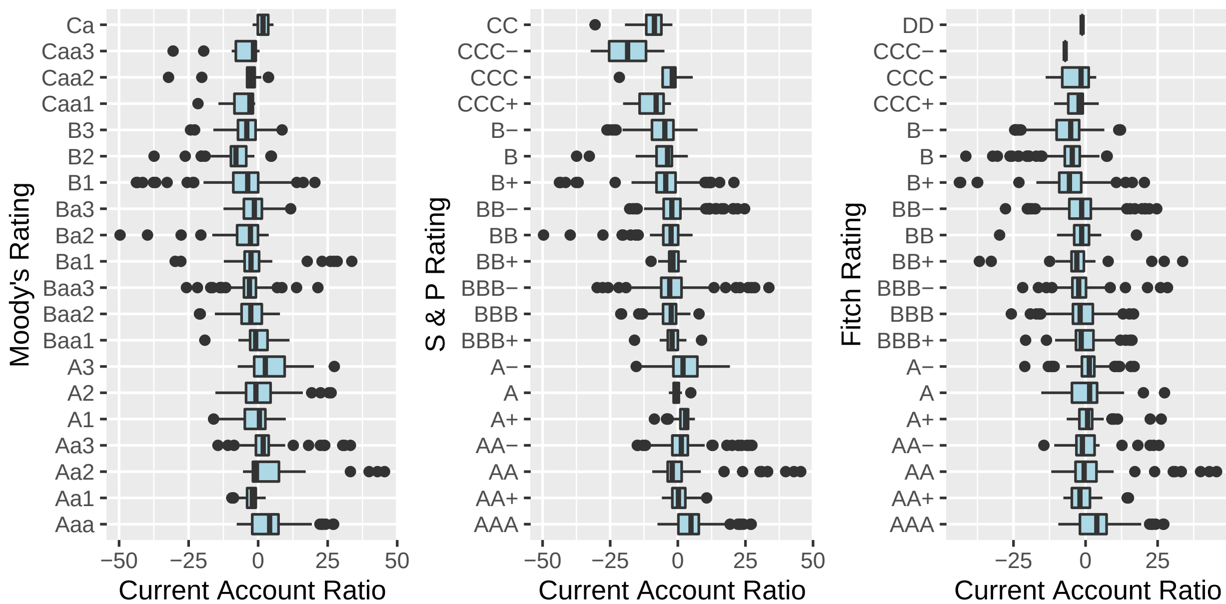

Figure 13: **Current Account Ratio vs Agency Rating Category**

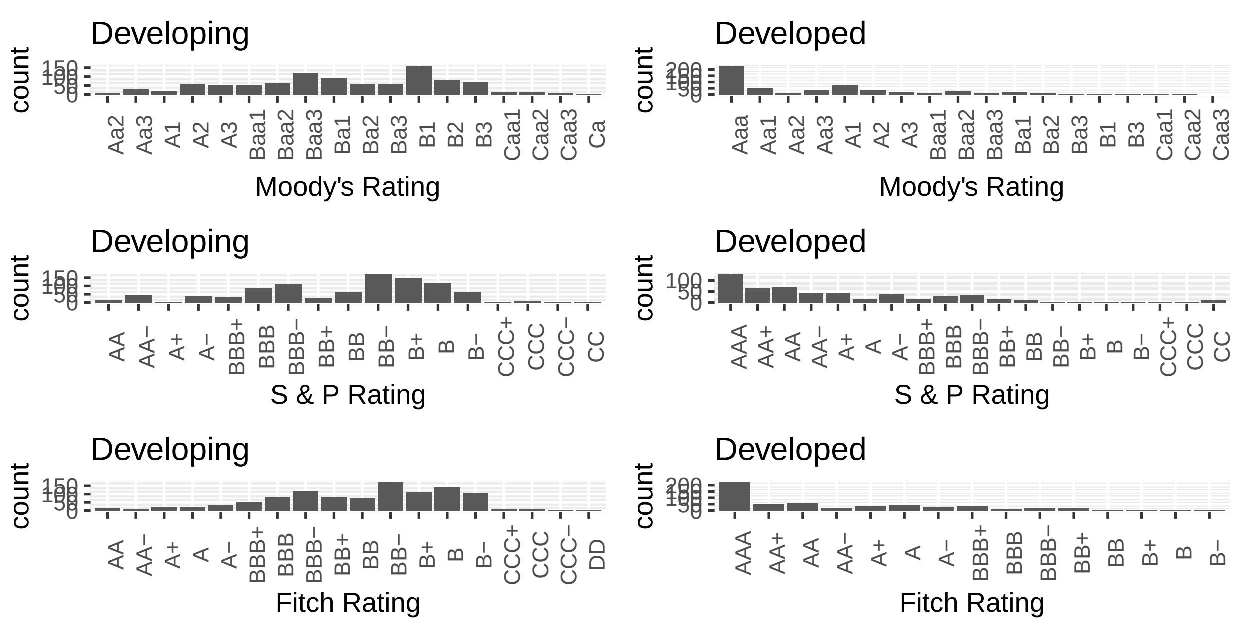

Figure 14: **Agency Rating Category Developing vs Developed**

is suboptimal because the extents of the classes of the ordinal response variable, or “class widths”, differ from class to class.

In the ordinal forest model, the class widths are the widths of J adjacent intervals in the range of an underlying continuous response variable; each of the J intervals correspond to the J classes of the ordinal response variable. The single assumption in this model is that, underlying the observed ordinal variable y, there exists a particular known or unknown refined continuous variable y^* that determines the values of the ordinal variable. The relationship between this continuous variable y^* and y is such that the higher the value of y^* is for an observation, the higher is the class of the ordinal response variable for that observation, e.e., if y^* falls into the jth interval of J adjacent intervals, y will take the value j.

The ordinal forest model is designed for the common situation in which the underlying continuous variable is not measured or is unknown. In ordinal forest, interval boundaries in y^* corresponding to the different classes of y are estimated or optimized by maximizing the OOB prediction performance of regression forests. Using score values that correspond to these optimized class intervals instead of using the class values $1, ..., J$ leads to an improvement in prediction performance. (The Out-of-bag (OOB) error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging) to sub-sample data samples used for training.)

A separate ordinal forest model is developed for each rating agency using the historical observations of sovereign credit ratings and corresponding economic and financial indicator values. Each model is trained using a randomly-selected subset of the historical observations, and the fitted model is next used to predict sovereign credit ratings using the indicators for a random complement test subset of the observations. The predicted credit ratings are then compared to the test subset credit ratings to assess the accuracy of the predictions.

2.3.2.1 Moody’s Ordinal Forest Model The Moody’s ordinal forest model is trained using a subset comprised of 1071 observations. The observations cover ratings for 92 countries across 20 rating categories. The summary output from training this model is shown below.

```
Ordinal forest

Number of observations: 1071, number of covariates: 5

Classes of ordinal target variable:
"Aaa" (n = 157), "Aa1" (n = 36), "Aa2" (n = 13), "Aa3" (n = 46), "A1" (n = 65), "A2" (n =

Forest setup:
Number of trees in ordinal forest: 500
Number of considered score sets in total: 100
Number of best score sets used for approximating the optimal score set: 10
Number of trees per regression forests constructed in the optimization: 100
Performance function: "probability"
```

The fitted model assigns a different level of importance to each of the indicator variables, as shown in the following table.

```
         per_capita_gni              growth_gni          cpi_inflation
            0.081852783             0.006795749            0.014330317
current_account_ratio                developing
            0.019265852             0.033865693
```

The table indicates that *per_capita_gni* is the first most important determinant of Moody’s sovereign credit ratings.

Table 3: Moody's Ordinal Forest Model Confusion Matrix

Actual Rating	Predicted Rating																					
	Aaa	Aa1	Aa2	Aa3	A1	A2	A3	Baa1	Baa2	Baa3	Ba1	Ba2	Ba3	B1	B2	B3	Caa1	Caa2	Caa3	Ca	C	WR
Aaa	64	12	4	8	9	4	4	2	3	1	0	2	0	0	0	0	0	0	0	0	0	0
Aa1	3	4	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0
Aa2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aa3	2	0	1	4	1	2	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0
A1	2	0	1	1	12	2	4	1	3	1	4	0	1	1	0	0	1	0	0	0	0	0
A2	0	0	0	1	3	6	1	2	3	2	2	2	1	2	0	0	0	0	0	0	0	0
A3	0	0	0	0	0	3	8	4	1	2	1	0	0	0	0	1	0	0	0	0	0	0
Baa1	0	0	0	1	0	2	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0
Baa2	0	0	0	1	3	1	2	3	7	7	3	2	2	5	0	1	0	0	0	0	0	0
Baa3	0	0	0	0	0	2	2	5	1	13	3	1	0	2	2	1	0	0	1	0	0	0
Ba1	0	0	0	0	1	2	1	3	2	4	9	4	0	1	1	3	0	0	1	0	0	0
Ba2	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	1	1	0	0	0	0	0
Ba3	0	0	0	0	0	0	0	0	0	1	0	0	6	0	0	0	0	0	0	0	0	0
B1	0	0	0	1	0	2	1	4	0	10	6	6	5	29	5	7	2	2	1	1	0	0
B2	0	0	0	0	0	0	0	0	0	0	0	1	0	3	11	1	0	1	0	0	0	0
B3	0	0	0	0	0	1	0	1	1	2	0	0	0	1	2	5	2	2	1	0	0	0
Caa1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Caa2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Caa3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Ca	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The fitted ordinal forest model is next used to predict the Moody's sovereign credit ratings based upon the randomized test set of indicators for each sovereign. The model calculates the probability of each credit rating category for each sovereign and its associated indicator values. The test set for the Moody's model contains 459 observations covering 89 sovereigns across 20 rating categories. The predicted ratings for the test set are compared to the actual test set ratings in the confusion matrix shown in Table 3. A visualization of the confusion matrix is presented in Figure 15.

2.3.2.2 Standard and Poors Ordinal Forest Model The Standard and Poors ordinal forest model is trained using a subset comprised of 1031 observations. The observations cover ratings for 91 countries across 20 rating categories. The summary output from training this model is shown below.

```
Ordinal forest

Number of observations: 1031, number of covariates: 5

Classes of ordinal target variable:
"AAA" (n = 96), "AA+" (n = 42), "AA" (n = 55), "AA-" (n = 68), "A+" (n = 33), "A" (n = 12)

Forest setup:
Number of trees in ordinal forest: 500
Number of considered score sets in total: 100
Number of best score sets used for approximating the optimal score set: 10
Number of trees per regression forests constructed in the optimization: 100
Performance function: "probability"
```

The fitted model assigns a different level of importance to each of the indicator variables, as shown in the following table.

```
       per_capita_gni          growth_gni      cpi_inflation
          0.068620433         0.005740243        0.012482946
current_account_ratio          developing
          0.020101995         0.035074298
```

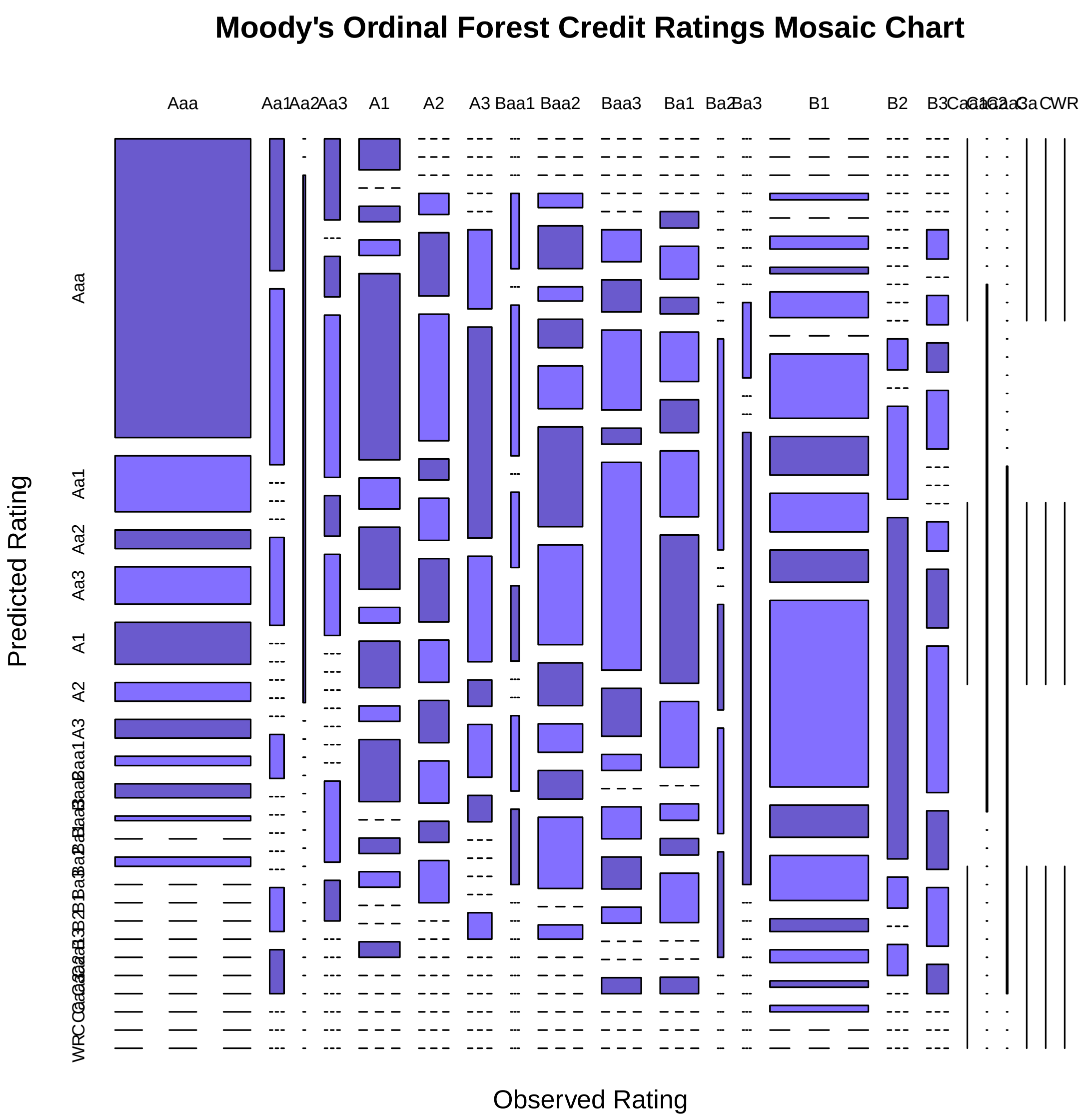

Figure 15: Moody's Ordinal Forest Model Prediction Mosaic

Table 4: Standard and Poors Ordinal Forest Model Confusion Matrix

Actual Rating	Predicted Rating																					
	AAA	AA+	AA	AA-	A+	A	A-	BBB+	BBB	BBB-	BB+	BB	BB-	B+	B	B-	CCC+	CCC	CCC-	CC	SD	D	NR
AAA	29	11	10	0	6	1	0	3	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0
AA+	1	7	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AA	3	4	12	1	1	1	1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0
AA-	0	0	1	7	1	0	10	0	3	3	1	1	3	0	0	0	0	0	0	0	0	0	0
A+	0	0	0	1	3	0	2	4	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
A	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A-	0	0	1	2	1	1	7	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
BBB+	0	0	0	0	1	0	0	2	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0
BBB	0	0	0	5	0	0	2	6	14	12	3	4	5	0	3	1	0	0	0	1	0	0	0
BBB-	0	0	0	1	0	0	5	2	4	16	1	2	3	1	1	3	0	1	0	0	0	0	0
BB+	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
BB	0	0	0	0	0	0	0	0	2	0	1	1	1	0	0	0	0	1	0	0	0	0	0
BB-	0	0	0	0	0	0	0	2	7	7	1	4	27	6	12	4	1	1	0	0	0	0	0
B+	0	0	1	0	0	0	0	0	0	1	0	2	6	26	7	4	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	1	1	2	0	0	4	6	11	0	0	0	0	0	0	0	0
B-	0	0	0	0	0	0	0	0	0	2	0	1	1	1	1	11	0	0	0	0	0	0	0
CCC+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCC	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
CCC-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CC	0	0	1	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0
SD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The table indicates that *per_capita_gni* is the most important determinant of Standard and Poors sovereign credit ratings.

The fitted ordinal forest model is next used to predict the Standard and Poors sovereign credit ratings based upon the randomized test set of indicators for each sovereign. The model calculates the probability of each credit rating category for each sovereign and its associated indicator values. The test set for the Standard and Poors model contains 441 observations covering 89 sovereigns across 20 rating categories. The predicted ratings for the test set are compared to the actual test set ratings in the confusion matrix shown in Table 4. A visualization of the confusion matrix is presented in Figure 16.

2.3.2.3 Fitch Ordinal Forest Model

The Fitch ordinal forest model is trained using a subset comprised of 1139 observations. The observations cover ratings for 99 countries across 20 rating categories. The summary output from training this model is shown below.

```
Ordinal forest

Number of observations: 1139, number of covariates: 5

Classes of ordinal target variable:
"AAA" (n = 149), "AA+" (n = 36), "AA" (n = 57), "AA-" (n = 18), "A+" (n = 41), "A" (n = 49

Forest setup:
Number of trees in ordinal forest: 500
Number of considered score sets in total: 100
Number of best score sets used for approximating the optimal score set: 10
Number of trees per regression forests constructed in the optimization: 100
Performance function: "probability"
```

The fitted model assigns a different level of importance to each of the indicator variables, as shown in the following table.

```
        per_capita_gni             growth_gni          cpi_inflation
           0.074603412            0.005569457            0.010320679
```

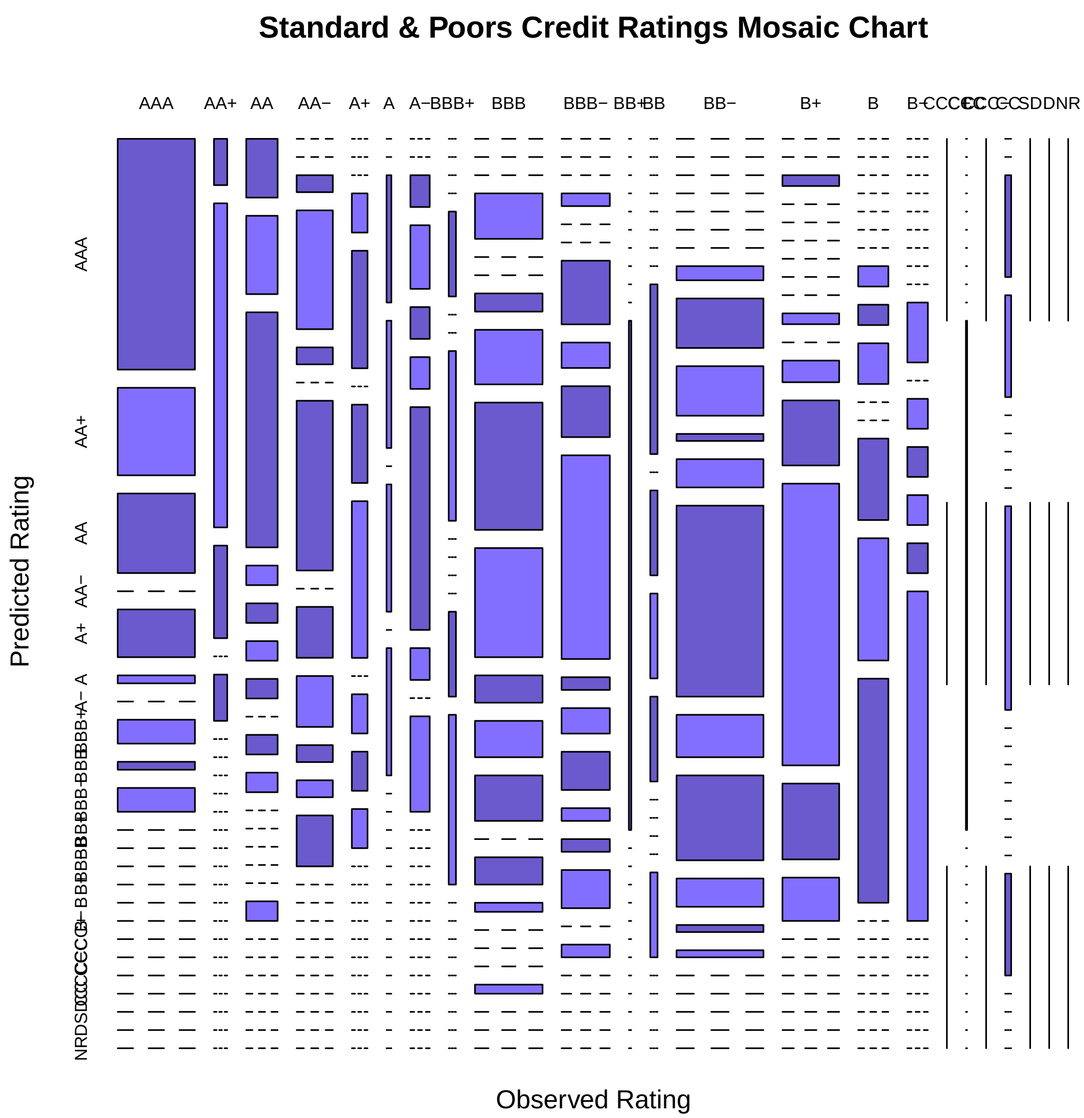

Figure 16: Standard & Poors Ordinal Forest Model Prediction Mosaic

Table 5: Fitch Ordinal Forest Model Confusion Matrix

Actual Rating	Predicted Rating																							
	AAA	AA+	AA	AA-	A+	A	A-	BBB+	BBB	BBB-	BB+	BB	BB-	B+	B	B-	CCC+	CCC	CCC-	CC	C	RD	D	DD	DDD
AAA	71	6	8	6	2	5	1	4	1	1	1	0	0	0	0	3	0	0	0	0	0	0	0	0	0
AA+	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AA	2	2	4	1	4	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
AA-	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A+	0	0	0	0	5	2	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
A	1	0	3	0	5	3	1	1	3	3	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
A-	0	1	0	0	0	1	7	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BBB+	0	1	1	0	1	3	3	3	1	5	2	1	1	0	1	0	0	0	0	0	0	0	0	0	0
BBB	0	0	0	0	0	1	0	4	10	2	2	3	7	0	4	0	0	0	0	0	0	0	0	0	0
BBB-	0	0	0	0	1	0	1	1	4	9	9	6	2	1	1	2	0	0	0	0	0	0	0	0	0
BB+	0	1	0	0	0	0	0	1	0	3	5	0	0	0	4	0	0	0	0	0	0	0	0	0	0
BB	0	0	0	0	0	0	0	1	0	1	0	4	0	0	1	1	0	0	0	0	0	0	0	0	0
BB-	0	0	0	0	3	0	1	1	1	9	10	7	34	9	7	8	0	2	0	0	0	0	0	0	0
B+	0	0	0	0	0	0	0	0	0	0	0	1	7	14	5	1	0	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	1	0	0	4	0	0	2	4	19	7	0	0	0	0	0	0	0	0	0
B-	0	0	0	0	0	0	1	0	3	3	0	1	1	3	6	13	0	2	0	0	0	0	0	0	0
CCC+	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
CCC	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
CCC-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DDD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

```
current_account_ratio          developing
          0.010818979         0.038972077
```

The table indicates that *per_capita_gni* is the most important determinant of Fitch sovereign credit ratings.

The fitted ordinal forest model is next used to predict the Fitch sovereign credit ratings based upon the randomized test set of indicators for each sovereign. The model calculates the probability of each credit rating category for each sovereign and its associated indicator values. The test set for the Fitch model contains 487 observations covering 99 sovereigns across 20 rating categories. The predicted ratings for the test set are compared to the actual test set ratings in the confusion matrix shown in Table 5. A visualization of the confusion matrix is presented in Figure 17.

2.3.3 Linear Discriminant Analysis Models

The second ordinal classification modeling approach considered in this report is *Linear Discriminant Analysis* (LDA). LDA is a generalization of Fisher's linear discriminant, a method used in statistics and other fields, to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier. Discriminant analysis works by creating one or more linear combinations of predictors, creating a new latent variable for each function. These functions are called discriminant functions. The number of functions possible is either $N_g - 1$, where N_g is the number of groups, or p (the number of predictors), whichever is smaller. The first function created maximizes the differences between groups on that function. The second function maximizes differences on that function, but also must not be correlated with the previous function. This continues with subsequent functions with the requirement that the new function not be correlated with any of the previous functions.

Given group j, with R_j sets of sample space, there is a discriminant rule such that if $x \in R_j$, then $x \in j$. Discriminant analysis then, finds "good" regions of R_j to minimize classification error, therefore leading to a high percent correct classified in the classification table. Each function is given a discriminant score to determine how well it predicts group placement:

- Structure Correlation Coefficients: The correlation between each predictor and the discriminant score of each function. This is a zero-order correlation (i.e., not corrected for the other predictors).
- Standardized Coefficients: Each predictor's weight in the linear combination that is the discriminant function. Like in a regression equation, these coefficients are partial (i.e., corrected for the other

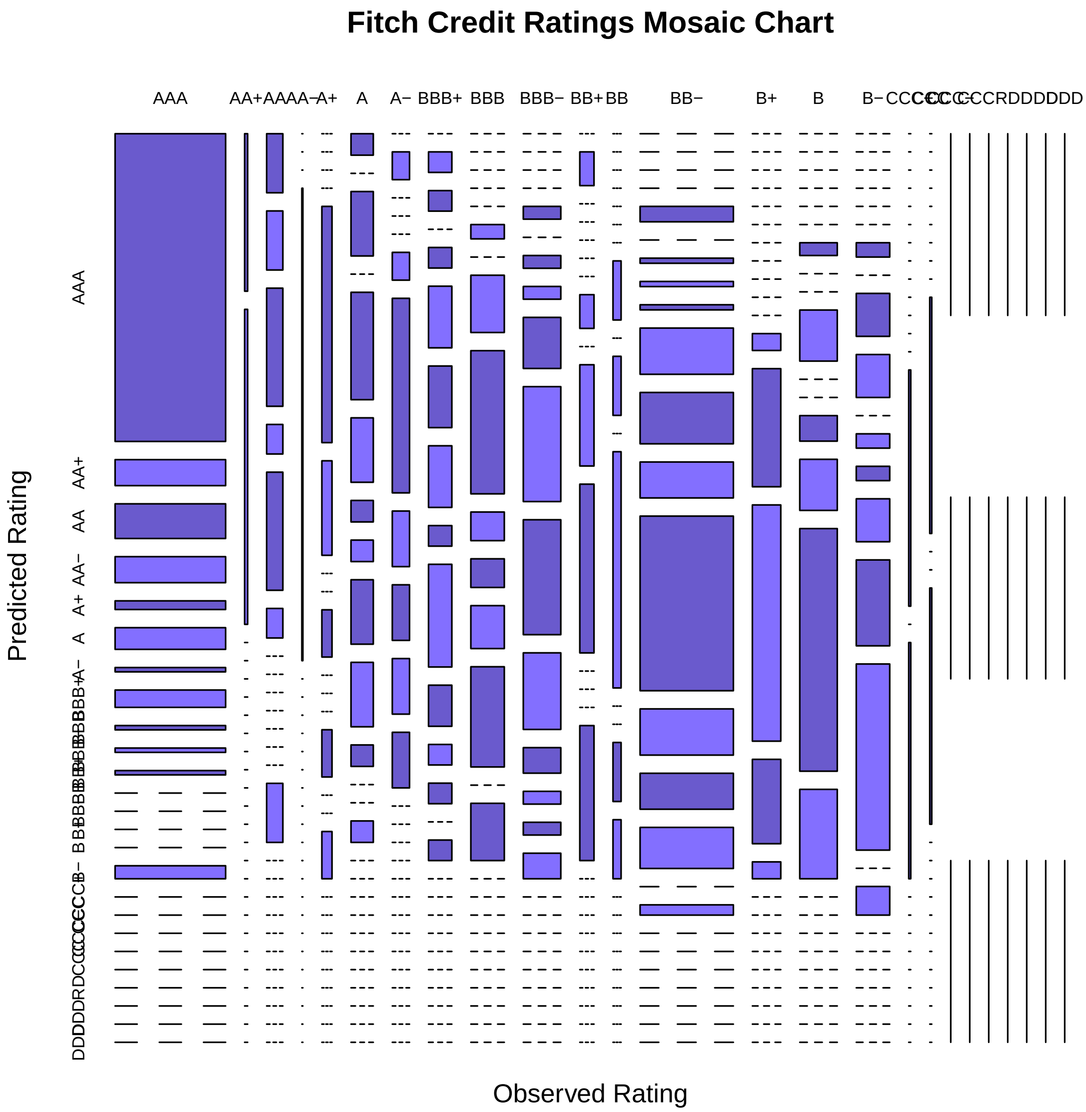

Figure 17: Fitch Ordinal Forest Model Prediction Mosaic

predictors). Indicates the unique contribution of each predictor in predicting group assignment.

- Functions at Group Centroids: Mean discriminant scores for each grouping variable are given for each function. The farther apart the means are, the less error there will be in classification.

2.3.3.1 Moody's Linear Discriminant Analysis Model The Moody's linear discriminant analysis model is trained using a subset comprised of 1071 observations. The observations cover ratings for 92 countries across 20 rating categories. The summary output from training this model is shown below.

Fitting the LDA model to Moody's observations produces the following results.

```
Call:
lda(rating ~ per_capita_gni + growth_gni + cpi_inflation + current_account_ratio +
    developing, data = moodys_observations_train_data)

Prior probabilities of groups:
          Aaa          Aa1          Aa2          Aa3           A1           A2
0.1465919701 0.0336134454 0.0121381886 0.0429505135 0.0606909430 0.0662931839
           A3         Baa1         Baa2         Baa3          Ba1          Ba2
0.0476190476 0.0317460317 0.0606909430 0.0877684407 0.0821661998 0.0457516340
          Ba3           B1           B2           B3         Caa1         Caa2
0.0429505135 0.1073762838 0.0550887021 0.0504201681 0.0112044818 0.0074696545
         Caa3           Ca
0.0065359477 0.0009337068

Group means:
     per_capita_gni growth_gni cpi_inflation current_account_ratio developing
Aaa       10.701823   2.299686      1.934706            4.46847562  0.0000000
Aa1       10.455612   1.375175      2.022328           -2.35952485  0.0000000
Aa2       10.756418   3.418739      2.523113            7.81510309  0.5384615
Aa3       10.513716   3.630720      2.010345            4.06318601  0.4782609
A1         9.991633   3.450273      1.937472           -1.07568690  0.2307692
A2         9.981508   3.119467      3.909772           -0.08986163  0.5774648
A3        10.104397   3.332042      2.275463            4.70031814  0.7254902
Baa1       9.744571   3.925975      2.894226            0.61063949  0.8529412
Baa2       9.740408   4.158209      3.891361           -3.00247851  0.7230769
Baa3       9.309100   3.806952      4.734341           -2.99109964  0.8829787
Ba1        9.164818   4.983525      4.847207           -1.06100671  0.8068182
Ba2        9.255302   3.448699      3.515380           -5.30525379  0.8571429
Ba3        8.822374   4.519311      5.951604           -2.30888340  0.9565217
B1         8.738565   5.010915      7.666306           -5.88582395  0.9913043
B2         8.431667   4.328464      6.801191           -8.15314100  1.0000000
B3         9.124754   2.609322      8.850616           -4.06978240  0.9629630
Caa1       9.212022   1.474215      6.835530           -5.23024595  0.9166667
Caa2       9.096823   3.457874      5.527706           -0.68156319  0.8750000
Caa3       9.182143   2.550586      4.850489           -4.70901495  0.7142857
Ca         9.334326   3.432910     13.912710           -1.99880043  1.0000000

Coefficients of linear discriminants:
                              LD1         LD2         LD3         LD4
per_capita_gni        -1.07267816 -1.14525516 -0.82833066  0.25536432
growth_gni            -0.02308327  0.01180311  0.12208195  0.10785764
cpi_inflation          0.02018583  0.02602544 -0.05493875 -0.13475245
current_account_ratio -0.03810579 -0.02916554  0.09862442 -0.07501589
developing             1.72260781 -2.81900127 -0.67193223  0.30734944
                              LD5
```

```
per_capita_gni           0.65299454
growth_gni               0.22027009
cpi_inflation            0.09777190
current_account_ratio   -0.04109281
developing              -0.07063482

Proportion of trace:
   LD1    LD2    LD3    LD4    LD5
0.8642 0.0897 0.0221 0.0166 0.0075
```

The separation of observations into different subsets by the linear discriminant functions is shown in Figure 18.

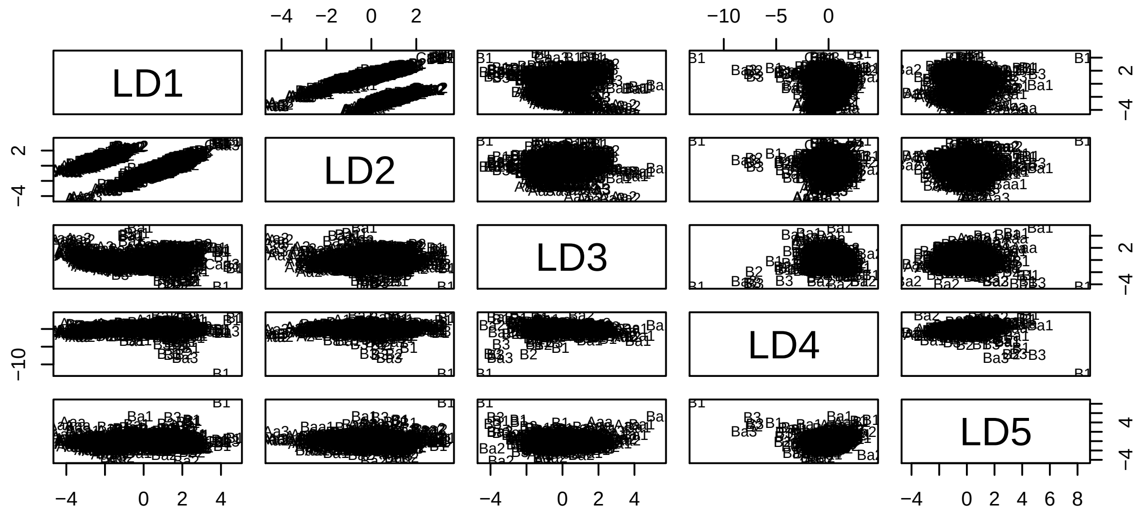

Figure 18: **Moody's Linear Discriminant Functions**

The fitted linear discriminant model is next used to predict the Moody's sovereign credit ratings based upon the randomized test set of indicators for each sovereign. The model calculates the probability of each credit rating category for each sovereign and its associated indicator values. The test set for the Moody's model contains 459 observations covering 89 sovereigns across 20 rating categories. The predicted ratings for the test set are compared to the actual test set ratings in the confusion matrix shown in Table 6. A mosaic plot of the confusion matrix is presented in Figure 19.

2.3.3.2 Standard and Poors Linear Discriminant Analysis Model The Standard and Poors linear discriminant analysis model is trained using a subset comprised of 1031 observations. The observations cover ratings for 91 countries across 20 rating categories. All of the observations in this subset are for developing countries, so the *developing* indicator is excluded from the training data. The summary output from training this model is shown below.

Fitting the LDA model to Standard and Poors observations produces the following results.

```
Call:
lda(rating ~ per_capita_gni + growth_gni + cpi_inflation + current_account_ratio +
    developing, data = sandp_observations_train_data)
```

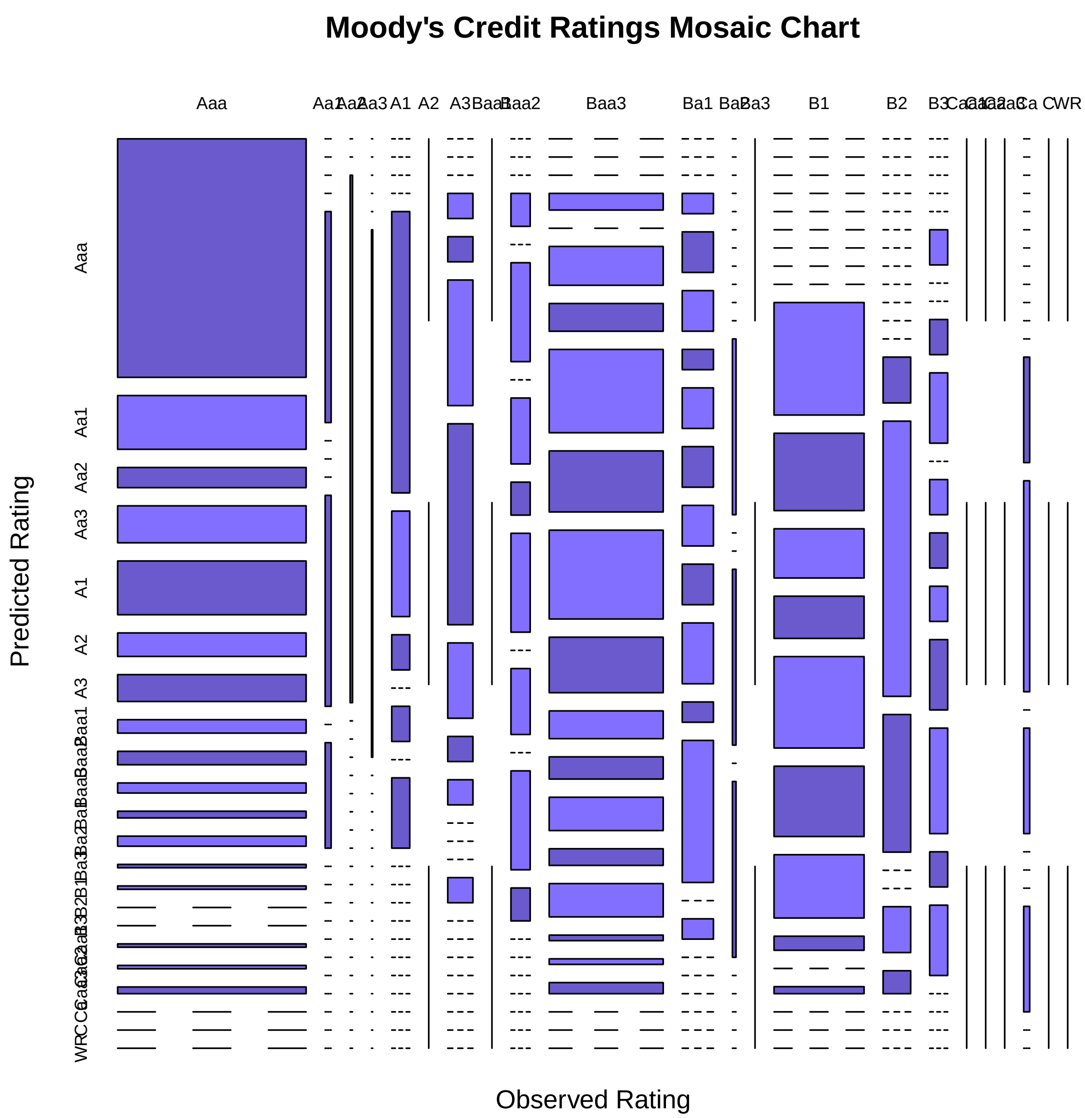

Figure 19: Moody's Linear Discriminant Model Prediction Mosaic

Table 6: Moody's Linear Discriminant Analysis Model Confusion Matrix

Actual Rating	Predicted Rating																					
	Aaa	Aa1	Aa2	Aa3	A1	A2	A3	Baa1	Baa2	Baa3	Ba1	Ba2	Ba3	B1	B2	B3	Caa1	Caa2	Caa3	Ca	C	WR
Aaa	71	16	6	11	16	7	8	4	4	3	2	3	1	1	0	0	1	1	2	0	0	0
Aa1	0	0	0	0	2	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0
Aa2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aa3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A1	0	0	0	0	8	3	1	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0
A2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	0	0	0	1	1	5	8	3	1	1	0	0	0	1	0	0	0	0	0	0	0	0
Baa1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Baa2	0	0	0	1	0	3	0	2	1	3	0	2	0	3	1	0	0	0	0	0	0	0
Baa3	0	0	0	3	0	7	5	15	11	16	10	5	4	6	3	6	1	1	2	0	0	0
Ba1	0	0	0	1	2	2	1	2	2	2	2	3	1	7	0	1	0	0	0	0	0	0
Ba2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0
Ba3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B1	0	0	0	0	0	0	0	0	0	16	11	7	6	13	10	9	2	0	1	0	0	0
B2	0	0	0	0	0	0	0	0	0	0	0	0	2	12	6	0	0	2	1	0	0	0
B3	0	0	0	0	0	1	0	0	1	2	0	1	1	1	2	3	1	2	0	0	0	0
Caa1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Caa2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Caa3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ca	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	1	0	0	0	1	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

```
Prior probabilities of groups:
         AAA          AA+           AA          AA–           A+            A
0.0931134821 0.0407371484 0.0533462658 0.0659553831 0.0320077595 0.0116391853
          A–         BBB+          BBB         BBB–          BB+           BB
0.0465567410 0.0300678952 0.0785645005 0.0902036857 0.0271580989 0.0562560621
         BB–           B+            B           B–         CCC+          CCC
0.1183317168 0.1086323957 0.0814742968 0.0426770126 0.0019398642 0.0058195926
        CCC–           CC
0.0009699321 0.0145489816

Group means:
     per_capita_gni growth_gni cpi_inflation current_account_ratio developing
AAA       10.731083   2.229957     1.7031909             5.3210104  0.0000000
AA+       10.671009   1.447070     1.8833925             0.5500303  0.0000000
AA        10.657856   2.495491     2.2814448             2.6055927  0.2181818
AA–        9.890075   4.630681     2.6183638             2.6652630  0.5441176
A+        10.358381   2.843894     0.9959287             2.2636478  0.1212121
A         10.410382   1.482365     1.7740810            −0.3567010  0.0000000
A–        10.044737   3.241478     3.6076567             1.9973113  0.3958333
BBB+       9.994721   2.639963     2.2124490            −2.3672362  0.7096774
BBB        9.537323   3.705721     4.5467564            −3.2135871  0.7160494
BBB–       9.600936   4.115717     4.3708446            −2.3260448  0.7741935
BB+        9.675618   3.181050     5.4286868            −2.2465455  0.6071429
BB         9.390299   3.133404     5.9919213            −4.8366683  0.8448276
BB–        8.932896   4.282406     5.6210808            −1.7382792  0.9918033
B+         8.396371   5.276848     7.6008025            −4.3180743  0.9732143
B          8.684060   3.917572     5.6783116            −5.2504704  0.9880952
B–         9.031783   3.244583    11.3665706            −7.0245797  0.9545455
CCC+       8.657295  −1.175623     2.1413171           −11.3823058  0.5000000
CCC        9.523749  −3.184051    20.0771427            −4.7257408  0.8333333
CCC–       9.010669   3.084322     2.8227581            −4.9296356  1.0000000
```

```
CC              9.604022    1.414166     6.1271329                -9.9220485   0.3333333

Coefficients of linear discriminants:
                               LD1          LD2          LD3          LD4
per_capita_gni         -0.979002803   1.23151798   0.64331742  -0.082982879
growth_gni             -0.033113551  -0.08990629   0.10290558   0.007222864
cpi_inflation           0.008390196   0.04063622  -0.07902044  -0.103863950
current_account_ratio  -0.038076487  -0.07097948   0.03759587  -0.083950403
developing              1.890823770   1.94294722   1.92215090  -0.308374178
                              LD5
per_capita_gni         -0.39372478
growth_gni             -0.23321302
cpi_inflation          -0.05805566
current_account_ratio   0.04992081
developing              0.45083745

Proportion of trace:
   LD1    LD2    LD3    LD4    LD5
0.8729 0.0579 0.0456 0.0165 0.0071
```

The separation of observations into different subsets by the linear discriminant functions is shown in Figure 20.

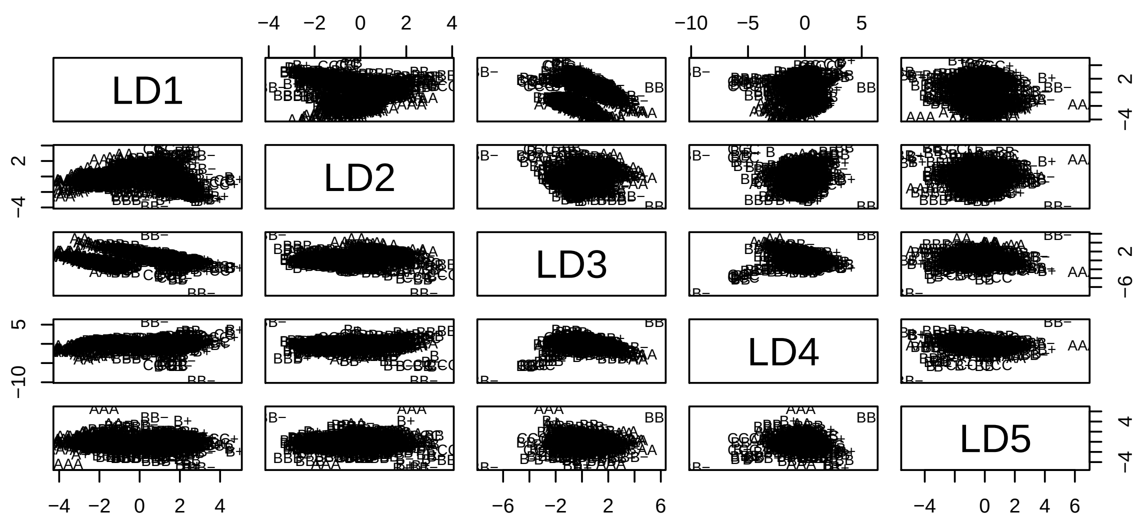

Figure 20: **Standard and Poors Linear Discriminant Functions**

The fitted linear discriminant model is next used to predict the Standard and Poors sovereign credit ratings based upon the randomized test set of indicators for each sovereign. The model calculates the probability of each credit rating category for each sovereign and its associated indicator values. The test set for the Standard and Poors model contains 441 observations covering 89 sovereigns across 20 rating categories. The predicted ratings for the test set are compared to the actual test set ratings in the confusion matrix shown in Table 7. A visualization of the confusion matrix is presented in Figure 21.

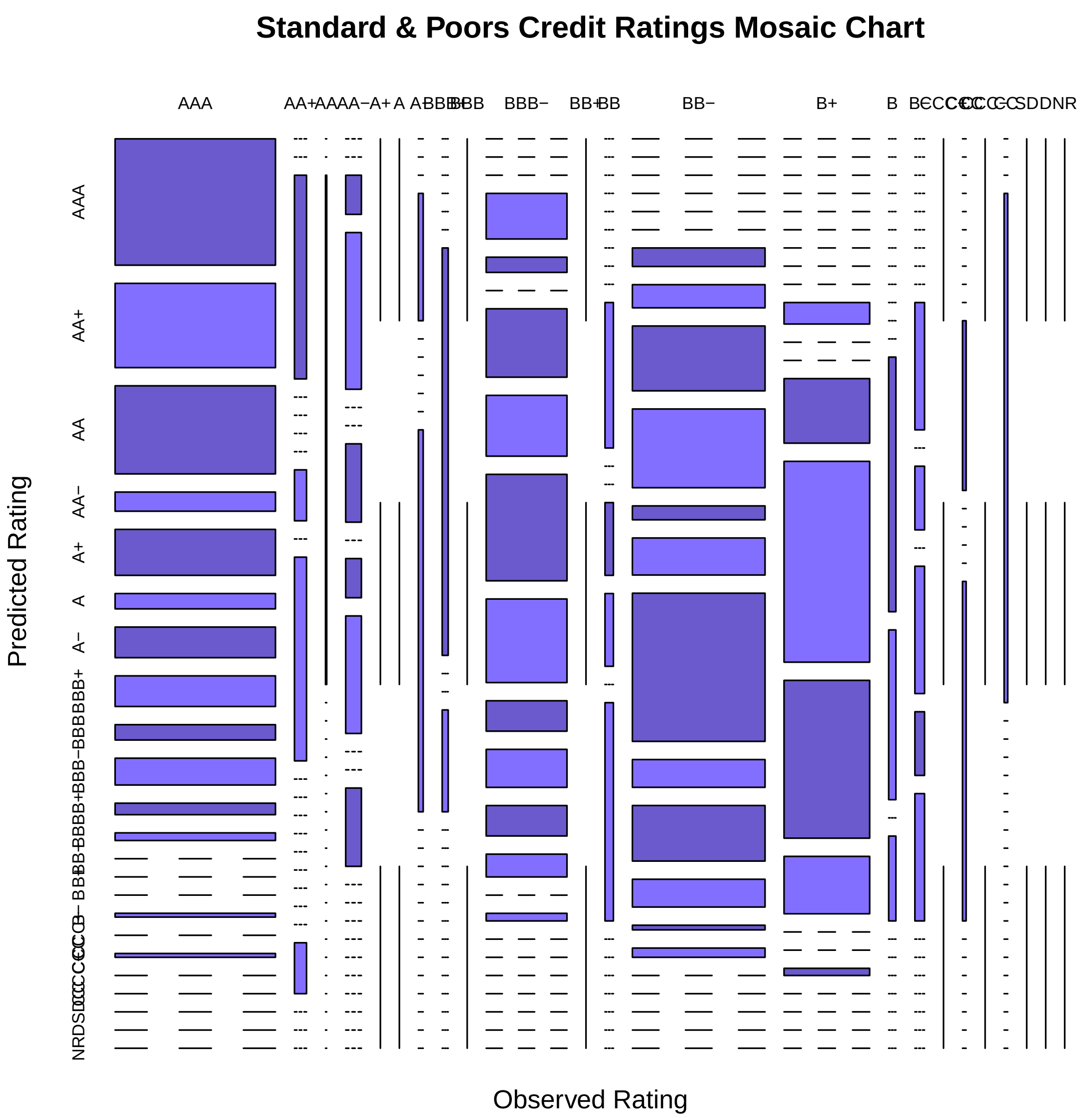

Figure 21: Standard & Poors Linear Discriminant Model Prediction Mosaic

Table 7: Standard and Poors Linear Discriminant Analysis Model Confusion Matrix

Actual Rating	Predicted Rating																					
	AAA	AA+	AA	AA-	A+	A	A-	BBB+	BBB	BBB-	BB+	BB	BB-	B+	B	B-	CCC+	CCC	CCC-	CC	SD	D	NR
AAA	33	22	23	5	12	4	8	8	4	7	3	2	0	0	0	1	0	1	0	0	0	0	0
AA+	0	0	4	0	0	0	0	1	0	4	0	0	0	0	0	0	0	0	0	1	0	0	0
AA	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AA-	0	0	1	4	0	0	2	0	1	3	0	0	2	0	0	0	0	0	0	0	0	0	0
A+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A-	0	0	0	1	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
BBB+	0	0	0	0	0	0	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
BBB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BBB-	0	0	0	6	2	0	9	8	14	11	4	5	4	3	0	1	0	0	0	0	0	0	0
BB+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BB	0	0	0	0	0	0	0	0	0	2	0	0	1	1	0	3	0	0	0	0	0	0	0
BB-	0	0	0	0	0	0	4	5	14	17	3	8	32	6	12	6	1	2	0	0	0	0	0
B+	0	0	0	0	0	0	0	0	0	3	0	0	9	28	22	8	0	0	1	0	0	0	0
B	0	0	0	0	0	0	0	0	0	0	0	0	3	2	0	1	0	0	0	0	0	0	0
B-	0	0	0	0	0	0	0	0	0	2	0	1	0	2	1	2	0	0	0	0	0	0	0
CCC+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCC	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0
CCC-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CC	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

2.3.3.3 Fitch Linear Discriminant Analysis Model

The Fitch linear discriminant analysis model is trained using a subset comprised of 1139 observations. The observations cover ratings for 99 countries across 20 rating categories. All of the observations in this subset are for developing countries, so the *developing* indicator is excluded from the training data. The summary output from training this model is shown below.

Fitting the LDA model to the Fitch observations produces the following results.

```
Call:
lda(rating ~ per_capita_gni + growth_gni + cpi_inflation + current_account_ratio,
    data = fitch_observations_train_data)

Prior probabilities of groups:
         AAA          AA+           AA          AA–           A+            A
0.1308165057 0.0316066725 0.0500438982 0.0158033363 0.0359964881 0.0430201932
          A–         BBB+          BBB         BBB–          BB+           BB
0.0395083406 0.0570676032 0.0649692713 0.0877963126 0.0597014925 0.0491659350
         BB–           B+            B           B–         CCC+          CCC
0.1027216857 0.0684811238 0.0825285338 0.0693590869 0.0061457419 0.0035118525
        CCC–           DD
0.0008779631 0.0008779631

Group means:
     per_capita_gni growth_gni cpi_inflation current_account_ratio
AAA       10.704720  2.0544744      1.871953            4.13425691
AA+       10.449491  1.8911329      2.099128           −1.73508141
AA        10.571061  2.2949477      1.959283            2.74370923
AA–       10.566039  2.2480649      2.845934            2.83503216
A+        10.111173  3.9288258      2.404227            0.67028724
A         10.159095  3.7835234      2.326003            0.63778154
A–         9.980429  4.7686197      2.264139            0.76305921
BBB+       9.818445  3.3448021      3.051300            0.00572473
BBB        9.723708  3.8266683      3.476741           −1.42369728
BBB–       9.381255  3.6184917      4.391949           −2.34274070
BB+        9.271346  3.4167676      4.029788           −2.84065832
```

BB	9.106131	3.9572505	4.750805	−0.97699446
BB−	8.781285	4.9986908	6.379353	−1.80383881
B+	8.410011	5.5092987	9.066712	−5.88951893
B	8.214908	4.6793416	6.703172	−5.67141566
B−	8.560674	3.2624401	8.038927	−7.27885894
CCC+	8.165289	2.9767164	21.419806	−3.21601747
CCC	8.208642	4.3883656	8.152790	−2.86737745
CCC−	8.935904	0.3090582	12.609508	−7.06479572
DD	8.299037	10.6721335	5.301244	−1.19146609

Coefficients of linear discriminants:

	LD1	LD2	LD3	LD4
per_capita_gni	−1.51090667	0.289595647	−0.15594793	0.65258092
growth_gni	−0.01119113	−0.127184513	0.09886437	0.20460529
cpi_inflation	0.04125019	0.141681480	0.05622317	0.06146301
current_account_ratio	−0.01429407	−0.003871553	0.11182156	−0.07641928

Proportion of trace:

LD1	LD2	LD3	LD4
0.9441	0.0265	0.0194	0.0100

The separation of observations into different subsets by the linear discriminant functions is shown in Figure 22.

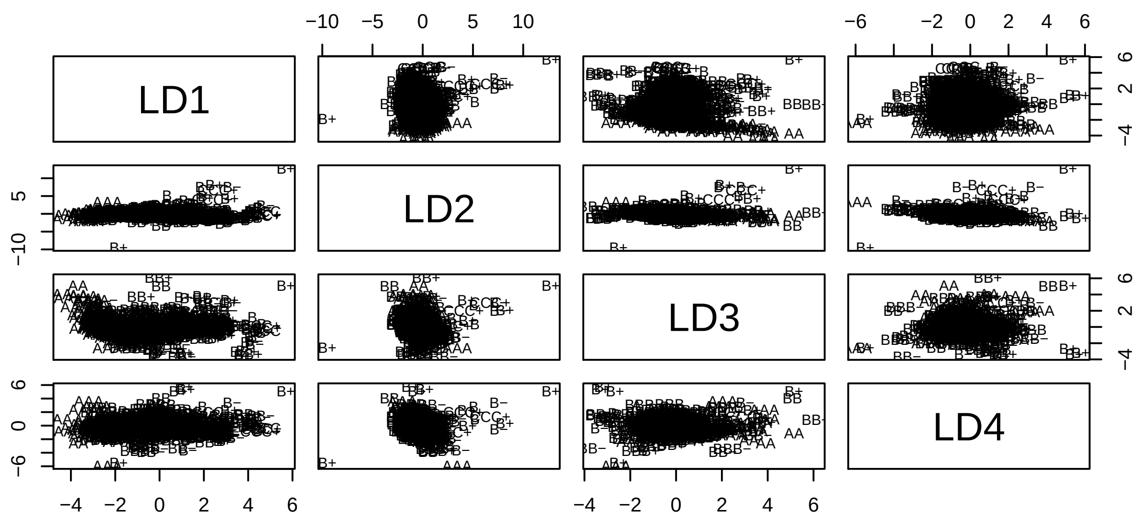

Figure 22: **Fitch Linear Discriminant Functions**

The fitted linear discriminant model is next used to predict the Fitch sovereign credit ratings based upon the randomized test set of indicators for each sovereign. The model calculates the probability of each credit rating category for each sovereign and its associated indicator values. The test set for the Fitch model contains 487 observations covering 99 sovereigns across 20 rating categories. The predicted ratings for the test set are compared to the actual test set ratings in the confusion matrix shown in Table 8. A visualization of the confusion matrix is presented in Figure 23.

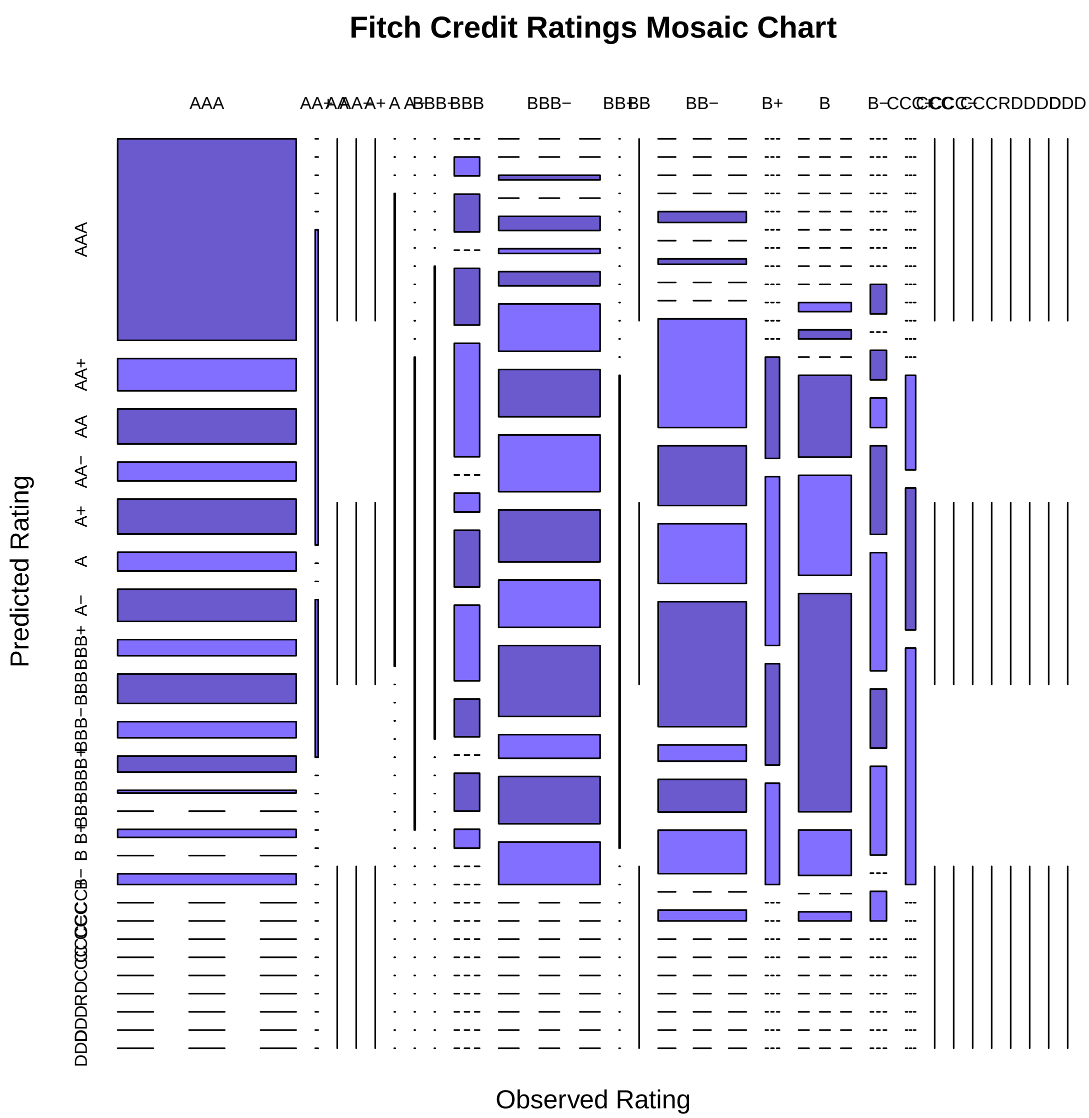

Figure 23: Fitch Linear Discriminant Model Prediction Mosaic

Table 8: Fitch Linear Discriminant Analysis Model Confusion Matrix

Actual Rating	Predicted Rating																								
	AAA	AA+	AA	AA-	A+	A	A-	BBB+	BBB	BBB-	BB+	BB	BB-	B+	B	B-	CCC+	CCC	CCC-	CC	C	RD	D	DD	DDD
AAA	75	12	13	7	13	7	12	6	11	6	6	1	0	3	0	4	0	0	0	0	0	0	0	0	0
AA+	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AA-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A-	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
BBB+	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BBB	0	1	2	0	3	6	0	1	3	4	2	0	2	1	0	0	0	0	0	0	0	0	0	0	0
BBB-	0	0	1	0	3	1	3	10	10	12	11	10	15	5	10	9	0	0	0	0	0	0	0	0	0
BB+	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
BB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BB-	0	0	0	0	2	0	1	0	0	20	11	11	23	3	6	8	0	2	0	0	0	0	0	0	0
B+	0	0	0	0	0	0	0	0	0	0	0	0	3	5	3	3	0	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0	1	1	0	9	11	24	5	0	1	0	0	0	0	0	0	0
B-	0	0	0	0	0	0	0	0	1	0	1	1	3	4	2	3	0	1	0	0	0	0	0	0	0
CCC+	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	5	0	0	0	0	0	0	0	0	0
CCC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCC-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DDD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

2.4 Performance Measurement

Performance of classification models with n categories is typically measured in terms of a confusion matrix $C = [c_{i,j}, i, j = 1, ...n]$, where i denotes the category of a predicted result, j denotes the category of an actual result for an observation, and c_{ij} is the number of observations where the predicted category is i and the actual category is j. In the previous section of this report, Tables 3 to 8 presented confusion matrices for the six credit rating prediction models developed there.

There are several alternative measures for classification model performance. The traditional measure is *Percentage Observed Agreement*, or P_o, which is the proportion of the confusion matrix values where the predicted result is the same as the observed result relative to the sum of all confusion matrix values, i.e.,

$$P_o = (\sum_{i=1}^{n} C_{i,i})/(\sum_{i=1}^{n}\sum_{j=1}^{n} C_{i,j})$$

which is the sum of all counts along the diagonal of the confusion matrix divided by the sum of all counts in all entries in the confusion matrix.

Percentage Observed Agreement is criticized due to its inability to take into account random or expected agreement by chance, which is the proportion of agreement that you would expect two raters to have based simply on chance. Cohen's *kappa* is a commonly used measure of agreement that removes this chance agreement (see Cohen (1960)), because it accounts for the possibility that raters actually guess on at least some variables due to uncertainty. This is accomplished by including *Percentage Chance Agreement* in the measurement of rater agreement. *Percentage Chance Agreement*, or P_e, is the sum of the products of the rows and columns marginal proportions in the confusion matrix, i.e.

$$P_e = \sum_{i=1}^{k} P_{i+} \cdot P_{+i}$$

where the row marginal proportion is

$$P_{i+} = \sum_{j=1}^{n} C_{i,j}/(\sum_{i=1}^{n}\sum_{j=1}^{n} C_{i,j})$$

Table 9: Kappa Statistic Strength of Agreement

Value of kappa	Strength
< 0	Poor
0.01 - 0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 1.00	Almost perfect

and the column marginal proportion is

$$P_{+i} = \sum_{i=1}^{n} C_{i,j} / (\sum_{i=1}^{n} \sum_{j=1}^{n} C_{i,j})$$

Cohen's *kappa* is then calculated as

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

For large sample sizes, the standard error (SE) of κ can be computed as follows (see Fleiss and Everitt)

$$SE_\kappa = \frac{P_e + P_e^2 + \sum_{i=1}^{n} P_{i+} \cdot P_{+i} \cdot (P_{i+} + P_{+i})}{(\sum_{i=1}^{n} \sum_{j=1}^{n} C_{i,j}) \cdot (1 - P_e)^2}$$

and a $100 \cdot (1 - \alpha)$ % confidence interval for κ may be computed using the standard normal distribution as follows:

$$\kappa \pm Z_{\alpha/2} \cdot SE_\kappa$$

In many applications, there is more interest in the magnitude of κ than in the statistical significance of κ. The classifications that have been suggested to interpret the strength of the agreement based on the Cohen's *kappa* value (see Landis and Koch (1977)) are shown in Table 9.

Cohen's *kappa* only counts strict agreement, where the same rating category is assigned by both the rating agency and the rating model. It takes no account of the degree of disagreement, all disagreements are treated equally. This is most appropriate when you have nominal variables, but for ordinal rating scales it may preferable to give different weights to the disagreements depending on their magnitude. *weighted kappa*, also originally developed by Cohen, addresses this issue by using a weighting scheme to take into account the closeness of agreement between rating categories. This is only suitable in the situation where you have ordinal or ranked rating variables, such as credit rating categories.

To compute *weighted kappa*, weights are assigned to each element in the confusion matrix, where the element weights range from 0 to 1, with weight = 1 assigned to all diagonal cells, corresponding to where the observed ratings and the model ratings agree. *weighted kappa* is calculated based on *Percentage Weighted Observed Agreement* and *Percentage Expected Chance Agreement*. *Percentage Weighted Observed Agreement*, or P_o^w, is the sum of weighted proportions across all elements of the confusion matrix, i.e.

$$P_o^w = \sum_{1=1}^{n} \sum_{j=1}^{n} W_{i,j} \cdot P_{i,j}$$

where $W_{i,j}$ is the weight assigned to each element $C_{i,j}$ of the confusion matrix, and

$$P_{i,j} = C_{i,j} / (\sum_{i=1}^{n} \sum_{j=1}^{n} C_{i,j})$$

is the proportion of observations associated with each element $C_{i,j}$ of the confusion matrix. *Percentage Expected Chance Agreement*, or P_e^w, is the sum of the weighted product of rows and columns marginal proportions, i.e.

$$P_e^w = \sum_{1=1}^{n} \sum_{j=1}^{n} W_{i,j} \cdot P_{i+} \cdot P_{+i}$$

where $W_{i,j}$, P_{i+}, and P_{+i} are defined above. *weighted kappa* is then calculated as

$$\kappa^w = \frac{P_o^w - P_e^w}{1 - P_e^w}$$

There are two commonly used weighting schemes for *weighted kappa*: i) linear weights with equal spacing between ratings (see Cicchetti and Allison (1971)), and ii) quadratic weights with spacing proportional to the square of the deviation between ratings (see Fleiss and Cohen (1973)). Linear weights are calculated as $W_{i,j} = 1 - (\mid i - j \mid)/(n-1)$, and quadratic weights are calculated as $W_{i,j} = 1 - (\mid i - j \mid)^2/(n-1)$, where $\mid i - j \mid$ is the distance between rating categories i and j.

The performance of credit rating prediction models can also be measured by additional metrics, including Kendall's W and Spearman's rank correlation coefficient.

Kendall's W (also known as Kendall's coefficient of concordance) is a non-parametric statistic that is also used to measure classification model performance. It is a normalization of the statistic of the Friedman test, and can be used for assessing agreement among different rating methods. Kendall's W ranges from 0 (no agreement) to 1 (complete agreement). If the test statistic W is 1, then all the raters have been unanimous, and each rater has assigned the same order to the ratings. If W is 0, then there is no overall trend of agreement among the raters, and their ratings may be regarded as essentially random. Intermediate values of W indicate a greater or lesser degree of unanimity among the various ratings.

Mathematically, Kendall's W is defined as follows. Suppose that observation i is given the rank $r_{i,j}$ by rater number j, where there are in total n observations and m raters. Then the total rank given to observation i is

$$R_i = \sum_{j=1}^{m} r_{i,j}$$

and the mean value of these total ranks is

$$\overline{R} = \frac{1}{n} \sum_{j=1}^{m} R_i$$

The sum of squared deviations, S, is defined as

$$S = \sum_{i=1}^{n} (R_i - \overline{R})^2$$

and then Kendall's W is defined as

$$W = \frac{12 \cdot S}{m^2 \cdot (n^3 - n)}$$

In the case of complete ranks, a commonly used significance test for W against a null hypothesis of no agreement (i.e. random rankings) is given by

$$\chi^2 = m \cdot (n-1) \cdot W$$

where the test statistic takes a chi-squared distribution with $(n-1)$ degrees of freedom.

Kendall's W is linearly related to the mean value of the Spearman's rank correlation coefficients between all $\begin{pmatrix} m \\ 2 \end{pmatrix}$ possible pairs of rankings between raters as follows

$$\overline{r_s} = \frac{mW - 1}{m - 1}$$

Table 10: Moody's Ordinal Forest Model Performance

Method	Value	Stat.Name	Stat.Value	P.value
Percentage agreement (Tolerance=0)	40.0871460		NA	NA
Cohen's Kappa for 2 Raters (Weights: unweighted)	0.3424172	z	24.817327	0
Cohen's Kappa for 2 Raters (Weights: equal)	0.3773650	z	13.621323	0
Cohen's Kappa for 2 Raters (Weights: squared)	0.3641277	z	8.005076	0
Kendall's coefficient of concordance W	0.7190344	Chisq(458)	658.635479	0
Spearman rank correllation coefficient r	0.4380687	z	9.744914	NA

Table 11: Standard and Poors Ordinal Forest Model Performance

Method	Value	Stat.Name	Stat.Value	P.value
Percentage agreement (Tolerance=0)	39.9092971		NA	NA
Cohen's Kappa for 2 Raters (Weights: unweighted)	0.3471886	z	25.39256	0
Cohen's Kappa for 2 Raters (Weights: equal)	0.4857897	z	16.34219	0
Cohen's Kappa for 2 Raters (Weights: squared)	0.5069665	z	10.68065	0
Kendall's coefficient of concordance W	0.7474774	Chisq(440)	657.78011	0
Spearman rank correllation coefficient r	0.4949548	z	11.02974	NA

Thus, Spearman's rank correlation coefficient $\overline{r_s}$ can be computed from Kendall's W using this relationship. The statistical significance of $\overline{r_s}$ is obtained from the z-score for $\overline{r_s}$, which is

$$z = \sqrt{\frac{(n-3)}{1.06}} \cdot F(\overline{r}_s)$$

where $F(r_s) = arctanh(r_s)$ is the Fisher transformation of r_s.

In the remainder of this section, the performance metrics outlined above are applied to each of the credit rating prediction models to determine the predictive power of the models.

2.4.1 Ordinal Forest Model Performance

This section contains the model performance results for the ordinal forest models developed and applied in the previous section.

2.4.1.1 Moody's Ordinal Forest Model Performance The performance results for the Moody's ordinal forest model are presented in Table 10. This model has modest *Percentage Agreement*, consistent with the dispersion shown in the mosaic chart in Figure 15. The model also shows fair strength of agreement between predicted and observed ratings based on κ values, and the z-statistics for the κ values indicate that the predicted and observed ratings are drawn from the same populations. Kendall's coefficient of concordance W indicates a high level of agreement between predicted and observed ratings, and the χ^2 statistic and associated p value show that the null hypothesis that there is no agreement between predicted and observed ratings can be rejected. The Spearman rank correlation coefficient $\overline{r_s}$ indicates a moderate positive relationship between predicted and observed rating ranking.

2.4.1.2 Standard and Poors Ordinal Forest Model Performance The performance results for the Standard and Poors ordinal forest model are presented in Table 11. This model has modest *Percentage Agreement*, consistent with the dispersion shown in the mosaic chart in Figure 16. The model also shows moderate strength of agreement between predicted and observed ratings based on κ values, and the z-statistics for the κ values indicate that the predicted and observed ratings are drawn from the same populations.

Table 12: Fitch Ordinal Forest Model Performance

Method	Value	Stat.Name	Stat.Value	P.value
Percentage agreement (Tolerance=0)	41.8891170		NA	NA
Cohen's Kappa for 2 Raters (Weights: unweighted)	0.3596422	z	26.01663	0
Cohen's Kappa for 2 Raters (Weights: equal)	0.4744336	z	17.25500	0
Cohen's Kappa for 2 Raters (Weights: squared)	0.4540351	z	10.02281	0
Kendall's coefficient of concordance W	0.7465168	Chisq(486)	725.61433	0
Spearman rank correllation coefficient r	0.4930336	z	11.54016	NA

Table 13: Moody's Linear Discriminant Analysis Model Performance

Method	Value	Stat.Name	Stat.Value	P.value
Percentage agreement (Tolerance=0)	28.7581699		NA	NA
Cohen's Kappa for 2 Raters (Weights: unweighted)	0.2057395	z	14.354633	0
Cohen's Kappa for 2 Raters (Weights: equal)	0.3280558	z	11.590253	0
Cohen's Kappa for 2 Raters (Weights: squared)	0.3855227	z	8.384859	0
Kendall's coefficient of concordance W	0.7268909	Chisq(458)	665.832040	0
Spearman rank correllation coefficient r	0.4537817	z	10.151729	NA

Kendall's coefficient of concordance W indicates a high level of agreement between predicted and observed ratings, and the χ^2 statistic and associated p value show that the null hypothesis that there is no agreement between predicted and observed ratings can be rejected. The Spearman rank correlation coefficient $\overline{r_s}$ indicates a moderate positive relationship between predicted and observed rating ranking.

2.4.1.3 Fitch Ordinal Forest Model Performance The performance results for the Fitch ordinal forest model are presented in Table 12. This model has modest *Percentage Agreement*, consistent with the dispersion shown in the mosaic chart in Figure 17. The model also shows moderate strength of agreement between predicted and observed ratings based on κ values, and the z-statistics for the κ values indicate that the predicted and observed ratings are drawn from the same populations. Kendall's coefficient of concordance W indicates a high level of agreement between predicted and observed ratings, and the χ^2 statistic and associated p value show that the null hypothesis that there is no agreement between predicted and observed ratings can be rejected. The Spearman rank correlation coefficient $\overline{r_s}$ indicates a moderate positive relationship between predicted and observed rating ranking.

2.4.2 Linear Discriminant Model Analysis Performance

This section contains the model performance results for the linear discriminant models developed and applied previously.

2.4.2.1 Moody's Linear Discriminant Analysis Model Performance The performance results for the Moody's linear discriminant analysis model are presented in Table 13. This model has modest *Percentage Agreement*, consistent with the dispersion shown in the mosaic chart in Figure 19. The model also shows fair strength of agreement between predicted and observed ratings based on κ values, and the z-statistics for the κ values indicate that the predicted and observed ratings are drawn from the same populations. Kendall's coefficient of concordance W indicates a high level of agreement between predicted and observed ratings, and the χ^2 statistic and associated p value show that the null hypothesis that there is no agreement between predicted and observed ratings can be rejected. The Spearman rank correlation coefficient $\overline{r_s}$ indicates a moderate positive relationship between predicted and observed rating ranking.

Table 14: Standard and Poors Linear Discriminant Analysis Model Performance

Method	Value	Stat.Name	Stat.Value	P.value
Percentage agreement (Tolerance=0)	25.1700680		NA	NA
Cohen's Kappa for 2 Raters (Weights: unweighted)	0.1764426	z	12.60413	0.00e+00
Cohen's Kappa for 2 Raters (Weights: equal)	0.3018815	z	10.94828	0.00e+00
Cohen's Kappa for 2 Raters (Weights: squared)	0.2727659	z	6.04477	0.00e+00
Kendall's coefficient of concordance W	0.6430507	Chisq(440)	565.88465	4.48e-05
Spearman rank correllation coefficient r	0.2861015	z	5.98269	NA

Table 15: Fitch Linear Discriminant Analysis Model Performance

Method	Value	Stat.Name	Stat.Value	P.value
Percentage agreement (Tolerance=0)	29.9794661		NA	NA
Cohen's Kappa for 2 Raters (Weights: unweighted)	0.2108882	z	14.388672	0
Cohen's Kappa for 2 Raters (Weights: equal)	0.3254973	z	11.917864	0
Cohen's Kappa for 2 Raters (Weights: squared)	0.3606759	z	8.310505	0
Kendall's coefficient of concordance W	0.7042132	Chisq(486)	684.495199	0
Spearman rank correllation coefficient r	0.4084263	z	9.267876	NA

2.4.2.2 Standard and Poors Linear Discriminant Analysis Model Performance The performance results for the Standard and Poors linear discriminant analysis model are presented in Table 14. This model has modest *Percentage Agreement*, consistent with the dispersion shown in the mosaic chart in Figure 21. The model also shows fair strength of agreement between predicted and observed ratings based on κ values, and the z-statistics for the κ values indicate that the predicted and observed ratings are drawn from the same populations. Kendall's coefficient of concordance W indicates a high level of agreement between predicted and observed ratings, and the χ^2 statistic and associated p value show that the null hypothesis that there is no agreement between predicted and observed ratings can be rejected. The Spearman rank correlation coefficient $\overline{r_s}$ indicates a low positive relationship between predicted and observed rating ranking.

2.4.2.3 Fitch Linear Discriminant Analysis Model Performance The performance results for the Fitch linear discriminant analysis model are presented in Table 15. This model has modest *Percentage Agreement*, consistent with the dispersion shown in the mosaic chart in Figure 23. The model also shows fair strength of agreement between predicted and observed ratings based on κ values, and the z-statistics for the κ values indicate that the predicted and observed ratings are drawn from the same populations. Kendall's coefficient of concordance W indicates a high level of agreement between predicted and observed ratings, and the χ^2 statistic and associated p value show that the null hypothesis that there is no agreement between predicted and observed ratings can be rejected. The Spearman rank correlation coefficient $\overline{r_s}$ indicates a moderate positive relationship between predicted and observed rating ranking.

3 Summary of Results and Conclusion

This report explored whether sovereign credit ratings by rating agencies can be predicted using current and historical economic and other data about each sovereign. Some economic and financial indicators that previous research indicated were relevant in predicting sovereign credit ratings were identified as potentially relevant to answering this question. Data for sovereign credit ratings and the economic and financial indicators was extracted from internet web sites to support the investigation, and a randomly-selected subset of this data was used to train two machine learning models for each rating agency. Each machine learning model was then used to predict sovereign credit ratings using a randomly-selected complementary subset of the data, and these predicted ratings for the complementaty subset were compared to observed ratings for this subset using confusion matrices and mosaic plots. Finally, various performance metrics for ordinal classification models were calculated to quantify the performance of each machine learning model.

The model performance results presented in the previous section show that both the ordinal forest and linear discriminant analysis models produce modest to moderate predictive performance measured by ordinal classification metrics such as Cohen κ, Kendall W, and average Spearman rank correlation $\overline{r_s}$. The models have a relatively low ability to exactly predict the same rating category as assigned by the corresponding rating agency, e.g. the Moody's ordinal forest model frequently predicts another rating category when the rating agency assigns an "A1" rating to a sovereign. On the other hand, the models appear to have a strong capability to produce sovereign rating predictions that rank different sovereigns consistently when measured, e.g. by Kendall's coefficient of concordance W. Thus, the reliability of the sovereign credit rating predictions produced by the two machine learning models is somewhat mixed and depends upon whether exact predictions or consistent rankings of sovereigns are most useful.

Undoubtedly with further research better models with higher predictive performance can be developed. There are several areas where further research is likely to be productive. A large collection of current and historical economic, financial, and other indicators for sovereigns is available from a variety of sources, including the World Bank, the International Monetary Fund, the Bank for International Settlements, and others. Some of these indicators may have a stronger statistical relationship to sovereign credit ratings than the indicators used in this report, and their incorporation into a credit rating classification model might improve predictive performance. In exploring these additional indicators, the extent of sovereign coverage will be an important issue, particularly if the coverage is uneven accross sovereigns with different rating categories, as was the case with *external_debt_ratio* in this report. A related issue is whether gaps in sovereign coverage can be closed by finding supplementary sources for the missing data, e.g., an additional source for *external_debt_ratio* data, or by some interpolation or similar scheme that could generate the missing data.

Alternative machine learning models might also produce better predictive results. Since this report only applies ordinal forest and linear discriminant analysis models to the sovereign credit rating prediction problem, it only investigates a small portion of currently available machine learning models. Other ordinal classification models are available, and these models might provide a better means of predicting sovereign credit ratings. Some of these alternative models were investigated during the research for this report, but they were not pursued for various reasons.

References

Cantor, Richard, and Frank Packer, 1996, Determinants and impact of sovereign credit ratings, *FRBNY Economic Policy Review*, pp 37–53.

Cicchetti, Dominic, and Truett Allison, 1971, A new procedure for assessing reliability of scoring eeg sleep recordings, *American Journal of EEG Technology* 3, 101–110.

Cohen, Jacob, 1960, A coefficient of agreement for nominal scales, *Educational and Psychological Measurement* XX, 37–46.

Eaton, Jonathan, 1996, Sovereign debt, repudiation, and credit terms, *International Journal of Finance and Economics* 1, pp 25–36.

Fleiss, Jacob Cohen, Joseph, and Brian Everitt, Large sample standard errors of kappa and weighted kappa, *Psychological Bulletin* 72, 327–332.

Fleiss, Joseph, and Jacob Cohen, 1973, The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, *Educational and Psychological Measurement* 33, 613–619.

Grothe, Magdalena, 2013, Market pricing of credit rating signals, *European Central Bank Working Paper Series.*

Hornung, Roman, 2020, Ordinal forests, *Journal of Classification* 37, pp 4–17.

Landis, JR, and GG Koch, 1977, The measurement of observer agreement for categorical data, *Biometrics* 33, 159–174.

Ozler, Sule, 1991, Evolution of credit terms: An empirical examination of commercial bank lending to developing countries, *Journal of Development Economics* 38, pp 79–97.

Reusens, Peter, and Christophe Croux, 2017, Sovereign credit rating determinants: The impact of the european debt crisis, *KU Leuven Faculty of Economics and Business Working Paper 1615*.

Valle, Cecilia Téllez, and José Luis Martín Marín, 2005, Sovereign credit ratings and their determination by the rating agencies, *Investment Management and Financial Innovations*, pp 159–173.