Climate Finance and Green Bond Evolution
Informing Policy with Machine Learning Text Analytics

Amber Jaycocks

This document was submitted as a dissertation in October 2019 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of Elvira N. Loredo (Chair), Walter L. Perry, and Thomas J. Sullivan.

This dissertation was partially supported by the Cazier Sustainability Dissertation Award.
To all my mentors and supporters,

To my colleagues and committee,

And to my friends and family,

Thank you.
Abstract

This research considers the case of financing the response to climate change, also known as climate finance, with emphasis on the labelled green bond market; climate finance is an exemplar of policy challenges in which private sector engagement is integral. This research aims to understand the evolution of themes associated with climate finance and green bonds to identify opportunities to enhance public-private cooperation and facilitate policymaking. The research exists at the intersection of policy analysis, climate science, environmental finance, and machine learning, and makes novel contributions across the data, method, and policy areas. The research employs topic modeling approaches in conjunction with sentiment and qualitative analyses on unstructured data to represent discourse surrounding (1) climate finance, and (2) green and climate bonds. The topic models aid in discovering interpretable, low-dimensional subspaces from corporas extracted from LexisNexis using a crowd-sourced search strategy. In the case of understanding the evolution of climate finance, dominant topics in climate finance news headlines are analyzed temporally and geographically. This is done using an unsupervised probabilistic generative topic model, Latent Dirichlet Allocation (LDA), along with an automated process for model selection and hyperparameter optimization. The LDA climate finance results indicate that topics representing the mobilization of capital and collective action are becoming more prevalent and are regarded more positively in recent years—suggesting a strong case for enhanced public-private partnerships. Labelled green bond opportunities are identified through news and blog articles that correspond to green bond sectors. Sector-specific topics are identified with Correlation Explanation (CorEx), an information-theoretic approach to topic modeling. In the semi-supervised version of CorEx, domain knowledge about the sectors is incorporated via topic anchors. The green bond topic results demonstrate the prevalence of certain investment areas, increasing interest that remains historically high, and market opportunities that may exist by focusing on industry and building sectors and consolidating water and pollution-control sectors. Overall, investments in market structuring and frameworks emphasizing monitoring, verification, and reporting will strengthen transparency and consistency, which will leverage the momentum in climate finance and assists in scaling up the green bond market. Furthermore, the methods and approaches herein have broad applicability to other complex policy settings.
### Table Of Contents

Abstract ii  
Table Of Contents iii  
Figures vii  
Tables xii  
Abbreviations xiv  
Executive Summary xvii

**1 Motivation And Policy Context: Climate Change, Climate Finance, And Green Bonds**  
1.1 Using Private Capital For Public Benefit 2  
1.2 The Global Issue Of Climate Change And Its Causes 4  
1.3 Two Facets Of Climate Change Action 6  
1.4 Major International Negotiations And The Birth Of Climate Finance 7  
1.5 Capital Cost And Flow Requirements For Adaptation And Mitigation 11  
1.5.1 Current Estimates Of Adaptation And Mitigation Costs 13  
1.5.2 Estimating Climate Finance Flows 16  
1.6 Climate Finance Sources, Intermediaries, And Instruments 17  
1.6.1 Alternative Sources And Private Sector Opportunities For Climate Finance 21  
1.7 The Emergence Of The Green Bond Market 22  
1.7.1 Green Bonds And The World Bank 23  
1.7.2 The History Of Bonds For Social Impact 28  
1.7.3 The History Of Green Bonds 30  
1.7.4 Green Bond Market Limitations 33  
1.8 Statement Of Research Objectives 34

**2 An Economic Framework: Socially Responsible Engagement For Climate Financing**  
2.1 Understanding The Concept Of Public Goods 38  
2.2 Challenges Of Modern Environmental Public Goods 39  
2.3 Understanding Corporate Social Responsibility 40  
2.3.1 Private Voluntary Action And The Triple Bottom Line 42  
2.3.2 The Debate Over CSR’s Private Voluntary Action 43  
2.3.3 Considering Scenarios Where CSR Could Be Beneficial 46  
2.4 Climate Finance As An Example Of CSR 49
2.5 Regulatory Considerations Associated With Private Engagement
   2.5.1 Policy Implications For CSR Addressing Climate Change

3 Corpora Creation: A Crowd-Sourced Search Strategy With NLP
   3.1 Keyphrase Relevance For Information Extraction - A Search Strategy
   3.2 Google Trends For Keyphrases Representing Collective Behavior
   3.3 Google Trends For Concept Exploration
   3.4 Creating Consolidated Keyphrases
   3.5 Sub-Corpora Extraction Using Keyphrases - Data For Analyses
   3.6 Sub-Corpora Processing Prior To Modeling
      3.6.1 Climate Finance Sub-Corpus
      3.6.2 Climate And Green Bond Sub-Corpus

4 Methods: Unsupervised And Semi-Supervised Topic Modeling, Hyperparameter Optimization, Evaluation, And Sentiment
   4.1 Topic Modeling Introduction
   4.2 Static Topic Models—Comparisons And Technical Details
      4.2.1 Latent Semantic Analysis
      4.2.2 Probabilistic Latent Semantic Analysis And Indexing
      4.2.3 Latent Dirichlet Allocation, LDA
      4.2.4 Correlated Topic Model
   4.3 LDA Topic Model Evaluation And Hyperparameter Selection
   4.4 Semi-Supervised Topic Modeling Approach
      4.4.1 Correlation Explanation, CorEx, Topic Model Details
   4.5 CorEx Topic And Anchor Selection
   4.6 Using Sentiment Along With Topic Models

5 The Evolution Of Climate Finance Using Topic Modeling
   5.1 Policy Background In Brief
   5.2 Research Questions And Approach
   5.3 Summary Of Climate Finance News Headline Data
   5.4 LDA Model Results
      5.4.1 Model Selection
      5.4.2 LDA Six-Topic Model Output
   5.5 Policy Implications
   5.6 Contributions And Discussion Of Private Sector Engagement
      5.6.1 Enhancing Private Sector Engagement On Climate Finance
      5.6.2 Public And Private Roles In Addressing Climate Change
6 Green-Climate Bond Opportunities Using Anchored Topic Modeling 126

6.1 Green Bond Background In Brief 126
   6.1.1 Green Bond Market Growth 127
   6.1.2 Labelled Green Bonds And Use-Of-Proceeds 130

6.2 Motivation And Research Questions 132

6.3 Green And Climate Bond Data 134
   6.3.1 Summary Of Green Bond Article Data 134
   6.3.2 Labelled Green Bond Investment Dollars 135

6.4 Modeling Approach 137

6.5 Results 140

6.6 Policy Considerations And Recommendations 146

7 Model Extensions And Future Work: Word Embeddings, Dynamic Topic Models, And Stance With Sentiment 153

7.1 Broad Applicability Of Text Summarization 153

7.2 Analytic Extensions 154
   7.2.1 Embeddings As Inputs 155
   7.2.2 Embeddings With Topic Models 156
   7.2.3 Dynamic Evolutionary Topic Models 157
   7.2.4 Other Topic Model Choices 159
   7.2.5 Sentiment And Stance 160

7.3 Additional Future Research 160

8 Concluding Discussion: Human Behavior, Ethics, And Social Change 163

8.1 The Manifestation Of Behavioral Economics 163

8.2 The Persistence Of Corporate Social Responsibility 166

8.3 The Ethical Dilemma Of Defining Responsible Behaviors 167
   8.3.1 Towards Economic Modeling More Reflective Of Human Behavior 168
   8.3.2 A Trend Towards Personalized Value-Based Investing And Its Implications 169

8.4 Expansion Of Social Metrics 169
   8.4.1 Strengthening Standards, Monitoring, And Verification 170

8.5 CSR As The Potential Preferred Avenue For Social Change 173

8.6 Concluding Remarks 173

Appendices 175

A: World Economic Forum Global Risks 176
B: Greenhouse Gas Emissions By Country And Industry 178
C: Conference Of The Parties 180
D: Parties And Observers From The UNFCCC
E: Growth In Green Bonds
F: World Bank Project Life Cycle
G: Details Of Belsey And Ghatak’s Economic Model
H: Twitter Related Terms
I: Common Topic Modeling Implementations
J: Topic Model Selection Using VEM Versus Gibbs Sampling
K: Rank Ordered Regions For Climate Finance Corpus
L: Additional Topic Model Evaluations
M: Topic Model Interactive Visualization
N: Sentiment Metric Variations
O: Explanation of Bonds - A Fixed Income Instrument
P: Climate Bonds Taxonomy
Q: Types Of Green Bonds
R: Emotion Category For Green Bond Corpus By Year

Citations
## Figures

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>CPI’s 2017 compilation of sources and intermediaries of public climate finance.</td>
<td>18</td>
</tr>
<tr>
<td>1.2</td>
<td>CPI’s 2017 compilation of sources and intermediaries of private climate finance.</td>
<td>19</td>
</tr>
<tr>
<td>1.3</td>
<td>CPI’s 2017 breakdown of climate finance flows by instrument in billions of USD.</td>
<td>20</td>
</tr>
<tr>
<td>1.4</td>
<td>Tenor, or duration, of certain bond classifications.</td>
<td>24</td>
</tr>
<tr>
<td>1.5</td>
<td>The World Bank project cycle.</td>
<td>27</td>
</tr>
<tr>
<td>2.1</td>
<td>Traditional economic delineation of types of goods.</td>
<td>38</td>
</tr>
<tr>
<td>3.1</td>
<td>High-level schematic of the keyphrase identification (i.e., search strategy) and sub-corpora extraction.</td>
<td>54</td>
</tr>
<tr>
<td>3.2</td>
<td>Google Trend interest score over time for searches: “climate change”, “green bonds”, and “environmental social and governance”.</td>
<td>57</td>
</tr>
<tr>
<td>3.3</td>
<td>Regional exploration of Google Trends.</td>
<td>57</td>
</tr>
<tr>
<td>3.4</td>
<td>The same time-series as shown in Figure 3.2, for “climate finance”, “environmental social and governance”, and “green bond” search interest.</td>
<td>58</td>
</tr>
<tr>
<td>3.5</td>
<td>Green bond term and related terms interest scores over time from Google Trends.</td>
<td>59</td>
</tr>
<tr>
<td>3.6</td>
<td>Climate finance term and related terms interest scores over time from Google Trends.</td>
<td>60</td>
</tr>
<tr>
<td>3.7</td>
<td>High-level schematic of the sub-corpora extraction and post-processing.</td>
<td>63</td>
</tr>
<tr>
<td>3.8</td>
<td>Bigram terms concatenated for analysis.</td>
<td>65</td>
</tr>
<tr>
<td>3.9</td>
<td>Graphical relationship of words with the term climate.</td>
<td>66</td>
</tr>
<tr>
<td>3.10</td>
<td>Example headlines that are converted to lowercase and lemmatized.</td>
<td>66</td>
</tr>
<tr>
<td>3.11</td>
<td>Green and climate bond articles (both news and blog) from 2010 onward where data could be assigned.</td>
<td>67</td>
</tr>
<tr>
<td>3.12</td>
<td>Green and climate bond articles by source.</td>
<td>68</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.13</td>
<td>Example of extracted article in raw HTML format and clean, processed version.</td>
<td>68</td>
</tr>
<tr>
<td>3.14</td>
<td>Example of extracted documents with the article contents, date, stemmed text, and lemmatized text.</td>
<td>69</td>
</tr>
<tr>
<td>4.1</td>
<td>High-level description of the data extraction and corresponding analysis.</td>
<td>71</td>
</tr>
<tr>
<td>4.2</td>
<td>Heatchart where every column corresponds to a document id and each row to a word.</td>
<td>75</td>
</tr>
<tr>
<td>4.3</td>
<td>Graphical representation of pLSA.</td>
<td>77</td>
</tr>
<tr>
<td>4.4</td>
<td>Graphical representation of LDA.</td>
<td>80</td>
</tr>
<tr>
<td>4.5</td>
<td>Graphical representation of correlated topic modeling modified from Lafferty and Blei (2006).</td>
<td>82</td>
</tr>
<tr>
<td>4.6</td>
<td>Example of obtaining the intersecting terms to determine sentiment from a document.</td>
<td>93</td>
</tr>
<tr>
<td>5.1</td>
<td>Distribution of the LexisNexis “Climate Finance” sub-corpus documents (news headlines) by year of publication.</td>
<td>98</td>
</tr>
<tr>
<td>5.2</td>
<td>Count of documents in climate finance corpus over time by month from 2010 to 2018.</td>
<td>99</td>
</tr>
<tr>
<td>5.3</td>
<td>Top publications from the sub-corpus originating from LexisNexis with the consolidated list of “climate finance” keyphrases.</td>
<td>100</td>
</tr>
<tr>
<td>5.4</td>
<td>Country and region rankings for climate finance news sub-corpus.</td>
<td>101</td>
</tr>
<tr>
<td>5.5</td>
<td>Grid-search with constant per-term topic distribution concentration hyperparameter, varying the per topic document distribution concentration hyperparameter and topic number.</td>
<td>103</td>
</tr>
<tr>
<td>5.6</td>
<td>Selection of topic number and parameters values through tuning.</td>
<td>104</td>
</tr>
<tr>
<td>5.7</td>
<td>Metric comparison for optimal and base parameter cases.</td>
<td>105</td>
</tr>
<tr>
<td>5.8</td>
<td>Boxplot out-of-sample comparison of topic models across a chosen K value using perplexity.</td>
<td>106</td>
</tr>
<tr>
<td>5.9</td>
<td>Hierarchical clustering of the topics from the six and nine topic models.</td>
<td>107</td>
</tr>
<tr>
<td>5.10</td>
<td>Heatchart displaying the similarity topics from the six and nine topic models.</td>
<td>108</td>
</tr>
<tr>
<td>5.11</td>
<td>Heatchart displaying the similarity of topics from the six LDA topic model.</td>
<td>110</td>
</tr>
<tr>
<td>5.12</td>
<td>LOESS smoothed average per-document topic proportions.</td>
<td>111</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.13</td>
<td>Percent of headlines with specific geographic labels corresponding to the specific topic.</td>
<td>114</td>
</tr>
<tr>
<td>5.14</td>
<td>LOESS smoothed average annual sentiment.</td>
<td>117</td>
</tr>
<tr>
<td>6.1</td>
<td>Public sector growth in the green bond market.</td>
<td>128</td>
</tr>
<tr>
<td>6.2</td>
<td>Recent green bond market growth is attributed to private institutions.</td>
<td>129</td>
</tr>
<tr>
<td>6.3</td>
<td>Geographic green bond market growth.</td>
<td>130</td>
</tr>
<tr>
<td>6.4</td>
<td>Labelled bonds by category.</td>
<td>131</td>
</tr>
<tr>
<td>6.5</td>
<td>Aggregated monthly counts for news and blog data from Green Bond sub-corpus with smoothed trend line.</td>
<td>134</td>
</tr>
<tr>
<td>6.6</td>
<td>Aggregated annual counts for news and blog data from Green Bond sub-corpus with smoothed trend line.</td>
<td>135</td>
</tr>
<tr>
<td>6.7</td>
<td>Aggregate annual values from 2013 to 2017.</td>
<td>136</td>
</tr>
<tr>
<td>6.8</td>
<td>Aggregate annual dollar values by sector for labeled green from</td>
<td>137</td>
</tr>
<tr>
<td>6.9</td>
<td>Illustrative example of the green bond sector investments areas that represent the latent variables, topics.</td>
<td>138</td>
</tr>
<tr>
<td>6.10</td>
<td>Green bond sector-specific anchor words used for CorEx.</td>
<td>139</td>
</tr>
<tr>
<td>6.11</td>
<td>LOESS smoothed average monthly sentiment across metrics for the green bond corpus.</td>
<td>140</td>
</tr>
<tr>
<td>6.12</td>
<td>NRC emotion categories for the entire green bond corpus from years 2009 to 2018.</td>
<td>141</td>
</tr>
<tr>
<td>6.13</td>
<td>LOESS smoothed average monthly counts of articles corresponding to the green bond investment sectors.</td>
<td>142</td>
</tr>
<tr>
<td>6.14</td>
<td>Contemporaneous annual green bond issuance and article topic counts within the energy sector.</td>
<td>143</td>
</tr>
<tr>
<td>6.15</td>
<td>Contemporaneous annual green bond issuance and article topic counts within the energy sector.</td>
<td>143</td>
</tr>
<tr>
<td>6.16</td>
<td>LOESS smoothed average monthly sentiment from articles corresponding to the green bond sectors using the Bing metric.</td>
<td>144</td>
</tr>
<tr>
<td>6.17</td>
<td>Contemporaneous annual green bond issuance and article topic counts corresponding to the industry &amp; building and land use sectors.</td>
<td>145</td>
</tr>
<tr>
<td>6.18</td>
<td>Contemporaneous annual green bond issuance and article topic counts corresponding to the water and waste &amp; pollution sectors.</td>
<td>146</td>
</tr>
<tr>
<td>Section</td>
<td>Text</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>6.19</td>
<td>Green bond deals are growing in number.</td>
<td></td>
</tr>
<tr>
<td>6.20</td>
<td>Liquidity is hindered by more small deals, especially in the developing market.</td>
<td></td>
</tr>
<tr>
<td>6.21</td>
<td>Recent growth has been fueled by the Asia-Pacific and Latin America regions.</td>
<td></td>
</tr>
<tr>
<td>6.22</td>
<td>The largest markets tend to have higher levels of reporting.</td>
<td></td>
</tr>
<tr>
<td>A.1</td>
<td>WEF Global Risks 2016</td>
<td></td>
</tr>
<tr>
<td>A.2</td>
<td>WEF Global Risks 2018</td>
<td></td>
</tr>
<tr>
<td>B.1</td>
<td>World’s greenhouse gas emissions from the World Resources Institute based upon the CAIT Climate Data Explorer.</td>
<td></td>
</tr>
<tr>
<td>B.2</td>
<td>Historical carbon dioxide emissions.</td>
<td></td>
</tr>
<tr>
<td>D.1</td>
<td>Annex I countries as of 2018.</td>
<td></td>
</tr>
<tr>
<td>E.1</td>
<td>Evolution of green bonds according to a European Investment Bank presentation based on data from JP Morgan.</td>
<td></td>
</tr>
<tr>
<td>H.1</td>
<td>Twitter hashtags related to #climatechange - based upon a sample.</td>
<td></td>
</tr>
<tr>
<td>H.2</td>
<td>Twitter hashtags related to #greenbonds - based upon a sample.</td>
<td></td>
</tr>
<tr>
<td>H.3</td>
<td>Twitter hashtags related to #ESG - based upon a sample.</td>
<td></td>
</tr>
<tr>
<td>H.4</td>
<td>Twitter hashtags related to #climatefinance - based upon a sample.</td>
<td></td>
</tr>
<tr>
<td>J.1</td>
<td>Topic model selection with Gibbs sampling and variational expectation maximization (VEM).</td>
<td></td>
</tr>
<tr>
<td>L.1</td>
<td>Hyperparameter optimization using the CaoJuan and Deveaud metric and a fixed per-term topic distribution parameter.</td>
<td></td>
</tr>
<tr>
<td>L.2</td>
<td>In-sample rank-ordering demonstrates consistent beta value or 0.01 and K of 6 with alpha between 6 and 8.</td>
<td></td>
</tr>
<tr>
<td>L.3</td>
<td>Selection of topic number and parameter values through gridsearch.</td>
<td></td>
</tr>
<tr>
<td>L.4</td>
<td>Optimal versus base hyperparameter comparison across K values.</td>
<td></td>
</tr>
<tr>
<td>L.5</td>
<td>Out-of-sample comparison using perplexity topic numbers between 4 and 11.</td>
<td></td>
</tr>
<tr>
<td>L.6</td>
<td>Topic model distributions using $K = 6$ with other candidate concentration values (alpha and beta).</td>
<td></td>
</tr>
<tr>
<td>M.1</td>
<td>Screenshot of the interactive visualization of the intertopic distance and...</td>
<td></td>
</tr>
</tbody>
</table>
top terms across the entire corpus.

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.2</td>
<td>Screenshot of the interactive visualization of the top terms for topic 1.</td>
<td>201</td>
</tr>
<tr>
<td>N.1</td>
<td>Sentiment analysis for climate finance topics using Syuzhet metric.</td>
<td>202</td>
</tr>
<tr>
<td>N.2</td>
<td>Sentiment analysis for climate finance topics using Afinn metric.</td>
<td>202</td>
</tr>
<tr>
<td>N.3</td>
<td>Sentiment analysis for climate finance topics using Bing metric.</td>
<td>203</td>
</tr>
<tr>
<td>N.4</td>
<td>Sentiment analysis for climate finance topics using NRC metric.</td>
<td>203</td>
</tr>
<tr>
<td>P.1</td>
<td>The taxonomy developed by Climate Bonds Initiative and used globally.</td>
<td>206</td>
</tr>
<tr>
<td>R.1</td>
<td>Emotion categories for the green bond corpus for each year.</td>
<td>208</td>
</tr>
</tbody>
</table>
Table Description

<table>
<thead>
<tr>
<th>Number</th>
<th>Table Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Key climate negotiation and climate finance milestones.</td>
<td>7</td>
</tr>
<tr>
<td>1.2</td>
<td>Summary of climate finance investment timeline.</td>
<td>10</td>
</tr>
<tr>
<td>1.3</td>
<td>Estimated annual climate finance needs for mitigation.</td>
<td>14</td>
</tr>
<tr>
<td>1.4</td>
<td>Estimated annual climate finance needs for adaptation.</td>
<td>14</td>
</tr>
<tr>
<td>1.5</td>
<td>Examples of World Bank mitigation and adaptation projects.</td>
<td>25</td>
</tr>
<tr>
<td>1.6</td>
<td>A list of Green Bond segments on stock exchanges as catalogued by Climate Bonds Initiative.</td>
<td>27</td>
</tr>
<tr>
<td>1.7</td>
<td>Snapshot of total US financial assets held by certain financial intermediaries.</td>
<td>31</td>
</tr>
<tr>
<td>3.1</td>
<td>Related keyphrases (Google search queries) for “green bonds”, “environmental social and governance”, and “climate finance” identified by Google Trends.</td>
<td>59</td>
</tr>
<tr>
<td>3.2</td>
<td>Consolidated set of keyphrases to be passed to LexisNexis database for information extraction.</td>
<td>60</td>
</tr>
<tr>
<td>3.3</td>
<td>Ten most common bigrams with the term climate.</td>
<td>64</td>
</tr>
<tr>
<td>4.1</td>
<td>The strengths, limitations, and computational requirements of several common topic models as summarized by S.Lee, Song, and Kim (2010). Some language is modified for clarify.</td>
<td>74</td>
</tr>
<tr>
<td>4.2</td>
<td>Comparison of two topic models that guide the model to learn topics of specific interest.</td>
<td>88</td>
</tr>
<tr>
<td>5.1</td>
<td>The six topics, brief descriptions, key example terms, and top terms for each.</td>
<td>109</td>
</tr>
<tr>
<td>6.1</td>
<td>Highlighted differences in topic modelling approaches used on the climate finance sub-corpus and the green and climate sub-corpus.</td>
<td>138</td>
</tr>
<tr>
<td>7.1</td>
<td>Sovereign Bond issuance by country as of 2018.</td>
<td>162</td>
</tr>
<tr>
<td>C.1</td>
<td>The location, session, and start date for each of the Conference of the Parties from 2010 to 2018.</td>
<td>180</td>
</tr>
<tr>
<td>K.1</td>
<td>Climate finance article counts ranked by geographic region.</td>
<td>194</td>
</tr>
<tr>
<td>L.1</td>
<td>Optimal alpha parameter for given metrics and topic number.</td>
<td>198</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------------------------</td>
<td>----</td>
</tr>
<tr>
<td>Q.1</td>
<td>Types of green bonds and their key differences.</td>
<td>207</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>ABS</td>
<td>Asset Backed Security</td>
<td></td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
<td></td>
</tr>
<tr>
<td>ASEAN</td>
<td>Association of Southeast Asian Nations</td>
<td></td>
</tr>
<tr>
<td>BOW</td>
<td>Bag-of-Words</td>
<td></td>
</tr>
<tr>
<td>CBI</td>
<td>Climate Bonds Initiative</td>
<td></td>
</tr>
<tr>
<td>CBOW</td>
<td>Common Bag-of-Words</td>
<td></td>
</tr>
<tr>
<td>CDM</td>
<td>Clean Development Mechanism</td>
<td></td>
</tr>
<tr>
<td>CoP</td>
<td>Conference of Parties</td>
<td></td>
</tr>
<tr>
<td>CorEx</td>
<td>Correlation Explanation</td>
<td></td>
</tr>
<tr>
<td>CP3</td>
<td>Climate Public-Private Partnership</td>
<td></td>
</tr>
<tr>
<td>CPI</td>
<td>Climate Policy Initiative</td>
<td></td>
</tr>
<tr>
<td>CSR</td>
<td>Corporate Social Responsibility</td>
<td></td>
</tr>
<tr>
<td>CSV</td>
<td>Comma-Separated Value</td>
<td></td>
</tr>
<tr>
<td>CTM</td>
<td>Correlated Topic Model</td>
<td></td>
</tr>
<tr>
<td>DFI</td>
<td>Development Finance Institution</td>
<td></td>
</tr>
<tr>
<td>DTM</td>
<td>Dynamic Topic Model</td>
<td></td>
</tr>
<tr>
<td>FSC</td>
<td>Forestry Sustainability Council</td>
<td></td>
</tr>
<tr>
<td>EIB</td>
<td>European Investment Bank</td>
<td></td>
</tr>
<tr>
<td>EIT</td>
<td>Economy In Transition</td>
<td></td>
</tr>
<tr>
<td>EM</td>
<td>Expectation-Maximization</td>
<td></td>
</tr>
<tr>
<td>ESG</td>
<td>Environmental, Social, and Governance</td>
<td></td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
<td></td>
</tr>
<tr>
<td>GAVI</td>
<td>Global Alliance for Vaccines and Immunization</td>
<td></td>
</tr>
<tr>
<td>GBP</td>
<td>Green Bond Principles</td>
<td></td>
</tr>
<tr>
<td>GCF</td>
<td>Green Climate Fund</td>
<td></td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
<td></td>
</tr>
<tr>
<td>GEF</td>
<td>Global Environmental Facility</td>
<td></td>
</tr>
<tr>
<td>GHG</td>
<td>Greenhouse Gas</td>
<td></td>
</tr>
<tr>
<td>GloVe</td>
<td>Global Vectors for Word Representation</td>
<td></td>
</tr>
<tr>
<td>HDP</td>
<td>Hierarchical Dirichlet Process</td>
<td></td>
</tr>
<tr>
<td>HTML</td>
<td>Hypertext Markup Language</td>
<td></td>
</tr>
<tr>
<td>IBRD</td>
<td>International Bank of Reconstruction and Development</td>
<td></td>
</tr>
<tr>
<td>IDA</td>
<td>International Development Association</td>
<td></td>
</tr>
<tr>
<td>IE</td>
<td>Information Extraction</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>IEA</td>
<td>International Energy Agency</td>
<td></td>
</tr>
<tr>
<td>IG</td>
<td>Information Gain</td>
<td></td>
</tr>
<tr>
<td>IFC</td>
<td>International Finance Corporation</td>
<td></td>
</tr>
<tr>
<td>IFFIm</td>
<td>International Finance Facility for Immunization</td>
<td></td>
</tr>
<tr>
<td>INDC</td>
<td>Intended Nationally Determined Contribution</td>
<td></td>
</tr>
<tr>
<td>IPCC</td>
<td>Intergovernmental Panel on Climate Change</td>
<td></td>
</tr>
<tr>
<td>JI</td>
<td>Joint Implementation</td>
<td></td>
</tr>
<tr>
<td>KL</td>
<td>Kullback-Leibler, as in divergence</td>
<td></td>
</tr>
<tr>
<td>LDA</td>
<td>Latent Dirichlet Allocation</td>
<td></td>
</tr>
<tr>
<td>LDC</td>
<td>Least Developed Countries</td>
<td></td>
</tr>
<tr>
<td>LDCF</td>
<td>Least Developed Country Fund</td>
<td></td>
</tr>
<tr>
<td>LOESS</td>
<td>Locally Estimated Scatterplot Smoothing</td>
<td></td>
</tr>
<tr>
<td>LSA</td>
<td>Latent Semantic Analysis</td>
<td></td>
</tr>
<tr>
<td>LSI</td>
<td>Latent Semantic Indexing</td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
<td></td>
</tr>
<tr>
<td>MTTM</td>
<td>Multiscale Topic Tomography Models</td>
<td></td>
</tr>
<tr>
<td>NLP</td>
<td>Natural Language Processing</td>
<td></td>
</tr>
<tr>
<td>NLU</td>
<td>Natural Language Understanding</td>
<td></td>
</tr>
<tr>
<td>ODI</td>
<td>Overseas Development Institute</td>
<td></td>
</tr>
<tr>
<td>OECD</td>
<td>Organization for Economic Cooperation and Development</td>
<td></td>
</tr>
<tr>
<td>pLSA</td>
<td>probabilistic Latent Semantic Analysis</td>
<td></td>
</tr>
<tr>
<td>POS</td>
<td>Part-of-Speech</td>
<td></td>
</tr>
<tr>
<td>SCCF</td>
<td>Special Climate Change Fund</td>
<td></td>
</tr>
<tr>
<td>SEB</td>
<td>Skandinaviska Enskilda Banken</td>
<td></td>
</tr>
<tr>
<td>SFI</td>
<td>Sustainable Forestry Initiative</td>
<td></td>
</tr>
<tr>
<td>SIB</td>
<td>Social Impact Bond</td>
<td></td>
</tr>
<tr>
<td>sLDA</td>
<td>supervised Latent Dirichlet Allocation</td>
<td></td>
</tr>
<tr>
<td>SRI</td>
<td>Socially Responsible Investing/ment</td>
<td></td>
</tr>
<tr>
<td>STEM</td>
<td>Science Technology Engineering and Mathematics</td>
<td></td>
</tr>
<tr>
<td>STM</td>
<td>Structural Topic Model</td>
<td></td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
<td></td>
</tr>
<tr>
<td>TC</td>
<td>Total Correlation</td>
<td></td>
</tr>
<tr>
<td>TF</td>
<td>Term Frequency</td>
<td></td>
</tr>
<tr>
<td>TF-IDF</td>
<td>Term Frequency Inverse Document Frequency</td>
<td></td>
</tr>
<tr>
<td>ToT</td>
<td>Topics over Time</td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
<td></td>
</tr>
<tr>
<td>UN</td>
<td>United Nations</td>
<td></td>
</tr>
<tr>
<td>UNFCCC</td>
<td>United Nations Framework Convention on Climate Change</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------</td>
<td></td>
</tr>
<tr>
<td>US or USA</td>
<td>United States of America</td>
<td></td>
</tr>
<tr>
<td>USD</td>
<td>United States Dollars</td>
<td></td>
</tr>
<tr>
<td>VCM</td>
<td>Voluntary Contribution Model</td>
<td></td>
</tr>
<tr>
<td>VEM</td>
<td>Variational Expectation Maximization</td>
<td></td>
</tr>
<tr>
<td>WEF</td>
<td>World Economic Forum</td>
<td></td>
</tr>
</tbody>
</table>
Executive Summary

Modern policy challenges and solutions often share some underlying commonalities. They seek to use novel data sources and methods, particularly from the digital realm. They address more complex transnational and multi-lateral issues than those seen in decades past in recognition of an interconnected world. They also consider the role of the private sector as well as traditional policymaking actors. This research explores a particular intersection of these three trends—examining the role of private sector actors in mobilizing finance to combat global climate change through the analysis of online textual information.

Climate change is a complex and multifaceted problem with existential risks and no simple solutions. Actors combat climate change by both mitigating, or reducing emissions and other causes of climate change, and adapting to reduce vulnerability to its impacts, such as building seawalls to withstand rising seas and stronger storms. Climate change impacts a myriad of industries and institutions, many of which are capital-intensive. Substantial amounts of funding are needed to transition to a climate-sustainable economy and adapt to the impacts of climate change.

A stable global climate, like many environmental benefits, is considered a public good in economics. As the benefits are shared publicly, there is little incentive for actors to bear the private cost of maintaining it. These dilemmas are usually addressed by government action. However, as the global climate transcends national boundaries, such action requires close cooperation, which is difficult. The economic incentive for the most capable nations is muted because historically rich nations tend to emit more yet be less impacted. Despite more than 25 years of international climate discussions, progress on adaptation and mitigation has been limited, with many experts concluding it is insufficient.

Likewise, long-standing efforts to enhance “climate finance”—or funding with the intention to address climate change—has underwhelmed. Observers note that estimates of the amount of funding required exceeds the amount provided by the international community, resulting in a funding gap. About 87 percent of the financing to combat climate change is spent in the country where it was raised, typically serving more affluent areas and not reaching the places where it can provide the most good. Similarly, more than 95 percent of funding goes towards mitigation, almost certainly ignoring many opportunities for high-utility adaptation projects.

In the past five to ten years, however, climate finance has grown rapidly, thanks in large part to greater interest from the private sector. This activity broadly parallels a trend towards socially responsible investing, in which private actors seek to maximize more than just monetary return on investment. The concept that private actors can and should play a role in providing public goods is not without controversy from both the economic and policy fields, which question both the utility and motive of such actions.
This research does not review questions pertaining to earth science or optimizing national strategies to combat climate change. Nor does this research address the ethical arguments for and against private participation in public policy. It begins with the assumptions that (1) climate change is a far-reaching and severe risk, impacting a public good, and should be addressed, and (2) private engagement is integral to climate finance and will continue to actively support climate finance given the scope of the challenge.

The overarching policy objectives require the public and private sectors to optimally cooperate, to position financing solutions appropriately, and to ensure climate change funding arrives where and when it is needed. Given the assumptions and policy motivations, the research explores two key research questions:

1. What insights can be gained about the evolution of climate finance to better facilitate policy making and private sector engagement?
2. Can opportunities for public-private cooperation be identified within green bond project financing?

To answer these questions, the research aims to understand conversational trends from news and blogs using machine learning, namely topic modeling and associated text analytics. Unstructured data, particularly text, continues to grow at an exponential pace. The potential insights to be gained from using machine learning with this data are vast. Online media that represents the global conversation is particularly useful to this policy debate because it addresses climate finance and instruments like the green bond from multiple sources and angles. This research applies unsupervised and semi-supervised topic modeling along with sentiment analysis to understand the evolution of the climate finance landscape, identify potential opportunities in particular sectors, and draw policy recommendations.

Particular attention is paid to the methodology, including information extraction and model selection processes, such that the research approach could be extended to other domains facing similar questions. Relevant textual data that includes news and blogs are extracted from the LexisNexis database using crowd-sourced search terms. This results in two corpus subsets that independently focus on climate finance and green bonds. Topic modeling is a machine learning approach that associates words and articles with latent ideas called topics. Sets of articles are classified into groups based on themes and commonalities found within the text. A probabilistic approach called Latent Dirichlet Allocation (LDA) is used for the climate finance corpus, while an information-theoretic approach called Correlation Explanation (CorEx) is used for the green bond corpus. LDA tends to identify dominant themes across the entire corpus, whereas CorEx offers the potential to identify less prominent topics guided by human input. Through hyperparameter optimization and evaluation of the LDA
model, geographical and temporal trends are deduced. With the CorEx model, temporal trends and contemporaneous project investment topics are surfaced.

Within the climate finance corpus, spikes in article frequency tend to appear around the Conference of Parties, the annual international meeting for climate change negotiations. The frequency of articles related to green bonds was observed to increase during the past decade. Both trends were anticipated, but a useful confirmation that the search strategy and information extraction approach was successful.

Six major climate finance topics were identified from the climate finance corpus using the LDA model. Applying human expertise to the identified topics, the topics are presented as assigning responsibility, reduction plans, collective action, multi-lateral (but not UN-led) action, addressing the causes of climate change, and mobilizing capital. The intensity of the mobilizing capital topic rose significantly from 2010 to 2018, becoming the dominant topic in news coverage and eclipsing the focus from the addressing causes topic, which dominated in 2010. This is likely attributable to the shift in dynamics at play during these crucial years. Consensus around climate change, its urgency, and its causes became much more commonplace in the global public square in the intervening years, and the question of how sufficient capital would be raised and where it would come from became a more pressing matter. Sentiment around the topics of mobilizing capital and collective action also trended positive during the same eight-year time frame, probably for similar reasons. There is strong indication of geographical differences of the headline frequency of the six identified topics, with several regions focusing more on mobilizing capital and others on responsibility.

Upon review of these geographical, temporal, and topical trends, policy insights emerge. The rising interest and positive sentiment towards mobilizing capital suggests an opportunity to advance private sector projects that engage a receptive public. Noting the positive sentiment towards collective action, projects are likely to be more successful when promoting partnerships with international organizations. Lastly, geographical differences should be scrutinized for strategic opportunities and geopolitical implications. Many of these insights could lead to future work examining predictive and causal relationships.

The CorEx topics from the green bond corpus were anchored on sector words associated with underlying green bond investments. These included energy, transport, water, waste, land use/marine, industry, and buildings. From 2010 to 2018, all topics increased in frequency with a peak between 2016 and 2017 and remain historically high. Energy appeared the most frequently of all topics through 2017, only to be eclipsed by industry in the following year. Interestingly, the amount of money issued to energy green bond projects was roughly four times that of industry green bonds in 2018. This suggests a substantial underinvestment in industry green bond projects given the attention and interest paid to them. Investments in water projects are funded frequently relative to their article counts, while waste and pollution projects are funded relatively infrequently. Given the overlap between water and waste and pollution sector classifications and certification criteria, consideration should be given to merge the
two to capitalize on the popular interest in water and reduce the additional certification costs. This could reduce friction and grow the market.

Policy insights specific to the green bond market are also apparent following this analysis. Strategic approaches to market structuring and sector identification and certification, in close coordination with subject matter experts, can identify emergent and underserved areas. The maintenance of transparency in green bonds, particularly in monitoring, reporting, and verifying investments to avoid perceptions of “greenwashing”, are key in sustaining interest over time. Lastly, leveraging user preferences for socially responsible investment can be applied in a framework for the investment community to accomplish both climate financing and broader sustainable development goals.

This research bridges the domains of policy, climate science, environmental finance, and machine learning. The research contributes to each in a number of ways; it illustrates a novel use of topic modeling methodology, informs a complex policy issue, and identifies opportunities for engagement. From the data side, it provides an approach to reduce a corpus to relevant information by leveraging crowd-sourced behavior and incorporating domain-specific custom dictionaries. In terms of methods, it demonstrates the value of unsupervised and semi-supervised topic modeling, and also provides a framework for model selection and hyperparameter optimization. In regard to policy, it connects climate finance to corporate social responsibility (CSR) in an economics setting and illuminates key climate finance relationships.

The analytic approach discussed in this research is applicable to other complex policy problems. By exploring previously untapped, unstructured data through topic modeling, policy stakeholders can review how products and policies are being received. Furthermore, stakeholders can discern when and how to improve associated actions like policy adjustments, investment timing and allocation, and strategic messaging. Given the complexity of the global and interconnected problems faced by policymakers today, private actor participation and CSR is likely to become more commonplace. Publicly available data such as media corporas can help surface trends, gauge interest, and identify opportunities for public-private partnerships. Further explorations of topic modeling applied to climate finance questions and additional policy conundrums will add valuable insights for the policy and finance communities.
1. Motivation And Policy Context: Climate Change, Climate Finance, And Green Bonds

Climate change is a classic example of a complex problem. It is global in nature, involving multiple sectors and industries, many of which are interlinked, and is resistant to simple solutions involving a subset of actors. It is also a classic example of a wicked\(^1\) problem—where policy makers suffer from incomplete information (e.g., earth science, technological progress, and emissions trajectories), contradictory requirements (e.g., satisfying human needs for food, water, and energy while reducing emissions), and frequent changes to the state of affairs (e.g., social, economic, and technological changes) (Hulme 2009). Finally, the consequences of the problem are existential, with human activity holding the potential within this century to push earth systems to a point where basic functions of human civilization (e.g., large-scale agriculture) become infeasible (Xu and Ramanathan 2017).

At the same time, climate change presents a multi-trillion-dollar investment opportunity to develop global infrastructure that is both carbon neutral and resilient to changing weather patterns during roughly the next decade (Blackrock 2016; New Climate Economy 2014; Tom Kerr, et al 2016). However, as there is no internationally agreed upon price on greenhouse gas (GHG) emissions, there is no market price signal that these are investment opportunities per se. Further, many of the areas which would benefit most substantially from a low carbon economy and protection from climate-linked damages are societies that can least afford it (Althor, Watson, and Fuller 2016). Thus, there is an active and concerted effort to direct funds towards these areas. The funding of efforts to facilitate climate mitigation—reducing the causes—and climate adaptation—reducing its potential effects—are commonly referred to as climate finance. Given the frequent framing of climate change as an international policy issue predominantly addressed within the context of United Nations (UN) negotiations, the financing of climate change adaptation and mitigation efforts was originally conceived as the domain of national governments. However, there is concern that current rates of investment have been and continue to be inadequate to address the challenges of climate change in a timely manner, resulting in what observers have called the financing gap.

Increasing awareness of the gap, along with increased understanding of the economic risks of climate change, the cost advantages of green technologies like renewables, and the encouragement of national governments and other climate advocates, major international and private financial entities now comprise a majority of financial flows (Padraig, Clark, and Meattle 2018). Climate finance presents a path to accelerate these investments and diversify the investment sources. On the other hand, analysts are keen to acknowledge the challenges of scale climate finance still faces. For example,

---

\(^1\) Meaning resistance to resolution, rather than actors being evil.
annual energy sector investments still dwarf investment in hydrocarbon energy by roughly two to one (Padraig, Clark, and Meattle 2018). Should climate finance move beyond a niche market, reaching a level of investor interest where private participation is able to operate independent of the support of public actors, there are public policy implications around how best to scale resources and cooperate with private actor engagement in this market.

The promotion of sustainable development-defined as economic development that does not exhaust environmental resources—is broader than climate finance but still a related topic, as the two often promote similar ideals. Climate finance and sustainable development initiatives are not mutually exclusive and often synergistic; many sustainable development policies have climate adaptation or mitigation co-benefits, and vice versa. The link between climate change and sustainability policies suggest the two could be tackled together, with climate resiliency development satisfying the objectives of both domains.

This research applies models and methods from machine learning (ML), and in particular natural language processing (NLP), to illustrate the evolution of topics associated with climate finance and to reveal broad thematic changes that inform areas of interest and potential opportunities in climate finance. Particular attention is paid to the evolution of the green bond market given its scale. Anticipating these evolving interests could help institutions and policy makers create climate finance instruments that simultaneously appeal to private investors, help bridge the financing gap, and adequately disseminate financing. While this research is specifically focused on climate finance, these methods can be applied to provide insight to a number of investment, marketing, or policy questions (e.g., ethically-sourced supply chains, infrastructure, and health care) with a sufficiently-sized corpus of text (news, social media, forums, etc.). Furthermore, identifying how to better engage private investors to complement public efforts can extend to most public policy areas where funding is judged to be inadequate. These broader implications are discussed in the conclusions of Chapter 8.

1.1 Using Private Capital For Public Benefit

Global public goods or shared goods, those which are enjoyed by all in the world and cannot be made exclusive, are starkly complex. Examples include the health of the ocean and harmonized technology standards. The traditional economic framing of public goods typically leaves their provision solely to governments to steward lest the forces of the private sector result in market failure. However, public-private partnerships are often leveraged for their subject matter expertise and financing in the case of global public goods. (Brinkerhoff and Brinkerhoff 2011) The same is true for the global public good of climate change; the myriad stakeholders and scale of the challenge have incentivized governments to pivot from a predominantly UN-based approach towards one reliant on public-private engagement for both finance and expertise. These private-public partnerships underscore a policy challenge common with global public goods: how can effective mechanisms be created to encourage private capital investment at the appropriate scope and scale? In addition to this being a challenging
endeavor, data that details private industry’s role is sparse and discussions about the appropriate role of private industry is often politically and socially sensitive (Shirley, Walsh, and Patrick 2000).

Markets that intrinsically value public good provision or public bad curtailment have evolved in recent decades. From food vendors touting the social utility of local- and organically-grown produce to major corporations promoting that their power is generated entirely from renewable energy, private institutions have taken concrete steps contrary to the short-term profit motive in order to respond to perceived value from what are often intangible gains, or to differentiate themselves from competitors.

In this era of proliferating digital information, the wealth of data offers observers of the financial world the opportunity to identify and respond to evolving investor preferences, beyond the traditional measures of rate of return and risk and more to the “triple bottom line” metrics that align with investor goals and values as well as short-term profits. Using machine learning and quantitative analytics, we can identify new signals to help private actors determine when and how to enter new markets more personalized to certain investors’ ethics and values.

In the case of climate finance, recent history shows governing bodies concluded the scope and scale of the problem are beyond the grasp of the governments alone. Following this recognition, innovative financing mechanisms were developed. Public actors are increasingly interested in leveraging private capital as they recognize the gap between required and available climate finance funding. Private sector investors—whether individual retail investors, private equity investors including venture capitalists, or larger institutional investors like pension funds, insurance companies, or sovereign wealth funds—have assets of trillions of dollars available for investment. These global, regional, and local financial institutions have the capacity to allocate capital and financial services to reduce this gap.

Green bonds represent a successful development in the realm of climate finance that incentivized private capital to serve a public good. The structure of these investment vehicles mobilized additional capital flows to supplement those of public institutions. They offer a mechanism for private institutions to play a role when governments are unwilling or unable to do so (examples of imperfect governments). Green bonds have demonstrated responsiveness to investor interests and demands. If green bonds can be structured such that they remain responsive to dynamic investor interests, it represents a way in which public-private partnerships can be sustainable and ensure longer-term viability of these markets (World Bank 2018).

The concepts of climate finance and green bonds are closely linked to socially responsible investing (SRI), impact investing, and sustainable investing—overlapping terms which all fall under the broad umbrella of corporate social responsibility. These concepts, and the arguments surrounding private actions intended for social good, are discussed in further detail in Chapter 2. In all these cases, the objectives are not only financial return but also positive social or environmental impact. Efforts to attain longer-term sustainability, however the term is defined and measured, are what link these endeavors. For some firms, it may mean ensuring an ethical supply chain to avoid a longer-term risk of public backlash against business practices. For others, it may mean following forestry protection
practices to ensure its resources remain viable in perpetuity. For others still, it may mean investing in climate finance to counter the drivers and impacts of climate change.

Corporate social responsibility offers the ability to meld the responsibility of public goods provision and public bad curtailment from public actors to private ones. Yet the research community has not concluded if the concept has an appropriate role to play in the provision of global public goods, and what that role should look like if so. Setting aside the moral and ethical component of ceding to public good protection and provision to private actors instead of government ones, this research looks to explore where private activity could focus to more effectively maximize a global public good. While this research focuses on climate finance and the use of a quantitative machine learning approach, the ability to better understand how policy makers and financial institutions can work together to construct mechanisms that sustainably engage, and position private capital extends to other policy realms.

1.2 The Global Issue Of Climate Change And Its Causes

The issue of climate change has increased in prominence on the international agenda over the last three decades. Domiciles are more familiar with the scope and challenge of the issue, as well as the inherent requirement that solutions require multi-lateral collective action. Most international organizations, governments, and scientists have galvanized towards the conclusion that anthropogenic activity² has exacerbated the current trend of warming over the past century, and that further warming during the remainder of this century would prove detrimental to human societies.

The consequences associated with rising temperatures and a changing climate can be viewed to have both near and long-term consequences. First, climate change is anticipated to compound current environmental risks in the coming decades—such as ecosystem degradation, habitat loss, food and water shortages, and increase the frequency and intensity of extreme weather. However, climate change is also anticipated to produce even more profound environmental dangers in the longer term—such as rising sea levels, altering long-standing weather patterns, and the elimination of water resources that were previously stored as snow and ice. International thought leaders, like the World Economic Forum (WEF), assessed climate change risks like the failure of climate resiliency action, water crises, extreme weather and natural disasters to be of high likelihood and global risks of great impact (Marsh & McLennan Companies 2016; Marsh & McLennan Companies 2018). For more information on global risk assessments, see Appendix A.

To investigate these risks, several intergovernmental organizations have been established. The Intergovernmental Panel on Climate Change (IPCC) routinely updates the international community on the general consensus of the physical science related to climate change and the associated impacts and implications. In 1990, two years after the creation of the IPCC, the panel concluded that “human activities were substantially increasing the concentrations of [certain gases, which] will enhance the

---

² Refers to human activity. This includes not only greenhouse gas emissions but also land surface changes such as agriculture and deforestation.
greenhouse effect, resulting on average in an additional warming of the Earth's surface.” (IPCC Working Group I 1990) In a synthesis report released in 2014, the IPCC remarked that human activity was “extremely likely to have been the dominant cause of the observed warming since the mid-20th century.” (IPCC 2014) There is support for the IPCC conclusions from the scientific community as well that global warming is largely attributed to humans—the majority (over 97 percent) of scientific publications taking a position on anthropogenic global warming agree with the IPCC’s conclusion that human influence is a major contributing cause (Cook et al. 2013, 2016). While perceptions about the veracity of these scientific claims in some countries are often closely associated with political and social group representation (Kempton 1997; Dunlap and McCright 2008), traditional scientific evaluation would place trust in the consensus of the scientific community (Oreskes 2018). The conclusions of these reports, combined with the educational efforts of scientific experts, have helped to shape the understanding of national governments when they opt to develop climate policies.

Nonetheless, government actors are beholden to their own wants, needs, political preferences, suspicions, ambitions, and intentions, which to date has made it exceptionally difficult to address the issue of climate change on the international stage at the rate and scale most scientists understand to be adequate for the problem. Briefly, the problem beseeches society to incorporate an economic externality—greenhouse gas emissions—into the global marketplace across disparate parties that view the costs of climate change and benefits differently. Many of the human activities which contribute most to climate change—particularly hydrocarbon combustion—have historically proven to be highly correlated with rising economic productivity and standards of living (Stern and Cleveland 2004; Cleveland et al. 1984). This is of particular concern for developing countries, who aspire to expand their own energy generation capacities in the decades ahead but are fiscally incapable of pursuing a buildout of renewable energy technologies and infrastructure as such systems are less proven and likely to cost more than traditional coal or gas power, even if they understand the climate vulnerabilities their countries face. This challenge is not limited to climate change; many environmental problems in the same countries are known to cause losses to health and productivity, yet the same citizens are often not willing to pay more for environmental protections (Magnani 2000; Franzen 2003).

In addition to international coordination, climate change poses other unique challenges as well. First, there is not an agreed-upon suite of policies or technologies that would direct an economy to a low-to-zero-carbon pathway without stark short-term economic tradeoffs. In other words, there is no example of a developing country raising gross domestic product (GDP) and standards of living to developed levels without also increasing consumption of carbon-intensive energy measured across decades. Next, it is a challenge of collective action, in which all parties share in the losses of climate change impacts, but there is scarce incentive for a party to unilaterally reduce their emissions if other parties do not sacrifice similarly. Lastly, it is a “slow burn” problem whose impacts are not immediately evident or obvious. As a result, a country's vulnerability to climate change often garners less attention than more imminent crises like conflict or geopolitical interests—ones whose goals are frequently at odds with climate change mitigation or adaptation efforts.
1.3 Two Facets Of Climate Change Action

Efforts to address climate change can be broadly categorized as either mitigation or adaptation, though such a list is neither mutually exclusive nor collectively exhaustive. Mitigation is defined as efforts to limit the magnitude of long-term temperature rise, either by reducing present and potential sources of greenhouse gases or amplifying and preserving sinks of greenhouse gases. Investing in climate change mitigation by transitioning developing countries to development pathways that are less carbon-intensive is likely to reduce the severity of future climate change impacts. Examples of mitigation include-switching to low-carbon energy sources, such as renewable and nuclear energy, improving the efficiency of energy usage such as modernizing electrical grids, and expanding forests or other carbon-sinks to remove greater amounts of carbon dioxide from the atmosphere.

By contrast, adaptation to climate change are actions taken to manage the presumptive (or unavoidable) impacts of global warming. Adaptation actions are taken to help communities and ecosystems effectively cope with the changing climate and associated impacts. Flooding from more severe weather events, amplified impacts from drought on account of warmer temperatures and changing weather patterns, and coastal erosion from sea level rise are just several of the many climate change impacts that require adaptation responses. Adaptation projects are often infrastructure projects such as flood protection improvements or water conservation. Certain efforts, such as reforestation that both stores carbon dioxide and reduces storm runoff, can be considered an effort with both mitigation and adaptation value, and inherently has co-benefits.

Mitigating or adapting to the effects of emissions and a changing climate often requires capital-intensive efforts are disproportionately felt by the poor who usually have fewer resources to withstand environmental shocks, and inevitably are concentrated in developing countries (United Nations Development Program 2004; Schipper and Pelling 2006; Independent Evaluation Group, World Bank 2004). Further, recent studies of climate vulnerability have identified that areas which have experienced and are likely to experience the most significant changes to regional climates are those regions with high concentrations of developing countries (Althor, Watson, and Fuller 2016; Jorgenson 2014).

While industrialized nations have historically contributed far more to total carbon emissions, recent trends show developing countries with the highest growth rate in carbon dioxide emissions (Marland et al. 2000; Raupach et al. 2007). Developing countries have routinely argued that wealthier industrialized countries should bear more of the responsibility in helping developing countries adapt to and mitigate the effects of climate change, citing the principle of “common but differentiated responsibilities” outlined in the United Nations Framework Convention on Climate Change (UNFCCC) charter. Appendix B provides more details on emissions by country, region, and industry. Policymakers have concluded that funding for both mitigation and adaptation is necessary, as some effects of climate change will be unavoidable.
1.4 Major International Negotiations And The Birth Of Climate Finance

International climate negotiations, which eventually instantiated climate finance, began in 1992. During the following three decades, the focus shifted from solely establishing GHG reduction targets to a suite of related issues, which include emissions trading and financing. Each set of negotiations and agreements aims to address emergent issues and resolve shortcomings of previous agreements. Progress on providing adequate funding for human societies to be more resilient to climate change has often been stymied by political, scientific, and financing disagreements. Table 1.1 below highlights key climate change negotiations and financing milestones.

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>United Nations Framework Convention on Climate Change drafted. First international agreement to address the problem.</td>
</tr>
<tr>
<td>1997</td>
<td>Kyoto Protocol drafted, which established targets for industrialized nations to reduce emissions</td>
</tr>
<tr>
<td>1998</td>
<td>First international market-based implementation mechanisms agreed upon. Provides funds for developing countries and allows for a trade in emissions credits.</td>
</tr>
<tr>
<td>2007</td>
<td>First international effort to assess climate finance needs established, which leads to the Bali Action Plan emphasizing the matter’s importance. Separately, European Investment Bank issues its first climate awareness bond, a predecessor for other vehicles.</td>
</tr>
<tr>
<td>2008</td>
<td>World Bank issues first green bond in response to interest from Scandinavian investment firm</td>
</tr>
<tr>
<td>2009</td>
<td>Copenhagen summit largely seen as a failure in that Kyoto Protocol replacement was not agreed upon. Financing identified as a major sticking point. Developed countries agree to finance $100 billion per year by 2020 to support low carbon development and adaptation.</td>
</tr>
<tr>
<td>2010</td>
<td>Green Climate Fund (GCF) established to help facilitate the achievement of the $100 billion per year target.</td>
</tr>
<tr>
<td>2015</td>
<td>Paris Agreement drafted, which serves as a bottom-up approach for all countries to set their own emissions reductions targets with the implicit idea that targets will become more stringent over time. In the agreement, parties also agree to establish a higher floor for climate finance by 2025.</td>
</tr>
<tr>
<td>2017</td>
<td>US announces it will withdraw from the Paris Agreement and not provide additional funding for the GCF.</td>
</tr>
</tbody>
</table>

Table 1.1: Key climate negotiation and climate finance milestones. See Appendix C for a list of the annual Conference of Parties and their locations.

In 1992, the first negotiations in Rio de Janeiro, Brazil concluded with the adoption of the UNFCCC charter with 154 countries as signatories, agreeing to stabilize GHG concentrations to prevent adverse interactions with the climate system. However, the charter contained no mandatory action or legally binding commitments, something that still evades international discussions (Oppenheimer and Petsonk 2005; United Nations 1992a). The UNFCCC charter also agreed to aid countries particularly vulnerable to the adverse impacts of climate change (United Nations 1992b). By 1994, 50 countries ratified the UNFCCC Treaty with the U.S. being the first industrialized nation and fourth overall. In 1995 at the Conference of Parties (COP), only three years after their adoption, the
UNFCCC mechanisms were deemed inadequate and Annex I countries\(^3\) (i.e. industrialized economies or economies in transition) agreed to make specific commitments (known as the Berlin Mandate), though not legally binding.

The UNFCCC in December 1997 adopted the Kyoto Protocol, the first attempt to control greenhouse gas emissions by making emissions targets legally binding and assisting countries in adapting to climate change. However, the Kyoto Protocol was not brought into force until 2005 (UNFCCC 2014) and even though the United States (along with 149 other countries) signed the agreement, it was never ratified by the United States Congress.

The Kyoto Protocol was generally regarded as ineffective. Industrialized countries did not reach their goal of reducing emissions by five percent from 1990 levels. Furthermore, non-industrialized countries, which were unregulated by the agreement, increased carbon emissions threefold. The International Energy Agency in 2015 concluded that Kyoto was “inadequate to deliver the global goal of limiting global temperature increase to less than 2 degrees Celsius above pre-industrial levels” (International Energy Agency 2015).

One year after the Kyoto Protocol, parties agreed to a two-year grace period to develop implementation mechanisms to facilitate countries meeting their emissions reductions targets. This process saw the implementation of the first market based mechanisms to address climate change mitigation: (1) the Clean Development Mechanism (CDM), (2) Joint Implementation (JI), and (3) Emissions Trading (UNFCCC 2009c). The CDM provides funding for emission-reducing sustainable development projects in non-Annex I countries\(^4\) (mostly low-income, developing countries) (UNFCCC 2009d), the JI funds sustainable projects in economies in transition (EITs) (UNFCCC 2009b), and emissions trading allows credits and allowance trading between Annex I countries (UNFCCC 2009a).

By 2000, all parties, except the U.S. agreed to the Kyoto Protocol mechanisms. In 2001, implementation of the Kyoto Protocol was agreed upon and adopted. With this first commitment of the Kyoto Protocol, adaptation and capacity-building efforts in exceptionally vulnerable developing countries began receiving financing through the Adaptation Fund (UNFCCC 2001a). The Adaptation Fund is funded through a combination of private donors, governments, and through a two percent levy on certified emissions reductions from CDM projects. Future agreements extend the funding to levies on international transfers from emissions trading and JI projects (UNFCCC 2001a). In this same year (2001), the first climate funds were established through the Global Environmental Facility (GEF)\(^5\), the Least Developed Country Fund (LDCF) and the Special Climate Change Fund (SCCF). The LDCF aims "to support a work program to assist LDCs carry out, inter alia, the preparation and implementation of national adaptation programs of action” (UNFCCC 2001b). The LDCF is intended

---

\(^3\) See Appendix D for Annex I countries or the UNFCCC for groups of countries by type and commitment at https://unfccc.int/parties-observers

\(^4\) See Appendix D for non-Annex I countries and definitions

\(^5\) The GEF was established in 1992 and is a partnership of 18 agencies, including but not limited to the United Nations agencies and multi-lateral development banks. The GEF is a financial mechanism for international environmental conventions including the UNFCCC.
for countries most vulnerable to the effects of climate change with “target sectors including water, agriculture and food security, health, disaster risk management and prevention, infrastructure, and fragile ecosystems” (GEF 2016). The SCCF complements the LDCF and aims to “finance projects relating to adaptation; technology transfer and capacity building; energy, transport, industry, agriculture, forestry and waste management; and economic diversification” (UNFCCC 2001c). Unlike the LDCF, the SCCF is available to all vulnerable nations for efforts across sectors. In addition to these two funding mechanisms, the Delhi Ministerial Declaration in 2001 established that developed countries would aid developed countries via technology transfer. Emissions reporting guidelines were formalized in 2003 and financing mechanisms became a central focus of international discussions within three years.

While it was generally understood that at least developing countries would be in need of financial assistance to prepare for climate change, little effort had still gone into quantifying what parties would need how much. The quantification of global cost assessments are discussed more in section 1.5.1; this section focuses on how initial efforts at estimating the requirements influenced the evolution of climate finance. In 2007, the UNFCCC made the first attempt to estimate the costs associated with climate mitigation and adaptation in 2030 to be 200-210 billion and tens of billions of US dollars (USD), respectively (UNFCCC 2007). Following these estimates, the international community recognized that initiatives to address financing would be necessary. At the Bali conference in December 2007, the international community created and adopted new strategies that heavily emphasized financial resource provision (UNFCCC 2008). During this period, the general awareness of climate finance increased and supranational banks started issuing debt products with an environmental focus; the European Investment Bank issued the world’s first Climate Awareness Bond in 2007 and in response to Scandinavian pension funds the World Bank issued its first Green Bond in 2008 (World Bank 2018). In 2008, negotiations on adaptation financing mechanisms began in earnest in Poland. The issue of climate financing became particularly prominent following the 2009 United Nations Climate Change Conference in Copenhagen.

The Copenhagen summit concluded with a pledge from developed countries to support low-carbon development and adaptation in developing countries with an annual USD 100 billion of new and additional finance by 2020 each year shared across the public and private sectors. The commitments included USD 30 billion of “fast-start” finance that was mobilized for more immediate needs (2010 to 2012). With 130 nations on board with the plan, this agreement is known as the Copenhagen Accord.

---

6 By 2017, the SCCF supported projects across 79 countries with a USD 350 million in funding (Global Environment Facility 2016)
7 While adaptation costs are estimated to be lower in the near term, they are projected to rise if mitigation is delayed because the magnitude of climate change and its impacts will increase over time.
8 The market for Green Bonds has grown substantially since this time. The market now includes municipal and corporate Green Bonds. See Appendix E. Green Bonds will be addressed more thoroughly in later Chapters.
In 2010, one year after the Copenhagen Accord, the framework for the Green Climate Fund (GCF) was established to help achieve the USD 100 billion commitment. The GCF was intended to be the centerpiece for expediting financing to less-developed countries for low-emission and climate-resilient projects. The GCF provides the funding to partner organizations (i.e., Accredited Entities) to implement projects that constitute the GCF’s project portfolio. The funding mechanisms in the GCF are intended to be flexible so it can be responsive to various environments and market conditions; the financial products offered through the GCF include grants, concessional loans, subordinated debt, equity, and guarantees. The establishment of the GCF served in part to ameliorate many developing country concerns as negotiations proceeded (Green Climate Fund 2018). Evidenced in 2011, by only the European Union (EU) agreeing to extend their GHG targets past 2012 while Russia, Japan, and Canada let theirs expire. In the years prior to the Paris Agreement (i.e. by the 17th and 18th CoP meetings in Cancun and Doha in 2010 and 2011 respectively), nations were coalescing towards meeting a global climate agreement fueled, in general, by a shared understanding that “the international community [faced] a major challenge in meeting developing country costs for climate change adaptation and mitigation” (Montes 2012). Climate finance is an important part of these negotiations as it is often the glue holding contentious agreements together and is a vital need for a number of communities.

<table>
<thead>
<tr>
<th>Year</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td><strong>United Nations Environment Program Finance Initiative</strong> launched along with Earth Week. It sought to promote sustainable financing collectively with banks, insurance agencies, and institutional investors. It is joined by about 230 financial institutions and investors.</td>
</tr>
<tr>
<td>2001</td>
<td><strong>Institutional investors Group on Climate Change</strong> founded to further low carbon investment by amplifying investor voice and collaborating with business, policymakers, and investors in Europe. It attracts over 160 investors in Europe.</td>
</tr>
<tr>
<td>2003</td>
<td>Non-profit <strong>Ceres</strong> hosts its first annual conference of investors, scientists, and others to launch responses to climate change, which attracts over 160 investors in North America. These conferences advance leading investment practices, corporate engagement strategies, and policy solutions to build an equitable, sustainable, global economy and planet.</td>
</tr>
<tr>
<td>2003</td>
<td>The <strong>Equator Principles</strong> are established and adopted by over 90 financial institutions and helps spur the development of other sustainable environmental and social management practices in the financial sector and banking industry.</td>
</tr>
<tr>
<td>2005</td>
<td>The <strong>Investor Group on Climate Change</strong>, comprising over 60 investors, is founded. It aims to encourage government policies and investment practices that address the risks and opportunities of climate change.</td>
</tr>
<tr>
<td>2006</td>
<td>About 2000 signatories join the UN-backed <strong>Principles of Responsible Investment</strong>, which encourage investors to use responsible and sustainable investment to enhance returns and better manage risks.</td>
</tr>
<tr>
<td>2011</td>
<td>About 30 organizations found the <strong>Asia Investor Group on Climate Change</strong>, a private forum of regional investors with the aim of peer to peer collaboration and learning about the impacts of the risks and opportunities climate change presents to their portfolios.</td>
</tr>
</tbody>
</table>

---

9 For contributions by country and current signed levels, see the Green Climate Fund https://www.greenclimate.fund/how-we-work/resource-mobilization
In December 2015, the UNFCCC agreed to a new framework agreement in Paris—the Paris Climate Accords. One of the most prominent policy decisions included in the Paris Accords, stated that governments agreed to hold global average surface temperature rise to “well below two degrees Celsius above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5 degrees Celsius” (UNFCCC 2015b). The time frame considered for this temperature limit is either 2100 or in perpetuity. For reference, the global average temperature as of 2018 is generally considered to have increased by about 1.0 degrees Celsius since the pre-industrial era. The GCF, established in 2010, was identified to serve as a primary financing mechanism to achieve this goal. Not only does the GCF enable various financing mechanisms but it is also able to engage and leverage financing from both public and private entities.

The Paris Accord also states that “Parties should continue to take the lead in mobilizing climate finance from a wide variety of sources, instruments, and channels, noting the significance of public funds, through a variety of actions, including supporting country-driven strategies, and taking into account the needs and priorities of developing country Parties.” (United Nations 2015)

The Paris Accords are notably different from Kyoto on several fronts, but most prominent of which is probably the “pledge and review” process, where each party puts forward an internally-determined five-year plan for mitigation and adaptation starting in 2020, with the intention for plan ambition to ramp up after each iteration. In absence of a top-down goal, Paris seeks for progress to be realized through bottom-up action (UNFCCC 2015c). Though not requiring particular emissions cuts or limits, the inclusion of developing countries in the Paris agreement is a substantial departure from the Kyoto Protocol. Further, the Parties to the Paris Accord agreed to set a new collective quantified goal higher than the previous floor of 100 billion USD per year prior to 2025 (UNFCCC 2019). As of publication of this research and analysis, the new goal still appears to be under negotiation.

1.5 Capital Cost And Flow Requirements For Adaptation And Mitigation

Due to differing assumptions, definitions, and methodologies, the cost estimates to adequately address climate change adaptation and mitigation vary significantly; this is the same for the estimates
of capital flows between countries. Given the complexity of these issues, such variance is not unexpected and several of the main reasons are discussed in greater detail in the following paragraphs.

Uncertainty is the predominant driver in the variations of mitigation and adaptation cost estimates (Narain, Margulis, and Essam 2011; Barbara Buchner, Angela Falconer, Morgan Hervé-Mignucci, Chiara Trabacchi, and Marcel Brinkman 2011). Because there is no clear blueprint to transition multiple, critical, and intertwined economic networks to carbon neutrality, any estimates are speculative in nature. Similarly, the evolution of climate-relevant technology over the next several decades is even more speculative; for example, the advances of lithium ion batteries and small modular nuclear reactors could drastically alter energy production and delivery, while advances in tangential areas like communications and virtual reality could disrupt energy demand. Further, projections about extreme weather events, sea level rise, and third order effects on human security also exhibit variation, which complicate estimates for what adaptations will be adequate. While climate finance cost estimates are an important exercise, all come with caveats about uncertainties that are necessary to understand.

While the definitions of mitigation and adaptation are easy to separate, the actions associated with each are not. (Huq and Grubb 2003) Consider reforestation, which is often associated with mitigation activities because forests act to store carbon dioxide. However, reforestation can also work to reduce erosion, slow and retain stormwater flows, and enhance surrounding water quality, also making it an effective adaptation to climate change as well. Replacing coal-fired power plants with multiple solar arrays may both help and harm adaptation efforts; society would be more resilient to the risk of associated water bodies being too hot to run thermal power plants like coal, but probably less resilient to responding to high energy demand scenarios like heat waves in a timely manner because the output of solar arrays cannot be adjusted up or down with demand. Effectively accounting for these activities is complicated, and no formalized internationally agreed-upon mechanisms exist.

Sector-specific estimates, and their interconnections with other sectors, further complicate estimate efforts. Some sector specific estimates, such as water, are sparser than others. This is because data and expertise at local levels is often less available. Climate analysis of the energy sector tends to be more frequent and robust on account of the strong ties between carbon emissions and the energy sector and the depth of industry expertise (IPCC Working Group 3 2014). The connectivity of sectors and their uncertain developments under climate change make mitigation and adaptation estimates more complicated. For example, the cost estimates for the water sector will vary significantly if the energy sector addresses climate mitigation through distributed solar energy or water-intensive carbon capture and storage systems. These linkages are commonly discussed in literature addressing the food-water-energy-nexus (World Economic Forum Water Initiative 2011). These gaps and uncertainties continue to challenge analysts in the climate finance space.

Lastly, in the matter of estimating scale and flows, a concise definition of what constitutes climate finance has not been wholly agreed upon by the international community. General definitions encompass all financial resources used to mitigate greenhouse gas emissions and adapt to climate change impacts. This lack of a formal definition and variation in what constitutes climate finance
results in variance when quantifying climate finance flows (J. T. Roberts and Weikmans 2017); different authorities will arrive at different conclusions on account of differences in how they choose to count.

The UNFCCC recommends “that long-term finance for climate change adaptation and mitigation should be an approximate mix of 15 percent public (bilateral and multi-lateral) and 85 percent private resources” (UNFCCC 2015b). Therefore, estimates of costs, required funding, and flows are relevant in describing the potential climate financing gap if the available capital to respond to climate change is unable to match those sums.

1.5.1 Current Estimates Of Adaptation And Mitigation Costs

Despite the challenges of quantifying the costs of responding to climate change, efforts to date have led to a few basic conclusions, and more clarity is often gained when cost estimates are considered in the context of specific industries or efforts. Over the course of roughly a decade, experts have been able to produce a general envelope of costs for responding to climate change which has carried the analysis forward. This section will discuss consensus agreements about overall cost and uncertainty, the breakdown of mitigation and adaptation, and sectoral breakdowns.

There is high confidence and high agreement that costs will be lower if action is taken sooner than in several decades, as threats will intensify the more greenhouse gases are able to accumulate and more will be lost if fewer adaptations are implemented (Council of Economic Advisors 2014). Over time, expert estimates of costs have generally trended upwards as more detailed information has become available. Early estimates in the 1990s broadly judged that the impacts to the United States of a 2.5 to 3 degrees Celsius rise would balance between positive and negative effects. As understanding has improved, assessments broadly concluded that impacts would be more negative than positive (Carey 2011). Costs also tend to increase over time, as more cost-effective high-return projects are addressed early and tougher projects are addressed in later decades. In the World Bank's World Development Report compiled a list of estimated costs needed per decade from 2010 to 2030, average costs tend to roughly double from 2010 to 2030. (World Bank 2010) These estimates are presented in Tables 1.3 and 1.4; it should be noted that these are a sample of several prominent estimates as of 2010 and not representative of all estimates within the literature nor of ones which may have been referenced in recent years. As a result, most experts now agree and recommend that more climate finance action now is a net positive and a check against potentially more negative outcomes.
Table 1.3: Estimated annual climate finance needs for mitigation. Estimates are in 2005 USD Billions from 2010 until the given year. A USD to Euro exchange rate of 5:4 is used.

<table>
<thead>
<tr>
<th>Author</th>
<th>Amount</th>
<th>Year</th>
<th>Measures Considered</th>
</tr>
</thead>
<tbody>
<tr>
<td>McKinsey &amp; Company</td>
<td>175</td>
<td>2030</td>
<td></td>
</tr>
<tr>
<td>Pacific Northwest National Laboratory</td>
<td>139</td>
<td>2030</td>
<td></td>
</tr>
<tr>
<td>International Institute for Applied Systems Analysis</td>
<td>63-165</td>
<td>2020</td>
<td></td>
</tr>
<tr>
<td>International Institute for Applied Systems Analysis</td>
<td>264</td>
<td>2030</td>
<td></td>
</tr>
<tr>
<td>International Energy Agency Energy Technology Perspectives</td>
<td>565</td>
<td>2050</td>
<td></td>
</tr>
<tr>
<td>McKinsey &amp; Company</td>
<td>300</td>
<td>2020</td>
<td></td>
</tr>
<tr>
<td>McKinsey &amp; Company</td>
<td>563</td>
<td>2030</td>
<td></td>
</tr>
<tr>
<td>Potsdam Institute for Climate Impact Research</td>
<td>384</td>
<td>2030</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.4: Estimated annual climate finance needs for adaptation. Estimates in 2005 USD Billions from 2010 until the given year. A USD to Euro exchange rate of 5:4 is used.

<table>
<thead>
<tr>
<th>Author / Source</th>
<th>Amount</th>
<th>Year</th>
<th>Measures Considered</th>
</tr>
</thead>
<tbody>
<tr>
<td>World Bank</td>
<td>9-41</td>
<td>2015</td>
<td>Development assistance, foreign and domestic investment</td>
</tr>
<tr>
<td>Stern Review</td>
<td>4-37</td>
<td>2015</td>
<td>Same as World Bank</td>
</tr>
<tr>
<td>United Nations Development Program</td>
<td>83-105</td>
<td>2015</td>
<td>Same as World Bank, plus poverty reduction strategies and strengthening disaster response</td>
</tr>
<tr>
<td>Oxfam</td>
<td>&gt;50</td>
<td>2015</td>
<td>Same as World Bank, plus national adaptation plans and organization’s projects</td>
</tr>
<tr>
<td>United Nations Framework Convention on Climate Change (UNFCCC)</td>
<td>28-67</td>
<td>2030</td>
<td>Agriculture, forestry, water, health, coastal protection, infrastructure</td>
</tr>
<tr>
<td>Project Catalyst</td>
<td>15-37</td>
<td>2030</td>
<td>Same as UNFCCC, plus capacity building, research, and disaster management</td>
</tr>
<tr>
<td>World Bank</td>
<td>75-100</td>
<td>2050</td>
<td>Agriculture, forestry, fisheries, infrastructure, water resource management, coastal zones, health, ecosystem services, extreme-weather events</td>
</tr>
</tbody>
</table>

Mitigation and adaptation costs are typically assessed separately, despite some overlaps. Mitigation costs are often one order of magnitude higher than adaptation costs, broadly because protections against extreme weather are expensive but exist, while zero-emission technologies for some industries have begun to emerge but no broad blueprints for total transformations exist. In the same World Bank report noted in Table 1.3 and Table 1.4, average costs for mitigation are in the hundreds of
billions, while costs for adaptation are in the tens of billions (World Bank 2010). The UNFCCC in 2007 estimated global mitigation to cost 200 to 210 billion by 2030 and the less precise estimated cost for adaptation is around several tens of billions (UNFCCC 2007). Separately, in 2013 the WEF estimated that mitigation costs to stay below a rise of 2 degrees would cost 0.7 trillion annually to 2020, while adaptation for the same goals and time frame would cost 0.1 trillion annually (World Economic Forum 2013).

An overwhelming majority of climate finance tends to be allocated towards mitigation efforts. Of the estimated 214 billion spent on climate finance by public sources in 2016, only 22 billion or just over ten percent went to adaptation projects (Peter Oliver et al 2018). Analysts noted that private sources did not have the same level of detailed accounting as public sources to estimate the overall spending on adaptation, but the large amount of funding allocated towards renewable energy projects suggests adaptation was similarly a small share of the total. This is probably because the economic case for some mitigation projects is stronger in the short-term; the economic value proposition for renewable energy and storage exists in many markets across the world, while adaptation projects such as seawalls pose immediate up-front costs with uncertain future savings.

There are many who argue that climate finance could benefit from significant expansion in the coming years. For example, current estimates for the amount of global climate finance for 2015 and 2016 is 463 Billion per year (Peter Oliver et al 2018), while the International Energy Agency (IEA) estimates investment will need to nearly double to USD 3.5 trillion annually until 2050 to boost low carbon energy and facilitate more efficient use of it (International Energy Agency 2017). Adaptation costs are often estimated by sector and total USD 30 to USD 450 billion annually based upon the sources, however, the costs to deal with residual effects might drastically increase the current estimates (UNFCCC 2007; Metz et al. 2007; Parry 2009; Montes 2012; Schalatek 2012). This is consistent with the most recent IPCC report that concluded mitigating climate change alone in the 21st century will require between a fraction to several percentage points of global GDP each year, depending on the timing and target of mitigation actions (Blanco et al 2014).

The 2007 UNFCCC cost estimates cited in the previous section identified global mitigation efforts at USD 200 to 210 billion by 2030 and a less precise estimated cost for adaptation is around several tens of billions (UNFCCC 2007) by the same year. Furthermore, the WEF estimates USD 5 trillion for business as usual if a two degree Celsius rise occurs and for growth (beyond business-as-usual) an additional USD 0.7 trillion (World Economic Forum 2013). These mitigation efforts often encompass capital and operating costs, with estimates reaching more than double the UNFCCC estimates. Regardless of the precise number, the quantities of capital described, especially in the recent era of slower economic growth and more austere budgets in much of the developed world, underscores the political challenge in facilitating the buildout experts judge is necessary to manage climate change.

---

10 All estimates are in USD.
The scale of certain markets suggest that climate finance could benefit from significant expansion in the coming years. For example, current estimates for the amount of global climate finance for 2015 and 2016 is 463 Billion per year (Peter Oliver et al 2018), while the International Energy Agency (IEA) estimates investment will need to nearly double to USD 3.5 trillion annually until 2050 to boost low carbon energy and facilitate more efficient use of it (International Energy Agency 2017). Adaptation costs are often estimated by sector and total USD 30 to USD 450 billion annually based upon the sources, however, the costs to deal with residual effects might drastically increase the current estimates (UNFCCC 2007; Metz et al. 2007; Parry 2009; Montes 2012; Schalatek 2012). This is consistent with the most recent IPCC report that concluded mitigating climate change alone in the 21st century will require between a fraction to several percentage points of global GDP each year, depending on the timing and target of mitigation actions (Blanco et al 2014).

1.5.2 Estimating Climate Finance Flows

Like estimating the cost of financing climate change, there is uncertainty surrounding how much financing certain bodies are investing or transferring to others (i.e., flows). As mentioned above, this variability is in part due to the myriad of accounting estimates associated with climate change flows for mitigation and adaptation and because no agreed upon definition among the community as to what constitutes climate finance (von Stechow T. Zwickel and J.C. Minx 2014). The independent think tank Climate Policy Initiative (CPI) identified climate finance to be financing that supports either mitigation or adaptation activity, and would include capacity building, research and development, low-carbon development and climate-resilient development (Barbara Buchner, Angela Falconer, Morgan Hervé-Mignucci, Chiara Trabacchi, and Marcel Brinkman 2011). This definition can be considered to include efforts that directly support mitigation and adaptation, and not just the projects themselves.

Private sources have been observed to comprise the majority of climate financing. For 2015 and 2016, CPI found that private financing accounted for roughly 54 percent of the total of climate financing. Project developers, such as those looking to fund and build renewable energy projects, tend to comprise the majority of private finance. Corporate, household, and institutional investors also comprise substantial portions of the overall private investment as well. In recent years, private equity, venture capital, and infrastructure funds have also made nominal investments in climate finance as well (Padraig, Clark, and Meattle 2018). The international community expects the private sector to play an even larger role in the years ahead. The UNFCCC recommends “that long-term finance for climate change adaptation and mitigation should be an approximate mix of 15 percent public (bilateral and multi-lateral) and 85 percent private resources”. The outsized role of private institutions in supporting a public good presents important policy questions, which are considered further in this research.

While the scale of climate finance has increased during the past several years, a majority of climate finance remains in its country of origin. From 2015 to 2016, more than eighty percent of climate finance was spent in the country in which it was raised. The East Asia and Pacific (predominantly China, Japan, and Korea), Western Europe, and the Americas (North America and
Chile) received almost three-fourths of all funding in those years. (Padraig, Clark, and Meattle 2018) This continues to be a challenge for developing countries in regions such as sub-Saharan Africa and Latin America. Such countries are unlikely to generate large amounts of in-country capital to finance climate projects, despite pressing needs for such, and are more likely to be reliant on public or international institutions for investment.

Sectoral differences across climate finance are also evident upon further scrutiny. For example, renewable energy has traditionally been the largest sector in climate finance. Energy efficiency investments, while harder to track, also tends to be a relatively large component of the overall total. The variance in estimates is expected given the complexity of the problem and the inherent uncertainty and variations in specifications associated with issues such as energy efficiency, capital costs for renewable energy, water supply projections, and sector-specific scenarios. In recent years, investment in sustainable transportation has increased dramatically, thanks to rising interest in and falling costs of battery electric vehicles. However, sectors such as forestry, agriculture, and land use, tend to receive relatively small sums, despite experts remarking that needs in those domains often exist. (Padraig, Clark, and Meattle 2018) The IPCC notes that there are probably a number of unique barriers preventing further expansion of climate finance into the latter sectors, as well as opportunities (Pete Smith 2014). The lack of an economic case as compelling as the cost-competitiveness of renewable energy almost certainly is a key factor. Unlocking additional capital potential in these sectors can provide great benefit to individuals most affected by climate change.

Although estimated costs and flows, vary and are inexact, certain conclusions appear to be consistent and agreed upon. Current climate financing is likely insufficient to meet the estimated funding requirements in the years and decades to come. Most funding tends to be allocated towards low-hanging fruit, where the economic case is strong and investments are more politically secure, as opposed to projects which address significant climate vulnerabilities. To adequately meet projected needs, climate finance must be internationally mobilized and diversified. The sheer volume of the current and future financing needs will require funding streams derived not just from governments or the public domain, but from public, private, and intermediary (e.g., multi-lateral organizations like the World Bank) sources. This engagement between public and private entities implies that neither governments nor the private sector alone have the capital and logistics to raise and allocate the necessary funding to effectively mitigate and adapt to the impacts of a changing climate.

### 1.6 Climate Finance Sources, Intermediaries, And Instruments

As an international initiative that spans public and private spheres, the landscape of climate finance is complex and evolving. New players and instruments appear relatively frequently in this young and burgeoning market. Measuring the sources and flows of climate finance, however, is important to understanding how effectively the global community is responding to the problem of climate change. Nevertheless, data is often difficult to come by, and as discussed in section 1.5, unclear
definitions of climate finance cascade to challenges of measuring its sources, intermediaries, and instruments. This section reviews a commonly accepted effort to quantify these climate finance measures, followed by a subsection that takes a closer look at some alternative sources and the potential for the private sector to play an even larger role in climate finance.

Each year, CPI publishes a report on this landscape utilizing data from various sources. While not authoritative, the think tank's stature is widely recognized in the climate finance community as rigorous and will be considered as a benchmark for the purposes of this research. The sources of climate finance are generally disaggregated into public and private sources, as these classifications are often relevant to international climate negotiations. According to the CPI, public financing includes donor governments (including their agencies), multi-lateral climate funds, and development finance institutions (DFIs)\(^{11}\) (Barbara K. Buchner, Padraig Oliver, Xueying Wang, Cameron Carswell, Chavi Meattle, and Federico Mazza 2017). DFIs are responsible for approximately 90 percent of all public funding and are a critical arm of climate finance. (Padraig, Clark, and Meattle 2018) Figure 1.1 details the contribution of public sources of climate finance during the past several years.

![Figure 1.1: CPI’s 2017 compilation of sources and intermediaries of public climate finance (Barbara K. Buchner, Padraig Oliver, Xueying Wang, Cameron Carswell, Chavi Meattle, and Federico Mazza 2017)\(^{12}\). National DFI refers to national development banks such as the US Export Import Bank. Source: Climate Policy Initiative](image)

In a crucially important development, according to the CPI, private financing has provided the majority of climate finance for at least the past five years. While CPI’s numbers have varied in individual years, private finance has accounted for between one-half to two-thirds of overall financing since 2012. (Barbara K. Buchner et al 2017; Peter Oliver et al 2018) Private financing includes financial commitments by corporations and project developers implementing new renewable energy projects, commercial bank project lending, institutional investors’ direct infrastructure investment, and households investing savings (Barbara K. Buchner, Padraig Oliver, Xueying Wang, Cameron Carswell,

\(^{11}\) DFIs include national, bilateral, and multi-lateral institutions. National DFIs refer to national development banks that lend overseas like the US Export Import Bank. Multi-lateral DFIs refer to international institutions like the World Bank or African Development Bank. Bilateral DFIs are either independent organizations or part of larger development banks and are generally associated with one particular country. One example is the CDC Group, an investment firm which is associated with the United Kingdom.

\(^{12}\) Updated analysis in 2018 adjusts totals for both 2015 and 2016 showing that national institutions estimates were more than double what was listed and rising. Other key trends unchanged (Padraig, Clark, and Meattle 2018).
Chavi Meattle, and Federico Mazza 2017). Private financing has increased significantly with investments from project developers and commercial financial institutions as interest in solar panels and electric vehicles rises. (Padraig, Clark, and Meattle 2018) Commercial financial institutions, which store deposits and make loans to businesses and individuals, have played a growing role during the past several years, more than doubling their investment rate from 21 billion a year to 42 billion USD a year by 2016. CPI also notes that institutional investors and private equity investors, despite being a relatively small subset of the overall landscape, have nonetheless been persistent in their support over the past several years. This suggests a maturation in the sector given these vehicles’ staying power. Figure 1.2 details the contributions of private sources and intermediaries13 to climate finance during the past several years.

![Figure 1.2: CPI's 2017 compilation of sources and intermediaries of private climate finance. Project developers in China, the U.K., and the United States mostly provided finance for projects in their own countries (Barbara K. Buchner, Padraig Oliver, Xueying Wang, Cameron Carswell, Chavi Meattle, and Federico Mazza 2017)14. Source: Climate Policy Initiative](image)

Climate finance can be implemented through various financial instruments such as donations, loans, grants, or bonds. The majority of the funding is non-concessional, meaning that the financier has a stake in the investment and will be paid back at a later date either through loan interest or project equity, as opposed to a donation or gift one makes expecting nothing in return. An example might be the funding of a renewable energy station and grid improvements in a country such as India, where the local utility receives climate financing, then routinely pays back bondholders and structures a project service charge onto its billing rates to account for the project. India's ability to pay, in addition to a clear and direct pathway for return on investment, make the project a prime candidate for non-concessional structuring. The utility might look for a loan (i.e., project debt) in the financial market, much like a city would raise money through issuing a municipal bond or allow investors a share of the project's potential profits through stocks (i.e., project equity). Alternatively, if the utility was a private

---

13 An intermediary is a third party that facilitates an agreement; in this case, between a source of funding and a climate-related project. For the private sector, commercial financial institutions, venture capital, and institutional investors would be classified as intermediaries. For the public sector, all entities that are not national government treasuries would be considered intermediaries.

14 2018 analysis of the last two years found project developer and commercial financial institution funding had declined slightly from 124 billion and 54 billion in 2015, respectively. No other key trends were observed to change (Peter Oliver et al 2018).
company, it might look to put the project on its balance sheet (i.e., balance sheet finance) and either look to issue new corporate bonds (i.e., debt) or stocks (i.e., equity). Each method has its own pros and cons depending upon the particular situation.

Concessional loans (low-cost project debt) or grants from public sources often fund riskier, yet necessary projects where private investment would be difficult to entertain. An example might be an afforestation project in Haiti, something that would be important for multiple sectors (e.g., health, flood control, job creation, soil preservation) across Haiti given its history of extensive deforestation, but unlikely to achieve substantial rates of capital return because of Haiti's level of impoverishment and the program's intangible benefits.

As of 2015 and 2016 according to CPI, market-rate debt (specifically project finance debt) is the most widely used instrument in climate finance. This is true for both public and private financing sources. As of 2018, CPI readjusted their project debt estimates for 2015 and 2016 to reach 190 and 215 billion USD, respectively. Project debt is a plurality, but not a majority, as balance sheet financing (i.e. equity or debt investments into companies) comprises roughly one third of all climate financing as an instrument (Barbara K. Buchner, Padraig Oliver, Xueying Wang, Cameron Carswell, Chavi Meattle, and Federico Mazza 2017; Padraig, Clark, and Meattle 2018). Figure 1.3 breaks down climate financing by instrument over time. Note that this figure includes 2017 numbers which are not updated with the 2018 project debt estimates but are still useful for comparison to other instruments.

![Figure 1.3: CPI's 2017 breakdown of climate finance flows by instrument in billions of USD. Balance sheet debt is typically raised by corporate actors and project developers (Barbara K. Buchner, Padraig Oliver, Xueying Wang, Cameron Carswell, Chavi Meattle, and Federico Mazza 2017)](image)

Figure 1.3: CPI’s 2017 breakdown of climate finance flows by instrument in billions of USD. Balance sheet debt is typically raised by corporate actors and project developers (Barbara K. Buchner, Padraig Oliver, Xueying Wang, Cameron Carswell, Chavi Meattle, and Federico Mazza 2017). Source: Climate Policy Initiative

The vast majority of climate finance has been and continues to be spent in its country of origin. This is most likely because country risks tend to be best understood and evaluated by developers and investors that are within the country with specific domain knowledge. The CPI notes that “domestic policy targeting investor risks may unlock investment at scale. While global climate finance flows are dominated by domestic investments in high or middle-income countries, developing countries rely relatively more on international capital in sourcing climate finance.” CPI does not have readily available statistics for flows by geography for each year for which it has reviewed climate finance, but

---

15 Updated analysis in 2018 found that project debt financing was significantly higher and rising from 190 to 215 Billion from 2015 to 2016. Other key trends were generally the same (Padraig, Clark, and Meattle 2018).
notes that more than four-fifths of all climate finance was spent domestically for the years of 2015 and 2016 (Barbara K. Buchner, Padraig Oliver, Xueying Wang, Cameron Carswell, Chavi Meattle, and Federico Mazza 2017; Padraig, Clark, and Meattle 2018). Of international flows, most went from Organization for Economic Cooperation and Development (OECD) countries to non-OECD countries—which mostly aligns with Non-Annex I Parties to the UNFCCC. The east Asia and pacific region receives the largest portion of climate finance, largely on account of Chinese domestic investments. The lack of funds to vulnerable regions like the Middle East and North Africa indicate persistent policy challenges as well as potential investment opportunities if climate finance instruments are structured correctly (Padraig, Clark, and Meattle 2018).

1.6.1 Alternative Sources And Private Sector Opportunities For Climate Finance

Despite the growth in climate finance over the past several years, the IPCC noted in its 2018 recent special report on the 1.5 degrees Celsius target that “significant gaps in green investment constrain transitions to a low-carbon economy aligned with development objectives.” (IPCC 2018) As a result of this shortfall, the panel concluded that unlocking new forms of public, private, and public-private financing would be essential not only to achieving the temperature target, but to ensuring “environmental sustainability of the economic system.” Consistent with the lack of spending in impoverished countries noted in section 1.6, the report also recommended improving access to financing for the least developed countries and small island states in order to avoid trade-offs with the sustainable development goals and national budget constraints. In short, the climate science community recommends that finance not only be expanded but take on new forms.

Fortunately, alternative funding sources that supported climate finance efforts existed even before the Green Climate Fund was conceived in 2010 (Green Climate Fund 2018). These investments, from public, private, and intermediary sources such as multi-lateral development agencies, are creating paths of sustainable growth and aiding the transition of many developed countries to low carbon economies. For example, the Climate Investment Funds started in 2008 and represents a notable effort in climate financing, 48 pilot countries, 14 contributing developed countries, five multi-lateral development banks, and stakeholders representing the public and private sectors as well as indigenous groups formally committed via signed agreement to a dedicated climate finance vehicle (Climate Investment Funds 2018). These blended approaches are important because they represent potential instruments that can bring the capability of many sectors to bear on some of the more critical problems in climate finance, that of scale and direction towards those most in need.

Furthermore, the insurance and reinsurance industries are trying to incorporate climate related risk into insurance pricing (e.g., company Munich Re) (Johnson 2010). The growth in climate change investment flows is no different in the financial management community, where incorporating innovative climate related products into portfolios is also on the rise. Even after the 2008 financial crisis, institutional investor support for climate change continued to grow. In 2018, the WEF’s Global Risk Landscape report concluded climate and cybersecurity predominate the risk landscape across a
ten-year time horizon (Marsh & McLennan Companies 2018), cap-stoning how climate and environmental risks have become more prominent in economic analyses in recent years. (Bloomberg et al 2014; Marsh & McLennan Companies 2016). The trend is likely to continue as socially responsible investing expands and more portfolios aim to incorporate non-financial environmental social and governance (ESG) metrics (HSBC 2019).

As noted in section 1.6, a diverse set of private sector actors already play a crucial role in supporting climate finance. Historically, some climate advocates were reluctant to engage corporations in financing solutions, concluding their motivations might only be self-serving and the corporate model antithetical to long-term climate change mitigation (Jones and Levy 2007). Whether private firms pursue involvement for the sake of organizational values or a competitive advantage, the private sector appears increasingly concerned about climate change, with many climate change stakeholders welcoming their assistance (Morgan, Levin, and Song 2015; Jones and Levy 2007). The rise in interest for the private sector could result in the private sector becoming starkly more important to climate finance in the years ahead.

Beyond interest, multiple aspects of the private sector present it with unique opportunities to play a leadership role in climate finance in the next several years. With corporations accounting for roughly two-thirds of the 100 largest economic entities in the world by revenue as of 2017 (Global Justice Now 2018), up from about half since the beginning of the century (S. Anderson and Cavanagh 2000), it is clear the private sector will continue to maintain a crucial stream of capital for tackling difficult problems if only by the nature of its size. In addition, private industry is often willing to use their access to capital in combination with technical and managerial expertise, especially when the capital is put towards an investment likely to generate profit. It is also possible, as vehicles yet to play a significant role in climate finance such as pensions and endowments might seek to square their investments with climate-related goals, that new financial instruments become much more relevant to the pool of climate finance and opening additional opportunities for more stakeholders. Combining the private sector’s speed, size, and expertise with public resources could more effectively tackle global development issues, like climate change adaptation and mitigation.

1.7 The Emergence Of The Green Bond Market

Given that green bonds, a particular type of market-rate debt instrument, are the most common of the largest classification of instruments in climate finance by dollar amount, (Padraig, Clark, and Meattle 2018) this paper reviews the history of one of the more well-known and longstanding investment vehicles in the industry. The vehicle’s evolution from institutional experiment to industry cornerstone provides useful insights about the state of climate finance and sets the scene for understanding important policy questions about its future explored in this paper.

Additionally, while this section demonstrates that the growth trajectory of green bonds has been impressive, empirical evidence that rising demand is correlated with their growing interest and positive receptivity has been unproven. It is important for policymakers to understand how much further their
growth potential might be in the hundred trillion-dollar global fixed income\textsuperscript{16} market. Understanding these dynamics will help tailor climate policy and determine whether these initiatives have sustainable momentum and should be encouraged. The analysis in this paper endeavors to provide these analyses, insights, and recommendations.

This section first details the operation of green bonds and the common instrument of World Bank\textsuperscript{17} Green Bonds specifically. Then it reviews the history and evolution of social bonds, and then narrows further into the history of green bonds. Additional observations from the financial and social communities on market trends in these sectors are also included. These observations will be considered further during the analysis in later chapters.

1.7.1 Green Bonds And The World Bank

Green bonds are an innovative financial product that demonstrate how capital markets can be utilized to leverage private capital to facilitate financing for climate resiliency. Green bonds and other climate-themed bonds are fixed income products that are structured like traditional bonds; put simply, an investor lends out a fixed amount of capital to a party, who proceeds to pay the capital back over time with interest. Most governments consider it standard practice to issue municipal bonds to fund infrastructure projects that they cannot finance solely through tax revenue (e.g., road expansion or water treatment plant upgrades). Historically, in addition to infrastructure projects, bonds have also been used to finance social programs like affordable housing and education. (Amadeo 2019)

Modern market practice and the use of multi-lateral intermediaries modify this process slightly. The investor (e.g., an underwriter like JPMorgan Chase) lends capital to the issuer (e.g., World Bank) to issue bonds for the debtor (e.g., Nigerian electric utility) who uses the capital to support projects. The bonds’ contract ensures the investor receives full payment plus interest by a particular date, called the maturity date. The majority of the green bonds issued are green “use of proceeds” or asset-linked bonds. Proceeds from these bonds are earmarked for green projects but are backed by the issuer's entire balance sheet. (Stanton 2000)

When developing countries need to raise bonds, they require an important public-private partnership with entities like the World Bank and its private partners to fund traditionally public projects because of their inability to raise the capital on their own. While any number of bonds can support sustainable development—often through the support of infrastructure-intensive projects in the energy, industrial, transport, and building sectors—green bonds are exclusively for projects with identified climate or environmental benefits. Green bond proceeds are used to primarily fund renewable

\textsuperscript{16} Fixed income refers to any type of investment under which the borrower is obliged to pay back the lender on a fixed schedule. Bonds are the most common fixed income vehicle.

\textsuperscript{17} The World Bank Group is a multi-lateral development agency that consists of five arms and International Bank of Reconstructions and Development (IBRD) is the arm responsible for market-rate loans and the primary issuer of project bonds including World Bank Green Bonds. The poorest countries often lack credit and cannot borrow from IBRD and receive credits from the bank’s concessional financing arm, the International Development Association (IDA). The World Bank tackles climate adaptation and mitigation through various channels, but this research will focus on their innovative Green Bond program.
energy and energy-efficient buildings projects. Over time, renewable energy projects have moved from a plurality to an outright majority of projects (Climate Bonds Initiative 2015, 2018), probably because of the projects’ growing cost-competitiveness. As many sustainable infrastructure projects can be capital-intensive, bonds have often been seen as a natural fit for financing.

The duration of the loan, also known as the bond’s tenor, can be of any length. As the green bond market has matured, the span of tenors has diversified. As of 2018, a majority of green bonds have a tenor of ten years or less (Climate Bonds Initiative 2018c). Green bonds are also more likely to have a shorter tenor than a similar plain vanilla bond¹⁸ without a green rating (CBI, HSBC 2017), as demonstrated in Figure 1.4. This is often a result of the most prominent infrastructure projects associated with green bonds like energy retrofits and electricity generation projects having relatively quick construction spans and returns on investment. However, the existence of green bond projects with tenors of 20 years highlights that the project universe is open to a multitude of proposals regardless of payback time horizon.

![Figure 1.4: Tenor, or duration, of certain bond classifications. Climate-aligned bonds, a separate classification that evolved after green bonds and briefly discussed in Chapter 6, tend to have longer tenors as issuers like rail companies tend to have longer construction and operation periods (CBI, HSBC 2017; Climate Bonds Initiative 2018c). Source: Climate Bonds Initiative](image_url)

World Bank Green Bonds are one of the oldest, largest, and most well-known originators of green bonds in climate finance. As with the overwhelming majority of World Bank projects, World Bank Green Bonds are all high credit quality and AAA-rated¹⁹. Similarly, World Bank Green Bonds typically require substantial project transparency in order to assure investors that the funds are

---

¹⁸ Plain vanilla is an investment term that describes the most basic or standard version of a financial instrument, such as options, bonds, futures, and swaps. Plain vanilla is the opposite of an exotic instrument, which alters the components of a traditional financial instrument, resulting in a more complex security.

¹⁹ AAA is the highest possible quality rating among the three major credit rating agencies.
supporting the environmental goals—a requirement that has generally permeated the entire green bond market. In World Bank’s criteria, World Bank Green Bonds are specifically allocated to projects that meet criteria that their completion would aid in the adaptation or mitigation of climate change. Table 1.5 represents examples of projects that World Bank Green Bonds might support.

<table>
<thead>
<tr>
<th>Mitigation</th>
<th>Adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solar and wind installations</td>
<td>Protection against flooding (including reforestation and watershed management)</td>
</tr>
<tr>
<td>Funding for new technologies that permit significant reductions in greenhouse gas (GHG) emissions</td>
<td>Food security improvement and implementing stress-resilient agricultural systems (which slow down deforestation)</td>
</tr>
<tr>
<td>Greater efficiency in transportation, including fuel switching and mass transport</td>
<td>Sustainable forest management and avoided deforestation</td>
</tr>
<tr>
<td>Waste management (methane emissions) and construction of energy-efficient buildings</td>
<td></td>
</tr>
<tr>
<td>Carbon reduction through reforestation and avoided deforestation</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.5: Examples of World Bank mitigation and adaptation projects (World Bank Treasury 2014).

World Bank Green Bonds have several components that make them more attractive to private investors. First, they are listed in relatively low-risk benchmark currencies. Issuance in US dollars, euros, or yuan where investment demand is high, as opposed to the project’s local currency, means investors do not need to exchange currencies in order to purchase bonds. Second, many of the bonds are AAA-rated, indicating strong credit with less risk. Third, and unlike other instruments, repayment is not tied to the performance of the bond-invested projects. Instead the investors rely upon the World Bank’s process to perform due diligence in identifying, selecting, and monitoring project managers with a strong history of creditworthiness, as well as its industry benchmark transparency and reporting requirements. Investors can instead focus on incorporating the green bonds into their portfolios based upon currency, size, coupon-rate, and maturity. (World Bank Treasury 2010) As a secondary effect, the security of World Bank Green Bonds are likely improving the attractiveness of green bonds within the market at large and drawing in more investor interest.

World Bank Green Bonds support a variety of projects mostly in middle-income countries, consistent with World Bank standards and goals (IBRD 2018). The poorest countries often lack credit and cannot borrow from the International Bank for Reconstruction and Development (IBRD) and were thus not eligible to qualify for early green-bonds-facilitated projects, a recognized project vulnerability since many countries severely impacted by a changing climate might not receive support via World

---

20 The World Bank classifies countries with a per capita income of between $1,026 and $12,475 as middle income, and can receive interest-free loans and grants, also known as non-concessional loans, from the IBRD. Some countries might be credit-worthy but have per capita income less than $1,026, in which case they are considered “blend countries” and can borrow from IBRD but also receive International Development Association (IDA) funding (World Bank 2019).
Bank Green Bonds due to their borrowing eligibility. In 2017, a separate arm of the World Bank worked with a European asset management firm to establish a fund dedicated to supporting the issuance of green bonds in developing country markets and address this gap (McNally 2017). As of the end of fiscal year 2018, the World Bank has funded 147 green bonds in 20 currencies for a total of 11 billion USD in project funding. There were 91 outstanding projects in 28 countries and a total of 15.4 billion USD in outstanding commitments. Thirty projects were completed over the ten-year span (Capital Markets Department 2019). As of 2015, most World Bank Green Bond purchases were made by institutional and private investors mostly from Europe, Japan, and the US (World Bank Treasury 2014; IBRD 2015). As the World Bank does not publish their list of investors every year, it is possible that this pool has diversified along with interest in the last four years.

The World Bank reviews all stages of the project life cycle, including early in the proposal process, to verify and monitor the Green Bond projects to ensure they are environmentally related and delivering desired impact (IBRD 2015; Capital Markets Department 2019). World Bank environmental specialists determine if a project meets the climate mitigation or adaptation goals prior to implementation. If a project meets World Bank Green Bond criteria, the projects are financed via disbursements from green bond capital. Capital disbursements occur over the project life-cycle, often spanning between one to ten years, and are received only when predetermined milestones are achieved (IBRD 2014). Throughout the project implementation and progress, outcomes and impact is monitored jointly by the project's home government and by the World Bank (IBRD 2014; Capital Markets Department 2019). Upon project completion, the World Bank's Independent Evaluation Group performs the project evaluations by comparing project outcomes against original objectives, and assess its sustainability and development impacts (IBRD 2014; Capital Markets Department 2019). Figure 1.5 demonstrates how the verification process typically unfolds, with additional details on the project life cycle available in Appendix F. In this manner, World Bank Green Bonds ensure capital is used to support environmental projects and has served as a benchmark for other green bonds and climate finance in general.
Once the World Bank issues the bonds, the manager (e.g., an investment bank) will offer the bonds to large institutional investors (e.g., pension funds, mutual funds, investment advisors, and large banks). Some institutions, such as Calvert Research and Management, incorporate the green bonds into a fund that offers an investment vehicle for individual investors (Vance Q2 2019). In recent years, several stock exchanges such as the London Stock Exchange have partnered with issuers to offer green bonds for individual purchase and secondary trading—that is, resale after purchase (Exchange 2017). Stock exchanges play a vital role in expanding the marketplace to a wider audience of institutional and individual investors, enhancing the maturity and development of the green bond market. Table 1.6 presents a list of stock exchanges which currently house green bond or climate-themed issuances.

---

Figure 1.5: The World Bank project cycle (World Bank 2012). See Appendix F for details on project life-cycle.
Source: World Bank
Table 1.6: A list of Green Bond segments on stock exchanges as catalogued by Climate Bonds Initiative (Climate Bonds Initiative 2018b).

<table>
<thead>
<tr>
<th>Name of Stock Exchange</th>
<th>Type of Dedicated Section</th>
<th>Launch Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oslo Stock Exchange</td>
<td>Green bonds</td>
<td>January 2015</td>
</tr>
<tr>
<td>Stockholm Stock Exchange</td>
<td>Sustainable bonds</td>
<td>June 2015</td>
</tr>
<tr>
<td>London Stock Exchange</td>
<td>Green bonds</td>
<td>July 2015</td>
</tr>
<tr>
<td>Shanghai Stock Exchange</td>
<td>Green bonds</td>
<td>March 2016</td>
</tr>
<tr>
<td>Mexico Stock Exchange</td>
<td>Green bonds</td>
<td>August 2016</td>
</tr>
<tr>
<td>Luxembourg Stock Exchange</td>
<td>Green, Social, and Sustainable bonds</td>
<td>September 2016</td>
</tr>
<tr>
<td>Borsa Italiana</td>
<td>Green and Social bonds</td>
<td>March 2017</td>
</tr>
<tr>
<td>Taipei Stock Exchange</td>
<td>Green bonds</td>
<td>May 2017</td>
</tr>
<tr>
<td>Johannesburg Stock Exchange</td>
<td>Green bonds</td>
<td>October 2017</td>
</tr>
<tr>
<td>Japan Stock Exchange</td>
<td>Green and Social bonds</td>
<td>January 2018</td>
</tr>
<tr>
<td>Vienna Stock Exchange</td>
<td>Green and Social bonds</td>
<td>March 2018</td>
</tr>
<tr>
<td>Nasdaq Helsinki</td>
<td>Sustainable bonds</td>
<td>May 2018</td>
</tr>
<tr>
<td>The International Stock Exchange</td>
<td>Green bonds</td>
<td>November 2018</td>
</tr>
<tr>
<td>Frankfurt Stock Exchange</td>
<td>Green bonds</td>
<td>November 2018</td>
</tr>
</tbody>
</table>

Green bonds still face several marketplace hurdles to further growth and maturation, particularly for individual investors where preferences for climate action are likely the strongest. The predominant challenge is one of scale. Though multiple funds and stock exchanges exist, the offerings for individual investors still remain somewhat limited. Analysts recognize that despite impressive growth, green bonds can still become a more “mainstream” product (Sean Kidney, Alex Vais 2010). Finally, investors lack benchmark indices through which to track the performance of green bonds. These indices are not only a useful reference for individual and institutional investors, but also a common indicator of maturity. These are indications that from the financial perspective, green bonds still have ways to develop if climate finance seeks to become more commonplace and integral to global finance.

Green bonds have typically produced returns at rates similar to vanilla bonds within the same risk categories, with some anecdotal suggestion that high consumer demand for these bonds has sometimes priced green bonds higher than average. (IBRD 2015; CBI, IFC 2017). This suggests that the premium for green bonds is placed not upon the purchasers, but on the issuers who must take the time and effort to meet transparency and reporting benchmarks but also market and demonstrate climate benefits of the project. This could pose a longer-term risk if issuers eventually judge, despite growing demand for green bonds, that the additional effort is not worth the return and their projects are better off funded with plain vanilla bonds.
1.7.2 The History Of Bonds For Social Impact

Beyond traditional municipal bonds, bonds aimed at supporting social good (referred to as social or cause bonds—henceforth social bonds) have emerged relatively recently in the public health and criminal justice areas to combat disease and recidivism. While the funding structure of green bonds blends ideas from the traditional and social bond markets they are more similar to traditional municipal bonds given their strong link to infrastructure projects. This section highlights several key financial innovations in the social good financing domain and describes how such instruments work.

A noteworthy innovation in bond financing is seen in the social bond market where vaccination programs are funded by leveraging private resources via what are called vaccine bonds. These bonds were first introduced in 2006 by the International Finance Facility for Immunization (IFFIm) and through July 2019 has raised more than 4.5 billion USD from capital markets as diverse as Australia, Europe, Japan, UK, and the US markets, as well as in Islamic finance (International Finance Facility for Immunisation 2019). With the IFFIm vaccine bonds, the World Bank plays the role of manager and advisor.

The vaccine bond is a type of cause bond that uses future commitments to fund bond payments—governments make pledges via the bond and the pledges are immediately converted to cash, available to support immunization programs. Specifically, the vaccine bond allows the Global Alliance for Vaccines and Immunization (GAVI) to borrow against future pledges from donors, such that the donors are guaranteeing vaccine prices once the vaccine is developed (IFFIm 2019a). GAVI represents a unique public-private partnership between immunization, health, and vaccination stakeholders that have facilitated multiple billions of dollars of funding for childhood vaccination, while also improving access to new and underused vaccines (IFFIm 2019b).

Vaccine bonds also represent a milestone for individual investor engagement. The bonds were first issued to Japanese investors in 2008 and claim to be one of the most influential deals in ethically-themed bond investments. The first two Japanese issuances of vaccine bonds were mostly purchased by investors aged 50 and over, equally by male and females, who reported seeing the bonds as tax-efficient with both psychological and financial returns (IFFIm 2008).

The success of vaccine bonds through GAVI prompted support for new and additional financial products in the vaccination domain. An idea to stimulate underfunded vaccinations that was originally grounded in academic circles was put into practice in 2009; this program aimed at combating the problem of undeveloped vaccines due to insufficient market demand and therefore lack of private investment. For the pilot program, many entities (Italy, UK, Canada, Russia, Norway, and the Bill and Melinda Gates Foundation) pledged financing to develop specific vaccines at sustainable, predetermined prices (Clemens et al. 2010). Even though this example is in the health arena, the success of this pilot program represents another frontier in innovative financial structuring that grew out of the success of an innovative fixed-income product.

Social impact bonds (SIBs) were developed several years following vaccine bonds once public policy experts recognized the success of the latter vehicle. Like other bonds, SIBs leverage private
capital to create a funding stream for public services or sectors that lack adequate funding. In the case of SIBs, the capital for social projects comes from private investors willing to assume more risk than the public sector or take on a project larger than the public sector is willing to fund or capable of funding. SIBs distribute risk to the private sector accompanied by the opportunity for additional return based upon performance. SIBs appeal to both the public and private sectors because both sides can benefit; the public sector is able to pay for riskier or resource-constrained services or programs and if predetermined performance targets are attained, private investors make a profit (Hathaway 2019).

SIBs were first used in the UK around 2010 and since in the U.S. at the federal, state, and municipal levels. The UK SIBs were used in the prison system for a program to reduce recidivism. In the US, the Department of Justice and Department of Labor have evaluated SIBs for social service programs. At the state level, New York and Massachusetts, where the financial sectors are strong, have solicited SIBs, and the District of Columbia has also evaluated the feasibility of SIBs. In one example, New York City partnered with investment bank Goldman Sachs and philanthropic organization the Bloomberg Foundation to generate 9.6 million USD in SIBs to attempt to reduce recidivism by funding a jail program that provides skills to inmates (Rudd et al. 2013). The program’s success is defined by a predetermined performance target - inmates must recidivate 10 percent below current levels. If the program is successful, the benefit to the public is obvious and the benefit to the private sector is realized by a return of principal (i.e., 9.6 million USD) and a 2.1 million USD profit (Rudd et al. 2013). If the performance target is not met and the program is deemed unsuccessful, the public still receives the services but does not pay back the principal or profit to the private sector investors. The private-public partnership provided by New York City’s SIBs is an example of the potential mutual benefits they offer.

While warmly received and welcomed as a tool to address social maladies, social bonds have demonstrated a vulnerability to financial disruptions. As the creditworthiness of the institutions guaranteeing the debt product diminishes, so does the credit quality of the actual product. For example, the vaccine bonds offered by IFFIm were originally AAA-rated but downgraded to AA+ in January 2012, accompanied by a warning that additional downgrades might follow (Finnegan 2012). The financing is linked to support from countries who received government debt downgrades. As the vaccine bonds were supported by countries like the UK, France, and Italy, the bonds’ creditworthiness fell when these countries received downgrades or poor economic outlooks. While an AA+ rating still signals a fairly safe investment, it nonetheless may determine whether certain funds must divest of the instrument on account of rules and regulations, and could discourage investors who seek the highest-rated products. This has negative impacts for the financial product, as it means the interest paid to investors must increase to accompany the increased risk and the interest paid to investors means less money is being spent on vaccine programs.
1.7.3 The History Of Green Bonds

Although the first issuance of World Bank Green Bonds took place in 2008 and is a more well-known milestone because the World Bank is now the longest-running and largest source of green bonds, its initiative was sparked by interest in climate financing from Scandinavian pension funds and the Scandinavian bank, Skandinaviska Enskilda Banken (SEB). The development of the Green Bond stemmed from the bank’s institutional investors desiring a liquid and plain vanilla investment product that would explicitly support climate-related products (Skandinaviska Enskilda Banken 2012; World Bank Treasury 2010). SEB worked with the World Bank in 2007 and 2008 to create an investment instrument with the fiduciary element of fixed income products with climate mitigation and adaptation awareness, giving mainstream investors access to climate-related investment opportunities. (Skandinaviska Enskilda Banken 2012; World Bank Group 2019). The success of the collaboration prompted the World Bank to consider the concept further.

The World Bank developed green bonds in 2008 out of a collaboration with a private industry at a time when the World Bank's Strategic Framework on Development and Climate Change was trying to support clients willing to use innovative financial products. (World Bank 2018). Following the success of their collaboration with SEB, the World Bank recognized the amount of capital managed by pension funds alone is in the trillions of USD, and if the World Bank developed a product to engage this capital, other asset managers and investors might follow (World Bank Group 2019; IBRD 2015). Table 1.7 provides more information about the size of certain financial intermediaries for the US market as an example. If successful, private sector collaboration with these much larger markets would diversify the World Bank's investor base while also mobilizing climate finance.

<table>
<thead>
<tr>
<th>Type of Institution</th>
<th>2006 Amount</th>
<th>2013 Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutual Funds</td>
<td>6,472.9</td>
<td>10,814.4</td>
</tr>
<tr>
<td>Private Pension Funds</td>
<td>4,875.7</td>
<td>7,782.1</td>
</tr>
<tr>
<td>Life Insurance Companies</td>
<td>4,478.7</td>
<td>5,855.6</td>
</tr>
<tr>
<td>State and Local Retirement Funds</td>
<td>2,173.5</td>
<td>4,802.2</td>
</tr>
<tr>
<td>Federal Government Retirement Funds</td>
<td>1,068.7</td>
<td>3,474.7</td>
</tr>
<tr>
<td>Money Market Mutual Funds</td>
<td>2,014.1</td>
<td>2,637.4</td>
</tr>
<tr>
<td>Security Brokers and Dealers</td>
<td>2,295.9</td>
<td>2,059.6</td>
</tr>
</tbody>
</table>

*Table 1.7:* Snapshot of total US financial assets held by certain financial intermediaries. Measured in USD Billions (US Federal Reserve 2019).

Following early successful issuances of World Bank Green Bonds, private financial institutions quickly took notice. In 2010, the U.S. banks and the Japanese bank Daiwa Securities began underwriting the World Bank's Green Bonds. In the following years, several intermediaries joined SEB and Daiwa to play pioneering roles, including Bank of America Merrill Lynch, Citibank, Credit Agricole, Daiwa, HSBC, JP Morgan, Morgan Stanley, TD Securities, and Westpac. Within four years,
the World Bank had issued over $4 billion through its Green Bond program (Barbara Buchner, Angela Falconer, Morgan Hervé-Mignucci, Chiara Trabacchi, and Marcel Brinkman 2011; World Bank Group 2019; World Bank Treasury 2014).

Increased investor demand and broader market appeal continued to engage new players in the green bond market in the years shortly thereafter, who proceeded to largely emulate the World Bank’s IBRD program. Two of the largest additions began green bond issuances in early 2013. First, the International Finance Corporation (IFC)—an arm of the World Bank which focuses on engagement with the private sector to achieve the Bank’s goals—issued 1 billion USD in green bonds. Around the same time, the European International Bank 21 issued 650 million Euro worth of green bonds (Reuters 2013a). Later in 2013, the Export-Import Bank of Korea—a government export credit agency—issued $500 million in green bonds (Wee 2013). Further, IFC declared that it planned for 20 percent or roughly $3.2 billion of its future bond issuances to be categorized as either green and SRI bonds by 2015 (Laura MacInnis, Alexandra Klopfer 2013).

Financial reporting suggests that demand for green investment products grew, especially in the U.S. market, following the successful issuances in these early years. Even though U.S. banks had a hand in early issuances, initial demand for green bonds came predominantly from European and Japanese investors. As late as 2017, Euro-denominated bonds still maintained the plurality of issuances (Cowan 2017). However, interest from U.S. investors has climbed in the past five years. As of 2018, the United States is the largest green bond market at $34.2 billion in issuances and has the largest issuer in housing institution Fannie Mae (Climate Bonds Initiative 2018c). California State Treasurer's Office is one of the first and largest U.S. investors and more recent market participants include pension funds (e.g., TIAA-CREF) and large asset managers such as Blackrock and Deutsche Wealth and Asset Management (California State Treasurer’s Office 2019; World Bank Group 2019). The most common investors are participants interested in the long-term financial products that offset shorter term risk and tend to be pension funds, mutual funds, insurance companies and sovereign wealth funds 22, according to financial reporting (Climate Bonds Initiative 2018c).

Municipal bonds and corporate bonds are more prevalent in developed economies yet much of the economic growth and infrastructure development requiring climate and environmental considerations are likely to occur in emerging economies. In July 2104, following the first state issuance in Massachusetts in 2013, Washington DC's municipal water department DC Water issued $350 Million in green bonds to finance projects to improve local water quality and flood mitigation. It was the first issuance of its kind for the US debt market (Water 2019). In the emerging economies where financial markets are less mature development and supranational banks are likely to play a larger role in the early stages of developing the green bond market. In 2017, private banks in Latin America

---

21 The European Union’s long-term non-profit investment bank for the purposes of European integration and social cohesion. (European Investment Bank 2019)

22 Sovereign wealth funds are state-owned investment funds that look to invest national wealth, often derived from proceeds from state-owned enterprises, into other vehicles.
(i.e., Colombian banks Bancolombia and Davivienda) issued green bonds. In the same year, Fiji issued a sovereign green bond - the first in emerging of which a number of others have followed.

Green bonds continued to experience exponential growth and diversity from a relatively low baseline during the last five years, possibly on account of rising popularity. As of 2018, new issuers were responsible for more than 30 percent of new issues by volume, and entities included corporate, government, development banks, sovereign wealth funds, and real estate investment trusts, among others (Climate Bonds Initiative 2018c). Though growth in volume has slowed since a dramatic rise after 2016, issuances still rose from roughly 162 Billion in 2017 to 167 Billion in 2018 (Climate Bonds Initiative 2018c). Further growth anticipated for 2019 as financial institutions and sovereign wealth funds continue their participation. As of June 2019, the IFC had issued 9.2 billion USD in bonds across 18 currencies (IFC 2019). Regardless of the economy (emerging vs developed) attracting private investment in the bond markets will be critical. Green bonds reduce friction for private investors to enter and engage in the climate business.

1.7.4 Green Bond Market Limitations

Despite the growth of interest and issuance of financial products like Green Bonds, it is important to recognize neither any individual issuance nor the product’s long-term success is guaranteed. Investors may not find the coupon rate offered to be a sound investment, and bond issuance failures can occur. Effective research and analysis prior to a bond’s issuance is critical, and failures are relatively common. Genesee, Michigan postponed the sale of $53 million in municipal bonds in August 2013 when not enough investors were found. Analysts concluded that despite an elevated coupon rate, investors still found it too low given the anticipated risk of default (Reuters 2013b). That same year, the Grafton Group withdrew the issuance of 50 million pounds after concluding that recent interest rate volatility and market uncertainty reduced the appetite for its bonds.

Likewise, certain types of products can fall out of favor if seen as unsuitable to the broader market. Traditionally, new financial products and instruments are released only when teams of market veterans identify strong potential demand. In his memoir, Good Derivatives, Richard L. Sandor recalls the process of creating new financial instruments such as interest rate futures and carbon dioxide emissions credits (Sandor 2012). The process often took months to years to complete, and relied heavily upon Sandor’s ability to understand latent market demands and communicate effectively with entities that would wish to issue and buy such products. While most new financial products developed by Sandor were successful, several failed to attract sufficient investor interest. The Chicago Carbon Exchange, for example, lasted from 2003 to 2010 when it was shuttered due to inactivity. Sandor concludes that a majority of products fail as the result of reluctance from potential market participants to change their status-quo behavior.

These examples highlight the need for additional metrics to gauge market and investor sentiment. Financial analysts at both investment banks and debt-seeking entities put significant time, effort, and capital into developing each bond issuance. Similarly, it is likely that policy experts and
advocates will spend similar efforts in developing initiatives to advocate for thematic investments to aid the public good in the years ahead. All parties stand to save substantially with the reduced risk of an issuance failure. Sentiment analysis of financial product demand could prove to be a useful tool for all parties involved, enabling teams to better time the initiation of new products and potentially prevent their failure.

With the acceleration and expansion of the Green Bond market, observers have raised concerns about transparency and standardization matching pace with growth. Transparency will make reporting clear and ensure product standards are met (e.g., reaping environmental benefits). Clear delineation from a central body that determines what “green” or “climate-themed” encompasses and if projects have sufficiently met these definitions would be necessary. An appropriate delineation of terms should also increase the scale of the market, likely leading to increased investor engagement. According to HSBC, both transparency and standardization will help scale the market, and by removing additional barriers to the market through securitization or guarantees private investors should be more willing to engage in the market (CBI, HSBC 2017).

1.8 Statement Of Research Objectives

While various aspects of climate change and environmental discussions can become politicized, the long-term benefits to productivity from sustainable, low-carbon, and climate-resilient development cannot be ignored. An increasing awareness of the implications associated with a changing climate is resulting in financing mechanisms for climate adaptation and mitigation activities. There is increasing funds flowing to address this complex policy problem, and there is a general consensus that additional financing is needed for financing the response to a changing climate. In summation, the global community is observing that a public good (i.e. a stable and habitable climate) is not being achieved, yet there is a growing and powerful set of tools (i.e. climate finance instruments) that are buttressed by the private sector. The global community will need to respond to this challenge both now and in the coming years, and its course of action could have ramifications across transnational policy issues.

By their nature climate change discussions cover a broad array of topics, and this research does not focus on either the earth science or political science discussions covered by many academics. Instead, the research focuses on the economic discussions surrounding climate change, financing the response. This research takes as given that many global actors seek to both mitigate and adapt to climate change and are often resource constrained. This research then recognizes that financing is being mobilized in response to this collective understanding of the problem of climate change, and that this financing consists of public, private, and public-private investor flows. Moreover, private and public-private streams of capital are integral to the climate finance landscape. Thus, there is a global public good problem which is underpinned by the private sector action that remains to be resolved by the policy world. It is this novel and very modern conundrum that this research seeks to explore, with the
This research aims to identify temporal, geospatial, and investment trends related to climate finance and to a specific climate financing mechanism, the green bond. The insights generated can inform financial product structuring, negotiation positioning, investment focus, geopolitical relationships, and policymaking. With this understanding, this information can be used to leverage (or engage) additional private investment or private partnerships for climate change adaptation and mitigation efforts. The underlying hypothesis of this research is that information obtained through the public discourse, which represents statements of both public and private actors, can inform areas of investment interest. Specifically, the research looks to explore whether an investigation of discourse can identify areas of investor interest in climate and environmental finance, or gaps between interest and capital deployment. If these areas or gaps can be identified, then such information can inform the following policy questions:

- How can a financing gap be bridged to tackle environmental (or other social) concerns?
- How should the public and private engagements work together?
- How should financing solutions be positioned?
- Will enough capital arrive at the right place at the right time?

To inform these overarching policy questions, the climate finance landscape and green bond investments are examined from multiple angles with exploratory and descriptive analyses using unstructured data from objective (news) and subjective (blogs) sources. The analysis is segmented into two parts that respond to the following key research questions:

- Part 1 - The Climate Finance Landscape: What can be learned about the evolution of climate finance to better facilitate policy making and private-sector engagement?
- Part 2 - Green Bond Investment Opportunities: Can opportunities be identified for public-private cooperation within green bond project financing?

These two parts represent connected research areas: (1) understanding the current status and landscape of climate finance and (2) an examination of specific actions using financial mechanisms for policy.

The first part of this research focuses on identifying and exploring dominant thematic trends within climate finance by using analytics that automatically summarize text-based data. To accomplish this, topic modeling (an unsupervised machine learning technique) is applied to international news publications, that are selected based upon aggregated, crowd-sourced, search history. These emerging and evolving topic trends represent evolving climate finance concepts from public and private actors in the news discourse. Topics are identified from within the news headlines and are used to describe the landscape of climate finance both statically (independent of time), temporally (measuring topics over
time), and geographically. As an additional lens, sentiment or text polarity is applied to gauge the general attitude or opinion of the topics.

The second part of the research highlights a specific fixed-income investment mechanism within climate finance—the labelled green bond. Unlike with the first part of the research, the desired topics are known and correspond to the green bond investment areas. Within text-based data from news and blogs, a subset of topics are identified that correspond to the investment areas. To do this a semi-supervised machine learning approach is used in conjunction with sentiment analysis. The inclusion of the sentiment metric is intended to evaluate whether positive perceptions of an investment area are informative for the use of proceeds within the financial product. The annual temporal topic and sentiment trends are qualitatively compared to the annual investments. This contemporaneous comparison is qualitative in nature yet highlights investment and institutional opportunities.

This research focuses on trends and opportunities present in the established realms of climate and environmental finance. Environmentally-themed and climate-focused financial products are similar to other new financial products (as are any emerging trend or technology), where a “critical mass” of interest from private investors (both institutional and retail) is often required in order to transition from a niche market to mainstream adoption. It is neither certain nor easy. The successful broad adoption and widespread interest in climate- and environmentally-themed investment vehicles, such as green bonds, is therefore noteworthy. This research seeks to provide insights to climate finance’s successful evolution, and evaluates potential opportunities in the industry. Suggested future work should build upon this analysis in a forward-looking and predictive manner, extend the focus of the corpus to other areas like adaptation, and explore causal mechanisms.

The remaining chapters are intended to inform the policy and research questions and while linked together could be read separately with basic familiarity of the research area. Chapter 2 provides an economic framework; it is an argument for engagement in socially responsible behavior from the private sector, like investing in climate finance or offering environmentally (or socially responsible) alternatives. Following this economic positioning, the data (Chapter 3) and methods (Chapter 4) used in the two research portions are described. This is followed by the analyses, results, and policy recommendations for the evolution of climate finance (Chapters 5) and green bond opportunities (Chapter 6). A discussion of future work and modelling extensions are contained in Chapter 7, and concluding social, behavioral, and ethical policy remarks can be found in Chapter 8.
2. An Economic Framework: Socially Responsible Engagement For Climate Financing

This chapter presents an economic framework to understand the role of private engagement in financing the response to climate change. A stable climate and healthy environment are considered a benefit to the public because they impact virtually all members of society and thus have characteristics indicative of public goods. As is discussed in this chapter, governments have in both theory and practice played a special role in the management and provision of so-called public goods. However, climate finance represents engagement by both private and public actors. When a public good is insufficiently provided by the public sector, private actors can engage in providing or contributing to that public good through practice called corporate social responsibility (CSR). CSR has been studied and debated extensively in public finance academics and popular press (Blomgren 2011; The Economist 2019; Epstein 1989; M. Friedman 1970; Hill et al. 2007). These discussions offer a perspective to understand when private or corporate engagement can at least supplement public institutions’ provision of public good, whether international organizations or local municipalities. With the rise in CSR over the last decade\textsuperscript{23}, understanding the instances where CSR implementation is most effective and most likely to be pursued has become valuable. This becomes even more important for the matters of climate change and climate finance because of the issues’ complexity and the multiple interacting forces at play.

The beginning of this chapter presents fundamental economic concepts and theory meant to be accessible regardless of expertise, whereas the later portion of the chapter assumes some familiarity with economics and public finance concepts. This chapter begins with a basic explanation of the term “public good” along with the general economic theory of the concept. The discussion is grounded in a basic public finance economic framework and is presented for those with little to no background in economics, as it is important in relation to the analysis and discussion of the subsequent chapters. With this foundation, the high-level economics of climate change and CSR are each discussed. Particular attention is paid to the historical academic debate among economists on whether CSR is an economically viable approach for private participation in supplying public goods. The chapter then considers CSR in the context of the response to climate change, and the particular application of private sector engagement in climate finance and green bonds, which highlights the pertinence of CSR to climate policy discussions today. Finally, the policy implications of CSR being a viable approach to effectively providing public goods are discussed, with particular consideration given to climate change policy and financing. These implications are discussed further in Chapter 8.

\textsuperscript{23} This corresponds to increasing interest in environmental, social, and governance (ESG) factors.
2.1 Understanding The Concept Of Public Goods

Economics is the study of the production, transfer, and consumption of goods and services. Goods can be broken down further into categories based on their qualities. A good is said to be rival if its consumption automatically excludes consumption by other consumers. Examples of rival goods include food and fuel. Alternatively, a good is considered non-rivalrous if its consumption does not preclude another party from its consumption. Internet data, for example, is a good that can be used by one firm without precluding use by other firms. A good is said to be excludable if there is an effective way of ensuring that it is only consumed by those who have paid for it, and non-excludable if it is impossible or otherwise infeasible to do so (Mankiw 2007). Figure 2.1 presents these concepts graphically and provides further examples.

![Figure 2.1: Traditional economic delineation of types of goods. Source: Lumen-Boundless (2019)](image)

Public goods are those goods which are both non-rival and non-excludable; goods whose benefits can be enjoyed by all and without exhaustion. Some common examples include national defense, a system of laws, and clean air. For instance, one person’s protection from foreign invasion does not impede another person’s ability to enjoy the same protection, nor can that protection not be provided for one or more individuals. In contrast, items in a store like shoes or food are provided only to paying customers and can only be enjoyed by the individual who purchased them.

Public goods will often be underprovided or underproduced by the free market on account of the challenges of these goods being non-rival and non-excludable. This is commonly referred to in economics as the “free rider problem.” Consider a local public radio station which relies entirely on donations to cover its operating costs. Anyone with a radio may listen, and there is no feasible way to ensure that only those who have sufficiently donated are able to listen, thus making the station a public good. As those who do not donate can enjoy the good as much as those who do, there is limited incentive to donate, and the radio station runs the risk of not receiving adequate funds to sustain

---

24 The examples listed are also known as pure public goods because they perfectly match the definitions of non-rival and non-excludable. Impure public goods, such as transportation networks, disease control, and clean water, fit the definition of non-rival and non-excludable more loosely but are nonetheless most accurately categorized as a public good.
operations. Those audience members who listen to the station routinely, but never make a donation during pledge drives would be considered “free riders” (Kim and Walker 1984).

Traditionally, because of these risks, public goods are delegated to the government to both fund and manage. As the sole authority that can both raise taxes and enact regulations, the government has tools to counter the free-rider problem. In the example above, the government could use tax revenues to ensure that the public radio station receives the funds it needs. In this way, the government can act as an effective steward of public goods. But as further examples will demonstrate, the complexities of some public goods challenge this traditional framing.

2.2 Challenges Of Modern Environmental Public Goods

Environmental public goods pose unique risks for market failure because of the difficulty in effectively measuring them. Citizens broadly benefit from the provision of a healthy environment but assessing the ways in which those benefits can be economically quantified is challenging and requires the use of experts, and possibly highly-valuable and complex tools. Groups of citizens are also likely to value certain environmental benefits more than others, complicating effective adjudication of the resource. For example, clean and safe water reduces health and disease risks for all citizens in a watershed, and provides secondary and tertiary benefits such as increased recreational opportunities and stronger fish and wildlife populations. Some citizens may highly value these secondary benefits, and others might not at all. None of these benefits are easily quantifiable, and the average citizen is unlikely to understand and price these benefits into their behaviors. Contrarily, the additional cost of buying and installing pollution prevention equipment is well understood for a firm, and the increase in prices to recover the associated investment could be easily understood by consumers. Thus, the incentive for the average citizen to be a free rider and develop a consumption pattern with little regard for its impact on water quality is strong. The traditional remedy is for the government to effectively set a price for the public good of water quality through regulations, penalizing activities that pollute, or by raising funds for projects that protect water sources through taxes (Seneca and Taussig 1974).

Climate change is widely regarded as one of the most complex environmental public good challenges in the modern era. While similar in nature to the water pollution example above, it presents several additional challenges. Foremost, a stable climate is a global public good that transcends national boundaries. Not only do free-riding incentives increase when the affected population is larger, effective management requires seamless coordination across sovereign nations, an effort that is challenging even in peaceful and prosperous times (Abbott 2012; Bulkeley et al. 2014).

The economic impacts of climate change also pose unique challenges to the very theory of public goods. Some regions are anticipated to experience economic gain, such as the Arctic which could see more commerce on account of more navigable sea routes from melting ice. Additionally, many northern latitudes are likely to experience climates more favorable to agriculture. However, most areas are likely to experience severe net-negative impacts, such as greater likelihood of floods, droughts, or heatwaves. Other areas face existential risks in the long-term; low-lying islands and coasts
can become permanently inundated and certain regions may become too hot and dry to sustain human livelihoods (M. Burke, Hsiang, and Miguel 2015; Althor, Watson, and Fuller 2016). As so much of commerce is interconnected across the globe, some academics raise the question of whether climate change can effectively be approached with an economic framework at all (Vanderheiden 2008).

Many risks are also decades or centuries into the future, raising ethical questions as to what an appropriate “discount rate” should be. The term discount rate refers to how much costs in the future should be “discounted” in order to derive an equivalent cost today; for example, the cost of a home renovation in five years may matter less to a homeowner today because they will have time to accumulate the capital to pay for it. While assessing current valuation of flood insurance losses in 30 years’ time may be difficult but feasible, assessing the same for complete inundation of countries in two centuries is an unprecedented undertaking. While literature has tried to achieve advances in this space, it remains contentious as arguments often contain an ethical component (Nordhaus 2019; Goulder and Williams 2012; Hepburn 2007).

Finally, while the scale and timing of changing climate risks are being forecast with greater resolution as computing power increases, risk profiles are not the same as impacts. Much as young people may view health insurance or a household might dismiss earthquake or flood insurance as too much cost for an unlikely benefit, many stakeholders who perceive they have not experienced severe climate-related impacts may be less likely to hedge against future damages (Knoblauch, Stauffacher, and Trutnevyte 2018). While there are some observations that this perception is changing over time in conjunction with demographics and a recent ability to identify the influence of climate change in extreme weather events in a timely manner, this has not shifted social perceptions on the whole towards immediate significant action to address climate change.

These reasons contribute to why scientists judge that progress to limit greenhouse gas emissions at an international level has almost certainly been inadequate to avoid dangerous impacts from a changing climate. Recent reporting from the Intergovernmental Panel on Climate Change remarks that emissions are likely to result in a global average temperature increase of 1.5 degrees Celsius or more this century, which will lead to dangerous and even existential losses for many communities (IPCC 2018).

2.3 Understanding Corporate Social Responsibility

In a traditional economics framework, firms or investors act as profit maximizing agents, concerned with complying with regulations and legal standards only insofar as noncompliance would lead to penalties and produce suboptimal profits. In this same framework, the concern for society and provision of public goods rests solely in the hands of the government. Each set of actors act with their own best interests in mind, maximizing their own equities (Seneca and Taussig 1974).

However, behavioral economics informs us that the behavior of economic actors is rarely this simple. Indeed, many actors make economic decisions that would be considered economically sub-
optimal or irrational, and oftentimes these decisions are linked to an intention to improve a social or public good (e.g., a consumer buying local produce to support local farms, and a firm obtaining an independent certification for safety quality or ethical behavior not required by law). The complex nature of human decision-making makes actors—including firms—inclined to make financially sub-optimal if not moral decisions (Shefrin and Statman 2003; Naqvi, Shiv, and Bechara 2006; E. Anderson 2000; Henrich et al. 2001). This opens the potential for actors at the firm-level to practice what is formally known as corporate social responsibility (CSR)—where firms engage in ethical or good will efforts even if such activities are less than profit-maximizing (Crane, Matten, and Spence 2019).

The inseparable dichotomy of private and public entity roles has been called into question since at least the 1970s when economists first explored a formalized concept of a blended stewardship of public goods by both types of actors (Bowman and Haire 1975). Some have argued that the concept dates back even further, when communities might often depend upon a firm to provide not just livelihoods but other means of social welfare improvement (The Economist 2019). Academics have noted that when corruption, incompetence, or wastefulness pervade a government, its ability to adequately and efficiently provide a public good is often hindered. In such instances, they assess that private enterprises could—or even should—step in and consider the impact of their actions on society (i.e., customers, the environment, communities, and employees). Thus, the private sector would bear some provision of public goods through socially responsible practices (Blomgren 2011).

Under this framing, it is theoretically possible that a public good is optimally provided not when firms are purely profit seeking (and the government acts as a policing actor who may detract from that good), but instead when individuals and firms enhance that good and governments incentivize their participation in such activities. Actors with socially responsible interests may be able to augment a government’s actions to deliver a public good; simultaneously, governments may be able to seek to establish policies which encourage greater private participation. In the public radio example above, government funding might be augmented by industry support, recognizing the utility that the radio station provides and the risk of loss. The private sector might conceivably lead, or in extreme circumstances replace, the public sector in the provision of certain public goods in situations where the government fails to provide a satisfactory level of that good. Security, garbage collection, and open source software are all examples where academic literature has observed private industry to demonstrate leadership (L. Burke and Logsdon 1996).

In situations where the private sector has a role to play in the efficient provision of public goods, the economics of CSR demonstrate the utility improvements especially when governments are considered “imperfect”. In public finance economics literature, Belsey and Ghatak (Besley and Ghatak 2007) characterize an “imperfect government” as one unable to either appropriately assess the values society places on certain goods or impartially monitor for cheating. The authors support their assessment with an economic model illustrating when CSR can provide more utility to the hypothetical society (Besley and Ghatak 2007); The model illustrates that in the presence of an “imperfect government”, enhanced utility is achieved through private actors engaging in CSR. Within this work
and academic work that follows, the threshold at which a government might be considered responsive and efficient enough to avoid the label of “imperfect” is left undiscussed. While the economic models represent a simplified version of complex societal systems and common knowledge understands that no government is “perfect”—or immediately responsive to the wants of its citizens and entirely free from waste, fraud, and abuse judging from the authors’ definition of the antonym.

In practice, the term “imperfect governments” can be generalized beyond Besley and Ghatak's (Besley and Ghatak 2007) definition to refer to autocracies, limited democracies, and unresponsive democracies. Economic literature tends to consider modern Western democracies as separate on account of strong functional institutions and presumes CSR would provide little to no improvement to overall utility (Lyon and Maxwell 2008; Florini 2013). However, it may be appropriate to consider a more flexible case-by-case approach to CSR. First, even strong functional democracies can find themselves under strains seen by less-developed counterparts. Corruption, political indecision, machine politics, corporate capture, and scandals of wastefulness are all possible examples of long-standing challenges that flourishing democracies continue to face. The persistence of these challenges suggests that a role for private actors engaged in CSR could be plausible in any government. Second, as noted above, the challenge of modern environmental regulation and financing climate change responses in particular vexes even the most capable governments. Many governments will not have the tools and scale to effectively value and regulate their natural environment, just as governments have not been able to achieve a plan that maintains greenhouse gas concentrations within the scientifically-recommended levels for human civilization. As such, economists should not look at the government first to determine whether or not CSR could be utility-enhancing, but rather the problem CSR is trying to address. Complex problems that interact with multiple government stakeholders might well present opportunities for CSR to be a more optimal solution, regardless of a government’s quality.

2.3.1 Private Voluntary Action And The Triple Bottom Line

While implementations and frameworks for CSR vary, a common denominator is that CSR provides a way for a private organization to adhere to societal obligations regardless of the economic functions. These societal obligations can produce positive externalities, yet the organization should still produce adequate profits in the eyes of its members or shareholders (Epstein 1989). Given this economic perspective, one could state that CSR can be classified as a private enterprise activity that generates public goods and/or curtails public bads in addition to the more traditional activity of producing private goods (Besley and Ghatak 2007).

Understanding why a profit-minded firm would commit to CSR, an act usually with upfront costs and uncertain benefits regarding profitability, is an interesting quandary. It is a subject of research where economic models are used to describe possible scenarios and explore the feasibility of firms engaging in CSR (described in section 2.3.3). Furthermore, when firms do partake in CSR, understanding the extent to which a public good can and will be provided is important. This is especially true since governments are most often the providers of public goods.
There are many reasons why a firm might engage in CSR. Considering the case of a firm that voluntarily reduces its air pollution below legal limits: its reasons may be as diverse as reducing employee health costs, avoiding price shock from anticipated regulation, or seeking to gain market share with a customer base with a higher willingness to pay for environmental protection. Regardless, greater public benefit is experienced by their actions than without. In this way, private enterprise can play a role in providing the public good (Hill et al. 2007; Epstein 1989).

In difficult economic times, the competitiveness of a firm becomes a greater priority for CEOs, executive boards, and investors. CSR could either increase or decrease a firm’s competitiveness (market-share and profits) depending on the time horizon under consideration and consumer preferences. The costs, benefits, and political landscape are all considerations leaders and firms make in the decision to engage in socially responsible practices. If waste and corruption permeate a government, or strong partisan difference can make certain policy preferences irreconcilable, its ability to solve social problems or adequately provide public goods becomes questionable. In situations where the government is perceived to be less willing to provide goods for public benefit, there is theoretical evidence that CSR can play a primary or supporting role in public good provision. Within this context, there is a possibility that private provision of climate finance, including green bonds, could play a key role in supporting the public good of mitigating and improving community resilience to climate change.

2.3.2 The Debate Over CSR’s Private Voluntary Action

Economist Milton Friedman began the debate over the relevance of CSR in 1970 in his New York Times article “The Social Responsibility of Business is to Increase Profits” in response to discussions of social responsibility by corporate leaders of that time (M. Friedman 1970). Today, almost 50 years later, the debate over CSR continues. The body of economic literature discussing CSR in an economic context places most economists as skeptics of CSR like Friedman, claiming that a competitive economy would not allow for CSR to exist for the long term as intense competition will not permit the sacrifice of profits (Baumol and Blackman 1991). Similarly, Friedman argued that companies have no incentive to go beyond minimal compliance with laws and regulations (and perhaps ethical obligations), and would instead be incentivized to maximize profits (Hill et al. 2007; M. Friedman 1970).

Other skeptics have claimed CSR is merely a marketing ploy for corporations to help clean up a flawed image. This accusation is especially relevant in the climate and environmental focuses of CSR, where standards and definitions of what constitutes clean, green, or sustainable frequently vary by stakeholder, and accusations of “greenwashing”—promoting negligible acts that appear environmentally friendly while disregarding environmental principles in one’s mainline business—can be common.

Proponents claim that CSR pays off not only for society but also for the firm and its stakeholders. Proponents accede that firms must pay upfront costs by either investing in new CSR-
related tools, methods, and programs (e.g., non-environmentally intensive machinery or certification training) or missing out on short-term profits. However, some business literature observes that this upfront cost is negated by a firm’s long-run profitability and social legitimacy and that the firm may even be less vulnerable to legislative changes or scandal (L. Burke and Logsdon 1996; Davis 1973; Hill et al. 2007).

An ample body of economic literature examines whether CSR-driven firms produce better or worse returns. However, there is no consensus definition for social responsibility or how it should be measured. Many of the studies employ a single metric to quantify socially responsible action. Metrics can include, but are not limited to, ratings from peer institutions, annual reports, and reputational indices; these metrics are then linked to the stock valuation or another economic performance indicator of a company. This is a persistent challenge and has led to debates over CSR’s efficacy.

Along with the issue of a single metric failing to capture a company’s social performance, finding statistical significance and meaningful results has also been problematic (L. Burke and Logsdon 1996). Several studies have shown that customers are willing to pay a premium for ethical or more socially responsible brands (Besley and Ghatak 2007; Geczy, Stambaugh, and Levin 2005) and that assets in social and environmental investments are increasing (Hill et al. 2007; Geczy, Stambaugh, and Levin 2005). Yet other surveys show that 70 percent of food shoppers base their purchasing decisions on price rather than ethics and only 5 percent actively purchase ethical products (Doane, 2005), suggesting that CSR is providing negligible to marginal utility and serves only a niche market of higher-end customers. Some may argue, that CSR is therefore unsuitable as a broad substitution for traditional government regulation, however, theory at the firm and individual levels argues that multi-attribute utility functions incorporate values; for the individual utility is not only based upon risk and return but also personal values (Bollen 2007) and for publicly traded companies, socially responsible activities can maximize the market value of the firm even when the present value of future cash flows are not maximized (Mackey, Mackey, and Barney 2007). Furthermore, the types of socially responsible activities undertaken can influence financial performance, with things like strengthening community relations bringing immediate positive returns but efforts like tightening labor and environmental standards presenting up-front costs for firms (Barnett and Salomon 2006).

Other proponents state that CSR can differentiate firms within an industry, attracting socially conscious clients, customers, and investors thus providing a substantial short-term reward for enacting CSR (Drumwright 1994). This is sometimes referred to as a company’s “good will” and can be related to its branding efforts. Some proponents also claim that CSR would improve the social image of firms that proactively adopt it, which in turn would increase profits by generating a more loyal or larger customer base (L. Burke and Logsdon 1996; Davis 1973; Hill et al. 2007). Still other proponents see CSR as a way for public goods to be underwritten when a market incentive is lacking (Baumol 1970). However, proponents of CSR acknowledge that executive leaders must not overemphasize social good and philanthropy to the extent that their primary goal to obtain a profit is compromised (Murphy 1994).
Yet opponents might still counter that the arguments above do not definitively prove that social goods would increase through a stricter regulatory approach instituted by an effective government. Though CSR has seen a rise in popularity and is increasingly discussed in business schools and media, empirical analyses of CSR initiatives do not paint a definitive picture whether the approach results in statistically significant improvements to either public goods provision or profitability (Ullmann 1985; Wood and Jones 1995).

Some of the salient literature grounding CSR in economic theory and price theory comes from (Besley and Ghatak 2007; M. Bagnoli and Watts 2003; Kotchen 2006). All three papers apply general economic models and public finance concepts and classify CSR activities as private provision of a public good25. This rather narrow area of literature sits within the broader research area focused on the occurrence of sufficient public good provision through private actions (including public finance economics and ecological economics), even with the existence of free-ridership (Warr 1983; Bergstrom, Blume, and Varian 1986)26. Looking more specifically at CSR, these authors ((M. Bagnoli and Watts 2003; Besley and Ghatak 2007; Kotchen 2006) are all able to demonstrate that CSR within profit-maximizing firms parallels other models of private provision of public goods—they all view CSR to be an act of private provision of public goods.

In general, theoretical literature tends to conclude that CSR is a less than optimal way to produce both profits and public goods, but caveat that these findings do not hold when the government works imperfectly. In an imperfect government situation, CSR may prove beneficial relative to the traditional economic model. Belsey and Ghatak modeled the feasibility and desirability of CSR in both perfect and imperfect government scenarios. Conclusions of the authors are discussed in detail in this research alongside the conclusions from their peers Bagnoli, Watts, and Kotchen (M. Bagnoli and Watts 2003; Besley and Ghatak 2007; Kotchen 2006).

Rarely is any government able to act in what literature would classify as a perfect manner, and complex public goods like the global climate only add to the difficulty, which means that CSR could be Pareto-improving27 in most instances. It is important to remember that these economic models are simplifications of reality and make many assumptions about entity behavior. And while the conclusions from the theoretical literature are rather consistent, the empirical literature, although limited in scope, presents mixed conclusions on the efficacy of CSR.

25 Slightly related are the competitive implications to firms from internal employee-activism associated with CSR (Baron 2001). For example, (Nyborg, Howarth, and Brekke 2006) examine the associations between motivated workers and CSR initiatives, while (Hemingway and Maclagan 2004) propose a theoretical framework for leader values motivating CSR adoption.

26 Refers to the use of a public good without paying a fair share to sustain it (Andreoni 1988).

27 Pareto efficiency or Pareto optimality refers to a scenario where resources are allocated such that no redistribution can occur without making at least one individual or preference worse off. See footnote 6. When this is not the case, an action is said to be Pareto-improving if it moves the scenario closer towards this state(Chappelow 2019).
2.3.3 Considering Scenarios Where CSR Could Be Beneficial

While governments are usually the providers of public goods and regulators of public bads, CSR offers the ability to shift at least a portion of public good provision and public bad curtailment into the private arena. While antithetical to traditional economic understanding of public and private roles, more recent publications in behavioral economics argue that this understanding could be closer to reality. It nonetheless may not be the most Pareto-efficient\(^{28}\) approach to maximizing profits and balancing them with the externalities of public goods and bads (E. Anderson 2000; Bânabou and Tirole 2010; Henrich et al. 2001).

This section reviews some of the literature to better understand if situations exist where shifting the provision of public goods to for-profit private sector actors engaging in CSR could be advantageous. The literature and associated models provide an economic motivation for why firms (or private industry) would engage in financing the climate change response. Specifically, the literature here shows that private industry can provide public goods via socially responsible practices (i.e., CSR) at or above the same level of a government. Additionally, there exist many government situations (described as “imperfect” in earlier parts of the chapter) when socially responsible engagement from the private sector is an appropriate mechanism that the government should incentivize. While this literature review examines the potential case for CSR in general, the findings are applicable to the matters of climate-related CSR, climate finance, and in particular the generation of green bonds.

Friedman and many economists analyze CSR under the assumption of a government being capable of regulating public goods and bads and doing so effectively. As the previous analyses have demonstrated, CSR is not Pareto-improving under this condition. However, in an imperfect government scenario, this conclusion may not necessarily hold. Imperfect governments are characterized as those that are unresponsive to their citizens’ needs and otherwise corrupt, inefficient, and/or wasteful, but it should be recognized that governments may be unable to sufficiently provide an appropriate level of public good and prevent a public bad for any number of more benign reasons, especially as we consider climate change and goods that transcend national boundaries. In particular, understanding of the public good’s value may be limited and subject to debate amongst a government’s citizens.

Besley and Ghatak (2007) propose a simple economic model to demonstrate that public goods can be produced at the same level under CSR as through a voluntary contribution equilibrium\(^{29}\) (Mark Bagnoli and Mckee 1991; Buchholz, Cornes, and Peters 2006; J. C. Cox and Sadiraj 2007). The model

\(^{28}\) “Pareto efficiency, or Pareto optimality, is an economic state where resources cannot be reallocated to make one individual better off without making at least one individual worse off. Pareto efficiency implies that resources are allocated in the most economically efficient manner, but does not imply equality or fairness.” (Chappelow 2019)

\(^{29}\) An economic term from the voluntary contribution model (VCM) in the study of the provision of public goods. The equilibrium achieved when then public good is financed by voluntary donations—not “Using the VCM mechanism it has been shown that when one increases the group size participating in the experiment the amount of ‘free riding’ increases. In addition, a reduction in the marginal return from contributing to the public good will also increase the amount of ‘free riding’ present” (Isaac and Walker 1988; EconPort 2018).
is described as the observation of “product market competition among firms for ‘ethical’ and neutral consumers…and allows [the authors] to investigate both feasibility and desirability arguments.” The model contains a public good and two private goods, where one private good is the numeraire\(^{30}\); both the public and private goods are valued by both consumers and producers. For additional model details, see Appendix G. This situation is similar to the case of climate financing where public and private financing (i.e., goods) are provided and both valued. Besley and Ghatak (2007) make a few notable observations based upon their simple model. In order for CSR to be logical, the government must be unable to provide the public good at the first-best level\(^{31}\).

- First, in this model, free-riding occurs in the CSR scenario just like in voluntary public good contributions. The public-good will then be underprovided (below the first-best level)\(^{32}\)—(M. Bagnoli and Watts 2003) also find the same conclusion when the first-best level is obtained through the Lindahl-Samuelson\(^{33}\) condition.
- Secondly, in a competitive CSR equilibrium\(^{34}\), consumers that care about the public good will prefer the firms providing and what they perceive as the ethical version of the private good. The allocation will be robust if each consumer’s valuation of a public good is private information and not publicly revealed.
- Thirdly, the allocations from the first-best outcome cannot be obtained, but a Pareto-improvement can be achieved by tightening CSR standards for firms catering to consumers who care about the public-good.

---

\(^{30}\) Numeraire goods allow for relatively good comparisons when a specific price is not relevant. Prices of all other goods are normalized by the numeraire good.

\(^{31}\) First-best is an economics concept based around Pareto efficiency to determine the optimal allocation between goods - if the most desirable (i.e., optimal and first-best) economic conditions cannot be satisfied, then the remaining conditions should (Paul Anthony Samuelson 1948; Paul A. Samuelson 1950). Conditions: “(i) The first-best Paretian conditions: between each pair of commodities, the marginal rate of substitution in consumption is equal to the marginal rate of transformation in production and (ii) the first-best distribution conditions: between each pair of individuals, income or wealth is distributed in a way that reflects some ethical judgment of the best distribution.”(Michie 2014)

\(^{32}\) “The theories of the first, second, and third best are concerned with public policies to achieve the best possible economic outcome, yet differ with respect to the treatment of the constraints. In the ‘first best,’ only resource constraints limit economic outcomes. In the ‘second best,’ distortions, e.g. from monopolies and externalities, also limit outcomes” (Michie 2014).

\(^{33}\) This is an optimal allocation where the sum of the quantity of private goods that consumers are willing to give up for an additional unit of public good will equal the quantity of the private good required to produce the additional unit of public good (Paul A. Samuelson 1954).

\(^{34}\) Equilibrium is a fundamental economics concept that equalizes supply and demand or prices such that markets clear. “Competitive equilibrium is a condition in which profit-maximizing producers and utility-maximizing consumers in competitive markets with freely determined prices arrive at an equilibrium price. At this equilibrium price, the quantity supplied is equal to the quantity demanded. In other words, all parties—buyers and sellers—are satisfied that they’re getting a fair deal” (Liberto 2019).
Fourth, the authors show that under CSR, crowding-out will occur if the government increases the public good provision; less-competitive firms will cease their CSR initiatives and instead comply with the new baseline set forth by the government.

Fifth, and finally, the authors consider “warm-glow” preferences for public goods like Bagnoli and Watts (M. Bagnoli and Watts 2003). They add the “warm-glow” preferences to the public-good-caring consumers and notice that the results are basically unchanged. They demonstrate that public goods can be provided at a level below the first-best at equilibrium.

These instances are counter to Friedman and other skeptics’ positions on CSR. While private institutions are at a disadvantage (in comparison to government) because they are unable to tax, they do have the ability to cater to subsets of the population, particularly minority groups (i.e., consumers that take particular care for certain public goods like those that seek out animal-cruelty-free products), and serve as a useful supplement to government policy that must balance the interests of all parties in its decision-making. Furthermore, when monitoring might be costly and difficult, CSR could play a vital role.

A smaller subset of academic research, (Kotchen 2006) and (Seema Arora and Gangopadhyay 1995), consider private sector efforts for social good particularly related to the “green” or environmental market. (Seema Arora and Gangopadhyay 1995) focus on competition for environmentally-conscious consumers void of public goods, which is less germane to the research line of effort of this paper. Kotchen (2006) examines consumer choice and the effects of an impure public good in a green market. Specifically, he looks at the choice between consumption of an impure public good that generates a joint private and public good or consumption of a private good followed by a contribution to a pure public good. He finds that in a voluntary contribution equilibrium, bundling a private and public good (i.e., the impure public good) does not change the provision of public goods at equilibrium, but does depend on whether the good is a complement to or substitute for a private good. In terms of the environmental market, private and public bundling will crowd out donations, it will always increase the provision level, and may not always be welfare-improving.

Doane (2005) examines CSR in a less theoretical framework, where CSR is instead always embedded in market structure. In this case, it will never be immune to market vagaries. Ultimately, a company must be incentivized through the market to invest in CSR initiatives. Her work claims that while firm reputation, protection of assets, consumer demands, and regulation avoidance are all reasons a firm might adopt CSR, market failure under CSR would require an entirely different approach. She further argues that CSR is associated with promises that do not materialize. First, she argues that CSR involving voluntary reporting will not actually improve performance. Secondly, she argues corporate

---

35 Andreoni (1989, 1990) developed the theory of “warm glow” giving based upon two motivations individuals have: (1) “altruism” based upon the work of Borro (1974) and Becker (1974) where individuals desire/demand more of the public good and (2) utility increases for an individual from giving by selfish motive - this is the warm glow.
behavior cannot be altered through voluntary ethical codes. And finally, she says that consumers and the investment community will not incentivize companies to act in a more socially responsible manner.

These cases have particular relevance to the climate problem, where recent evidence suggests that many firms are acting faster than governments to new knowledge of climate risk. Proponents of action on climate change have also argued that governments are insufficiently monitoring their own greenhouse gas emissions (particularly outside the energy and transportation sectors), much less and responding to climatic risk in order to reach internationally agreed targets like limiting warming to two degrees Celsius. Thus, imagining governments as imperfect in the face of a complex problem like climate change is not difficult.

These studies, though not attaining a consensus, provide useful insights for further investigation into financing the response to climate change, also known as climate finance. It is also easy to imagine tiers of consumers that value carbon offsets or financial products differently or not at all, as in Belsey and Ghatak’s model. The value that a stable climate provides as a public good - often attempted to be distilled as the social cost of carbon - has proven difficult to resolve, especially as its impacts are harder to perceive than safe air and water, making the valuation of zero for a portion of the population, though scientifically impractical, politically plausible. Empirical evidence could be used to identify whether Belsey and Ghatak’s five conclusions are borne out in practical experience or one could look to determine whether Kotchen’s conclusions of donation crowding are borne out by the evidence; answers here could determine whether models of behavior under CSR are accurate or perhaps still missing elemental components in explaining a complex phenomenon.

2.4 Climate Finance As An Example Of CSR

CSR can address many facets of public policy; health consequences of tobacco, conflict-free minerals, and equitable wages are a few among the many examples discussed in academic literature (Jenkins 2004; Hirschhorn 2004; Welford 2005). As described earlier in this chapter, environmental impacts are a common theme for CSR due to the challenge of measuring environmental benefits. This section focuses on the environmental challenge presented by climate change because challenges CSR stakeholders face are unique on account of being urgent, uncertain, transnational, and multifaceted. These characteristics are reflective of a growing number of policy challenges in the modern world.

As detailed in Chapter 1, the response to climate change will take on several forms, most of which are categorized as mitigation and adaptation. However, carbon-neutral and climate resilient technologies are either not yet available or have not yet scaled to the point where they are economically competitive or even economically viable, although advancements are anticipated during the next several decades. As a result, resources and funding will need to be provided to stakeholders in order to ensure such mitigation and adaptation occurs, because the long-term economic incentives to do so often conflict with shorter-term economic operations and many developing countries lack the capacity to do so themselves.
Under traditional economics framing, national governments could tax or regulate greenhouse gas emissions and mandate or incentivize the development of climate resilient infrastructure (including financing mechanisms) immediately. However, governments have been reluctant to take this step out because of the comprehensive challenge it presents. Some countries have started to regulate GHG emissions domestically through measures such as carbon emission prices or through less direct measures like renewable energy or car efficiency regulations, but efforts in many major emitting countries as well as in the international agreements have trended towards voluntary measures. In fact, one key difference between the Kyoto Protocol and the Paris Accord was the pivot from countries agreeing to comply with legally-binding reduction targets to countries volunteering their own individual targets. The prevalence of voluntary mechanisms in the global effort to address climate change at least contributes to an environment where CSR can provide key contributions. Some countries have started to regulate GHG emissions domestically through measures such as carbon emission prices or through less-direct measures like renewable energy or car efficiency regulations, but efforts in many major emitting countries as well as in the international agreements have trended towards voluntary measures.

The international community acknowledges that to date governments have not taken the actions necessary to meet the internationally-agreed target of limiting greenhouse gas emissions such that global average temperatures do not exceed 2 degrees Celsius above the pre-industrial average. Furthermore, capital flows show that governments alone are not on track to provide the USD 100 billion a year by 2020 to developing countries pledged during the Copenhagen summit of 2009. Given the prevalence of voluntary mechanisms and the acknowledgement of failures in achieving its goals, it is fair to conclude the UN-coordinated response to climate change, and in particular delivering funding to those who are most vulnerable, has been a demonstration of governance that is “imperfect.” This in turn suggests that climate change is a policy space where CSR has a role to play in addressing the gaps in the global community’s response to climate change, particularly in providing financing.

The CSR proponents’ arguments that implementation promotes longer-term profitability is of particular relevance to CSR in the climate space, as there is a general consensus among many observers, particularly major energy firms, that carbon emission regulations are likely to intensify at least over the span of decades, providing those firms who are taking proactive voluntary approaches to low-carbon energy adoption as being more prepared for profits over the long term. Similarly, by facilitating the finance of climate-resilient activities, firms may presume that their CSR activities in the climate finance space are entering into a growth market that future government action will only further encourage. In this way, firms engaged in facilitating climate finance vehicles like green bonds may not only view their activities as ethically fair and brand-building but a sound business growth strategy for the long term.
2.5 Regulatory Considerations Associated With Private Engagement

Many regulatory questions will arise relating to whether CSR enables private companies to produce public goods at levels superior to the government alone. If corporations can indeed produce superior social and environmental outcomes through voluntary actions, such as CSR, then the government’s optimal approach to regulating those public goods changes significantly. When CSR is Pareto-improving and firms are willing to invest in CSR, the situation implies that some companies are willing to voluntarily perform above legal requirements. This voluntary compliance reduces the need for standards and regulations for at least a subset of firms and may suggest regulation has moved too slowly regarding the industry overall. If voluntary compliance is able to “pull” the rest of the industry towards greater provision of a public good without the impetus of regulation by an imperfect government and at lower costs, it could have profound effects across many policy domains (e.g., climate policy and emissions standards, labor policy, working conditions, and wage standards).

However, if CSR is ineffective in delivering solutions from private industries to global problems such as climate change and poverty, perhaps CSR is not a panacea as claimed by proponents. Perhaps it is not worth eliminating (provided it does provide some good), but rather worth prioritizing other initiatives and policies. As Doane (2005) explains, CSR was the answer to the question “how can business minimize its negative impacts on society and the environment?” but perhaps this is the wrong question. Instead “what institutions, organizations, or actions do we need to deliver a sustainable society?” might be a more appropriate question, and the answer involves institutional and behavioral reform where CSR is one piece of the puzzle.

Most of the economic models discussed in this chapter deal with public goods in a broad and general sense. However, the conclusions can be applied to many policy scenarios. For example, Besley and Ghatak’s (2007) model was applied to the product market context but could be extended to apply to donations in the labor market and ethical investments in the capital market. Kotchen’s (2006) application touches on the environment through green technologies, but does not explicitly address climate change policy or bonds that finance renewables. As demonstrated in this section, these broad conclusions have direct implications for how private-led action can affect climate change policy and financing.

2.5.1 Policy Implications For CSR Addressing Climate Change

In an era beset by multiple environmental concerns and gradual steps in addressing climate change, perhaps society is expecting the private sector to assume responsibilities traditionally assigned to the public sector. Climate policy may indeed be one of the biggest beneficiaries of CSR. Private industry accounts for almost half of GHG emissions and is a central player and target in climate regulation (Oikonomou, Patel, and Worrell 2006). More than half of greenhouse gas emissions now come from developing countries36 (Johannes Friedrich 2017), where corruption and waste tend to be

---

36 Classified in the Kyoto Protocol as Annex II countries (see Appendix D for a list of Annex II countries).
more prevalent—in these cases, the imperfect government classification applies. All the while, awareness of the climate change issue continues to expand globally and becomes more prominent in many corporate and social statements. This presents an environment and a challenge where CSR may thrive. In this context, the private sector may be incentivized to adopt CSR, seeking not only a potential sound long-term business strategy but also acquisition of an expanding market segment, which would benefit climate policy and GHG emission reduction. In some countries and industries, the Kyoto protocol has already spawned this reaction where offset-trading in the clean development mechanism (CDM) and joint-implementation (JI) markets greatly incentivized private entity participation in low-carbon projects.

These questions of efficacy of private engagement in socially responsible engagement become even more pressing and challenging when considering CSR as it relates to climate change. Climate regulation is relatively young, evolving, and regulations are likely to become more stringent during the next several decades. Both governments and corporations (i.e., public and private) are responding to public concerns about the problem. Still, the observers are dogged by a lack of information as to whether climate initiatives such as funding a transition of corporate facilities to 100-percent renewable energy have improved profitability or provided a sound social good in mitigating climate change. In addition, lack of consensus around an outcome metric could lead to different conclusions regarding the efficacy of government or corporate actions. On the profitability side, the rapidly changing economics of wind, solar, and energy storage (along with divisions between adaptation and mitigation) have complicated the debate as to whether certain corporate moves are CSR-related or merely cost-saving. While observers may welcome the social good it provides regardless, policy analysts care deeply if the actions being observed are indications of a new paradigm about public good provision or simply business as usual, and what lessons if any may apply to other policy domains. Policy action in this realm will carry significant implications for governments, corporations, and investors in the next several decades, as will an understanding as to which policies are economically optimal.

---

37 Some options for the social good may be measured by kilowatt hours of renewable energy generated, emissions rates of carbon dioxide, concentration of greenhouse gases in the atmosphere, or global average temperature.
This research uses textual data extracted as a relevant subset from the LexisNexis database, a global aggregation of thousands of historical news, business, and public record sources (e.g., blogs, legal cases, or patents). The sub-corpora contain content from the following categories: newspapers, industry trade press, web-based publications, newswires and press releases, magazines and journals, news transcripts, newsletters, and aggregate news sources. Due to the diversity in content and geography, this research assumes the sources contained within LexisNexis represent public discourse.

This chapter describes the methodology used to obtain the datasets for the analyses presented in subsequent chapters. The final datasets (two sub-corpora) represent portions of the LexisNexis database (the parent corpus) that contain information pertinent to climate finance and green bonds. These two data subsets create the source data for further analysis, namely topic modeling variations (e.g., static, evolutionary, and dynamic), to answer the research questions.

The sheer size of the LexisNexis database presents computational complexity, and more importantly is not fully accessible to the end user. Additionally, the entirety of information from all aggregated news and blog sources within the LexisNexis database span a multitude of concepts, many of which are not relevant. In the case of this research, the concepts of interest are climate change, climate finance, green bonds, and environment social and governance; these concepts represent larger ideas that embed related topics. Extracts need to be obtained that contain the concepts of interest.

The process of creating a relevant extract requires identifying the best set of keyphrases that represent climate finance and green bond concepts. And then using these keyphrases to extract subsets of pertinent data representative of the public discourse around those two concepts. Representation of the public discourse comes from the parent corpus (LexisNexis Database). This parent corpus requires filtering by matching content (news headlines, news article text, or blog text) to the specified keyphrases corresponding to the concepts of interest. For the purpose of this analysis, a simplified approach to keyphrase identification is pursued - primarily using crowd-sourced internet search queries made available through Google Trends (Google 2018b). The relevant subset of information that is returned using the keyphrases, constitutes the sub-corpora used for analysis. This process of identifying salient keyphrases and extracting the sub-corporas that represent climate finance and green bonds is outlined in Figure 3.1. The left portion of the figure outlines the steps and the right portion provides examples or illustrations of the steps.

---

38 Note that this chapter uses terminology common in the information extraction and natural language processing fields, where appropriate explains the terms in footnotes.
Figure 3.1: High-level schematic of the keyphrase identification (i.e., search strategy) and sub-corpora extraction. Green check boxes indicate confirmation and the input above is passed to LexisNexis.

The remainder of this chapter provides context and details on each stage of the process. Section 3.1 will review the background of applying keywords or keyphrases in information extraction. Section 3.2 builds on section 3.1 by introducing keyphrases from Google Trends and research (industry and academic) use of tools. Section 3.3 makes relative comparisons of the temporal trends related to the concepts and initial keyphrases. Section 3.4 presents the process of incorporating crowd-sourced behavior (Google Trends and Twitter) to generate a consolidated set of keyphrases. Lastly, section 3.5 describes the sub-corpora (and associated meta-data) returned from the parent corpus, the LexisNexis database, using the consolidated keyphrases.

3.1 Keyphrase Relevance For Information Extraction - A Search Strategy

Tackling unstructured data brings together research from many fields spanning Information Extraction (IE), Natural Language Processing (NLP), machine learning (ML), computation, and linguistics (Berry and Kogan 2010). The declining cost of storage and advancements in technology (e.g., file storage systems and parallel processing) have made applications that combine structured and unstructured data more accessible. However, the problems associated with identifying structure from inherently noisy data is still challenging.

Information extraction is a burgeoning field, due in part to the increased volume of unstructured data, the speed at which this data is produced, and the emergence of methods to analyze and extract
useful information from the data. Information extraction is grounded in the processes of NLP with focus on extracting entities, entity relationships, and entity attributes primarily from unstructured text (Sarawagi 2008). Entities represent common, nominal things and are typically nouns (e.g., “climate”, “finance”, “climate finance”, “bond”, “green bond”, etc.). In this research the concepts of interest (e.g., “climate change”, “green bonds”) might be represented by multiple entities and these entities could be a keyword (e.g., “climate” or “bond”) or a keyphrase (e.g., “climate finance” or “climate bond”). For the remainder of this chapter, the concepts of interest are the initial keyphrases - specifically (1) “climate change”, (2) “climate finance”, (3) “environment social and governance” (as in ESG investing), and (4) “green bonds”.

This research draws upon the advancements in the field of NLP to collect, process, and analyze data to answer applied policy questions. The research questions studied here require a link between a concept and a set of keyphrases. Entire areas of research can focus on identifying these salient keyphrases from a large corpus of documents; in this research the process of identifying salient keyphrases to be used to extract a sample from a larger database is called a search strategy.

3.2 Google Trends For Keyphrases Representing Collective Behavior

Various methods exist to obtain the pertinent keywords or keyphrases. One avenue relies upon the methodology built into free tools offered by companies like Google (Keyword Planner and Google Trends). There are also paid services and companies that will identify these keyphrases using proprietary methods and algorithms. The academic approaches within the information retrieval domain, rely upon keyword extraction research. In the academic setting, the approaches include unsupervised (Kang, Domeniconi, and Barbara 2005) and supervised text-mining applications (Wan, Yang, and Xiao 2007; K. Zhang et al. 2006; C. Zhang 2008) and graph-based approaches (Litvak and Last 2008); in general these approaches are used to find one or more keywords that best describe the text within a document or collection of documents. These more rigorous academic approaches are useful when a body of literature already exists for the desired concept (unlike in this research).

39 Some frequent NLP tasks around entities are entity linking and named entity recognition. “Entity linking is the ability to identify and disambiguate the identity of an entity found in text (for example, determining whether the “Mars” is being used as the planet or as the Roman god of war). Named entity recognition is the ability to identify different entities in text and categorize them into predefined classes.”

40 This idea of search strategy or process of keyphrase identification spans many applications, but one of the most widespread occurs in the private sector. For companies and corporate product marketing, a common exercise involves identifying terms related to a concept, as well as, terms that co-occur with the related concept. The company or product represents the concept and a common goal is to improve search engine query results or product placement by automatically identifying the keywords that optimize search results. Take for example a financial evaluation company building brand reputation around sustainability ratings. This company will want to identify common search engine terms related to sustainability ratings and ensure their company is near the top of the returned websites (i.e., “optimize search results” so the company is related to this concept). The search terms, related to the concept “sustainability ratings”, will likely involve the following collection: “environmental social governance ratings”, “esg”, “socially responsible”, and “sri”.

55
Past studies have applied data from Google Trends to predict (H. Choi and Varian 2012) or inform behavior related to financial markets (Da, Engelberg, and Gao 2011, 2015). This includes correlating Google search behavior with stock market moves (Curme et al. 2014; Preis, Moat, and Stanley 2013) and volatility (Vlastakis and Markellos 2012/6) and using it as an input for forecasting stock demand (Carrière-Swallow and Labbé 2013). Signos used abnormal changes in Google search behavior to identify alterations in firm structure via merger and acquisition activity (Siganos 2013/9) and earnings releases (Drake, Roulstone, and Thornock 2012). In addition to this private sector behavior, Vosen and Schmidt show Google trends can inform individual consumer demand behavior (Vosen and Schmidt 2011). Based upon this evidence across various areas of the financial markets, applying Google Trends (Google 2018b) data to inform keyword extraction appears to be a reasonable approach.

Google Trends data are compiled using an unbiased sample from all the searches in Google since 2004. These trends are considered to represent collective behavior. Google collects the searches, anonymizes and categorizes the data and then connects it to a topic (Google 2017). They remove duplicate searches, sparse or searches with low frequency, and special characters (Google 2017). The data are also normalized by the time and location so that places or times with only the most volume do not always rank highest (Google 2017).

3.3 Google Trends For Concept Exploration

This section summarizes the use of Google Trends to explore four concepts: (1) climate finance, (2) environment social governance (ESG) investing, (3) green bonds, and (4) climate change. The temporal trends revealed by Google Trends represent collective public interest based upon Google users’ search queries. The climate change and ESG concepts provide relative comparisons to the other two concepts (climate finance and green bonds); the research here is narrowed to focus on the climate finance (Chapter 4) and green bond (Chapter 5) concepts.

The normalized Google Trends scores are displayed over time in Figure 3.2. The numbers in the figure are created by Google Trends and represent the interest from searches relative to the highest point over a particular time. The upper-bound of 100 represents the peak popularity and a term half as popular would have a score that is half (in this case 50) and a zero score would indicate popularity only one percent that of the peak. Relative to the other concepts, climate change is the concept with greatest interest, as measured by Google Trends. The climate change search term dwarfs the other three concepts, so much so that the climate finance keyphrase cannot be seen in the figure.

---

41 According to Google Trends documentation, Google applies the following data normalization: “Each data point is divided by the total searches of the geography and time range it represents, to compare relative popularity. Otherwise places with the most search volume would always be ranked highest. The resulting numbers are then scaled on a range of 0 to 100 based on a topic’s proportion to all searches on all topics. Different regions that show the same number of searches for a term will not always have the same total search volumes.”
Figure 3.2: Google Trend interest score over time for searches: “climate change”, “green bonds”, and “environmental, social, and governance” (red). Searches are in green, yellow, and red, respectively. Note the search “climate finance” does not appear here because the “interest” or trend score is so much lower relative to the search phrase “climate change”.

Source: Extracted and modified from Google Trends (Google 2018b).

This time-series “interest” results displayed in Figure 3.2 are not surprising; it is likely that climate change is the most colloquial topic and the other three topics are related to climate change but less well-known among the general population. This hypothesis is evidenced by the broad geographic popularity of the climate change search phrase (shown in green in Figure 3.3a) as compared to climate finance that is US-centric (Figure 3.3b). Similar to climate finance, both green bonds (Figure 3.3c) and ESG (Figure 3.3d) searches show popularity in the UK in addition to the US. It can be inferred from this geographic dispersion the concept of climate change has global reach, crossing development boundaries, whereas the climate investment concepts tend to be more localized in developed financial economies.

Figure 3.3a: Geographic distribution of “climate change”

Figure 3.3b: US-centric geographic distribution for “climate finance”

Figure 3.3c: US-UK localization for “green bonds”

Figure 3.3d: US-UK localization for “environment, social, governance”
Note for Figure 3.3: All figures created directly from the regional exploration function in Google Trends, which allows one to view the geographic distribution of origin in the corresponding search terms. This display is international because this research is focused on international policy responses (Google 2018a). Source: Google Trends Region Exploration

In Figure 3.4, with the removal of the climate change phrase an increasing trend among the finance related terms appears. This increasing trend is not evident in Figure 3.2 because the scores normalization includes the expansive and globally used search for climate change. Figure 3.4 shows that from 2011 to 2013 the popularity (as measured by Google Trend interest) of climate finance and green bonds was comparable. Before 2013, both climate finance and green bonds scores were very low, implying that attention to these concepts was minimal relative to ESG and climate change. In more recent time the popularity of all finance concepts appears to be increasing, with the rate of ESG and green bonds exceeding that of climate finance.

**Figure 3.4:** The same time-series as shown in Figure 3.2, for “climate finance”, “environmental, social, and governance”, and “green bond” search interest. With the axis rescaled and the climate change interest score removed, the three concepts show similar trends. Source: Extracted and modified from Google Trends (Google 2018b).

### 3.4 Creating Consolidated Keyphrases

During the creation of the sub-corpora, the keyphrases used to match and extract the content should capture the entire concept of interest. Google Trends related queries functionality helps ensure the keyphrases will encompass variations in keyphrases representing the concept used by the public. Figure 3.1 illustrates the expansion of the keyphrase list for each concept using the related query functionality in Google Trends. The search strategy uses these related keyphrases to arrive at the final consolidated keyphrase list.

Based upon the concept exploration (Section 3.3), related keyphrases are not identified for climate change because the concept of climate change appears broader than the desired financial focus. For each of the financial concepts, Google Trends is leveraged to obtain related keyphrases. Table 3.1 shows each initial keyphrase (i.e., concept) and the related and trending keyphrases as identified through Google Trends.
Table 3.1: Related keyphrases (Google search queries) for “green bonds”, “environment social and governance”, and “climate finance” identified by Google Trends. These are phrases Google identifies as related based upon aggregated user searches where capitalization does not impact results - related results are returned uncapitalized. The related keyphrases might change over time due to alterations of Google’s underlying algorithm (Google 2017).

<table>
<thead>
<tr>
<th>climate finance Google Trend Keyphrases</th>
<th>green bonds Google Trend Keyphrases</th>
<th>environmental social and governance Google Trend Keyphrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>green climate fund</td>
<td>green bond(s) principles</td>
<td>morningstar esg ratings</td>
</tr>
<tr>
<td>oecd climate finance</td>
<td>climate bond(s)</td>
<td>msci esg index</td>
</tr>
<tr>
<td>climate policy initiative</td>
<td>green bond(s) india</td>
<td>esg etf</td>
</tr>
<tr>
<td>ontario green bond(s)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figures 3.5 and 3.6 show the time-series of the Google Trends’ interest score for the initial and related keyphrases. From Figure 3.4, it appears that the keyphrase “green bonds” has predominantly more public interest but some interest still originates from the related keyphrases “climate bonds” and “green bond principles”. Figure 3.6 shows that expanding the keyphrases for “climate finance” to include “green climate fund” is capturing a keyphrase with greater public interest (as measured through Google Trends) as opposed to using only the keyphrase “climate finance”. By expanding the keyphrases to include related keyphrases the content matched to the keyphrases should be more inclusive of the public interest surrounding the concepts.

Figure 3.5: Green bond term and related terms interest scores over time from Google Trends. Source: Extracted and modified from Google Trends (Google 2017)

42 “morningstar” and “msci” are rating/evaluation companies and “etf” “stands for exchange traded fund. The list contains select terms and removes non-financial search queries like “esg vs dat” and “esg vs hr”. These two phrases are unlikely finance terms and judging from Google search results refer to live-streaming video game competitions.
The keyphrase list is further winnowed because not all of the words in the keyphrases identified in Table 3.1 are required to do the content matching for the sub-corpora extraction. For example, the keyphrase of “green bond” will match content with either the plural or singular form and longer phrases including either form—hereafter, specified by green bond(s). The keyphrase “green bond” will not capture “climate bond” but does capture all of the following: green bond(s), green bond(s) india, ontario green bond(s), and green bond principles. Hence, a consolidated keyphrase list of “green bond” and “climate bond” is sufficient to capture all relevant keyphrases. The consolidated list of keyphrases are identified in Table 3.2. These are the keyphrases used to match the content from within the parent corpus (LexisNexis) and extract the sub-corpora.

<table>
<thead>
<tr>
<th>Consolidated Climate Finance Keyphrases</th>
<th>Consolidated Green Bond Keyphrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>climate finance</td>
<td>green bond</td>
</tr>
<tr>
<td>green climate fund</td>
<td>climate bond</td>
</tr>
<tr>
<td>climate policy initiative</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: Consolidated set of keyphrases to be passed to the LexisNexis database for information extraction. The plural forms are also returned, as well as, any variation in capitalization.

In addition to using the crowd-sourced Google Trends related keyphrases, a sample of Twitter hashtags was evaluated to determine if there were additional keyphrases. This analysis identified no new keyphrases. This investigation looked at a sample of Twitter and identified similar hashtags to those hashtags corresponding to the initial keyphrases, using a publicly available online tool. All hashtags that were identified as similar were permutations in capitalization of the initial keyphrases. For example, for “climatefinance” the related hashtags are “climateFinance” and “climatechange”. Due to the broad nature of climate change this term was already excluded and identified as a separate concept. The capitalization seen in related Twitter keyphrases (like in climate finance) were ignored because this will not impact the content matching.

---

43 See Appendix H for the figures from the online tool used to make this evaluation.
3.5 Sub-Corpora Extraction Using Keyphrases - Data For Analyses

As shown in Figure 3.1, the subsets of data are extracted using the consolidated list of keyphrases identified in Table 3.2. The consolidated keyphrases, that comprise the search strategy, leverage aggregate public search queries (see section 3.2). In addition to the unstructured data contained in the headline and article, meta data such as the publication source, date, and geographic focus of the article are available. The sub-corpora are extracted to cover a time-period from January 2010 to October 2018[44] and with no restrictions on the geographic focus of the contents. As shown in the bottom right of Figure 3.1, the sub-corpus used for research in Chapter 4 contains headlines (similar to titles) from the content, returned in multiple files in a comma-separated value (csv) format. Whereas the research in Chapter 5 uses the entire body of text from the news articles as well as more subjective information originating from blogs—this is all returned in a less structured Hypertext Markup Language (HTML) format.

In the analysis, the dates play a key role in understanding the temporal trends while the geographic meta-data serves to identify regional relationships. The time-stamps (dates) nearly always exist for the documents. However, the geographic meta-data of the document’s content is sparser and requires some post-processing. It is important to note that the geographic data field does not correspond to the document’s published location but rather the geographies referenced in the article. This data field that corresponds to the document’s geographic focus is a multi-class label assigned to each document by LexisNexis. The classes span countries and regions (both can be contained in the field). Each class label (geography) is assigned with a probability and all label probabilities do not necessarily sum to one. The meaning of this probability is somewhat ambiguous -- it comes from an internal model and logic applied by LexisNexis[45]. Despite the ambiguity, the closer the probability of the geography class is to one the more likely the document will pertain to said geography. When the value is closer to zero it is less likely the geography class is associated with the document. This research converts these class labels to binary values with a threshold of 50 percent. The labels assigned to the documents are multi-class and multi-label[46]. For example, consider a document which contains a geographic label of France with a probability of 60 percent, the US at 20 percent, and Western Africa at 50 percent. This conversion will result in a document label where France is one, the US is zero, and Western Africa is one—all other geographies not included in the document’s geographic data field will also be zero[47]. Additional post-processing of the unstructured data fields (e.g., stop-word removal, stemming, and lemmatization) is specific to each sub-corpus and discussed in the pertinent chapters that follow.

[45] The LexisNexis user guide states that “all news sources accessed by the [LexisNexis News Search] undergo the indexing process. They are analyzed for ... geographical locations identified,” but the guide does not provide further detail on how the geographical locations are identified or analyzed (LexisNexis 2018).
[46] Multi-class refers to problems where there are more than two outcomes. The multi-label problem allows an observation to belong to more than one class. Multi-class, multi-label has N classes where N is the number of geographies and documents can be assigned to more than one of the geographies.
[47] Sometimes referred to as one-hot encoding.
3.6 Sub-Corpora Processing Prior To Modeling

From the expansive corpus that contains both subjective and objective text sources in the LexisNexis database, two sub-corporas are created for use within this climate finance research. These two sub-corporas represent the general ideas surrounding the relevant material or areas of interest: (1) climate finance and (2) climate and green bonds. As depicted in Figure 3.7, the aforementioned consolidated keyphrases from the search strategy are used to extract pertinent sources of information. In the first case of climate finance, the titles associated with the news articles are consolidated in structured .csv files. In the second case of climate and green bonds, the titles and article contents of news and blog sources are extracted, parsed, and consolidated from HTML format. The article contents (from both news and blogs) are added in the second corpus because the number of articles are smaller, and the modeling used in conjunction with this dataset (i.e. the sub-corpus) relies on identifying more granular information (i.e. specific terms referenced in each article).

In both sub-corporas, post-processing steps are applied to “clean” and prepare the data for use as inputs to the machine learning topic models. These post-processing steps applied to the sub-corporas are outlined in Figure 3.7. In general, the first five post-processing steps serve to normalize terms and ensure similar terms are represented by the character strings. The last two steps, tokenize and vectorize, serve to prepare the data for a format the models can use. The application of these steps to the sub-corporas are described in Sections 3.6.1 and 3.6.2. These steps are common when working with text-based data, yet each step is a choice that can influence model outcomes. Despite the commonality of the post-processing steps, some of the approaches and implementations are novel and provide insight for follow-on research in machine translation and text extraction.

48 It is important to note that with deep-learning models or word embeddings, these steps may not be necessary or could change.
3.6.1 Climate Finance Sub-Corpus

The Climate Finance sub-corpus is the collection of news titles extracted and processed from the more expansive corpus within the LexisNexis database. This sub-corpus is based upon the consolidated keyphrases identified in Table 3.2. To prepare the text data as inputs to the models, the Climate Finance sub-corpus is processed at the word-level using standard NLP techniques. These techniques, which are illustrated in Figure 3.7, include: assessing relationships among words, creating a custom dictionary, creating custom stop-words, removing stop-words, lemmatizing, and filtering by term frequency and word length. After identifying the relevant articles from the keyphrases and applying the NLP processing steps, a sub-corpus consisting of 8,168 documents (i.e., the news article title or headline) and a vocabulary of 1,083 terms that come from 641 unique publications.

This processed sub-corpus serves as an input to the textual machine learning models that help describe the evolution of climate finance trends. The methods for these machine learning models are discussed in Chapter 4 and the results are discussed in Chapter 5. The modeling approach used for this research, known as “bag-of-words” (BOW) considers each word or bigram to be representative of a concept. To account for common concepts that may be represented by more than one term, some two and three-word terms (bigrams and trigrams) are concatenated. These bigrams are identified through frequency, graphical clusters (see Figure 3.9) and subject matter expertise. Take for example the word climate - by itself it has a meaning and that meaning changes slightly when considering bigrams such as “business climate”, “climate finance”, “Paris climate”, and “climate talks”. Table 3.3 shows the ten
most common bigrams containing the term climate and the frequency of these bigrams within the sub-corpus.

<table>
<thead>
<tr>
<th>Bigram With Term Climate</th>
<th>Frequency In Headlines</th>
</tr>
</thead>
<tbody>
<tr>
<td>climate change</td>
<td>1513</td>
</tr>
<tr>
<td>climate fund</td>
<td>471</td>
</tr>
<tr>
<td>climate talk</td>
<td>324</td>
</tr>
<tr>
<td>green climate</td>
<td>293</td>
</tr>
<tr>
<td>paris climate</td>
<td>239</td>
</tr>
<tr>
<td>climate finance</td>
<td>226</td>
</tr>
<tr>
<td>climate deal</td>
<td>166</td>
</tr>
<tr>
<td>climate action</td>
<td>106</td>
</tr>
<tr>
<td>climate summit</td>
<td>102</td>
</tr>
<tr>
<td>global climate</td>
<td>89</td>
</tr>
</tbody>
</table>

Table 3.3: Ten most common bigrams with the term climate.

Rather than adding the term “climate” to a stop-word list, it was concatenated in a number of instances to better capture the concept during the modeling. Figure 3.8 displays those terms that were concatenated.

49 The modeling is further described in Chapter 4 and uses bag-of-words. Alternative approaches might have considered n-grams or embedding vectors as the modeling input.
In addition to concatenating terms, common terms with similar meaning are mapped to a canonical form (Conghui Zhu 2007; Filip et al. 2006). This process, which makes terms like President Obama and Barack Obama consistent through word substitution or mapping, is known as text normalization and is executed through a custom dictionary. For example, “U.S.”, “U.S.A”, “United States”, and “United States of America” are all mapped (or updated) to “United States”. Similarly, “Paris Accord”, “Paris Talks”, “Paris Agreement”, and “Paris Treaty” are all mapped to “Paris Agreement”. Many of these terms were examined and identified using graphical plots. Figure 3.9 is an example of one such plot that displays the relationship and density between groups of words. Other terms were added to the custom dictionary through manual inspection and subject matter expertise50.

---

50 Note that future work that uses word embeddings, either trained on this corpus or pre-trained on a broader (and larger) corpus, could overcome the need for this preprocessing step.
Figure 3.9: Graphical relationship of words that include the term climate. The direction of the arrow indicates the word ordering - for example the arrow pointing from “prime” to “minister” indicates “prime” is the first word of the bigram. The darkness of the arrow indicates the strength of the relationship (i.e. greater frequency among all bigrams). This figure uses the R igraph package.

Prior to concatenation, words are converted to lowercase and lemmatized\textsuperscript{51} as shown in Figure 3.10. The remaining processing is done during the creation of the document term matrix. These rules include the following:

- Remove any punctuation
- Remove stop-words\textsuperscript{52}
- Use terms that have more than two characters and less than 50
- Ensure the terms appear in ten or more documents

Figure 3.10: Example headlines that are converted to lowercase and lemmatized.

The extraction and data processing procedures are performed with the R statistical software. It creates the final dataset that contains news headlines, a date of publication, a source of publication, publication type, and region associated with the article content. The lemmatized terms from the news

\textsuperscript{51} Lemmatization uses the TreeTagger engine with English dictionary from the R koRpus package.

\textsuperscript{52} Stop-word dictionary includes the following custom words: "p8", "02", "03", "09", "20", "200", "201", "201to", "24", "26", "27", "28", "2nd", "30", "50", "500", "60", "NA", "r1403003"
headlines (i.e. the left-most column in Figure 3.10) serve as the topic model inputs in the form of word vectors corresponding to each article. Further summary statistics of this sub-corpus can be found in section 3 of Chapter 5.

3.6.2 Climate And Green Bond Sub-Corpus

While the Climate and Green Bond sub-corpus is similar to the Climate Finance sub-corpus, it differs in a few notable ways. For one, there are fewer news and blog articles in this sub-corpus. Using the consolidated keyphrases, the 2,878 articles (1,607 news and 1,271 blog) identified are roughly 35 percent of the climate finance volume. Another differentiator is the inclusion of article contents that increases the total vocabulary. Lastly, and most pertinent to the post-processing steps, the HTML format of the articles extracted from LexisNexis requires cleaning. All the processing for this sub-corpus was performed using Python.

The concept of green and climate bonds is relatively new and emerging around the time of increased coverage associated with the Paris Agreement. Figure 3.11 shows this rising trend in reporting for both news and blog data after 2010.

![Graph showing Green & Climate Bond Article Counts (News & Blog)](image)

**Figure 3.11:** Green and climate bond articles (both news and blog) from 2010 onward where a data could be assigned. The majority of articles (2780 of the 2878) occur after 2010.

When the news and blog articles are broken out by type, as shown in Figure 3.12, both domains show a similar upward trend beginning around 2014.
As previously noted, extracting the dates, headlines, and article content from HTML requires identifying relevant strings that precede the desired content. While the HTML strings are valuable for identifying content, they are not relevant to the article meaning and are added to the stop-word dictionary (e.g., <DIV CLASS="c4">). Figure 3.13 illustrates an example of raw HTML extracted from the corpus (which appears in the top portion of the figure) and the cleaned version of the text.

Figure 3.12: Green and climate bond articles by source. Left panel is from news sources and the right panel is from blog sources.

Following the HTML processing, the news and blog articles relating to climate and green bonds are combined into one corpus with a date (extracted from the HTML) attached. Figure 3.14 illustrates this processing step, with the article contents in the first column, date in the second, comparisons of different stemmers such as porter and snowball and a lemmatized version (all tokenized).
Like with the climate finance corpus, processing of this corpus uses lemmatized terms and proceeds with the same bi and tri-gram concatenations, stop-word and punctuation removals, tokenizing, and vectorizing. This is processed dataset used in the anchoring topic modeling work to identify opportunities in the green bond market as discussed in Chapter 6.

This chapter explains the text analytics employed for this research. A machine learning approach known as topic modeling is leveraged to examine both the climate finance and green bond landscapes. The concept of topic modeling, the broad applications, and the historical grounding of topic modeling are introduced in section 4.1. What follows is a moderately technical overview and comparison of unsupervised topic modeling—the type of topic modeling applied in Chapter 5 on the climate finance corpus. Chapter 5 presents the findings from a generative statistical topic model called Latent Dirichlet Allocation (LDA); LDA and the precursors to LDA are described in Section 4.2. This section is followed by a discussion of model evaluation and selection, including the selection of hyperparameters for the LDA model and the evaluation criteria used against other topic models. The chapter proceeds to a discussion of semi-supervised topic modeling, where topics are seeded with words of interest (Section 4.4). A semi-supervised approach is the foundation of the research presented in Chapter 6 on green bonds. Unlike the generative process assumed by LDA, the semi-supervised topic model uses an information theoretic approach to topic modeling and incorporates human input via anchor words - the model is called CorEx short for correlation explanation. The discussion of the CorEx model is followed by a description of the model selection and the evaluation process (Section 4.5). The chapter closes with a brief discussion of sentiment, which is used to classify the polarity of the text documents within topics. Figure 4.1 displays the analysis process (corresponding to each corpus) and where the methods described in this section apply.
4.1 Topic Modeling Introduction

The rapidly growing pool of internet data—projected to grow to 163 zettabytes of data by 2025, according to industry research firms Gartner and The International Data Corporation—is approximately 80 percent unstructured, and much of this consisting of text (Rizkallah 2017; Reinsel, Gantz, and Rydning 2018), challenging researchers’ ability to extract insights and information compared to its structured counterparts. Examples of unstructured data can be found in nearly every industry and include data such as log-files, e-mails, message boards, tweets, blogs, patents, and news articles. The increasingly inexpensive data storage costs is a key contributor to the proliferation and accessibility of unstructured data. A strong appetite exists to provide insightful information from these unstructured data sources. Research across computation-related disciplines—from statistical learning, machine learning, information extraction and retrieval, and natural language processing—has proposed methods, algorithms, and techniques to uncover structure and derive potentially useful insights from unstructured data. Many of these approaches provide automatic summarization of the content in a structured manner, facilitating insights to the unstructured data.

Topic modeling is one such approach frequently applied to help summarize text-based data and identify relevant concepts (Gallinucci, Golfarelli, and Rizzi 2015; Hisano et al. 2013; Yang, Torget, and Mihalcea 2011). When analyzing unstructured textual data, topic modeling provides a convenient approach to understand clusters of related terms and documents. Topic
models are based on the assumption that each document contains a mixture of topics and that each topic is characterized by a collection of words. For instance, a topic model might assign a James Bond novel the topics of “espionage”, “romance”, and “adventure” judging by the collection of words that appear in the novel, and each topic can be assigned with varying levels of relevance. In this way, topic modeling serves as a method to provide structure at the aggregate corpus-level. Because the topics are not directly observed and only uncovered by the modeling, they are considered latent or hidden variables. These latent variables provide conceptual understanding of the corpus and are called the topics.

Topic models and all the variations focus on identifying patterns in the data using words, terms, or concepts that tend to occur together. Many of the topic modeling approaches, including those employed in this research, have origins in vector space models (VSMs)—where each document is mathematically represented as a vector of identifiers, such as words (Salton 1989). Inherent in the VSM is the “bag-of-words” (BoW) assumption, which assumes the document is simply a collection of terms or words whose structure and order are ignored. By ignoring the syntactic structure of the text, order can be altered without impacting the results of the analysis. In this research, certain common but important ordered terms become concatenated (e.g., “climate finance” is “climatefinance”). In this case, the concatenated term “climatefinance” preserves the meaning of this concept captured through the sequence of these two words. Without the concatenation, there would be no distinction between “climate finance” in the following two sentences: (1) “The volatile market in finance is creating a frenzied climate” and (2) “The World Bank is tackling climate finance with green bonds”.

Most topic modeling is an unsupervised learning approach where the algorithm does not rely on any categorization of the data by humans to arrive at the suggested clusters (i.e. topics). Like partitioning clustering approaches (MacQueen 1967), topic modeling requires the specification of the number of topics by the algorithm designer. As no human intervention guides the algorithm’s effort at clustering, the identified topics typically eschew simple classification, and often require subject matter expertise to identify a unifying theme or concept to the topic. In most real-world cases, the appropriate number of topics is unknown. In this research, a strategy to automatically obtain a reasonable estimate of the number of topics is implemented. As a simple example, while a person might cluster stories from a given day’s newspaper into the sections like news, sports, and business, topic modeling might instead cluster articles with themes of attack, trade, and negotiate, based on the types of terms encountered in stories.

---

53 Document refers to any single unit of textual analysis like a news headline or tweet
54 A closely related term is exchangeability in the generative language model
55 Note supervised and semi-supervised variations do exist, but are less commonly used in practice.
In the realm of topic modeling, methodological approaches fall into a few general areas: static topic modeling, topic evolution modeling, and guided topic modeling. The former covers methods that include latent semantic indexing and analysis (LSI and LSA) (Deerwester et al. 1990; Landauer, Foltz, and Laham 1998), probabilistic latent semantic analysis (pLSA) (Hofmann 1999), latent Dirichlet allocation (LDA), hierarchical Dirichlet process (HDP) (Chong Wang et al. 2011), and correlated topic models (CTM) (Blei and Lafferty 2007). While these are methods foundational to static topic modeling, they are not exhaustive and there are many variations of each. The second area incorporates a temporal component into topic modeling. LDA serves as a foundation for most of these models, which include dynamic topic models (DTMs) (Blei and Lafferty 2006a; C. Wang, Blei, and Heckerman 2012), topics over time (ToT) models (X. Wang and McCallum 2006) , multiscale topic tomography\textsuperscript{56} models (MTTM) (R. M. Nallapati et al. 2007; Iwata et al. 2010), and many other variations. These temporal models operate on the premise that time is an important confounding variable; the models assume that as time progresses, the collection of documents (the corpus) and their content will also evolve, and thus time is an important feature to capture\textsuperscript{57}. The final set of models include both semi-supervised and supervised approaches. In the semi-supervised case user selected terms are used to “seed” or “guide” the topics to converge around the desired terms - examples include the anchored correlation explanation (CorEx) (Gallagher et al. 2016; Reing et al. 2016; Ver Steeg and Galstyan 2014) and the guided LDA approach (Jagarlamudi, Daumé, and Udupa 2012; Vikash 2017). In a supervised setting, like supervised LDA (sLDA) (Mccauliffe and Blei 2008; T. Nguyen et al. 2015; Ramage et al. 2009), a response variable is associated with each document and often does not influence the topics discovered as with the semi-supervised approaches.

4.2 Static Topic Models—Comparisons And Technical Details

For roughly the past 40 years, there has been a major advancement in topic modeling about every decade, with each advancement overcoming a previous limitation. The most widely used topic modeling technique today is the generative model known as latent Dirichlet allocation (Blei, Ng, and Jordan 2003). As previously mentioned, precursors to this work include latent semantic indexing or analysis (Deerwester et al. 1990; Steyvers and Griffiths 2007) and probabilistic latent semantic analysis (Fuhr 1992; Girolami and Kabán 2003; Hofmann 1999, 2017; Ding 2005). Many other topic models, like the Correlated Topic Model (Blei and Lafferty 2007), are extensions of LDA and grounded in similar principles with relaxed assumptions (Blei

\textsuperscript{56} According to the author’s talk tomography comes from the Greek words "tomos" (to cut or section) and "graphein" (to write) (R. Nallapati 2007). Unlike LDA that looks at topic distribution in documents (normalized by document), topic tomography models look at how topics are distributed across documents (normalized by topic) (R. Nallapati 2007).

\textsuperscript{57} Future research (outlined in Chapter 7) should consider using evolutionary or dynamic topic models.
between these models (common programmatic implementations can be found in Appendix I). Each of these models are summarized in more detail in subsequent sections of this chapter, and the table below from Lee et al. (2010) highlights some differentiators between these models (common programmatic implementations can be found in Appendix I).

<table>
<thead>
<tr>
<th>Model</th>
<th>Characteristics</th>
<th>Limitations</th>
<th>Conditions for High Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>[LSA] Latent Semantic Analysis</td>
<td>● Reduces dimensionality using Singular Value Decomposition</td>
<td>● Difficult to determine the number of topics</td>
<td>● Documents have to contain high-order co-occurrences and transitive relations</td>
</tr>
<tr>
<td></td>
<td>● Captures synonyms of words</td>
<td>● Difficult to interpret loading values with probability meaning</td>
<td>● Exploits unique structure as factors (Kontostathis and Pottenger 2006)</td>
</tr>
<tr>
<td></td>
<td>● Weak statistical background</td>
<td>● Difficult to label a topic using words in the topic</td>
<td>● Can get good performance when documents represent a single topic, have peculiar terms in one topic, or are void of style modifiers (Papadimitriou et al. 2000)</td>
</tr>
<tr>
<td>[pLSA] Probabilistic Latent Semantic Analysis</td>
<td>● Mixture components are multinomial random variables that can be viewed as representations of “topics”</td>
<td>● No probabilistic model at the level of documents</td>
<td>● Beside high-order co-occurrence and transitive relations, generative models require conditional independence. Can be checked with correlation analysis or separation causal map (Daphne Koller and Friedman 2009)</td>
</tr>
<tr>
<td></td>
<td>● Each word is generated from a single topic; different words in a document may be generated from different topics</td>
<td></td>
<td>● If a document is long, many words in the document probably have relation within the document but also with other documents, which reduces high-order co-occurrence. Lengthy documents are not appropriate</td>
</tr>
<tr>
<td></td>
<td>● Partially handles polysemy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[LDA] Latent Dirichlet Allocation</td>
<td>● Provides full generative model with multinomial distribution over topics</td>
<td>● Incapable of model relations among topics</td>
<td>● Theoretically, LDA can handle mixed-length documents. However, because of exchangeability, words in a topic can be collected from any part of a document, and the topic meaning will be ambiguous.</td>
</tr>
<tr>
<td></td>
<td>● Handles long-length topics</td>
<td></td>
<td>● To enrich transitive relations and high-order co-occurrence, in some cases, LDA is applied to many small documents</td>
</tr>
<tr>
<td></td>
<td>● Shows adjectives and nouns in topics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[CTM] Correlated Topic Model</td>
<td>● Considers relations among topics using logistic normal distribution</td>
<td>● Requires complex computations - may not be able to handle a large corpus</td>
<td>● Topic relations can be extracted from many co-occurrence relations. Thus, documents in online discussion forums or question and answer formats might produce meaningful relations among topics.</td>
</tr>
<tr>
<td></td>
<td>● Allows the occurrences of words in other topics</td>
<td>● Contains too general words in topics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>● Allows topic graphs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: The strengths, limitations, and computational requirements of several common topic models as summarized by S. Lee, Song, and Kim (2010). Some language is modified for clarity.
4.2.1 Latent Semantic Analysis

Latent Semantic Analysis (LSA) evolved from the term Latent Semantic Indexing (LSI), which was originally introduced around the late 1980s as a method for use in information retrieval (Deerwester et al. 1990); both terms are roughly interchangeable. The goal of LSA is to find similarities between text using vector representations, akin to document clustering (Dumais 2005). Given a corpus of \( m \) documents, each distinct word\(^{58} \) within those documents would be assigned to \( n \) rows, and each element in the \( n \times m \) matrix is then weighted in a manner proportional to the number of times each word appears in each corresponding document. This results in what is called a term-document matrix, \( A \in \mathbb{R}^{n \times m} \). With this matrix, similarity between text can either be determined across all documents or words; the normalized dot product based on the magnitude of the row vectors is the word similarity, while the dot product of the column vectors will represent the document similarity.

The vector representation creates a lower-dimensional space where similar documents appear close to each other, as shown in Figure 4.2.

![Heatmatrix where each column corresponds to a document id and each row to a word. A cell stores the weighting of a word in a document by using term-frequencies or term-frequency inverse document frequencies. The dark cells indicate higher weights, suggesting the word may serve as a possible topic marker. The low dimensional representation groups similar documents together based upon the term weights. For example, the documents with high word counts related to environmental quality are in the top left corner (air, pollution, power, and environmental terms make the topic) and documents in the first four rows best represent this topic; those with high word counts for national security are in the bottom right. The resulting patterns are used to detect the latent components representing topics (Wikipedia contributors 2018). Source: Wikipedia](image)

---

\(^{58}\) Used interchangeably with “term” and can include any term or word the researcher desires. Common practice removes articles, conjunctions, and disjunctions and treats these as “stop-words”.

75
LSA uses any variation of the term-frequency (TF) or a term-frequency inverse document frequency (TF-IDF) measurement for weighting the vectors and then applies Singular Value Decomposition (SVD) (Golub and Kahan 1965) to the matrix to get a low-dimension representation. Often the term-document matrix is sparse, noisy, and high-dimensional so the dimensions of $A$ are reduced and approximated using SVD or truncated SVD (Hansen 1987).

With truncated SVD, $A$ is approximated by $A \approx U_kS_kV_k^T$ where $U$ and $V$ have orthonormal columns and $S$ is a non-negative diagonal — $k$ is a parameter specifying the number of topics — only the $k$ largest singular values are selected along with the $k$ columns of $U$ and $V$. The columns of $U$ and $V$ associate the latent variables to the documents and words. Specifically, the document-topic matrix is $U \in R^{m \times k}$ and the term-topic matrix is $V \in R^{n \times k}$.

LSA is computationally quick and efficient to implement and is often applied to cross-language information retrieval and text classification or summarization. Although storing the full-rank term-document matrix in memory can be a computational hindrance, recent low-memory implementations are now available. If the original full-rank term-document matrix is too large to store, is noisy, or is overly sparse, then the low-rank matrix serves as an approximation or de-noised matrix of higher quality. In addition to the difficulty interpreting the output (embeddings), another drawback of the methodology is that it will group documents together that might have similar words but not necessarily co-occurring words, which may lead to inaccurate results. To overcome this, a very large corpus and vocabulary set is needed.

### 4.2.2 Probabilistic Latent Semantic Analysis And Indexing

To overcome the lack of statistical foundation in LSA, probabilistic models for LSA and information retrieval were introduced roughly a decade later in the late 1990’s (Ding 2005; Fuhr 1992). The main model for pLSA was introduced around the year 2000 (Hofmann 1999, 2001). Unlike LSA that uses SVD, pLSA uses a method that looks to find a probabilistic model that contains latent topics that would be likely to generate the data. In this case, the data are those contained in a document-term matrix (TF or TF-IDF). Unlike the LSA approach, pLSA defines a proper generative model for the data that disambiguates words with multiple meanings (polysemy) and groups words that are likely to share context together. The results of the probabilistic approach (pLSA) are generally superior to those of LSA, and the clusters of topics generated by pLSA tend to be more interpretable as well.

---

59 Truncated SVD is not an exact decomposition of $A$ because only the largest singular values from $S$ are calculated, this is $k$, and only the $k$ corresponding column vectors of $U$ and the $k$ row vectors of $V^T$ are retained. In SVD, $U$ will be $m \times m$, $V$ will be $n \times n$, and $S$ will be $m \times n$. Whereas in truncated SVD, $U_k$ is $m \times k$, $V_k$ is $k \times n$, and $S_k$ is $k \times k$. 
pLSA is based on a statistical model known as an aspect model. An aspect model is a latent variable model for co-occurrence data, which associates unobserved class variables—in this case, the likelihood of occurrence of a word—with each observation (Tuomo Kakkonen et al. 2008). LSA assumes words and documents form a joint Gaussian distribution, when the empirical evidence shows a Poisson distribution is actually observed and so pLSA does not assume a Gaussian distribution.

In this case, the goal is to model the probability that for any document, \(d\), and word, \(w\), each document is a probability distribution over topics, \(c\), where the topics are classes. Given a document, \(P(c \mid d)\) is the probability that a topic is contained in the document. Similarly, \(P(w \mid c)\) is the probability that a word is drawn from a topic. pLSA is depicted by the graphical model below in Figure 4.3. In this and all graphical models the “nodes denote random variables; edges denote dependence between random variables. Shaded nodes denote observed random variables; unshaded nodes denote hidden random variables. The rectangular boxes are plate notation, which denote replication” (Blei and Lafferty 2009).

![Graphical representation of pLSA. P(d) is determined from the corpus. P(c | d) and P(w | c) are modeled as multinomial distributions and trained using the expectation-maximization (EM) algorithm.](image)

pLSA adds the probabilistic treatment to the topics and words in the following process to estimate the probability that a word is contained within a document:

1. Determine the probability of a document:
   Select a document \(d\) with probability \(P(d)\)

2. Determine distribution of topics by document:
   Pick a latent topic (class) \(c\) with probability \(P(c \mid d)\)
   Generate a word \(w\) based on the topic with probability \(P(w \mid c)\)

3. Determine how likely the word is within a document:
   Estimate \(P(w \mid d)\)
While pLSA is more flexible than LSA, it too has drawbacks. Like LSA it suffers from scalability; as the number of documents increases, so do the parameters needed to estimate the model. The parameters for $P(c \mid d)$ grow linearly along with the number of documents and can lead to overfitting as well as the generation of too many topics. The lack of parameters for $P(d)$ is another issue when assigning a probability to a new document is unknown. Lastly, pLSA is often not used alone—LDA is an extension of pLSA and tends to perform better.

4.2.3 Latent Dirichlet Allocation, LDA

Like pLSA, LDA is a generative probabilistic model that attempts to identify salient and related topics within text or text collections; to date, it is one of the most successful topic models (Blei, Ng, and Jordan 2003; Steyvers and Griffiths 2007). LDA stems from the seminal work of LSA and is the Bayesian extension of pLSA, serving as the textual analog to principal components analysis (PCA) for textual data (Buntine and Jakulin 2004). LDA does not restrict latent textual discovery to a single topic or a set of correlated topics. Instead LDA treats a topic as a distribution over a fixed vocabulary (of terms or words) and documents as a distribution over the topics where each document contains each topic with a different proportion. Blei describes this as “natural assumption to make because documents in a corpus tend to be heterogeneous, combining a subset of main ideas or themes that permeate the collection as a whole” (Blei and Lafferty 2009).

These “themes” mentioned by Blei and Lafferty (2009) are the topics represented by the latent (or hidden variables) that the model learns. These hidden random variables (i.e., “themes”, topics, or classes) interact with the observed data (i.e., vocabulary and documents) in order to learn the structure of the distribution using posterior probabilistic inference (Blei and Lafferty 2009). This distillation of the corpus into topics—the posterior distribution of the hidden variables given the observed documents—is what allows the corpus to be thematically (or topically) examined. Hence, the topics are modeled from the given data; within each document (e.g., a news headline, tweet, article, or forum post), the observed data are represented by the document’s terms and the hidden variables comprise the latent topic structure. Based upon the collection of documents, a posterior distribution of the hidden variables representing the topics is estimated given the observed documents and terms within the documents. The computational difficulty lies in estimating this posterior distribution.

The statistical basis of LDA is best described as a multi-level hierarchical Bayesian model where observed data interacts with hidden random variables that are drawn from a Dirichlet distribution. The Dirichlet distribution is a generalization of a multivariate beta distribution. It can be thought of as a distribution over discrete distributions where sampling
occurs over a probability simplex (Kotz, Balakrishnan, and Johnson 2004; Olkin and Rubin 1964). The topics, c, the per-document topic proportions, \( \theta \), and the per-word topic assignments, \( \phi \), make this a mixed-membership model (Erosheva, Fienberg, and Lafferty 2004) where the documents represent a mixture of topics rather than only one topic (McLachlan and Peel 2000; Blei and Lafferty 2009).

The document-topic and topic-word distributions use Dirichlet priors, \( \alpha \) and \( \beta \), respectively. A larger value of \( \alpha \), the document-topic density, implies the documents are more strongly comprised of multiple topics (the probabilities are more evenly distributed), whereas a smaller value of \( \alpha \) implies a document will be identified with a more specific topic or at higher probabilities. For example, when the prior is small, one topic will represent a document with a high probability while all other topics will be associated with a much lower probability. If the value is less than one it can push the values towards extremes; in a trivariate Dirichlet distribution this small value will concentrate mass in the corners. Alternatively, with a relatively higher value prior a document could be distributed over all topics equally, so if there are three topics each will occur with about equal probability, about 33 percent; in a trivariate Dirichlet distribution this large value will concentrate mass in the center. Similarly, a high \( \beta \), topic-word concentration, implies more words are associated with a topic and a lower \( \beta \) fewer words strongly represent a topic. Often a symmetric Dirichlet is used in LDA, where each parameter component is the same value, so the priors influence these topic or word distributions equally.

The probabilistic process in LDA is used to model the random process that is assumed to have generated the topic structure. Thus, all the documents in the collection, \( M \), are represented as random mixtures over latent topics, \( c \), where each topic is comprised of a distribution of words or terms, \( w \), over all terms that comprise the entire vocabulary, \( N \). The per-topic word proportions and the per-document topic proportions, \( \phi \) (distributions of the vocabulary over topics) and \( \theta \) (distributions over topics) respectively, are random Dirichlet variables. For a particular document a random sample is drawn from the topic Dirichlet distribution, \( \text{Dir}(\alpha) \). From this topic distribution, \( \theta \), a topic, \( c \), is selected based upon the distribution. Then for the words, the word distribution, \( \phi \), of topic \( c \) is obtained by taking a random sample from the word Dirichlet distribution, \( \text{Dir}(\beta) \), and then from \( \phi \) a word, \( w \), is chosen. The process is depicted with a directed graph in Figure 4.4 and the generative process follows:
Figure 4.4: Graphical representation of LDA. $K$ is the number of topics; $N$ is the number of words or terms in the document; $M$ is the number of documents to analyze; $\alpha$ is the Dirichlet prior concentration parameter of the per-document topic distribution; $\beta$ is the same parameter of the per-topic word distribution. $\phi$ or $\phi(k)$ is the word distribution for topic $k$ and is a Dirichlet distribution; $\theta$ or $\theta(i)$ is the topic distribution for document $i$ and is a Dirichlet distribution; $c$ or $c(i,j)$ is the topic assignment for $w$ or $w(i,j)$ and is multinomial; $w$ or $w(i,j)$ is the $j$-th word in the $i$-th document and is multinomial.

The LDA generative process for each word $w_j$ (from a vocabulary of size $N$) in document $d_i$ follows the following steps (Blei, Ng, and Jordan 2003; Hornik and Grün 2011):

1. Select, $K$, topics for the model, and the topic, $\beta$, and proportions, $\alpha$, parameters
2. Determine the term distribution for each topic:
   - Draw $\phi_k \sim Dir(\beta)$, where $k = 1, \ldots, K$; $\phi_k \in \Delta_V$
   - $\phi_{k,w} =$ probability of word $w \in \{1, \ldots, N\}$ in topic $k \in \{1, \ldots, K\}$
3. For each document determine the topic distribution:
   - Draw $\theta_i \sim Dir(\alpha)$, where $i = 1, \ldots, M$; $\theta_i \in \Delta_K$
   - $\theta_{i,k} =$ probability that document $i \in \{1, \ldots, M\}$ has topic $k \in \{1, \ldots, K\}$
4. For each word of the $V$ words in the entire vocabulary:
   - Draw a topic $c_{i,j} \sim Mult(\theta_i)$, where $c_{i,j} \in \{1, \ldots, K\}$
   - Draw a word conditioned on topic $w_{i,j} \sim Mult(\phi_{c_{i,j}})$, where $w_{i,j} \in \{1, \ldots, N\}$

As previously mentioned, approximating this posterior distribution is difficult yet crucial to using LDA for topic-trend exploration, document similarity, or prediction. The observed and hidden variables are contained in the LDA joint distribution while the topic structure is described by the posterior distribution in equation 1 (Blei, Ng, and Jordan 2003; Blei and Lafferty 2009):

$$p(\theta_{1:M}, c_{1:M,1:N}, \phi_{1:K} \mid w_{1:M,1:N}, \alpha, \beta) = \frac{p(\theta, c, \phi \mid w, \alpha, \beta)}{\int_{\phi} \int_{\theta} \sum_{c} p(\theta, c, \phi \mid w, \alpha, \beta)}$$ (1)
Provided $M$ documents where the documents have $N$ words in the vocabulary and each of the $w$ words are generated from a set of $K$ topics provides the joint posterior probability of $\theta$ (distribution of topics, one for each document), $c$ (topics for each document), $\phi$ (distribution of words, one for each topic) given all the words from the corpus (all documents) and the concentration parameters, $\alpha$ and $\beta$.

Several different techniques—each with their trade-offs in speed, complexity, and accuracy—have been developed to solve this approximation problem for the intractable posterior (with integrals in the denominator shown above in Equation 1). Most techniques try to closely match the true posterior with a known probability distribution, some of these include: mean field variational inference (Blei, Ng, and Jordan 2003), expectation propagation (Minka and Lafferty 2002), variational expectation maximization (Hornik and Grün 2011), and Gibbs sampling (Darling 2011; Porteous et al. 2008; Steyvers and Griffiths 2007). The two implementations considered in Appendix J of this research are Gibbs sampling and variational expectation maximization (VEM). Gibbs sampling, a Markov Chain Monte Carlo technique, is used for all results presented in this chapter (Hornik and Grün 2011).

4.2.4 Correlated Topic Model

The correlated topic model (CTM) builds on the foundations of LDA but allows for correlation to exist within the topic proportions $^{60}$ (Lafferty and Blei 2006; Blei and Lafferty 2007). Like LDA, documents modeled using CTM will contain multiple topics with varying proportions, however, CTM leverages correlation in the terms and latent variables (i.e., themes or topics). The process for CTM is very similar to the LDA process outlined above with the notable exception that the topic proportions, $\theta$, (from step 2 in LDA above) are not drawn from a Dirichlet distribution (Blei and Lafferty 2007). Here $\theta$ is the $\log(\theta_i/\theta_K)$ and mapped to the topic proportions as follows: $f(\theta) = \frac{exp(\theta)}{\sum_i exp(\theta_i)}$ (Lafferty and Blei 2006; Blei and Lafferty 2007). The mean and covariance, $\{\mu, \Sigma\}$, are $K$-dimensional vectors and a $K \times K$ matrix, respectively. The topics are $K$ multinomials of the vocabulary $V$. The correlated topic model assumes that an $N$-word document arises from the following generative process that creates the document:

(1 CTM) For each document, $m \in \{1, \ldots, M\}$ determine the topic distribution:

$$\text{Draw } \theta_m \mid \{\mu, \Sigma\} \sim N(\mu, \Sigma)$$

(2 CTM) For each word in the vocabulary:

- Draw topic assignment $c_{m,n} \mid \theta_m$ from Mult($f(\theta_m)$)
- Draw word $w_{m,n} \mid \{c_{m,n}, \beta_{1:k}\}$ from Mult($\beta_{c_{m,n}}$)

$^{60}$ This is via the logistic normal distribution. Refer to Blei and Lafferty’s work for additional detail.
In an application of the model to science articles, Blei and Lafferty (Blei and Lafferty 2007) describe the dependencies that allow for correlation between the topics as shown in Figure 4.5: “The CTM draws a real valued random vector from a multivariate Gaussian and then maps it to the simplex to obtain a multinomial parameter. The covariance of the Gaussian induces dependencies between the components of the transformed random simplicial vector, allowing for a general pattern of variability between its components... [the logistic normal] provides a more expressive document model” (Blei and Lafferty 2007)\textsuperscript{61}.

4.3 LDA Topic Model Evaluation And Hyperparameter Selection

As with other machine learning problems hyperparameter choice influences the results. In topic modeling, the way in which the number of topics and concentration parameters are chosen is a modeling choice. For both the LDA topic model and its variations (like CTM), a common approach is to fit the model to the document-term matrix using various values of parameters set in advance (hyperparameters)\textsuperscript{62} and then determine the “best” resultant fit; Determining the “best” requires selecting a metric or criteria to use for the evaluation. Unlike other model parameters, hyperparameters are not learned in the machine learning process, instead they are chosen apriori. This section will describe the LDA hyperparameters and then discuss how they are selected in this research—referred to as hyperparameter optimization. The hyperparameter optimization process offers a framework that uses two metrics for an in-sample evaluation and then applies an out-of-sample evaluation along with a comparison as a confirmatory check.

These selection choices of the hyperparameters may differ by application—some applications will require topics that make sense to the human and are coherent, others may require more granular topics containing very homogenous documents, and sometimes coarser,

\textsuperscript{61} For a visualization see the example from Science in Blei and Lafferty (2007).
\textsuperscript{62} The hyperparameters are parameters that require a value to be set in advance of the learning process with model training. The hyperparameters differ from learned model parameters learned from the data and training process - the hyperparameters will influence the model’s learned parameters.
generic topics (yet fewer in number) will suffice. The selection of the priors (the topic and word concentration parameters—$\alpha$ and $\beta$), discussed in the LDA section, requires consideration of the application as well—the belief that few words are associated with a topic or few topics are contained in a document will influence these values. When using Gibbs sampling to estimate the LDA or CTM models, these values need to be specified in advance like the topic number.

Ongoing and active discussion exists around the hyperparameter selection process for the LDA topic model. Many convenient approaches were developed with varying considerations in mind (Arun et al. 2010; Cao et al. 2009; Deveaud, SanJuan, and Bellot 2014; Griffiths and Steyvers 2004; Ponweiser 2012; Zhao et al. 2015). Most of these techniques select the hyperparameters using data-driven approaches that fit several models with different parameter values. Nikita (2016) implements four of these approaches to help choose the optimal topic number, $K$, of in a given collection: (1) density selection based upon topic distances (Cao et al. 2009) (2 and 3) maximizing or minimizing Kullback-Leibler (KL) divergence (Deveaud, SanJuan, and Bellot 2014) (Arun et al. 2010) and (4) Bayesian model based expectation maximization method (Griffiths and Steyvers 2004). Each of these make use of the entire sample to make the parameter selection - this is referred to as in-sample evaluation. In practice, usually only one of these metrics is used and it is often used to determine only the number of topics, not the concentration parameters.

Instead of only using one metric to choose the optimal number of topics, this research makes use of two metrics with different evaluation approaches (Cao et al. 2009; Deveaud, SanJuan, and Bellot 2014). This research also uses these metrics to not only choose the number of topics but to select the density parameters as well. By using two metrics, the selection of the parameters ($K$, $\alpha$, and $\beta$) is more robust to the deficiencies or biases exhibited by a single metric. For this application, the research demonstrates that using two of the metrics in conjunction, yields rather consistent results for the selected hyperparameters.

To select the optimal combination of hyperparameters, the full-factorial combination of the three hyperparameters are evaluated: the number of topics ($K$), the Dirichlet prior for topic concentration ($\alpha$), and the Dirichlet prior for word concentration ($\beta$). For each combination of the hyperparameters the density-based (Cao et al. 2009) and the KL divergence (Deveaud, SanJuan, and Bellot 2014) metrics are computed. Metrics implemented by Nikita (2016), including Cao’s (Cao et al. 2009) (referred to as “CaoJuan” or “CaoJuan2009” in the analysis) and Deveaud’s (Deveaud, SanJuan, and Bellot 2014) (referred to as “Deveaud” or “Deveaud2014”) are used in this research. The entire dataset is used to build (train) the LDA model for the specified parameters and then evaluate it against the metric (this is done for each set of the parameters in the full-factorial combination). Then the absolute and rank-ordered metric values for each parameter combination are compared. The highest rank-ordered
combination is selected and the values for $\alpha$ and $\beta$ are used; from the top rank-ordered combinations a reduced candidate set of topic numbers, $K$, are obtained and then compared in a cross-validation setting.

In this research the hyperparameter selection above occurs over the entire sample, representing the in-sample evaluation. The alternative approach to an in-sample evaluation is to train on a portion of the data and then evaluate on a separate sample that the model has not seen before (i.e., out-of-sample and cross validation). In addition to this in-sample evaluation, the topic number, $K$, is also validated with an out-of-sample performance metric (i.e., perplexity using a training and testing set) described later in this section. The out-of-sample approach with cross-validation is used as a complement to the in-sample approach and to confirm the parameters chosen in both approaches are consistent.

Both the in-sample and out-of-sample evaluation processes compare the results of the model using hyperparameter optimization with suggested or default values for the concentration parameters ($\alpha$ and $\beta$ priors). The suggested default values, as described by Griffiths and Steyvers (2004), are $50/K$ for $\alpha$, where $K$ is the number of topics, and 0.1 for $\beta$. For a specific $K$, the topic number, the in-sample and out-of-sample performance metrics are compared for the models using these default priors to those models where the models are optimized. An example and further elaboration can be found within Chapter 5.

In information theory and text mining perplexity is a commonly used evaluation metric to perform an out-of-sample evaluation of the model. Very simply, perplexity measures how well a probability model predicts a sample - the model is constructed using the training data set and then the perplexity of the held-out sample (i.e., testing data set) is evaluated. The perplexity measure is equivalent to the geometric mean per-word likelihood (Hornik and Grün 2011) and is related to cross-entropy (P. F. Brown et al. 1992). Like entropy, a lower perplexity indicates a more desirable model with less uncertainty in the unseen sample (words or documents). Whereas when evaluating the likelihood of the unseen observations higher likelihoods imply better models—for further discussion and more specific details refer to (Hornik and Grün 2011; Wallach et al. 2009) (2011) and Wallach et al. (2009).

The specific formulas for perplexity, $Perplexity(w)$, and likelihood with Gibbs Sampling, $\log(p(w))$, used in this implementation are the same as those used by Hornik and Grün (2011) in their R package implementation. The perplexity is described by the following equation, where $w$ is a term contained in the entire vocabulary of $V$ with $w \in \{1, \ldots, V\}$ across all documents $d$ in the corpus $M$ where $d \in \{1, \ldots, M\}$ and $n^{(jd)}$ is the $j$-th term within the $d$-th document (Hornik and Grün 2011):
\[
\text{Perplexity}(w) = \exp \left[ -\frac{\sum_{d=1}^{M} \sum_{j=1}^{V} n^{(jd)}}{\sum_{d=1}^{M} \sum_{j=1}^{V} n^{(jd)}} \right] 
\]

Using Gibbs sampling and the perplexity equation above the likelihood is determined by the following (Hornik and Grün 2011; Newman et al. 2009):

\[
\log(p(w)) = \sum_{d=1}^{M} \sum_{j=1}^{V} n^{(jd)} \log \left( \sum_{k=1}^{K} \theta_k^{(d)} \beta_k^{(j)} \right) 
\]

For every observation in the testing-set, half the words at random are designated for use (i.e., fold-in), and the remaining words are used as a test set. The document-topic mixture, \( \theta \), is learned using the fold-in part, and log probability of the test words is computed using this mixture, ensuring that the test words are not used in estimation of model parameters. The LDA model perplexity is computed using the standard practice (V. A. Nguyen, Boyd-Graber, and Resnik 2014) of averaging over multiple samples within the logarithm (Griffiths and Steyvers 2004; Hornik and Grün 2011). The perplexity metrics are computed and compared for the optimized and base hyperparameters to ensure the hyperparameters demonstrate superior performance compared to the base parameters.

Research has shown that perplexity can be unstable in topic modeling (Zhao et al. 2015) or inconsistently produces human-interpretable results (Chang et al. 2009). Because of these concerns, this research applies the perplexity metric in conjunction with the two in-sample evaluation metrics “CauJuan” (Cao et al. 2009) and “Deveaud” (Deveaud, SanJuan, and Bellot 2014). Perplexity is applied in this research as described and implemented by (Hornik and Grün 2011) and (Newman et al. 2009) with five-fold cross validation and fitting the LDA with Gibbs sampling. With several draws, as in cross-validation, the mean is taken over the samples from within the logarithm (Hornik and Grün 2011). This combined approach is fully automated and serves as a contribution for topic model selection, especially when sensitivity exits to hyperparameters like with LDA.

### 4.4 Semi-Supervised Topic Modeling Approach

The unsupervised topic modeling approaches mentioned in previous sections—LSA, pLSA, LDA, and CTM—are ideally suited for situations when dominant topics describing the text collection are desired. All these methods, combined with hyperparameter optimization, are scalable and can quickly identify high-level topics that represent dominant term and document relationships. However, in situations when less dominant or niche topics are desired these

\[^{63}\text{In LDA this is achieved by maximizing the probability of the observed data.}\]
approaches are less suitable. In essence, the performance of the rare topics is sacrificed for global performance in modeling the most probable term and document relationships. In LDA (and variations), the topics are inferred by making assumptions about how the documents were written—the probabilistic generative model—which requires making assumptions (the hyperparameters) about the data generating process (documents in the corpus). In the information-theoretic framing of CorEx, these assumptions are not required.

In situations when the topic structure should be informed by existing domain or human knowledge, supervised or semi-supervised approaches can overcome some of the short-comings inherent in the unsupervised approaches. Many of the semi-supervised approaches bring a human into the learning loop. This is done by leveraging human domain knowledge such that the quality of the desired topics (i.e., the modeling output) can be improved. Despite the ability of supervised models to build highly accurate and precise classification models, the document labels are often unknown or are too costly to obtain (Lacoste-Julien, Sha, and Jordan 2009; Mcauliffe and Blei 2008; Ramage et al. 2009). This is where semi-supervised approaches demonstrate strength; semi-supervised approaches offer a balance between the need to understand content at scale and the need to leverage knowledge of the data and domain64. A semi-supervised approach can infuse the models with domain expertise by using human input to guide the models to form topics around desired concepts (Andrzejewski, Zhu, and Craven 2009; Andrzejewski and Zhu 2009; Gallagher et al. 2016; Y. Hu et al. 2014; Jagarlamudi, Daumé, and Udupa 2012). The desired concepts are known by the human researcher and identified through terms representing the concept, where the researcher is responsible for identifying the terms.

For context the supervised topic modeling came into the picture about three years after LDA and about three years later this was followed with semi-supervised approaches that “seed” topics. The semi-supervised work has continued with expansions beyond LDA. Andrzejewski and Zhu (2009) made use of words that pertain to specific topics and that are restricted to appearing in some subset of all the possible topics. A second model by Andrzejewski et al. (2009) uses relationships between words to partition muddled topics. Jagarlamudi et al. (2012) proposed SeededLDA, a model that seeds words into given topics and guides, but does not force, these topics towards the seeds. SeededLDA is most similar to work where anchor words are used in non-negative matrix factorization (S. Arora, Ge, and Moitra 2012; Sanjeev Arora et al. 2012). Like Jagarlamudi et al. (2012), Halpern et al. (2014; 2015) and Gallagher et al. (2016) apply domain expertise and semi-supervised topic modeling to discover desired latent variables. Additional automation was introduced to find anchor words without human knowledge, manually adjust topics by inserting words, incorporate additional latent information, and to make

---

64 These approaches are related to those that learn semantic lexicons using bootstrapping (Thelen and Riloff 2002) or prototype learning (Haghighi and Klein 2006).
the algorithms more scalable (Sanjeev Arora et al. 2013; Foulds, Kumar, and Getoor 2015; Mimno and Lee 2014; Thang Nguyen, Hu, and Boyd-Graber 2014). Jagarlamudi et al. (2012) also demonstrate how feature selection can be used to ensure the topic structure aligns with the class structure.

These topic models that allow the topics to be anchored, seeded, or guided with desired terms all fall within the semi-supervised topic modeling. The semi-supervised approach offers a balance between the ease of execution and classification. These models are most suitable for the goals of the research contained in Chapter 6 where key concepts related to green bonds are tracked. For this application, an algorithm called CorEx (Gallagher et al. 2016) is used over the SeededLDA\textsuperscript{65} approach (Y. Hu et al. 2014; Jagarlamudi, Daumé, and Udupa 2012). There are a handful of implementations in this area but the CorEx or SeededLDA implementations are easily accessible through common computing environments. Table 4.2 highlights some of the characteristics, limitations, and benefits of the anchored CorEx model (used in Chapter 6) compared to an alternative semi-supervised approach, SeededLDA.

\textsuperscript{65} SeededLDA is sometimes called GuidedLDA.
Table 4.2: Comparison of two topic models that guide the model to learn topics of specific interest. For the generative process and graphical depiction refer to Jagarlamudi et al. (2012) for SeededLDA and Gallagher et al. (2014), Ver Steeg and Galstyan (2014), and Ver Steeg (2017) for specific modeling details on CorEx.
4.4.1 Correlation Explanation, CorEx, Topic Model Details

The CorEx topic model relies upon a principle called Total Correlation Explanation (Ver Steeg 2017; Ver Steeg and Galstyan 2014) to identify topics in a corpus that explain structure through dependence found in the data. CorEx is grounded in the concept of total correlation that is described by Ver Steeg and Galstyan (2014). The interpretation differs from the more common understanding of correlation and can be better understood as dependence. The total correlation is an information-theoretic measure and allows the model to learn “maximally informed topics” (Gallagher et al. 2016). The idea of “maximally informed topics” applies information gain (IG) (Mitchel, 1997), obtained through entropy, to get the features that differentiate topic structures within a corpus. The IG of word $w$ in a topic $z$ is defined in equation 4 as:

$$\text{IG}(w, z) = H(z) - H(z | w)$$

(4)

Where $H(z)$ is the topic entropy and $H(z | w)$ is the conditional entropy of the topic, given the word. This approach is unlike LDA which requires specification of the hyperparameters that characterize the generative process. The CorEx algorithm and associated topic model versions, like the anchored version (Gallagher et al. 2016), use the concepts of entropy and mutual information to describe the dependence through total correlation without additional assumptions.

In the CorEx interpretation a total correlation (dependence) a discrete random variable, $X$, represents the terms in corpus and the entropy of $X$ is defined as $H(X)$ with mutual information of random variables, $X_1$ and $X_2$, defined in equation 5 as:

$$I(X_1 : X_2) = H(X_1) + H(X_2) - H(X_1, X_2)$$

(5)

The sub-collection of terms associated with a particular topic is defined as $X_G$ where $G \subseteq \{1, \ldots, K\}$. The probability of observing this group of terms is $p(X_G = x_G)$ and shortened to $p(x_G)$. Per Gallagher et al. (2016), the total correlation (TC) is expressed as the multivariate mutual information for the group of random variables, $X_G$ (i.e., the groups of terms associated with the topics) as defined in equation 6:

$$\text{TC}(X_G) = \sum_{i \in G} H(X_i) - H(X_G) = D_{KL} \left( p(x_G) || \prod_{i \in G} p(x_i) \right)$$

(6)

In the TC equation above, $D_{KL}$ is the Kullback-Leibler (KL) Divergence and will be zero if no dependence exists between the random variables. As noted by Gallagher et al. (2016), this does not resemble correlation as defined by the Pearson correlation coefficient or similar and is more similar to the dependence.
Ver Steeg and Galystyan (2015) extend this definition to include topics by conditioning TC on another random variable, \( Y \), that represents topics; specifically, in the author’s formulation, \( Y \) represents binary latent topics, \( Y_1, \ldots, Y_K \) that have corresponding word groups as defined by, \( X_G \). Such that the total correlation conditioned on \( Y \) becomes:

\[
TC(X_G ; Y) = TC(X_G) - TC(X_G | Y) = \sum_{i \in G} I(X_i : Y) - I(X_G : Y) \tag{7}
\]

Given these definitions Ver Steeg and Galystyan (2015) maximize the explanation of the word dependencies through the latent topics by maximizing the lower bound \( TC(X_G ; Y_1, \ldots, Y_K) \):

\[
\max_{g_j, p(y|x_G)} \sum_{j=1}^{K} TC(X_{G_j} ; Y_j) \tag{8}
\]

The closer \( TC(X_G | Y) \) is to zero the more the topics, \( Y \), explain the dependencies seen in the word groups, \( X_G \). For implementation, the authors allow each word to be associated with only one topic (i.e., single-membership)\(^{66}\). The intuition here is that some of the total correlation is explained by each topic and the topics, the latent factors \( Y \), are optimized given the specific dataset and vocabulary.

In the Anchored CorEx version of CorEx the topics are learned from “anchor” terms provided through human input. The research results in Chapter 7 relies upon the Anchored CorEx version. Mathematically, Gallagher et al. (2016) describe this as the information bottleneck, where a parameter controls the influence of the anchor words in learning the topic representations. While the model can be used with and without these anchor terms, using the anchored terms is ideal if topics representing specific terms or ideas or known a priori. In addition to specifying anchor terms, a weight can be specified that represents the confidence or certainty that the anchor terms represent a topic in the corpus. This weighting specification is especially useful when a topic is of interest but only present in a subset of the documents. In this case, the small subset of documents will be associated with the topic and in a binary manner, rather than a probability distribution (as is the case with LDA) from all documents.

\^{66} For additional detail on the implementation refer to Gallagher et al. (2016) and Ver Steeg’s publications (academic, online, and code repositories).
4.5 CorEx Topic And Anchor Selection

Like many topic modeling algorithms (e.g., LDA), CorEx requires the hyperparameter for the number of topics, $K$, to be chosen. Unlike the LDA hyperparameter optimization discussed previously, the CorEx hyperparameter optimization for this research is more concerned with representing the domain knowledge in a subset of topics. This domain knowledge comes in the form of topics that correspond to predetermined investment areas, specified through anchor terms. In this setting, the exact number of topics is not as important as having a subset of topics that distinctly and accurately represent the investment areas. The topic number, $K$, is obtained by producing CorEx topic models across a range of values and then identifying the $K$ that produces the highest percent of topics described by groups of anchor words (i.e., investment areas). As multiple topic models with different values of $K$ identify most of the investment areas, with equal percentages, an additional step is taken that examines document clusters within topics. The homogeneity of the documents that correspond to the subset of investment area topics are compared.\(^{67}\)

Future implementations of a CorEx hyperparameter optimization could examine a combined measure of topic homogeneity and topic quality based upon coherence. Gallagher et al. (2016) implement a measure of semantic topic coherence that correlates with human interpretation (Mimno et al. 2011)). This type of measure, rather than perplexity, could be used to compare models like CorEx, LDA, CTM, and HDP. Gallgher et al. (2016) explains that “CorEx does not explicitly attempt to learn a generative model and, thus, traditional measures such as perplexity are not appropriate for model comparison against LDA.”

In the Anchored CorEx model, the strength of the groups of anchor terms also needs to be set. As with the number of topics, CorEx topic models are created with the determined $K$ value over a range of values that determine the influence of the anchor weight. Using the various anchor weights, the average Jaccard measure of similarity (Kosub 2019; Levandowsky and Winter 1971; Niwattanakul and Singthongchai 2013) is computed between the anchor terms and top terms describing the topic. While this process is performed iteratively, it could also be performed as a grid-search similar to the LDA model selection process.

Constructing the groups of anchor terms is an iterative process. The initial groups of anchor terms consist of all terms that were sub-categories of the investment areas. Following stemming, terms in the anchor groups that were not found in the corpus vocabulary were removed. Then words that appeared related or appeared in other topics were added to the corresponding anchor group. To further improve topic quality and reduce topic ambiguity, words that were less germane to an investment area topic but sometimes related were added to separate

\(^{67}\) A similar type of measure could be used to compare document clusters in LDA by assigning a document to the topic with the highest probability or to a topic if the probability meets a threshold value.
anchor groups. This expanded list of topics and anchor word groups leads to more interpretable topics that represent the subject matter expert’s idea of the topic—in this case investment areas.

While some approaches to automate this process exist, human discretion is valuable to the conceptual understanding of the topic. Improvements to this process need to maintain the understanding while also automating the anchor groups. One approach would be to add or remove anchor words to groups based upon statistical model improvement. These endeavors towards additional automation in constructing the groups of anchor terms is a suggested route for future research and is not implemented in the research here.

4.6 Using Sentiment Along With Topic Models

Text information consists of two broad categories, factual information and opinions. While facts are objective, opinions are subjective and consist of sentiment related to an object or object’s feature set. Text mining and textual related research is well established on the factual side and is grounded in the areas of information extraction, natural language processing, and supervised and unsupervised text classification. However, the area of research associated with the computational study of opinions, sentiment, and emotion in textual forms is known as sentiment analysis and as a subjective side of textual analysis, sentiment analysis is a less refined field. Opinion passages or opinionated portions text can be expressed towards both the object and features by the opinion holder or opinion source, (i.e., message poster and news articles, respectively). These opinions can contain positive, negative or ambiguous information regarding the object or feature which is known as sentiment orientation or polarity. With the ever-increasing amounts of information and methods with which to convey opinion such as blogs, social media (e.g., Twitter), and opinion news, automating sentiment analysis is crucial to understanding how a product, policy, or idea is faring in the public.

The majority of the research in sentiment classification has gone into document level classification, mostly using supervised learning methods (Chesley et al. 2006; Y. Choi et al. 2005; Hatzivassiloglou and McKeown 1997; Hatzivassiloglou and Wiebe 2000; Wiebe et al. 2004). Sentiment classification can occur at the document level, known as coarse sentiment (similar to sentiment analyzed from news and blog contents in Chapter 6), or at the sentence level, known as granular sentiment (similar to sentiment analyzed at the headline level in Chapter 5). Sentiment analysis at the word level is less common but possible. In document-level classification, the assumption is that the document being examined contains opinionated text

An unsupervised method that started much of the research in this area is based upon part-of-speech (POS) tagging adapted from document-level analysis—Turney (2002) applies this method to movie reviews. The method extracts two-word phrase combinations using POS tagging; two-word combinations consist of noun phrases surrounded by an adjective or adverb. Significant
phrases receive either positive, neutral, negative or opinion, no-opinion classification. In previous research, these two-word combinations have been shown to indicate subjectivity. The method then estimates sentiment polarity based upon the coocurrence probabilities of the two-word phrases and polarity. Turney uses excellent and poor as the reference words to signal polarity and calculate the co-occurrence probability. The sentiment polarity is then averaged over all two-word phrases in the document. If the sentiment polarity is positive it indicates a close association with excellent or positive polarity and contrary holds for poor.

There are various methods available for sentiment analysis and the most common include training a known dataset, creating classifiers, and using predefined lexical dictionaries. Most are based on word polarities identified in a dictionary, where some words are positive, others negative, and some neutral. All the lexical dictionaries compare a document’s terms to the lexicon and identify words within the dictionary and then an aggregate measure of polarity is computed. See Figure 4.6 for an illustrative example where the term document could represent a longer text selection (e.g., document or paragraph) or a short piece of text (e.g., headline, tweet, or sentence).

**Figure 4.6:** Example of obtaining the intersecting terms to determine sentiment for a document. Source: Medium, The Startup (A 2018).

For this research, a new sentiment measure is not created, rather predefined sentiment measures which rely upon pre-trained models are applied and compared. A prebuilt sentiment analyzer, the syuzhet package in R\(^\text{68}\), is used with a few different lexical dictionaries (M. L. Jockers 2015). The common lexical dictionaries compared follow\(^\text{69}\):

- Syuzhet created from Nebraska Literary Lab by Matthew L. Jockers (M. L. Jockers 2015)
- Afinn created as Afinn Word Database by Finn Arup Nielsen (Nielsen 2011)

\(^{68}\) Using sentiment extraction from Stanford’s coreNLP software (Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky 2014).

\(^{69}\) For specific details please refer to the associated papers or the R package implementation.
Bing created as the Opinion Lexicon by Minqing Hu and Bing Lui (M. Hu and Liu 2004; Liu, Hu, and Cheng 2005)
NRC created the NRC Emotion Lexicon by Saif M. Mohammad and Peter D. Turney (S. M. Mohammad and Turney 2010)

Previous literature demonstrates social media’s ability to make stock market predictions and, more broadly, to gauge interest and disinterest in the product market (Cui, Mittal, and Datar 2006; Pang, Lee, and Vaithyanathan 2002; Pang and Lee 2008). In these cases, the market prediction and product assessment use positive and negative sentiment detected from social media text. The sentiment analysis (or polarity) exposes contextual attitudes and reactions within the text. In this research, the average sentiment (attitudes and reactions) is calculated by topic for the most relevant documents using the metrics mentioned above (Syuzhet, Afinn, Bing, NRC) and one chosen for qualitative analysis. Sentiment is used as an additional lens to identify interest from the public (where the private sector has a presence) in the climate-related financial products.
5. The Evolution Of Climate Finance Using Topic Modeling

Chapter 1 summarized key issues and milestones associated with addressing climate change and related financing initiatives. Over time, the foci of issues associated with managing climate change have evolved. So too have the public discussions around financing efforts to address climate change. This chapter assesses and explains these changes using LDA topic modeling.

This chapter records how the portrayal of climate finance in media, particularly global news outlets, has evolved over time. Textual data containing climate finance relevant terms are extracted from media sources and indexed by time. This data is fed into a model that uncovers latent meaning from unstructured data, which then identifies themes related to climate finance. The relative influence of certain themes and topics over time is assessed in the overall media conversation, along with their rise and fall and geographic importance.

This chapter begins with a brief review of the motivation and summary of the policy context\textsuperscript{70}, prior to introducing the research questions. The research questions are followed by descriptive statistics of the input data. The summarization of the data, along with some metadata, places the data in context of world events related to climate change. Next, the modeling output and results are discussed in conjunction with climate finance policy implications. Note, that the high-level review of the methodology and evaluation applied in this chapter is outlined in Chapter 4. The chapter ends with a deeper discussion of the generalized results and considerations for textual analysis in public policy.

5.1 Policy Background In Brief

Since the inception of the United Nations Framework Convention on Climate Change in 1992, the key objectives of international climate talks have been stabilizing emissions and preventing adverse impacts of a warming climate. Within the international framework, there are two primary components of responding to climate change: (1) adaptation, which involves structural and behavioral changes to reduce the human and environmental impacts of climate change, and (2) mitigation, which involves human interventions to reduce or remove GHG emissions. Within international discussions, both typically involve assistance to developing countries since most of them have historically contributed relatively little to GHG emissions. The first decade of climate negotiations focused on mitigation via emission reduction commitments

\textsuperscript{70} Refer for Chapter 1 for additional policy context.
(Ciplet, Roberts, and Khan 2013). As the subsequent aftermath of both the Kyoto Protocol and Paris Accord demonstrate, global adherence to emission reduction targets has been at best inconsistent, and substantial achievement on this front largely remains elusive.

Subsequent years of climate negotiations have expanded the discussion to include other matters, including financing the global response to climate change. Following the development of financing products and a consensus agreement that all countries should maintain quality of life by adapting to climate-related changes, some emphasis of more recent climate talks have shifted towards ensuring sufficient funds and promoting types of financing mechanisms (Ciplet, Roberts, and Khan 2013). At the same time, nations acknowledged it was unclear how developing countries would find the resources to become both carbon neutral economies and resilient in the face of the physical impacts of climate change (Montes 2012), especially since developed countries often wrestle with resource challenges.

Of the climate finance raised to date, most has been spent in the country in which it was raised, presenting a severe hurdle for low-income countries. Further challenges are faced on account of varying definitions for what constitutes climate finance, which exacerbates uncertainties and variance of estimates for financial needs and finances provided (J. T. Roberts and Weikmans 2017). Regardless of uncertainties, the sheer volume of the estimated financing requirements will likely require funding streams derived not just from governments or the public domain, but also from public, private, and intermediary (e.g., multi-lateral organizations like the World Bank) sources. Both public and private institutions are mobilizing increasing amounts of climate finance because it is widely understood the current climate finance flows do not meet the mitigation and adaptation targets outlined in the Paris Agreement (Kawabata 2019). For further discussions about climate finance, international agreements to address it, and the evolving role of the private sector, see Chapter 1.

Broadly speaking, the current policy goal in climate finance is to ensure adequate financing flows (whether public, private, or partnerships) are available, such that everyone can capably address climate related impacts, transition to sustainable carbon-neutral economies, and create climate-resilient communities. Increasing climate financing, addressing overlooked sectors and industries, and creating avenues for the neediest of countries to access funding are examples of sub-components of the same overarching policy goal.

5.2 Research Questions And Approach

This research does not intend to solve or propose a solution to the policy problem of climate finance, or the subcomponents described, but better equip policy-makers in their

---

71 Reducing the uncertainties around finance needs is an example of a policy problem related to climate finance but not addressed in this paper.
decision-making processes for such problems. The substantial role of the private sector and market mechanisms in facilitating climate finance (Barbara K. Buchner, Padraig Oliver, Xueying Wang, Cameron Carswell, Chavi Meattle, and Federico Mazza 2017; J. Brown and Jacobs 2011) emphasizes the need to understand this landscape and be more responsive to its changing dynamics. With additional insight into the global discourse surrounding climate finance, these financial institutions, financial intermediaries, and policy makers can become better positioned to tailor a message or roll out new initiatives. Furthermore, understanding the counter position (in this instance, negative sentiment towards financing as a concept) could help stakeholders clarify points of contention, adjust collective conversations, or help reach a compromise. Take for example the frequently contentious matter of who bears responsibility for the impacts of climate change in developing countries; understanding the topics most associated with these countries may help policy actors prioritize and incorporate developing country concerns (in this research, surrounding climate finance). Finally, by incorporating metadata like geographic and temporal components into their analyses, policy actors could qualitatively understand if the timing or location of climate finance initiatives are going to be more or less likely to be received positively.

This research combines quantitative and qualitative analyses to uncover themes within the global climate finance landscape as described by general news sources compiled by LexisNexis. As described in Chapter 3, news headlines are used as a proxy to represent the discussion surrounding climate finance. Using these selected news headlines as the data source, and natural language processing and topic modeling as the analytical methods, the research aims to answer the following questions, more broadly defined in Chapter 1:

- What major topics do exist within the climate finance landscape?
- How has the emphasis of these topics evolved as the industry has matured?
- Do the topics reveal geographic and temporal relationships a policy maker can use to tailor their climate finance activities?
- Based upon the topics, can interest be gauged in new climate finance products?
- Is analysis of news data a reasonable method to understand the climate finance landscape and changes to it?

Answers to these questions are sought by fitting a standard unsupervised latent Dirichlet allocation or LDA topic model (discussed further in Chapter 4), optimizing the model’s parameters, qualitatively reviewing the output to create labels for the latent aspects (i.e., the themes or topics identified by the model), and then incorporating metadata into the discretized

72 In practice, topics discovered by modeling represent broad themes within climate finance.
model output to observe trends over time and geography. From this output, insights into the above research questions are obtained.

### 5.3 Summary Of Climate Finance News Headline Data

The data used in this analysis corresponds to the sub-corpus\(^{73}\) obtained from LexisNexis using the consolidated list of “Climate Finance” keyphrases as specified in Chapter 3: “climate finance”, “green climate fund”, and “climate policy initiative”. The unit of importance is the word or phrase (concatenated words) that create the news headline. This corpus contains news headlines between the years 2010 and 2018. Sources where the consolidated keyphrases were found within the content (e.g., news article) body or the headline were added to the corpus. This results in a total of 8,168 news headlines, distributed by year as shown in Figure 5.1. The cause of the spike in 2015 cannot be confirmed but could be attributed to the discussions leading up to the signing of the Paris Agreement (ratified in 2016 but agreed to in 2015).

![Figure 5.1: Distribution of the LexisNexis “Climate Finance” sub-corpus documents (news headlines) by year of publication.](image)

By examining the corpus at a more granular level (i.e. monthly), it is observed that the spikes in volume tend to correspond to world events surrounding climate change action and international meetings. Many of the volume spikes align with various Conferences of the Parties that occur at the end of each year (see Appendix C for a complete list). Figure 5.2 displays this

---

\(^{73}\) This is the article headlines from the search strategy identifying key phrases for climate change funding.
time series and the labels attached to the spikes are qualitatively assigned. The spikes in volume can be considered increased discourse or conversation and the annual meetings are represented by the Conference of the Parties (CoP).

**Figure 5.2:** Count of documents (y-axis) in climate finance corpus over time (x-axis) by month from 2010 to 2018. Many spikes in volume correspond to the global meetings and world events. CoP refers to the annual international climate summit, called the Conference of the Parties. Note in 2013, during CoP19 parties agreed to establish Intended Nationally Determined Contributions (INDCs)\(^\text{74}\); these INDCs were in anticipation of the Paris meeting in 2015.

Given the geographic distribution of search interest in climate finance (shown earlier in Figure 3.3), it is unsurprising that two of the top publication sources originate from the US and the UK. Non-western press outlets within the top sources (Figure 5.3) are publications from Pakistan (Daily Pak Banker), India (The Times of India - electronic and print), and Bangladesh (The Financial Times). There are 641 unique publication sources contained in this sub-corpus.

\(^{74}\) According to the World Resource Institute (World Resources Institute 2014) “Countries publicly outlined what post-2020 climate actions they intended to take under the new international agreement, known as their Intended Nationally Determined Contributions. The climate actions communicated in these INDCs largely determine whether the world achieves the long-term goals of the Paris Agreement: to hold the increase in global average temperature to well below 2°C, to pursue efforts to limit the increase to 1.5°C, and to achieve net zero emissions in the second half of this century.”
Using the geographic data field discussed in Chapter 3, the countries and regions to which the articles pertain are shown in Figure 5.4. Each of the panels are rank-ordered by the volume over the entire time period. Bars corresponding to the early time period (2010 to 2014), more recent period (2015 to 2018), and both periods are also shown for the regions and countries. It is important to note that not all news headlines contain a geographic label. In the regional view, any label that corresponds to a country is rolled-up to the region. It is nonetheless interesting to note that while North American and South and Central Asia contain the highest counts of pertinent news over the entire time period and sub periods, some regions appear to have received less attention over time. In particular, Africa (East, North, Southern, Western), Australia and New Zealand, South America, and the Middle East observed declines in counts from the early part of this decade to the later part. At the country-level there are many that exhibit similar phenomena, such as Australia, South Africa, Peru, and Mexico). On the contrary, France, Pakistan, and Germany are three countries whose headline count increased over time. It is possible that these countries have become more relevant to climate finance discussions in recent years.
Figure 5.4: Country (left) and region (right) rankings for climate finance news sub-corpus. A document can pertain to more than one country or region (i.e., multi-class labels). The dark grey bar covers the full time period (2010 to 2018), the medium grey bar is recent time (2015 to 2018), and the light grey bar covers a time period from 2010 to 2014. For a table by region, see Appendix K.

5.4 LDA Model Results

This section is divided into two sections: (1) outlining the process for selecting the model parameters and final model used to describe the topics in the climate finance corpus and (2) presenting the model results, which includes the six topics discovered in the corpus as well as their temporal and geographic relationships, and placing them within a policy context. The following sections in this chapter build upon these results by describing some policy implications and extending the discussion to the role of private engagement.

5.4.1 Model Selection

The model selection process aims to determine the parameters that are used in the LDA topic model and to validate the model with the chosen hyperparameters (by evaluating in-sample and out-of-sample performance). As described in Chapter 4, the three hyperparameters are \( K \), the number of topics, \( \alpha \) the two concentration parameters for the Dirichlet priors, \( \beta \) the per-topic document distributions and \( \beta \) the per-term topic distribution. Through this selection and

---

Refer to Chapter 4 for the details on LDA.
evaluation process, an LDA model with six (6) topics was chosen, as were concentration hyperparameters that provide improved model performance compared to default hyperparameters. These hyperparameters are discussed in Chapter 4 and details of the selection process with results are discussed in further detail in the subsequent sections.

To select the LDA model hyperparameters, a grid-search strategy was implemented. In the grid-search, an LDA model is evaluated for each of the possible hyperparameter combinations. These hyperparameter combinations represent a three-dimensional grid identified by the number of topics ($K$, from 4 to 20), the per-term topic distribution ($\beta$, 25 increments from 0.01 to 0.5), and the per-topic document distribution ($\alpha$, 25 increments from 0.05 to 12.5). As described in Section 4.2, metrics quantifying the quality of the model, called the CaoJuan (Cao et al. 2009) and Devaud (Deveaud, SanJuan, and Bellot 2014) metrics in the figures, are evaluated for each LDA model with a corresponding parameter combination (Hornik and Grün 2011).

By rank-ordering the models, the per-term topic distribution hyperparameter, $\beta$, is consistently at the lower end of the range at 0.01. Using this constant value for $\beta$, and varying the topic number, $K$, and the per-topic document distribution hyperparameter, $\alpha$, the optimal models tend to be concentrated in the mid-range of $\alpha$ and with the number of topics between six (6) and nine (9). The set of heatcharts in Figure 5.5 displays the output of the models evaluated over these possible parameter combinations; darker cells indicate more optimal models. As shown with the red circle, a “hotspot” of optimal models exist in both the CaoJuan and Devaud metrics when the topic number is equal to six and the per-topic document distribution parameter is 6.275. Additional evaluations showing consistency in the optimal model using these two metrics can be found in Appendix L.
Figure 5.5: Grid-search with constant per-term topic distribution concentration hyperparameter (0.01)$^{76}$, varying the per topic document distribution concentration hyperparameter (from 0.05 to 12.5) on the y-axis, and topic number (from 4 to 20) along the x-axis. Cells represent LDA model evaluation results using the CaoJuan (left) and Deveaud (right) metrics (discussed in chapter 4). Values are normalized by metric. Darker cells represent superior model results. The cells contained within the red-circle represent relatively good performing models within the displayed metric and across the two metrics.

The values obtained for the hyperparameters ($K = 6, \alpha = 6.275$ and $\beta = 0.01$) with the grid-search (above) are considered the “optimal values”. To see the sensitivity of the optimal number of topics with these concentration hyperparameters, the performance metrics are compared to default or base concentration values. The base values are $50/K$ and 0.10 (Steyvers and Griffiths 2007), for $\alpha$ and $\beta$ respectively where $K$ is the number of topics. The top panel of Figure 5.6 displays the optimal parameters and the bottom panel the base parameters. Each panel contains two plots corresponding to the two metrics used - CaoJuan (upper and optimal $K$ is a minimum) and Deveaud (lower and optimal $K$ is a maximum). $K$ is either minimized or maximized on account of each metric. In both situations, the optimal number of topics is consistently determined to be six. When the optimal and base time-series are compared and normalized within the metrics, the optimal parameters indicate better model performance than the base parameters, and again confirm that six is the optimal number of topics as demonstrated in Figure 5.7.

$^{76}$ All other values of the per-term topic distribution were evaluated. The models performing the best (rank ordered) always had a per-term topic concentration value of 0.01.
Figure 5.6: Selection of topic number and parameter values through tuning. The top two plots contain the normalized CaoJuan and Deveaud metrics by topic number (y-axis) using the optimal concentration parameters. The blue rectangle indicates the optimal topic number is six for both the CaoJuan and Deveaud metrics. The bottom two plots depict similar output using the default concentration parameters; again, the optimal number of topics are consistent across the metrics as well as with the optimal hyperparameters.
Figure 5.7: Metric comparison for optimal and base parameter cases. The top plot compares the CaoJuan metric from the LDA model using the topic number (y-axis) and concentration parameters for the optimal parameters (circle) and default parameters (square) - the best topic number is chosen when this metric is minimized. The bottom plot compares the Deveaud metric from the LDA model using the topic number (y-axis) and concentration parameters for the optimal parameters (triangle) and default parameters (cross) - the best topic number is chosen when this metric is maximized. These time-series represent the same series in Figure 5.6. In all cases, a $K=6$ is the best LDA model.

As discussed in Chapter 4, the out-of-sample performance is also compared across topic numbers, by base and optimal concentration hyperparameters ($\alpha$ and $\beta$), and between the LDA and correlated topic model (CTM). A lower perplexity score is the desirable result. In Figure 5.8, the LDA models (regardless of concentration parameter choice) perform better than the CTM models using topic numbers from four (4) to 11. Using an LDA model with 5, 6, 7, or 8 topics shows that the optimal hyperparameters are superior to the default or base hyperparameters. This confirms the value of using a model with either of these topic number values. Despite the perplexity being lowest for a $K$ value equal to 7 or 8, a model with 6 topics is still used due to the consistency in the in-sample performance (using the CaoJuan and Deveaud metrics) for 6 topics.
Figure 5.8: Boxplot out-of-sample comparison of topic models across a chosen K value (number of topics on x-axis) using perplexity. Lower perplexity (y-axis) is desirable. LDA models different concentration parameters. The optimal values for both the CaoJuan and Deveaud metrics are the same for topic numbers 4 to 11; the LDA Optimal bars represent these LDA models. The LDA Base models vary given the number of topics.

From the grid-search, a candidate model very similar to the six (6) topic model has nine (9) topics and a slightly larger $\alpha$ value. When models with more topics are compared to models with fewer, there are clusters of documents representing the topics that are very similar. Figure 5.9 shows how the documents cluster in the LDA models with 6 and 9 topics. For every topic in the six-topic model, there is a very similar topic in the nine model; three topics from the nine-topic model are isolated (topics 1, 4, and 5 - LDA9_1, LDA9_4, and LDA9_5 respectively).

---

77 Based upon the symmetric Dirichlet distribution, this implies more topics and more similar topic distributions ($\alpha$ is larger)
Figure 5.9: Hierarchical clustering of the topics from the six and nine topic models. The similarity between the topics is measured using the Hellinger distance (Nikulin 2001). The clustering is agglomerative and minimizes the variance using the Ward method (Ward 1963).

Each of these three topics are relatively similar to more than one of the topics in the six topic model—Figure 5.10 shows the similarity based upon the Hellinger distance measure (Hellinger 1909; Rus, Niraula, and Banjade 2013). As the number of topics increases, new topics are combinations of pre-existing ones. In Figure 5.10’s heatchart, lower values represent more similar topics. For topic 1 of the nine-topic model (LDA9_1) it is most similar to topic 1 (LDA6_1, referred to as the Responsibility (1) topic) and topic 3 (LDA6_3, referred to as the Collective Action (3) topic) in the six-topic model. Topics 4 and 5 of the nine-topic model are most similar to topic 2 in the six-topic model (LDA6_2 called Reduction Plans (2)), and then topic 3 and topic 4 (LDA6_4, referred to as the Multi-lateral Action topic), respectively. These topics make plausible sense as subsets of broader ideas. In this instance, the new topics in the nine-topic model would refer to collective action to address the global responsibility for climate change, emissions reduction through collective action, and emissions reduction through multi-lateral action, respectively. Given that these new topics were not deemed exceptionally relevant to the policy questions at hand, and the strong performance and consistency of the six-topic model was observed before, the six-topic model was retained for further analysis.

The Hellinger distance $HD(P, Q)$ for two probability distributions $P$ and $Q$, is the probabilistic analog to the Euclidean distance measure and is used to measure the difference between two probability distributions. In this case the probability distributions for LDA are represented by $\theta$, for example $P = \theta_{LDA9_1}$ and $Q = \theta_{LDA6_1 - Responsibility}$. Unlike the Kullback-Leibler divergence (relative entropy), the Hellinger distance metric is symmetric.
5.4.2 LDA Six-Topic Model Output

This section presents the six topics discovered in the climate finance corpus. The topics represent the latent information from the LDA model with hyperparameter optimization. Descriptive titles are provided for each topic as well as policy-relevant descriptions and trends. The output is linked to the research questions discussed earlier in the chapter.

What major topics exist within the climate finance landscape?

The final model chosen for the analysis contains six topics; these six topics represent the climate finance conversation within the chosen corpus. The topics were identified through common words and documents (i.e. news titles) that have greater proportions corresponding to the topic. For each of the six topics, simple qualitative labels were assigned, and these labels were used throughout the chapter in place of topic numbers. The six topics, brief descriptions, and their most frequent terms are shown in Table 5.179:

---

79 For the interactive output and LDA visualization, please see Appendix M.
<table>
<thead>
<tr>
<th>Topic Name</th>
<th>Responsibility</th>
<th>Reduction Plan</th>
<th>Collective Action</th>
<th>Multi-lateral Action</th>
<th>Addressing Causes</th>
<th>Mobilizing Capital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic Details</td>
<td>Developed vs Developing Responsibility (Manage and Fund Climate Related Impacts)</td>
<td>Carbon Emission Reduction Plans (Deals and Summits)</td>
<td>Global Collective Action (Paris Accord) Policy and Regulatory Changes</td>
<td>Individual, bilateral, and multi-lateral action</td>
<td>Addressing climate change causes (environmental, power, industry, economic)</td>
<td>Mobilizing capital - climate investments and financing partnerships</td>
</tr>
<tr>
<td>Salient Phrases</td>
<td>climate change, cap, fighting, least, contribute, real, priority</td>
<td>Paris, cut, global warm, make, year, country, emission</td>
<td>California, regulatory update, European, service, file, regard, public utility commission</td>
<td>live, Canberra, president, politics, parliament, statement, development</td>
<td>industry, coal, power, deep, UK’s, fire</td>
<td>GCF, declaration, grant, partnership, approve, ADB, million</td>
</tr>
<tr>
<td>Top Topic Terms</td>
<td>climatechange greenclimatefund fund talk country pledge need nation help must meet developnation fight set can</td>
<td>unitednations climatesummit paris india world emission deal carbon trump year cut leader globalwarm target climatedeal</td>
<td>new global energy paraisagreement issue conference report agreement environment european international meeting document pacific europeunion</td>
<td>washington will unitedstates china minister pakistan summit development state make g20 president obama climateaction pariscমিল্পাকালিমেটsummit</td>
<td>climate say call climatefund urge action aid take policy coal economy big power pact seek</td>
<td>green project finance climatefinance plan investment support africa billion bank growth million lead worldbank get</td>
</tr>
</tbody>
</table>

*Table 5.1:* The six topics, brief descriptions, key example terms, and top terms for each.
In Figure 5.11, the similarity (using the term proportions) between these six topics is displayed. The values closer to zero indicate more similar term distributions. The values are relatively close to one and within a narrow range because the term distributions cover the entire vocabulary. The Responsibility and Reduction Plan topics are the most dissimilar from all other topics, whereas the Multi-lateral Action and Collective Action topics are the most similar, followed by the Addressing Causes topic.

![Figure 5.11: Heatmap displaying the similarity of topics from the six LDA topic models. The similarity between the topics is measured using the Hellinger distance that takes on values between 0 and 1 (Nikulin 2001) from the word proportions describing the topics. A lower value indicates more similar topics. The dendrogram at the top illustrates how the topics cluster based upon the similarities. The narrow range of values is expected given the higher value and the incorporation of all documents, rather than a subset of the documents that best represent the topic.](image)

**How has the emphasis of these topics evolved as the industry matured?**

To answer this research question, the average document proportions of the total corpus ($\theta$ from the LDA model) were smoothed and plotted over time. This depiction is shown in Figure 5.12, followed by extended descriptions of these topic trends along with a representative headline. Note that headlines may still retain language alterations as a result of processing. Over time, we see a change in which topics become more or less prominent. Around 2010, the dominant topic was Addressing Causes. Most recently, the dominant topic has become Mobilizing Capital.
Responsibility: This topic corresponds to the division in responsibility for climate change action, like emissions reductions and support for financing, between developed vs developing nations. The temporal trends show this is a prominent topic of conversation and consistently present in international climate discussions. The topic is more prominent during discussions about the Green Climate Fund but declined as discussions focused around the Paris Agreement. Discussions around this topic have slightly increased since the ratification of the Paris Agreement.

Example Headline: Those who contribute least to climate change fight for survival; On the frontline of climate change, effects are real and measurable. Paris talk must make most vulnerable countries top priority and cap warming at 1.5°C.

Reductions Plan: This topic is primarily focused on emission reductions, particularly rates and timelines. Over time, this is a relatively steady and persistent topic. It plays a more prominent role in the discussion early in the decade and became less of a focus during a time progressed and as the Paris Agreement framework provided for countries to determine their own emissions.
plans and targets. Focus on the topic has continued to decline in the following years since ratification.

*Example Headline:* EU wants Paris climate deal to cut carbon emissions 60% by 2050; A major UN climate summit in Paris later this year should call on countries to make tough carbon cuts to avoid dangerous global warming, EU document says.

**Collective Action:** This topic is particularly focused on cooperation within the United Nations. While focused internationally, this topic also contains individual entities (countries, regions, states) looking to keep the spirit of global cooperation together, like California and Europe (and does not include the phrase United Nations). This topic shows a stark increase in the run up, signing, and ratification of the Paris Agreement in 2015 and 2016. International cooperation has persisted as a more prominent topic in spite of the US notice to withdraw from the agreement, and general populist and nationalist sentiments worldwide. The ongoing prevalence of this topic may represent the resilience of the Paris Agreement and the ability of the United Nations framework to weather the stresses of current events.

*Example Headline:* Register of Commission document: European Parliament non-legislative resolution of 18 April 2018 on the draft Council decision on the conclusion on behalf of the Union of the Framework Agreement between the European Union and its Member states, of the one part, and Australia, of the other part.

**Multi-lateral Action:** Individual transactions, Bilateral agreements, and Non-United Nation actions are all relevant to this topic. Headlines tend to focus on big players (US/China) or actors that could impede progress on broader UN-led cooperation (Pakistan/Canberra). The actions of individual countries buying in and committing to the run-up of the Paris Agreement was more prominent up until 2015, then waned after the Paris Agreement was passed. With the agreement in place, interest in the topic has sharply declined.

*Example Headline:* China and Australia sign historic free trade agreement - politics live; With the G20 now done and dusted, MPs are gathering in Canberra for a special sitting of the parliament. China’s president Xi Jinping has addressed the parliament and inked a new free trade deal with Canberra. All the developments, live.

**Addressing Causes:** Headlines within this topic are frequently associated with the energy and power/coal sector. The topic dominated early in the decade and declined over time as attitudes and actions instead focused on whether or not a compromise and agreement could be achieved in Paris. With that agreement now in place, the topic again appears to be on the rise. The
proportion of this topic is rising along with the Responsibility topic, and these are both appearing more prominently in United Nations climate negotiations.

*Example Headline:* Future of coal in the UK looks increasingly bleak, says industry leader; The UK's last deep mine set to close and coal fired-power stations phased out, but industry warning over power shortage.

**Mobilizing Capital:** This topic contains all things relevant to funding. This includes who will raise, allocate, and spend the capital needed to respond to all the challenges put forth by climate change. This topic was not the focus of several talks early in the decade, but finance issues rose steadily in the runup to and following the 2012 Doha talks that established the Green Climate Fund, which is reflected in the figure. The topic held relatively steady prior to the Lima discussions, and continued to rise as the Paris conference moved towards an agreement. Currently, it is the most prominent topic of all six. Interestingly, the challenges faced by international action on climate change since 2017 did not impede the rise of Mobilizing Capital as a topic of interest, much like how Collective Action had not decreased in relevance as a topic. This suggests the topic has become the most pressing for stakeholders in the past several years and could persist for several more.

*Example Headline:* EIB80 backs World's First Emerging Market Green Bond Fund with USD 100 Million Investment.

**Do the topics reveal geographic and temporal relationships a policymaker can use to tailor climate finance activities?**

Given the information above, the temporal relationships are useful to understand the context of the global trends related to climate finance. For instance, the rising interest in the Addressing Causes and Responsibility topics suggest there may be more receptivity to climate finance aimed at reforming the energy sector and supporting projects that go from developed to developing countries. Positioning agreements, plans, or strategic paths can also be aided by linking the geographic metadata to the topics. The heatchart in Figure 5.13 depicts the percent of articles within a geography that correspond to each topic. Some regions have a high percent of articles corresponding to a particular category and others have a low percent. Inference can be made by looking at regions where a topic is more or less prominent or at the key topics across the region.

---

80 In the headline, EIB refers to the European Investment Bank
Within the Mobilizing Capital topic for example, Southeast Asia, East Africa, Caribbean, Central Africa, Central America, and the Balkans all have a larger percent of articles related to this topic. Most of these regions also coincide with a strong representation in the Responsibility topic. This implies that financing should be positioned in a way to explicitly address what and who is responsible for the climate related impacts being resolved by funds. Given the fact that most financing stays in the country where it was raised, these regions might be ideal regions to explore and encourage cross-border investments. Because these regions are more attuned to the need for capital and historical responsibilities, they may be more receptive to foreign investment vehicles aimed at responding to these issues. Specifically, the Caribbean is vulnerable to cyclones and sea-level rise, which might suggest an effort for international and local financing for sea-wall or mangrove forest adaptations and low-carbon energy sources like wind and solar power. The high attention from the Balkans on Reduction Plans and Mobilizing Capital is rather
surprising; this should be investigated further, and policymakers may look to tie initiatives to the region’s local emission reduction plans. The percent of articles contained in the Mobilizing Capital topic are lower for higher income regions like New Zealand/Australia, Western Europe, and North America, each with ample available capital and innovative financial markets. Perhaps by raising awareness about capital mobilization in these regions, additional financing could be obtained. For Western Europe, this conversation should incorporate discussion around emission reduction plans given the prevalence of that topic. Conversely, discussions in North America and New Zealand/Australia should include more emphasis on multi-lateral action. For Polynesia/Micronesia, another island region like the Caribbean, the analysis suggests that any financing would benefit from explaining how they address the causes of climate change impacting the region.

**Based upon the topics, can we gauge interest in new climate finance products?**

Topics corresponding solely to specific climate finance products, such as green bonds, do not emerge in the six-topic model. With an alternative model, such as the anchored topic model (used in Chapter 6), it is possible to more concretely analyze climate finance topics. However, some insights can still be garnered from the chosen LDA model. Using a measure of positive and negative sentiment\(^{81}\), and basing the results upon the most salient documents in each topic of each year, a coarse measure of interest in each topic is obtained. Figure 5.14 shows this measure of sentiment smoothed over the years. From this rather rough measure, insights related to each topic can be obtained. However, because most sentiment measures focus on lexicons that classify words as positive or negative, there is uncertainty in all these insights. In addition, alternative sentiment metrics (see Appendix N) produce differing trends, however, the relative conclusions tend to be similar.

- **Mobilizing Capital:** Language indicates positive sentiment exists throughout time, which is likely representative of a general feeling that emphasis should be placed on capital or financing to combat climate change. As coverage of the topic increases, the positive nature of the language rises. This probably indicates there is an increasingly mature and confident climate finance market globally.
- **Collective Action:** Like Mobilizing Capital, positive sentiment rises over time. This suggests both growing confidence and desire around nations being able to act together on a global scale. The withdrawal of the US from the Paris Agreement appears to galvanize the effect.

\(^{81}\) Refer to Chapter 4 for a description of sentiment.
- Reduction Plans: The negative sentiment in this topic corresponds to the extension of the Kyoto Protocol (which was widely seen as a failure to develop more ambitious emission plans). Prior to the Paris talks, as nations developed concrete plans, sentiment turned more positive. Positive sentiment peaks after ratification of the Paris Agreement, and continues to decline in the years that follow. This could be attributed to a lack of commitment in the plans.

- Multi-lateral Action: The topic’s changing sentiment follows its temporal trend of proportion seen in the LDA model. It increases as nation-to-nation discussions help facilitate confidence in the Paris Agreement. Following ratification, multi-lateral action appears to be associated mostly with negative sentiment.

- Responsibility: This topic is always regarded as negative because there is often “finger pointing” (including claims that developed nations are not doing enough) among countries and industries. Negative sentiment declines in more recent years and could be attributed to the Paris Agreement being considered a relative success story.

- Addressing Causes: In general, this topic contains language that indicates negative sentiment. The time-series peaks represent neutral (but not positive) sentiment. This is likely due to discussion related to the causes of climate change, which often refer to the fossil fuel industry as dirty and how the problem remains on the whole unsolved. The sentiment’s uptick in more recent time could be associated with the rising cost-competitiveness of renewable energy.
Is analysis of news data a reasonable method to understand the climate finance landscape and changes to it?

Upon examining the output of the LDA topic model and placing it in a policy context, this topic modeling approach does provide valuable insights. These insights are temporal and geospatial and could also provide valuable geopolitical context. The implications of the insights are discussed in greater detail in subsequent sections. In summary, this research answers the question in the affirmative: these results do yield value for policymakers and actors from both public and private sectors.

5.5 Policy Implications

A comprehensive review of LDA topic modeling analysis indicates that these results provide policymakers and other stakeholders a valuable additional framework through which to observe changing trends and interests in discussions surrounding climate change and financing the response. The incorporation of sentiment demonstrates both positive and negative language is used in the discourse, where some topics contain more negative sentiment and others more.

---

82 Any news related to climate change denial would probably fall into the topics focused on addressing the causes or the responsibility to respond with financing. Both of these topics are relatively more negative in sentiment.
positive. Several of these overarching themes carry climate change policy implications for
government and private industry actors alike.

The most prominent conclusion illuminated by this topic modeling analysis of global news articles is the attention to mobilizing climate finance. This topic has increased in recent years, particularly after the passage and ratification of the Paris Agreement. It is observed that mobilization of finance as a topic appeared in the background of discussions around international climate negotiations earlier this decade, would wax and wane depending on other trends of the time, then finally became the dominant component of conversation towards 2016. The elevation of mobilization as a topic within the global media conversation around climate change suggests that the modern moment is as near to an ideal time for interested parties to encourage participation in climate finance, given its rapidly rising share of the global discussion. This timing is perhaps the most crucial takeaway for stakeholders; it is feasible to imagine a comprehensive campaign to recruit more firms and institutions to participate in climate finance rolled out in 2014 would have been received less warmly as attention was more focused on ensuring multi-lateral climate action.

The focus on bilateral and multi-lateral actions on climate change—i.e. outside the UN process—has declined sharply with the ratification of the Paris Agreement. This is understandable, as the culmination of the agreement effectively assured that the UN became the predominant venue in which discussions related to climate change would be held. Nevertheless, the analysis also shows that the actions or inactions of local, state, and regional actors can still drive other topics and trends. This would suggest that policymakers focused on climate finance should seek to focus on the UN process as the key venue for their actions.

Analysis also indicates that matters of responsibility and the root causes of climate change, particularly around emission-intensive energy, still remain in the global discussion even if they are not the most prominent. Their persistence even after important conferences suggest that the matters may not be satisfactorily dealt with within certain circles and could become issues of contention at future conferences. Policymakers would be wise to ensure their envoys are capably prepared to speak to these concerns effectively in international discussions on climate change.

Topic modeling can further help climate finance watchers more effectively tailor a campaign to encourage public-private partnerships to enhance climate finance by assessing what structures will resonate more with investors and the public. Enhanced focus on collective action suggests that partnership and support of the UN-backed agreement process would likely be a net benefit for enhancing investor and public interest. Likewise, the relatively constant interest in addressing climate change root causes and the responsibility of the impacts as topics of discussion suggest that projects geared towards increasing low-carbon energy or those geared
towards partnering developed country capital with developing country projects could serve to enhance momentum with investors.

Additionally, as the topic of multi-lateral (but not global collective action) has declined in recent years, a current green bonds issuance (discussed in Chapter 6) would probably want to avoid financing, certification, or other association with single developed countries or institutions such as ASEAN (Association of Southeast Asian Nations) or the OECD (Organization for Economic Co-Operation and Development), as data suggests these would resonate less than a UN-backed initiative. Emission reduction plans have also fallen out of favor, suggesting that green bond projects should look to play down any association with how the supported projects might support a country’s commitment to the Paris agreement, but rather more directly speak to its climate or developmental benefits.

Lastly, topic saturation by geography can provide climate finance stakeholders the granular insights allowing them to better target campaigns. It is first important to recognize that not all articles in the corpus are geotagged, which increases uncertainty in the conclusions derived but can still suggest trends for further investigation. Mobilization of capital appears most prominently in news articles from the Balkans, Central Africa, and Southeast Asia. Two of these regions are consistent with expert observations; Central Africa and Southeast Asia are highly susceptible to climate risks such as drought and storm surge, respectively, and are very keen to access capital to address these climatic stresses that are nearer-term and more existential than other regions. Interest from the Balkans is unexpected and warrants further investigation. The prominence of the topic suggests that actors within the Balkan region could be more receptive to serving as a willing partner to either provide capital or develop local projects. Public and private entities may wish to redouble efforts in these regions to capitalize on rising interest in climate finance. Alternatively, climate finance outreach coordinators may wish to focus on areas where climate finance has resonated the least, so as to educate those unfamiliar with the topic and spurn laggards instead of preaching to the converted. North America and Australia / New Zealand are two areas with higher income levels (CIA 2018) and lower attention to the topic of Mobilizing Capital, suggesting a coordinated attempt to raise awareness could unlock substantial new streams of climate finance. Likewise, the Eastern Europe and South America’s relatively low level of attention combined with their own climate vulnerabilities could suggest an area where state and local agencies would welcome both adaptation and mitigation projects. Additional parsing and further analysis could help policymakers identify both where and how to deliver a message to encourage greater local engagement in climate finance.

It should be noted that the topics identified by the LDA topic model did not bifurcate climate finance into mitigation and adaptation topics. The topic Addressing Causes could be considered to focus on mitigation, but no corresponding topic related to adaptation, nor a topic
focused on finance for mitigation, was identified in this analysis. This could be indicative of the relatively small amount of coverage adaptation gets within climate discussions, despite its importance as understood by experts. It could also imply that focus is still on finding and implementing successful mitigation measures since adaptation measures are usually taken with mitigation efforts fail. Improving the profile of adaptation is likely to remain a challenge for policymakers. Future work could examine topics more concisely focused on adaptation and mitigation using another modeling approach (e.g., anchored topic model) or with a corpus containing information pertinent to these areas.

5.6 Contributions And Discussion Of Private Sector Engagement

This research contributes to the ongoing discussion of generating sustained capital flows to support low-carbon transitions and linking climate finance availability to climate finance needs. Understanding the emergence of new topics in the public sphere (e.g., in news media), as they relate to climate finance, could aid in the conceptualization, modification or introduction of new products for institutional and private investors. This would allow policy makers to more effectively leverage the capital of private institutions and reduce the financing gap. Additionally, identifying areas where private interest is lacking could signal a need for public financing or public policy solutions. This section discusses two broad policy-relevant matters raised by the results of this chapter—the desire to enhance current private sector engagement in climate finance, and the appropriate role of the private sector in addressing climate change and global public goods in general.

It is also important to recognize that many of the key takeaways presented in the previous sections, while consistent with expert judgment about international discussions, are not immediately intuitive to the lay reader following climate change news in press reports. Judging from popular press reporting, particularly the US and UK sources (L. Friedman 2019; Pulgar-Vidal 2019), it might be reasonable for an educated generalist to conclude that international discussions may focus on endeavoring to preserve multi-lateral cohesion in light of the United States’ withdrawal from the Paris Agreement and several other countries wavering in their commitment to the spirit of the agreement. However, as this analysis illustrates, multi-lateral action has largely faded from international discussion, with far more attention being paid to matters related to action on climate change, and climate finance in particular. This quantitative analysis mirrors journal articles which have observed international negotiations for years (Urpelainen and Van de Graaf 2018) but could appear contrary to popular press reporting. In this way, text analysis and topic modeling can serve as an effective check on assumptions about broader global sentiment based on a cursory review of headlines. Indeed, they can add rigor to analytic judgments made through qualitative approaches.
5.6.1 Enhancing Private Sector Engagement On Climate Finance

Unlocking private sector finance is widely regarded as the solution to achieving the unprecedented task of transitioning the global economy to a sustainable, equitable, and low-carbon one (Development Bank Consortium 2015). The reasons are predominantly twofold. The first is the scope and scale of the funding gap; while large uncertainties persist and are dependent upon what climate policies are implemented, estimates put the needs for climate mitigation and adaptation for the globe at a floor of hundreds of billions of dollars per year, with broad agreement that resources to address climate change need to be ramped up considerably over the next several decades (Fankhauser et al. 2016; Gupta, S., Harnisch, J., et al 2014). Such outlays challenge the treasuries of public institutions. Further, there is a long-standing awareness that funding for environmental and climate efforts is scarce (Ferraro and Pattanayak, 2006, James et al., 1999).

Private sector investors—whether individual investors, private equity including venture capitalists, or larger institutional investors like pension funds, insurance companies, or sovereign wealth funds—have assets under management representing several trillions of dollars globally. In addition, global, regional, and local financial institutions have the capacity to provide much needed capital and financial services to finance privately-developed climate change projects—if the terms are right. This has resulted in a common refrain in development literature which viewed private finance as a vehicle to potentially be “unlocked” for the purposes of climate finance (World Bank, 2015).

The second matter is credibility and sustainability. Given the paucity of funds, development finance often succumbs to politicized and strategic funding by donors (Berthelemy 2006). This undermines trust in the long-term viability of initiatives if they are subject to political shifts in donor countries. In the realm of climate finance, the US decision to cease support for the Green Climate Fund particularly underscored this point (Kotchen 2017). Fostering private participation in low-carbon markets not only addresses near-term development needs, but also supports the general viability of climate finance products in the inevitable event of political priorities from major donors shifting, if private markets are structured correctly. Thus, a unique opportunity exists for public and private actors to work together to increase climate-change-related private capital flows to developing countries.

Concurrently, the challenge of mobilizing capital for climate finance has become one of the most salient issues within the global news coverage of climate change, as this research demonstrates. This corroborates other observations in the private investment space. Growth in socially responsible investing and more integration of environmental, social and governance metrics (Stewart 2015) is evidence of increased interest in this investment domain that includes climate finance. As private investors look to combine financial return with social return,
investing in climate finance that fosters sustainable growth should be appealing. Industry reports note that climate finance continues to increase as it is driven by rising private sector interest in low-carbon energy (Padraig, Clark, and Meattle 2018). Understanding investor demand for both the type and specification of climate finance products is important. This understanding can aid in developing products like green bonds that facilitate public-private collaboration that ultimately help to bridge the financing gap or distribute capital to areas in need.

What remains for policymakers is determining how precisely to capitalize on this opportunity, and which strategy best fits with their goals. Several options exist to combine the insights derived from this topic modeling work with strategic plans. For instance, if policymakers determine they should accelerate the sum total of climate finance as fast as possible, policymakers could focus on expanding governance support for financing energy projects. Leading industry reports from the Climate Policy Initiative, show increased private sector investments are driven by renewables (Padraig, Clark, and Meattle 2018; Barbara Buchner, Angela Falconer, Morgan Hervé-Mignucci, Chiara Trabacchi, and Marcel Brinkman 2011; Barbara K. Buchner, Padraig Oliver, Xueying Wang, Cameron Carswell, Chavi Meattle, and Federico Mazza 2017), the cost of renewable83 technologies including power storage continue to decline and become more cost effective than comparable fossil fuel sources (Kåberger 2018), and this chapter’s research indicates growing coverage of the transition away from fossil energy—particularly coal—appearing as a news topic in the past 24 months. The combination of rising interest, sustained investment interest, and favorable economics could help to rapidly build momentum around climate finance as a concept as investment dollars multiply.

Alternatively, policymakers may seek to steer new investments towards addressing deficiencies that previous iterations of climate finance have missed. The existence of financing need is often undisputed, however, prioritizing where to place the next tranche of dollars is contested. A traditional approach is to look at geographic locales and sort by mitigation needs, adaptation needs, and capability to access to existing financial markets (Fankhauser et al. 2016). The research in this chapter presents potentially new metrics—including news coverage about mobilizing climate finance within the region and news coverage about energy transition within the region; both of these measures can further detect potential mismatches between investment appetite, coverage, and potential. The research here explored, but does not include, metrics related to sentiment or stance towards a concept. The incorporation of these measures could provide additional value. Other analyses call for alternative frameworks for climate finance, considering more diverse concepts such as distributive justice (Grasso 2010), hazard based forecasting (Coughlan de Perez et al. 2015), or to address less-prominent approaches such as

---

83 Renewables often fall into the mitigation category because they curb emissions, but they can also be considered adaptation efforts (Ley 2017).
decentralized electricity grids (Neha Rai 2016). Topic modeling of news coverage can help in any of these approaches by further assessing to what extent these approaches might gain traction with investors. It is left to the policymakers to determine how their institution envisions an appropriate goal for climate finance.

In regard to how an organization should carry out the work of incentivizing private industry, development organizations like ODI (Overseas Development Institute) have recommended a CP3 (Climate Public Private Partnership) fund model particularly to mobilize climate mitigation projects. CP3 funds provide limited amounts of public funding to support substantially larger private investments by ensuring that projects can develop financial returns if structured correctly. Lack of capital is addressed through the early stage equity provided by private investors, while the high risk perception typically associated with developing countries and some low carbon projects is addressed through public institution support, whose higher credit ratings can support higher risk and longer term initiatives (J. Brown and Jacobs 2011). By effectively mitigating up-front risk, public institutional support of private sector capital can be more effectively allocated to projects that both recipients need, and a class of investors want to support.

Regardless of the approach, anyone that leverages the methods within this paper should use them holistically—that is, in conjunction with other domain knowledge or available metrics. Following trending topics alone risks the development of financial products which solely trace the trend of the day, which contravenes the fundamental concept of CSR in which value is identified despite market and popular forces seeing otherwise. The work herein should not be taken in isolation. Prominent and emergent trends within news media need to be balanced with policy objectives, investor interest, investment risk and return, and humanitarian need.

5.6.2 Public And Private Roles In Addressing Climate Change

As noted in Chapter 2, climate change is an example of a public goods problem where CSR can play an effective role given that government responses to the problem to date are widely regarded as inadequate, and thus this is an instance of governments acting imperfectly. Given that financial resources will always face some constraints and any actor has limited capacity to solicit participation and myriad constraints on time, climate finance actors will wish to maximize investor interest. This could be done by ideally timing and structuring their products with their initial presentations instead of subsequent iterations, thereby increasing the optimization of the global public good that is a stable climate. Through topic modeling’s analytic insights, we observe how stakeholders might use them to encourage more social participation and ultimately increase a global public good.
As the analysis in this chapter concludes, greater attention is being paid to mobilizing finance to address climate change. This brings to light the awareness of the need for climate finance and the existing gap between what societies need to address climate change and what societies have allocated to the problem to date. There is also concurrence between the persistent participation from the private sector and the associated market failure to address the issue. Academic articles, and indeed UN documents (Jachnik, Caruso, and Srivastava 2015; Climate Finance sub-programme and Finance Technology and Capacity-building programme 2018; Z. Zhang and Maruyama 2001), speak to the importance of leveraging the private sector to address the climate finance gap; this suggests there is a consensus that the private sector has a role—and almost certainly a vitally important one—to play in providing the global public good that is a stable climate.

This analysis has generally accepted the proposition that private sector finance will play a role in sufficiently mobilizing climate finance as part of the global response to climate change during this century. This premise is built on the observation that the private sector has historically contributed the majority of climate finance to date and international organizations have encouraged their participation, as well as the scale of capital required (World Bank, 2015). This presents an interesting political science dilemma, as it effectively cedes that private enterprise has a role—and likely a key one—in providing a global public good. As previously noted, this can run contrary to the philosophies of some economists and political scientists. Further, the determination of what authorities should a government be allowed to take on intrinsically will vary across societies. Rare are other policy examples where national governments and international organizations are as welcoming to private enterprise, presenting researchers a unique challenge in presenting policy recommendations.

It follows that, as climate finance research proceeds, researchers must ask themselves which set of actors are the intended recipients of these insights (i.e., public or private) and how might they be best used. For example, in recognizing that renewable energy transition has persisted as a topic while mobilizing capital has risen, should this tailor the campaign of diplomats in packaging how central banks and treasuries can optimally support new projects, or that of financiers looking to engineer the most optimal climate finance product offering to the market? Both may work, and each actor has his/her own set of strengths. How should researchers navigate this issue?

Several guiding thoughts are offered here for the policy process as various disciplines inevitably continue to wrestle with the challenge of the most optimal and ethical role for each set of actors. First would be to begin every research project with a particular customer or actor in mind. Knowing an actors’ capabilities, reach, and intentions to play in the climate finance space can effectively resolve some of the toughest challenges. It is ultimately recommended to first
defer to a public actor to determine when and where a private actor should be brought in, even if the private actor would be presumed to take a leadership role. This cedes to the fact that the state has sole authority over legal frameworks and enforcement, and effectively can define the distinction through rule of law. Private actors, on the other hand, cannot. Lastly, researchers should consider the political capacity in a given region or space. While some literature observes that the past experiences of early climate adapters can be widely applied across income levels (Araos et al. 2016), certain government actors will not have the capabilities to design, allocate, and execute climate finance as the whole of the private sector. These capacity factors will influence what public and private actors can realistically undertake, even if political preferences might seek a greater public role. These considerations may serve as a helpful outline as this unique policy space with hybrid actors continues to develop over time.
6 Green-Climate Bond Opportunities Using Anchored Topic Modeling

As introduced in Section 1.7, the green bond is a specific financial mechanism used to finance the response to climate change by funding projects that will generate environmental benefits. This chapter builds on the green bond overview in Chapter 1 by providing details on the types of bonds considered in this research and specifying the underlying investment areas. The idea of using fixed income instruments for social good dates back more than a decade, yet the market for climate and green bonds continues to develop and evolve. As the market for socially beneficial and ESG-linked investments has matured, so too have green and climate bonds. Many of these bonds support specific investment areas like energy, transport, and water. Further evolution should be anticipated as community understanding and perceptions of climate vulnerability change, and stakeholders need to capitalize on such trends as much as possible if they wish to address the chronic financing gap.

This chapter describes emerging trends in media linked to green bonds and identifies potential opportunities in these investment areas by using Correlation Explanation or CorEx to model topics in a corpus containing both objective (news) and subjective (blogs) text. The research questions are framed within the policy context and followed by the temporal trends of the corpus. The approach (at a high-level, with additional details in Chapter 4) is presented along with the anchors that construct the topics corresponding to the investment areas. The methods and analyses of this research attempt to assess the feasibility of using social media and news to preemptively gauge investor interest of innovative ESG products or areas, like focusing on particular industries or regions.

The subset of relevant topic results from CorEx are presented in tandem with the investment dollars and description of sentiment trends. The relative temporal trends across all green bond investment areas are similar, but in particular cases, unanticipated results emerge, presenting possible opportunities for increased issuance for projects in these areas. The chapter closes with a discussion of potential opportunities and policy considerations in the labelled green bond market, given that investors will likely seek new investment products that support such initiatives based on the growth of socially responsible investment.

6.1 Green Bond Background In Brief

Collaboration efforts between environmental-climate policy and finance typically promotes the incorporation of ESG criteria into investment decisions. Understanding investment
decisions, however, ultimately depends on the type of investor being considered. Examples might include: 1) an individual investor who invests in solar and wind equities in order to help reduce the negative impacts of climate change, 2) an institutional investor who generates a portfolio of clean technology, biomass, and other “green” funds for its clients, and 3) a corporation that incorporates socially responsible and environmental practices into its business model and wishes to document its practices for potential investors. While the motivations for these investment decisions may not be the same – these actions are all types of socially responsible investments or impact investments that promote environmental sustainability. These actors almost certainly have several preferences behind these decisions, such as seeking profit above the market’s average return (known as alpha), improving reputation, minimizing risks, improving long-term prospects, and supporting a social good. While the exact definition of impact investing varies, a seminal report on the topic from JP Morgan states: “impact investments are investments intended to create positive impact beyond financial return” and defines socially responsible investing as investments which “generally seek to minimize negative impact” (O’Donohoe 2010).

Green bonds represent one avenue through which to execute impact investing. Green bonds and other climate-themed bonds are fixed-income products and are structured like most other bonds where their issuance raises large upfront capital. At their simplest, all bonds, including green bonds, are long-term debt securities issued by governments or other organizations that offer fixed-interest payments periodically for a period of more than a year. For more information on the bond market, please refer to Appendix O. Various bonds can support sustainable development, often through the support of infrastructure-intensive projects in the energy, industrial, transport, and building sectors. However, green bonds are unique in that they are issued exclusively for projects with environmental and/or climate benefits.

### 6.1.1 Green Bond Market Growth

Ensuring the global economy aligns with the goals of sustainable development cannot be addressed with government aid and philanthropy alone because the scale of the challenge overwhelms the capabilities of these actors. Addressing climate change and financing the response to it is a fundamental component of sustainable development, the private sector is integral to current strategies, and its role is unlikely to change. Over the past decade, government and philanthropic actors have staged a vigorous campaign to enlist the aid of private investors in pursuing sustainability and social impact as corporate goals. The number of investors stating they are interested in ESG has grown dramatically, and many in the financial world reject the notion that choosing between maximizing risk-adjusted returns and providing social value must be
binary. See Chapter 2 for more discussion on corporate social responsibility, and Section 3.3 for more about trends in ESG.

Green and climate aligned bonds have become a consistent and growing component of climate finance. They have come to represent one of the more mature markets used to finance a low-carbon transition. Since the first issuance from the EIB in 2007 and the World Bank in 2008 (both multinational banks) the rate of issuance in green bonds has outpaced that of the rest of climate finance. As described in Chapter 1, green bonds are fixed income debt instruments. Proceeds from green bonds or climate bonds are used to finance assets that provide benefit to the environment or climate resiliency. For more discussion about how bonds are structured and function in the financial market, see Appendix O.

The IFC accelerated market growth when it issued its first green bond in 2013. In the same year, Massachusetts issued the first green municipal bond. Figure 6.1 illustrates the public sector growth which was initially led by development banks (EIB, World Bank, and IFC), but then followed by local governments and sovereign wealth funds.

![Figure 6.1: Public sector growth in the green bond market. Growth has shifted from the initial development banks to a broad mix with more emphasis on government issued sovereign bonds (Climate Bonds Initiative 2018c). Source Climate Bonds Initiative](image)

Private sector interest followed and soon outpaced public sector issuances. In the same year that IFC and Massachusetts issued their bonds, a Swedish company issued the first corporate green bond, and solarCity issued a green asset backed security (ABS)\(^{84}\). Figure 6.2

\(^{84}\) ABS are collateralized pools of assets.
demonstrates how corporate private entities (ABS, Financial corporate, and non-financial corporates) now make up the majority of the green bond space.

![Figure 6.2: Recent green bond market growth is attributed to private institutions (Climate Bonds Initiative 2018c). Source: Climate Bonds Initiative](image)

This growth in the private sector has led to the development of green bond indices - metrics created by firms (e.g., Barclays MSCI, Bank of America Merrill Lynch, S&P Dow Jones) to track the green bond market - and listings of green bonds on multiple stock exchanges (e.g., Oslo, Stockholm, Luxembourg and London). This market growth has necessitated the desire for assessments from rating agencies (e.g., Moody’s) as well. There is also sustained growth in the retail green bond market with availability in Japan, New Zealand, South Africa, the UK, and the United States (mostly through municipal bonds). To complement the retail market, a number of technology players that leverage crowd-sourcing are providing platforms to fund green projects.

The green bond market started mostly in North America (the United States) and Europe but is now seeing the fastest growth in emerging markets, as demonstrated in Figure 6.3. Most notably, growth has predominantly occurred in the Asia-Pacific region. Within this region China is the largest player and has started to play an active role in regulation and reporting guidelines for green bonds.
Despite the green bond market’s tremendous growth, diversity, and geographically broad adoption, green bonds still represent roughly one percent of the global bond market size (Capepe 2017). While the sustained high growth rate suggests green bonds can become a larger fraction of the market, its relative size underscores its niche market status. This suggests the market will remain challenged by perceptions of unfamiliarity, exoticism, and additional (unknown or increased) risk. Thus, the challenge for the green bond market is twofold: (1) it must maintain or accelerate growth to erase the niche label, and (2) at the same time establish standards and regulations to guarantee its sense of maturity and trust among investors. Ultimately, these two challenges are interdependent. In order for the market to attain maturation, standards for risk management, structure, and other policies are key. Furthermore, such actions will be influential and likely to shape the market and its trajectory for decades. It is therefore key for policymakers to engage at this particular moment to craft a sustainable future of green bonds.

6.1.2 Labelled Green Bonds And Use-Of-Proceeds

With rising interest in sustainable and socially responsible investing inevitably comes the question of what measures ensure that financial instruments are what they claim to be. This is particularly true for the fixed income market given the growth rate of environmentally themed products (e.g., climate, green, social, and sustainability bonds). Transparency measures beyond traditional financial filings such as additional reporting and review can ensure bonds are used for their intended environmental causes, prevent greenwashing, and instill confidence in the marketplace. This is often met by having an independent party review, verify, and ultimately “label” the bond as legitimate. Bonds that fail to meet these criteria, opting instead to “self-label”
as green or climate bonds, likely do not provide adequate assurance for investors of their environmental benefits.

Labeled green bonds refer to those green bonds which have undergone independent review and certification to verify their environmental benefits. While labelled green bonds are a substantial portion of the broader climate-aligned (or climate-themed) bond universe, they only make up roughly one-third (200 billion USD) of its total (600 billion USD) size (see Figure 6.4). The data cited in Section 6.1.1 refer explicitly to labelled green bonds only. The green label can be applied to bonds of any form, to include privately placed bonds, securitized bonds, unsecuritized bonds, covered bonds\(^{85}\), medium-term notes\(^{86}\), and sukus\(^{87}\). Labelled green bonds can be issued by any institution, such as a multi-national, national, or private bank, a corporation, or a government.

![Figure 6.4: Labelled bonds by category. Green bonds remain the majority of labelled bonds. With market maturity, diversity is increasing (Climate Bonds Initiative 2018c). Source: Climate Bonds Initiative](image)

The International Capital Market Association created the voluntary Green Bond Principles (GBPs) from (ICMA 2019) to offer guidelines for issuers and investors in response to this need for transparency. The voluntary nature led to further standardization by the Climate Bonds Initiative, who developed the Climate Bonds Standard and Certification Scheme where assurance, second-party, or third-party verification ensures the bonds are consistent with the Paris Agreement objectives. The Climate Bonds Standards integrate the GBPs consist of a

85 “Covered bonds are highly-regulated securities with superior credit ratings. They achieve lower funding cost than unsecured debt thanks to a dual recourse structure whereby bond investors have a general claim against the issuer, as well as a claim over a dedicated ‘cover’ pool of assets.” (Climate Bonds Initiative 2018c)

86 “MTN structures are a valuable tool as they can be deployed by repeat issuers to facilitate access to the market and lower deal costs. They can be structured to allow for a variety of currency denominations and debt formats to increase their flexibility.”(Climate Bonds Initiative 2018c)

87 Sukuk abide by the Islamic Securities Guidelines; they are backed by a pool of assets that are Shari’ah compliant.
certification process, requirements both pre and post issuance, and guidance for sector eligibility (Climate Bonds Initiative 2014; “Overview: Climate Bonds Standard” 2015). The Climate Bonds Standard and Certification Scheme helps to label and prioritize investments and is used globally.

Along with this global framework and corresponding with the sector eligibility, Climate Bonds Initiative developed a taxonomy with key sectors: energy, transport, water, waste, nature-based assets, industry, buildings (see Appendix P for a depiction of the taxonomy). For each of the sectors, there is criteria derived from international experts in both academia and industry. This allows the bonds to be certified under both the Climate Bonds Standard and Sector Criteria. Bonds that are certified under both are used in this research.

To date, the majority of labelled green bonds are considered use-of-proceeds bonds, meaning the proceeds from the bonds support multiple green activities or projects and are backed by the issuer’s balance sheet. These projects are reviewed to ensure they have complied with the GBPs and other standards. With use-of-proceeds bonds, pricing is similar to ordinary vanilla bonds (i.e., flat pricing because the bonds share similar credit profiles), making them an attractive option for institutional investment houses who can sell and market them to a broad audience of investors. Thus, use-of-proceeds labelled green bonds are an important component of sustaining market growth.

6.2 Motivation And Research Questions

Analysts at the intersection of policy and finance note that the challenges faced by sustainable development cannot be addressed with government aid and philanthropy alone. Over the past decade, financial observers have noted a vigorous attempt to use private investors to help aid in the development of environmental sustainability. There has also been a general shift in sentiment among investors, many of whom reject the notion that return on investment and advancing social goods are mutually exclusive.

Broad opinion is that green bonds are playing an important role in reducing the climate financing gap and spurring private sector investment. It follows that the gap could be further diminished by successfully identifying new areas or products of interest to private investors. As investor demand has shown consistent growth, it is likely that demand will persist as the climate and ESG markets expand and mature into different products and asset classes. There are likely other investment vehicles, structures (e.g., securitization), and specific regions or industries where investors are willing to engage. Climate finance facilitators would benefit from identifying these specific areas where investor appetite is high.

88 Appendix Q discusses other types of green bonds.
In traditional financial markets, textual data like news and social media are being used to inform investment decisions. Previous studies demonstrate the link between social media sentiment (i.e., positive or negative polarity automatically identified in text) of stocks of companies and trading volume and returns (Deng et al. 2018). In certain active online message boards, social media sentiment is able to track market returns as well as volatility (Oliveira, Cortez, and Areal 2017). Given the market’s trajectory, an increase is anticipated in both discussion of green bonds and positive sentiment (i.e., favorable opinions and positive speech terms associated with green bonds)\(^89\).

This research seeks to identify trends and relationships between conversation around investment areas (i.e., semi-supervised topics) within the green bond corpus and the issuance size (i.e., investment dollars). This research is exploratory in nature, and aims to answer the following questions:

1. From a methodological perspective, can topics be anchored such that they represent underlying green bond investment sectors?
2. How do the strengths of the conversational pulse in green bond investment sectors (i.e., the historical topic composition) compare relative to one another and do they correspond to underlying investment sizes?
3. Are there sectors that have faced underinvestment and appear to represent an opportunity for increased engagement or new product development?
4. Are there projects supported through green bonds that can be shifted to attract institutional investors?

Answers to these questions will contribute to the ongoing discussion around the ability to leverage private investment to decrease the gap between climate finance availability and climate finance requirement. Additionally, identifying areas where private interest is lacking could signal a need for public financing, market restructuring, consolidation, or public policy solutions.

This research represents a new application of semi-supervised topic modeling and sentiment analysis in the social media sphere. If sentiment is used to identify favorable products and predict stock market volume, applying this methodology at the intersection of these two domains could yield interesting and policy-relevant results. This is especially true in a market where investors seek innovative financial products and the public sector cannot afford to tackle global problems without private sector support. It is expected that increased, sustained article volume and positive sentiment corresponds to strong sector interest. This could serve as a signal\(^89\).

\(^89\) Future research could focus on the predictive nature and evaluating leading indicators.
that private sector interest in investment is increasing, and the market could likely expand with more issuances or more varied financial products.

6.3 Green And Climate Bond Data

In the previous chapter, a global pulse was taken on the conversation surrounding climate finance. In this chapter, a localized pulse is taken on specific parts of the conversation around green bonds. This is done by exploring the temporal trends of topics related to green and climate bond areas in news or social media to understand if green bond investment areas have historically been successful in adapting to investor interests. Although the research in this paper examines annual issuance volume from the green bond program rather than daily trading volume, this is more appropriate for a relatively niche market.

The research in this chapter makes use of two data sources: (1) the green bond sub-corpus described in Chapter 3 and (2) total annual issuance dollars for labelled green bonds with sector classification from the Climate Bonds Initiative.

6.3.1 Summary Of Green Bond Article Data

The Green and Climate Bond article data are extracted from LexisNexis and processed as outlined in Chapter 3. The contents of the articles are post-processed and used to fit the anchored topic model using CorEx. Figure 6.5 presents the monthly article counts and smoothed trend-line from January 2010 to October 2018. The increasing trend is not surprising the market did not see diversification in bond issuances until roughly 2013.

Figure 6.5: Aggregated monthly counts for news and blog data from Green Bond sub-corpus with smoothed trend line. Uses consolidated keyphrases to extract the data from the LexisNexis database.
Figure 6.6 presents the same data aggregated by year. When comparing the annualized article counts, the trends between investment data (shown in Section 6.1) and the article volume are similar. The strong uptick in 2014 is expected as this follows the IFC’s issuance, along with the first corporate, municipal, and ABS bond issuances. The 2018 data lacks the last quarter. Given the monthly volumes of previous years and the strong contribution of article volume generated by the annual conference of parties in the last quarter, we would expect the 2018 volume to exceed that of 2017.

6.3.2 Labelled Green Bond Investment Dollars

To obtain estimates of labelled green bond investment dollars, this research used annual aggregates by sectors from the Climate Bond Initiative. Data beginning in 2014 is used because the Climate Bonds Standard taxonomy was not developed until after the tracking began in 2009 and has been updated over time. These aggregates include labelled green bonds that meet screening requirements as described in their database methodology (Climate Bonds Initiative 2018a). Per Climate Bonds Initiative, these aggregates include “bonds with at least 95 percent use-of-proceeds financing or refinancing green/environmental projects - social bonds are not included. [The] bonds which are broadly aligned with the Climate Bonds Taxonomy.” (Climate Bonds Initiative 2019c)
The annual aggregate dollar values across all sectors per year are displayed in the left panel of Figure 6.7. The trend is similar to that of the article counts for the corresponding years displayed in the right panel of Figure 6.7. However, it does appear that article counts are a leading indicator for investment dollars, as the former rose sharply three years before the latter. Note that 2018 is excluded in this comparison because the year’s article count does not include the last quarter and the year’s dollar values use an estimation method inconsistent with that of the previous years.

![Figure 6.7: Aggregate annual values from 2013 to 2017. Left Panel - Dollars in USD Billions, across all sectors, for labelled green bonds from Climate Bonds Initiative. Right Panel - Article volume for Green Bond sub-corpus from LexisNexis extract.](image)

In Figure 6.8, the aggregate annual dollar values by sector for labelled green bonds are displayed. Due to data availability, the relatively new internet communications and broadband sector were omitted and the industry and building sectors were combined. This data was obtained through calculation of the percent breakdowns by sector from each year and the total dollar value for the years. It is unsurprising that the energy sector garners the largest percent of assets; renewable energy has experienced dramatic price declines and is a key focus of national plans to meet the Paris Agreement reduction targets.
Examples of sector projects are provided for clarification:

- Energy: Solar, wind grid, bioenergy, and geothermal hydropower
- Transport: Low emission vehicles, electric vehicles, rail, and bus rapid transit
- Water: Sustainable water management, water infrastructure, and storm adaptation
- Waste-Pollution: Water and air pollution, recycling, and methane reduction
- Land-Use: Agriculture, forestry, food supply, wood, and paper
- Industry and Buildings: low-carbon buildings, industrial facility, and processing (e.g., cement)

6.4 Modeling Approach

The CorEx topic model (discussed in Chapter 4) is able to identify topics of interest from the green bond sub-corpus (discussed in Chapter 3). In Chapter 5 a global pulse on the conversation around climate finance was taken by identifying dominant topics through a probabilistic, entirely unsupervised learning approach with LDA. In this chapter, a localized pulse is taken on pre-identified (considerably niche) topics using a semi-supervised learning approach—localizing the conversational pulse on specific ideas through an information-theoretic approach.
Given the aforementioned green and climate bond sub-corpus, the approach tries to construct topics corresponding to underlying bond investment areas (see Figure 6.9). The CorEx model does this by anchoring the topics with terms that represent the sector investment area\(^{90}\). Trends that reflect global interest emerge when the volume of documents corresponding to the topics (i.e., investment areas) are analyzed alongside sentiment and investment dollars.

![Figure 6.9: Illustrative example of the green bond sector investment areas that represent the latent variables, topics.](image)

Like LDA, the goal of the topic model is to infer latent variables (i.e., the topics), where the topics are a collection of related words and documents are collections of topics. Details of the CorEx model can be found in Chapter 4, while Table 6.1 summarizes the differences between the topic models used in this research.

<table>
<thead>
<tr>
<th>Global Pulse on Dominant Climate Finance Topics</th>
<th>Local Pulse on Niche Green Bond Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA - Chapter 5</td>
<td>CorEx - Chapter 6</td>
</tr>
<tr>
<td>Unsupervised approach</td>
<td>Semi-supervised approach</td>
</tr>
<tr>
<td>Identifies topics without human input</td>
<td>Identifies topics with anchor words</td>
</tr>
<tr>
<td>Often identifies dominant topics</td>
<td>Can isolate less prominent topics</td>
</tr>
<tr>
<td>Every document is a mixture of all topics</td>
<td>Every document may not be associated with all topics</td>
</tr>
<tr>
<td>Every topic is a mixture of words</td>
<td>Every topic contains related words</td>
</tr>
<tr>
<td>LDA is a mathematical method for estimating both word and topic mixtures simultaneously</td>
<td>CorEx estimates the likelihood a document belongs to a topic given the words in the document</td>
</tr>
<tr>
<td>Treats data as though observations arise from a probabilistic process</td>
<td>Injects topics with domain knowledge to guide towards topics of importance to the user</td>
</tr>
</tbody>
</table>

**Table 6.1:** Highlighted differences between topic modeling approaches used on the climate finance sub-corpus and the green and climate bond sub-corpus.

\(^{90}\) The sector topics represent a subset of all topics identified, which total 31.
CorEx topic models from two sets of topic anchor words are compared: single words that describe the sector and sets of words representing the sector. For the set of anchor words, synonyms are also added to the list of anchors. The words selected for each green bond sector as anchors are then confirmed to exist in the corpus vocabulary and are lemmatized such that they correspond to the words. The original words used as anchors are shown in Figure 6.10, with the single words on the left and set of words on the right.

![Figure 6.10: Green bond sector-specific anchor words used for CorEx](image)

In regard to the corpus, the TF-IDF is used with stop-words removed and terms lemmatized. Thresholds were set for the minimum number of articles in which a word must appear to ten and the saturation across articles to a maximum of 50 percent. Topic coherence was compared using the TF and TF-IDF with superior performance from the TF-IDF.

Average monthly sentiment across the entire corpus and by sector is evaluated in a manner similar to that described Chapter 5. The Bing metric is used to compare the temporal trends because it appears representative. Measuring sentiment is outlined in Sector 4.6, and the metrics compared use pre-trained dictionaries which contain positive or negative connotations represented in the text. In addition, the NRC emotions (S. M. Mohammad and Turney n.d., 2010, 2013) are evaluated over all time and by year for the entire green bond corpus.

---

91 The Information Communications and Technology—ICT—is used as an input but CorEx is not able to identify a topic with a collection of terms that describe this sector.
6.5 Results

Given the bottom-up approach from the 189 countries submitting national commitments for the Paris Agreement, there is an opportunity to use green bonds or fixed-income instruments to help achieve the intended goals. Examining labelled green bonds using the evolution of sector topic volume and sentiment represents one lens to view potential opportunities. By combining these insights with sectoral and geographic experts as well as financial market analysts and economists, policymakers can better gauge the likelihood of success for green finance initiatives, enhancing the business environment and getting closer to sustainable development goals create business and markets for green and climate business. This section presents the results from the CorEx topic model, sentiment, and NRC emotion categories for the green bond corpus.

Sentiment for the entire green bond corpus has trended positive across the entire timespan. As displayed in Figure 6.11, the positivity has been increasing since 2012 with the greatest rate between 2014 and 2016. This trend of positive sentiment is likely to continue into the future, judging from rising investment levels and related trends like climate finance. However, if green bonds become plagued by greenwashing, liquidity issues, transaction cost increases, poor impact results, or negative press, then this could introduce additional negative sentiment and reverse this trend.

![Green Bond Average Sentiment By Metrics](image)

**Figure 6.11:** LOESS smoothed average monthly sentiment across metrics (Syuzhet, Afinn, Bing, NRC) for the entire green bond corpus.
The most pervasive NRC emotion categories observed across the corpus are Positive and Trust. The NRC emotion category rank orders are consistent year over year as shown in Appendix R. Confirming the positive sentiment from the metrics in Figure 6.11, Figure 6.12 shows the article counts associated with each category. While the positive and trust categories are the most pervasive, the negative and anticipation categories follow. Indicating the green bond market, for all its benefits, carries with it some negative emotion. It is possible that these do not relate to the instruments themselves but rather the perils of climate change or and the risk of acting too slowly. They may also point to the risks and barriers in the market that prevent greater confidence and participation. Future research could identify representative text strings that represent these emotion categories, helping policymakers to further understand what is generally behind these emotions.

Figure 6.12: NRC emotion categories for the entire green bond corpus in years 2009 to 2018.

Topics for six green bond sectors are identified using CorEx. The CorEx topic model with the highest intra-cluster homogeneity identifies a total of 31 topics - the focus here is on the subset of sector-specific topics. The subset of topics corresponding to the investment sectors are as follows: energy, water, industry-buildings, transport, land-use, and waste-pollution. One topic is identified for both the industry and building sectors. The topics most representative of the green bond sectors used the set of word descriptors as anchors rather than a single word. This implies that differentiation between the two may be minimal at least when popularly discussed. Figure 6.13 plots the LOESS smoothed volume of articles (news and blogs) associated with the sector topics over time.
Energy is the most discussed sector while water is the least. All sector trends are similar, rising after the IFC issuance. Most peak between 2016 and 2017 following the conclusion and ratification of the Paris Agreement, as demonstrated in Figure 6.13. The former observation was anticipated given the rapid maturation of decentralized renewable energy and energy storage technologies. In climate negotiations, much emphasis has been placed on encouraging renewables and phasing out coal. Indeed, energy as a topic accelerates during the run up to the Paris agreement. However, the diminutive status of what many experts anticipated to be a higher-profile topic in the environmental community was surprising. It is possible that the monetary challenges surrounding the sector, which is typically far less lucrative than energy and electricity firms, undercuts its interest and presence in the corpora. This is in spite of the fact that a new type of bond—the water bond—was created and the volume representing interest seems lower than expected. However, water related discussions sometimes overlap with discussion on pollution control. If these two were combined the trend would be on par with the energy sector. Separately, the transport sector peaked early in 2016, likely due to investment around rail.

The energy sector may be oversaturated, and the transport sector may soon reach saturation. When comparing the total annual dollars issued in billions against annual sum or articles associated with the corresponding sector topic (Figure 6.14), investments in the energy
sector continue to rise while interest (measured through the count of articles) is declining. The transport sector (Figure 6.15) is not as dramatic, but significant decline in interest occurs after 2016 while investment dollars continue to steadily rise. This should be read by financial executives and policymakers with caution. It is possible that the market surrounding green energy and transport is resilient to its level of attention, perhaps in part thanks to maturing technology. But if not and market oversaturation were to occur, capital flight could severely undercut not just forthcoming issuances in that sector, but ripple through the green bond market as a whole.

Figure 6.14: Contemporaneous annual green bond issuance and article topic counts within the energy sector.

Figure 6.15: Contemporaneous annual green bond issuance and article topic counts within the transport sector.
**Sentiment across all sectors trends generally flat or positive.** Industry, water and transport increase the most over the eight-year time frame as shown in Figure 6.16. Reasons for this are uncertain; it is possibly a function of the overall market maturing and more investment opportunities becoming apparent in these sectors after years of less consideration. Interestingly, average sentiment across the energy sector tended the least positive and stayed the most neutral of all six sectors. It is unclear if this is a result of energy generally being perceived as an exacerbating factor for climate change. These again reinforce the hypotheses discussed above regarding the energy and water sectors, along with the same note of caution and possible unexplored opportunity, respectively.

**Figure 6.16:** LOESS smoothed average monthly sentiment from articles corresponding to the green bond sectors using the Bing metric.

**Market opportunity may exist in the industry and building sector.** Figure 6.17 shows the industry and building sector alongside the land-use sector. The former has a substantially higher article count than the latter, yet both tend to raise relatively little money compared to the more prominent sectors like energy. This augurs an opportunity to identify new projects or expand issuance in the industry and building sector given the steady trends in interest and sentiment and relatively small total annual issuance. The article count for the land-use sector appears to be declining, suggesting that something should be done to address interest in the topic or have the topic’s certification reconsidered.
The water and waste and pollution sectors face different challenges, presenting a consolidation opportunity. The water sector receives a relatively steady but small amount of annual issuance and sees relatively few articles each year. Contrarily, the waste and pollution sector maintains a relatively high but declining article volume and has only begun to experience a noticeable amount of issuance. For both sectors, the article volume to investment dollars are less than other sectors. It is unclear what factors drive these trends in article volume; even high-profile news stories like the Flint Michigan water crisis did not appear to influence the article count. When considering the sector certification prioritization water pollution of byproducts from pollution that affect water could be prioritized. Secondly, ensuring the criteria is overarching may accelerate investments.
6.6 Policy Considerations And Recommendations

Several policy recommendations are provided in light of the conclusions in this chapter and the previous literature review. The results indicate an opportunity to identify underserved and emerging areas early as well as to capitalize on the sustained and increasing interest and positive sentiment in this market. At a high-level the recommendations involve engagement in the public and private sectors with cooperation across the two. From the public side the goals are to increase issuance, increase the investor base, and focus on the market structure. On the private side the goals are to increase liquidity, maintain competitive risk-return profiles, and standardize definitions, disclosures, and reporting across the industry. Any sector certification should be strategic and prioritized alongside expert opinion.

Focus development project sourcing in the areas of interest. When prioritizing the development of new sectors or criteria for specific types of projects (e.g., hydropower in energy) the interest could be considered in the prioritization. In this research the focus is on the specific sectors associated with green bonds, but one could imagine expanding the focus to securitized bonds or indices that support specific endeavors such as water management. If social media sentiment can be identified as a precursor to investor interest, especially for a financial instrument in its relative infancy, intermediaries (i.e., multi-lateral development agencies) and investment firms could use this information to determine if and when new products should be structured or released to the market. If successful, this analysis could serve to guide the
development of new ESG markets, such as water quality and sustainable forestry credits, or further refining green bonds to focus on specific topics of investor interest.

Focus on expanding the market and reducing barriers could help overcome the liquidity issue. By consolidating sector criteria and reducing barriers for project identification larger issuances might be more accessible. As shown in Figure 6.19, the deals across all issuance sizes are growing but notable in the 1 billion or more category. Greater liquidity can be provided through these larger deals which can in turn attract more investors (i.e., more private capital).

**Figure 6.19:** Green bond deals are growing in number (Climate Bonds Initiative 2018c). Source: Climate Bonds Initiative

Furthermore, the growth in the average size (Figure 6.20) is attributable to these larger issuances (Figure 6.19). However, the rise in the median size especially indicates larger deals are issued by more issuers. This is again positive for the growth of the market and corresponds to the trends noted earlier. For those regions accessing capital markets and private investors through smaller issuance, they may choose to concentrate the issuance in the sectors with opportunity.
Europe and Asia-Pacific are driving growth and reporting/verification. The regions with large issuance (Figure 6.20) and growth over time (Figure 6.21) have an opportunity to shape criteria development and market structuring in underserved sectors. More recently we see countries developing national green bond regulations – a key example being China.
Often there is a first mover advantage to establishing a market. If a country or region captures a large enough market share and develops the reporting and verification standards, other regions are likely to follow instead of developing their own and encouraging others to follow their lead. This phenomenon carries geopolitical implications as well, as such actions may move the market structure towards favorable or unfavorable conditions depending on one’s geopolitical preference. Accepted standards that do not follow the long-standing tradition of transparency and verification common in international development would inevitably be a step backwards for climate finance and undermine its sustainability. Issuance by country is driven by USA, China, France, Germany, and Netherlands (Figure 6.22); given the national size of these respective markets along with their maturity, standards and guidelines probably will likely be reinforced by national policies and not in general alignment.

Figure 6.22: The largest markets tend to have higher levels of reporting (Climate Bonds Initiative 2018c). Source: Climate Bonds Initiative

Advance private engagement by addressing common concerns. As investor interest in sustainable goals is suggested to be increasing, based on the analysis in this chapter, broadening the use of instruments familiar to retail investors and portfolio managers is beneficial. Broad investor bases are Green bonds and related instruments like social bonds and water bonds, are well-suited to a broad investor audience; the instrument of a bond is familiar to many in the investment community and its operation and behavior is easy to understand, thus increasing the likelihood of uptake. The concern that green bonds still remain something of a niche market relative to the trillions of dollars traded in the global fixed income market, also known as a liquidity problem, still causes hesitation among many investors. Breaking this cycle - where investors are hesitant to invest because the market is not mature enough with adequate volume, short tenures, and larger issuances - presents a challenge that requires action on multiple levels.
Encourage further research into sustainable bond market interest and alignment. The research in this chapter clearly demonstrates that both interest and capital flows towards sustainable investments like green bonds is increasing. Complementarily, national and international organizations have made strong pushes for sustainability in their own realms. Nowhere is this more evident than in the United Nations establishment of the SDGs, which are inclusive of the targets facilitated by green bonds but more expansive. In the near-term, issuers and institutions might look to mine news data to discover what type or types of bond might attract sufficient attention in the retail market. Chapter 7 further discusses topic modeling methods and advancements that can assist in identifying emergent or trending topics. Policymakers should encourage such exploration, possibly through research grants or investment incentives. As many of these types of bonds track with or support the UN SDGs, a longer-term goal of the issuance market could be to market as many types of bonds that track with each topic listed in the SDGs. Future research might also focus on identifying hotspots of geographic interest given both rising interest from the general investment community and an increase in nations deploying sovereign wealth funds and other government-back institutions to issue, invest in, and otherwise make use of green bonds. For instance, US government-backed mortgage institution Fannie Mae is now the largest issue incorporating such bonds into renovation loans to projects that qualify as green or sustainable improvements. Before governments take action, identifying where regional interest and attention is high would be of strong value for investors.

Harmonize multiple bond frameworks. With multiple types of bonds addressing similar goals such as green, climate, sustainability and water bonds, investors are challenged. Secondly, nations are also developing criteria that may not necessarily extend or harmonize across borders. Standardizing the frameworks surrounding the definition, pre and post issuance criteria, reporting standards, and impact metrics of these bonds will go a long way in alleviating investor concerns. As mentioned before, aligning the bond sectors with more of the UN Sustainable Development Goals (SDGs) will be beneficial in a number of ways by likely: (1) making the investments more attractive, (2) increasing attention due to the support and framing provided by a preeminent international organization, and (3) serving as an effective way for the investment community to standardize efforts at funding multiple sustainability facets, like renewable energy, water, and gender empowerment. As mentioned in Chapter 2, a standard framework and guidelines, crossing nations, for sector criteria development and prioritization will aid in trust and transparency to avoid potential greenwashing criticisms. Lastly, a standard framework could facilitate the flow of capital from developed to developing countries, as project developers and financiers would better understand the requirements for certification.
**Manage transaction and issuance costs.** Furthermore, transaction and issuance costs will always be associated with reporting required to ensure standards compliance. However, multiple standards bodies and external review increase the potential for added transactional costs, particularly if certain markets or clients follow separate standards. Harmonized standards reduce this cost for projects, issuers, and investors. While reporting, third-party verification, and disclosure practices can prevent greenwashing, they can also increase the issuance costs. When expanding issuance in the sectors with increasing interest standard bodies should aim to increase the issuance and investor bases while balance the costs incurred.

**Augment provisions for climate bonds and related vehicles.** One key concern for investors is the long-term time horizon for green bond investments to mature (typically tens of years for payment return, and the environmental benefit potentially difficult to fiscally quantify), as compared to the many unique short-term risks projects may face. These include regulatory, where a new political regime opts to modify their priorities for climate and environmental ends, and political, such as civil upheaval and corruption. These risks are particularly acute for developing countries most in need of climate financing. The additional risk also discourages investment as the additional risk premium tends to increase price. This does however present an opportunity to construct a financial investment vehicle that incorporates political risk insurance for green bond projects. Using insurance to help pool the risk across multiple countries, sectors, and time horizons, the minimal price differential between green bonds and vanilla bonds can be maintained and ensure the former’s competitiveness.

**Engage institutional investors.** It is important to recognize that institutional investors such as pension funds, reinsurance companies, and sovereign wealth funds serve to benefit from the optimization of costs and balancing of risks and returns. It is these institutional investors who have the scale of capital to provide significant support to climate finance, even if political interest and momentum may be strongest with individual retail investors. If 85 percent of long-term climate finance is meant to come from private sources, as proposed in the Copenhagen Accord, institutional investors will need to be involved given their outsized presence in the global financial market. Private interests such as project developers and banks can further their outreach with such private financial actors. Within the public sector, policy discussions surrounding the appropriate methods to incorporate environmental and political externalities into financial risk management, which would influence investment behavior at all levels.

---

92 These risks compound with those bond investments traditionally face such as currency fluctuations and project failure, which could make the bond less valuable.
**Engage investors with higher risk capacity.** While the potential of the green bond market is vast, it is important to keep in mind that green bonds aren’t suitable for every type of environmental project, especially if institutional investors are the primary buyer of the debt. In general, green bonds are not ideal for supporting new and unproven technologies, due to higher risk of default. Moreover, institutional investors have demonstrated little interest in bonds with returns tied to the performance of an asset (asset-backed bonds) and direct investment, due to risk, illiquidity and issues relating to technical capacity. Investments into unproven markets are more suited to investors with higher risk tolerances, such as venture capitalists. Given the results in specific sectors like land-use, buildings, and transport.
7. Model Extensions And Future Work: Word Embeddings, Dynamic Topic Models, And Stance With Sentiment

The results in Chapters 5 and 6 summarize trends using two variations of topic models - LDA and CorEx. These two approaches, combined with the text extraction and text analysis, provide promising results in the climate finance policy context. However, recent breakthroughs in NLP and Natural Language Understanding (NLU), such as advances in text classification, transfer learning, machine translation, content understanding, text summarization, and text generation, have matured rapidly and should be considered in future work. Applying some of the temporal or evolutionary topic models (e.g., topic models over time) and using inputs that leverage a larger conversation to understand concept relationships (e.g., pre-trained word embeddings) could provide complementary ways of understanding a policy domain. These advances may also overcome some of the shortcomings associated with the approaches applied here. This chapter will describe how these methods might apply to other policy research areas (7.1), modeling advancements and alternative approaches to those applied in this paper (7.2), and how these advancements would apply specifically to further research in climate finance. (7.3).

7.1 Broad Applicability Of Text Summarization

The text summarization achieved through topic modeling, as applied in this research, has broad applicability to multifaceted policy topics. With appropriately curated textual data (corpora), the methods applied here likely have the capability to model changing trends and sentiments of any number of broad policy arenas. As discussed in this research, understanding trends around global issues can enrich discussions and perhaps enhance the probability of policy success. Such insights could help policymakers identify whether or not to advocate for certain agenda items given those topics’ resonance in the current environment. As a hypothetical example, policymakers following the topic of education might want to bide their time on presenting an overarching plan to enhance international competitiveness if the predominant discussion of the day centers around inequality of service. Alternatively, policymakers could use topic modeling insights to tailor messaging, national goals and objectives, or policy specifics in light of public sentiment and discussion. To carry forward the educational example, tailoring a message about enhancing science, technology, engineering, and mathematics (STEM) training as a means to address underlying systemic inequality by broadening the opportunities for high-value careers could have a greater chance at success than a campaign which ignored popular
concerns. Overall, one can imagine topic modeling as a means to serve a broad range of policy domains to improve insight generation and political strategy.

However, these applications are not without limitation. As with any approach, be it qualitative or quantitative in nature, policymakers and stakeholders will want to recognize the limitations. As deep learning and NLP techniques become open-sourced and widely used, understanding their limitations is particularly important. While it is of utmost importance for scientists and modelers to make design and modeling decisions with intellectual integrity, it is equally important that the policymakers benefiting from their research be able to understand the modeling implications, the strengths of automated approaches, and the drawbacks. Benefit can be derived through the cross-pollination of policymakers and researchers. Furthermore, success in integrating new approaches in a policy environment will require a policy analyst to recognize the nuances of what machine learning approaches can and cannot conclude and the role human input and subject matter experts will inevitably play. For example, the output from a single topic model can be sensitive to hyperparameters or seed word choice; these corporas might have intrinsic biases which resultantly skew the eventual results (Bolukbasi et al. 2016). By examining the results and assumptions, these choices can be validated or revoked with input from experts. The observations made by topic modeling should probably be corroborated with additional lines of analysis about public and investor sentiment, such as dialogue with experts, qualitative review, polling, and press and academic reports. As applied to this paper, the structuring and marketing of green bonds would require extensive subject matter expertise even in light of the insights herein. While these tools might offer suggestions, critical policy decisions probably will remain in the realm of human judgment, as analysis-derived automation of such tasks in the policymaking arena is unlikely to be a near-term evolution of the technology.

7.2 Analytic Extensions

Advances in NLU and NLP, combined with the ease of access to vast compute resources for non-specialists, has given rise to many opportunities for methodological extensions in a broad variety of fields, including policy and social science. This section will focus on some of the most salient extensions and modifications to topic modeling as applied to policy settings. Future research could include advancements in topic modeling, word relationships (embeddings), evaluation criteria, and topic coherence93.

---

93 The descriptions of the modifications in this section are at a high-level; for greater detail recent papers or textbooks should be consulted.
7.2.1 Embeddings As Inputs

The LDA and CorEx topic models used in this research are applied to corporas of modest size and manageable vocabularies allowed for expert oversight. In the bag-of-words (BoW) approach, the dimensions increase with each unique word, even if the unique word is a synonym or spelling variation. In the BoW context, the dimensionality and sparse features are handled by ensuring the terms appear in a number of articles above a certain threshold and by concatenating terms or using n-grams. If the corpus and associated vocabulary were to expand and modeling constraints were focused instead on training time and compute resources, reducing the dimensionality by predefining the vector space to a fixed dimension that represents every word or document could be valuable - this is known as word or document embedding (Yoshua Bengio, Ducharme, and Vincent 2001; Collobert and Weston 2008; Mikolov, Sutskever, et al. 2013; Mikolov, Chen, et al. 2013; Le and Mikolov 2014). Instead of high-dimensional term space, dense vectors in a lower-dimensional space are created using neural networks. These vectors encode semantic relationships that can make understanding intuitively similar words without distinct rules explaining their relationship, such as “cat”, “kitten”, “whisker”, and “meow”, possible. That means terms (as in word embedding models) or documents (as in document or paragraph embedding models) with similar meaning, regardless of exact word choice, will appear closer together.

Trade-offs in using counts, term frequencies, or weightings (e.g., TF-IDF) should be balanced with the limitations of embeddings. With a large enough corpus, word embeddings can capture the semantic and syntactic nature of terms. Importantly, increasing corpus size can always reduce the dimensionality and reduce the intensity of compute resource needs. When considering dimension reduction in policy settings, a drawback of embeddings can be the interpretability. In embedding models, words or documents are mapped to a continuous vector space and their similarities are then computed by comparing the difference between vectors. However, interpreting how those vectors are relevant to the policy topic at hand is not intuitive. In this paper, ascribing topic labels from the climate finance and green bond corporas was a feasible task with the domain knowledge and top terms within topics. Performing a similar assessment across vector space poses additional challenges.

Word embedding models also face unique language challenges. In the BoW approach, using n-grams or word concatenation can partially address polysemy (i.e., words or phrases that are spelled the same but have different meanings) or homonymy (i.e., words or phrases that have

---

94 Originally referred to as learning a distributed representation of words.
95 Two architectures exist for creating word embeddings to capture context: (1) the Common Bag-of-words (CBOW) and (2) the Skip-Gram (reverse process of CBOW). Skip-gram uses a target word as the input and is generally better for small-datasets and rare words (like this research) whereas CBOW uses the context words as the input and is generally preferred for speed or emphasis on frequent words.
both the same spelling and pronunciation with different meanings). While the sequence of words is informative, improving embedding models to handle polysemy and homonymy is challenging - this is an active field of research in computer science. Recent breakthroughs use deep-learning that use what is known as a “transformer” architecture\(^{96}\) to improve the performance of embedding models and address shortcomings (such as the polysemy and homonymy issues) (Vaswani et al. 2017).

Embeddings generally take either of two approaches: (1) the use of pre-constructed embeddings or (2) the construction of a domain-specific embedding model. The pre-trained embedding models (Collobert and Weston 2008) are often easy to use, but in the case of niche areas with specific vocabularies (e.g., in finance where the terms alpha and beta take on different meanings than in other fields) pre-trained models may not boost performance. Some common pre-trained models are word2vec - built by Google from 100 billion words from news data (Mikolov, Chen, et al. 2013; Mikolov, Sutskever, et al. 2013), GloVe (Global Vectors for Word Representation) - built by Stanford NLP using co-occurrence probabilities (Pennington, Socher, and Manning 2014), and fastText - built by Facebook and uses subwords (Joulin et al. 2016). More recent deep learning models with advanced architectures (as compared to the aforementioned) include BERT - built by Google and using the aforementioned transformer architecture and looks at inputs bidirectionally (Devlin et al. 2018), the GPT series of models developed by OpenAI which often seek to employ more parameters than competing models (T. B. Brown et al. 2020), and Grover - a model built by the Allen Institute initially to identify artificially generated content (Zellers et al. 2019). In future research, the quality of topics or predictive performance from using embeddings (both trained on the corpus and pretrained versions) could be evaluated and compared.

7.2.2 Embeddings With Topic Models

The word embeddings described above come from the research area of deep learning and neural networks, whereas the topic models used in this research have their foundation in Bayesian mathematics. These are two different research communities, yet the underlying idea of representing words mathematically is similar. Considering the probability of a word given a topic, the topic model looks like a word embedding model. Probabilistic topic models may be improved by combining them with deep learning embeddings mentioned earlier.

Following the word2vec model in 2013 (Mikolov, Chen, et al. 2013), an LDA variation called lda2vec was introduced in 2016 (Moody 2016). lda2vec trains both word and document

---

\(^{96}\) The transformer architecture is an alternative approach using neural networks that focuses entirely on the attention or importance of particular words relative to others in the phrase or document. The approach is also parallelizable and reduces training time (Vaswani et al. 2017).
vectors at the same time and both are used to predict an associated word; the topic representations in a document are built on word embeddings using both the word and context. For example, word2vec might produce similar words to “France” as “Germany”, “Denmark”, whereas with lda2vec the input of “France” and a document with the topic of food might produce words like “cheese”, “bread”, and “wine”. Word embeddings have also been proposed for use with correlated topics (Xun et al. 2017). In this climate finance research, the correlated topic model was compared to the LDA model using the perplexity measure. Despite the superior performance of the LDA model with the climate finance corpus, using embeddings in conjunction with the correlated topic model in the future might prove even better.

Traditional topic models, like LDA, provide human-interpretable topics. Interpretability of topics, while in some unique cases may not be relevant to the end user, was desired for the climate finance research presented here and is often crucial for policy making decisions. lda2vec also produces interpretable topic vectors and sparse document weight vectors. If prediction of topics or classification is the main goal lda2vec could be more appropriate. Given the topic models presented here, lda2vec is likely to improve prediction of topics in a new time period. In the case of shorter text inputs (like tweets and news headlines) where topic models can sometimes suffer due to limited term co-occurrences, the embeddings would serve as complementary information in generating the topics. Training lda2vec can be computationally intensive but the context word vectors are an added benefit and could be used in other models.

### 7.2.3 Dynamic Evolutionary Topic Models

Adding a temporal lens to topic modeling provides a valuable way to look at composition changes over time. This is called a dynamic or evolutionary topic model. In this research, the time dimension is incorporated by training the topic models over the entire time horizon and then using article time stamps to look at changes over time. In this case, the time dimension is not factored into the Bayesian framework so new topics that may have emerged during the time series are not identified. A dynamic topic model (DTM) allows the words most strongly associated with a given topic to vary over time. Alternative models that incorporate a time dimension either in discrete time increments over a continuous spectrum are available and might surface new topics in climate finance given that terms associated with events like the Paris Accords or natural disasters might ebb and flow over time.

In Blei and Lafferty’s landmark dynamic topic modeling paper (Blei and Lafferty 2006b), they demonstrate the usefulness of a temporal dimension by showing how fields of science can influence one another or when new fields of science (e.g., computational biology) emerge from collaborations between scientists. To find the emergence of these new fields, they “chain” topics

---

97 lda2vec is implemented in Python: https://lda2vec.readthedocs.io/en/latest/?badge=latest
and the respective topic distributions together to get sequential topic models segmented over time. The documents in each time slice, $t$, evolve based upon the topics at the previous time, $t-1$. As such, the topic distribution is estimated at each iteration and previous time-periods are predictive of future periods. The generative process is an extension of the logistic normal distribution. One shortcoming associated with this DTM is that it does not model correlation between topics. Nonetheless, the approach can uncover a topic’s lineage and evolution over time, as well as provide an understanding of how one area influences another. Such topic model variations could be informative for the policymaking community, especially given social concerns that overlap in multiple areas like environmental justice which originates from the domains of environmental protection and racial inequality.

Dynamic or evolutionary topic modeling is a growing field of research and many of the following variations could be considered as alternative models to those employed in this paper. In general, the model alternatives fall into two categories: (1) those that examine variation within topics and (2) those that look at variation between topics. While DTMs are effective at topic discovery and trend identification, all implementations are batch algorithms that scan the entire dataset before making model updates. Bhadury et al. (2016) propose a more scalable inference algorithm combining Gibbs sampling and stochastic gradient Langevin dynamics. Within this area of research, topic evolution can be examined as variation within topics or between topics; Blei and Lafferty use their model to discern how article topics published in the journal Science vary over time and how new topics are constructed.

Both continuous and discrete versions of DTMs are available. For example, a continuous dynamic topic model was developed by Wang, Blei, and Heckerman (2008, 2012). The model uses Brownian motion to model the topics through time-ordered documents. They applied the model to predict the timestamp of documents and to help overcome granular or short and consecutive time windows. Could be pertinent to corporas containing headlines or tweets. Another useful model is topics over time (ToT) (X. Wang and McCallum 2006), a non-markov continuous time model that incorporates time into the word co-occurrence pattern by using its continuous distribution over the time period. This distribution expands and contracts with the duration of the co-occurrence strength - a narrow time-distribution with a strong word co-occurrence over a short period and a wide time-distribution when the relationship endures. A non-parametric approach was also developed using discrete time-horizons called Dynamic Correlation Detection (He et al., 2009). In this work, the vocabulary size is assumed to be equal across documents. To map the words to a lower-dimension, represented by topics, a temporal prior is applied. This prior captures the correlations from the latent space where the topics are constructed. A Gaussian process is used to infer the relationships between documents as well as
between topic sets (topic correlations). The topics, their temporal behaviors, and their correlations with other topics provides an interesting additional facet.

These model variations can help to address different variations of policy questions. While this research considered topics that spanned the entire time window of the corpus, it is possible that some topics may have originated within that time span and others fell out of favor after a strong early presence. Adopting more evolution approaches for examining both climate change and investment trends might yield new appreciation for how topics like green bonds emerged from the financial world and converged with other sustainable investment topics. Similarly, dynamic modeling approaches on text corpora could be very useful for policy analysts looking to understand negotiation dynamics, following how topics, themes, and relationships between parties transform over time.

7.2.4 Other Topic Model Choices

There are a number of models that bridge characteristics of the static and dynamic topic modeling approaches. The Structural Topic Model (STM) can be considered as an extension of a CTM where it includes covariates at the document level. In STMs, the relationships between topic prevalence and the covariates can be identified (M. E. Roberts et al. 2014). For example, the influence or relative importance of individual topics over time can be evaluated by including a date. Like the ToT, the topics are fixed and do not change. Multiscale topic tomography is an extension to a DTM, which provides increased control of topic evolution at different time-steps and associates the word generation process with Poisson parameters (R. M. Nallapati et al. 2007).

Along with the aforementioned unsupervised learning approaches to topic modeling, the research community has also actively developed supervised learning approaches. For example, Mcauliffe and Blei developed a supervised learning approach to LDA (Mcauliffe and Blei 2008). Supervised learning models incorporate a response that is associated with each document. In this case, the documents and responses are jointly modeled. Latent topics are identified by optimizing parameters that best predict the response variables for future unlabeled documents. This is especially useful for online forums with ratings such as stars given to a movie or the category of a restaurant.

A key area of potential improvement in topic modeling approaches is the automation of hyperparameter selection, given their relevance to the outputs. Identifying the number of topics can be done by iterating over various metrics, as was done in this research, or taking an ensemble approach. As with unsupervised clustering, the topic cluster variance (inter, intra, or a combination) could be evaluated with the analogous measure in text such as coherence or perplexity (e.g., TC-W2V coherence measure from O'Callaghan et al. (2015)). Future work could
also focus on more efficient hyperparameter optimization and auto-identification of symmetric versus asymmetric priors in the topic models. None of the approaches should be considered better or worse, but rather more or less appropriate for a given research question, underlying data set, and objective.

7.2.5 Sentiment And Stance

In this research, pretrained sentiment classifiers were applied to representative documents corresponding to a topic. Future research could focus on creating a new sentiment measure using labelled data within the climate and ESG space; this measure could consist of an individual or ensemble classifier (Das and Chen 2007). Another option may be to use sentiment as a response in a supervised LDA setting.

In conjunction with the sentiment measure stance could be substituted or included jointly with sentiment. Stance provides a quantitative measure of the perception towards a target entity. Stance can take on meaning beyond the polarity identified in text sentiment (S. M. Mohammad 2020), yet sentiment does influence stance. Stance like sentiment generally requires a classifier and labelled training data or an expert set of rules; standard emotion labels do exist for training (e.g., SemEval 2016 Twitter stance and sentiment corpus (S. Mohammad et al. 2016)) and combined sentiment-stance research efforts could be replicated and incorporated alongside the topic modeling (S. M. Mohammad, Sobhani, and Kiritchenko 2016; Sobhani, Mohammad, and Kiritchenko 2016).

Sentiment and stance analytics have the potential to provide extensive utility to policy research questions, particularly as they relate to communication dynamics between parties and response to policy decisions. Applying these emergent tools to an appropriately scoped-policy question and news corpus presents computer science and policy researchers with a greenfield of research opportunities.

7.3 Additional Future Research

Reasonable next steps to further understanding on climate finance involve the incorporation of several of the analytic extensions discussed in the previous section. These include comparisons of various topic models, sentiment indicator construction, using stance measures, and further automation of the hyperparameter optimization.

Separately, new data sources could be incorporated, and effort could be made to understand, evaluate, and possibly account for any biases derived from the data extraction strategy. When constructing topic models, both the context of the data and the interpretation of the results are important considerations. The models herein rely upon the relationships in terms and documents. The context of terms, both individually and collectively, can vary based upon
many factors (e.g., the geographic distribution, the language used in the corpus, and how the corpus is curated). The methods used for data collection (i.e., the search strategy and data extraction) can also influence the interpretation of the topics. The interpretation and descriptions contained here are quantitatively grounded, but also nuanced; they are influenced by the researcher’s qualitative understanding of the domain, including policy context, geopolitical relationships, climate science and financial markets. A review of the results in light of these implicit biases might help understand where to add caveats to conclusions and investigate these nuances further.

Based upon the anticipated growth in socially responsible investments and the climate-related investment area, the private sector and investors will likely seek new investment products or products that focus investment on particular industries or regions. The methods and analyses of this research attempt to assess the feasibility of using news and social media (i.e., blogs) to understand the conversation and trending topics as well as identify opportunities. Ideally, the next step will involve using textual data sources, alone or alongside other data sources, to preemptively gauge investor interest in a novel ESG product like the green bond or focusing on a specific sector like wastewater or geographic region like the Caribbean. In the future, this research could extend into time-series analysis and predictive analytics to predict bond volume or pricing. Some specific research questions may be:

- Can future investment allocations from green bonds (i.e., use of proceeds) be predicted using metrics like the intensity of a topic’s presence, sentiment about an investment area, and past investments?
- Can potential market uptake of green bonds and related financial instruments within a region be forecasted with topic models?
- Is changing regional sentiment in climate finance or green bonds predictive of issuance?
- Can an understanding of private investor opinion help institutions create more successful products?
- Does a current political environment impact the success of climate finance?
- What is the effect of positive or negative sentiment towards a topic or entity preceding green bond issuance?
- Is changing regional sentiment in climate finance or green bonds predictive of issuance?
- Does topic sentiment and stance from public news sources improve predictive performance?
One of the remarkable trends is the recent surge in sovereign bond issuance (Table 7.1) that contains a geospatial component. This geographic information could be incorporated in numerous ways, such as by extracting a corpus from within a specific region, performing cross-country (or region) comparisons, or examining the composition of a topic by geography. Furthermore, using an anchoring model within a region or examining countries that emerge in a topic over time model could be informative. On the predictive analytics side, one could imagine using trends in sentiment and topics from countries that have issued sovereign green bonds to predict new countries where the environment (as measured through sentiment and topics) is receptive to issuance.

<table>
<thead>
<tr>
<th>Nation</th>
<th>Total Issued in USD</th>
<th>Year of First Issuance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>5.5 billion</td>
<td>2018</td>
</tr>
<tr>
<td>Fiji</td>
<td>49 million</td>
<td>2017</td>
</tr>
<tr>
<td>France</td>
<td>16.7 billion</td>
<td>2017</td>
</tr>
<tr>
<td>Indonesia</td>
<td>2 billion</td>
<td>2018</td>
</tr>
<tr>
<td>Ireland</td>
<td>3.5 billion</td>
<td>2018</td>
</tr>
<tr>
<td>Lithuania</td>
<td>24 million</td>
<td>2018</td>
</tr>
<tr>
<td>Nigeria</td>
<td>30 million</td>
<td>2017</td>
</tr>
<tr>
<td>Poland</td>
<td>2 billion</td>
<td>2016</td>
</tr>
<tr>
<td>Seychelles</td>
<td>15 million</td>
<td>2018</td>
</tr>
</tbody>
</table>

Table 7.1: Sovereign bond issuance by country as of 2018 (Climate Bonds Initiative 2018c). More countries intend to issue in the near term (e.g., Spain, the Netherlands, Hong Kong and Egypt).

Funding a nation’s commitments to the Paris Agreement can be assisted by incorporating green bonds into a country’s sovereign wealth fund and by doing this the country’s climate plan becomes more visible.

Lastly, extensions to other ESG products, new sovereign bond markets, the inclusion of political environments, or specific areas of focus within climate finance all present opportunities for future research. As previous sections highlight there is potential value for topic modeling in various policy analysis communities. In general, the specific methods and future work plans herein would be particularly relevant for policy domains where private sector investments may play an important role in public goods provision - especially pertinent in realms of health care, education, and the environment.
Taking a broad step back, there are strategic implications for policymakers to consider for economics and governance at large. This research demonstrates that far from being a niche market or transient trend, climate finance, and in particular the sale of green bonds, has been a high growth market for more than a decade with staying power. Both the topic and its common instruments like green bonds have evolved substantially from just several years ago. Both government and private sector approaches to public goods have evolved during the same time frame. Maturation is likely to continue as these products appear to resonate with investors and there are receptive audiences globally. Succinctly, the public and the marketplace are reacting favorably to the involvement of private actors in public good provision, as evidenced by their sustained and growing support. This can be observed from pronouncements like the Business Roundtable in 2019 announcing it should pursue more than shareholder returns (Gelles and Yaffe-Bellany 2019) or the UN Principles of Responsible investment receiving 500 news signatories in 2018 and 2019 (Rust 2019). This is an important point to recognize, and as previously discussed in the literature review, a point not immediately intuited by all economists and political scientists. Several important overarching conclusions become evident from this observation. This chapter explores several of these lessons learned, then looks at broad and general trends applicable across domains, and finally delves into specific lessons for the financial world and social impact investments.

8.1 The Manifestation Of Behavioral Economics

The trend of growth in climate finance serves an exemplar of behavioral economics at play, and the value of the approach to 21st century problems. In classical economic theory, investors are rational economic actors looking to maximize their returns on investment. Asset pricing and the choices made thereafter follow this framework. Seeking to gain the greatest return with the least amount of risk in the least amount of time with the limited amount of information they have, investors would choose to invest in profit-maximizing vehicles strictly on their risk and reward preferences. Personal preferences, much less ethics, morals, and values, would play virtually no role in the investment decisions of individuals, firms, and organizations.

And yet it has been demonstrated that consumers are willing to pay for value-aligned products. More firms view sustainability as sustainable business, and sustainable business does not need to imply increased cost or decreased quality and profits. Pension stakeholders are
demanding divestitures from fossil fuel companies (Mooney 2017). Energy firm shareholders are demanding carbon risk disclosures in quarterly shareholder reports (Cushman 2017). There has been a marked rise in climate finance investment from the individual to the international level (Padraig, Clark, and Meattle 2018). As this research and analysis concludes, public interest in climate finance is increasing and the public’s stance towards investments is growing more positive. Furthermore, in the green bond market, the investment in new types of instruments aligns with public discourse; new fixed income products are seeing life in the form of blue bonds (i.e., water focus) and climate resiliency, as well as new financial avenues in corporates, sovereign wealth, and municipalities.

These data points are counterintuitive to traditional economics, when the cost of fossil fuels remains competitive if not cheaper than renewable alternatives, the strongest climate-linked damages remain in the distant future, and the plausibility of drastic changes to carbon pricing from political actors in the near and medium term remain unlikely. Therefore, the motive for these economic decisions are not being made solely on account of investors’ best fiscal interests or because they have incomplete information, but rather on account of personal preferences and values. Namely, investors want or need to see positive environmental outcomes. This aligns with the increasing availability of ESG metrics and ESG incorporation investment decisions. Looking more broadly, actors are making economic decisions as the result of non-economic intentions and goals, at least as such goals have traditionally been measured via market valuation.

For economists, this demonstrates strong support for the works of authors like Daniel Kahneman and Amos Tversky, whose insights melded the work of human psychology and economic science, underscoring the importance of heuristics and biases in decision making under uncertainty (Shefrin and Statman 2003) Notably, Khaneman and Tversky noted that individuals tended to focus on singular information—namely stories and their details—as opposed to base rate information - such as statistical data. Applied to the fiscal realm, this entailed that people would often overestimate future returns of investments that had performed well vice versa. The observations herein suggest that these concepts can be explored even further in the realms of social preference. It is plausible that actors might seek socially responsible investments because they believe that the market has inherently underpriced them, and future valuations are likely to correct the error in price. Alternatively, it is also plausible that the investment is sought more for moral or ethical reasons, in seeking an “alternative bottom line” than return on investment. With the help of qualified psychological and economic experts, these important questions can be further explored to aid in the understanding of this nascent field.

For policymakers, these observations suggest that entirely new frameworks of incentive structures could be developed in the policy realm. Indeed, many policies have already sought to incorporate the lessons learned from behavioral economics. For example, it is well-established in
financial literature that structuring a 401(k) investment system for a business where employees are automatically enrolled to contribute a small percentage of their paycheck that increases gradually over years produces higher participation rates and greater retirement savings long-term when compared to a structure of employees must make the choices themselves. As average employee preferences for both short- and long-term capital—as well as the time spent to explore the 401(k) platform—are weak, most will opt to do nothing and accept the base case. In the parlance of behavioral economics, the system that requires one to opt-out has increased participation compared to a system that requires one to opt-in. Therefore, structuring the base case to maximize the greatest good is an important tool available to policymakers at all levels. (Richard Thaler And 2008).

This research and analysis suggests that the structuring of incentives has only begun to be adequately explored. If actors are often making fiscal decisions based on separate value structures, these can effectively be leveraged to achieve other outcomes. In some ways, this is intuitive and has been explored by humankind for centuries; advertising more often appeals to the emotional side rather than the rational. However, this has evolved in many ways that the public sector has yet to fully appreciate. To consider the 401(k)-platform example above, might participation rates change if employees judged that their investments could go towards causes they cared about as well as supporting their retirement? The trend of “gamification”—where the use of game-design elements like measurements of progress is applied in non-game scenarios—has been leveraged for behavioral change in arenas as diverse as financial wellness, shopper loyalty rewards programs, and personal fitness (Deterding et al. 2011). Might these approaches also help foster socially responsible ends?

Government policy could also look to leverage the lessons learned about human behavior and increasing interest in ESG, SRI, or values-based investing from these and other examples. For instance, the mobile application Forest is a tool to help improve mindfulness, attention, and moderation of screen time largely geared towards the environmentally conscious. The Forest app is an example of a tool that leverages non-economic preferences, in particular, the user base’s commitment to the particular public good of afforestation, to incentivize users to pursue and achieve personal goals. It is conceivable that public policy could leverage similar incentive structures to help achieve complementary policy goals as well. At a simple level, a government organization might offer to match funds or donations given to qualified non-profits to encourage such behavior by the citizenry. At a more complex level, a city government may offer to divest its city pension holdings of tobacco firms if a sufficient fraction of its populace attend stop-smoking rehabilitation sessions at local hospitals. Similarly, it could offer to allocate resources to parks and recreation should school enrollment or public transportation ridership numbers reach
certain thresholds. Leveraging the skillsets and resources of private sector partners could also help expand the opportunities that a public sector actor might employ.

While such examples are relatively crude measures and not without legal considerations before implementation, they serve to demonstrate the unexplored potential to appeal to the general public’s social and ethical preferences and not just their economic ones. As preferences are likely to vary in both magnitude and direction, it is also likely that some efforts will need to be more tailored than others; for example, a program that incentivizes public-private action on habitat restoration will likely appeal more to the populace that self-identifies as nature-lovers than those who do not. Some advocates may also criticize government actors for not taking social and ethical action outright, instead making it transactional. Still, innovatively energizing the public’s core interests in the provision of public goods by any means remains low-hanging fruit for policymakers.

8.2 The Persistence Of Corporate Social Responsibility

In the summer of 2019, press articles noted with surprise that almost 200 CEOs of major firms released a statement announcing that their firms should serve more than just owners and shareholders, and should consider entities like staff, the community, and the environment (assorted 2019; Gelles and Yaffe-Bellany 2019; The Economist 2019). These actions exemplify the current interest in the private sector addressing pressing public goods challenges of the moment like climate change. In part, this can be attributed to the failure of governments to reach political consensus or compromise; returning to classical economics framing - governments are becoming increasingly less perfect, allowing for more opportunities for private actors to fill. The recent press also notes the idea of CSR was more commonplace in the mid-20th century in America, when firms would often support retirement programs, communities, or basic science research for the greater good. Causal understanding of why these trends ebbed is uncertain. Some point to the rise of the Chicago School framework of a more libertarian approach to economics, while others argue the stagnation of the 1970s forced a reevaluation. The 2019 Business Roundtable Declaration is just one example that leads to the conclusion that CSR is having a resurgence.

Even if this moment is short lived, it should be put into appropriate context. CSR was never entirely dismissed as a concept for over several decades, and it would be improper to suspect it would disappear after its trending moment today. It is also useful to recognize that CSR’s fate is not identically aligned with challenges of governance and international cooperation. Climate finance evolved for more than a decade, and public-private partnerships strengthened over that same time frame. During moments of both strong international cooperation and crises of confidence, climate finance continued its growth.
Popular reaction to the current moment appears somewhat mixed. Some argue such structures undercut fiscal accountability and creative dynamism, which both risk undermining long-term prosperity (The Economist 2019). Others argue that the last several decades of industrial practices were anomalous from the long-term trend of more holistic thinking.

Looking ahead, it is likely that CSR will continue to play a role in the strategies of major firms. Popular sentiments probably will cause its prominence to wax and wane, as history has demonstrated in past decades, but as a concept it is almost certain to persist. As the ability to accumulate and analyze data about a target customer set, supply chains, and the world around us, it is probable that more firms seek to experiment with CSR strategies at a minimum as the barriers to entry diminish. Several decades after journals debated whether or not the concept should exist, this research and analysis suggests that CSR and its many manifestations are here to stay.

8.3 The Ethical Dilemma Of Defining Responsible Behaviors

The potential abundance of ethical metrics brings forth a potential dilemma—what arbiters will determine the appropriate ethics to measure and how? This challenge is separate from the “greenwashing” challenges, wherein a firm markets ethical behavior but accomplishes little in practice (described in further detail later), and speaks more fundamentally to what ethics private actors or public-private partnerships should seek to adopt. Oftentimes, CSR does not have metrics, and those that are emerging are those which have come from ESG, the measures which have been around the longest. While higher standards of environmental care and review may be a widely accepted net positive for a firm to pursue, stances on more socially divisive issues or those which may appear incongruous—for example promoting efforts to counter cyber-bullying and also promoting the strengthening of free speech protections—will prove more difficult to navigate. In the case of public actors, the challenge simply becomes one of policy—critical and fundamental questions of ethics can be settled through drafting regulations and law.

In the case of private actors and public-private partnerships, where efforts are meant to go beyond legal and regulatory minima to derive greater public and private good, a firm must consider both policy and shareholder ends. The problem intensifies even further for international firms, as CSR preferences are observed to vary in between nations as a result of unique national business systems (Chapple and Moon 2005). Without independent arbitration bodies which seek to set standards, firms will ultimately then define ethics and responsibility for themselves, and likewise mediate potential conflicts of goals on their own.

The experience of climate finance—in particular the mature market of green bonds—could provide a useful guide for public private partnerships to navigate this uncertain territory. As green bond demand grew rapidly, various certification mechanisms arose to allow for
granularity, continuity in assessment and verification and validation of environmental benefits. In the beginning, the main concern in the market acceleration and expansion of green bonds was transparency and standardization. Transparency makes reporting clear and ensures product standards (in the case of green bonds for example, that GHG emissions were reduced as a result of project implementation) are met. Early on in the green bond market, key players realized clear delineation from a central body or recognized group would need to determine or certify “green” or “climate-themed” standards in climate products. This delineation has led to a market that encompasses three categories of climate-related bonds: certified bonds, self-labeled bonds and aligned bonds\(^98\). Both transparency and standardization have helped scale the market. By removing additional barriers to the market through securitization or guarantees private investors should be even more willing to engage in this market.

Further standardization of certifications for socially aligned investment products is projected to improve market certainty and help the market expand even further (Ehlers and Packer 2017). This self-selected move towards standardization is likely to have resonance far beyond the green bond industry. Climate accounting also has a long history of tools such as life cycle assessments and integrated assessment modeling (Patel and Calvin 2019) that allows for projects and initiatives to be considered to their fullest extent. As mitigation practices may have implications for adaptation and vice versa, the lessons learned from such evaluations can be extended to broader considerations of the interplay of CSR goals. As climate finance has been a pioneer in CSR, so too is it likely to guide the development of standardizing its definition as interest in CSR continues to grow.

8.3.1 Towards Economic Modeling More Reflective Of Human Behavior

As this discussion demonstrates, investors’ moral and social preferences do matter in their decision making, but taking reliable measures outside of laboratory experiments to assess how such preferences will manifest is exceptionally difficult given the confluence of other variables involved (P. Cox, Brammer, and Millington 2004; Mackey, Mackey, and Barney 2007; Petersen and Vredenburg 2009). Nevertheless, such estimations are possible, and the data-mining opportunities presented by online digital platforms, whose advances in understanding continue to be leveraged effectively by online advertising firms, (Solomos et al. 2019; Liu-Thompkins 2019) probably will be an important source to consider for deriving investment preferences as our economic understanding of the world becomes more nuanced. With the growing interest in values-based investing and the nascent opportunity for analysis of how investors behave when

\(^{98}\) Certified bond issuances request and receive certification by standards setting bodies such as Climate Bonds Initiative. Self-labeled bond issuances declare their intention to be considered as a climate-aligned product, but no standards review has taken place and could be subject to the charge of greenwashing. Aligned bond issuances have neither declaration nor certification, but independent observers allow them to potentially qualify as climate aligned.
presented with moral and economic choices, the academic community will likely find substantial value in advancing their understanding of moral and economic choices under uncertainty. If adequate protection of personally identifiable information can be derived, the tools and applications that will likely come from the CSR and ESG community in the next several years should prove valuable to enhancing understanding and developing the next generation of economic models. The impact will only be amplified if the approaches used to understand aggregate preferences and trends can be combined with the individual-level data.

8.3.2 A Trend Towards Personalized Value-Based Investing And Its Implications

As the digital revolution has transformed many industries, the financial sector is no different - at the institutional and retail levels. Personalized investment has been a service many firms have offered in an automated fashion for almost two decades thanks to the aid of digital recordkeeping and analysis (Peters and Weiss 2003). Many investment firms already strive to personalize clients' portfolios. Investment and retirement platforms common in the US ask clients to answer questions such as their time horizon, diversification preferences, risk tolerance, and financial goals. These questions go beyond an individual's current earnings, savings, and target nest egg size, but rather seek to tailor investments more towards individual preferences and constraints. In fact, a common criticism of many platforms is that they do not provide enough personalization, and that they steer users to a handful of funds that are far too generic for their wants and needs. While there is evidence of interest in climate finance products like green bonds, many are available at the institutional level. Retail investors are able to access these through indexes or ESG screened investments.

Despite strong interest particularly from younger entrants into the market, investors seeking firms that embrace CSR are often left with a handful of vague choices labelled “socially responsible” (Huang 2016) or required to conduct research on their own. The confluence of available data, interest, and a successful exemplar in green bonds suggest this is a strong market opportunity for a firm with the right technology, social science, and financial expertise. While the industry still has several challenges to overcome, personalized value-based investments probably will become a more commonplace approach for the financial sector in the coming years.

8.4 Expansion Of Social Metrics

Values-based investing must diversify and strengthen its standards in order to evolve and become more mainstream in today’s financial markets. This section examines lessons learned from the modern and historical social investment movements, and in particular lessons from this
research and analysis, and considers possible near-term opportunities for the values-based investment industry as well as several implications for policymakers.

Though the financial sector has seen the greatest breakout successes from environmentally focused initiatives and products like green bonds, the environment is by no means the sole value on which investors at large wish to direct their funds. The approach, framing, and vehicles developed for climate finance and described herein are applicable to other social initiatives as well. Historically, socially responsible funds can be traced back decades and many initially focused on funds opposed to investing in vice industries. (Huang 2016) SRI initiatives that focus on, for example, eliminating conflict minerals from the supply chain or rehabilitating opioid-stricken communities could well attain a critical mass of investor interest in the next several years. In this research, the application of machine learning topic modeling approaches demonstrated the rising interest in specific green bond areas and climate finance, so too can these tools determine the “investment readiness” of other social initiatives. Private sector actors and public-private partnerships can monitor such trends to more effectively identify where investor interest is coming from, what particular niche they want the investment to serve, and what underlies the motive for investing. By expanding the suite of mature issues and themes on which an investor can base their moves, the investment pool will increase.

8.4.1 Strengthening Standards, Monitoring, And Verification

As described earlier, definitions surrounding ethics and responsibility can vary widely between individuals. The terms CSR, ESG, and SRI have become the most prominent within the industry, but the determination of what governance and responsibility specifically means to an organization can vary widely. As interest in the space rises, lack of clarity as to how to measure and identify what investment vehicles are or are not meeting social responsibility standards presents a potential to undermine market confidence and deflate the industry’s momentum. Strong leadership by organizations that put forward clear, transparent, and actionable standards for corporate social responsibility will eliminate the change for firms to receive certification while concurrently not abiding by the spirit of CSR, undermining trust in the endeavor and potentially destroying the CSR marketplace before it has a chance to grow.

Corporations will always have the opportunity to portray themselves as more responsible than their actual business practices in order to reap public relations benefits. As CSR and ESG become more commonplace with increasing public pressures, the incentive for firms to do so will only grow stronger. The environmental community has experienced this challenge for decades with “greenwashing” - when a firm presents a message that they are engaged in environmentally responsible behavior but their corporate practices do not match the rhetoric (Duber-Smith and Rubin 1988). The term was coined in a 1986 essay by Jay Westerveld who
noted that the hotel industry’s presentation that its towel conservation service helped them save
the environment did not match the many ways in which industry practices were detrimental to
the environment (Watson 2016). Greenwashing is often a judgment in the eye of the beholder,
some may judge certain corporations such as oil majors to be environmentally unsound on
account of the industries in which they participate (Peter Steele 2017). Without an authoritative
set of standards and arbiters to determine what practices accurately define tiers of ESG, the
problem will only get worse.

Without an agreed-upon clearinghouse to serve as a standards or certifying body for
tranches of socially responsible investment, organizations are likely to crowd each other out and
reduce consumer and investor confidence. Unclear or opaque standards also offer the opportunity
for firms with bad intentions to use the confusion to paper over their behaviors and still appear as
upstanding citizens. Forestry certification serves as an effective example. Following strong
interest in the 1992 Rio Earth Summit, two groups aiming to promote forest sustainability and
stewardship: The Forestry Sustainability Council (FSC) and the Sustainable Forestry Initiative
(SFI). The lack of differentiation between the two organizations, intensified by the fact that one
(SFI) was an industry group and the other (FSC) was led by scientific advocates, often caused
consumers to have low confidence in forest stewardship labeling. Consumers feared that one
label was engaged in “greenwashing”—even though environmental analysis did not necessarily
find SFI to be lacking in their standards—consumers were unable to distinguish between the two
(Fernholz 2011). The similarity between the two councils, and the lack of an arbiter to determine
how standards should be set, led to confusion and a lack of trust in the marketplace (Annealtor
2017). As such, the incentive to further strengthen forestry standards was diminished. Proactive
engagement on socially responsible investment standards should help to avoid such an outcome.

A public-private partnership may be ideal for accelerating a harmonization of standards.
By lending the authority of a national or international governing body to standards setting,
particularly with the cooperation of leading firms in the industry, the industry is more likely to
coalesce around one set of standards. As described earlier, a lack of agreed-upon definitions
about what constitutes climate finance has not just challenged consensus-building on accurate
measurements but hobbled trust and provided a headwind for growth. Guidelines from
authoritative bodies like the UNFCCC allow for considerable discretion in accounting
approaches (Weikmans and Roberts 2019). As the majority of climate financing tends to be spent
in the same country it was raised (Padraig, Clark, and Meattle 2018) and this research
demonstrates a strong interest in climate finance across several specific geographic regions, a
lack of transparency and verification raises the risk that funds are directed only in part towards
climate-related ends. However, no research has confirmed or denied this hypothesis. Regardless,
the persistent lack of clarity around definitions has left open the potential for greenwashing (Berensmann and Lindenberg 2016).

The experience of the green bond market and its relative success in self-regulation offer a promising example for the larger CSR industry to consider. Upon recognition that not all bonds were meeting the same standards for transparency and climate-related return, many in the industry feared the vehicles could be commandeered by ill-intentioned actors as a greenwashing effort just as interest was taking off. Several institutions such as Climate Bonds Initiative began to put forward standards on what constituted a green or climate bond, how a project could get certified, and how technical and industry working groups would further refine these definitions and address questions as the industry evolved. Critically, the organization ensured bond issuers, governments, and market actors all agreed to their proposal in order to encourage comprehensive stakeholder buy-in as opposed to a process that favored industry or advocacy groups. (Climate Bonds Initiative 2014) As a result, more than $30 billion in Climate Bonds Initiative certified climate bonds had been issued in 2019 as of September, and certification is available for projects as diverse as land conservation and energy efficiency upgrades. (Climate Bonds Initiative 2019b) Certified bond issuances are diverse both in region and in type of sector; some examples include the country of Chile, German auto-maker Porsche, and the New York Metropolitan Transit Authority (Climate Bonds Initiative 2019a). Non-aligned bonds mostly originate from China and are generally in decline (“China Definitions” 2016). In this regard, green bonds have effectively avoided the pitfalls of what some new markets with limited governance structures might experience.

Moves towards attempts to better define climate finance underscore the importance of strong standards and bodies to verify and enforce them. The Paris Agreement’s “enhanced transparency framework”—requires more reporting on financial support that has been provided, raised, or received and commissions a technical expert review of the information. Despite persisting gaps in definitions, the Paris Agreement’s decision text offers the potential for improvements in transparency and validation to strengthen a growing industry (Weikmans and Roberts 2019). It is work indicative of the outsized and positive role public policy actors can play in a space where private sector funding dominates.

A recent analysis of financial mechanisms to strengthen climate adaptation for urban environments by a major global financial center noted the importance of “soft” infrastructure in supporting such initiatives. The term soft infrastructure refers to legal and market features such as the strength and capacities of domestic markets, credit ratings agencies, regulations, enforcement, bureaucratic transparency, urban planning, and rule of law. Where this infrastructure is weak, the authors note, creditworthiness concerns become exacerbated and can critically undermine opportunities for areas like the Global South where a long history of
borrowing and repayments are uncommon. (Hindlian 2019) As these institutions encourage confidence in the market as a whole, institutions that promote monitoring, verification, and standard setting diminish the incentive for firms to produce disingenuous self-promotion, helping to enhance confidence in the market and increase investments.

Other CSR sectors are almost certain to face similar challenges in standardization, messaging, and uncertainty. The experiences and lessons learned from the climate finance space can help to avoid a worst-case scenario where unorganized marketplaces fail to accommodate rising interest that quickly subsides after it becomes disillusioned.

8.5 CSR As The Potential Preferred Avenue For Social Change

Rising interest in CSR both as an investment strategy and as a means to provide public goods could substantially alter social norms. Engaged citizens may look to CSR as a key pathway to facilitate desired social outcomes. The rationale is simple to imagine. If the public sector remains gridlocked and subject to regulatory capture, the private sector would be the ideal alternative. Indeed, private sector or public-private partnerships may often be able to circumnavigate bureaucratic, judicial, and legal challenges that might hobble the public sector. Furthermore, the private sector is often able to move faster, more efficiently, and respond to new changes and trends, which are likely to arise in a complex interconnected world.

This new dynamic of roles and responsibilities might resemble the social norms and roles of previous generations, where firms were often responsible for the social welfare of the communities they employed (assorted 2019; Gelles and Yaffe-Bellany 2019; The Economist 2019). The different social and environmental challenges faced in the new era, as well as different employment patterns and the complication of digital connectivity, mean that the manner in which firms might provide more CSR will play out unlike past eras.

Further leveraging of the private sector could help address some public good needs like the climate finance gap. However, risks will also become more manifest because private actors will often maintain their own profit-minded motivation. Indeed, if one of the concerns with public actors is that they are subject to regulatory capture, entrusting private actors with even more jurisdiction over public goods could cause their further erosion. As awareness of CSR (and associated provision of public goods through private enterprise) increases, considerations need to be made regarding the trade-offs that come with ceding potential power to private actors.

8.6 Concluding Remarks

This research considered trends and their implications to a complex policy landscape - financing the response to climate change. In this chapter, the extensions of these research
conclusions to the economic, ethical, human behavior, and policy domains were discussed. While this research addressed society’s response to climate change, focusing on the financing aspect (climate finance) and more specifically on labelled green bonds, it is not the only policy realm to which this type of data and approach can be applied. Many other problems have similar characteristics, similar challenges, and probably will uncover similar opportunities to inform policy.

As discussed in previous chapters, machine learning approaches were employed to understand a particular problem with beneficial public private partnerships. A suite of methods novel to policy research were applied to widely available unstructured data sources to take a global conversational pulse on climate finance and pulses on niche topics related to green and climate bonds. While appropriate caution should be taken in recognizing that these methods are exploratory in nature and cannot infer causality, their ability to assess topics, trends (temporal and geographic), and relationships is noteworthy. The results indicate public private partnerships were viable, enduring, and held opportunities for improvement in this policy domain. These conclusions help to provide direction, both for policy decisions and for methods to employ in later research, and present actionable insights.

Looking ahead, this research can serve as foundational for several future avenues of exploration. Additional data sources beyond news, blogs, and social media can be utilized and combined to seek unique context and insights depending upon the policy arena. As noted in Chapter 7, rapid development of new and complementary methods are available to explore the policy questions herein more deeply or from different angles. More importantly, the structure of this research can be used as a template for policy researchers focused on unrelated complex policy domains ranging from disinformation to public health. With appropriate cooperation between policy and computer science experts, opportunities for exploration abound.
Figure A.1: WEF Global Risks 2016 (Marsh & McLennan Companies 2016)
Figure A.2: WEF Global Risks 2018 (Marsh & McLennan Companies 2018)
Figure B.1: World’s greenhouse gas emissions from the World Resources Institute based upon the CAIT Climate Data Explorer (Johannes Friedrich 2017)
Figure B.2: Historical carbon dioxide emissions. Note that the predominant geographical sources have changed in recent decades, but cumulative emissions remain largely from the US and Western Europe (CDIAC 2017).
### Table C.1: The location, session, and conference start date for each Conference of the Parties from 2010 to 2018 (UNFCCC 2018).

<table>
<thead>
<tr>
<th>Location</th>
<th>Session</th>
<th>Conference Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancun, Mexico</td>
<td>COP 16</td>
<td>November 2010</td>
</tr>
<tr>
<td>Durban, South Africa</td>
<td>COP 17</td>
<td>November 2011</td>
</tr>
<tr>
<td>Doha, Qatar</td>
<td>COP 18</td>
<td>November 2012</td>
</tr>
<tr>
<td>Warsaw, Poland</td>
<td>COP 19</td>
<td>November 2013</td>
</tr>
<tr>
<td>Lima, Peru</td>
<td>COP 20</td>
<td>December 2014</td>
</tr>
<tr>
<td>Paris, France</td>
<td>COP 21</td>
<td>November 2015</td>
</tr>
<tr>
<td>Marrakech, Morocco</td>
<td>COP 22</td>
<td>November 2016</td>
</tr>
<tr>
<td>Bonn, Germany</td>
<td>COP 23</td>
<td>November 2017</td>
</tr>
<tr>
<td>Katowice, Poland</td>
<td>COP 24</td>
<td>December 2018</td>
</tr>
</tbody>
</table>
**D: Parties And Observers From The UNFCCC**

The UNFCCC categorizes countries into three main groups, Annex I, Annex II, and non-Annex I. The information below is directly from the UNFCCC (UNFCCC 2015a).

**Annex I** Parties include the industrialized countries that were members of the OECD (Organization for Economic Co-operation and Development) in 1992, plus countries with economies in transition (the EIT Parties), including the Russian Federation, the Baltic States, and several Central and Eastern European States.

![Annex I Countries](image)

**Figure D.1:** Annex I countries as of 2018.
**Annex II** Parties consist of the OECD members of Annex I, but not the EIT Parties. They are required to provide financial resources to enable developing countries to undertake emissions reduction activities under the Convention and to help them adapt to adverse effects of climate change. In addition, they have to "take all practicable steps" to promote the development and transfer of environmentally friendly technologies to EIT Parties and developing countries. Funding provided by Annex II Parties is channeled mostly through the Convention's financial mechanism.

**Non-Annex I** Parties are mostly developing countries. Certain groups of developing countries are recognized by the Convention as being especially vulnerable to the adverse impacts of climate change, including countries with low-lying coastal areas and those prone to desertification and drought. Others (such as countries that rely heavily on income from fossil fuel production and commerce) feel more vulnerable to the potential economic impacts of climate change response measures. The Convention emphasizes activities that promise to answer the special needs and concerns of these vulnerable countries, such as investment, insurance and technology transfer.
E: Growth In Green Bonds


Figure E.1: Evolution of green bonds according to a European Investment Bank presentation (Kreivi 2016) based upon data from JP Morgan.
F: World Bank Project Life Cycle

The following information is directly summarized from World Bank (Bank) documents (World Bank 2014).

Country Strategy and Project Identification: The World Bank works with a borrowing country's government and other stakeholders to determine how financial and other assistance can be designed to have the largest impact. After analytical work is conducted, the borrower and the Bank produce strategies and priorities for reducing poverty and improving living standards. Identified projects can range across the economic and social spectrum from infrastructure, to education, to health, to government financial management. The World Bank and the government agree on an initial project concept and its beneficiaries, and the Bank’s project team outlines the basic elements in a Project Concept Note. This document identifies proposed objectives, imminent risks, alternative scenarios, and a likely timetable for the project approval process. Two other Bank documents are generated during this phase. The Project Information Document contains useful public resources for tailoring bidding documents to the proposed project, and the publicly available Integrated Safeguards Data Sheet identifies key issues related to the Bank’s safeguard policies for environmental and social issues.

Project Preparation: The borrower government and its implementing agency or agencies are responsible for the project preparation phase, which can take several years to conduct feasibility studies and prepare engineering and technical designs, to name only a few of the work products required. The government contracts with consultants and other public sector companies for goods, works and services, if necessary, not only during this phase but also later in the project’s implementation phase. Beneficiaries and stakeholders are also consulted now to obtain their feedback and enlist their support for the project. Due to the amount of time, effort and resources involved, the full commitment of the government to the project is vital. The World Bank generally takes an advisory role and offers analysis and advice when requested, during this phase. However, the Bank does assess the relevant capacity of the implementing agencies at this point, in order to reach agreement with the borrower about arrangements for overall project management, such as the systems required for financial management, procurement, reporting, and monitoring and evaluation. Earlier screening by Bank staff may have determined that a proposed project could have environmental or social impacts that are included under the World Bank’s Safeguard Policies. If necessary, the borrower now prepares an Environmental Assessment
Report that analyzes the planned project’s likely environmental impact and describes steps to mitigate possible harm. In the event of major environmental issues in a country, the borrower’s Environmental Action Plan describes the problems, identifies the main causes, and formulates policies and concrete actions to deal with them. From a social point of view, various studies aimed at analyzing a project’s potentially adverse effects on the health, productive resources, economies, and cultures of indigenous peoples may be undertaken. An Indigenous Peoples Plan identifies the borrower’s planned interventions in indigenous areas that may be needed, with the objective of avoiding or lessening potential negative impacts on the people. These plans are integrated into the design of the project.

Project Appraisal: Appraisal gives stakeholders an opportunity to review the project design in detail and resolve any outstanding questions. The government and the World Bank review the work done during the identification and preparation phases and confirm the expected project outcomes, intended beneficiaries and evaluation tools for monitoring progress. Agreement is reached on the viability of all aspects of the project at this time. The Bank team confirms that all aspects of the project are consistent with all World Bank operations requirements and that the government has institutional arrangements in place to implement the project efficiently. All parties agree on a project timetable and on public disclosure of key documents and identify any unfinished business required for final Bank approval. The final steps are assessment of the project’s readiness for implementation and agreement on conditions for effectiveness (agreed upon actions prior to implementation). The Project Information Document is updated and released when the project is approved for funding.

Project Approval: Once all project details are negotiated and accepted by the government and the World Bank, the project team prepares the Project Appraisal Document (for investment lending) or the Program Document (for development policy lending), along with other financial and legal documents, for submission to the Bank’s Board of Executive Directors for consideration and approval. When funding approval is obtained, conditions for effectiveness are met, and the legal documents are accepted and signed, the implementation phase begins.

Project Implementation: The borrower government implements the development project with funds from the World Bank. With technical assistance and support from the Bank’s team, the implementing government agency prepares the specifications for the project and carries out all procurement of goods, works and services needed, as well as any
environmental and social impact mitigation set out in agreed upon plans. Financial management and procurement specialists on the Bank’s project team ensure that adequate fiduciary controls on the use of project funds are in place. All components at this phase are ready, but project delays and unexpected events can sometimes prompt the restructuring of project objectives. Once underway, the implementing government agency reports regularly on project activities. The government and the Bank also join forces and prepare a mid-term review of project progress. In addition, the World Bank’s Report on the Status of Projects in Execution, a brief summary of all Bank-funded projects active at the end of each fiscal year, is available to the public. As projects close during the fiscal year, they are removed from this report, since their individual Implementation Completion and Results Reports are publicly disclosed at that time. The project’s progress, outcomes and impact on beneficiaries are monitored by the government and the Bank throughout the implementation phase to obtain data to evaluate and measure the ultimate effectiveness of the operation and the project in terms of results.

*Project Completion*: When a project is completed and closed at the end of the loan disbursement period, a process that can take anywhere from one to ten years, the World Bank and the borrower government document the results achieved; the problems encountered; the lessons learned; and the knowledge gained from carrying out the project. A World Bank operations team compiles this information and data in an Implementation Completion and Results Report, using input from the implementing government agency, co-financiers, and other partners/stakeholders. The report describes and evaluates final project outcomes. The final outcomes are then compared to expected results. The information gained during this exercise is also often used to determine what additional government measures and capacity improvements are needed to sustain the benefits derived from the project. In addition, the evaluation team assesses how well the entire operation complied with the Bank’s operations policies and accounts for the use of Bank resources. The knowledge gained from this results measurement process is intended to benefit similar projects in the future.

*Evaluation*: The Bank’s Independent Evaluation Group assesses the performance of roughly one project out of four (about 70 projects a year), measuring outcomes against the original objectives, sustainability of results and institutional development impact. From time to time, IEG also produces Impact Evaluation Reports to assess the economic worth of projects and the long-term effects on people and the environment against an explicit counterfactual.
The model from Besley and Ghatak (2007), as referenced in Chapter 2, contains a public good and two private goods and is set up as follows: a public good, \( g \), a private good that is produced, \( x \), and another private good that is the numeraire\(^{99} \) (valued by both consumers and producers), the number of private good consumers, \( N \), and the utility, \( b > 0 \), received by each consumer from consumption of the private good. The consumers are of two types: (1) those who care about the value of the public good, taken together as \( n \leq N \), with a strictly concave and increasing valuation function of \( f(g) \) and (2) those who are indifferent about the public good’s value, which as a whole equal \( N-n \). Preferences for each individual, \( i=1\ldots N \), are represented through the utility function, \( V \):

\[
V_i(p,g) = b - p - \Upsilon f(g). \quad (G.1)
\]

where \( \Upsilon_i=1 \) for the \( n \) public-good-valuing consumers, \( \Upsilon_i=0 \) for the \( N-n \) indifferent consumers, and \( p \) is price. There is free entry, the private good is homogenous and the public good can be offered at heterogeneous levels. While the public good can be offered at different levels, each firm offers the public good at a single level, \( \theta \), where \( \theta \geq 0 \). The cost to produce the private good is then \( c + \alpha \theta \). Firms first choose \( p \) and \( \theta \) and are profit-maximizing.

Then the consumers, who are utility maximizing, move by picking from which firm to buy (or not). The consumer equilibrium solely involves shopping decisions, whereas the producer equilibrium is conditional on the consumers’ shopping decisions and the strategies of other firms and constitutes prices and firm strategies. The market equilibrium is characterized by two variable sets consisting of price, \( p \), and public good purchases, \( \theta \). The indifferent consumers take no value in the public good and seek the lowest price, thus their set is G.2:

\[
\begin{align*}
p^*_n &= c \quad (G.2a) \\
\theta^*_n &= 0 \quad (G.2b)
\end{align*}
\]

\(^{99}\) Numeraire goods allow for relatively good comparisons when a specific price is not relevant. Prices of all other goods are normalized by the numeraire good.
For the consumers who care about the public good, they take the public good and seek to pay in their fair share. Their set is G.3:

\[ p^*_c = c + \alpha \theta^*_c \quad (G.3a) \]
\[ f'(n \theta^*_c) = \alpha \quad (G.3b) \]

The consumers who do not care about the public good are not adversely impacted by its presence, but there is utility improvement for those that do care, thus a Pareto-improvement is generated in this CSR scenario (as opposed to only indifferent product production).
H: Twitter Related Terms

Figure H.1: Twitter hashtags related to #climatechange - based upon a sample.

Figure H.2: Twitter hashtags related to #greenbonds - based upon a sample.
Figure H.3: Twitter hashtags related to #ESG - based upon a sample.

Figure H.4: Twitter hashtags related to #climatefinance - based upon a sample.
I: Common Topic Modeling Implementations

The most extensive implementations of these methods are in the static topic model space. In the evolutionary space, implementations might occur in one language with a wrapper in another language.

**Python:** A popular package for topic modeling in Python is Gensim. Gensim is a free Python library that is aimed at automatic extraction of semantic topics from documents. The input of Gensim is a corpus of plain text documents. There are several algorithms in Gensim, including LSI, LDA, and Random Projections to discover semantic topics of documents. Once the semantic topics are discovered, the plain text documents can be queried for topical similarity against other documents [https://radimrehurek.com/gensim/](https://radimrehurek.com/gensim/) Another python library for LDA is also available here: [http://pythonhosted.org/lda/](http://pythonhosted.org/lda/)

**R:** The R package topicmodels currently provides an interface to the code for fitting the LDA, CTM, and STM models. The R package lda (Chang 2010) provides collapsed Gibbs sampling methods for LDA and related topic model variants, with the Gibbs sampler implemented in C. All models in package lda are fitted using Gibbs sampling for determining the posterior probability of the latent variables. Wrappers for the expectation-maximization (EM) algorithm are provided which build on this functionality for the E-step. Note that this implementation therefore differs in general from the estimation technique proposed in the original papers introducing these model variants, where the VEM algorithm is usually applied. The R package topicmodels currently provides an interface to the code for fitting an LDA model and a CTM with the VEM algorithm as implemented by Blei and co-authors and to the code for fitting an LDA topic model with Gibbs sampling written by Phan and co-authors. Package topicmodels builds on package tm (Feinerer, Hornik, and Meyer 2008; Feinerer 2011) which constitutes a framework for text mining applications within R. The STM package has a reputation of being fast and intuitive, and includes functions to choose the most appropriate number of topics [http://www.structuraltopicmodel.com](http://www.structuraltopicmodel.com). The LDA and CTM packages: ([https://cran.r-project.com/package=lda](https://cran.r-project.com/package=lda) and [https://cran.r-project.org/package=topicmodels](https://cran.r-project.org/package=topicmodels))

**Java:** MALLET is a Java-based package for natural language processing, including document classification, clustering, topic modeling, and other text mining applications. There are implementations of LDA and LDA variations (e.g., Hierarchical LDA in the MALLET topic modeling library and toolkit. The MALLET Topic Modeling library: [http://mallet.cs.umass.edu/topics.php](http://mallet.cs.umass.edu/topics.php)
**Scala:** Stanford TMT was written in the Scala language by the Stanford NLP group. It is designed to help social scientists or other researchers who wish to analyze voluminous textual material. The input of Stanford TMT can be text in Excel or other spreadsheets. There are several algorithms in TMT, including LDA, Labelled LDA, and Probabilistic LDA.

**C/C++:** David Blei’s Lab at Columbia University provides many freely available open-source packages for topic modeling. These open-source packages have been regularly released at GitHub and include the dynamic topic model in C language, a C implementation of variational [spell out EM]EM for LDA, an online variational Bayesian for LDA in the Python language, variational inference for collaborative topic models, a C++ implementation of HDP, online inference for HDP in the Python language, a C++ implementation of supervised LDA, hierarchical LDA, and a C implementation of the CTM. 
http://www.cs.columbia.edu/~blei/topicmodeling_software.html

**C++ and Python:** For Dynamic Topic Modeling, the implementation from Blei’s lab is in C++. However, a Gensim wrapper was also created making this also available in Python.
Figure J.1: Topic model selection with Gibbs sampling and variational expectation maximization (VEM). Both approaches are consistent in providing an optimized K value (i.e., number of topics) of 6.
Table K.1: Climate finance article counts ranked by geographic region. Note that Western Europe does not include the British Isles.
**L: Additional Topic Model Evaluations**

Figure L.1: Hyperparameter optimization using the CaoJuan and the Deveaud metrics and a fixed per-term topic distribution parameter. Variation is by number of topics and distribution parameter alpha. The optimal parameters are consistent across both metrics and concentrated around a $K$ value (i.e., number of topics) of 6.

**Figure L.2:** In-sample rank-ordering demonstrates consistent beta value or 0.01 and $K$ of 6 with alpha between 6 and 8.
Figure L.3: Selection of topic number and parameter values through gridsearch. The top two plots contain the normalized CaoJuan and Deveaud metrics by topic number (y-axis) using the optimal concentration parameters. The blue rectangle indicates the optimal topic number is consistently nine for both the CaoJuan and Deveaud metrics. The bottom two plots depict similar output using the default concentration parameters however, there is not consistency between the metrics - eight and six are options.
Figure L.4: Optimal versus base hyperparameter comparison across $K$ values. The metrics from the LDA model using the topic number (y-axis) and concentration parameters for the optimal parameters (circle in top panels for CaoJuan and triangle in bottom for Deveaud) and default parameters (square for CaoJuan and cross for Deveaud).

The top two plots where $K=6$ match Figure 5.7. With 6 topics, the optimal hyperparameters demonstrate improvement as the CaoJuan values are above the base and the Deveaud values are below the base. It is also the case for $K=9$, but there is consistency in the optimal $K$ across the optimal and base values when $K=6$. 
Figure L.5: Out-of-sample comparison using perplexity topic numbers between 4 and 11. Lower perplexity is preferred. Perplexity using the optimal parameters (green) is superior for $K = 5, 6, 7,$ and $8$. Compared to the CTM (blue) the LDA topic model (using base and optimal) have superior perplexity scores. The minimum occurs when $K$ is 7 or 8 - this does not alter the results much as shown from the clustering figures.

<table>
<thead>
<tr>
<th>Candidate Topic Number</th>
<th>Cao Metric</th>
<th>Deveaud Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Topics</td>
<td>10.425</td>
<td>10.425</td>
</tr>
<tr>
<td>5 Topics</td>
<td>6.275</td>
<td>6.275</td>
</tr>
<tr>
<td>6 Topics</td>
<td>6.275</td>
<td>6.275</td>
</tr>
<tr>
<td>7 Topics</td>
<td>4.719</td>
<td>4.719</td>
</tr>
<tr>
<td>8 Topics</td>
<td>4.719</td>
<td>4.719</td>
</tr>
<tr>
<td>9 Topics</td>
<td>8.869</td>
<td>8.869</td>
</tr>
<tr>
<td>10 Topics</td>
<td>7.313</td>
<td>7.313</td>
</tr>
<tr>
<td>11 Topics</td>
<td>6.794</td>
<td>6.794</td>
</tr>
</tbody>
</table>

Table L.1: Optimal alpha hyperparameter for given metrics and topic number. At both 5 and 6 topics, there is agreement across the two metrics.
Figure L.6: Topic model distributions using $K = 6$ with other candidate concentration values (alpha and beta, called eta in above figure). For each panel in the figure, the alpha and beta values correspond to the optimal values when $K = 4, 6, \text{ or } 9$. The last panel has values that correspond to the recommended values from the literature.
Figure M.1: Screenshot of the interactive visualization of the intertopic distance and top terms across the entire corpus. The topics are well separated across the two dimensions - this is desirable.
Figure M.2: Screenshot of the interactive visualization of the top terms for topic 1. The red bar is the frequency within the topic and blue bar frequency across the corpus. The fact that the red bar nearly covers the blue bar indicates these terms are well captured by this topic.
**N: Sentiment Metric Variations**

**Figure N.1:** LOESS smoothed sentiment analysis for climate finance topics using Syuzhet metric.

**Figure N.2:** LOESS smoothed sentiment analysis for climate finance topics using Afinn metric.
Figure N.3: LOESS smoothed sentiment analysis for climate finance topics using Bing metric.

Figure N.4: LOESS smoothed sentiment analysis for climate finance topics using NRC metric.
**Explanation of Bonds - A Fixed Income Instrument**

Bonds and the suite of fixed-income products are essentially the rights to receive payment on a loan. In the instance of bonds, it is a government (treasury and municipal bonds), corporation (corporate bonds), or similar entity that is seeking a loan in order to fund their planned operations. Like typical individual loans with which most individuals are familiar— car loans or home mortgages—the schedule of repayments for bonds are fixed when the contract is established. This leads to the term fixed-income. Similarities also exist in the origination and transfer of these contracts as well. Unlike individual loans, however, there are substantial differences in the corresponding roles entities play in the bond market, and these differences warrant further mention.

When an entity seeks to issue bonds, it will consult with an investment bank as to how the bond should be priced and structured. As traditional commercial banks review an individual’s history and need to determine a feasible mortgage, so too will the investment bank review the entity’s history and needs to set an interest rate, the terms of the bonds’ issuance called an indenture agreement, and ultimate amount of money (or principal) to be sought. In the case of World Bank Green Bonds, the World Bank is the issuer that consults with the manager, such as SEB; SEB acts as the broker providing input to the World Bank regarding potential demand from clients (e.g., pension funds) that will aid in setting the terms of the loan (i.e., Green Bond).

One of the more important terms to both the entity issuing the bonds and the eventual buyer of the bonds is the bond’s interest, called a coupon rate. This rate signifies how much interest the bond holder will earn over the life of the bond. The coupon rate is based upon the entity's likelihood of paying the bond back in full relative to other issuers, the duration of the bond, and other market conditions. Typically, the better credit rating an entity has, the lower the coupon rate of their bond will be as investors are willing to accept less of a risk premium payment along with their interest. Similarly, the longer the duration of the bond, the higher the coupon rate will often be as payment is deferred for a longer period of time (Kennon 2003).

Unlike a mortgage bank, the investment bank seldom intends to hold onto the bonds they structure for their clients. Instead, the bank will initiate (i.e. issue) the sale of the bonds through its own clearinghouse. This is called the primary market, as the first transaction takes place in this manner. Typically, large institutional investors such as pension funds and bond-holding mutual funds make up the entirety of primary market purchases, as this is the most cost-effective manner for all parties involved. Both the bank and bond issuer make a concerted effort to advertise the issuance to ensure sufficient
demand (Kennon 2003; Stanton 2000). The one exception in the bond market to the method of issuance listed above is treasury bonds, which are lent to cover national government debt. The United States Department of Treasury does not rely on any financial institution to issue Treasury Bonds, and instead issues them through their own auctions. These auctions are scheduled throughout the year and are highly publicized. Individual investors are invited to participate, though most individual investors elect to buy bonds non-competitively and accept the outcome of the auction. Bidders present the rates they will accept for the bonds, and the Treasury accepts all bids in order of most-favorable to least-favorable (i.e. lowest to highest coupon rate) until the quantity of outstanding debt is satisfied. The last bid with the highest rate is set for all bonds issued (Treasury Direct 2017). Similar procedures are followed by other foreign governments, like the United Kingdom Debt Management Office (UK Debt Management Office 2012). For the scope of this summary, it is succinct and accurate enough to state that currency reserve banks, also known as central banks, ensure that all outstanding national government debt is purchased at each auction.

Once a bond has been sold at a set price in the primary market, it may then be sold again by the bondholder at any price. These transactions typically occur through a financial broker or an investment bank’s exchange and are called the secondary market, as the transaction takes place after the primary market sale. Numerous factors may lead a bondholder to wish to trade the bond, such as changes to interest rates, changes in the entity who was lent the money, or changes in the bondholder’s needs. For example, if a bondholder bought a $1000 bond issued by Company X at 10 percent for one year, then found interest rates for similar companies were found to be at 5 percent for the same time horizon, the bondholder might wish to realize the new value of his bond immediately, and offer to sell it for $1048. Bonds are traded on what are called over the counter (OTC) markets – as opposed to regulated exchanges like the stock market – because of the substantial variation in every bond’s quality (e.g., duration, date of issuance, coupon rate, structure, tax benefits), making most bond trades significantly different from each other (“SEC” 2017). World Bank Green Bonds are rarely, if ever, traded on the secondary market and are generally bought and held until maturity; they are similar to municipal bonds in this way.
**P: Climate Bonds Taxonomy**

![Climate Bonds Taxonomy](image)

**Figure P.1:** The taxonomy developed by Climate Bonds Initiative and used globally (Climate Bonds Initiative 2018c). Source: Climate Bonds Initiative
### Q: Types Of Green Bonds

<table>
<thead>
<tr>
<th>Type</th>
<th>Proceeds</th>
<th>Debt recourse</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of Proceeds</td>
<td>Earmarked for green projects</td>
<td>To the issuer: same credit rating applies as issuer’s other bonds</td>
<td>EIB Climate Awareness Bond, backed by EIB</td>
</tr>
<tr>
<td>Use of Proceeds Revenue</td>
<td>Earmarked for green projects or their refinance</td>
<td>Revenue streams from the issuers through fees, taxes, etc. are collateral for debt</td>
<td>State of Hawaii, backed by a fee on state electricity bills</td>
</tr>
<tr>
<td>or ABS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project</td>
<td>Ring-fenced for the specific underlying green project(s)</td>
<td>Only to the project’s assets and balance sheet</td>
<td>Invenergy, backed by firm’s Campo Palomas wind farm</td>
</tr>
<tr>
<td>Securitization (ABS)</td>
<td>Earmarked for green projects or the refinance of green project portfolios</td>
<td>To a specified group of projects bundled together</td>
<td>Tesla Energy, backed by residential solar leases</td>
</tr>
<tr>
<td>Covered</td>
<td>Earmarked for eligible projects included in the covered pool</td>
<td>To the issuer; if the issue is unable to repay, then to the covered pool</td>
<td>Berlin Hyp, a German financial firm</td>
</tr>
<tr>
<td>Loan</td>
<td>Earmarked for eligible projects or secured on eligible assets</td>
<td>For unsecured loans, to the borrower(s); for secured loans, to the collateral, with some limited recourse to the borrower(s) possible</td>
<td>MEP Werke, a German solar firm, backed by residential solar lease receivables</td>
</tr>
<tr>
<td>Other Debt Instrument</td>
<td>Earmarked for eligible projects</td>
<td>varies</td>
<td>Convertible bonds, commercial paper, sukuk</td>
</tr>
</tbody>
</table>

Table Q.1: Types of green bonds and their key differences. Information comes directly from Climate Bonds Initiative (Climate Bonds Initiative 2018c).
R: Emotion Category For Green Bond Corpus By Year

NRC Emotion Categories For Year 2010

NRC Emotion Categories For Year 2011

NRC Emotion Categories For Year 2012
NRC Emotion Categories For Year 2013

NRC Emotion Categories For Year 2014

NRC Emotion Categories For Year 2015
Figure R.1: NRC emotion categories for the green bond corpus for each year. Order of emotions by prevalence does not change over time.


https://doi.org/10.1145/2213977.2213994.

https://doi.org/10.1016/0167-2681(95)00037-2.

https://doi.org/10.1109/FOCS.2012.49.

https://doi.org/10.1007/978-3-642-13657-3_43.


JBE 52 (1): 27–43. https://doi.org/10.1023/B:BUSI.0000033105.77051.9d.
http://amj.aom.org/content/16/2/312.short.


Hu, Yuening, Jordan Boyd-Graber, Brianna Satinoff, and Alison Smith. 2014. “Interactive Topic...


Montes, M. 2012. “Understanding the Long-Term Finance Needs of Developing Countries.” Bonn: UNFCCC.


Nikita, Murzintcev, and Maintainer Murzintcev Nikita. 2016. “Package 'ldatuning.'” cloud.r-


https://dl.acm.org/citation.cfm?id=77013.


———. 2014 "World Bank Project Cycle.”
