Undergraduate Economic Review

Volume 16 | Issue 1 Article 9

2019

Recessions Or Partisanship: What Explains Climate Skepticism in the U.S.?

Abhishek S. Sambatur
DePauw University, abhishek.sambatur@gmail.com

Recessions Or Partisanship: What Explains Climate Skepticism in the U.S.?

Abstract

This paper investigates the variations in public mood pertaining to climate skepticism and attempts to empirically assess whether economic recessions or partisanship help explain aggregate-level trends and movements across a 16-year time horizon. Public survey data from the iPoll and Gallup Organization were used to construct the Climate Change Skeptic Index (CCSI) that served as a proxy to capture public opinion trends in skepticism across the U.S. A two-part vector autoregressive model suggests that while economic recessions might be causally linked to climate skepticism, partisanship plays a more influential role in explaining it over time. The key result is that holding all included variables constant, anti-climate change statements by Republican Congresspersons made three quarters ago raise the CCSI by 0.17 percentage points on average in the current quarter.

Keywords

vector autoregression, climate change, climate skepticism, dyad ratios, recession, partisanship, political elite cues, iPoll+ Roper Datacenter, ordinary least squares

Cover Page Footnote

This senior economics thesis wouldn't be possible without the immense support and guidance from my advisor Dr. Humberto Barreto. His belief and continual encouragement to work harder kept me going for the past year and helped me overcome several setbacks. I'd also like to thank the economics department at DePauw University for the opportunity to pursue this independent thesis as an alternative to the senior seminar. Professor Salil Benegal at DePauw University provided advice and shared his impressive research papers on the topic of climate change and skeptical attitudes to motivate me and formalize my thesis proposal. Reverend Katherine Smanik was extremely helpful in providing the core dataset used in the regression analyses. Finally, I'd like to thank Dr. Jarrod Hunt, and my two friends, Kiara Goodwine and Shuto Araki for reading my thesis and providing useful feedback along the way.

Table of Contents

Back to Contents

Introduction

In 2006, U.S. District Judge Gladys Kessler took a year to produce a monumental, 1,652-page opinion piece. This opinion detailed the highly sophisticated strategies used by tobacco companies to deny the science behind the harmful effects of smoking. Agnotology is the study of culturally induced doubt or ignorance, particularly through the publication of misleading scientific data. Similar to the strategies used by the tobacco industry, the evolution of climate skepticism provides an intriguing example of agnotology. Over the past two decades, the phenomenon of climate skepticism represents the massive gap between the scientific community's consensus view on climate change and the U.S. public's divided opinion on climate change.

Discussions of public opinions on socio-economic, political, or other stimulating topics are usually heralded from three types of sources: a perspective that agrees with or argues for the subject in discussion, a perspective that disagrees with or argues against the subject, and a neutral perspective that assumes the unbiased stance. One way to contextualize climate skepticism at an aggregate level is by categorizing and summarizing American public opinions on climate change. This method assumes that there is such a thing as a "public mood" on climate skepticism that isn't static but can be dynamically influenced by other factors over time. Contemporary research in Political Science usually tends to focus on the effects of the American public's partisan values on climate skepticism while controlling for other factors. The dominating influence of partisanship to explain climate skepticism is so strongly backed that researchers have even tested it as a "moderating variable" to learn if it could overwhelm other explanatory factors (Egan & Mullin 2016, 216). Studies by Malka (2009), McCright & Dunlap (2011), Guber (2013), and Hamilton (2015) have shown that informational factors such as education, self-rated knowledge, and science comprehension are positively related to climate belief for Democrats (and liberals) and vice-versa for Republicans (and conservatives). Despite the plethora of research focused on partisanship, others have argued that economic recessions provide a unique, alternative perspective to understand climate skepticism. Scruggs and Benegal (2012) find convincing evidence that the onset of the Great Recession in 2008 was an important contributing factor to explain public opinion trends on skepticism. My curiosity to understand the influencers of climate skepticism and the debate between its prevailing explanatory factor, partisanship, and niche explanatory factor, recessions, sets up the key research question for this thesis.

An empirical research thesis can choose to go any number of ways to conceptualize the relationships between climate skepticism and the two key explanatory factors. I chose a multivariate timeseries method called vector autoregression analysis to unearth these interrelationships and find an answer to the research question: “Recessions or Partisanship: What explains climate skepticism in the U.S.?” Data for the explanatory variables is relatively easy to find with help from past research and well-developed institutional platforms like the FRED Economic Data. Computing aggregate-level climate skepticism is a bit more challenging and is constructed using a novel strategy with inspiration from Brulle et al. (2012) and the aid of the Dyad Ratios algorithm. I also used existing literature and made a few subjective decisions on what variables best represent economic recessions and partisanship. For recessions, I used variables that are often espoused in research and media to capture the declining state of the economy. For partisanship, I focused on “political elite cues” similar to Brulle et al. (2012) to understand how statements and voting patterns of Democratic and Republican Congresspersons could shift public opinion on climate change. Through the empirical analysis, I find that the effects of economic recessions on climate skepticism are not clearly discernible for the target period (2000 – 2015). The causality tests introduced later in the paper indicate that recessionary factors might be causally linked to skepticism, but do not provide enough evidence to make definitive claims without further explication and analysis. In terms of partisanship, Republican (Congressperson) statements and voting patterns supply consistent evidence to suggest causality and explain the variance in climate skepticism. Democratic (Congressperson) statements aren't significant at explaining trends in skepticism but their voting patterns demonstrate causal effects systematically and the regression model contributes to explaining the variance in climate skepticism.

The focus of this empirical thesis is to find an answer to the research question. But along the way, I've attempted to replicate and intuitively understand some of the sophisticated algorithms and data generating processes to justify using these techniques. These sections, that serve as a quasi-knowledgebase to demonstrate my learning and reference for curious readers, are *Section 3* and *Section 4.1* in the table of contents and can be skipped by readers only interested in the empirical analysis. The remainder of my paper is arranged as follows. Section 2 is a *Literature Review* of other research pertaining to climate skepticism. Section 3 contains the *Methodology Theory* that discusses the construction of the Climate Change Skeptic Index. Section 4 is the heart of the thesis comprising of the *Empirical Strategy* that is further broken down into (4.1) *DGP Overview* that describes the CCSI and replicates the vector autoregression, (4.2) *Applying the Model to the CCSI* which provides descriptive statistics and defines the key hypothesis tests, and (4.3) *Data Analysis* that highlights and interprets the empirical results. Section 5 consists of *Conclusion*, that summarizes my answer to the research question and explores further avenues for research, *Citations*, and the *Appendix* which hosts supplementary information.

Back to Contents

Literature Review

The business of doubt is not a contemporary issue. Humans have repeatedly used it to derive economic value and psychological pleasure in the markets and political economy. In fact, at a micro-level, good parenting involves introducing healthy skepticism into a child's mind and institutions of higher education often proclaim their ability to develop contrarians. Nevertheless, a phenomenon with the scope to create economic value comes with innate accountability and responsibility to its stakeholders. Big tobacco indulged in such a business of doubt in the late 1990s and found itself at the brink of collapse when their deceiving business model was eventually trumped by overwhelming scientific evidence. Similarly, individuals and stakeholders that witnessed the impending trade-offs mandated by climate change research in the 1990s, sought to pull out the same stops (sometimes, incredulously, using the same lawyers as the tobacco industry) and inject doubt in the public's mind. Powell (2011) asserts that this "anatomy of denial" isn't novel and humans have used such "rhetorical devices" since the time of the Greeks (Powell 2011, 170). One of the fifteen methods described by Powell (2011) pertains to this idea of "manufacturing doubt" in the common person. Equating climate denial to a civil trial, Powell states that "a defense attorney (climate denier) has to prove nothing – only sow enough doubt to weaken the prosecution's (climate scientist) case" (Powell 2011, 127).

The scientific evidence for climate change is plentiful and the potential long-term risks are well-documented. A sweeping threat like climate change is tough to ignore and finding effective and sustainable solutions should have become the norm since the famous Charney report, "Carbon Dioxide and Climate: A Scientific Assessment", caught the attention of top government officials, scientists, business professionals, and the public in 1979. Contrary to this expectation, global warming and climate change emerged as controversial economic and contentious political issues creating a deep divide about climate science amongst the American public. Egan and Mullin (2017) observes this divide statistically and states: "by 1997, concern (about global warming) had dropped sharply among Republicans compared to Democrats, the beginning of a gap between partisans that has widened over time and currently stands at more than 40 percentage points" (Egan & Mullin 2017, 217). McCright and Dunlap (2011) provides groundbreaking insights on the rise of partisanship and the subsequent polarization of climate change. They explain the growing polarization problem within climate change with the "party sorting" theory. The key idea propagated by the theory involves developing friction among party elites on a controversial issue which results in "party sorting", i.e. dividing the public and driving them to assume conflicting positions. McCright and Dunlap (2011) test this theory by using an empirical strategy comprising of a multivariate logistic regression model to examine Gallup polling data on climate change opinion from 2001 to 2010 for evidence on three distinct areas: the political divide on global warming beliefs and concern, the moderating effect of political orientation, and ideological and partisan polarization. Positioning political ideology and party identification as explanatory variables and controlling for demographics, temperature and nine other variables, McCright and Dunlap (2011) finds statistically significant and positive results to support their hypotheses that "self-identified liberals and Democrats are more likely to report beliefs about climate science consistent with the scientific consensus (hypothesis 1) and express personal concern about global warming (hypothesis 2)" (McCright & Dunlap 2011, 170). Furthermore, to study the moderating effect of political orientation, McCright and Dunlap (2011) utilizes interaction terms that combine party identification and ideology with educational attainment to verify the results of previous studies on this subject. The analysis is consistent with the theory once again as McCright and Dunlap (2011) asserts that "the effects of educational attainment and self-reported understanding on beliefs about climate science and personal concern about global warming are positive for liberals and Democrats, but are weaker or negative for conservatives and Republicans" (McCright & Dunlap 2011, 175). The concluding part of their analysis is dedicated to answering the question: "Has the polarization and ideological divide on climate change, tested in the previous hypotheses, grown larger among the public?". McCright and Dunlap (2011) relies on another interaction effect, this time between "political orientation x year" to gather insights on this question (175). The regression model used to test this specification finds a statistically significant result and McCright and Dunlap (2011) emphasizes that the polarization trend has grown consistently over time and state that differences in global warming belief diverged from an 18-point difference in 2001 to a 44-point difference in 2010 between liberals and conservatives.

The influence of partisan differences on climate change is not a problem unique to the US as a "meta-analysis of 25 polls and 171 studies in 156 countries showed that aligning with conservative party ideology consistently predicted climate change skepticism across political settings" (Egan & Mullin 2017, 216). But, since 1997, the consistently widening gap of climate change opinions between partisan groups in the US has provided researchers with an interesting phenomenon to consider. Brulle (2013) asserts a strong correlation between targeted foundation funding to proliferate climate skepticism. Furthermore, these conservative think tanks, trade associations, and foundations that form a part of the larger "climate change counter-movement (CCCM)" are used by Brulle (2013) in the final analysis that results in "140 foundations making 5,299 grants totalling $558 million to 91 (CCCM) organizations" over a period of 7 years from 2003-2010 (Brulle 2013, 684). Predominantly, media coverage and academic literature of the CCCM has been limited to a few key organizations and simplistic discussion of their activities. On the contrary, Brulle (2013) approaches this issue holistically and following a comprehensive definition of the climate change counter-movement, considers questions like, "How are these organizations financially maintained?" and "How do these organizations and their funders interact to form a social movement?" (Brulle 2013, 682). Sticking essentially to a consistent technique of 'following the money', Brulle (2013) uncovers several big CCCM donors such as the Donors Trust ($78.8 million) and Scaife Affiliated Foundations ($39.6 million) and recipients of these CCCM funds including well-known conservative think tanks like the American Enterprise Institute for Public Policy Research ($86.7 million) and Heritage Foundation ($76.4 million). Another interesting link is implied when Brulle (2013) highlights that the rise of Donors Trust/Capital and the subsequent decline of ExxonMobil and Koch coincides with targeted environmental campaigns criticizing Koch and Exxon by the Union of Concerned Scientists and Greenpeace. Literature on the climate change counter-movement is not scarce and researchers have deployed varying methods to study this occurrence. Jacques and Dunlap (2008) provides another fascinating approach by performing a quantitative analysis of the link between conservative think tanks (CTTs) and environmental skepticism. The analysis involves "141 environmentally skeptic books published between 1972 and 2005" and were chosen if they "denied or downplayed the seriousness of problems such as climate change" and eight other categories of environmental issues (Jacques et al., 2008, 358). On the other end of this analysis, CTTs were identified with the help of the Heritage Foundation's web portal that stores a database of other CTTs espousing similar conservative values and were filtered by using specific keywords to extract the ones focused on environmental issues and policy. Their findings show that of the 141 chosen books, "130 books (92.2%) have a clear link one or more CTTs – either via author affiliation (62 books) or because the book was published by a CTT (5 books) or both (63 books)" (Jacques et al., 2008, 360). Finally, they scan 50 CTT websites and find that 45 (90%) of them espouse environmentally skeptic values.

Academic research pertaining to climate skepticism might be disproportionately focused towards the politicization and partisan themes involved in this matter, but alternate theories suggest economic factors could have a substantial say in influencing public opinion on climate change. Scruggs and Benegal (2012) argues that the impact of the great recession on public opinion of climate change may have been overlooked; a quick glance at nationally recognized survey polls from Gallup and Pew indicate that agreement over whether there is "solid evidence of warming" declined from 77% in 2007 to 57% in October 2009 (Scruggs and Benegal, 2012, 2). Scruggs and Benegal (2012) attempts an aggregate level and an individual level analysis of public survey responses to examine the influence of weather, media, and economic indicators on climate skepticism. For the aggregate level analysis, survey responses from Pew, Gallup, and Stanford are pooled together as a proxy to measure public opinion and an Ordinary Least Squares (OLS) regression model is fitted to observe the influencers of climate skepticism. The aggregate level analysis showed evidence that weather and economic indicators, specifically unemployment rate, had a greater impact on public support for climate change than the media. Scruggs and Benegal (2012) finds that a 2.1 point increase in unemployment rate leads to a 4-percentage point decline in public support for climate while holding weather, media, and the consumer confidence index constant. The individual level analysis utilizes a binary response variable to gather evidence on the question: "Is there solid evidence that the Earth is warming?". The authors decide that a logistic regression model is a better estimator than an OLS with a binary response variable and include various demographic controls in the model. This model gathers insights on the extent of partisan influence and states that while climate change belief rates are lower for Republicans and fell from 60% in 2006 to 38% in 2010, the climate change belief rates didn't fare much better for Democrats during the same period and fell by 10 percent from 90%. Scruggs and Benegal (2012) attempts to raise the importance of an economic crisis and bring it amidst the climate change conversation without side-lining or ignoring the impact of partisanship, weather, media, and misinformation campaigns. Finally, further research and similar regression analyses on European countries show that public opinion of climate change in countries with low partisan differences can still be negatively impacted due to economic indicators.

Scruggs and Benegal (2012) shows that an aggregate level analysis of climate skepticism is possible, and researchers can empirically assess the societal mood towards the issue in a chosen year. Brulle et al. (2012) considers a longer time horizon to assess the changes in public mood on climate change. Here the authors build a "Climate Change Threat Index (CCTI)" to develop a macro-level measure of the American public's consensus on the threat attributed by climate change to their lives. Brulle et al. (2012) tests seven model specifications using a vector autoregression model (VAR) to assess their influence on the CCTI. The variables used to estimate the VAR fall into six broad categories: extreme weather events, scientific information, mass media coverage, media advocacy, political elite cues, and economic controls. Their research finds that public statements in support of climate change by Democrats, positive trends in GDP, and New York Times mentions of *An Inconvenient Truth* are the three strongest positive predictors of change in the CCTI. On the other hand, the level of anti-environmental Republican voting patterns and the unemployment level are the strongest negative predictors of changes in the CCTI (Brulle et al., 2012, 14). In the next section, I'll explore various empirical methods available to researchers interested in estimating and analyzing the aggregate level opinion trends on climate skepticism and substantiate the approach chosen to find answers to the salient questions posed in this paper.

Back to Contents

Methodology Review

In the literature review, we discussed different individual and aggregate level approaches to gather insights on climate skepticism in the US. Since this paper attempts to discern and find empirical results to answer the key question of whether recessions or partisanship influence the U.S. public mood on climate skepticism, an aggregate level model inspired by the approach used in Brulle et al. (2012) is the effective way moving forward. An individual level analysis isn't helpful since the data sample in such a technique pertains to individual respondents and hence does not facilitate specific observations related to the causal effects between macroeconomic factors and the skeptic attitudes of the US public. Scruggs and Benegal (2012) shows us results from both an aggregate-level approach and an individual-level analysis. The key difference in this instance pertains to the nature of results obtained as the aggregate-level approach allows the researcher to present a generalized argument on how media, weather, and economic conditions influence public opinion about global warming. On the other hand, the individual-level survey analysis provides empirical answers that allow researchers to make specific arguments on how the race, party affiliation, education, income and other key demographics affect an individual's likelihood of responding positively or negatively to the global warming question. Hence, an aggregate-level analysis channels the focus of an empirical study to consensus estimates that provide a framework to critically answer macro-level questions about climate skepticism in the US.

Though the aggregate-level approach is determined as the best way to proceed, the researcher now has several options to build a narrative that connects their theoretical claims to an empirical result. The main obstacle involves finding a consistent and reliable measure to define the climate skeptic attitudes and capture the altering public opinion trends. A potential source for such a measure comes from public polling institutions that have collected data on climate change for many years. Since their target audience (the public, politicians, journalists, etc.) usually cares about trending topics and stats of a shorter time horizon, most of the questions posed through these polls do not repeat and lack consistent question wording over time. Furthermore, there is considerable disagreement over whether the differences between a climate change 'skeptic' and a climate change 'denier' are significant. For this thesis, I've combined them and avoided differentiating between these sub-groups. Herein I run into the classic aggregation problem, a problem on how to obtain a proxy that best represents climate skepticism across the US without excluding relevant polling results and including irrelevant survey questions.

❖ What is the best proxy for climate skepticism in the US?

The rationale behind the shorter time horizon of most survey questions conducted by polling institutions was mentioned earlier. This is a pertinent problem because independent factors that could influence climate skepticism need to be analysed over time and contribute little if looked at through a cross-sectional study. For instance, a key agenda of the paper is to determine whether economic recessions are causally linked to heightened climate skepticism. To include a time period with at least two recessionary periods, the proxy needs to capture a full decade from early 2000s (the dot-com crash) to 2010 (which includes the financial crisis of 2008). In short, we are looking for a proxy that captures the level of climate skepticism by aggregating information from relevant survey questions spanning the two recessions of 2001 and 2008.

Simple or Weighted Average: The mean is perhaps the most common measure of center used in aggregation problems. Barreto and Howland (2006) highlights the pertinence of using the average as a linear estimator and states that when the "data generating process follows the classical econometric model (CEM), then the sample average is the best linear, unbiased estimator" (Barreto & Howland 2006, 346). This theorem, known commonly as the Gauss-Markov theorem, fails to validate the usage of the average for my purposes because the sampling process and the climate change polling questions do not fulfil the requirements of the CEM. Among several requirements of the CEM, a key one states that the error terms in the model are distributed independent and identical to one another. Since the chosen method involves a time-series analysis, this requirement for the error terms may not hold true and the conditions of a CEM are violated. The average may not qualify as the best unbiased estimator for my purported model due to discrepancies to the Gauss-Markov theorem, but I can still use it as a comparative measure to test the robustness of other unbiased estimators. The weighted average holds a slight advantage over the simple average in this model since it will proportionately assign weights to survey responses with larger sample sizes instead of treating them equally. The advantage arises because larger sample sizes result in smaller standard errors. But, the average runs into problems and becomes less reliable as an estimator when missing questions enter the time series model. The missing questions problem arises because most polling institutions ask questions based on issues that are of current importance and either change the wording of the question or discontinue the questions over subsequent periods. Even the same question is not asked every single year. For example, Gallup Polls posed the question: "Is the seriousness of global warming generally exaggerated, generally correct, generally underestimated?" in the years: 2000, 2004, 2010, 2013, 2015. Now, if the chosen timeline for the model extends from 2000 to 2015, the variable representing this question in the time series will have 10 missing values in its respective matrix. This example is not an exception as most survey questions measuring skepticism face a similar issue. The average struggles to perform consistently with missing values and hence might not be the best proxy for capturing the macro-level estimate of climate skepticism in the US.

Dyad Ratios Method: Stimson (2017) recognizes this shortcoming in using the average as an estimator of underlying mood based on survey data and proposes the dyad ratios algorithm as an alternative solution to better estimate the "latent structure" underlying a given dataset

(Stimson 2017, 5). The dyad ratios algorithm deals with the missing values dilemma in a creative way that involves building “dyads”, where all the survey results in a given timeseries are transformed into ratios reliant on other existing values and the resulting matrix is used to capture the latent structure of the dataset through a recursive estimation technique. The next section dives deeper into the dyad ratios algorithm and compares its timeseries output of the polling data used in this paper to the simple and weighted averages. Furthermore, I will briefly review the Item Response Theory alternative to the Dyad Ratios method and critically discuss their merits and shortcomings to justify the chosen approach for this paper.

Back to Contents

Stimson's Dyad Ratios Algorithm

At the onset of this paper, the psychology and prevalence of climate skepticism is discussed extensively leading to a theoretical case for the creation of an aggregate-level measure that captures the underlying public mood. The concept of public mood isn't abstract, researchers have used it time and again to create generalized dispositions of public opinion to predict the changes in their theorized model through external shocks. To clarify, phenomena like climate skepticism might have specific factors predicting outcomes in any given situation. But there is a subset of unknown, generic factors that underpin the public mood or latent structure of such issues. For instance, Brulle (2013) mentions the contributions of conservative think tanks and institutions to fund the counter climate change movement that is directly targeted at influencing climate skepticism in the US. On the other hand, the occurrence of an economic downturn is a general trend of a boom-and-bust economic cycle but may still instigate climate skepticism even if it isn't purported to influence the public mood on climate change.

❖ Why was the Dyad Ratios Algorithm created?

On the issue of estimating public mood, another estimation technique often used is called the Principal Components Analysis (PCA). It is a statistical tool used to compress a large variable set to a small set while preserving most of the original information. A mathematical procedure, using a square symmetric matrix, is conducted to transform a range of potentially correlated variables into a subset of uncorrelated variables which gives the analysis its name of principal components. But PCA is unreliable in estimating public mood on issues like climate skepticism because it requires a completed matrix for the mathematical procedure to work and create a consistent estimation. There is an abundance of data through public polls on climate skepticism, but the pain point arises due to the irregularity and inconsistency of the polls resulting in an incomplete matrix. The Dyad Ratios Algorithm was created as an alternative to the PCA and contends as a data extraction technique that can estimate a latent structure while attempting to gauge public opinion based on irregular survey data. Next, I take a detailed look at the logic of the dyad ratios algorithm by walking through the key steps of building dyads, implementing the recursive estimation procedure, using an iterative process for validity estimation and rounding it off with the bootstrapping of standard errors. Finally, this section will conclude with a discussion of the advantages and disadvantages of this approach and a quick glance at the Item Response Theory alternative proposed by McGann (2014).

❖ How does the Dyad Ratios algorithm work?

The algorithm begins with an assertion that changes in survey marginals from one period to the next indicates changes in the underlying public mood, assuming the chosen survey adequately captures the mood. In the case of climate skepticism, we have several reliable survey questions that have been administered by different polling institutions over time. A rigorous selection process gives us a subset of questions that can potentially capture the variance of public mood on climate skepticism on a time series. A crucial observation is the entry of missing survey marginals for questions that are either discontinued or modified over subsequent periods. To begin the construction of the Dyad Ratios algorithm, I have a subset of irregular questions that can estimate the latent structure, i.e. climate skepticism, over a chosen period. Next, I explore the implementation of dyads and lay out the foundation of the algorithm.

A) Dyad Ratios and Matrix Formation: To simplify the dyad ratios creation, the algorithm necessitates that all survey marginals are scored in the same direction, i.e. a higher number indicates a greater indication of the latent structure and vice-versa (Stimson 2017, 8). The ideal way to clarify the complexities of the dyad ratios algorithm is to create a small example that helps us understand the logic and various moving parts. The small example can be followed along using the "SmallExample-4x4.xls" file. In *Figure 1.1,* I have a 4x3 complete matrix with hypothetical survey question items x_i, x_j, and x_k that are administered in Time (T) 1 to 4. Starting with a complete matrix will help understand the dyad ratios estimation technique when missing values enter the equation.

Complete Matrix

	Question Items		
Time	x_i	x_j	x_k
1	30	40	50
2	40	50	60
3	45	55	65
4	50	60	70

Figure 1.1: Complete Matrix

A dyad can be defined as the value obtained from a ratio of a given item over any two time points. Stimson (2017) argues that using dyads to make relative comparisons among survey marginals of different question items has two advantages. First, the missing values scenario doesn't impact the usefulness of the dyads as they are relative measures built using known values of the item. Secondly, with missing values, descriptive statistics of variables like the mean become unreliable. But, the expected value for dyads of a given item across multiple periods is equal to 1 and this improves the consistency of the data used to estimate the latent structure. This relative nature of dyads results in an exponential growth every time a new question item or a time period is added to the matrix. Moreover, the recursive estimation technique of the algorithm uses forward and backward recursion that employs a given item's value at each time period twice (excluding the first and last values in the timeseries), the numerator and the denominator positions of the dyad ratio. Using the 4x3 example matrix, a deeper look into the recursive estimation process will clarify the ideas discussed above, provide a side-by-side comparison to contextualize the "best proxy of climate skepticism" debate in the previous section, and reveal why dyad ratios provide a better estimate of the latent structure than a simple average.

B) Recursive Estimation: As discussed, the 4x3 matrix results in multiple combinations of dyads and *Figure 1.2* illustrates all the potential ratio combinations using the forward and backward recursion processes for the items x_i, x_j, and x_k.

Forward recursion

	Ratios x_i		
x_i2 / x_i1	1.33	1.13	x_i3 / x_i2
x_i3 / x_i1	1.50	1.25	x_i4 / x_i2
x_i4 / x_i1	1.67	1.11	x_i4 / x_i3
	Ratios x_j		
x_j2 / x_j1	1.25	1.10	x_j3 / x_j2
x_j3 / x_j1	1.38	1.20	x_j4 / x_j2
x_j4 / x_j1	1.50	1.09	x_j4 / x_j3
	Ratios x_k		
x_k2 / x_k1	1.20	1.08	x_k3 / x_k2
x_k3 / x_k1	1.30	1.17	x_k4 / x_k2
x_k4 / x_k1	1.40	1.08	x_k4 / x_k3

Backward recursion

	Ratios x_i		
x_i1/x_i4	0.60	0.75	x_i1/x_i2
x_i2/x_i4	0.80	0.67	x_i1/x_i3
x_i3/x_i4	0.90	0.89	x_i2/x_i3
	Ratios x_j		
x_j1/x_j4	0.67	0.80	x_j1/x_j2
x_j2/x_j4	0.83	0.73	x_j1/x_j3
x_j3/x_j4	0.92	0.91	x_j2/x_j3
	Ratios x_k		
x_k1/x_k4	0.71	0.83	x_k1/x_k2
x_k2/x_k4	0.86	0.77	x_k1/x_k3
x_k3/x_k4	0.93	0.92	x_k2/x_k3

Figure 1.2: Forward and Backward Recursion Dyads for Complete Matrix

To understand the logic of the dyads, let's walkthrough a slightly altered matrix which includes missing values and is a better representation of real survey samples. Consider the following incomplete matrix in *Figure 1.3* below. There are 3 missing values, one for each item across different time periods. To reiterate, the recursive estimation process is a better solution to combine information across time than an average when missing values are involved.

Incomplete Matrix

	Questions Items		
Time	x_i	x_j	x_k
1	30	40	
2	40		60
3		55	65
4	50	60	70

Figure 1.3: Incomplete Matrix

Backward Recursion Process:

1. To build the dyad ratios using backward recursion, the first step involves starting at the final period (T4, in this case) for an item, x_i for instance, and building ratios relative to T4 for x_i at all preceding time periods (T3, T2, & T1). Next, we further this process by

moving one time period backwards ([T4]-1 = T3) and building ratios relative to T3 for x_i at all preceding time periods (T2 & T1). This process is repeated until we reach the first time period (T1) and there are no preceding time periods with item values in the dataset. *Illustration 1* below is a visual description of this dyad creation process for the backward recursion process. This gives a subset of ratios from the incomplete matrix which looks similar to the backward recursion table in *Figure 2*, but now I have 3 missing dyad values for each item induced by the missing values in the incomplete matrix. The backward recursion table in *Figure 1.4* illustrates the dyads that will be used to complete the rest of the recursive estimation process.

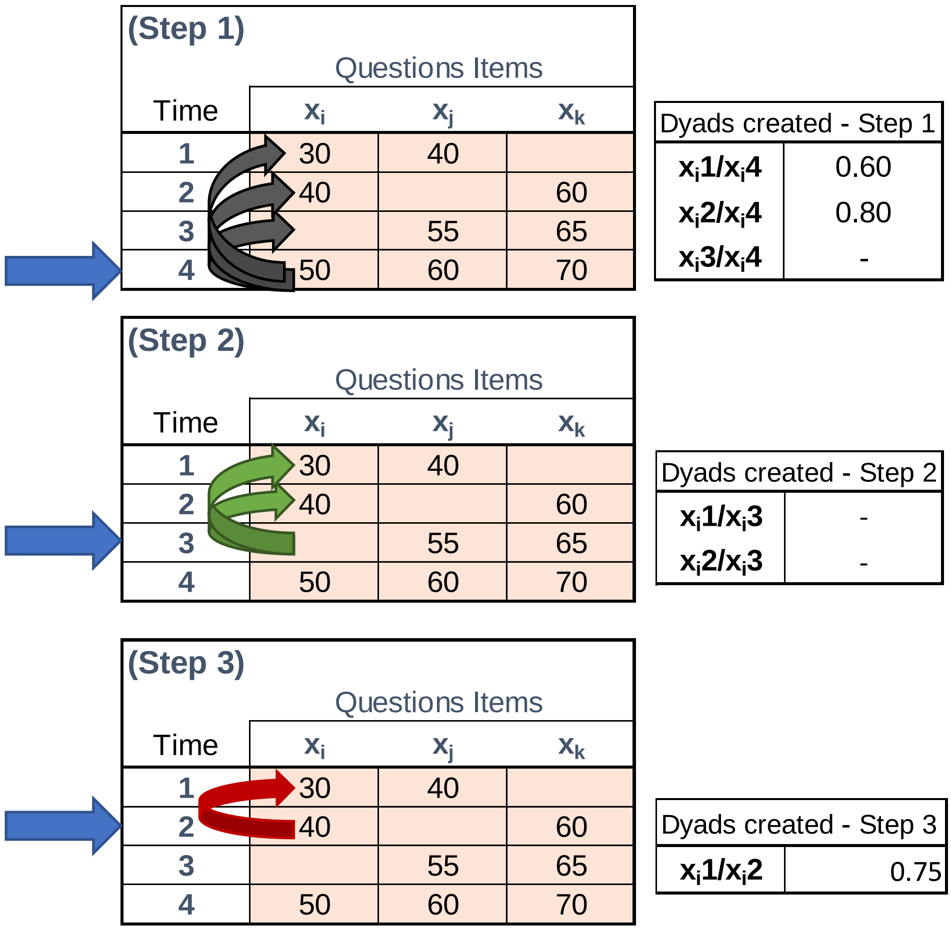


Illustration 1: Backward Recursion Dyads Walkthrough

Forward recursion

	Ratios x_i		
x_i2 / x_i1	1.33333		**x_i3 / x_i2**
x_i3 / x_i1		1.25	**x_i4 / x_i2**
x_i4 / x_i1	1.66667		**x_i4 / x_i3**
	Ratios x_j		
x_j2 / x_j1			**x_j3 / x_j2**
x_j3 / x_j1	1.375		**x_j4 / x_j2**
x_j4 / x_j1	1.5	1.09091	**x_j4 / x_j3**
	Ratios x_k		
x_k2 / x_k1		1.08333	**x_k3 / x_k2**
x_k3 / x_k1		1.16667	**x_k4 / x_k2**
x_k4 / x_k1		1.07692	**x_k4 / x_k3**

Backward recursion

	Ratios x_i		
x_i1/x_i4	0.60	0.75	**x_i1/x_i2**
x_i2/x_i4	0.80		**x_i1/x_i3**
x_i3/x_i4			**x_i2/x_i3**
	Ratios x_j		
x_j1/x_j4	0.67		**x_j1/x_j2**
x_j2/x_j4		0.73	**x_j1/x_j3**
x_j3/x_j4	0.92		**x_j2/x_j3**
	Ratios x_k		
x_k1/x_k4			**x_k1/x_k2**
x_k2/x_k4	0.86		**x_k1/x_k3**
x_k3/x_k4	0.93	0.92	**x_k2/x_k3**

Figure 1.4: Forward and Backward Recursion Dyads for Incomplete Matrix

2. To estimate the latent structure at each time period, I consider an arbitrary value of 100 for the final period T4 in our dataset. This is done while computing ratios in the previous step since I had no available information to equate the value of any item to a respective value in a time period after T4. Thus, I was only able to use that final time period (T4) in the denominator of the computed ratios, i.e. as a relative measure for other item values. Continuing the backward recursion, the next step is to estimate C_{T-1}, the latent structure at the penultimate period, which is T3 or the T4 – 1 period by using the absolute values for all items with existing values in T3 and in turn, existing dyad ratios for the T3 / T4 periods. A quick glance at the incomplete matrix in *Figure 1.3* shows that only items x_j and x_k have existing values for T3. Hence, C_{T-1} is estimated by averaging the dyad ratios for all existing item values with T3 / T4 dyads and projecting it based on the final period (T4) that was assigned the arbitrary value of 100. The calculation of C_{T-1} is shown in *Figure 1.5*, where I obtain the "Avg T4-3" value by averaging the dyads "x_j3/x_j4" and "x_k3/x_k4" and multiply this average with the arbitrary value of 100 to obtain C_{T-1}. At the

end of this step, I have two values estimating the latent structure, C_T which is the arbitrary value of 100 and C_{T-1} which is the "data determined value reflecting the true ratio of T (T4) and T-1 (T3) estimated from all of the existing data" (Stimson 2017, 11). This step is crucial and justifies the claim that dyad ratios uses only *existing values* of items and restricts the missing values from affecting the latent structure estimation.

3. The process in Step 2 is extended and repeated until I have latent structure estimates for the remaining time periods. In the example, this involves estimating C_{T-3} and C_{T-2} to conclude the backward recursion process. While the averaging process of the ratios is the same as stated in Step 2, the only difference is that instead of T3 / T4 dyads I am using all existing items with T2 / T3 dyads and projecting C_{T-2} by using the data determined value of C_{T-1} instead of the arbitrary value of 100 chosen for C_T. Steps 2 and 3 are repeated until the backward recursive process hits the first period in the dataset. A final step to finish this process is to measure the percentage change in the latent structure C from one period to the next as shown in the "Latent C (% change)" column in *Figure 1.5*. This is done to transform the latent C obtained by the dyad ratios algorithm into a comparable form to the latent C estimated by a simple average. Since the estimation technique involves averaging

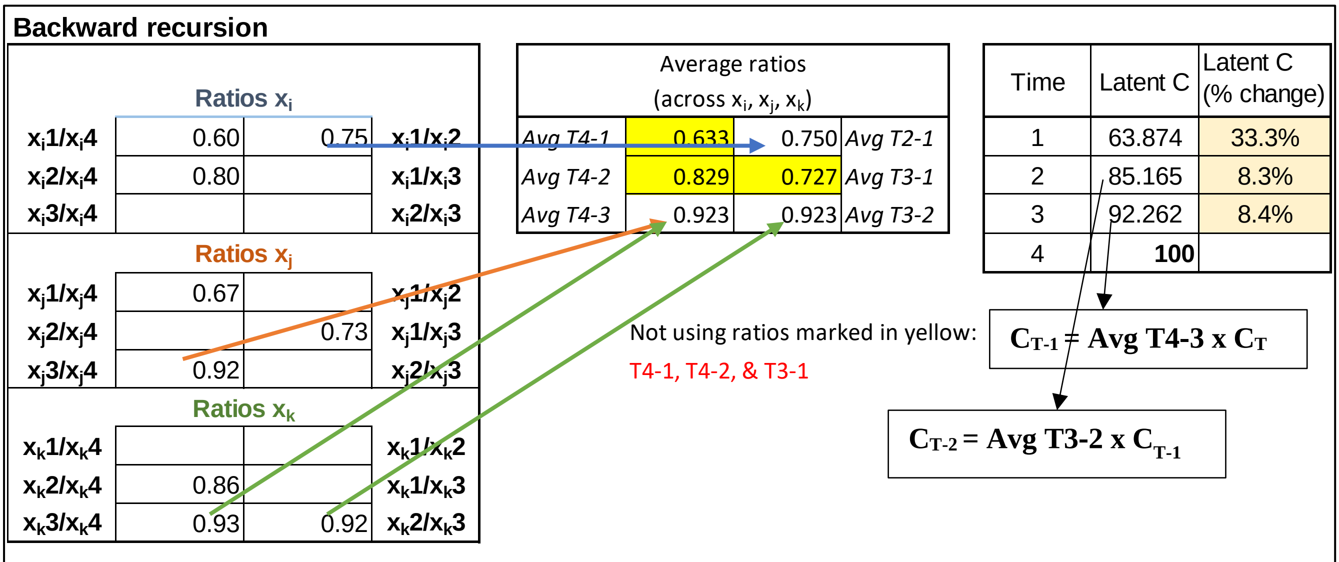


Figure 1.5: Complete Walkthrough of Backward Recursion Process

dyad values and projecting them *relative to other time periods*, the latent C needs to be interpreted as percent changes rather than an absolute level.

Forward Recursion Process:

Once the backward recursion process is understood, the forward recursion process is simple to understand. Steps 1, 2 and 3 from the backward recursion process are mirrored and tweaked to build the dyad ratios, averaging the existing dyads, and projecting the latent structure. Here, I note only the key differences when implementing the following steps for the forward recursion process and *Figure 1.6* and *Illustration 2* will serve as visual guides for the process.

(Step 1)

	Questions Items		
Time	x_i	x_j	x_k
1	30	40	
2	40		60
3		55	65
4	50	60	70

Dyads created - Step 1	
x_i2/x_i1	1.33
x_i3/x_i1	-
x_i4/x_i1	1.6666667

(Step 2)

	Questions Items		
Time	x_i	x_j	x_k
1	30	40	
2	40		60
3		55	65
4	50	60	70

Dyads created - Step 2	
x_i3/x_i2	-
x_i4/x_i2	1.25

(Step 3)

	Questions Items		
Time	x_i	x_j	x_k
1	30	40	
2	40		60
3		55	65
4	50	60	70

Dyads created - Step 3	
x_i4/x_i3	-

Illustration 2: Forward Recursion Dyads Walkthrough

1. In contrast to the backward recursion, the first step here involves starting at the first period (T1) for an item, x_i for instance, and building ratios relative to T1 for x_i at all successive periods (T2, T3, & T4). Next, we continue the process by moving one period forward ([T1] + 1 = T2) and building ratios relative to T2 for x_i at all successive periods (T3 & T4). This process is repeated until the final period (T4). The results of the dyad ratios creation for all items can be seen in the forward recursion table in *Figure 1.4.* Once again, the effect of the initial missing values is evident as the resulting subset of ratios is similar to the forward recursion table in *Figure 1.2* but features 3 missing values for each item.

2. The averaging process for forward recursion mirrors the one used in the backward recursion. But instead of starting with the final period, the forward recursion begins with an arbitrary value (100) at the first period (T1) and moves forward in time. Thus C_{T+1}, the latent structure at the T2 period, is estimated by using the absolute values for all items with existing values at T2 and in turn, existing dyad ratios for the T2 / T1 periods. *Figure 1.6* shows the calculation of the averages for all items with existing values and how these averages are used in projecting the forward recursive estimates of the latent structure at all time periods, C_{T+1}, C_{T+2} and C_{T+3}. The percentage change of the latent structure over subsequent periods is shown in the column "Latent C (%change)" in *Figure 1.6* and the formula is reversed to reflect the earlier periods starting with C_T (T1) as the initial number. Stimson (2017) highlights another important quality of the Dyad Ratios when he says "When using backward recursion later periods tend to dominate the solution. Forward recursion has the reverse weighting of backward, earlier items contribute more to the solution than do later ones" (Stimson 2017, 11).

Forward recursion

	Ratios x_i		
x_i2 / x_i1	1.33333		x_i3 / x_i2
x_i3 / x_i1		1.25	x_i4 / x_i2
x_i4 / x_i1	1.66667		x_i4 / x_i3
	Ratios x_j		
x_j2 / x_j1			x_j3 / x_j2
x_j3 / x_j1	1.375		x_j4 / x_j2
x_j4 / x_j1	1.5	1.09091	x_j4 / x_j3
	Ratios x_k		
x_k2 / x_k1		1.08333	x_k3 / x_k2
x_k3 / x_k1		1.16667	x_k4 / x_k2
x_k4 / x_k1		1.07692	x_k4 / x_k3

	Average ratios (across x_i, x_j, x_k)		
Avg T1-2	1.333333	1.083333	*Avg T2-3*
Avg T1-3	1.375	1.208333	*Avg T2-4*
Avg T1-4	1.583333	1.083916	*Avg T3-4*

Time	Latent C	Latent C (% change)
1	**100**	
2	133.333	33.3%
3	144.444	8.3%
4	156.566	8.4%

Not using ratios marked in yellow:
T1-3, T1-4, & T2-4

C_{T+1} = Avg T1-2 x C_T

C_{T+2} = Avg T2-3 x C_{T+1}

Figure 1.6: Complete Walkthrough of Forward Recursion Process

<u>Consolidated Latent Structure Estimate</u>:

The final step in the Recursive Estimation process involves averaging the Forward and Backward recursive estimates of the latent structure, "C_F and C_B", calculated in *Figure 1.5* and *Figure 1.6.* Stimson (2017) notes that the first advantage of averaging C_F and C_B stems from using all the available information from the question items transformed into their respective dyads and avoiding the pitfalls of adverse selection. Furthermore, the averaging of C_F and C_B tackles the differences in weights produced by the backward and forward recursions discussed previously and gives the user a single summary score that balances the weights of the earlier and later periods in the dataset.

❖ Is this a holistic description of the Dyad Ratios algorithm?

The motive behind this section is not to replicate every single step and provide an exact description of the dyad ratios algorithm. The algorithm doesn't end with the recursive estimation process discussed above as it involves three more stages that Stimson (2017) explains in more precise terms. I focus on the recursive estimation process and describe it in greater detail with an example as it signifies the crux of the algorithm. At most, the recursive estimation exposition provided here is a simplified version of the dyad ratios' true formula to capture the spirit of the algorithm and rationalize its use in constructing the Climate Change Skeptic Index. The three additional stages carried out by the software package (Wcalc) that supports the dyad ratios are:

1) Smoothing: Stimson (2017) argues that in combining the forward and backward recursive estimates (C_F and C_B), a smoothed approximation is better than data-determined estimates which have sampling error baked into their estimation process. The chosen model for the dyad ratios algorithm is an exponential smoothing model of the form: $y_t = \alpha x_t + (1-\alpha)\, x_{t-1}$.

2) Validity Estimation: During the recursive estimation, I assumed without proof that the items included in the dyad ratios to estimate the latent structure are valid indicators of it. Stimson (2017) suggests there are three alternatives for validity estimation: assuming perfect validity, estimating from the R^2 of multiple regressions of an item as dependent on all other items, and iterative estimation. Stimson (2017) uses the iterative estimation approach that essentially creates a weighted average for the dyad ratios at each time point using their validity estimates "μ_i^2" as the weights. The validity estimates for each item i (μ_i^2), is the amount of variance shared between the item and the latent structure. Stimson (2017) notes that the iterative solution for this process is obtained when the difference in the μ_i^2 between past and present iterations differs by less than .001. Based on the theory of vector decomposition, Stimson (2017) argues that if the true values of the μ_i^2 were known, then we can state that the squared correlation between the latent structure and the item would be equal to μ_i^2. Thus, by comparing the squared correlations between the latent structure and each item for all N, and verifying that they differ by a small amount (.001) across time, we can build an iterative solution to gauge the respective item's validity and use it as a weight to estimate the respective dyad ratio.

3) Bootstrapping Errors: The need for bootstrapping arises from the lack of readily available standard errors automatically generated by the estimator. Stimson (2017) asserts that bootstrapping is empirically a “second best” alternative than other options and espouses its statistical foundation when he states “the fundamental idea of bootstrapping is that we can subject the estimator to variation of known magnitude in data input and then observe its behavior” (Stimson 2017, 16). The inferred variations in data input improves with the number of observations and is simply a protracted description for the standard deviation of the distribution.

❖ **Alternative estimators and the rationale for using Dyad Ratios:**

In the Methodology section, I briefly contrasted the dyad ratios with the simple and weighted averages as tools to build the potential proxy for climate skepticism. Despite the reproduction of the recursive estimation process of the Dyad Ratios in this section, there are several elements of the algorithm that I failed to replicate on excel. This inability to replicate every aspect of the Dyad Ratios algorithm on a spreadsheet layout renders its black box qualities and raises questions about its applicability for estimating the latent structure of climate skepticism. The graphs in *Figure 1.8* are a crucial first step to justify the use of Dyad Ratios. *Figure 1.8* plots the timeseries estimates of the latent structure of climate skepticism varying from Q1 2000 to Q4 2015 provided by the three estimators I’ve already discussed. The Climate Change Skeptic Index (CCSI) is a name inspired by the Climate Change Threat Index created by Brulle et al. (2012) and will be discussed in greater detail in the next section. The actual survey data used in the empirical analysis section of the paper was used to create a simple average estimate, weighted average estimate, and a dyad ratios estimate. The simple and weighted averages were created on an excel sheet using simple INDEX & MATCH functions to filter the survey questions by year, average all the marginals and adjust their weights by the sample size. For the dyad ratios estimation, a software package called

Wcalc (described in the CCSI section) was used to produce the timeseries. Coming back to *Figure 8,* this side-by-side comparison with the averages provides an initial screening and shows that the data generating process of the dyad ratios is consistent and produces close

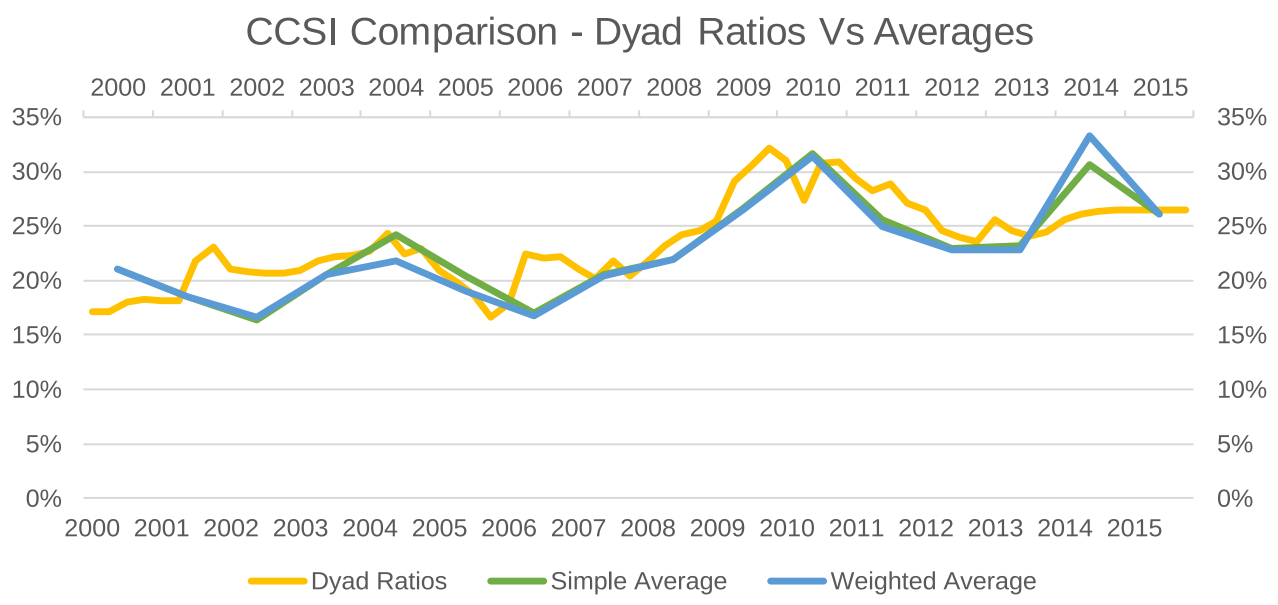


Figure 1.8: Dyad Ratios Justification

estimates to the average. Next, I look at an Item Response Theory alternative to build the latent structure and conclude with the rationale for sticking with the Dyad Ratios algorithm.

❖ McGann's Item Response Theory and Criticism of the Dyad Ratios

Item Response Theory (IRT) models have been a standard method used predominantly in the field of psychology but have found particular use cases in contemporary political science research. IRT was considered an important innovation for researchers in psychometrics as it provided an alternative for the Classical Test Theory and captured the interactions between survey items and individual-level responses in a similar manner as probit regression models. Implementing an IRT model allows the researcher to render an S-shaped curve to analyze dichotomous items using estimation techniques like maximum likelihood and other Bayesian methods. Despite these use cases, McGann (2013) highlights the fact that most of the existing IRT approaches only work with individual-level response and cannot be applied to aggregate level data in the same way as the Dyad Ratios algorithm. Next, I briefly discuss McGann's IRT approach that can deal with aggregate level data and describe his criticisms of the Dyad Ratios algorithm.

IRT Model of Policy Mood: Unlike the Dyad Ratios algorithm, McGann's IRT approach assumes item validity, i.e. the items chosen by the researcher to estimate the latent structure (or mood) are assumed to be valid indicators. Beginning with this assumption, the model states that there is a probability function to assess the respondent's answer and categorize it as "correct" (for instance - estimating climate skepticism, correct would equate to answering as a skeptic). Next, each question has two parameters: difficulty and discrimination that affect the probability of the correct response. The parameters instigate the variations in the probability of the correct response and McGann (2013) explains that if the variable measuring the position of the respondent is greater than the difficulty parameter, the probability of a correct response is greater than 0.5. Similarly, a low discrimination parameter coupled with a respondent's greater ability to answer a question correctly (i.e. the respondent's position variable > difficulty parameter) will render a probability that is closer to 1. *Figure 1.9* below depicts the varying probabilities of answering correctly captured by three question items. McGann (2013) uses this foundation to develop the IRT model, run other mathematical transformations to improve the estimation results and implements it using a Bayesian inference software such as JAGS or BUGS (McGann 2013, 120).

Criticism of Dyad Ratios: The main criticism of the Dyad Ratios by McGann (2013) relates to the apparent "asymmetry" between dyad ratios due to differences induced by choosing the left-wing (or non-skeptical) responses versus right-wing (or skeptical) responses as the object of the ratio. McGann (2013) uses this example to show that a shift from 20% to 60% gives a ratio of 1:3 but a shift from 80% to 60% gives a ratio of 4:3. This might not be the best representation of the relative changes in policy mood. There is a concession that the problem could be mitigated by the Dyad ratio algorithm as it repeatedly "reweights items based on commonalities", meaning it verifies the weight of each question item by its ability to indicate the latent structure.

Stimson's Defence of Dyad Ratios: Stimson (2017) doesn't directly address the concerns raised by McGann (2013) but implies that the comparison between the IRT and Dyad Ratios might not be a point-by-point comparison since it's a case where both the mathematical model and the input data of the approaches vary. But, Stimson (2017) shows that the latent structure estimates produced by both Dyad Ratios and IRT converge and hence can be used to model the empirical approach for similar purposes.

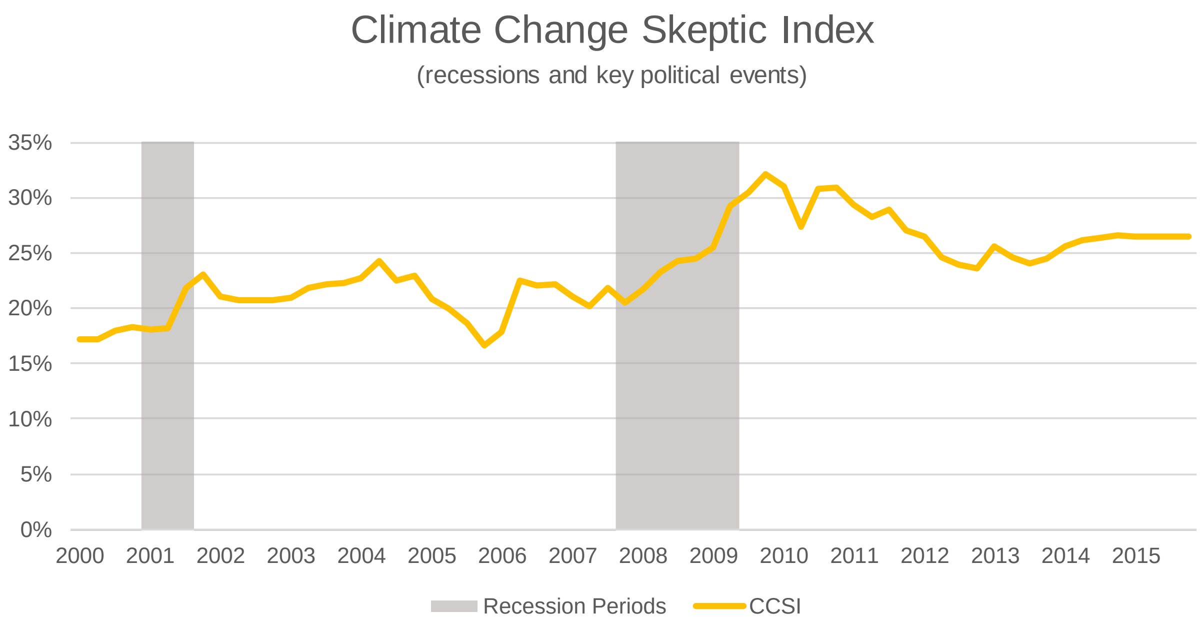
The observed differences between the two approaches and the criticisms of the Dyad Ratios algorithm do not discourage its use case for this paper. The marginal benefits stemming from McGann's IRT code, deciphering and implementing it for this empirical study do not supersede its incremental costs. The Dyad Ratios provides a convenient way of building a consistent latent structure estimate using the Wcalc software and helps achieve the key agenda of evaluating trends by running regression analyses. Ideally, I would have liked to demystify the "black-box nature" of certain components in the Dyad Ratios algorithm and replicate it perfectly. But the rational approach is to consider the scope of this academic paper, the opportunity costs involved in exploring the IRT (or other valid methods) and make the pertinent trade-off for the greater good.

Empirical Strategy

The crux of every serious claim made in this research paper rests on the clarity of the empirical strategy. Theoretically, the key goals of the paper are straightforward. I've discussed a few broad ideas involving the existence of an aggregate-level measure of climate skepticism and refer to contemporary literature to theorize the potential factors that influence this phenomenon over time. Empirically, the agenda is manifold but can be summarized into five overarching stages: 1) to build this aggregate-level measure of climate skepticism in a logical and realistic manner, 2) to find the best way to combine this measure analytically with data pertaining to the key explanatory variables, 3) to assess the quality of the results obtained from the model, 4) to interpret the results and contextualize them within the project definition and 5) to critically discuss the implications of the findings by making focused, statistically justified observations about the factors explaining climate skepticism. To tackle this agenda, the empirical section will be divided into three sub-sections. First, I examine the data generation process (DGP) that renders the Climate Change Skeptic Index (CCSI), the aggregate-level measure of climate skepticism in the US. This part also discusses the well-known multivariate time series model, vector autoregression (VAR), which will be used to study the relationships between the CCSI and variables representing recessionary economic factors and partisanship. The second sub-section will take a comprehensive look at all the data used in the analysis and state the various hypotheses tests that will be assessed through the VAR. Lastly, I conclude with an interpretation and discussion of the results to address the key takeaways from the analysis.

❖ DGP: Climate Change Skeptic Index

During the ideation phase of this paper, the intent was to test empirical claims of factors influencing public opinion on climate change in the US. Further exploration led to discovering extensive literature on the culture of climate skepticism. Initially, as I brainstormed plausible ways of formulating a thesis from these intriguing ideas, the task of building an aggregate-level measure of climate skepticism seemed insurmountable. But this persistent search for plausible estimators of climate skepticism led to the climate change threat index (CCTI) (Brulle et al., 2012). The inspiration and guidance for formulating the CCSI came directly from Brulle et al. (2012) and the methodology used to create the CCTI. In the *Stimson Algorithm* section, I explored the underlying dyad ratios method that directs the process of the Wcalc program. Here I will discuss the stylistic elements of the data used to build the CCSI, including the criterion used to qualify survey questions for the index and their respective descriptive stats.



<u>CCSI Construction and General Facts:</u> The raw survey data for the CCSI was gathered from two different sources and merged into one master dataset to ease the recoding process needed to meet the requirements of the Wcalc program. The first source was the iPoll+ database hosted by Cornell University's Roper Center. The Roper Center is home to a vast amount of public opinion data collected from prominent polling institutions such as Pew, Gallup, NY

Times, etc. A database search with the keywords “climate change”, “global warming”, and “greenhouse” yielded a total of 268 questions for the timeline filter: 1st January 2000 – 31st December 2015. Next, I manually verified these 268 questions to exclude all items irrelevant to the construction of the CCSI. The manual selection was verified multiple times to maximize the effort to avoid human error and bias. To standardize the selection process, I established two broad screening categories to ensure that questions related to these themes made it into the CCSI. The first category captured questions on climate skepticism and was further divided into three sub-categories. 1) “Belief”- these types of questions test whether the respondent simply believes that climate change/global warming is real/occurring. 2) “Science” – these types of questions test whether the respondent disagrees with the scientific consensus on climate change or disputes basic scientific facts. 3) “Attitude” – these types of questions attempt to gauge the respondent’s attitude towards climate change and check for skeptic/denial responses to questions about climate change action. The second category includes slightly tougher questions related to the nature of the cause of climate change. Essentially, this category aims to capture questions that test the respondent’s belief that climate change (or global warming) is not anthropogenic, i.e. not caused by human activities and attributed to natural changes. Another critical decision involved determining which survey marginal responses to include in the formulation of the CCSI for the relevant questions. Most questions had approximately five responses along these lines: strongly agree, agree, neither agree nor disagree, disagree and strongly disagree. Based on the question content, the survey marginals (percentage of total respondents choosing a particular option) of both positive (“strongly agree” and “agree”) or negative (“strongly disagree” and “disagree”) responses were combined to record the respective item’s skeptic response value. This question-filtering process from the first source resulted in a dataset with 101 total question items coming from 69 unique questions. 18 polling organizations asked these questions for the target period of Q1 2000 to Q4 2015 and the dataset has a sample size of 115,355 respondents. Though the iPoll+ database hosted questions from the Gallup organization, I was able to find 8 other questions related to climate skepticism on their official website which matched the selection criterion. These questions were re-coded in a similar manner and the resulting dataset consisted of 85 total questions from the 8 unique questions administered to 84,000 respondents for the target time period. After combining these two datasets, we are ready to use the Wcalc program to create the CCSI.

Basic Instructions - Wcalc Program: The Wcalc program was created by James Stimson for researchers to input data and build a timeseries latent estimate based on his Dyad Ratios algorithm. Wcalc is very specific on how it reads the input file and essentially requires all data classified into four categories: 1) Date, 2) Variable Name, 3) Marginal Score, and 4) Sample Size. The raw dataset from iPoll+ already provides three of the four filters for the data and I already explained how the remaining one (survey marginals) was built. An extensive documentation of how to use Wcalc has been provided by the author, Stimson (2017). *Table S2* and *S3* in the *Appendix* provides descriptive stats like the mean and standard deviation for the questions inputted into Wcalc. A comprehensive list including full text of chosen questions, polling organization, CCSI Iteration History by Wcalc, etc. can be found in the *Appendix* as well. As a reminder, *Table S3* only shows 19 questions since the dyad ratios algorithm requires a minimum repetition of 2 cases over the chosen time period.

Now that we have understood the process and steps involved in the construction of the CCSI, we take a deeper look at the data generating process of the VAR model. I believe that understanding the data generating process is an essential precursor to reliably interpret the results of the empirical model and confidently discuss the findings.

❖ DGP: Vector Autoregression Model

I have previously mentioned that the methodology used in Brulle (2012) was a big inspiration for the empirical strategy of this paper. But prior to making the decision of using a VAR model, I researched other ways to model the key theoretical questions about climate skepticism. Some of the alternative empirical strategies were discussed in the literature review. The closest alternative idea was to run simple OLS or logit regression analyses immediately before and after recession periods to study its impact on climate skepticism. A drawback of this approach is that it fails to account for past periods of variables influencing current period estimates. This is especially important when considering that recessionary effects might take time to impact people's lives and public opinion on issues like climate change are never static. Another drawback comes from the inability to account for uncertainty with respect to the nature of the variables included in the model, i.e. whether a certain variable and all previous period estimates of it are truly exogenous to the specified model. The VAR system mitigates these problems and turns out to be a valuable tool in analyzing variations in climate skepticism over longer periods of time. The natural question to consider is: "What is a VAR and what does it mean?". Our work in this section is not to simply restate definitions and equations that can be easily found elsewhere. I will strive to provide an explicit answer to this question while being cognizant of the layperson's needs and the scope of this paper. For this purpose, the VAR model can be broken down into five steps: 1) model specification, 2) pre-estimation steps, 3) estimation of VAR, 4) post-estimation causality steps, and 5) post-estimation diagnostics. My focus will be on the first three steps to ensure there is sufficient clarity. The final two steps are important but have been covered extensively by academics and are done mostly to analyze the relevance of our results and gain confidence on the estimations.

1. Model Specification

The VAR system is often chosen when researchers are not sure of the exogeneity of the included variables in the model. For instance, one set of variables used in this paper pertain to media coverage of climate change. There is no discernible way to say that environmental and conservative magazine articles are strictly exogenous, i.e. are independent and unaffected by current levels and past values of other variables in the model. To help understand and walkthrough the rest of the VAR model, I will use a variable subset from the main data as a guide for the rest of this section. Readers can follow along using the "Stata VAR Excelification.xlsx" and "VAR Excelification (dead).xlsx" files. A bivariate third-order model consisting of unemployment rate (y) and the CCSI (z) will be used for reference. The "'AR" part of the VAR model stands for "autoregressive" or variables that can be influenced by past values of their own sequence. The "third-order" simply indicates the number of lags (3) that will be included in the model. The number of lags refers to how many previous periods of the variables will be included in the model. Since all variables in the VAR model are endogenous, they will each appear on the left-hand side of the equation once and will be estimated at current levels (i.e. time, t) by regressing past realizations of their own values (3 lags = t-1, t-2, t-3) and past and current realizations of other included variables in the system.

VAR (3rd Order Model):

y_t (unemployment) $= a_{10} + a_{11}.y_{t-1} + a_{12}.y_{t-2} + a_{13}.y_{t-3} + a_{14}.z_{t-1} + a_{15}.z_{t-2} + a_{16}.z_{t-3} + e1t$

z_t (CCSI) $= a_{20} + a_{21}.y_{t-1} + a_{22}.y_{t-2} + a_{23}.y_{t-3} + a_{24}.z_{t-1} + a_{25}.z_{t-2} + a_{26}.z_{t-3} + e2t$

Figure 2.1: Bivariate VAR: Standard Form Equations

The bivariate third-order model is represented in its standard form in *Figure 2.1*. The properties of the error (also known as innovations) terms (e1t and e2t) are crucial and represent white-noise processes that are stationary (we will learn why this is important in the next step) and correlated with one another. The sequences for the error terms are recreated in excel using the formula "=NORMINV(RAND(),0,1)". The NORMINV function returns the inverse of a probability corresponding to the normal cumulative distribution (i.e. RAND(), which returns a random number > = 0 and less than 1) for the specified mean (0) and standard deviation (1). If the VAR model isn't specified in its standard form, I cannot apply ordinary least squares (OLS) techniques to estimate the coefficients. The primitive form of the VAR disallows using OLS because "y_t has a contemporaneous effect on z_t and z_t has a contemporaneous effect on y_t" (Enders 2011, 285). To be precise, the y_t sequence will influence the z_t sequence and vice-versa during the same time period that we try to estimate their parameter values and hence runs into a multicollinearity problem with the regressors and the errors terms ending up correlated. Now that I have specified the VAR model, let's move onto a few pertinent pre-estimation steps.

2. Pre-Estimation Steps

Stationarity tests and optimal lag length selection tests are the two key pre-estimation steps that must be done before we can proceed onto estimating the VAR. Though I have already specified that the third-order model was chosen, let's take a look at the intricacies involved in that lag length selection process. First, I tackle two important questions: "What does it mean for a process to be stationary?" and "Why does it matter for our VAR model?". There are three necessary conditions for a timeseries to be considered stationary. Firstly, the expectation of our process, let's call it y_t, needs to be equal to some constant, μ. Secondly, the variance of y_t needs to be equal to σ^2, again a constant. Finally, the covariance for y_t with y_{t+h} is some function of h (f[h]), and not a function of time. Essentially, the key thing to remember here is that our process, y_t, comes from some data generating process (DGP) that is similar across all time periods. A process that satisfies these conditions is classified as stationary and this basically assures that the y_t process is not generated by different DGPs from one period to the next and is consistent irrespective of time.

To explain why stationarity matters for our VAR and understand how to run a stationarity test on Stata and replicate it in excel, I reintroduce the bivariate model (see *Figure 2.1)* discussed previously but only consider a first-order model (1 lag) to narrow our focus. My steps can be followed along using the excel file "VAR Excelification (dead)". *Figure 2.2* is a screenshot of the data table you will find in the "VAR_DFTest" sheet. There are two main reasons for the stationarity conditions in the VAR model (and most timeseries models). First, stationarity helps estimate any linear interdependencies between the included variables for the given period. For instance, in *Figure 2.2* if the CCSI or the unemployment series were nonstationary, then I would struggle to accurately interpret the coefficient estimates and describe the relationships shared by the two series across the target time period. The second reason is a theoretical one and according to Lambert (2013), without stationary timeseries we wouldn't be able to leverage the Law of Large Numbers and the Central Limit Theorem for inference purposes.

VAR- Stationarity Replica			**Constants:**	a_{10}:	0.303	**Parameters:**	a_{11}:	0.989	a_{12}:	-0.009
Model: Unemployment Rate-> CCSI				a_{20}:	3.247		a_{21}:	0.3	a_{22}:	0.788
Dates	CCSI	Unemployment Rate	e1t	e2t	y_t (unemployment)	▲y_t	z_t (CCSI)	▲z_t	y_{t-1}	z_{t-1}
2000Q1	17.145	4.03	-0.991	0.959	0	0	0	0	0	0
2000Q2	17.145	3.93	-0.746	-0.095	-0.443	-0.443	3.152	3.152	0	0
2000Q3	17.985	4.00	0.864	-1.335	0.700	1.143	2.508	-0.644	-0.443	3.152
2000Q4	18.268	3.90	0.994	-0.665	1.966	1.266	3.886	1.378	0.700	2.508
2001Q1	18.095	4.23	0.203	1.945	2.416	0.450	7.907	4.021	1.966	3.886
2001Q2	18.157	4.40	-1.776	0.388	0.846	-1.570	7.911	0.004	2.416	7.907
2001Q3	21.808	4.83	0.729	0.511	1.798	0.952	6.798	-1.113	0.846	7.911
2001Q4	23.044	5.50	-1.338	-0.461	0.682	-1.116	6.242	-0.556	1.798	6.798
2002Q1	21.038	5.70	1.449	-2.158	2.370	1.689	3.499	-2.743	0.682	6.242
2002Q2	20.76	5.83	0.797	-0.219	3.413	1.042	5.946	2.447	2.370	3.499
2002Q3	20.721	5.73	-1.445	-0.529	2.180	-1.233	7.191	1.245	3.413	5.946
2002Q4	20.735	5.87	0.672	1.109	3.066	0.886	8.231	1.039	2.180	7.191

Figure 2.2: Screenshot of bivariate (first order) VAR used for Step 2

Running and Replicating Stationarity Tests: A common way econometricians test for non-stationary timeseries is by running a Dickey-Fuller test for the presence of a unit root process. The mathematics behind the unit root process is beyond the scope of this thesis and the reader can refer to the appendix for information and resources to understand the mechanics behind it. But, in short, the "unit root problem is concerned with the existence of characteristic roots of a time series model on the unit circle" (Tsay 2008, 1). The existence of a unit root in the VAR process is not a good sign as it indicates that the series may be nonstationary and carries negative implications. The stationarity test begins with a null hypothesis that the chosen timeseries variables, y_t and z_t, are nonstationary and the alternative hypothesis that they are stationary. The Dickey-Fuller test is used in lieu of an ordinary t-test because under the null hypothesis, the Central Limit Theorem (CLT) fails, and I'm unable to use an ordinary t-distribution to test the t-statistic. To tackle this issue, I take the first differences of the timeseries as it has a better chance of producing a stationary process. At the minimum, even if the right-hand side processes are nonstationary, I am in a better position with the first differenced series on the left. The delta-y_t and delta-z_t columns in the excel file show this transformation and help us continue the stationarity test. The first-differencing improves our situation but I still cannot use an ordinary t-test. I run into the same problem as before because under the null hypothesis, y_{t-1} and z_{t-1} are still considered to be nonstationary and thus the t-statistic is still not comparable to the t-distribution (CLT fails). This is where Dickey and Fuller (1979) enter the scene and save us the hassle of tabulating the asymptotic distribution of the least squares estimator for a_{11}, a_{12}, a_{21}, and a_{22}, the coefficients of y_{t-1} and z_{t-1}, under the null hypothesis that these processes are unit root. The final step for checking whether the series are nonstationary is straightforward as I simply compare the t-statistic with the Dickey-Fuller distribution. *Figure 2.3* shows the results of the Dickey-Fuller test on Stata and depicts our attempt at replicating the same test in the "VAR_DF Test" sheet of the "VAR Excelification(dead)" excel file. If you hit F2 in the t-stats table in Excel, you can see how the LINEST function is built using the first differences of y_t and zt. The t-statistics found via excel match the ones given by Stata and this confirms that the replication method is accurate.

t-statistic for DF Tests	y_t (unemployment rate) equation	
-2.278	-0.15712	0.05408
	0.06897	0.14056
	0.0784	1.09784
LINEST	5.18914	61
	6.25418	73.5198
	z_t (CCSI) equation	
-3.280	-0.25646	1.39237
	0.07818	0.43513
	0.14995	1.44157
	10.7608	61
	22.3624	126.766

```
. dfuller ytunemployment

Dickey-Fuller test for unit root                   Number of obs   =        63

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.278            -3.562            -2.920            -2.595
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.1792

. dfuller ztCCSI

Dickey-Fuller test for unit root                   Number of obs   =        63

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.280            -3.562            -2.920            -2.595
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0158
```

Figure 2.3: Excel Replication of Stata DF Test Results

The second pre-estimation step is the optimal lag length selection test. This step is a crucial part of the identification process needed before I can reliably estimate the VAR. Let's revert back to the third order bivariate VAR shown in *Figure 2.1* since I only used the single order model to simplify the DF tests excel replication.

Optimal Lag Lengths: While it is technically possible to allow different variables in the VAR to have varying lag lengths, it does not help use OLS estimation techniques since they require identical regressors on the right-hand side for each equation. Selecting the lag length is also important because it determines how many coefficients I'll have to estimate from the model. A model with p lag lengths and n equations will contain $n*p$ coefficients plus the intercept term. Enders (2011) captures the perils of not choosing the optimal lag length and states: "If p is too small, the model is mis specified; if p is too large, degrees of freedom are wasted" (Enders 2011, 303). As you may have guessed, I had run the optimal lag length test on Stata using the "varsoc" command and the Akaike Information Criterion (AIC) picked 3 lags for the model in *Figure 2.1*. Additional information on the AIC and other similar tests like it is provided in the *Appendix*. To maintain focus on the VAR steps that can explicitly be replicated on Excel and to avoid resharing theoretical knowledge already stated by academics, I push ahead to the next section. Now that the VAR model is specified and the pre-estimation due diligence is completed, I'm ready to estimate the model specified in *Figure 2.1*.

3. Estimation of the VAR

The standard form equations of the model are specified in *Figure 2.1*. An important thing to note is that we always specify our VAR model in *levels* and not in *differences*. Specifying the VAR in differences will result in mis-specification because the right-hand side variables will vary across all equations in the model. Another key thing to remember is that the results from the Dickey-Fuller test determine whether the VAR model can be constructed as specified. This means that if the raw series is nonstationary, it must be stationary after first difference (integrated of order 1, I(1)) or else it cannot be included in the model. Both the CCSI and unemployment series are stationary. The estimation of the VAR model in Stata is done using the command "var ytunemployment ztCCSI, lags(3)". This command spits out the estimation results as shown in *Figure 2.4.* Stata gives us 6 coefficients and one intercept term for each equation (number of equations (n) * lags (p)) and the estimates table is divided into two mini sections with the bolded variable representing the key endogenous variable for the respective equation. The first coefficient, 1.561, adjacent to the "L1" row of the ytunemployment is basically the a_{11} estimate depicted in the standard form equations in *Figure 2.1.* The second coefficient, -0.371, corresponds to the a_{12} estimate and so on.

Vector autoregression

Sample:	2000q4 - 2015q4	Number of obs =	61
Log likelihood =	-95.07846	AIC =	3.576343
FPE =	.1227777	HQIC =	3.766209
Det(Sigma_ml) =	.0774264	SBIC =	4.060806

Equation	Parms	RMSE	R-sq	chi2	P>chi2
UnempRate	7	.225042	0.9850	4017.976	0.0000
CCSI	7	1.43409	0.8618	380.3805	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
UnempRate						
UnempRate						
L1.	1.561047	.1268445	12.31	0.000	1.312437	1.809658
L2.	-.3714483	.2303677	-1.61	0.107	-.8229606	.0800641
L3.	-.2621824	.1272457	-2.06	0.039	-.5115793	-.0127854
CCSI						
L1.	.0061154	.0206343	0.30	0.767	-.034327	.0465579
L2.	-.0299973	.0270255	-1.11	0.267	-.0829663	.0229717
L3.	.0395877	.0216977	1.82	0.068	-.0029391	.0821145
_cons	.0971024	.2256394	0.43	0.667	-.3451426	.5393475
CCSI						
UnempRate						
L1.	.6398767	.8083216	0.79	0.429	-.9444045	2.224158
L2.	.4322604	1.468027	0.29	0.768	-2.44502	3.309541
L3.	-.8796307	.8108783	-1.08	0.278	-2.468923	.7096616
CCSI						
L1.	.8762241	.1314928	6.66	0.000	.6185029	1.133945
L2.	-.2033428	.1722211	-1.18	0.238	-.5408899	.1342043
L3.	.1523714	.1382697	1.10	0.270	-.1186321	.423375
_cons	3.05282	1.437896	2.12	0.034	.234595	5.871044

Figure 2.4: Stata VAR Estimation Results

In terms of interpreting all these numbers, some researchers can intuitively jointly interpret all the lagged values of the regressors. But the conventional way is to interpret them in the same manner as an OLS regression. For instance, the a_{11} estimate can be interpreted as the first lag of the unemployment rate (y_{t-1}) having a positive impact of 1.56 percentage points on average (at the 1% significance level) on the current level of unemployment (y_t) holding all its lagged values and those of other variables constant. While this type of interpretation is valid and gives the researcher some insights, most econometricians prefer running causality checks that help interpret the causal link of the combined lags of a variable with other variables. Other use a visual tool to interpret these results and engage in forecasting by seeing how the variables respond to different shocks in the model. This tool is called an impulse response function and I exclude it from this thesis since the focus is less on forecasting and more so on finding the causal connections between CCSI and the key explanatory variables. For an intuitive way to understand the effect of the coefficients, open the "Stata VAR Excelification" excel file and scroll right in the "VAR_Replication" sheet to column AI. Hit F9 and observe the graph to visually see how the underlying coefficients impact the predicted ytunemployment in the graph. Estimating a VAR and being able to interpret it on Stata is great, but it does not help us understand the underlying data generating process of the model. To truly understand what is going on, I replicate this VAR estimation in excel and run simulations to draw comparisons with the Stata output.

The excel replication discussion can be followed along by using the file "Stata VAR Excelification.xls". As seen in *Figure 2.4*, I begin with the known timeseries of the CCSI, built using the Wcalc program, and the unemployment rate from Q1 2000 to Q4 2015 which gives an n = 64. The errors terms 'bounce' as they are built using the same function as the Dickey-Fuller tests to mimic white-noise processes. The bounce in the error terms comes because the function pulls a random number, using the specified parameters, from a probability distribution every time I run an operation in Excel. The data in columns F-M are *live* primarily due to these errors. Hitting F9 on your computer allows you to observe this live data and the effects of the sampling process. Next, to populate the F-M columns, which contain the lagged values of y_t and z_t, I need to incorporate the standard form equations from *Figure 2.1*. But the constraint to applying these equations is that I do not know the true parameter values and don't possess any valid estimates. Hence, I use the coefficient estimates found through the Stata VAR estimation as starting values to build our excel replication. The named range "Parameter_Table" spans Cells F1 thru U2 and includes the constants and parameter estimates for the y_t equation and the second row includes parameter estimates for z_t equation. The downside is that I'm making a big assumption that the Stata estimates are reliable and unbiased. Let's skip the first three cells of the y_t column and assume zero values since I need three lagged values of each variable to include in the standard form equations (yt-1, yt-2, yt-3, zt-1, zt-2, and zt-3). Thus, I begin with cell F7 and the formula inputted in the cell can be viewed by hitting F2 in excel. Essentially, I transform the standard form equations formula into excel here. The same step is repeated for z_t and I enter its corresponding values in the formula. After I have copied the formula for all the cells in y_t and z_t, I can complete the remaining steps for the lagged values by either copying and pasting the y_t and z_t values in the respective lagged value columns or referencing the appropriate cells (hit F2 on the lagged value cells) for each lagged variable to their y_t or z_t equations.

Now that I have the full series of variables and their lagged values, I can use the built-in LINEST function in excel to estimate the model. Since LINEST can only estimate one equation at a time we run it separately for y_t and z_t. To estimate the y_t equation, I highlight the 7 columns and 5 rows, [P4:V8] range (LINEST always includes 5 rows but the number of columns depends on the coefficients being estimated) and input the formula (hitting F2 anywhere in the table displays the formula but remember to exit by hitting the ESC key). I repeat the process in the [P10:V15] range for the z_t equation. In the y_t equation's LINEST table, cell U4 is the coefficient estimate of the a_{11} parameter and the estimates flow in a reverse chronological order ending with P4 estimating the a_{16} parameter. Notice that the estimates in the table continue to bounce since the data is still live. To verify our excel VAR replication effort, I can run a simulation that repeats these OLS estimates multiple times and plots the resulting sampling distribution. A 10,000 repetitions simulation of the U4 cell returns an average (coefficient estimate) of 1.505 (see the "a11_sim" sheet) and an approximate SE of 0.1332. At first glance it might seem tough to make a statement of whether 1.505 is close enough (to the assumed a_{11} parameter of 1.56) but it becomes evident that this estimate is biased when I look at the value in cell M7 (highlighted in yellow). I did a simple calculation in cell M7 to test whether the approximate SE is large enough to allow for the error in the coefficient estimate to form an interval that included the a_{11} parameter value of 1.56. Unfortunately, the resulting estimate is biased, and running similar simulations (see sheets under “Simulations” tab) of other coefficients garners mixed results as some simulations produce unbiased estimates (a_{13}). Consistency of the estimator is another important quality that one should care about, and *Figure 2.5* illustrates that the excel VAR model is consistent. I arrive at this conclusion because as n rises (from 32 to 64), the estimates converge to the true mean.

Coefficient Estimate of a11 (cell U17) with n=32		Coefficient Estimate of a11 (cell U4) with n=64		Notes	
Average	1.402	Average	1.505	n = 32 Vs n = 64	
SD	0.2379	SD	0.1332	See Sheet: VAR_Replication	
Max	2.263	Max	2.018	0.00133	
Min	0.359	Min	0.947	Parameter value:	1.56

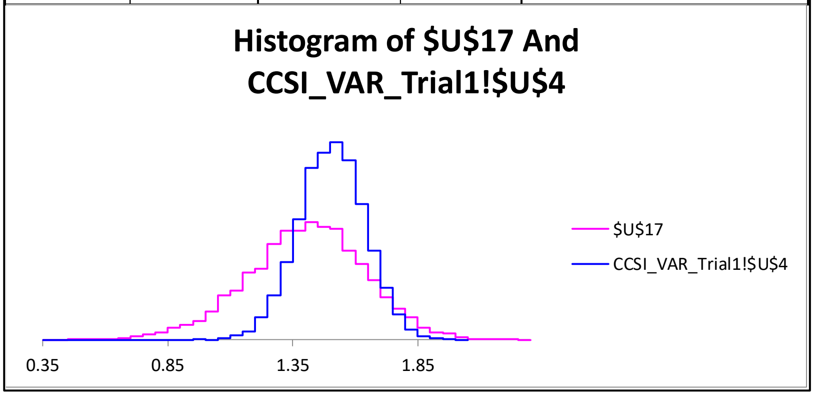


Figure 2.5: Proof of the Estimator’s Consistency

<u>Excel Replication with Dead Data</u>: Replicating the Stata VAR model with live data on excel gave us a lot of insights on the quality of the estimator and helped understand the black box nature of the DGP. But it still leaves a gnawing doubt of whether my excel replication is following the exact DGP and estimation technique of the Stata model. To clarify these doubts, I run the VAR estimation again in *Figure 2.6* but this time using *deadened* data instead of live data. The VAR result shown in *Figure 2.6* is available in the same file used to run the DF tests, “VAR Excelification (dead)”. This is done by simply copying the error series (e_{1t} & e_{2t}) and pasting them as values. The rest of the process remains the same on

excel and the results of the LINEST function can be seen in *Figure 2.6.* A sure-fire way of verifying the excel replication of VAR is by importing this deadened data seen in *Figure 2.6*

CCSI_VAR- Replica			Constants:	a_{10}:	0.09	Coefficients:	a_{11}:	1.56	a_{12}:	-0.37	a_{13}:	-0.26	a_{14}:	0.006	a_{15}:	-0.02	a_{16}:	0.03	
Model: Unemployment Rate-> CCSI				a_{20}:	3.05		a_{21}:	0.63	a_{22}:	0.43	a_{23}:	-0.87	a_{24}:	0.87	a_{25}:	-0.2	a_{26}:	0.15	
Dates	CCSI	Unemployment Rate	e1t	e2t	yt (unemployment)	zt (CCSI)	yt-1	yt-2	yt-3	zt-1	zt-2	zt-3	Excel Replica	yt (unemployment rate) equation					
2000Q1	17.145	4.03	-0.110	0.540	0	0	0	0	0	0	0	0	-0.0886	0.16186	-0.0485	-0.1671	-0.5232	1.58966	-0.0008
2000Q2	17.145	3.93	-0.907	-1.180	0	0	0	0	0	0	0	0	0.06734	0.11393	0.08556	0.16305	0.24849	0.13179	0.36705
2000Q3	17.985	4.00	-0.060	-0.788	0	0	0	0	0	0	0	0	0.97759	0.85628	#N/A	#N/A	#N/A	#N/A	#N/A
2000Q4	18.268	3.90	0.540	2.117	0.6305	5.1673	0	0	0	0	0	0	414.367	57	#N/A	#N/A	#N/A	#N/A	#N/A
2001Q1	18.095	4.23	-1.132	-1.657	-0.0278	6.2854	0.630	0	0	5.167	0	0	1822.91	41.7929	#N/A	#N/A	#N/A	#N/A	#N/A
2001Q2	18.157	4.40	0.355	0.727	0.1024	8.4655	-0.028	0.630	0	6.285	5.167	0	**LINEST**	**zt (CCSI) equation**					
2001Q3	21.808	4.83	1.153	-0.037	1.3296	9.4004	0.102	-0.028	0.630	8.465	6.285	5.167	0.03749	0.05239	0.79827	-0.8927	0.29599	0.69996	1.94643
2001Q4	23.044	5.50	-0.629	-0.075	1.5803	11.3092	1.330	0.102	-0.028	9.400	8.465	6.285	0.09373	0.15856	0.11909	0.22694	0.34585	0.18342	0.51085
2002Q1	21.038	5.70	-1.075	-0.105	1.0957	13.6518	1.580	1.330	0.102	11.309	9.400	8.465	0.98512	1.19177	#N/A	#N/A	#N/A	#N/A	#N/A
2002Q2	20.76	5.83	-1.520	0.085	-0.5133	14.3734	1.096	1.580	1.330	13.652	11.309	9.400	629.114	57	#N/A	#N/A	#N/A	#N/A	#N/A
2002Q3	20.721	5.73	1.789	0.874	0.4143	14.1679	-0.513	1.096	1.580	14.373	13.652	11.309	5361.22	80.9576	#N/A	#N/A	#N/A	#N/A	#N/A

Figure 2.6: Screenshot of deadened data and LINEST table

into Stata and running the VAR estimation command on it. After importing the data and completing the necessary recoding I obtain the results shown in *Figure 2.7* using Stata's VAR command. The estimated coefficients in the Stata VAR table precisely match the ones I computed on the excel LINEST table! This is exciting since it clarifies any doubts over the excel-VAR replication DGP matching Stata's DGP.

Equation	Parms	RMSE	R-sq	chi2	P>chi2
ytunemployment	7	.856276	0.9776	2791.528	0.0000
ztCCSI	7	1.19177	0.9851	4238.242	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ytunemployment						
ytunemployment						
L1.	1.589659	.1243729	12.78	0.000	1.345893	1.833425
L2.	-.5232138	.2345096	-2.23	0.026	-.9828442	-.0635835
L3.	-.167093	.1538778	-1.09	0.278	-.468688	.1345019
ztCCSI						
L1.	-.0485405	.0807483	-0.60	0.548	-.2068044	.1097233
L2.	.1618621	.1075154	1.51	0.132	-.0488643	.3725884
L3.	-.0885884	.0635518	-1.39	0.163	-.2131475	.0359708
_cons	-.0007741	.3463915	-0.00	0.998	-.6796891	.6781408
ztCCSI						
ytunemployment						
L1.	.699962	.1731025	4.04	0.000	.3606874	1.039237
L2.	.2959869	.3263911	0.91	0.364	-.3437278	.9357017
L3.	-.892735	.2141675	-4.17	0.000	-1.312496	-.4729744
ztCCSI						
L1.	.7982679	.1123858	7.10	0.000	.5779959	1.01854
L2.	.0523948	.1496402	0.35	0.726	-.2408947	.3456843
L3.	.0374933	.0884515	0.42	0.672	-.1358684	.2108551
_cons	1.946432	.4821086	4.04	0.000	1.001517	2.891348

Figure 7: Stata VAR Estimation of deadened data

With the excel replication done, I have now covered the key focus areas by investigating the three steps of specification, pre-estimation and, estimation of a simple VAR model. I will very briefly touch upon the last two steps of the model to avoid digression and repeating what others have explained better. Then I move onto the heart of my empirical analysis in the next section by describing the data used in this paper and setting up the main hypotheses.

4. Post-estimation Steps

The two main post-estimation steps that need to be done are causality checks and diagnostic tests. Equations in the VAR model can have high R-squared values without implying any causal connections between the dependent variable and its regressors. When the VAR model is applied to the CCSI in the next section and null hypothesis significance testing (NHST) is used to measure the significance of recessionary factors and political elite cues, the causality checks will determine whether I have enough evidence to reject the null. Finally, the diagnostic tests are done to ensure model integrity and gives us greater confidence and additional reliability to justify the model's results.

Causality Checks: There are three causality checks that can be used to determine a causal link between the CCSI and its regressors. First, Stata's VAR table provides "p-values" for every coefficient estimated by the model. A low p-value is often interpreted as a statistically significant result. For instance, a p-value < 0.05 corresponds to a statistically significant result at the 95% significance level and a p-value < 0.01 corresponds to a statistically significant result at the 99% significance level. The main question here is: "How do you interpret a low p-value intuitively?" A low p-value tells us that there is enough evidence in the sample to reject the null for the underlying population. Assuming a true null hypothesis, a low p-value technically implies the probability of obtaining a resulting effect at least as extreme as the one in the sample data.

The p-values only provide the statistical significance of individual coefficients and don't provide an easy way to jointly measure the causality of all lags of a given regressor in the equation. The Granger Causality test is one way to solve this issue. For simple OLS regression models, I run F-tests that determine whether the increase in R-squared caused by adding new independent variables was significant and a comparison of the whole-model to the restricted model provided enough evidence to reject the null (of no significance). The Granger Causality test acts in a similar manner and gives us a criterion to determine whether the regressors and their lagged values "Granger cause" the key endogenous variable in the equation (i.e. CCSI, in our main models).

Lastly, the third causality check is called a "linear test of parameter estimates" and is usually known as the "Wald test". The Wald test is another way to determine whether a regressor or any of its lagged values are causally linked to the key endogenous variable. To avoid delving into the complexities of the Wald test that have been better explained elsewhere, I explain only a simple scenario. In a Wald test with a univariate variable, the Wald test can be written as: $\frac{(\theta - \delta_0)^2}{var(\theta)}$, where "$\theta$" is the maximum likelihood estimate and "δ" is the parameter estimate and under the null hypothesis being true, this equation is chi-squared distributed with 1 degree of freedom.

<u>Diagnostic Tests</u>: As stated previously, the diagnostic tests provide mathematical justification to support the interpretation of the results from the VAR model. The three key areas that will be targeted using the diagnostics are: autocorrelation, normality of innovations (error terms), and stability. Firstly, the issue of autocorrelation (or serial correlation) of the errors terms can be defined as a condition when the covariance of an error, e_{1t}, and some other error, e_{2t}, is not equal to 0 and the two errors are not equal to each other. To keep things simple, when the VAR equations have autocorrelated errors then OLS is no longer the best linear unbiased

estimator (BLUE) since there are other estimators that perform better with this condition. The first diagnostic check, to test for autocorrelation, is called the Lagrange Multiplier test and can be run on STATA using the “varlmar, mlag(number of lags)” command. Running this command results in a table like those seen in the *Data Analysis* where I can use the p-values in the rightmost column to test the null hypothesis of no autocorrelation at the respective lag order. A low p-value indicates that the model specification doesn’t provide enough evidence to *not reject the null of no autocorrelation*.

The second diagnostic is done to check if the errors are normally distributed. There are many reasons why normality is important but most prominently, a normally distributed probability distribution makes the analysis easier and feasible. The true probability distribution is almost impossible to ascertain for the dataset but testing for normality in the errors helps capitalize on proven mathematical facts about the Central Limit theorem. The Jarque-Bera normality test can be done on STATA using the command “varnorm, jbera”. The output is a table that looks like the ones in the *Data Analysis section* and I can test the null hypothesis of normality using the p-values in the rightmost column again. A low p-value in this test indicates that there is evidence to reject the null of normally distributed errors in the respective equation.

Lastly, the third diagnostic checks for stability of our VAR model. The details of eigenvalue stability conditions are beyond the scope of this paper and will only be briefly discussed. In the realm of differential equations, solutions can be represented as summations of periodic contributions bounded by exponential functions. Eigenvalues represent the powers of these exponential functions (Dawkins 2019). The stability test can be run on Stata using the command “varstable” and the resulting table will clearly indicate whether all the eigenvalues lie inside the unit circle, an indication that the VAR satisfies the stability condition.

Back to Contents

Applying the Model to CCSI

Previously, in the empirical strategy section, I provided a general overview of the main DGP used to create the CCSI and using practical examples held a theoretical discussion of the VAR. Before I proceed onto the data analysis using the VAR, I will describe all the other endogenous variables that are theorized to influence climate skepticism and by extension, the VAR model. Lastly, this section will include a few hypothesis tests that will come in handy during the post-estimation phase of the analysis and help us in answering the key questions posed in this paper.

Variables: The CCSI will be the key contemporaneously endogenous variable in every model that is tested. This is obvious because we are trying to capture the effects and variations in climate skepticism with correspondence to other endogenous variables. To recap, the CCSI is a timeseries (spanning from 2000 to 2015), built using survey marginals of polling data to estimate the aggregate-level of climate skepticism in the US. Based on inspiration from Brulle et al. (2012), contemporary literature (see *Literature Review*) and other theorized propositions, I have grouped my data into five categories: recessionary economic indicators, political elite cues, scientific information, extreme weather, and media coverage and advocacy. Elite cues and structural economic variables will serve as my key explanatory variables while the rest will act as controls.

1. <u>Recessionary Economic Indicators</u>: An economic recession is not consistently defined by all economists and there is a debate on what factors should be considered as predictors and indicators. Kenton (2019) and others often cite an approximate definition of recession as "two consecutive quarters of negative economic growth as measured by the country's gross domestic product" (Kenton 2019). Moreover, economists often tout that weaknesses in

Real GDP Percent Change

	Percentiles	Smallest		
1%	-8.4	-8.4		
5%	-2.1	-4.4		
10%	-1	-2.3	Obs	64
25%	.6	-2.1	Sum of Wgt.	64
50%	2.2		Mean	1.95
		Largest	Std. Dev.	2.531641
75%	3.5	5.1		
90%	4.7	5.4	Variance	6.409206
95%	5.1	7	Skewness	-1.140739
99%	7.5	7.5	Kurtosis	6.442718

Unemployment Rate

	Percentiles	Smallest		
1%	3.9	3.9		
5%	4.033333	3.933333		
10%	4.433333	4	Obs	64
25%	4.966667	4.033333	Sum of Wgt.	64
50%	5.716667		Mean	6.301563
		Largest	Std. Dev.	1.780449
75%	7.766667	9.633333		
90%	9.3	9.633333	Variance	3.169998
95%	9.633333	9.833333	Skewness	.6800838
99%	9.933333	9.933333	Kurtosis	2.192759

Industrial Production Index

	Percentiles	Smallest		
1%	-5.6	-5.6		
5%	-2.8	-4.3		
10%	-1.3	-3.3	Obs	64
25%	-.15	-2.8	Sum of Wgt.	64
50%	.6		Mean	.1609375
		Largest	Std. Dev.	1.392453
75%	1	1.6		
90%	1.4	1.6	Variance	1.938926
95%	1.6	2	Skewness	-1.985469
99%	2	2	Kurtosis	7.735718

Total Nonfarm Payroll

	Percentiles	Smallest		
1%	-1.7	-1.7		
5%	-.7	-1.4		
10%	-.4	-1.2	Obs	64
25%	-.05	-.7	Sum of Wgt.	64
50%	.3		Mean	.146875
		Largest	Std. Dev.	.480069
75%	.5	.6		
90%	.6	.6	Variance	.2304663
95%	.6	.6	Skewness	-1.887373
99%	.6	.6	Kurtosis	6.812809

Dates of Recession

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	64
25%	0	0	Sum of Wgt.	64
50%	0		Mean	.15625
		Largest	Std. Dev.	.3659625
75%	0	1		
90%	1	1	Variance	.1339286
95%	1	1	Skewness	1.893459
99%	1	1	Kurtosis	4.585185

Price of Oil

	Percentiles	Smallest		
1%	1.197333	1.197333		
5%	1.433333	1.232		
10%	1.475667	1.432333	Obs	64
25%	1.678167	1.433333	Sum of Wgt.	64
50%	2.612333		Mean	2.560432
		Largest	Std. Dev.	.8299873
75%	3.354333	3.781667		
90%	3.689333	3.809	Variance	.688879
95%	3.781667	3.849	Skewness	.0357608
99%	3.901	3.901	Kurtosis	1.706355

Figure 3.1: Summary Stats for Recessionary Economic Indicators

industrial production and employment are historical indicators of an economic recession. Based on these observations, we've chosen five variables that can potentially capture the recessionary attributes of the economy. These include Real GDP (% change from preceding quarter), unemployment rate, industrial production index, total nonfarm payrolls (%change from preceding quarter), and a dummy variable for official recession dates (1 = quarter officially recorded as a recessionary period). *Figure 3.1* provides summary statistics on each of these timeseries variables. As you can see, *Figure 3.1* has a sixth variable, the price of oil. The inclusion of this variable is partly inspired by Brulle et al. (2012), but also because it is an important structural economic factor that has the potential of psychologically influencing a population. The *Appendix* contains specific information on the sources and recoding efforts of these variables.

Democrat Statements Favoring CC Action

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	1	0	Obs	64
25%	2.5	0	Sum of Wgt.	64
50%	5		Mean	10.4375
		Largest	Std. Dev.	11.36672
75%	14.5	34		
90%	27	38	Variance	129.2024
95%	34	43	Skewness	1.329278
99%	43	43	Kurtosis	3.830419

Republican Statements Opposing CC action

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	64
25%	0	0	Sum of Wgt.	64
50%	0		Mean	3.6875
		Largest	Std. Dev.	7.285286
75%	3	22		
90%	12	23	Variance	53.0754
95%	22	29	Skewness	2.627598
99%	35	35	Kurtosis	9.651506

House Hearings on Climate Change

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	64
25%	1.5	0	Sum of Wgt.	64
50%	4.5		Mean	6.265625
		Largest	Std. Dev.	6.69338
75%	9	19		
90%	14	23	Variance	44.80134
95%	19	27	Skewness	1.748667
99%	32	32	Kurtosis	6.279115

Senate Hearings on Climate Change

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	64
25%	2	0	Sum of Wgt.	64
50%	4		Mean	4.78125
		Largest	Std. Dev.	4.968053
75%	6	16		
90%	11	18	Variance	24.68155
95%	16	22	Skewness	1.917835
99%	22	22	Kurtosis	6.529546

LCV Dem Pro-Env. Score

	Percentiles	Smallest		
1%	.7514231	.7514231		
5%	.7514231	.7514231		
10%	.7777647	.7514231	Obs	64
25%	.8174419	.7514231	Sum of Wgt.	64
50%	.8446465		Mean	.8460453
		Largest	Std. Dev.	.0475323
75%	.884031	.9324028		
90%	.9084402	.9324028	Variance	.0022593
95%	.9324028	.9324028	Skewness	-.1382561
99%	.9324028	.9324028	Kurtosis	2.397081

LCV Rep. Anti-Env. Score

	Percentiles	Smallest		
1%	.0380359	.0380359		
5%	.0380359	.0380359		
10%	.0509456	.0380359	Obs	64
25%	.105027	.0380359	Sum of Wgt.	64
50%	.1270453		Mean	.1255761
		Largest	Std. Dev.	.0434647
75%	.1690227	.1802133		
90%	.1779352	.1802133	Variance	.0018892
95%	.1802133	.1802133	Skewness	-.5168059
99%	.1802133	.1802133	Kurtosis	2.306758

Figure 3.2: Summary Stats for Political Elite Cues

2. Political Elite Cues: Political elite cues are the second set of variables that will serve as key explanatory models in my data analysis. While the choice of these variables was inspired by Brulle et al. (2012) again, there is a strong rationale to include them based on contemporary research and my hypotheses tests that will be discussed next. Amongst commonly referenced

papers, McCright and Dunlap (2011) finds polarization along political lines and their empirical analysis shows that Democrats hold consistent beliefs with the scientific consensus while Republicans' beliefs represent a mismatch. The six variables in this category that I have chosen to test my hypothesis include: congressional hearing statements of Democrats favoring climate change action, congressional hearing statements of Republicans opposing climate change action, house hearings on climate change, senate hearing on climate change, league of conservation voters (LCV) Democrats pro-environmental score, and LCV Republicans anti-environmental score. *Figure 3.2* provides summary statistics on these timeseries variables. Again, more information including descriptions, sources and recoding efforts of these variables are available in the *Appendix*.

Typically, in most regression analyses, the researcher includes a range of control variables that help reduce the well-documented omitted variable bias in the model and improve the accuracy of the estimator. The multiple categories of data included as controls might tend to cause overparameterization of the model but there is sufficient research and theory backing correlations between climate skepticism and these variables. Moreover, as Enders (2011) states: "a VAR will be overparameterized in that many of these coefficients will be insignificant. However, the goal is to find the important interrelationships among the variables" (Enders 2011, 290). Thus, my approach is justified in continuing with these control variables as the prospect of finding crucial relationships between these series is greater than the risk of losing degrees of freedom and overparameterization. As I've already stated, please refer to the *Appendix* for further documentation on these variables. The remaining three categories of data are discussed below, and their summary statistics are available in the *Appendix*. All the variables data can be explored first-hand since the Stata and excel files to replicate the analysis are included in the paper.

3) Scientific Information: This category is self-explanatory as it helps factor in the influence of the scientific community's contribution toward climate change in the model. I include two variables here: popular scientific magazine articles on climate change and a dummy variable noting periods which saw a release of major scientific reports on climate change (1 = quarters when at least one such report was released).

4) Extreme Weather Conditions: Variables pertaining to extreme weather are one of the most commonly appearing in empirical papers. Researchers differ in what metrics they choose to use to capture the effects of this category but for my purposes I will stick with Brulle et al. (2012) and include: overall climate extremes index, US percentage areas (very warm), US percentage areas (very cold), and drought levels (using the Drought Severity Classification Index).

5) Media Coverage and Advocacy: Finally, the media coverage and advocacy category is a natural contender as a control variable due to its widespread influence on the public's perception of climate change. Since the nature and medium of media coverage has become so widespread, it makes sense to compile these various sources into one proxy variable. This is exactly what Brulle et al. (2012) does and I follow the same logic to build an additive index by compiling data from major television networks and weekly news magazines. Media advocacy is split into two segments to capture the "pro" and "anti" climate change sentiments often espoused by the competing media factions. These two segments are environmental magazines and conservative magazines on climate change.

❖ Hypothesis Tests

Given the vast dataset and a complex list of potential relationships between climate skepticism and the discussed variables, I make purposeful decisions to rein in the empirical strategy. Consequently, the Hypothesis tests are designed to narrow the focus on finding interdependencies between climate skepticism and the key explanatory variables falling in either the recessionary economic indicators category or political elite cues. Before turning to the analyses, I set up two hypothesis tests to represent and test my initial expectations based on existing research and theory.

The Role of Recessionary Factors in Inducing Climate Skepticism

Imagining a causal link between economic recessions and climate skepticism might seem arbitrary at first glance, but a deeper consideration of people's priorities and their short-sightedness on such issues might unearth these complex relationships. Researchers and scholars have provided reasonable evidence to believe that the Great Recession of 2008 negatively influenced public opinion on climate change (Scruggs & Benegal 2012). The threat perception of climate change and the issue salience of the matter is particularly low and public opinion surveys such as Gallup (2016) has stated that only about 1% polltakers name any environmental issue when asked to identify the most important problem facing the country. Thus, though my dataset includes only two major recessionary periods, I expect a delayed influence on skepticism since the negative effects and magnitude of recessions vary and take time to manifest in people's lives.

Hypothesis 1: Economic indicators of recession (such as unemployment rate, declining GDP, etc.) are more likely to heighten the aggregate-level climate skepticism in the US.

The Role of Partisanship and Political Elite Cues in Inducing Climate Skepticism

I've referenced multiple papers and sources that have researched and theorized the increased polarization along the lines of partisanship and political ideology. The divide between Democrats and Republicans on climate change existed back in the late 1980s but has grown larger over time and records as one of the most polarizing issues since the Great Recession (Egan & Mullin 2017, 218). Despite all the existing research, I hope to acquire a unique perspective from my model and understand the influence of political elites on the public mood regarding climate change. The interpretation of the results from this model will differ because I estimate trends between statements and voting patterns of Democrats and Republicans in the Congress with climate skepticism. For the sake of clarity and simplicity, I split the political elite cues hypotheses into two parts to capture unilateral relationships between these dichotomous variables and the CCSI.

<u>Hypothesis 2(a)</u>: Pro-climate change statements and voting patterns of Democratic Congresspersons are more likely to lower the aggregate-level climate skepticism in the US.

<u>Hypothesis 2(b)</u>: Anti-climate change statements and voting patterns of Republican Congresspersons are more likely to raise the aggregate-level climate skepticism in the US.

Back to Contents

Data Analysis

At the beginning of the Empirical Strategy section, I established a 5-step agenda to simplify the complex DGP that the thesis has undertaken and extrapolate empirical findings to shed light on the theoretical discussions and existing literature on climate skepticism. The *DGP overview* and the *Applying the Model to CCSI* sections fulfilled the first couple of steps in the agenda and this data analysis section will cover the rest. The methodology discussion of the CCSI, replication attempts of the VAR, and definition of the hypothesis tests provide critical information in understanding the data analysis and interpreting the results at the end. Given the arsenal of timeseries data that I have to predict the CCSI at my disposal, the challenge becomes how to avoid overparameterization of the model while extracting relevant knowledge pertaining to the interrelationships among variables. I split the analysis into 2 comprehensive models, one specified to identify any potential causal connections between recessionary economic indicators and climate skepticism. The second one serves a similar function but focuses on likely relationships between partisan politics and skepticism. Given that the VAR analysis has multiple pre and post-estimation steps attached to it, I will walkthrough each model separately, make specific observations of Stata's output for each step, and interpret the estimation results from both models together at the end of the section. The data analysis procedure for the VAR models can be summarized as follows:

1. Model Specification and Basic Setup
2. Pre-estimation steps:
    a. Stationarity test (Augmented Dickey-Fuller)
    b. Optimal lag length determination (AIC, etc.)
3. VAR model estimation
4. Post-estimation steps:
    a. Diagnostic Tests (Autocorrelation, Normality, and Stability tests)
    b. Causality Checks (Granger causality and Wald tests)

❖ Model 1: Economic Recessions and CCSI

Step 1 - Model Specification: To follow along, open the Stata file "VAR_CCSI.dta" and the do-file associated with Model 1, "Model1_Recession.ado". The do-file has helpful comments corresponding to each step but it's particularly useful for those seeking to replicate and run their own version of my VAR analyses. I discussed the challenges of imposing restrictions and specifying the model to best estimate the VAR. I know that the chosen variables in the reduced-form VAR need to pass stationarity tests before I can run the analysis. So, the question that needs to be answered is: "How do we specify the VAR and impose restrictions without knowing the stationarity conditions of the included timeseries variables?" This issue wasn't specifically addressed in the *DGP: Vector Autoregression* section because there are several ways to tackle it and exploring it deeply would detract from the focus of my thesis. In short, econometricians usually impose restrictions in two ways. The first one is a Choleski decomposition and is sometimes criticized for its *ad hoc* nature and diminishing the role of the economist to one that merely suggests appropriate variables to include in the VAR (Enders 2011, 313). The second one is called a structural decomposition and imposes restrictions by combining economic theory with vector analysis. For the purposes of this paper, I use the first approach since I'm focused on uncovering the potential relationships between recessionary indicators and skepticism. The standard forms of Model 1 are shown in *Figure 4.1* and the first four variables represent recessionary indicators while the next three are controls for media coverage, scientific information, and extreme weather. The standard practice is to order the key contemporaneously endogenous variable, i.e. the CCSI, last while the least endogenous variable, Recession Dates dummy, is ordered first in our reduced-form VAR. Stationarity tests and lag lengths of the model are discussed next.

Model 1 VAR (4th Order Model):

y = CCSI, x = Recession Dates, z = Unemployment rate, r = Real GDP, i = IPI

y_t (CCSI) = $a_{10} + a_{11}.y_{t-1} + a_{12}.y_{t-2} + a_{13}.y_{t-3} + a_{14}.y_{t-4} + a_{15}.x_{t-1} + a_{16}.x_{t-2} + a_{17}.x_{t-3} + a_{18}.x_{t-4} + a_{19}.z_{t-1} + a_{20}.z_{t-2} + a_{21}.z_{t-3} + a_{22}.z_{t-4} + a_{23}.r_{t-1} + a_{24}.r_{t-2} + a_{25}.r_{t-3} + a_{26}.r_{t-4} + a_{27}.i_{t-1} + a_{28}.i_{t-2} + a_{29}.i_{t-3} + a_{30}.i_{t-4}$ + *Lags*(Controls: Media Index) + *Lags(Controls: Scientific Reports)* + *Lags(Controls: CEI)* + e1t

Figure 4.1: Standard Form Equation for Model 1

Step 2(a) – Stationarity: The keen observer may have noticed that the Non-farm payrolls variable was excluded from our VAR and this is simply because it is nonstationary and failed the Augmented Dickey-Fuller (ADF) test. Stata commands for ADF tests are single-line commands that specify the timeseries and the chosen number of lags. Complete results of my ADF tests are shown in the data analysis section of the *Appendix* and I can move onto the next step since all the included variables are integrated of order 1 (I[1]), that is they are stationary after first difference. As a refresher to the *DGP: Vector Autoregression* section, the ADF test can be interpreted by comparing the p-value against the Dickey-Fuller critical values. Remember, I reject the null hypothesis of the presence of a unit root if the p-value is lower than the DF critical values (at least the 10% level). Since all the chosen variables pass the stationarity tests, I can proceed with my analysis.

Step 2(b) – Optimal Lag Length: The optimal lag length was given away when I specified the model using 4 lags. The results of Stata's *varsoc* command to determine the optimal lag

```
varsoc RecessionDates UnempRate RealGDP IPI SciReport MediaIndex CEI CCSI

 Selection-order criteria
 Sample:  2001q1 - 2015q4                    Number of obs      =        60
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-599.663				.086503	20.2554	20.3646	20.5347
1	-212.701	773.92	64	0.000	1.9e-06	9.49002	10.4731*	12.0032*
2	-143.636	138.13	64	0.000	1.8e-06	9.3212	11.1781	14.0684
3	-63.6477	159.98	64	0.000	1.4e-06*	8.78826	11.519	15.7694
4	14.1995	155.69*	64	0.000	1.7e-06	8.32668*	11.9312	17.5418

```
 Endogenous:  RecessionDates UnempRate RealGDP IPI SciReport MediaIndex CEI
              CCSI
  Exogenous:  _cons
```

Figure 4.2: varsoc command results in Stata

length is shown in *Figure 4.2* and I choose the AIC information criterion that specifies 4 lags. In practice, I run the *varsoc* command first on the specified model to identify the ideal number of lags and to include this number in my stationarity tests to run Augmented DF tests.

Step 3 - Model Estimation: When you run the "var" command as specified in the do-file, Stata spits out a total of 32 coefficients for *each* equation in the VAR. This isn't surprising because I already know that each variable appears on the left-hand side once and has the same set of regressors as the other timeseries variables. *Table 1* hosts only the *main* VAR estimation results (the equation with CCSI as the dependent variable) as I only care about the relationships between the CCSI and its regressors and want to learn if recessionary economic indicators are statistically significant and help predict the CCSI. The Stata table output is slightly different from my table here which is formatted differently to highlight key information. The Stata table provides coefficient estimates, standard errors, p-values, and 95% confidence intervals. A quick glance to check for p-values shows that only some of the hypothesized recession variables are statistically significant. The null hypothesis significance testing for this model and the partisan politics model are based on the hypothesis I set up in the *Applying the Model to CCSI* section. I defer the hypothesis testing and interpretation of these coefficients to the *VAR Model Interpretation and Discussion* section where I can holistically discuss both models in tandem with each other.

Step 4 (a) – Diagnostic Tests: I briefly discussed the diagnostic tests and called them "a form of post-mortem analysis of the VAR model." Without these tests, the discussions of my findings become baseless and will lack any conviction among econometricians. The three diagnostics listed in my do-file are: the Lagrange-multiplier test for autocorrelation (Stata command: *varlmar*), the Jarque Bera test for normality (Stata command: *varnorm, jbera*), and a test for stability (Stata command: *varstable*). The Lagrange-multiplier test for our model is shown in *Figure 4.3* and I fail to reject the null hypothesis of no autocorrelation at all lag orders since the chi-squared value is low and gives us a p-value that isn't significant at the 5% level. Overall, I can comfortably state that the Lagrange multiplier test finds no evidence of the existence of autocorrelation in Model 1. This result is a positive one because given the complexity of predicting the CCSI, my VAR model does well to escape the perils of autocorrelation. Potential reasons for the occurrence of autocorrelation and their drawbacks are discussed in the *DGP: Vector Autoregression* and *Appendix* sections.

The second diagnostic test is the Jarque-Bera test for normality of the errors or innovation terms in the VAR. Notice that in the do-file, I run the Stata commands for all the diagnostic tests following the *var* command, so Stata knows which regression to use. *Figure 4.4* displays the results of the test and observing the p-value column, I cannot reject the null hypothesis of normally distributed errors for all variables except Recession Dates. The last row of the table shows that the null cannot be rejected for all the error terms combined and this makes my estimation process easier and gives confidence in the interpretations.

```
. varlmar, mlag(4)
```

Lagrange-multiplier test

lag	chi2	df	Prob > chi2
1	73.8372	64	0.18759
2	68.0408	64	0.34143
3	80.6223	64	0.07836
4	67.2216	64	0.36737

H0: no autocorrelation at lag order

Figure 4.3: Lagrange-Multiplier Test Result

Jarque-Bera test

Equation	chi2	df	Prob > chi2
RecessionDates	9.370	2	0.00923
UnempRate	0.130	2	0.93724
RealGDP	1.056	2	0.58984
IPI	1.946	2	0.37790
SciReport	0.671	2	0.71486
MediaIndex	3.046	2	0.21806
CEI	0.056	2	0.97259
CCSI	0.145	2	0.93001
ALL	16.420	16	0.42406

Figure 4.4: Jarque-Bera Test Result

Finally, I test the stability condition of the model using the Eigenvalue stability condition. *Figure 4.5* lists all the eigenvalues and the modulus for the model and the most important observation here is that the VAR model satisfies the stability condition. Please refer to the *DGP: Vector Autoregression* and *Appendix* sections for more information on what this entails. Let's proceed onto the last step of running causality checks for the model.

```
varstable

 Eigenvalue stability condition

      Eigenvalue                    Modulus

   .9817121 +   .1503089i          .993152
   .9817121 -   .1503089i          .993152
   .9918256                        .991826
   .8784881 +   .4439502i          .984293
   .8784881 -   .4439502i          .984293
 -.02308477 +   .8942311i          .894529
 -.02308477 -   .8942311i          .894529
  -.4690816 +   .7571773i          .890705
  -.4690816 -   .7571773i          .890705
  -.8529977 +   .1368424i          .863904
  -.8529977 -   .1368424i          .863904
   .5475769 +    .645728i          .846643
   .5475769 -    .645728i          .846643
    .733557 +   .3924499i          .831939
    .733557 -   .3924499i          .831939
  -.2564336 +    .757835i          .800045
  -.2564336 -    .757835i          .800045
  -.7110046 +   .3512509i          .793035
  -.7110046 -   .3512509i          .793035
   .2571655 +   .7440755i          .787263
   .2571655 -   .7440755i          .787263
   .4888408 +   .5980164i          .772392
   .4888408 -   .5980164i          .772392
  .04814916 +   .7606343i          .762157
  .04814916 -   .7606343i          .762157
   .7133701                         .71337
  -.5613292 +  .05872573i          .564393
  -.5613292 -  .05872573i          .564393
  -.2593367 +   .2922743i          .390743
  -.2593367 -   .2922743i          .390743
   .2582666                        .258267
  -.1359189                        .135919

 All the eigenvalues lie inside the unit circle.
 VAR satisfies stability condition.
```

Figure 4.5: Eigenvalues – Stability Test Result

Step 4 (b) Causality Checks: The three basic ways to check for causal links between my key explanatory variables and the CCSI were discussed earlier. I've already acquired the p-value and the t-statistics of the included variables through the var estimation process and I can use this to test individual coefficients for causality. Second, the Granger causality test is particularly useful in jointly determining causality of all the variables in the equation. *Figure 4.6* is a table of the Granger causality test for this model and I can reject the null hypothesis of no Granger causality at the 5% significance level for all variables except unemployment rate and release of scientific reports. Together, as seen on the last row of the table, it can be stated that all the regressors Granger cause CCSI at the 1% significance level and any inferential arguments of causality during the results discussion ought to be contemplated seriously. Another common way for researchers to assess causality is to use the direction and classify it under a certain type of causality. The three common ones are unidirectional (x Granger causes y but y doesn't Granger cause x), bidirectional (if both x and y Granger cause one another), and independent (when neither x or y Granger cause each other). The complete excerpt of the Granger Causality Wald tests can be obtained by following the Do-file and executing the "vargranger" command. Interestingly, of the four explanatory variables, three possess unidirectional causality. So, I can conclude that at the 5% significance level and lower, lagged values of Recession Dates, Real GDP, and the Industrial Production Index *Granger cause* the CCSI.

vargranger

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
CCSI	RecessionDates	11.059	4	0.026
CCSI	UnempRate	5.0151	4	0.286
CCSI	RealGDP	9.7982	4	0.044
CCSI	IPI	22.332	4	0.000
CCSI	SciReport	3.7444	4	0.442
CCSI	MediaIndex	10.715	4	0.030
CCSI	CEI	21.563	4	0.000
CCSI	ALL	142.66	28	0.000

Figure 4.6: Granger Causality Wald Test Result

Lastly, to strengthen my conviction and add greater value to the results discussion I run the Wald tests on individual regressors, specifically the four recession indicators. *Figure 4.7* is a collage of all these linear tests of the parameter estimates and I notice that only the unemployment rate is not significant when tested individually but the rest of the variables are statistically significant and yield low p-values which reject the null hypothesis of no causality and implies statistical significance at the 5% level.

```
. test ([CCSI]: L.RecessionDates L2.RecessionDates L3.RecessionDates L4.RecessionDates)

 ( 1)  [CCSI]L.RecessionDates = 0
 ( 2)  [CCSI]L2.RecessionDates = 0
 ( 3)  [CCSI]L3.RecessionDates = 0
 ( 4)  [CCSI]L4.RecessionDates = 0

           chi2(  4) =   11.06
         Prob > chi2 =    0.0259

.
. test ([CCSI]: L.UnempRate L2.UnempRate L3.UnempRate L4.UnempRate)

 ( 1)  [CCSI]L.UnempRate = 0
 ( 2)  [CCSI]L2.UnempRate = 0
 ( 3)  [CCSI]L3.UnempRate = 0
 ( 4)  [CCSI]L4.UnempRate = 0

           chi2(  4) =    5.02
         Prob > chi2 =    0.2858
```

```
 test ([CCSI]: L.RealGDP L2.RealGDP L3.RealGDP L4.RealGDP)

 ( 1)  [CCSI]L.RealGDP = 0
 ( 2)  [CCSI]L2.RealGDP = 0
 ( 3)  [CCSI]L3.RealGDP = 0
 ( 4)  [CCSI]L4.RealGDP = 0

           chi2(  4) =    9.80
         Prob > chi2 =    0.0440


 test ([CCSI]: L.IPI L2.IPI L3.IPI L4.IPI)

 ( 1)  [CCSI]L.IPI = 0
 ( 2)  [CCSI]L2.IPI = 0
 ( 3)  [CCSI]L3.IPI = 0
 ( 4)  [CCSI]L4.IPI = 0

           chi2(  4) =   22.33
         Prob > chi2 =    0.0002
```

Figure 7: Wald Test Results Collage

❖ Model 2: Partisan Political Influence and CCSI

Step 1 – Model Specification: The same Stata file (“VAR_CCSI.dta”) can be used to follow along again and the do-file associated with Model 2 is: “Model2_PoliticsControls.ado”. The standard form equations for Model 2 are listed in *Figure 4.8.* The chosen lag lengths (4) corresponds with the optimal lag length tests which will be discussed in the next step. Another noticeable aspect of this model specification are the omitted control variables that were discussed in the *Applying the Model to CCSI* section. This is done purposefully as I ran multiple specifications to test for collinearity among the endogenous variables and unfortunately none of the series in the Media advocacy and coverage category made the cut as they failed the diagnostic tests. Since my focus is on the partisan influence of politics, I proceed with this reduced form VAR and analyze how well my model can predict the variance in the CCSI. The first 6 variables in the model represent my key explanatory variables that will be used to draw inferences in the results section.

Model 2 VAR (4th Order Model):

y = CCSI, x = Dem. Statements, z = Repub. Statements, r = LCV Dems Score, i = LCV Repub. Score, k= House Hearings, s = Senate Hearings

y_t (CCSI) = $a_{10} + a_{11}.y_{t-1} + a_{12}.y_{t-2} + a_{13}.y_{t-3} + a_{14}.y_{t-4} + a_{15}.x_{t-1} + a_{16}.x_{t-2} + a_{17}.x_{t-3} + a_{18}.x_{t-4} + a_{19}.z_{t-1} + a_{20}.z_{t-2} + a_{21}.z_{t-3} + a_{22}.z_{t-4} + a_{23}.r_{t-1} + a_{24}.r_{t-2} + a_{25}.r_{t-3} + a_{26}.r_{t-4} + a_{27}.i_{t-1} + a_{28}.i_{t-2} + a_{29}.i_{t-3} + a_{30}.i_{t-4} + a_{31}.k_{t-1} + a_{32}.k_{t-2} + a_{33}.k_{t-3} + a_{34}.k_{t-4} + a_{35}.s_{t-1} + a_{36}.s_{t-2} + a_{37}.s_{t-3} + a_{38}.s_{t-4}$ + *Lags*(Controls: US% Warm Areas) + *Lags(Controls: Scientific Reports)* + e2t

Figure 4.8: Standard Form Equation for Model 2

Step 2 (a) - Stationarity: The results of the ADF tests are shown in the *Appendix* and 7 of the 9 timeseries variables are stationary after first difference (i.e. Integrated of Order 1). This is still acceptable because I care if the stationarity conditions are met by the included series and the restrictions mandate that the VAR model is only constructed if this stationary condition is satisfied to avoid spurious regressions. After running the Dickey-Fuller commands on Stata, I can use the ADF results to test the null hypothesis of the presence of a unit root. I reject the null hypothesis of the presence of a unit root for all 9 series since the respective p-values are lower than the DF critical values.

Step 2 (b) - Optimal Lag Length Selection: The *varsoc* command on Stata returns the table output with the asterisks marking the chosen lag lengths under different information criterion. *Figure 4.9* displays these results and as I did with Model 1, I choose the AIC option of 4 lag lengths. This marks the completion of the pre-estimation steps and we proceed with the

varsoc DemsStatmnts RepubStatmnts HouseHearings SenateHearings LCVDems LCVRepubs USWarmAreas SciReport CCSI

Selection-order criteria
Sample: **2001q1 - 2015q4** Number of obs = **60**

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-374.795				2.9e-06	12.7932	12.9161	13.1073
1	-158.102	433.39	81	0.000	3.2e-08*	8.27005	9.49887*	11.4116*
2	-97.9869	120.23	81	0.003	7.8e-08	8.96623	11.301	14.9351
3	-14.4617	167.05	81	0.000	1.2e-07	8.88206	12.3228	17.6783
4	94.3259	217.58*	81	0.000	1.5e-07	7.9558*	12.5024	19.5794

Endogenous: DemsStatmnts RepubStatmnts HouseHearings SenateHearings LCVDems LCVRepubs USWarmAreas SciReport CCSI
Exogenous: _cons

Figure 4.9: Optimal Lag Length Test Result

estimation and post-estimation steps to round up this walkthrough.

Step 3 – Model Estimation: Again, the main *relevant* regression results are displayed in *Table 2*. Running the *var* command on Stata spits out a ton of equations with symmetrical regressors and coefficient estimates. *Table 2* only focuses on the CCSI and a glance at the table shows us that there are quite a few explanatory variables with statistically significant coefficient estimates. As mentioned in the Model 1 walkthrough, I defer my analysis and interpretation of the results to the *VAR Model Results and Interpretation* section.

Step 4 (a) – Diagnostic Tests: I run the same three diagnostic tests (Stata commands: [*varlmar*], [*varnorm, jbera*], and [*varstable*] as I did for Model 1 and discuss their implications here. First, the Lagrange-Multiplier test shown in *Figure 4.10* is similar to the one I conducted for Model 1. Again, this time around the p-values at all lag orders aren't low enough and I fail to reject the null hypothesis of no autocorrelation as a result. I can proceed with this positive result that the sample of my model doesn't show enough evidence to reject the null of no autocorrelation at all lag orders.

```
varlmar, mlag(4)

 Lagrange-multiplier test
```

lag	chi2	df	Prob > chi2
1	93.6558	81	0.15900
2	85.7075	81	0.33904
3	86.9555	81	0.30537
4	101.8985	81	0.05819

H0: no autocorrelation at lag order

Figure 4.10: Lagrange-multiplier Result

Second, the output of the Jarque-Bera (JB) test for normality of errors or innovations in the VAR model is available in *Figure 4.11*. The JB test results are not all positive and raise a few red flags especially with the Republicans LCV score. This series has low p-values and imply a rejection of the null hypothesis which assumes the presence of normally distributed errors. There is a positive tone to our results as I cannot reject the null for the rest of my variables, and I move on to the last diagnostic test with an awareness of the issues arising from the series with non-normally distributed innovations in the model.

```
varnorm, jbera

 Jarque-Bera test
```

Equation	chi2	df	Prob > chi2
DemsStatmnts	5.871	2	0.05310
RepubStatmnts	1.240	2	0.53806
HouseHearings	4.427	2	0.10934
SenateHearings	1.416	2	0.49271
LCVDems	1.016	2	0.60173
LCVRepubs	48.563	2	0.00000
USWarmAreas	1.415	2	0.49284
SciReport	1.624	2	0.44398
CCSI	0.147	2	0.92925
ALL	65.717	18	0.00000

Figure 4.11: Jarque-Bera test Result

Lastly, *Figure 4.12* lists the eigenvalues and the modulus for this model and once again, Stata tells us that all the eigenvalues are inside the unit circle and the VAR model satisfies the stability condition. This concludes the diagnostics for Model 2 and I move onto the penultimate step leading up to the results interpretation and discussion section.

varstable

Eigenvalue stability condition

Eigenvalue	Modulus
.8202381 + .4773718i	.949039
.8202381 - .4773718i	.949039
-.9451846	.945185
.9041822 + .2219715i	.93103
.9041822 - .2219715i	.93103
.9164868	.916487
-.3925537 + .8010098i	.892029
-.3925537 - .8010098i	.892029
.8851254 + .08559016i	.889254
.8851254 - .08559016i	.889254
-.7534544 + .4492786i	.877237
-.7534544 - .4492786i	.877237
.4185698 + .7534159i	.861879
.4185698 - .7534159i	.861879
.5868895 + .6178072i	.85213
.5868895 - .6178072i	.85213
.01567127 + .8439139i	.844059
.01567127 - .8439139i	.844059
.6808766 + .4806357i	.833429
.6808766 - .4806357i	.833429
.09365774 + .8274657i	.832749
.09365774 - .8274657i	.832749
-.0819743 + .8053043i	.809466
-.0819743 - .8053043i	.809466
-.6336358 + .3856001i	.741742
-.6336358 - .3856001i	.741742
-.236091 + .6543252i	.695615
-.236091 - .6543252i	.695615
.5144663 + .4644696i	.693114
.5144663 - .4644696i	.693114
-.6303124 + .2781568i	.688959
-.6303124 - .2781568i	.688959
-.55751 + .1776081i	.585117
-.55751 - .1776081i	.585117
-.05003814 + .3070108i	.311062
-.05003814 - .3070108i	.311062

All the eigenvalues lie inside the unit circle.
VAR satisfies stability condition.

Figure 4.12: Eigenvalues Stability Result

Step 4 (b) – Causality Checks: The p-values and t-statistics of individual coefficients in my model will be discussed in the next section. The Granger causality test is conducted in the same way as Model 1 and the results are displayed in *Figure 4.13.* The data from the causality test is promising as only one of the key explanatory variables (Senate Hearings) has a high p-value and I cannot reject the null hypothesis of no Granger causality. For the rest of the included variables, the null hypothesis is rejected and referencing back to the *DGP: Vector Autoregression* section, I know that Granger causality allows econometricians to measure whether past values of a series help predict the current value of the key dependent variable. For instance, in *Figure 4.13* the variables measuring LCV score for Democrats and Republicans are said to *Granger cause* CCSI as their past values help explain the current level of the CCSI. As I did for Model 1, the Granger Causality Wald test table gives us information to classify the nature of the causality shared between these variables. *Table 3* provides a summary of the different types of causality shared between the 6 key explanatory variables and the CCSI in Model 2. The unidirectional classification for the Senate Hearings variable might be confusing, but it's essentially telling us that CCSI Granger causes Senate Hearings. This is an intriguing finding that will be discussed more in the results section. Democratic statements, Republican statements, LCV Republicans, and House hearings all share a bi-directional causality with CCSI. This classification makes intuitive sense because past values of CCSI potentially result in more statements by Democrats and Republicans for or against climate change. Similarly, past values of CCSI seem to impact the voting patterns of Republicans and the number of House hearings on climate change. Another surprising result is the unidirectional causality of LCV Democrats (since CCSI doesn't Granger cause the LCV Democrats score). The optimist might claim this result suggests that the voting patterns of Democrats *stays consistent* regardless of the past levels of climate skepticism in the US. While this line of reasoning is fascinating, it is extremely hard to prove or verify given the complexity of the model and prematurely stating it as fact might end up in committing a type II error.

vargranger

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
CCSI	DemsStatmnts	10.707	4	0.030
CCSI	RepubStatmnts	19.757	4	0.001
CCSI	HouseHearings	15.602	4	0.004
CCSI	SenateHearings	6.865	4	0.143
CCSI	LCVDems	14.38	4	0.006
CCSI	LCVRepubs	15.942	4	0.003
CCSI	USWarmAreas	6.7525	4	0.150
CCSI	SciReport	8.5787	4	0.073
CCSI	ALL	112.61	32	0.000

Figure 4.13: Granger Causality Wald Test Results

Finally, I run the third causality check, a Wald test to perform linear hypothesis tests of the parameters in the model. *Figure 4.14* is a similar collage of all the Wald tests and I can immediately notice that these are in sync with the Granger causality results unlike my Model 1 test results. The test results show that I can reject the null of no causality at a 5% significance level for 5 of the 6 explanatory variables. The estimation results and any relationships between the Senate hearings variable and the CCSI should be interpreted with this awareness due to its unidirectional causality found through the Granger causality tests

```
. test ([CCSI]: L.DemsStatmnts L2.DemsStatmnts L3.DemsStatmnts L4.DemsStatmnts)

 ( 1)  [CCSI]L.DemsStatmnts = 0
 ( 2)  [CCSI]L2.DemsStatmnts = 0
 ( 3)  [CCSI]L3.DemsStatmnts = 0
 ( 4)  [CCSI]L4.DemsStatmnts = 0

           chi2(  4) =   10.71
         Prob > chi2 =    0.0301

. test ([CCSI]: L.RepubStatmnts L2.RepubStatmnts L3.RepubStatmnts L4.RepubStatmnts)

 ( 1)  [CCSI]L.RepubStatmnts = 0
 ( 2)  [CCSI]L2.RepubStatmnts = 0
 ( 3)  [CCSI]L3.RepubStatmnts = 0
 ( 4)  [CCSI]L4.RepubStatmnts = 0

           chi2(  4) =   19.76
         Prob > chi2 =    0.0006

. test ([CCSI]: L.HouseHearings L2.HouseHearings L3.HouseHearings L4.HouseHearings)

 ( 1)  [CCSI]L.HouseHearings = 0
 ( 2)  [CCSI]L2.HouseHearings = 0
 ( 3)  [CCSI]L3.HouseHearings = 0
 ( 4)  [CCSI]L4.HouseHearings = 0

           chi2(  4) =   15.60
         Prob > chi2 =    0.0036
```

```
. test ([CCSI]: L.SenateHearings L2.SenateHearings L3.SenateHearings L4.SenateHearings)

 ( 1)  [CCSI]L.SenateHearings = 0
 ( 2)  [CCSI]L2.SenateHearings = 0
 ( 3)  [CCSI]L3.SenateHearings = 0
 ( 4)  [CCSI]L4.SenateHearings = 0

           chi2(  4) =    6.87
         Prob > chi2 =    0.1432

. test ([CCSI]: L.LCVDems L2.LCVDems L3.LCVDems L4.LCVDems)

 ( 1)  [CCSI]L.LCVDems = 0
 ( 2)  [CCSI]L2.LCVDems = 0
 ( 3)  [CCSI]L3.LCVDems = 0
 ( 4)  [CCSI]L4.LCVDems = 0

           chi2(  4) =   14.38
         Prob > chi2 =    0.0062

. test ([CCSI]: L.LCVRepubs L2.LCVRepubs L3.LCVRepubs L4.LCVRepubs)

 ( 1)  [CCSI]L.LCVRepubs = 0
 ( 2)  [CCSI]L2.LCVRepubs = 0
 ( 3)  [CCSI]L3.LCVRepubs = 0
 ( 4)  [CCSI]L4.LCVRepubs = 0

           chi2(  4) =   15.94
         Prob > chi2 =    0.0031
```

Figure 4.14: Wald Linear Test Results Collage

and noncausality result from the Wald tests.

❖ VAR Model Results and Interpretation

Before I can begin interpreting the coefficients and drawing inferences from the VAR models, I revisit my agenda for the empirical strategy section and contextualize the discussions thus far. I broke this agenda down into five stages and have completed the first three: building an aggregate-level measure of climate skepticism (CCSI), finding a way to combine the CCSI with explanatory variables that represent our thesis (VAR), and assessing the quality of these results from the VAR models (pre and post-estimation steps: stationarity, autocorrelation, etc.). Technically, I began the fourth stage of interpreting the results when I discussed causality in the VAR model walkthrough. Now, I can extract the estimation results from both models, interpret the results and critically examine them to answer the key questions raised in the paper.

Recessionary Factors and Climate Skepticism in the US

In the build-up to the data analysis, I used existing literary findings and empirical research to set up a null hypothesis that stated: “Economic indicators of recession (such as unemployment rate, declining GDP, etc.) are more likely to heighten the aggregate-level climate skepticism in the US”. The Model 1 specification included four variables to estimate the influence of economic recessions on climate skepticism. Recession dates is my dummy variable that represents quarters officially classified as recession periods between 2000 - 2015. The Lag 4 coefficient of Recession Dates says that on average, the CCSI is 1.42 percentage-points lower in the current quarter if the T-4 period (four quarters ago) was undergoing a recession (versus a no-recession quarter) holding all other lagged values of Recession Dates and lagged values of other included variables constant. The p-value is significant at the 5% level so I can reject the null that this result was obtained by chance alone. The estimated standard error is small at .007. None of the other lags of Recession Dates is statistically significant. The Recession Dates dummy doesn’t provide much evidence and I cannot definitively state that an economic recessionary period will predict a significant shift in the CCSI. On the contrary, the Lag-4 coefficient suggests that in the aftermath of a recessionary quarter, the CCSI is likely to fall by 1.42 percentage-points. This result fails to reject the null hypothesis that recessionary economic factors *do not* heighten the CCSI.

Unemployment rate, Real GDP, and the Industrial Production Index have a combined total of 3 (out of 12) coefficients that have a statistically significant result that are worth considering; Lag 3 of Real GDP, Lag 3 and Lag 4 of IPI. Lag 3 of Real GDP has a coefficient of 0.0000648 that is significant at the 5% level and suggests that on average, the CCSI goes up by 0.006 percentage points in the current quarter for a 1% increase in real DGP in the T-3 period, holding all other included variables and their lags constant. Again, this coefficient doesn't provide much evidence to support the claim that real GDP growth and CCSI might be inversely related. The coefficient of Lag 3 of IPI says that on average, the CCSI falls by 1.2 percentage points in the current period when the IPI goes up by 1% in the T-3 period (or three quarters ago) holding all other included variables and lagged values constant. While this result may be interesting and suggests that IPI and CCSI are inversely related, my confidence is shaky because the result is predicting an outcome based on an extended time period (3 quarters). But, the Lag-3 coefficient of the IPI does provide some evidence to suggest that I can reject the null hypothesis. Moreover, as previously stated, the IPI has a narrower definition conceptually when compared to the GDP. The Lag-4 coefficient of IPI is smaller in magnitude relative to Lag-3 and the coefficient of 0.009 implies a positive impact on the CCSI. Since both these coefficients are highly statistically significant, these results do carry causal weight. But the extended timeline of the prediction coupled with inconsistent effects across Lags 3 and 4 result in a cautionary approach while rejecting the null hypothesis. Before I can comment on the causal links between economic recessionary variables and the CCSI, I will refer to the two other causality checks performed in the previous section.

The Granger Causality tests and Wald tests provide intriguing results that might contradict my interpretations of the OLS coefficients. The results of Model 1 might be a good example for why experienced econometricians use multiple causality checks and tools to determine causal links instead of relying solely on the VAR regression results. The results of these supplementary tests suggest that all key explanatory variables, barring the unemployment rate, are causally linked to the CCSI. Granted the Granger Causality only provides evidence on the direction of causality and not magnitude, the more important finding is that Recession Dates, Real GDP, and IPI all *Granger cause* CCSI. Both the

Granger and Wald tests allow me to reject the null that economic recessionary indicators aren't causally linked to the CCSI. While the VAR regressions results were a mixed bag, the Granger causality test and Wald test have given me a more definitive result. Rejecting the null for both these tests means that the sample suggests there is enough evidence to reject the null for the population. These results will be contextualized and given a definitive answer with regards to the hypothesis alongside the Model 2 results in the *Conclusion* section.

Partisanship / Political Cues and Climate Skepticism

There is extensive literature specifically regarding the connections between political ideology / partisan values and climate skepticism. In the *Applying the Model to CCSI* section, I created a two-part hypothesis to test claims like these: "opinion on global warming has become increasingly polarized across partisan and ideological lines since the 1990s" (McCright & Dunlap 2011, 178). The Model 2 specification and our data is capturing a slightly different relationship than most papers. The statements made by Democrats and Republicans in Congress and their voting patterns aren't representative of direct opinions or expressions of

Main VAR Regression Results:

VAR Results Table 1	Model 1	
Key Dependent Variable- **CCSI**	**Coefficients**	**Standard Errors**
Explanatory Variables		
Recession Dates		
Lag-1	0.0133	(0.00737)
Lag-2	0.000663	(0.00826)
Lag-3	-0.00980	(0.00790)
Lag-4	-0.0142*	(0.00724)
Unemployment Rate		
Lag-1	-0.271	(0.583)
Lag-2	0.161	(0.491)
Lag-3	0.236	(0.409)
Lag-4	0.498	(0.465)
Real GDP		
Lag-1	0.00000533	(0.0000214)
Lag-2	-0.0000236	(0.0000264)
Lag-3	0.0000648*	(0.0000256)
Lag-4	-0.0000351	(0.0000220)
IPI		
Lag-1	-0.00131	(0.00267)
Lag-2	0.00255	(0.00389)
Lag-3	-0.0126**	(0.00437)
Lag-4	0.00979***	(0.00284)
Controls		
Media Index		
Lag-1	-0.0000752	(0.0000432)
Lag-2	-0.0000159	(0.0000525)
Lag-3	0.000102*	(0.0000500)
Lag-4	0.0000811*	(0.0000402)
Scientific Reports on CC		
Lag-1	-0.000315	(0.00331)
Lag-2	-0.00426	(0.00355)
Lag-3	0.00174	(0.00349)
Lag-4	0.00383	(0.00364)
CEI		
Lag-1	-0.0346	(0.0198)
Lag-2	-0.0336	(0.0203)
Lag-3	-0.0389	(0.0214)
Lag-4	-0.0713**	(0.0230)
CCSI		
Lag-1	0.472***	(0.119)
Lag-2	0.0130	(0.131)
Lag-3	-0.107	(0.158)
Lag-4	-0.0330	(0.132)
constant:	0.129	(0.0994)
Sample Time:	2001 Q1 - 2015 Q4	
Observations:	60	
R^2 :	0.9522	
Standard errors in parentheses		
*p<0.05 **p<0.01 ***p<0.001		

VAR Results Table 2	Model 2	
Key Dependent Variable- **CCSI**	**Coefficients**	**Standard Errors**
Explanatory Variables		
Dem. Statements Pro-CC		
Lag-1	-0.000356	(0.000241)
Lag-2	-0.000445	(0.000245)
Lag-3	0.000277	(0.000246)
Lag-4	0.000420	(0.000263)
Repub. Statements Anti-CC		
Lag-1	0.000354	(0.000517)
Lag-2	0.000122	(0.000483)
Lag-3	0.00179***	(0.000462)
Lag-4	-0.000749	(0.000573)
LCV Dems. CC Voting Score		
Lag-1	0.0654	(0.0712)
Lag-2	-0.268**	(0.0979)
Lag-3	0.352**	(0.108)
Lag-4	-0.217*	(0.0994)
LCV Repubs. CC Voting Score		
Lag-1	0.217	(0.125)
Lag-2	-0.658***	(0.186)
Lag-3	0.308	(0.212)
Lag-4	0.0485	(0.172)
House Hearings on CC		
Lag-1	0.00186***	(0.000488)
Lag-2	0.000151	(0.000532)
Lag-3	-0.0000847	(0.000470)
Lag-4	0.0000640	(0.000409)
Senate Hearings on CC		
Lag-1	-0.000575	(0.000637)
Lag-2	-0.000765	(0.000709)
Lag-3	-0.000872	(0.000631)
Lag-4	0.000327	(0.000590)
Controls		
US (%) Very Warm Areas		
Lag-1	-0.00742	(0.0158)
Lag-2	0.0387*	(0.0152)
Lag-3	-0.00256	(0.0139)
Lag-4	-0.000636	(0.0133)
Scientific Reports on CC		
Lag-1	-0.00163	(0.00326)
Lag-2	0.00154	(0.00340)
Lag-3	0.00655*	(0.00309)
Lag-4	-0.00386	(0.00280)
CCSI		
Lag-1	1.169***	(0.144)
Lag-2	-0.567**	(0.203)
Lag-3	-0.0546	(0.183)
Lag-4	0.174	(0.103)
constant:	0.121	(0.0623)
Sample Time:	2001 Q1 - 2015 Q4	
Observations:	60	
R^2 :	0.9439	
Standard errors in parentheses		
*p<0.05 **p<0.01 ***p<0.001		

individuals with party affiliations. Since my CCSI is built using survey polls, it's likely that these individuals are the ones that *hold* these climate change beliefs which are classified as skeptical. The VAR model is intended to capture how these political "elites" that represent common interests in Congress can potentially predict the level of the CCSI at a given time. The two-part hypothesis suggested that: "Pro-climate change statements and voting patterns of Democratic Congresspersons are more likely to lower the aggregate-level climate skepticism in the US" and "Anti-climate change statements and voting patterns of Republican Congresspersons are more likely to raise the aggregate-level climate skepticism in the US".

The first of my six explanatory variables in *Table 2* is the pro-climate change statements made by Democrats (Stata label: DemsStatmnts) in Congress. None of the Lags of DemsStatmnts are statistically significant and thus these results don't provide much evidence in support of my hypothesis. The second variable is the anti-climate change statements made by Republicans in Congress (Stata label: RepubStatmnts). The only lagged value that is statistically significant here is Lag 3. The 0.0017 coefficient might appear to be relatively low magnitude but it suggests that on average, the CCSI goes up by .17 percentage points in the current quarter if the anti-climate change statement made by a Republican Congressperson increased by 1 in the T-3 period (3 quarters ago) holding all other variables and their lagged values constant. This result is consistent with the hypothesis and since it is statistically significant at the 1% level, I can reject the null hypothesis that Republican anti-climate change statements do not impact the CCSI. Thus, I can state that while pro-climate change statements made by Democrats aren't statistically significant, anti-climate change statements made by Republicans in Congress are effective and causally linked to the CCSI.

Before skipping to the LCV scores, I take a quick look at the House and Senate hearings. The House and Senate hearings were included to test the "political cues" aspect of the hypothesis and a quick glance at the respective coefficients shows that only Lag 1 of the

House hearings on climate change had a statistically significant result. This coefficient suggests that House hearings and CCSI are positively related and I can reject the null of no relationship between House hearings and CCSI for Lag-1. Finally, the last couple of explanatory variables are the LCV scores of Democrats and Republicans. To reiterate, the LCV scores are an aggregated average of individual Democrats (or Republicans) scores based on their votes for or against specific climate change related legislations in the House and Senate. A higher score indicates a voting pattern that is supportive of climate change action. For the LCV score of Democrats (Stata label: LCVDems), Lags 2 and 3 are highly statistically significant (at the 1% level) results. The Lag 2 coefficient of LCVDems tells us that on average, the CCSI falls by approximately 26 percentage points in the current period for a 1 percentage point increase in the LCV score in the T-2 period (2 quarters ago) holding all lagged values of LCVDems and other included variables constant. This is an intriguing finding and instantly demands further exploration. The estimated standard error is large and suggests that there is a high margin of error for this estimate. Barring the high estimated SE, my confidence in this result grows due to the postestimation causality checks mentioned previously. The Granger causality tests discussed in the previous section indicate that LCVDems *Granger causes* the CCSI. The Wald tests provided a similar result as the low p-value implies that I can confidently reject the null of no causality for LCVDems.

Lags 3 and 4 for LCVDems both produce significant results but the positive effect in Lag 3 raises a concern. Does the sign change imply a reverse effect compared to the other two lags? Without any further information or specialized knowledge about the complications of sign switching, I am forced to interpret it as a reactionary contradicting effect due to factors that isn’t completely knowable over the T-3 extended timeline. Again, this goes beyond the scope of the thesis and is nothing beyond an educated guess. The Lag 3 coefficient implies an average positive effect of 35 percentage-points on the CCSI in the current quarter for a 1% increase in LCVDems in the T-2 period, holding all other variables constant. The Lag 4 coefficient implies the same negative effect as Lag 2, but slightly lower at 21 percentage-points, for the T-4 period.

The last explanatory variable for Model 2 is the LCV score of Republicans (LCVRepubs) and only both Lag-2 produces statistically significant results. The coefficient of Lag 2 is perhaps the most fascinating result of the model as it implies that on average the CCSI falls by *65 percentage-points* in the current quarter for a 1%-point increase in LCVRepubs in the T-2 period, holding all other variables and their lags constant. While this result is significant at the 0.001 level, implying a one in a thousand chance of being wrong, the high estimated SE and the non-normally distributed errors (see JB test result) induce some caution about the magnitude of the effect. Regardless, the Lag-2 coefficient suggests a strong causal link between voting patterns of Republican Congresspersons and the CCSI. The Granger Causality and Wald tests produce consistent findings and further my conviction about a causal link between the Republican LCV scores and CCSI.

The *VAR Model Results and Interpretation* section broadly discussed the magnitude and causality of all my explanatory variables but didn't tie these results back to the initial hypotheses. *Table 3* is an attempt to consolidate the results of the three causality checks across both models for the key explanatory variables. In the next section, I will attempt to contextualize the vast information available in the two models, explain the results in tandem to the hypothesis tests, and provide a final answer to the main research question.

Despite multiple theoretical predictions and discussions of causal connections between economic recessions and CCSI, I only found a few meaningful connections in the model. A few important things to note are that the sample size for Model 1 only possesses two periods of extended recessionary effects. This might have lowered the probability of sniffing out any trends and patterns in the CCSI using a model specified with recessionary indicators. Furthermore, both models didn't include all the control series available in the dataset since the resulting VAR model failed diagnostic tests and rendered unreliable estimates. In the final few sections, I briefly discuss the economic and political significance of my findings and suggest a few avenues for further research on this fascinating topic.

Table 3: Overview of Causality Test Results

Causality Check Summary	Model 1			Causality Check Summary	Model 2		
Explanatory Variables	*Test #1: OLS P-Values*	*Test #2: Granger Causality*	*Test #3: Wald*	*Explanatory Variables*	*Test #1: OLS P-Values*	*Test #2: Granger Causality*	*Test #3: Wald*
Recession Dates	✓ Lag 4 (P < 0.05)	✓ (Unidirectional)	✓ (P < 0.05)	Democrats Statements	x	✓ (bidirectional)	✓ (P < 0.05)
Unemployment Rate	x	x	x	Republican Statements	✓ (P < 0.001)	✓ (bidirectional)	✓ (P < 0.001)
Real GDP	✓ Lag 3 (P < 0.05)	✓ (Unidirectional)	✓ (P < 0.05)	LCV Dems Score	✓ Lag 2 (P < 0.01) Lag 3 (P < 0.01) Lag 4 (P < 0.05)	✓ (unidirectional)	✓ (P < 0.01)
IPI	✓ Lag 3 (P < 0.01) Lag 4 (P < 0.001	✓ (Unidirectional)	✓ (P < 0.001)	LCV Repubs. Score	✓ Lag 2 (P < 0.001)	✓ (bidirectional)	✓ (P < 0.01)
				House Hearings	✓ (P < 0.001)	✓ (bidirectional)	✓ (P < 0.01)
Note: All Causality Tests Results summarized from Granger Causality Tests				Senate Hearings	x	x	x

Back to Contents

Conclusion

The famed theoretical physicist, Albert Einstein once said: "The grand aim of all science is to cover the greatest number of empirical facts by logical deduction from the smallest possible number of hypotheses or axioms" ("Albert Einstein", 2016). Obviously, I'm not a scientist and this paper doesn't talk about the theory of relativity. But my empirical strategy renders countless small and large-scale findings while attempting to answer a question based on two broad hypotheses. I'll aim for a conclusion that is succinct and gathers the key findings from the *Empirical Strategy* section, that is nothing short of an abyss filled with information.

First, I'll consider the VAR regression results from *Table 1* of the *Main VAR Regression Results* and the summary of the causality checks for Model 1 in *Table 3* to prove or disprove *Hypothesis 1*. The three causality checks for the four chosen explanatory variables, clearly indicates a trend and a potential relationship between recessionary factors and CCSI. The only variable that fails to show any significant result whatsoever is unemployment rate. Results from IPI provide the best claim to establish a causal link with the CCSI. The IPI variables has multiple lags (3 and 4) with highly statistically significant results across the three causality checks that give me enough evidence to reject the null hypothesis of no causality. The low p-values suggests the likelihood of my dataset given a true null hypothesis; i.e. economic recessionary factors *do not* explain an increase in the CCSI. Referring to *Table 3* again, I can reject this null for both the Recession Dates and the Real GDP as well since these variables produce consistent results across the three causality checks but at a lower level of significance compared to the IPI. For Hypothesis 1, I find a positive result with some mixed information but *find evidence to reject the null that economic recessions don't explain an upward trend in CCSI.*

Next, let's turn to the VAR regression results for Model 2 in *Table 2* of the *Main VAR Regression Results* and the summary of the causality checks in *Table 3* in order to prove or disprove the Hypothesis 2 (a & b). The null hypothesis for 2 (a) states that pro-climate change statements and voting patterns of Democratic Congresspersons has *no impact* on lowering the CCSI. The Democratic statements variable produces mixed results since the regression doesn't provide any lags with significant results and the p-values for Granger Causality and Wald tests are significant at the 5% level. Based on this mixed result, I don't find enough evidence to reject the null hypothesis. Next, the LCV Dems variable, representing Democrats' voting patterns, has three coefficients at lags 1, 2, and 3 that are statistically significant, and their magnitude indicates a strong negative relation with the CCSI. Since, the other two causality checks give consistent statistically significant results, there is enough evidence to reject the null hypothesis. Again, the low p-values for LCV Dems across all causality checks suggest my sample provides enough evidence to reject the null for the target population (U.S. public). For Hypothesis 2 (a), I find a mixed result where *I have little confidence to reject the null that pro-climate change statements made by Democrats don't explain a negative trend in CCSI, but find enough evidence to reject the null that pro-climate change voting patterns of Democrats don't explain a fall in the CCSI.*

Finally, I can use the same results from Model 2 to test the null hypothesis of 2 (b) which states that anti-climate change statements have no impact on raising the CCSI and pro-climate change voting patterns of Republican Congresspersons has no impact on lowering the CCSI. The Lag 3 coefficient of Republican statement is highly statistically significant ($p < 0.001$) and positively affects the CCSI. Moreover, the causality checks indicate consistent results across all tests, and this is sufficient evidence to reject the null. Lastly, the LCV Republicans score will help understand the existence of any causal effects between voting patterns of Republican Congresspersons and the CCSI. *Table 3* clearly indicates that all three causality checks produce statistically significant results, and this is enough evidence to reject the null hypothesis. A key observation from the *Model Results and Interpretation* pertained to the magnitude and direction of causality for the Lag-2 coefficient of LCV Republicans. The negative effect might seem confusing, but it makes sense since the LCV score allots higher scores for pro-climate change voting patterns. One interpretation is that a higher LCV score for Republicans is *more effective* in lowering the CCSI than a higher LCV score for Democrats. For Hypothesis 2 (b), I find a consistent result where *I can reject the null that anti-climate change statements made by Republican elites don't explain a positive trend in CCSI, and also find enough evidence to reject the null that pro-climate change voting patterns of Republicans don't explain a fall in the CCSI.*

Main Takeaways: Before I answer the main research question, here is a shortlist of insights and empirical results that I find the most intriguing:

1. I initially expected more evidence from the Recession Dates and Unemployment Rate variables. But it was IPI, with relevant coefficients, that helps explain the variance of the CCSI and bolster the case for a causal link between economic recessions and climate skepticism.
2. House hearings share a causal link with the CCSI and are more relevant in explaining the variance in the CCSI than Senate hearings. The bidirectional causality for House Hearings and CCSI (*Table 3*) implies that there is evidence to suggest that higher levels of CCSI has a causal impact on the number of house hearings too.
3. Republican anti-climate change statements share a more consistent causal link with the CCSI than Democratic pro-climate change statements (*Table 3*). One potential theory why this might be the case is because the data used to construct the CCSI comprises of skeptical public responses belonging to people from conservative backgrounds or identify as Republican.

4. The unidirectional causality of LCV Democrats suggests that changes in CCSI do not influence the voting patterns of Democrats. Contrary to political statements, voting patterns of Democrats and Republicans are equally influential in explaining the CCSI. The higher magnitude impact in mean changes of the CCSI with respect to LCV Republicans suggests that pro-climate change voting from Republicans is more effective, since it might elicit a stronger decline in CCSI compared to Democrats' voting patterns.

Final Answer to Research Question

Given the target time period of Q1 2000 to Q4 2015, the key decisions made to compute the CCSI, and including/excluding variables to represent recessions and partisanship, there is some evidence to back the claim that economic recessions explain climate skepticism. But, the VAR results for Model 1 do not show consistent patterns and strong effects across all recession variables. The VAR results from Model 2 identify a more systematic pattern with a larger magnitude impact on the CCSI. Thus partisanship, represented through the lens of political elite cues, emerges as a clear winner in explaining the variance and movement in aggregate-level climate skepticism in the U.S.

Avenues for further research

While there are many creative ways one could extend this thesis or reframe the research question, two potential areas that could be interesting for further research are:

1. Studying interaction effects and analyzing the CCSI with greater context on the demographics of the underlying population. One way to do this might be to build state-wise CCSI based on local and state survey questions and running data analysis on these aggregates, but more focused measures of climate skepticism.
2. Introducing an element of forecasting which uses the VAR regression results to make predictions of possible influencers of climate skepticism. This could help with public policy recommendation and can refocus the problem of climate skepticism from an "agnotology" perspective.

Back to Contents

Citations

"Albert Einstein." *Albert Einstein - Oxford Reference*, 31 Oct. 2016, www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00003988.

Barreto, Humberto, and Frank Howland. "Introductory Econometrics." 2006, doi:10.1017/cbo9780511809231.

Brulle, Robert J. "Institutionalizing Delay: Foundation Funding and the Creation of U.S. Climate Change Counter-Movement Organizations." *Climatic Change*, vol. 122, no. 4, 2013, pp. 681–694., doi:10.1007/s10584-013-1018-7.

Brulle, Robert J., et al. "Shifting Public Opinion on Climate Change: an Empirical Assessment of Factors Influencing Concern over Climate Change in the U.S., 2002–2010." *Climatic Change*, vol. 114, no. 2, 2012, pp. 169–188., doi:10.1007/s10584-012-0403-y.

Dawkins, Paul. "Eigenvalues and Eigenfunctions." *Differential Equations - Eigenvalues and Eigenfunctions*, 2019, tutorial.math.lamar.edu/Classes/DE/BVPEvals.aspx.

Dickey, David A., and Wayne A. Fuller. "Distribution of the Estimators for Autoregressive Time Series With a Unit Root." *Journal of the American Statistical Association*, vol. 74, no. 366, 1979, p. 427., doi:10.2307/2286348.

Egan, Patrick J., and Megan Mullin. "Climate Change: US Public Opinion." *SSRN*, 15 May 2017, papers.ssrn.com/sol3/papers.cfm?abstract_id=2968069.

Enders, Walter. *Applied Econometric Time Series*. Wiley, 2011.

Guber DL. 2013. A cooling climate for change? Party polarization and the politics of global warming. Am. Behav. Sci. 57:93–115

Hamilton LC,Hartter J, Lemcke-StamponeMD, MooreW, StaffordTG. 2015. Tracking public beliefs about anthropogenic climate change. PLOS ONE 10:e0138208. doi:10.1371/journal.pone.0138208

Jacques PJ, Dunlap RE, Freeman M. 2008. The organization of denial. Environ. Polit. 17:349–85

Kenton. "Recession Definition." *Investopedia*, Investopedia, 6 May 2019, www.investopedia.com/terms/r/recession.asp.

Lambert, Ben. (2013, September 11). *Stationary Series Summary* [Video file]. Retrieved from https://www.youtube.com/watch?v=ZIWyGjrAlks

Malka A, Krosnick JA, Langer G. 2009. The association of knowledge with concern about global warming: trusted information sources shape public thinking. Risk Anal. 29:633–47

McCright AM, Dunlap RE. 2011. The politicization of climate change and polarization in the American public's views of global warming, 2001–2010. Sociol. Q. 52:155–94

Mcgann, Anthony J. "Estimating the Political Center from Aggregate Data: An Item Response Theory Alternative to the Stimson Dyad Ratios Algorithm." *Political Analysis*, vol. 22, no. 1, 2014, pp. 115–129., doi:10.1093/pan/mpt022.

Powell, James Lawrence. "Anatomy of Denial." *The Inquisition of Climate Science*, 2012, pp. 170–179., doi:10.7312/columbia/9780231157193.003.0015.

Scruggs, Lyle, and Salil Benegal. "Declining Public Concern about Climate Change: Can We Blame the Great Recession?" *Global Environmental Change*, vol. 22, no. 2, 2012, pp. 505–515., doi:10.1016/j.gloenvcha.2012.01.002.

Stimson, James A. "The Dyad Ratios Algorithm for Estimating Latent Public Opinion." *Bulletin of Sociological Methodology/Bulletin De Méthodologie Sociologique*, vol. 137-138, no. 1, 2017, pp. 201–218., doi:10.1177/0759106318761614.

Tsay, R. *Lecture 11: Unit Root and Unit-Root Test.* Bus 41910, Time Series Analysis, University of Chicago 2008.

Back to Contents

Appendix

4. Empirical Strategy - DGP: Climate Change Skeptic Index

Table S1: Climate Change Skeptic Index Input Data (2000 – 2015)

S.No	Variable Name	Date	Summary Score (% Sceptic)	Sample Size
1	GallupClimateWorry	4/9/2000	12	72000
2	GallupSeriousness	12/31/2000	30	72000
3	GallupPersonalWorry	3/5/2001	5	72000
4	GallupClimateWorry	3/5/2001	13	72000
5	GallupGWEffects	3/5/2001	7	72000
6	GallupGWNews	3/5/2001	30	72000
7	GallupGWScientists	3/5/2001	4	72000
8	USPSRA	4/26/2001	14	1202
9	USPS	4/26/2001	25	1202
10	USGALLU	6/1/2001	40	1011
11	USCBSNYT	6/20/2001	17	1050
12	USCBSNYTT	6/20/2001	22	1050
13	USCBSNT	6/20/2001	32	1050
14	USPSRAA	10/24/2001	13	1281
15	GallupPersonalWorry	3/5/2002	6	72000
16	GallupClimateWorry	3/5/2002	17	72000
17	GallupGWEffects	3/5/2002	9	72000
18	GallupGWNews	3/5/2002	31	72000
19	USWASHP	7/8/2002	19	1402
20	GallupPersonalWorry	3/5/2003	10	72000
21	GallupClimateWorry	3/5/2003	17	72000
22	GallupGWEffects	3/5/2003	10	72000
23	GallupGWNews	3/5/2003	33	72000
24	GallupGWHumanActs	3/5/2003	33	72000
25	GallupPersonalWorry	3/5/2004	7	72000
26	GallupClimateWorry	3/5/2004	19	72000
27	GallupGWEffects	3/5/2004	11	72000
28	GallupGWNews	3/5/2004	38	72000
29	USGREEN	4/13/2004	10	1610
30	USUMARY	6/25/2004	23	753
31	USUMARY1	6/25/2004	54	753
32	USUMARY2	6/25/2004	19	753
33	USUMARY3	6/25/2004	29	753
34	USUMARY4	6/25/2004	30	753
35	USPSRAA	8/18/2004	12	2009
36	GallupSeriousness	12/31/2004	38	72000
37	USUMARY5	1/18/2005	29	801
38	GallupPersonalWorry	3/5/2005	6	72000
39	GallupGWEffects	3/5/2005	9	72000
40	GallupGWNews	3/5/2005	31	72000
41	USUMARY	7/5/2005	21	812
42	USUMARY1	7/5/2005	44	812
43	USUMARY6	7/5/2005	28	812
44	USUMARY7	7/5/2005	6	812
45	USUMARY8	7/5/2005	13	812
46	USUMARY9	7/5/2005	28	812
47	USPSRAA	11/17/2005	10	2006
48	USABCWPPP	1/29/2006	35	1002
49	USPSRAAA	2/28/2006	8	2000
50	USPSRAAAA	2/28/2006	14	2000
51	USPSR	2/28/2006	6	2000
52	USPSRAAAAA	2/28/2006	24	2000
53	USPSRA7	2/28/2006	18	2000
54	USPSRA8	2/28/2006	26	2000
55	GallupPersonalWorry	3/5/2006	5	72000
56	GallupClimateWorry	3/5/2006	15	72000
57	GallupGWEffects	3/5/2006	8	72000
58	GallupGWNews	3/5/2006	30	72000
59	GallupGWHumanActs	3/5/2006	36	72000
60	GallupGWScientists	3/15/2006	3	72000
61	USCBSNYT	5/9/2006	30	1241
62	USCBSNYTTT	5/9/2006	6	1241
63	USIPSOSRR	12/22/2006	9	1000
64	USNBCWSJ	1/29/2007	33	1007
65	GallupPersonalWorry	3/5/2007	6	72000
66	GallupClimateWorry	3/5/2007	16	72000
67	GallupGWEffects	3/5/2007	8	72000
68	GallupGWNews	3/5/2007	33	72000
69	GallupGWHumanActs	3/5/2007	35	72000
70	USGALLUP	4/19/2007	26	1007
71	USGALLUPP	4/19/2007	30	1007
72	USCBSNYT	4/26/2007	12	1052
73	USCBSNYTT	4/26/2007	20	1052
74	USCBSNYTTT	4/26/2007	3	1052
75	USCBSNY	4/26/2007	9	1052
76	USCBSNYTTTT	4/26/2007	9	1052
77	USORCCC	5/9/2007	19	1028
78	USORC	5/31/2007	42	1028
79	USIPSOSR	6/30/2007	7	1001
80	USICR	7/1/2007	20	2140
81	USPSRNEW	8/31/2007	39	1002
82	USPSRNEW1	8/31/2007	42	1002
83	USPSRNEW2	8/31/2007	10	1002
84	USPSRNEW3	8/31/2007	42	1002
85	USPSRNEW4	8/31/2007	13	1002
86	USPSRNEW5	8/31/2007	17	1002
87	USPSRNEW6	8/31/2007	18	1002
88	USPSRNEW7	8/31/2007	17	1002
89	USIPSOSR	9/26/2007	20	1001
90	USCBS	10/18/2007	15	1282
91	USCBSNYT	12/31/2007	15	1133
92	GallupPersonalWorry	3/5/2008	7	72000
93	GallupClimateWorry	3/5/2008	17	72000
94	GallupGWEffects	3/5/2008	11	72000
95	GallupGWNews	3/5/2008	35	72000
96	GallupGWScientists	3/5/2008	7	72000
97	GallupGWHumanActs	3/5/2008	38	72000
98	USSRBI	5/8/2008	24	1502
99	USORC	6/6/2008	45	1035
100	USPSRNEW8	6/20/2008	19	1010

S.No	Variable Name	Date	Summary Score (% Sceptic)	Sample Size
101	USSRBI3	8/5/2008	18	1502
102	USABCWPPPP	12/20/2008	20	1003
103	USCBS	2/28/2009	22	864
104	USCBSSS	2/28/2009	32	864
105	GallupPersonalWorry	3/5/2009	7	72000
106	GallupClimateWorry	3/5/2009	20	72000
107	GallupGWEffects	3/5/2009	16	72000
108	GallupGWNews	3/5/2009	41	72000
109	USPAF	4/3/2009	27	1001
110	USPAFF	4/3/2009	37	1001
111	USPAFFF	4/3/2009	9	1001
112	USPAFFFF	4/3/2009	12	1001
113	USORCC	5/5/2009	17	2019
114	USPSRA6	7/9/2009	13	2001
115	USSRBI	10/22/2009	32	1500
116	USSRBI1	10/22/2009	5	1500
117	USSRBI	10/31/2009	32	1500
118	USABCWP	11/24/2009	17	1001
119	USSRBI2	12/3/2009	51	2000
120	USORC	12/7/2009	54	1041
121	USORCCCC	12/7/2009	24	1041
122	USCBSNY	12/14/2009	27	1031
123	USABCWPP	12/18/2009	62	1003
124	USABCWPPPPP	12/18/2009	29	1003
125	GallupPersonalWorry	3/5/2010	7	72000
126	GallupClimateWorry	3/5/2010	29	72000
127	GallupGWEffects	3/5/2010	19	72000
128	GallupGWNews	3/5/2010	48	72000
129	GallupGWScientists	3/5/2010	10	72000
130	GallupGWHumanActs	3/5/2010	46	72000
131	USCBSNYT	4/14/2010	29	1580
132	USVIRGCU	5/27/2010	42	1001
133	USVIRGCU1	5/27/2010	45	1001
134	USVIRGCU2	5/27/2010	49	1001
135	USGALLUPPP	6/30/2010	20	1014
136	USCBS	8/31/2010	26	847
137	USCBS	10/31/2010	25	1253
138	GallupSeriousness	12/31/2010	48	72000
139	GallupPersonalWorry	3/5/2011	7	72000
140	GallupClimateWorry	3/5/2011	28	72000

S.No	Variable Name	Date	Summary Score (% Sceptic)	Sample Size
141	GallupGWEffects	3/5/2011	18	72000
142	GallupGWNews	3/5/2011	43	72000
143	GallupGWScientists	3/5/2011	8	72000
144	GallupGWHumanActs	3/5/2011	43	72000
145	USCBS	4/30/2011	21	1021
146	USORC	9/15/2011	51	1038
147	USCBSNY	9/16/2011	12	1566
148	GallupPersonalWorry	3/5/2012	7	72000
149	GallupClimateWorry	3/5/2012	23	72000
150	GallupGWEffects	3/5/2012	15	72000
151	GallupGWNews	3/5/2012	42	72000
152	GallupGWScientists	3/5/2012	7	72000
153	GallupGWHumanActs	3/5/2012	41	72000
154	USCBSNYT	6/30/2012	27	990
155	USCBSNYT	6/30/2012	25	990
156	USCBSS	10/31/2012	9	1132
157	USPRRI	12/13/2012	34	1018
158	USORC	1/31/2013	47	814
159	USCBSS	1/31/2013	10	1052
160	GallupPersonalWorry	3/5/2013	8	72000
161	GallupClimateWorry	3/5/2013	23	72000
162	GallupGWEffects	3/5/2013	15	72000
163	GallupGWNews	3/5/2013	41	72000
164	GallupGWScientists	3/5/2013	6	72000
165	GallupGWHumanActs	3/5/2013	39	72000
166	GallupGWNatural	3/10/2013	40	72000
167	USCBS	4/30/2013	21	977
168	USCBS	4/30/2013	15	977
169	USCBSS	4/30/2013	10	977
170	USCBSS	4/30/2013	8	977
171	GallupSeriousness	12/31/2013	41	72000
172	YPCCCtaxdividendOppose	2/15/2014	24	13000
173	YPCCCCO2limitsOppose	2/15/2014	34	13000
174	YPCCCregulateOppose	2/15/2014	23	13000
175	GallupPersonalWorry	3/5/2014	10	72000
176	GallupClimateWorry	3/5/2014	24	72000
177	GallupGWEffects	3/5/2014	18	72000
178	GallupGWNews	3/5/2014	42	72000
179	GallupGWScientists	3/5/2014	8	72000
180	GallupGWHumanActs	3/5/2014	40	72000
181	GallupPersonalWorry	3/5/2015	10	72000
182	GallupClimateWorry	3/5/2015	24	72000
183	GallupGWEffects	3/5/2015	16	72000
184	GallupGWNews	3/5/2015	42	72000
185	GallupGWScientists	3/5/2015	8	72000
186	GallupGWHumanActs	3/5/2015	41	72000
187	GallupSeriousness	12/31/2015	42	72000

Table S2: CCSI Iteration History (Provided by WCALC)

Iteration	Convergence	Criterion	Items	Reliability	AlphaA	AlphaB
1	0.2296	0.001	19	0.616	0.568	0.656
2	0.021	0.001	19	0.715	0.501	0.659
3	0.0037	0.001	19	0.72	0.501	0.663
4	0.0005	0.001	19	0.721	0.5	0.664

Table S3: Threat Index Loadings and Descriptive Variable Information

Variable Loading	Variable Name	Cases	Dimension 1 Loading	Mean	Standard Deviation
1	GallupClimateWorry	15	0.98	19.8	5.009
2	GallupSeriousness	4	0.99	39.25	6.457
3	GallupPersonalWorry	15	0.435	7.2	1.6
4	GallupGWEffects	15	0.952	12.667	4.044
5	GallupGWNews	15	0.967	37.333	5.594
6	GallupGWScientists	9	0.911	6.778	2.043
7	USCBSNYT	6	0.774	21.5	7.089
8	USCBSNYTT	2	-1	21	1
9	USPSRAA	3	0.967	11.667	1.247
10	GallupGWHumanActs	10	0.915	39.2	3.682
11	USUMARY	2	1	22	1
12	USUMARY1	2	1	49	5
13	USCBSNYTTT	2	1	4.5	1.5
14	USCBSNY	3	0.809	16	7.874
15	USORC	5	0.999	47.8	4.261
16	USIPSOSR	2	1	13.5	6.5
17	USCBS	6	0.948	21.167	3.804
18	USSRBI	2	1	28	4
19	USCBSS	3	0.88	9.333	0.471

Dimension 1 Information:
Eigen Estimate: 1.53 of possible 1.89
Percentage Variance Explained: 80.73
Weighted Average Metric: Mean - 23.63, Std. Dev - 3.80

Table S4: Descriptive Statistics of Data Used in Analysis

Variables	Measures	Mean	Standard Deviation	Minimum	Maximum
	CCSI	23.62745	3.83207	16.669	32.154
Economic Recessionary Data	Recession Dates	0.15625	0.3659625	0	1
	Unemployment Rate	6.35938	1.80271	4	10
	Real GDP (constant 2012)	15205.02	1257.655	12924.18	17456.22
	Industrial Production Index	98.04934	4.911275	87.5984	106.3359
	Total Nonfarm Payrolls	134279.6	3550.869	129804	143125
	Price of Oil	2.5604	0.82998	1.1973	3.901
Partisanship / Political Elite Cues	Democrat Pro-CC Statements	10.4375	11.3667	0	43
	Republican Anti-CC Statements	3.6875	7.285286	0	35
	LCV Democrats Score	84.60453	4.75323	75.14231	93.24028
	LCV Republicans Score	12.55761	4.34647	3.80359	18.02133
	House Hearings on CC	6.265625	6.69338	0	32
	Senate Hearings on CC	4.78125	4.968053	0	22
Controls: Extreme Weather	US % Warm Areas	19.47042	11.80179	1.14	46.56
	US % Cold Areas	4.1663	6.41102	0	35.13667
	Climate Extremes Index	21.33281	7.67746	8.58	45.26
	Drought Levels	109.9875	37.20504	39.76923	207.3077
Controls: Media	Media Coverage Index	115.3125	54.84549	31	314
	Environmental Magazines	11.46875	5.887756	2	27
	Conservative Magazines	4.40625	3.910583	0	20
Controls: Scientific Information	Science Magazines	78.45313	37.19273	23	163
	Release of Major Scientific Report	0.546875	0.5017331	0	1

Table S5: Survey Questions Used in Construction of CCSI (by WCALC)

Full Question Text	Variable Name	Dates Administered	Source
I'm going to read you a list of environmental problems. As I read each one, please tell me if you personally worry about this problem a great deal, a fair amount, only a little or not at all. First, how much do you personally worry about: Global Warming & Climate Change	GallupClimateWorry	April 2000, March: 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015	Gallup Organization
Is the seriousness of global warming generally exaggerated, generally correct, generally underestimated?	GallupSeriousness	December: 2000, 2004, 2010, 2013	Gallup Organization
Next, I'm going to read a list of problems facing the country. For each one, please tell me if you personally worry about this problem a great deal, a fair amount, only a little or not at all? How much do you personally worry about the quality of the environment?	GallupPersonalWorry	March: 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015	Gallup Organization

Which of the following statements reflects your view of when the effects of global warming will begin to happen -- they have already begun to happen, they will start happening within a few years, they will start happening within your lifetime, they will not happen within your lifetime, but they will affect future generations (or) they will never happen?	GallupGWEffects	March: 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015	Gallup Organization
Thinking about what is said in the news, in your view is the seriousness of global warming --generally exaggerated, generally correct or is it generally underestimated?	GallupGWNews	March: 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015	Gallup Organization
Just your impression, which one of the following statements do you think is most accurate -- most scientists believe that global warming is occurring, most scientists believe that global warming is NOT occurring or most scientists are unsure about whether global warming is occurring or not?	GallupGWScientists	March 2001, 2006, 2008, 2010, 2011, 2012, 2013, 2014, 2015	Gallup Organization
Do you think global warming is an environmental problem that is causing a serious impact now, or do you think the impact of global warming won't happen until sometime in the future, or do you think global warming won't have a serious impact at all?	USCBSNYT	June 2001, April 2007, December 2007, April 2010, June 2012, May 2006	CBS News/New York Times
Do you think it is necessary to take steps to counter the effects of global warming right away, or isn't it necessary to take steps yet?	USCBSNYTT	June 2001 and April 2007	CBS News/New York Times
(As I read a list of possible long-range foreign policy goals which the United States might have, tell me how much priority you think each should be given. Do you think this should have top priority, some priority, or no priority at all?)... Dealing with global warming	USPSRAA	October 2001, August 2004, November 2005	Princeton Survey Research Associates International
And from what you have heard or read, do you believe increases in the Earth's temperature over the last century are due more to -- the effects of pollution from human activities (or) natural changes in the environment that are not due to human activities?	GallupGWHumanActs	March: 2003, 2006, 2007, 2008, 2010, 2011, 2012, 2013, 2014, 2015	Gallup Organization
There is a controversy over what the countries of the world, including the US (United States), should do about the problem of global warming. I'm going to read you three statements. Please tell me which statements comes closest to your own point of view....Until we are sure that global warming, is really a problem we should not take any steps that would have economic costs. The problem of global warming should be addressed, but its effects will be gradual, so we can deal with the problem gradually by taking steps that are low in cost. Global warming is a serious and pressing problem. We should begin taking steps now even if this involves significant costs.	USUMARY	June 2004, July 2005	Program On International Policy Attitudes, University of Maryland
Which of the following statements is closest to your own opinion?...There is a consensus among the great majority of scientists that global warming exists and could do significant damage. There is a consensus among the great majority of scientists that global warming does not exist and therefore poses no significant threat. Scientists are divided on the existence of global warming and its impact.	USUMARY1	July 2005, June 2004	Program On International Policy Attitudes, University of Maryland
Global warming is a term used to describe changes in the temperature of the earth's atmosphere which could result in changes in the environment. How much have you heard or read about global warming--a lot, some, not much or nothing at all?	USCBSNYTTT	May 2006, April 2007,	CBS News/New York Times

Which comes closer to your view?...Global warming is a very serious problem and should be one of the highest priorities for government leaders. Global warming is serious but does not need to be a high priority. Global warming is not serious and can be addressed years from now.	USCBSNY	April 2007, December 2009, September 2011	CBS News/New York Times
Which of the following statements come closest to your view of global warming?...Global warming is a proven fact and is mostly caused by emissions from cars and industrial facilities such as power plants and factories. Global warming is a proven fact and is mostly caused by natural changes that have nothing to do with emissions from cars and industrial facilities. Global warming is a theory that has not yet been proven.	USORC	May 2007, June 2008, December 2009, September 2011, January 2013	ORC International
If nothing is done to reduce global warming in the future, how serious of a problem do you think it will be for the world?...Very serious, somewhat serious, not so serious, not serious at all	USIPSOSR	June 2007, September 2007	Ipsos-Public Affairs
Do you think global warming is an environmental problem that is causing a serious impact now, or do you think the impact of global warming won't happen until sometime in the future, or do you think global warming won't have a serious impact at all?	USCBS	October 2007, February 2009, August 2010, October 2010, April 2011, April 2013	CBS News
In your view, is global warming a very serious problem, somewhat serious, not too serious, or not a problem?	USSRBI	May 2008, October 2009	Abt SRBI
Which statement comes closest to your view about global warming?...Global warming is caused mostly by human activity such as burning fossil fuels. Global warming is caused mostly by natural patterns in the earth's environment. Global warming does not exist.	USCBSS	October 2012, January 2013, April 2013	CBS News

NOTE: *A summary of all the variable data used in the VAR analysis is available in the "Stata VAR Excelification.xlsx" file in the "Data" tab*

Applying the Model to CCSI

Detailed Information on Variable Sources and Recoding

1. Economic Recessionary Data: 6 measures of economic recession indicators are used in the thesis. This category of data was easy to find as the St. Louis FRED Economic Research website hosted all this information and I downloaded quarterly data for the desired time period (Q1 2000 – Q4 2015). Data can be found at: https://fred.stlouisfed.org/

2. Partisanship/Political Elite Cues: 6 measures of partisanship were included in the model and I followed the same steps as Brulle (2012). The sources and recoding information of these variables is borrowed from Brulle (2012) Supplementary Information since they are exactly replicated for my thesis:-

- Congressional action statements on climate change issued by Republicans and Democrats identified by a keyword search of Lexis-Nexis Congressional (Sellers 2010: 79-80). Each statement was coded as either supporting, opposing, or neutral regarding Congressional legislative action to address climate change.

- Number of Congressional hearings on climate change reported in the Proquest Congressional Data Base (3/1/2019) under "Global Climate Change," "Greenhouse Effect," “CO2” and "Carbon Dioxide."

- Senate and House roll call votes on climate change bills identified in the League of Conservation Voters National Environmental Scorecard (see Lindaman and Haider-Markel 2002: 97). Data can be found online at: http://scorecard.lcv.org/scorecard?year=all

3. Controls – Extreme Weather Data: 4 measures of extreme weather as measured by drought levels, percentage of warm and cold areas, and an overall extreme weather index were compiled using the same strategy suggested by Brulle (2012). The recoding description is again borrowed from Brulle (2012) Supplementary Information:-

- Overall Climate Extremes Index – arithmetic average of six indicators of climatic extremes across the U.S. Data can be found online at: https://www.ncdc.noaa.gov/extremes/cei/graph

- Extremes in Maximum Temperature % of U.S. with maximum temperatures much above normal. Data can be found online at: https://www.ncdc.noaa.gov/temp-and-precip/uspa/

• Drought Levels - % of U.S. in severe drought based on the Palmer Drought Severity Index. Data can be found online at: https://www.drought.gov/drought/search/data?f%5B0%5D=field_data_coverage%3A157

4. Controls - Media Coverage & Advocacy: 3 measures of media coverage and advocacy were built using strategies recommended in Brulle (2012). The recoding information and description are partly borrowed from Brulle (2012) Supplementary Information:

- The Media Coverage Index is an additive index (alpha=0.649) used to represent media coverage on climate change. The Index is based on three types of media sources (TV, Newspaper, & Magazines):
  - Number of stories on climate change on the nightly news shows of the major broadcast TV networks (NBC, CBS, ABC) based on a Boolean keyword search of the Vanderbilt Television Archives using "global warming," "climate change," “greenhouse” and "sea level". Data can be found online at: https://tvnews.vanderbilt.edu/
  - Number of stories on climate change in the New York Times - The count of stories on climate change in the NY Times was collected by a Lexis-Nexis Academic Search using the same set of keywords as above.
  - Number of stories on climate change in the three major weekly magazine stories (Newsweek, Time, and USA Today) - gathered using “Readers' Guide Full Text Select (H.W. Wilson)”, access provided by DePauw libraries. The search was again made using the following terms: "climate change" or "global warming" or greenhouse or "atmospheric carbon dioxide".
- Number of stories on climate change in 7 major environmental magazines (listed below) - gathered from “Readers' Guide Full Text Select (H.W. Wilson)”, access provided by DePauw libraries. Search was made using the following terms: "climate change" or "global warming" or greenhouse or "atmospheric carbon dioxide". The 7 environmental magazines are: *American Forests*, *E: The Environmental Magazine*, *Environment*, *National Parks*, *Oceanus*, *Sierra*, and *The Mother Earth News*.
- Number of stories on climate change in 4 major conservative (listed below) - gathered from “Readers' Guide Full Text Select (H.W. Wilson)”. Search was made on the following terms: "climate change" or "global warming" or greenhouse or "atmospheric carbon dioxide". The magazines are: *Human Events, National Review, Reason*, and *The American Spectator*.

5. Controls - Scientific Information: 2 measures of scientific information were used based on recommendations from Brulle (2012) again. These are:

- Count of articles in different types of scientific outlets including: Magazines, Academic Journals, Biographies, and Peer-Reviewed Articles. The search phrases used were: “global warming” OR “climate change” OR greenhouse”. Used the same source as Media advocacy: Reader's Guide Full Text Select (H.W. Wilson).
- Release of major climate change assessment reports - The release of major climate change assessment reports scored as a dummy variable (yes = 1) for the quarters in which a report was released. The following reports were included: 1) Intergovernmental Panel on Climate Change (IPCC) Reports (varies years), 2) US Global Change Research Program (USGCRP), and 3) the America’s Climate Choices report released by the NRC. Data for IPCC can be found at: https://www.ipcc.ch/reports/. Data for the other two can be found at: https://www.globalchange.gov/browse/reports?f%5B0%5D=field_report_organization%3A175

Data Analysis

These are all the results of the Augmented Dickey Fuller tests for both Model 1 and Model 2 given by Stata. The commands for these tests are available in the associated Do-Files.

```
. dfuller UnempRate, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.259            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.1855

.
. dfuller RealGDP, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -0.750            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.8332
```

```
. dfuller IPI, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.018            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.2787

.
. dfuller SciReport, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.033            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0319
```

```
. dfuller NonfarmPayrolls, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -0.772            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.8272

.
. dfuller RecessionDates, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.909            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0443
```

```
. dfuller MediaIndex, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.504            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.5314

.
. dfuller CEI, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.802            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0580

.
. dfuller CCSI, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.937            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3150
```

Figure S1: *Augmented Dickey Fuller Tests for Model 1*

*not showing ADF tests at first differences for nonstationary variables

```
. dfuller DemsStatmnts, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -2.506            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.1141

.
. dfuller RepubStatmnts, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -2.338            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.1599

.
. dfuller HouseHearings, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -1.766            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3974
```

```
. dfuller SenateHearings, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -1.909            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3281

.
. dfuller LCVDems, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -1.377            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.5931

.
. dfuller LCVRepubs, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -0.742            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.8355
```

```
. dfuller USWarmAreas, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -3.289            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0154

.
. dfuller SciReport, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -3.033            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0319
```

Figure S2: *Augmented Dickey Fuller Tests for Model 2*

*not showing ADF tests at first differences for nonstationary variables

Resources and Additional Information

Disclaimer: All the knowledge shared about the Vector Autoregression is self-taught using the Applied Econometric Time Series (4^{th} Edition) book by Walter Enders (2011), several journal articles, and some really helpful videos by Ben Lambert.

DGP: Vector Autoregression – Pre-estimation Steps

Unit Root Process: Detailed scholarly information about the unit root process can be found on Enders (2011) – Chapter 4 (Models with Trend), Unit Roots and Regression Residuals (Unit 3).

For those seeking for a quick intuitive explanation of the Dickey Fuller test for unit root, please refer to Lambert (2013) video: “Dickey Fuller Test for Unit Root” https://youtu.be/2GxWgIumPTA.

Optimal Lag Length Tests: Again, detailed information and a discussion of the optimal model selction criterion to render parsimonious models is provided by Enders (2011) on page 69, in the *Model Selection Criteria* subsection of the chapter, “Sample Autocorrelation of Stationary Series”.

Again, a shorter visual discussion of these methods are discussed in this 11-minute video by Lambert (2013): “Evaluating model fit through AIC, DIC, WAIC, and LOO-CV. https://youtu.be/xS4jDHQfP2o

DGP: Vector Autoregression – Post-estimation Steps

Wald Test: Lambert (2013) provides a short, under 7-minute, introduction to the Wald test for those interested in learning more about it. See the video: https://youtu.be/TFKbyXAfr1M