

APPLICATION PAPER  

Machine learning and feature selection: Applications in economics and climate change

Berkay Akyapı 

Department of Information Systems and Operations Management, University of Florida, Gainesville, FL, USA
Email: berkayakyapi@ufl.edu

Received: 13 February 2023; **Revised:** 01 August 2023; **Accepted:** 23 October 2023

Keywords: Causality-based feature selection; climate change; economics; feature selection; machine learning

Abstract

Feature selection is an important component of machine learning for researchers that are confronted with high dimensional data. In the field of economics, researchers are often faced with high dimensional data, particularly in the studies that aim to understand the channels through which climate change affects the welfare of countries. This work reviews the current literature that introduces various feature selection algorithms that may be useful for applications in this area of study. The article first outlines the specific problems that researchers face in understanding the effects of climate change on countries' macroeconomic outcomes, and then provides a discussion regarding different categories of feature selection. Emphasis is placed on two main feature selection algorithms: Least Absolute Shrinkage and Selection Operator and causality-based feature selection. I demonstrate an application of feature selection to discover the optimal heatwave definition for economic outcomes, enhancing our understanding of extreme temperatures' impact on the economy. I argue that the literature in computer science can provide useful insights in studies concerned with climate change as well as its economic outcomes.

Impact Statement

Understanding the effects of climate change to the economy requires interdisciplinary research. This article reviews the literature in computer science, and aims to introduce recent developments in machine learning that can assist earth scientists and economists to uncover the effects of climate change to the welfare of countries.

1. Introduction

There has been an interest in understanding the effects of greenhouse gas (GHG) emissions to the climate for many years.¹ However, there were claims suggesting uncertainties regarding anthropogenic climate change, questioning whether fossil fuels genuinely contribute to climate changes. For example, the New York Times' 1997 article titled "A Degree of Uncertainty" talks about two factors for nations to move prudently about implementing policies to address climate change: "[t]here is a high degree of uncertainty over the timing and magnitude of the potential impacts that man-made emissions of GHGs have on

  This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

¹ For example, in 1896 Svante Arrhenius wrote an article about the effect of Carbonic Acid to average temperatures (Arrhenius, 1896).

climate [and] the emission-reduction policies being considered carry with them very large economic risks” (Mobil Corporation, 1997). As of 2023, the uncertainty over the effects of human contribution to the changing climate is no longer considered as an uncertainty even for the same corporation writing the aforementioned article.² However, concerns persist about the potential harmful effects on the economy that may arise from regulating or taxing the fossil-fuel industry.

In an attempt to address this second concern, there is ongoing research showing that changing climate is harming the economy. Therefore, switching from the fossil-fuel industry will benefit the economy through the reduction of GHG emission, which in turn will limit increasing temperatures. Hence, the negative impacts from taxing the fossil-fuel industry might be covered from the benefits of limiting the climate change. There has been extensive research that aims to quantify the effects of climate change to the overall economy, usually measured through analyzing how gross domestic product (GDP) is affected from average temperatures or certain extreme events caused from increasing temperatures.³

To better comprehend the impact of increasing temperatures on the economy, two key phenomena need to be understood. Firstly, the relationship between rising temperatures and extreme weather events must be explored, along with the magnitude and frequency of their changes. However, this necessitates a deep understanding of earth science to establish causality and accurately identify the specific types of events influenced by temperature. Secondly, it is important to determine which extreme weather events have significant economic implications, as not all events may affect the economy. By studying these phenomena, a clearer understanding of the economic effects of temperature increase can be obtained, enabling informed decision-making regarding mitigation strategies.

To formulate effective policies mitigating the economic impacts of climate change, interdisciplinary research is crucial. The advancements in technology have resulted in abundant data on weather events like temperature and precipitation. Thus, leveraging machine learning (ML) techniques capable of capturing complex patterns can aid in understanding the changing climate. However, ML researchers may lack insights into the social and environmental intricacies underlying the data being analyzed. Hence, promoting multi-disciplinary research in ML can yield more efficient algorithms in this context. In this review, I assess various articles from computer science (CS) literature that can address two key challenges in studying climate change’s adverse effects on the economy: identifying the causal impact of rising temperatures on extreme weather events and selecting relevant events that influence the economy.

Recent development in ML has already attracted the interest of fields other than CS. An example of such field is economics; where researchers have long been interested in using prediction tools (econometrics) to analyze data and infer if the theoretical models are in line with what is being observed. As such, some econometricians aimed to introduce ML tools to economists. For example, Athey and Guido (2019) introduce well known algorithms in ML and suggest that some economic problems could benefit from the use of these tools. Similarly, Imbens (2020) compares potential outcome framework (e.g., inference through randomized control trials) that is widely used in economics to directed acyclic graphs (DAGs) which is being used in the CS literature to infer causality.

In this article, I explore recent feature selection algorithms developed in the CS literature and propose their application to enhance practices aimed at mitigating the economic impacts of extreme climatic events. I start by introducing terminology commonly used in economics and establish connections with the terminology employed in CS literature. Subsequently, I argue that noncausal feature selection algorithms such as the Least Absolute Shrinkage and Selection Operator (LASSO) can be used to understand how weather events caused by climate change can affect the economy. Causality can be imposed by economists to feature selection algorithms through economic theory. For example, a weather shock may affect the GDP of a country, but it is unlikely that the GDP in that year is going to affect the

² Mobil and Exxon merged on November 30, 1999. ExxonMobil “share the ways in which [they] remain determined to tackle head-on the challenge of strengthening energy supply security and reducing emissions to support a net-zero future” in their “2023 Advancing Climate Solutions Progress Report”.

³ Dell et al. (2014) and Newell et al. (2021) provide literature reviews about this area of research.

occurrence of a weather shock in that country.⁴ Therefore, noncausal feature selection algorithms such as the LASSO (Tibshirani, 1996) or adaptive LASSO (Zou, 2006), which are already proven to be efficient, can be used to select the weather events that affect the economy.

Later, I argue that, if Earth scientists guide the development of the algorithms, causality-based feature selection algorithms can be used to reveal how changing climate is affecting the magnitude and frequency of extreme weather events. Causality-based feature selection is a method that identifies and selects features in a dataset that have a causal relationship with the target variable of interest. As stated above, computer scientists developing causality based feature selection algorithms can be unaware of nuances that exist in the social or environmental processes that this technology is being applied to analyze (Imbens, 2020). Hence, they attempt to develop algorithms that can bypass the need for informed experts by taking an exhaustive approach. However, lack of input from experts may result in unsuccessful attempts of algorithm development, for example by not considering an important feature that is relevant for the study area. The efficiency could improve by working together with specialists in different disciplines and focusing development efforts on the specific problems that would improve the ability of the algorithms.

Finally, I present an application that enhances our understanding of how heatwaves influence economic outcomes. Heatwave definitions can vary across different applications in the literature. Considering all combinations leads to 32 distinct measures for heatwaves. For each definition, I generate 9 distinct measures. I use Group LASSO to select the heatwave definition that best explains the variation in personal income per capita among US counties during the 21st century. Subsequently, I employ Sparse Group LASSO to identify a single event impacting the economy.

The findings reveal that heatwaves can significantly and negatively affect the growth of personal income per capita in counties. Specifically, an additional heatwave occurrence can decrease personal income per capita growth by 0.126%. For medium-sized counties, GDP ranged from \$2.3 billion, to \$42.3 billion in 2021.⁵ As a result, one more occurrence of a heatwave could have costed between \$2.9 million to \$53.3 million dollars for a single medium-sized county.

The rest of the article is organized as follows. In [Section 2](#) I introduce some jargon used in the economics literature that can differ from CS literature and the econometric problems that researchers are facing when trying to understand the effects of climate change to the economy. In [Section 3](#) I introduce the developments in the CS literature that aims to select features using different techniques where I particularly focus on LASSO and causality based feature selection. I also introduce some recent developments regarding the usage of artificial neural networks (ANNs) for feature selection. In [Section 4](#), I offer an illustrative example using feature selection techniques to enhance our comprehension of the economic impact of heatwaves. Finally, in [Section 5](#) I conclude the paper.

2. Background

2.1. Bridging the terminology gap between economics and computer science

It is important to note the distinction in jargon between CS and economics. In ML, a dataset can be represented as an $N \times K$ matrix, with N denoting the number of observations (rows) and K denoting the number of distinct variables or features (columns). In CS context, the term “feature” refers to one of these K columns, representing an input or predictor variable used in a model. In economics, the more commonly used term is “variable”, which encompasses any measured or observed quantity. Each row is referred to as “observations” in economics, but can be termed “instances” in CS and ML literature. For the purpose of this text, I use feature and variable interchangeably, as well as observations and instances.

⁴ It can be argued that higher GDP may lead to higher Green House Gas (GHG) emissions, and that could lead to climate change. However, first GHG emissions have a delayed effect of changing the climate (IPCC, 2021), and second the GHG emissions of a country does not necessarily effect the occurrence of extreme weather events on that specific country. Therefore, the effect of higher GDP to the occurrence of weather events can be controlled for by using time fixed effects in the regressions.

⁵ <https://www.bea.gov/news/2022/gross-domestic-product-county-2021>.

In economics, researchers distinguish between control variables and treatment variables. Treatment variables, also known as explanatory variables, are factors that are manipulated or naturally vary in an analysis to assess their impact on the outcome of interest. Control variables, on the contrary, are held constant or considered to isolate the relationship between the treatment variable and the outcome. They help control for potential confounding factors and enhance the accuracy of estimating the causal effect of the treatment variable. Even though this analysis can be conducted through a linear regression, the control variables can enter to the equation as higher-order polynomials to capture the relationship accurately. For example, an increase in income may decrease the probability of committing violent crime, but this relationship may be nonlinear. The decrease in the probability of violent crime may differ for income increases from \$10,000 to \$20,000 versus increases from \$100,000 to \$200,000. Therefore, higher order polynomials may capture the relationship between violent crimes and income more accurately.

In this context, the functional form of control variables pertains to the specific mathematical relationship or equation used to model the association between the control variable and the outcome. Economists aim to carefully select the functional form to accurately represent the presumed relationship between the control variable and the outcome variable. The choice of functional form can have implications for the estimated effects and the interpretation of the results.

To select the correct functional form of control variables in the regressions economists use LASSO. An example is Belloni et al. (2012), which develop a new algorithm for LASSO to choose among many features to infer the causality of a variable of interest in a specific setting. This algorithm is further developed by Belloni et al. (2014b) and used in other examples in Belloni et al. (2014a). Even though these articles exploit the potential of feature selection in economics, there seems to be a detachment between the economics and the CS literature. For example, Belloni et al. (2012) and Belloni et al. (2014b) use simulations to show the convergence properties of their algorithm and choose a hyper-parameter (the parameter for the penalty term in LASSO) that maximizes the R-squared of their prediction with simulated data. Later on, they use the same penalty parameter for other three exercises in Belloni et al. (2014a). However, the CS literature states that one cannot know a-priori the optimal hyper-parameters and has to search over different hyper-parameters for each problem in hand.⁶

Establishing causality is a fundamental goal in economics. Causality refers to the relationship between cause and effect, specifically demonstrating that changes in the treatment variable lead to changes in the outcome variable while ruling out alternative explanations. However, endogeneity poses a challenge in establishing causality. Endogeneity refers to a situation where the relationship between variables is influenced by factors that are not adequately accounted for in the analysis, leading to biased or inconsistent estimates. Exogeneity, in contrast, implies that the variables being studied are unaffected by such omitted factors and can be treated as independent of the error term or other variables in the model.

For example, assume we want to uncover the causal effect of some weather events (X_1) to GDP per capita (y), but we do not consider controlling for other variables denoted as Z_1 . Omitting Z_1 from the regressions may cause an endogeneity problem, that will prevent unveiling the true causal effect of X_1 on y . To clarify, assume that we do not observe Z_1 and we run a linear regression to find the coefficient of X_1 on y :

$$y = \beta X_1 + Z_1 w + \varepsilon \quad (1)$$

As we do not observe Z_1 , we will run the regression given in the following equation:

$$y = \beta X_1 + u \quad \text{where } u = Z_1 w + \varepsilon \quad (2)$$

In this case we will estimate β as follows:

⁶Recent articles in the economics literature begin considering hyper-parameter optimization in their work; for example Bianchi et al. (2022).

$$\begin{aligned}
 \hat{\beta} &= (E[X_1 X_1^T])^{-1} E[X_1 y] \\
 &= (E[X_1 X_1^T])^{-1} E[X_1 (\beta X_1 + Z_1 w + \varepsilon)] \\
 &= \beta_1 + \underbrace{(E[X_1 X_1^T])^{-1} E[X_1 Z_1^T]}_{\text{Omitted Variable Bias}} w
 \end{aligned}
 \tag{3}$$

Causality can still be inferred even if we omit Z_1 from the regression under certain conditions. Firstly, if Z_1 is uncorrelated with X_1 (i.e., $E[X_1 Z_1^T] = 0$) or secondly, if Z_1 is uncorrelated with y (i.e., $w = 0$), we can infer the causal effect of X_1 on y . For example, when investigating the effect of droughts on GDP per capita, if Equation 1 satisfies $E[\varepsilon|X_1, Z_1] = 0$, no co-linearity exists, and we have a sufficiently large sample, then an unbiased estimator for β is possible as long as $E[X_1 Z_1^T] = 0$, even if Z_1 affects GDP per capita (i.e., $w > 0$).

The differentiation between control variables and treatment variables holds significance during feature selection. When we aim to choose the appropriate treatment variables among many options but have certain variables that must be included as controls regardless of the selection, we can enforce the control variables into the selection process using projection matrices. By using projection matrices, the coefficient of X_1 remains the same in both Equations (1) and (4). Therefore, if we want to incorporate the control variables into the selection, we can enforce them into the process using the following technique, provided we observe Z_1 :

$$\begin{aligned}
 y - Z_1 (Z_1^T Z_1)^{-1} Z_1^T y &= \beta X_1 + Z_1 w - Z_1 (Z_1^T Z_1)^{-1} Z_1^T X_1 \beta - \underbrace{Z_1 (Z_1^T Z_1)^{-1} Z_1^T}_{I} Z_1 w \\
 (I - Z_1 (Z_1^T Z_1)^{-1} Z_1^T) y &= (I - Z_1 (Z_1^T Z_1)^{-1} Z_1^T) X_1 \beta
 \end{aligned}
 \tag{4}$$

In other words, to incorporate the control variables into the selection process, we can use $(I - Z_1 (Z_1^T Z_1)^{-1} Z_1^T) y$ as the dependent variable and $(I - Z_1 (Z_1^T Z_1)^{-1} Z_1^T) X_1$ as the relevant potential features to be selected. This helps ensure that the control variables are included in the selection during the other feature choices. I use this technique in the application that I present in Section 4.

To summarize, feature selection can play a crucial role in both controlling for confounding factors and identifying relevant variables associated with the outcome variable. By carefully selecting features, the precision of estimating causal effects and comprehending climate-economy relationships can significantly improve as I discuss in the following sections.

2.2. Weather events and the economy

It is well established in the economic literature that geography, including a region’s climate, is not the main characteristic affecting a country’s development (Acemoglu and Robinson, 2012). As a result, the literature that aims to understand the effect of climate change to the economy uses panel data, where one observes several replications of the time series across a panel of observations, and analyze the effects of weather shocks to within-country economic variation. For example, Dell et al. (2015) use a panel data analysis to understand how inter-annual variations in temperature affected countries’ GDP per capita growth and Burke et al. (2015) follow a similar strategy to assess the functional form of average temperature changes by including higher order polynomials of annual average temperature per country.

However, relying on one climatic variable estimated as an annual average does not capture all weather events that affect a countries’ economy. For example, Figure 1a shows the maximum temperatures observed on the earth surface in June 21, 2019, a date that marks the June solstice and onset of summer in the Northern hemisphere.⁷ The average temperature in United States for this day is 18.93°C, and the

⁷ Figure 1 was created using a satellite-derived gridded dataset measuring maximum temperature in each pixel. The maps in Figure 1 are generated by the author, and a variation of them are also used in Akyapi et al. (2022).

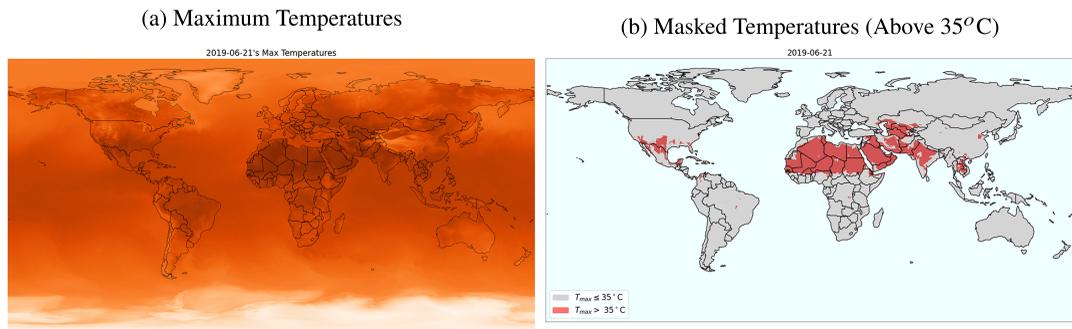


Figure 1. Hourly temperature data for June 21, 2019 obtained from ERA5 data set provided by European Centre for Medium-Range Weather Forecasts. Panel (a) provides a color map showing the maximum temperatures observed on this day in each grid cell. Panel (b) shows grid cells that have temperatures above 35°C in red, and temperatures below 35°C in gray.

average temperature for South Africa is 13.71°C.⁸ United States has a higher average temperature than South Africa; however, South Africa has a participating day heatwave for this specific date but United States does not.⁹ Therefore, a study that focuses only on average temperatures per country would not be able to capture such an event.

Figure 1b shows the grid cells with maximum temperatures above 35°C in red, and grid cells with temperatures below 35°C in grey. Even though South Africa has a participating heat wave on this day, no grid cells were above 35°C. On the contrary, even though United States does not have a participating heat wave on this day, 7.96% of the country experienced temperatures above 35°C, which is potentially harmful to the economy.

Overall, Figure 1 provides an example showing that average temperatures by themselves may not be able to capture events that may be harmful for a country's economy. First, focusing on temperature averages can neglect the complexities of weather shocks, events where countries experience anomalous conditions, such as the heatwave in South Africa observed in June 21, 2019. Second, if we were to try understanding if United States had experienced temperatures above 35°C after taking the average, we would have missed this event, because 92.04% of the country had temperatures below this threshold. Moreover, this inspection also shows that using the same data source (hourly temperature data) but different aggregation methodologies, a researcher can generate many weather events that can be relevant to a country's growth.

Therefore, the approach taken by Dell et al. (2012) and Burke et al. (2015) create a risk of omitting some important linkages between a region's climate and economic activity. Nevertheless, it is critical for adaptation that we understand the exact channels from which weather is affecting the economy. Owing to the recent developments in data acquisition and accessibility, we can observe temperature and precipitation on high frequency and high spatial resolution at a global scale. Concurrent use of high resolution data with synthesis of knowledge from the Earth science literature enhances our ability to generate weather events and indices necessary for understanding the socio-environmental processes influencing a country's economy.¹⁰ For example, European Centre for Medium-Range Weather Forecasts (ECMWF)

⁸ The average temperatures for each country are calculated by the author by assigning a grid-cell (pixel) to a country if the centroid of the cell (pixel) is within the country boundaries.

⁹ I use heatwave definition from Perkins and Alexander (2012), where a day heatwave event occurs if the maximum temperature exceeds the 90th percentile of the 1979-2019 distribution in a 15-day window centered on each day, for at least three consecutive days.

¹⁰ See Perkins and Alexander (2012); Kim et al. (2020) Lai et al. (2020) for some examples to generate weather events that can be relevant to GDP growth.

provides hourly estimates of a large number of climatological variables that cover the Earth on a 30km grid from 1979 to 2023, which makes billions of observations. Furthermore, as it is discussed above, one can construct many variables using the distribution of these measures. Two examples are Kotz et al. (2021) and Kotz et al. (2022), where the former shows that the variance of temperature affects countries' GDP growth and the latter shows the same for rainfall changes. Hence, the first main problem to overcome in understanding the effects of changing climate is to find the right causal climate event affecting the welfare of countries, which is rarely straightforward.

Even though the aforementioned problem can be solved by analyzing past occurrences of weather events, there is also interest in predicting potential effects for the future. However, without knowing how increasing temperatures due to GHG is going to affect the frequency and magnitude of extreme events, it is not feasible to extrapolate past results and try to predict the future. Therefore, a second challenge in applying ML to model the relationship between climate and economies is to understand how changing climate is going to affect the magnitude and frequency of extreme weather events.

A relevant literature in CS to overcome these two main problems (causality and feature selection) is the literature regarding Feature Selection. This literature has different algorithms to tackle down different type of problems. In the next section, I discuss two algorithms that can overcome the specific difficulties that are summarized above: LASSO and causality-based feature selection. The latter uses DAG to describe and solve the problem. DAGs are graphical representations of causal relationships between variables where each variable can be represented as a node, and arrows between nodes indicate the causal relationships. DAGs are used in causal inference when direct experimentation is not feasible. For example, in understanding the causal relationship between anthropogenic climate change and extreme events, such as floods or hurricanes, it is impossible to directly observe a counterfactual scenario where the climate remained unchanged.

I introduce an example DAG in Figure 2.¹¹ This is an example of an acyclic graph where each edge directed from a node to another does not form a closed loop (where the nodes are temperature, precipitation, M weather events (where i th weather event is written as WE_i), GDP per capita and Z_1 (an unobserved variable affecting GDP per capita) and edges are shown with arrows). It shows that certain weather events are a result of changing temperature and/or precipitation. Additionally, these weather events may create other extreme events or they may directly affect GDP per capita. In this context, GDP

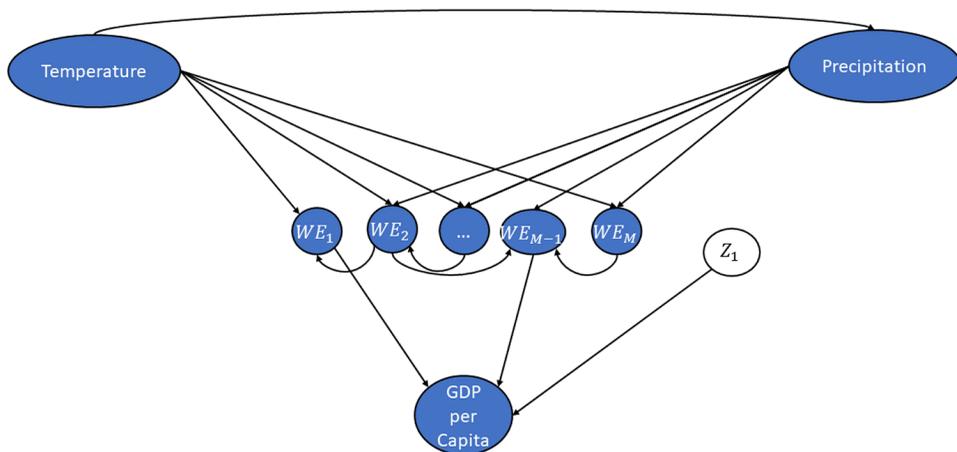


Figure 2. Hypothetical directed acyclic graph (DAG) for weather events and GDP per capita.

¹¹ Note that Figure 2 is only presented as an example to summarize the problem. The causality of increasing temperature on precipitation or on other climate events needs to be assessed by climatologists. For example, the type of soil within a grid cell might also affect the probability of a flood event when there is an extreme event of precipitation and it is not shown in Figure 2.

per capita is the *child* of WE_1 and temperature is the *ancestor* of GDP per capita and WE_{M-1} is the *spouse* (another *parent* of GDP per capita) of WE_1 .

3. Causal and Noncausal Feature Selection Literature

This section reviews the literature in CS regarding Feature Selection. First, categorizations that can be helpful in understanding feature selection algorithms are introduced. Second, some recent developments regarding LASSO and causality-based feature selection are introduced together with discussions on how these can be helpful in the area regarding climate change and the economy. Finally, some recent developments regarding feature selection using ANNs are summarized.

3.1. Categorization in feature selection algorithms

When a researcher has high dimensional data (i.e., data that has a high number of features or control variables) it can be difficult to choose the relevant ones to answer the question in hand. For example, as discussed in Section 2 there are many weather events that might be affecting the welfare of countries and it is not feasible to try out all potential events in order to find the most relevant features. Assume we have n number of features and we want to select d features among them. If we were to try all potential combinations, we would have to search over $\binom{n}{d}$ possibilities (Jain and Zongker, 1997). Therefore, algorithms to search efficiently without the need of doing an exhaustive search are developed and they are referred as Feature Selection Algorithms.

There are different set of algorithms for feature selection and several categorizations are proposed to differentiate between them. The first type of categorization that is commonly mentioned in this literature is regarding the features. Features can be divided into three main categories: Strongly relevant features, weakly relevant features, irrelevant features (Yu and Liu, 2004; Yu et al., 2021). Strongly relevant features are the essential features that should not be removed during a feature selection process. Weakly relevant features could be selected under certain conditions; however, they can be replaced with other features.¹² Irrelevant features are the ones that should not be selected during a feature selection process because they are not relevant to the outcome. The aim of a feature selection algorithm is to choose all relevant features and a subset of weakly relevant features while dropping the irrelevant and redundant ones.

The second type of classification is regarding feature selection algorithms. These algorithms are generally classified into three main categories in the literature: standard filter, wrapper and embedded methods (Jovic et al., 2015; Yu et al., 2021).¹³ Standard filters select the features without taking into account the model to be used. In other words, this method aims to give information about strongly relevant, weakly relevant and irrelevant features which then can be used for classification, clustering or regression analysis. The criteria for relevancy are typically based on correlation with the target variable or information gain. Most causality based feature selection algorithms are under this type of algorithms (Yu et al., 2020a, 2021) and they are presented in Section 3.3. The wrapper method selects the variables during the modeling process. For example, a clustering algorithm (e.g., K-means clustering) searches over all the parameters and chooses the features that help the most on defining the clusters of each instance. This method can be more effective on selecting the most relevant features, yet they may require higher computational costs (Jovic et al., 2015). The literature regarding the effects of climate change to the economy is mostly interested in regression analysis. Therefore, I do not introduce details about these type of methods in this article because these methods are more relevant to classification or clustering

¹² Jovic et al. (2015) define a fourth category as redundant features; features that are weakly relevant but can be dropped if another weakly or strongly relevant feature already captures the relevance of the redundant feature

¹³ This classification can differ between different articles. For example, Gui et al. (2017) state that there are only two classes of feature selection algorithms: standard filter and wrapper.

problems.¹⁴ Finally, the embedded method combines both standard filter and wrapper methods. Some examples are LASSO or Elastic Net where a penalty parameter is put into the regressions to provide sparsity when conducting the analysis. This types of feature selection algorithms can be widely used in economics literature as discussed in Section 2.

Additionally, standard filter algorithms are classified in different ways. For example, Jain and Zongker (1997) classify them under two main classes: forward and backward methods. In the forward method the algorithm begins with an empty set and keeps adding features. The backward method starts with a full set of features and deletes the features as it proceeds. An advantage of forward method is that one can have more features than observations during the selection process (Jovic et al., 2015; Yu et al., 2020a) suggest that there are two additional methods (in addition to forward and backward) for feature selection algorithms that lie under standard filter methods: bidirectional (or simultaneous) selection and heuristic feature subset selection. The bidirectional selection starts from both sides (i.e., with an empty set and with a full set of features) and simultaneously consider larger and smaller sets of features, whereas the heuristic feature selection generates a starting subset based on a heuristic and begins the exploration from this subset.

Finally, Gui et al. (2017) classify sparsity inducing feature selection algorithms that belong to embedded methods into two: vector based feature selection and matrix based feature selection. Vector based feature selection includes models such as LASSO where the sparsity is achieved by using penalty parameters of l_1 -norm. Matrix based feature selection algorithms are similar to the vector based algorithms in the sense that they penalize the inclusion of additional features. However, instead of using an l_1 -norm, they use $l_{2,1}$ -norm, which helps solving multiclass problems. As vector-based algorithms are well-suited for regression analyses, they hold greater relevance in economics and climate change applications. Therefore, I focus on vector-based algorithms to explore their potential for effectively addressing feature selection challenges in regression analysis.

Different methods have varying advantages among different tasks. Jovic et al. (2015) provide some background for best methods to be applied in different tasks such as clustering, classification and regression. This work focuses on methods that can be helpful for regressions. As stated in the introduction, one of the most common ML tools in the economics literature is LASSO. In the following subsection other forms of LASSO (such as Elastic Net, Group LASSO and Sparse Group LASSO) that are not commonly encountered in the economics literature are introduced. Finally, some recent developments regarding causality based feature selection (that lies under standard filter methods) that can be useful for climate change analysis are discussed.

3.2. Different forms of LASSO

The most common LASSO algorithm used in the economics literature is the one where the mean squared error is being minimized with features whose effect in minimizing the error term is higher than the penalty of including an additional variable. This type of feature selection is formulated as in the following equation:

$$\min_w L(w) + \lambda \|w\|_1 \quad (5)$$

where $L(w) := (y - Xw)^2$ is the mean squared error, $penalty(w) := \|w\|_1 = \sum_{i=1}^d |w_i|$ is the penalty term and d denotes the number of features. Note that it is allowed to have $d > n$; that is, it is allowed to have more features than number of observations (Belloni et al., 2012). The relevance of variables or features is defined based on their impact on the model's performance, such as assigning nonzero coefficients to features that effectively explain the variations in the outcome variable. The penalty terms shrinks the coefficient estimates toward zero, effectively selecting the most relevant features and setting irrelevant ones to exactly zero.

¹⁴ Classification and clustering could be helpful in this area of research as well. However, so far the studies focus mainly on regressions. Therefore, I leave these topics as future work. See Jovic et al. (2015) for more detailed information about wrapper methods.

The penalty term is not continuous hence no closed form solution exists to this problem. However, algorithms are suggested to solve these type of problems that generally begin with Ridge regression; where the penalty term is $penalty(w) = \sum_i^d |w_i|^2 = \sum_i^d w_i^2$ which makes the problem have a closed form solution.

LASSO has been used by Akyapı et al. (2022) to select among large number of distinct weather events generated using high-resolution-high-frequency geospatial data. However, other applications are possible. For example, LASSO could be used to select weather events affecting the economy for islands and non-island countries separately. This way, it could give information about the type of events that islands are especially sensitive which could guide adaptation strategies of these countries.

Other extensions to LASSO are provided by Gui et al. (2017). One of them is adaptive LASSO where the penalty term becomes $\sum_i^d a_i |w_i|$, that is, the coefficients (w_i) are weighted by a_i . A version of adaptive LASSO in the economics literature is presented in Belloni et al. (2014b). However, there are three other extensions of LASSO that could be useful in the economics literature but are being applied more rarely. One of them is the Elastic Net regularization, where the penalty term is written as $penalty(w) = \alpha \sum_{i=1}^d |w_i| + (1 - \alpha) \sum_{i=1}^d w_i^2$ ($0 \leq \alpha \leq 1$). In words, elastic net regression is a mixture of LASSO and Ridge estimator. This could be especially useful in applications where some features are strongly correlated, in which case LASSO may choose only one of them (Gui et al., 2017). Elastic Net can be useful in applications where a researcher tries to infer the causal effect of one of the features. If the algorithm does not choose a feature that is correlated with the dependent variable and with the variable whose causality the researcher is interested in, one may have a biased estimation of the true coefficient (see Equation 3).

Other two relevant extensions to LASSO are Group LASSO and Sparse Group LASSO. The Group LASSO optimization problem can be written as in 6

$$\min_p \left(\|y - \sum_{i=1}^k X_k w_{G_i}\|_2^2 + \lambda \sum_{i=1}^k \beta_i \|w_{G_i}\|_q \right) \quad (6)$$

where w is written as k disjoint groups (e.g., $w = \{w_{G_1}, w_{G_2}, \dots, w_{G_k}\}$) and β_i is the weight for the i^{th} group.

One application of Group LASSO, would be grouping weather events that are mostly a cause of high temperatures versus another group that have measures on weather events that are mostly a cause of extreme precipitation. A second approach could be to group the weather events with events that are measured as deviations from country specific long term averages (e.g., heatwaves) versus another group that measures weather events that are higher or lower than an absolute threshold (e.g., temperatures above 35°C). In Section 4, I present an application example where I use Group LASSO to select the most relevant definition of heatwaves concerning economic outcomes.

A drawback of Group LASSO is that it chooses groups but all features would have nonzero coefficients. This would work in a setting where only a subset of climate variables are being considered, but including other climate variables might be problematic because of the *curse of dimensionality*. A solution to this problem is using Sparse Group LASSO, where the penalty term is written as $penalty(w) = (1 - \alpha) \|w\|_1 + \alpha \sum_{i=1}^k \beta_i \|w_{G_i}\|_q$ (where $0 \leq \alpha \leq 1$). Sparse Group LASSO can be used for feature selection and group selection simultaneously. This could make it a very useful tool for climate and economics literature where many climate variables are constructed to understand the exact channel from which changing climate is affecting the economy.

3.3. Causality-based feature selection

Understanding causality can contribute to efficient feature selection as well as to the efficiency of other ML algorithms because it can provide more robust and transferable learning (Scholkopf et al., 2021). Hence, there are many examples where researchers use causal inference techniques to improve the performance of their methods. For example, Arya et al. (2021) use causality inference algorithms to identify the root causes of events that create interruptions in IT operations. This subsection begins with the

introduction of several definitions that are widely used in this literature and that are mentioned frequently throughout this subsection.

A widely used causality definition in this literature is Granger Causality, which states that “a variable is the cause of another if past values of the former are helpful in predicting the future values of the later” (Liu et al., 2010, p. 2). This notion is widely used in panel data settings which is relevant for climate change and economics literature. In fact, the developed techniques that learn using the notion of Granger causality are tested with climate or economics data. For example, Liu et al. (2010) use temporal causal graphs for climate modeling in the United States and Jangyodsuk et al. (2015) use a Bayesian Network Learning Technique that uses Granger Causality for flood prediction. Additionally, Basu et al. (2015) use these techniques to assess the risks faced by banks using a panel of banks’ balance sheet information.

Two other important definitions in the literature regarding causality based feature selection algorithms are Bayesian Networks (BN) and Markov Blanket (MB) which require the Faithfulness assumption (Jangyodsuk et al., 2015; Yu et al., 2020a; Ling et al., 2021; Yu et al., 2021). The structure of BN is a graphical model to represent dependencies of a set of variables and can be represented by a DAG. For example, Figure 3 is an example of causal BN derived from Figure 2. A MB “implies the local causal relationship between the class variable and the features [are all included] in its MB” (Yu et al., 2020a, p. 3). For example, in Figure 3 an MB of WE_1 would include Temperature and WE_2 (parents), GDP per capita (child) and WE_{M-1} (spouse).¹⁵ The MB of a BN is unique under the faithfulness assumption, which requires that “the independence facts true of the distribution are all and only those entailed by the network structure” (Meek, 1995, p. 1). In other words, the faithfulness assumption states that the conditional independence relationships represented in the network are consistent with the true causal relationships in the underlying system. For example, if two variables are independent in the data given a set of other variables, then there is no direct causal relationship between them in the real world.

As the causality is being inferred through observed data from a certain distribution, it is important to connect causality with statistical dependence. If two observed variables X and Y are correlated with each other, we can have four potential causality options. It can be the case that (1) X is causing Y , (2) Y is causing X or (3) there is another event Z that is the cause of both X and Y . In the presence of the third case, we can use conditional independence. For example, if X and Y are correlated without conditioning on Z , but they are not correlated when we condition on Z we could say that Z is the cause of X and Y . This is an example regarding the notion of conditional independence which is widely used to obtain an underlying

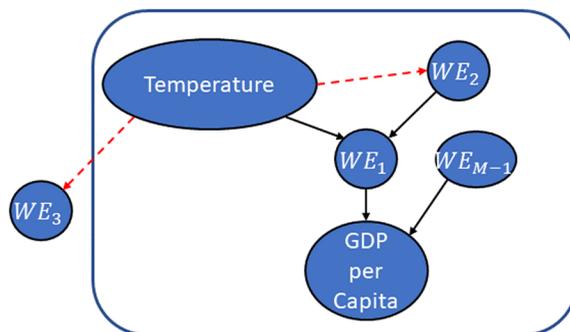


Figure 3. Markov blanket (MB) of WE_1 . In this example, temperature and WE_2 are the parents of WE_1 , GDP per capita is the child and WE_{M-1} is the spouse. The dashed arrows are not part of the MB of WE_1 , but they would be part of the MB of temperature.

¹⁵ Ignoring the unobservable variables such as Z_1 . If we were to include Z_1 , it would be in the MB of WE_1 as the spouse. However, as Equation 3 shows we can neglect the spouses that are uncorrelated with weather events.

graph of certain events (e.g., [Figure 2](#)).¹⁶ Finally, (4) it could be the case that there is a reverse causality, that is, X is causing Y and Y is causing X at the same time. In such situations, instrumental variables (IV) can be employed to overcome this issue and estimate the causal effect of X on Y . An IV is a third variable that is correlated with X but does not directly influence Y , except through its impact on X . It serves as an “instrument” for X , enabling the isolation of the causal relationship between X and Y . While there are attempts to synthetically generate IV (Dzhumashev and Tursunaliyeva, 2021), identifying valid instruments necessitates a profound understanding of the subject area.

Other widely used definitions are information gain and mutual information. Information gain is the reduction of entropy after an adjustment is made to a BN and it is calculated by comparing the entropy before and after a change is made.¹⁷ Mutual information calculates the statistical dependence between two variables and is the name given to information gain for the applications regarding feature selection (Ling et al., 2021). For example, if the information gain is higher when we use the assumption of X is causing Y versus Y is causing X , we can infer that the former is the true causality.

3.3.1. Standard filter methods for causality based feature selection

Standard filter methods focus on finding relevant and irrelevant features independent of the model to be applied (Jovic et al., 2015; Yu et al., 2020a). A natural question regarding the usage of these methods for feature selection would be about the connection between causal and noncausal feature selection. Yu et al. (2021) compare several aspects of causal and noncausal feature selections. First, they find that both causal and noncausal feature selection algorithms have the same objective function (for classification), where the objective function aims to maximize the mutual information between features and the outcome(s) of interest. However, they assert that the approximations and assumptions to solve the problems can differ between both methodologies. They conclude that causal feature selection algorithms perform better in finding the true relationship between the features and outcome(s) of interest while non causal feature selection algorithms are computationally more efficient and need less number of observations (or instances).

This finding has important implications for the research regarding climate change and economics. As it can be seen in [Figure 2](#), there are two important aspects to be addressed in this research area. The first one is regarding the effects of increasing temperature and changing precipitation in creating different climate events through direct effects (by being a parent) or indirect effects (by being an ancestor), that is, by being “a cause of the cause of the focused effect” (Jangyodsuk et al., 2015, p. 2). This part of the research has billions of observations and high dimensionality. Hence, causal feature selection algorithms seem to be a good candidate to tackle down causality between different climate events. The second part of the research is about understanding the effects of climate variables to the economy. This part has lesser observations and LASSO (and its different forms) seem to be a good candidate for this part of the problem. Feature selection using LASSO is introduced in [Section 3.2](#). Therefore, the rest of this subsection focuses on causality based feature selection algorithms that use standard filter methods.

An example of causality based feature selection algorithms is presented by Yu and Liu (2004). They point out that focusing on relevance of features with the outcome of interest may result in redundant feature selection. This is an important point for the analysis of climate variables and economics. For example, one can calculate heat waves during day and night separately (Kim et al., 2020) and both heat wave measures might have strong correlation with each other within country within a given year. If heatwaves are effective on a country’s economy, and we do not address redundancy, both of them might be chosen as a parent of GDP per capita. In a regression, this may cause near multicollinearity and prevent a researcher to tackle down the effect of heat waves on GDP per capita. To address this type of problems Yu and Liu (2004) suggest a methodology to explicitly handle feature redundancy. Their method first conducts a relevance analysis and removes irrelevant features. Later on, their method conducts redundancy analysis by doing a separate correlation analysis of between features and between features and class.

¹⁶ Conditional independence can also be referred as X and Y being d-separated (Ling et al., 2021).

¹⁷ Entropy is defined as $\sum_i -P_i \log(P_i)$ and it is a positive quantity because $0 \leq P_i \leq 1$.

There are two main strategies for causality-based feature selection: Standard Forward-Backward Feature Selection (SFBF) and Interleaving Forward-Backward Feature Selection (IFBF) (Yu et al., 2020a). SFBF begins with an empty set and adds features in the forward phase until a certain criterion is met and then removes false positives in the backward stage. IFBF performs both stages simultaneously. When a new feature is added in the forward phase, the backward phase is automatically triggered and begins searching for false positives. Both methods use either a Constraint-Based Method, where the algorithms give decisions according to a statistical independence test, or Score-Based Method, where the algorithms decide on the structure of the DAG by a scoring function, such as a measure of fitness between the DAG and the dataset (for example information gain).

An example that introduces the algebraic characterization of DAGs for score-based algorithms can be found in Zheng et al. (2020), where the authors present a framework of score-based algorithms that can be applied in various non-parametric settings. Yu et al. (2020a) provide an extensive literature review about causality based feature selection. They conclude that even though many algorithms were proposed so far, there are still many open problems to be addressed. Some open issues can be mitigated by supervising the causality based feature selection algorithms. For example, Yu et al. (2020a) state that it is difficult to distinguish between parents and children (PC) in causality-based feature selection algorithms. However, some structure could be imposed by researchers in relevant areas to ease the search of the algorithms. For example, it is known that a parent of heatwaves would be temperature or a parent of floods would be precipitation. If a relationship is found between temperature and heatwaves, an informed scientist could understand parent–child relationship between both nodes.

Another important application of causality based feature selection algorithms that is relevant for the climate change and economics literature is presented in Yu et al. (2020b). They analyze feature selection in an environment where similar features can be obtained through multiple sources and propose Multi-Source Causal Feature Selection (MCFS) algorithm to choose among features from datasets that might have different distributions. Their method uses the concept of causal invariance, which assumes that conditional distributions will remain unchanged from different potential interventions. Later on, they define a search criterion using mutual information. This is relevant in the climate change literature because spatial data regarding temperature and precipitation could be obtained through Weather Stations or through satellite data. For example, World Bank dataset¹⁸ on climate is obtained through weather stations, whereas ECMWF provide temperature and precipitation information derived from satellite data. Therefore, MCFS algorithm could be helpful to understand the causality of different weather events using different sources of data.

3.3.2. Critiques about causality-based feature selection

Several economists argue that causality based feature selection “has not shown much evidence of the alleged benefits for empirical practice in settings that resonate with economists” (Imbens, 2020, p. 1131). One of the main reasons of these critiques is because these methods may not be able to capture unobserved causes of an outcome, for example by omitting an important variable in the analysis. Therefore, it is important to develop these algorithms with experts in the area to be studied. For example, to study the causality between weather events and changing climate it is important to have a research team with earth scientists that can guide algorithm developers in using the right set of variables in the analysis.

3.4. Artificial neural networks for feature selection

The Universal Approximation Theorem states that any function can be approximated by ANNs. Therefore, ANNs are being used in many applications because they can provide better prediction or classification compared to other methods that rely on heuristics. For example, Yeh et al. (2020) show that one can infer the development in Africa by using satellite imagery and deep learning. Feature selection

¹⁸ <https://data.worldbank.org/topic/climate-change>.

and ANNs are complementary, in the sense that each can be used to enhance methodological approaches of the other tool. First, feature selection algorithms can be used to prune ANNs and reduce the computational burden. An example is Koneru and Vasudevan (2019), where they suggest a method to decrease the interconnectedness of ANNs by using a LASSO technique. As pointed out in Section 3.2, LASSO has no closed form solution because the absolute value function is not differentiable at the origin. To improve the efficiency of LASSO, Koneru and Vasudevan (2019) proposes a smoothing function to achieve sparsity in ANNs more efficiently.

On the contrary, one can use deep learning for causality based feature selection. Luo et al. (2020) review the literature that uses ANNs to construct causal DAGs. They first introduce the articles that transform the discrete DAG constraints into continuous functions. This approach turns the optimization problem into a differentiable one; which makes the usage of gradient descent algorithms feasible. The literature regarding the usage of ANNs for causality based feature selection is still developing, yet it has potential of significantly improving the performance and capabilities of causality based feature selection (Luo et al., 2020; Yu et al., 2020a).

4. GDP Impact Assessment of Heatwaves within the United States in the 21st Century

4.1. Potential definitions of heatwave

The literature presents various definitions of extreme heat. Perkins and Alexander (2012) defines day (night) heat waves as exceeding the 90th percentile of maximum (minimum) temperatures within a 15-day window centered on each day, for at least three consecutive days. On the contrary, Kim et al. (2020) defines “Warm Spell Duration” as temperatures above the 90th percentile of maximum temperatures within a 5-day window centered on each day, for at least six consecutive days.

Both Perkins and Alexander (2012) and Kim et al. (2020) establish a baseline distribution from 20 to 40 years of initial observations to calculate fixed thresholds for heatwave occurrences. However, Kahn et al. (2021) advocates considering adaptation when assessing the impact of rising temperatures on the economy and hence use a moving window when considering the temperature distribution. For instance, Moscona and Sastry (2022) show that “innovation reacts to climate change and shapes its economic impacts.” Consequently, using a moving average to calculate heatwave thresholds could offer valuable insights into the economic effects of extreme heat, as it considers adaptive responses to the changing climate.

Unlike Perkins and Alexander (2012) and Kim et al. (2020), Bilal and Rossi-Hansberg (2023) uses the fraction of days with temperatures above the 95th percentile of the national annual mean temperature distribution to identify extreme heat days. Which definition should be used to calculate the effects of extreme heat on the economy? Researchers can consider whether to use the 90th or 95th percentile, and whether to define a heatwave based on 3 consecutive days or 6 consecutive days. The window size can be either 15 days or 5 days, and researchers may decide between using a moving average or a constant threshold derived from initial years of the distribution. Additionally, the relevance of day heatwaves (based on maximum daily temperature) versus night heatwaves (based on minimum daily temperature) can play different roles in affecting the economy. I suggest using Group LASSO to address this problem.

4.2. Weather data

The relevant definition of heatwave can be determined by examining past occurrences and identifying the definition that best explains the variation in economic outcomes. To achieve this, I consider all potential combinations of heatwave definitions in Section 4.1, resulting in 32 distinct definitions for each county. Within each definition, I calculate three measures of heatwaves as suggested by Perkins and Alexander (2012): the number of days with a heatwave, the length of the longest heatwave, and the total number of heatwaves in a year. Additionally, I extend the calculations to cover three-month periods: January, February, and March; April, May, and June; and July, August, and September. However, to avoid

collinearity, I exclude October, November, and December from the analysis. Hence, within each group, there are 9 potential measures whose effects may be relevant to the economy.

In the following, I use d to denote calendar days and $j = 1, \dots, J$ to denote grid cells in every county. For ease of notation, I do not index variables by county and year.

I use weather data from the ERA5 dataset covering the period from 1979 to 2021. The original ERA5 dataset provides hourly data, but for this analysis, I use the aggregated the data at the daily level using Google Earth Engine (GEE).¹⁹ Specifically, I use the minimum temperature ($TN_{j,d}$) and maximum temperature ($TX_{j,d}$) within each day constructed by selecting the lowest and highest values, respectively, from the 24 measurements. By leveraging these daily grid-cell data points, I construct all the variables necessary for my analysis in county level using geemap package in Python (Wu, 2020).

In the analysis for the year 2000, I derive heatwave thresholds using temperature data from 1980 to 1999. I calculate these thresholds using 15-day or 5-day windows and considering both the 90th and 95th percentiles for maximum and minimum temperatures. Once established, I apply these thresholds to calculate heatwaves for each county from 2000 to 2019 for the definition of heatwaves without moving averages as in Perkins and Alexander (2012) and Kim et al. (2020).

To account for adaptation, I update the thresholds using a moving window approach. For instance, to calculate heatwaves for the year 2001, I determine the thresholds using temperature data from 1981 to 2000. For the year 2002, I use data from 1982 to 2001, and so on. Table 1 provides a summary of the definitions for each variable considered in the analysis.

Figures 4 and 5 show the number of heatwaves in 2012 for 8 distinct definitions for day heatwaves, that is, heatwaves calculated using maximum daily temperature. In Figure 4, all panels use definitions requiring at least 6 consecutive days above the defined thresholds, while in Figure 5, all panels use definitions requiring at least 3 consecutive days above the thresholds. A comparison between the two figures reveals that the definitions with 3 consecutive days result in a higher number of heatwaves as expected.

The top two panels in Figure 4 illustrate the difference in the number of observed heatwaves when considering only the moving average in threshold calculation while keeping everything else constant. Both panels display the count of intervals lasting at least 6 consecutive days, with maximum temperatures exceeding the 95th percentile of the distribution calculated using a 15-day window centered on a day within 2012. Comparing the panels reveals that the use of a moving average alters the distribution of heatwaves when calculating the heatwave thresholds.

The two panels on the left-hand side of Figure 4 demonstrate the impact of the window size centered on the day of focus when considering only that aspect while keeping everything else constant. Both panels show the count of intervals lasting at least 6 consecutive days, with maximum temperatures exceeding the 95th percentile of the distribution calculated using a moving average for heatwave thresholds. Comparing the panels reveals that the window size also influences the distribution of heatwaves, albeit to a lesser extent than when using a moving average (as seen in the top-right panel of Figure 4).

The top-left and bottom-right panels of Figure 4 illustrate the influence of percentiles on the number of observed heatwaves when considering only that aspect while keeping everything else constant. Both panels display the count of intervals lasting at least 6 consecutive days, with thresholds calculated using a moving average and a 15-day window. As anticipated, using a lower percentile as a threshold increases the number of observed heatwaves in specific locations. Similarly, some areas where no heatwaves are observed when the 95th percentile is used may exhibit heatwaves when 90th percentile thresholds are applied.

4.3. Economic data

I use personal income per capita data from the Bureau of Economic Analysis (BEA) found in the CAINC1 County and MSA personal income summary tables. The data covers the period from 1969 to 2019 in current dollars. To standardize all values to 2015 terms, I apply the Consumer Price Index (CPI) from

¹⁹ https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_DAILY#description.

Table 1. Definitions of heatwaves

Variable	Definition
Temperature	
$TN_{j,d}$	Daily temperature minimum: temperature minimums almost always occur at night
$TX_{j,d}$	Daily temperature maximum: temperature maximums almost always occur in daytime
TN_d	Daily minimum T °C: $\sum_j TN_{j,d} / J$
TX_d	Daily maximum T °C: $\sum_j TX_{j,d} / J$
$TNp(k)_d$	p^{th} percentile of the 20-year distribution of TN_d in a k -day window centered on d beginning one year before to the year being analyzed
$TXp(k)_d$	p^{th} percentile of the 20-year distribution of TX_d in a k -day window centered on d beginning one year before to the year being analyzed
$TNp(k)_d(NM)$	p^{th} percentile of the 1980–1999 distribution of TN_d in a k -day window centered on d
$TXp(k)_d(NM)$	p^{th} percentile of the 1980–1999 distribution of TX_d in a k -day window centered on d
Heatwaves	
# of HW (Xp-k-c)	Number of intervals of at least c consecutive days in which $TX_d > TXp(k)_d$. Calculated annually, and separately for three consecutive month intervals: January, February, and March (JFM); April, May, and June (AMJ); and July, August, and September (JAS)
# of HW (Np-k-c)	Number of intervals of at least c consecutive days in which $TN_d > TNp(k)_d$. Calculated annually, and separately for three consecutive month intervals: January, February, and March (JFM); April, May, and June (AMJ); and July, August, and September (JAS)
# of HW (Xp-k-c-NM)	Number of intervals of at least c consecutive days in which $TX_d > TXp(k)_d(NM)$. Calculated annually, and separately for three consecutive month intervals: January, February, and March (JFM); April, May, and June (AMJ); and July, August, and September (JAS)
# of HW (Np-k-c-NM)	Number of intervals of at least c consecutive days in which $TN_d > TNp(k)_d(NM)$. Calculated annually, and separately for three consecutive month intervals: January, February, and March (JFM); April, May, and June (AMJ); and July, August, and September (JAS)
# of HW days (Xp-k-c)	Number of days in which $TX_d > TXp(k)_d$ for at least c consecutive days. Calculated annually, and separately for three consecutive month intervals: January, February, and March (JFM); April, May, and June (AMJ); and July, August, and September (JAS)
# of HW days (Np-k-c)	Number of days in which $TN_d > TNp(k)_d$ for at least c consecutive days. Calculated annually, and separately for three consecutive month intervals: January, February, and March (JFM); April, May, and June (AMJ); and July, August, and September (JAS)
# of HW days (Xp-k-c-NM)	Number of days in which $TX_d > TXp(k)_d(NM)$ for at least c consecutive days. Calculated annually, and separately for three consecutive month intervals: January, February, and March (JFM); April, May, and June (AMJ); and July, August, and September (JAS)

Continued

Table 1. Continued

Variable	Definition
# of HW days (Np-k-c-NM)	Number of days in which $TN_d > TNp(k)_d(NM)$ for at least c consecutive days. Calculated annually, and separately for three consecutive month intervals: January, February, and March (JFM); April, May, and June (AMJ); and July, August, and September (JAS)
Longest HW (Xp-k-c)	Number of days in the longest period during which $TX_d > TXp(k)_d$ for at least c consecutive days
Longest HW (Np-k-c)	Number of days in the longest period during which $TN_d > TNp(k)_d$ for at least c consecutive days
Longest HW (Xp-k-c-NM)	Number of days in the longest period during which $TX_d > TXp(k)_d(NM)$ for at least c consecutive days
Longest HW (Np-k-c-NM)	Number of days in the longest period during which $TN_d > TNp(k)_d(NM)$ for at least c consecutive days

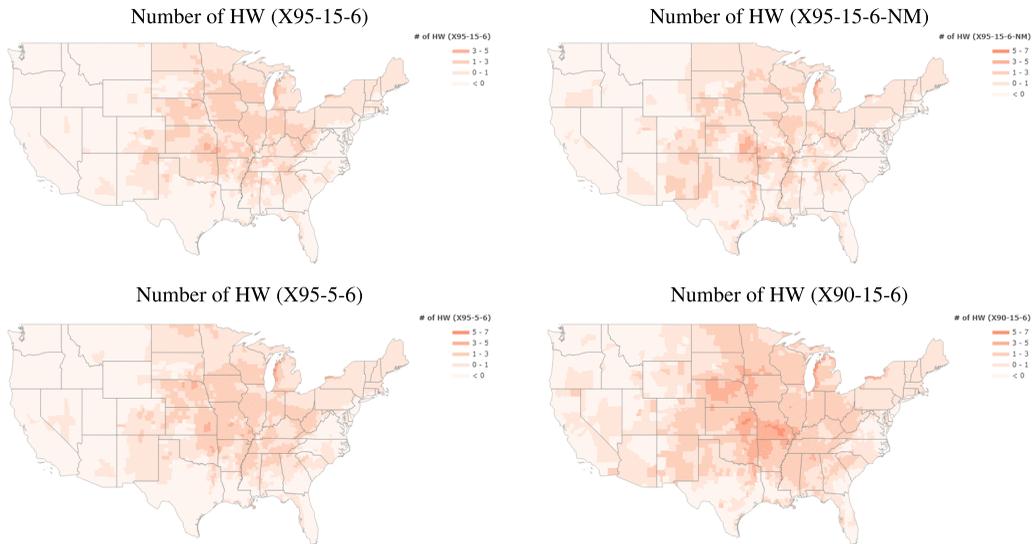


Figure 4. Heatwaves in 2012—comparison of definitions requiring 6 consecutive days.

Federal Reserve Economic Data (FRED). My focus is on analyzing the effects of heatwaves on the economy during the 21st century; therefore, I concentrate on data beginning after 1998.²⁰

To enhance the robustness of the data, I apply trimming by excluding the upper and lower 1 percentiles. Specifically, I calculate the 99th and 1st percentiles of GDP per capita growth across the entire sample, and remove observations that fall above or below these thresholds. Additionally, if lagged observations exceed these thresholds, they are also omitted from the analysis. Finally, after merging the data, I exclude the District of Columbia along with Alaska and Hawaii from the analysis. Table 2 presents the summary statistics both before and after this trimming process.

The left panel of Figure 6 displays a map showcasing the county-level GDP growth between 2018 and 2019. Note that certain counties in Virginia do not align with weather event data due to variations in the

²⁰ As explained in Section 4.4, I use the first differences and first two lags of the data. To include the second lag of the first difference appearing in 2001, data from 1998 is needed.

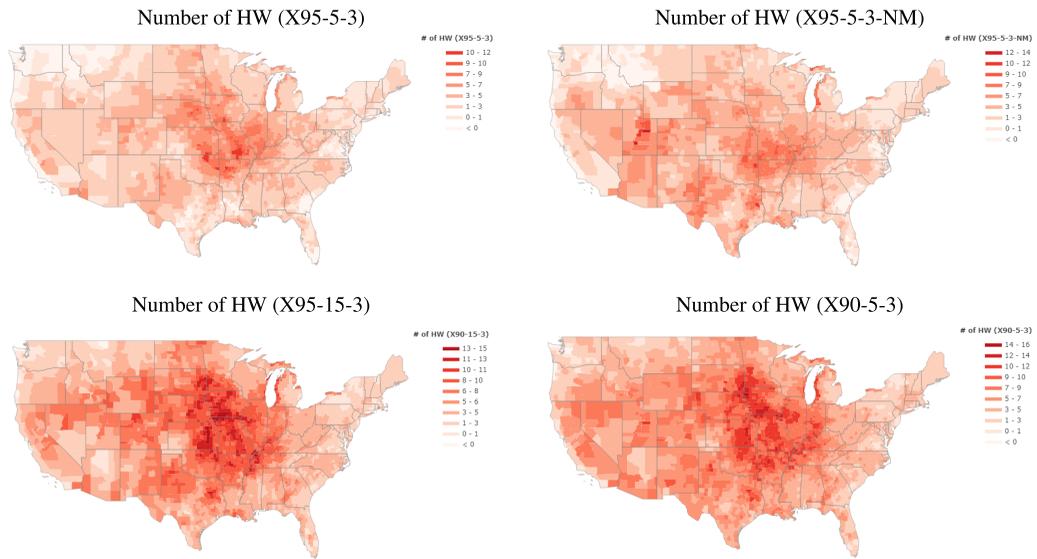


Figure 5. Heatwaves in 2012—comparison of definitions requiring 3 consecutive days.

Table 2. Summary statistics of GDP growth before and after trimming

	GDP growth	First lag of GDP growth	Second lag of GDP growth
Before trimming			
Count	58,059	58,059	58,059
Mean	0.01	0.01	0.01
Std	0.06	0.06	0.06
Min	−0.89	−0.89	−0.89
%	−.01	−.01	−.01
%	.03	.03	.03
Max	0.87	0.87	0.87
After trimming			
Count	55,382	55,382	55,382
Mean	0.01	0.01	0.01
Std	0.04	0.04	0.04
Min	−0.15	−0.15	−0.15
%	−.01	−.01	−.01
%	.03	.03	.03
Max	0.19	0.19	0.19

definition of counties by the BEA.²¹ The right panel of Figure 6 displays the map reflecting the trimmed data. As depicted in the figure, certain counties are dropped from the analysis for this particular year due to

²¹ Specifically, their footnote writes “Virginia combination areas consist of one or two independent cities with 1980 populations of less than 100,000 combined with an adjacent county. The county name appears first, followed by the city name(s). Separate estimates for the jurisdictions making up the combination area are not available. Bedford County, VA includes the independent city of Bedford for all years.”

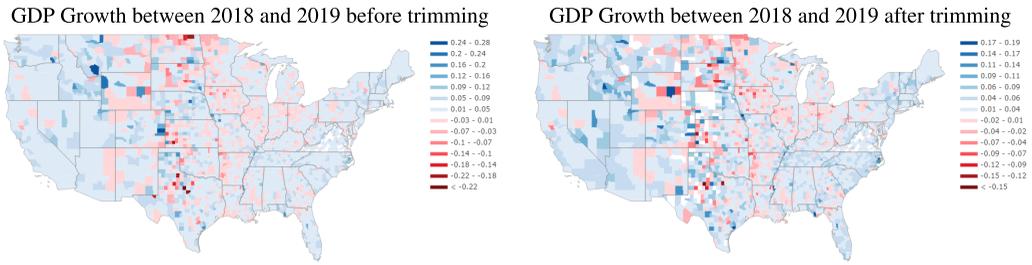


Figure 6. GDP Growth in US counties between 2018 and 2019. Each panel shows the first difference of the $\Delta \log(\text{Personal Income per Capita})_t$ (in 2015 terms) for $t = 2019$ before and after trimming the upper and lower 1 percentile.

the trimming procedure. This trimming approach helps to mitigate the potential influence of extreme observations and improve the reliability of the results.

4.4. Econometric model and feature selection

The econometric model’s objective is not to forecast personal income per capita for counties based on heatwave occurrences. Instead, its primary goal is to assess the impact of extreme heatwaves on the economy. To achieve this, I control for county-specific attributes that remain constant over time. For instance, factors like a county’s state affiliation and geographical location can influence its economic growth trajectory and weather events. Accounting for these time-invariant attributes is crucial as they may be correlated both with the variable we want to analyze causally and with the outcome variable, potentially leading to biased estimates if omitted (see Equation 3).

To account for yearly factors that are common to all counties and weather events, I include year fixed effects in the econometric model. These effects help capture simultaneous influences on both climate and economic data, such as El Niño events or global recessions.

The econometric specification, as shown in Equation 7, incorporates county and year fixed effects denoted by κ_i and τ_t , respectively. In the equation, y_{it} represents the log of personal income per capita for county i in year t . The variable Δy_{it} corresponds to the growth in personal income per capita, calculated as $\Delta y_{it} = \log(gdp_{it}) - \log(gdp_{it-1}) = \log\left(\frac{gdp_{it}}{gdp_{it-1}}\right)$. Additionally, ΔX_{it} denotes the first difference of heatwave variables introduced in Section 4.2. Taking these first differences removes the mean from the heatwaves, effectively controlling for serial correlation and non-stationarity of levels.²² The first two lags of the dependent variable are also included on the right-hand side as potential control variables to account for growth dynamics. Finally, ε_{it} represent the error terms clustered by county.

$$\Delta y_{it} = \kappa_i + \tau_t + \Delta X_{it}\beta + \Delta y_{it-1} \cdot \alpha_1 + \Delta y_{it-2} \cdot \alpha_2 + \varepsilon_{it} \tag{7}$$

To identify the most relevant heatwave definition among the 32 considered, I employ the Group LASSO technique introduced in Section 3.2. Since our objective is not to forecast personal income per capita solely based on heatwave occurrences, I choose a hyperparameter that selects only one heatwave group rather than optimizing the penalty weight to forecast out-of-sample observations. This allows to determine which of the heatwave definitions best explains the variations in personal income per capita growth. Each heatwave group comprises the 9 distinct variables outlined in Table 1. Additionally, there is a 33rd group containing the first two lags of the dependent variable, which is also included as a potential explanatory group.

²² Refer to Kahn et al. (2021) for a comprehensive discussion on taking first-differences of the right-hand side variables.

Before conducting the Group LASSO, it is crucial to ensure that the selection process incorporates the inclusion of the fixed effects in Equation 7. Hence, I use projection matrices to enforce the fixed effects. Assume K is the matrix representing county fixed effects, T is the matrix representing year fixed effects and $B = [K \ T]$. Notice that the following regression have the exact same coefficients as the regression in Equation 7:

$$\Delta y_{it} - B(B^T B)^{-1} B^T \Delta y_{it} = \left(I - B(B^T B)^{-1} B^T \right) \cdot \Delta X_{it} \cdot \beta + \left(I - B(B^T B)^{-1} B^T \right) \cdot \Delta y_{it-1} \cdot \alpha_1 - \left(I - B(B^T B)^{-1} B^T \right) \cdot \Delta y_{it-2} \cdot \alpha_2 + \varepsilon_{it}$$

I use the `asgl` package in Python for the Group LASSO analysis (Mendez-Civieta et al., 2021). When setting $\lambda = 0.034$ in Equation 6, the two selected groups are the one containing the first two lags of the dependent variable and the heatwaves defined using $TX95(15)_6$, representing day heatwaves that exceed the 95th percentile of maximum temperatures within a 15-day window centered on each day, for at least six consecutive days, using a moving average.²³

The Group LASSO’s selection of the heatwave definition using the moving window average to calculate thresholds is intriguing. This choice indicates that considering adaptation in the heatwave definition explains a greater portion of the variation in personal income per capita growth. Consequently, caution is necessary when projecting the impacts of changing climate into the future, especially when coefficients are obtained solely from focusing on average temperatures and historical data.

Another important observation is that the most extreme definition of heatwaves is being selected. As discussed in Section 4.2, requiring six consecutive days and calculating the threshold using the 95th percentiles results in fewer heatwaves. This finding suggests that as the extremity of events increases, their effect on the economy becomes more significant.

4.4.1. Regression results

Finally, to find out the heatwave variable that is dropped the last among this group I conduct a Sparse Group LASSO (SGL). The SGL has two hyper-parameters α and λ as shown in Equation 8, where I keep the notation in Equation 6:

$$\min_p \left(\left\| y - \sum_{i=1}^k X_k w_{G_i} \right\|_2^2 + \alpha \lambda \|w\|_1 + (1 - \alpha) \lambda \sum_{i=1}^{33} \beta_i \|w_{G_i}\|_q \right) \tag{8}$$

By setting $\lambda = 0.0355$ and $\alpha = 0.465$, the Sparse Group LASSO selects the first two lags of the dependent variable and the first difference of # of HW (X95-15-6). After this selection, I perform the regression in Equation 9 using nonstandardized variables:

$$100 * \Delta y_{it} = \kappa_i + \tau_t + \beta \cdot \Delta \# \text{ of HW}(X95 - 15 - 6) + \alpha_1 \cdot \Delta y_{it-1} + \alpha_2 \cdot \Delta y_{it-2} + \varepsilon_{it} \tag{9}$$

The regression yields a coefficient of $\beta = -0.126$ with a clustered standard error of 0.0198. Since we multiplied the dependent variable by 100, this implies that one more heatwave occurrence corresponds to a decline of 0.126% on average in personal income per capita growth. The coefficient is negative and statistically different from zero at the 99.9th confidence interval.

5. Conclusion

In this article, I first introduced common issues to understanding the channels from which climate change is affecting the economic welfare of countries and in evaluating the climatic events that arise due to increasing temperature. I surveyed articles in the CS literature that can be helpful for two key issues that

²³Note that I standardized both the left and right hand side variables such that they have 0 mean and a standard deviation of 1 before the Group LASSO.

researchers are faced with: feature selection and causality. Lastly, I presented an application showcasing how feature selection can be effectively used in economics research to address climate change concerns.

Throughout the article, I discuss several tools from the CS literature with potential to assist in our understanding of the ways in which rising temperatures impact the occurrence of climatic events, which in turn influence economic activities. These tools could be tuned according to the current needs in climate change research that can address some open problems in feature selection algorithms, which could then be used to infer the causality between different climate events.

I employed feature selection techniques to analyze the impact of heatwaves on economic outcomes. I considered multiple heatwave definitions and generated 32 distinct measures, each with 9 individual metrics. Using Group LASSO, I identified the optimal heatwave definition that explains the variation in personal income per capita growth in U.S. counties during the 21st century, revealing a significant negative impact.

As discussed throughout the article, a key objective in estimating the detrimental effects of extreme events on the economy is to calculate the price of GHG emissions. This calculation guides policymakers in determining the appropriate carbon tax or the amount of government revenues that should be allocated to mitigate climate change. To achieve this, the next step is to determine the effect of GHG emissions on the number of heatwaves, using the heatwave definition selected by the Group LASSO.

For instance, we aim to discern how many of these heatwave occurrences are attributed to anthropogenic climate change and how many would have occurred even without human-induced climate alterations. Unraveling this causal effect requires input from scientists knowledgeable about the complex earth system. The collaboration of earth scientists and computer scientists, using causality-based feature selection techniques and the abundance of available data, can serve as a guide to address this question effectively.

To conclude, an interdisciplinary approach that includes earth scientists, computer scientists and economists would be beneficial in devising feasible and effective policy suggestions capable of mitigating the effects of climate change and adopting to these effects.

Acknowledgments. I am grateful for the helpful comments of Tamer Kahveci and Emanuele Massetti. I would like to thank Jim Hoover for providing access to the HiperGator computational resources at the University of Florida. I also thank to the participants of 92nd Southern Economic Association (SEA) conference for suggestions. I am also grateful for the valuable feedback provided by two anonymous referees and the editor.

Author contribution. Conceptualization: B.A.; Data curation: B.A.; Methodology: B.A.; Writing – original draft: B.A.; Writing – review & editing: B.A.

Competing interest. The author declares none.

Data availability statement. The data used for Figure 1 are taken from ERA5 dataset and country border information is taken from Food and Agricultural Organization of the United Nations and calculations are done using Google Earth Engine and Python. The code to generate Figure 1 and the application presented in Section 4 can be found in <https://github.com/bakyapi/Figure-1-of-Machine-Learning-and-Feature-Selection-Applications-in-Economics-and-Climate-Change>.

Ethics statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study countries.

Funding statement. This work received no specific grant from any funding agency, commercial or not-for-profit sectors.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/eds.2023.36>.

References

- Acemoglu D and Robinson JA** (2012) *Why Nations Fail: The Origins of Power, Prosperity and Poverty*, Vol. 1. New York: Crown, pp. 1–529.
- Akyapi B, Bellon M and Massetti E** (2022) Estimating Macro-Fiscal Effects of Climate Shocks from Billions of Geospatial Weather Observations. *IMF Working Papers* 2022/156, 1–70.
- Arrhenius S** (1896) On the influence of carbon acid in the air upon the temperature of the ground. *Philosophical Magazine and Journal of Science* 41(5), 237–276.

- Arya V, Shanmugam K, Aggarwal P, Wang Q, Mohapatra P and Nagar S (2021) Evaluation of causal inference techniques for AIOps In *8th ACM IKDD CODS and 26th COMAD*, pp. 188–192, <https://doi.org/10.1145/3430984.3431027>.
- Athey S and Guido I (2019) Machine learning methods that economists should know about. *Annual Review of Economics* 11 (2019), 685–725.
- Basu S, Shojaie A and Michailidis G (2015) Network granger causality with inherent grouping structure. *Journal of Machine Learning Research* 16(31), 417–453.
- Belloni A, Chen D, Chernozhukov V and Hansen C (2012) Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80(6), 2369–2429.
- Belloni A, Chernozhukov V and Hansen C (2014a) High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives* 28(2), 29–50.
- Belloni A, Chernozhukov V and Hansen C (2014b) Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies* 81(2), 608–650.
- Bianchi F, Ludvigson S and Ma S (2022) Belief distortions and macroeconomic fluctuations. *American Economic Review* 112(7), 2269–2315.
- Bilal A and Rossi-Hansberg E (2023) *Anticipating Climate Change Across the United States*. NBER Working Paper 31323.
- Burke M, Hsiang S and Miguel T (2015) Global non-linear effect of temperature on economic production. *Nature* 527, 235–239.
- Dell M, Jones B and Olken B (2012) Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics* 4(3), 66–95.
- Dell M, Jones B and Olken B (2014) What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature* 52(3), 740–798.
- Dzhamashev R and Tursunaliyeva A (2021) *Synthetic Instrumental Variables*. SSRN Working Paper.
- Gui J, Sun Z, Ji S, Tao D and Tan T (2017) Feature selection based on structured sparsity: A comprehensive study. *IEEE Transactions on Neural Networks and Learning Systems* 28(7), 1490–1507.
- Imbens G (2020) Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *IEEE Transactions on Neural Networks and Learning Systems* 58(4), 1129–1179.
- IPCC (2021) Summary for policymakers. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York: Cambridge University Press. In press, <https://doi.org/10.1017/9781009157896>.
- Jain A and Zongker D (1997) Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(2), 153–158.
- Jangyodsuk P, Seo D, Elmasri R and Gao J (2015) Flood prediction and mining influential spatial features on future flood with causal discovery. In *2015 IEEE 15th International Conference on Data Mining Workshops*. Atlantic City, NJ, USA. <https://doi.org/10.1109/ICDMW.2015.111>.
- Jovic A, Brkic K and Bogunovic N (2015) A review of feature selection methods with applications. In *38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. Croatia: Opatija. <https://doi.org/10.1109/MIPRO.2015.7160458>.
- Kahn ME, Mohaddes K, Ryan NCN, Pesaran MH, Raissi M and Yang J-C (2021) Long-term macroeconomic effects of climate change: A cross-country analysis. *Energy Economics* 12, 1–13.
- Kim Y, Min S, Zhang X, Sillmann J and Sandstad M (2020) Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. *Weather and Climate Extremes* 29, 1–15.
- Koneru BNG and Vasudevan V (2019) Sparse artificial neural networks using a novel smoothed LASSO penalization. *IEEE Transactions on Circuits and Systems II: Express Briefs* 66(5), 848–852.
- Kotz M, Levermann A and Wenz L (2022) The effect of rainfall changes on economic production. *Nature* 601, 223–227.
- Kotz M, Wenz L, Stechemesser A, Kalkuhl M and Levermann A (2021) Day-to-day temperature variability reduces economic growth. *Nature Climate Change* 11, 319–325.
- Lai P, Zhang M, Ge Z, Hao B, Song Z, Huang J, Ma M, Yang H and Han X (2020) Responses of seasonal indicators to extreme droughts in Southwest China. *Remote Sensing* 12(5), 1–17.
- Ling Z, Yu K, Wang H, Li L and Wu X (2021) Using feature selection for local causal structure learning. *IEEE Transactions on Emerging Topics in Computational Intelligence* 5(4), 530–540.
- Liu Y, Niculescu-Mizil A, Lozano A and Lu Y (2010) Learning temporal causal graphs for relational time-series analysis. In *Proceedings of the 27th International Conference on Machine Learning*. Madison, WI: Omnipress, pp. 687–694.
- Luo Y, Peng J and Ma J (2020) Using feature selection for local causal structure learning. *Nature Machine Intelligence* 2, 426–427.
- Meek C (1995) Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence (UAI'95)*. San Francisco, CA: Morgan Kaufmann Publishers Inc., pp. 411–418.
- Mendez-Civieta A, Aguilera-Morillo MC and Lillo RE (2021) Adaptive sparse group LASSO in quantile regression. *Advances in Data Analysis and Classification* 15, 547–573.
- Mobil Corporation (1997) *Climate Change: A Degree of Uncertainty*. New York Times.
- Moscona J and Sastry K (2022) Does directed innovation mitigate climate damage? Evidence from US agriculture. *Quarterly Journal of Economics* 138(2), 637–701.
- Newell RG, Prest BC and Sexton SE (2021) The GDP-temperature relationship: Implications for climate change damages. *Journal of Environmental Economics and Management* 108, 1–26.

- Perkins SE and Alexander LV** (2012) On the measurement of heatwaves. *Journal of Climate* 26(13), 4500–4517.
- Scholkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, Goyal A and Bengio Y** (2021) Toward causal representation learning. *Proceedings of the IEEE* 109(5), 612–634.
- Tibshirani R** (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society* 58(1), 267–288.
- Wu Q** (2020) Geemap: A python package for interactive mapping with Google earth engine. *The Journal of Open Source Software* 5(51), 1–3.
- Yeh C, Perez A, Driscoll A, Azzari G, Tang Z, Lobell D, Ermon S and Burke M** (2020) Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications* 11, 1–11.
- Yu K, Guo X, Liu L, Li J, Wang H, Ling Z and Wu X** (2020a) Causality-based feature selection: Methods and evaluations. *ACM Computing Surveys* 53(5), 1–36.
- Yu L and Liu H** (2004) Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research* 5, 1205–1224.
- Yu K, Liu L and Li J** (2021) A unified view of causal and non-causal feature selection. *ACM Transactions on Knowledge Discovery from Data* 15(4), 1–46.
- Yu K, Liu L, Li J, Ding W and Le TD** (2020b) Multi-source causal feature selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42(9), 2240–2256.
- Zheng X, Dan C, Aragam B, Ravikumar P and Xing EP** (2020) Learning sparse nonparametric DAGs. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*. Palermo, Italy, Vol. 108, 1–11.
- Zou H** (2006) The adaptive Lasso and its Oracle properties. *Journal of the American Statistical Association* 101(476), 1418–1429.