Estimating Market Risk Resulting from Climate Change Physical Risk

University of Chicago

Department of Mathematics

Chintan Singh, Clara Xue, Som Mookherji, Tony Zu, Xinyu Liu

Faculty Advisor – Mark Hendricks

Executive Summary

Research

- ✓ Potential sector-wise economic impact of climate change on macroeconomic health indicators (CPI)
- ✓ Statistical modeling approaches in the ESG and market risk spaces (attributional models)

Modeling

- ✓ Costs associated with natural resources, waste incineration, air pollutant emissions & water costs have a collective tangible impact on CPI growth rate (Confidence Interval > 90%)

Testing

- ✓ Benchmarking against models in literature, industry specific standards, and intuition
- ✓ Discernible & reasonable statistical inferences obtained

Agenda

Literature Review

Goal: Estimate how price levels are impacted by climate change

01

Understanding, surveying, and exploring the possible fundamental physical risk drivers of climate change

02

Addressing and evaluating the potential price level impact resulting from physical risk-based damage functions as drivers of climate change

Literature Review (Part 1 of 2)

Literature	Hypothesis	Results	Key Influence & Implications
Parker M. et al. (2018): “Economics of Disasters and Climate Change”	Impact of natural disasters on CPI in developed economies	GHG emissions would intensify disasters and therefore affect price levels as the migration to alternative sources of energy continues	Natural events have tangible impacts on economic growth
Michael F. et al. (2013): "Proposed carbon tax policy in South Africa: learning from the experience of other countries and effect on consumer price index"	Quantitative estimation of the impact of carbon tax policy on South Africa’s consumer price index	Extensive multivariate hypothesis tests show that the introduction of a carbon tax in South Africa eventually led to a higher CPI	Primary drivers of CPI
Kim Matthes et al. (2021): "Relationship between Extreme weather index and CPI in US"	Survey on frequencies of extreme weathers and their potential influence on inflation as estimated using CPI	Extreme weather is attributable to GHG, waste landfill, sewage costs etc. have impact mainly on food and energy prices	Build an aggregate indicator of extreme weather using candidate environmental damage functions to augment model robustness

Literature Review (Part 2 of 2)

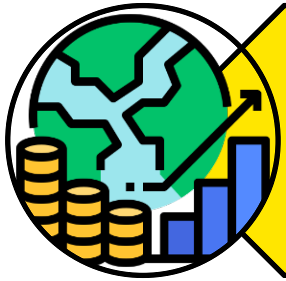
Literature Review	Model
Parker, M. (2018) ''Economics of Disasters and Climate Change''	$$\pi_{i,t} = \sum_{j=0}^{p} \beta_j D_{i,t-j} + \mu_i + \lambda_t + \nu_{it}$$
Michael (2013) "Proposed carbon tax policy in South Africa: learning from the experience of other countries and effect on consumer price index"	$$\gamma_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$
Kim, Matthes, Phan (2021) "Relationship between Extreme weather index and CPI in US"	$$\mathbf{y}_t = (1 - \tilde{z}_{t-1}) \left(\mathbf{m}_1 + \sum_{\ell=1}^{C} \mathbf{A}_{\ell,1} \mathbf{y}_{t-\ell} + \boldsymbol{\Sigma}_1 \mathbf{e}_t \right) + \tilde{z}_{t-1} \left(\mathbf{m}_2 + \sum_{\ell=1}^{C} \mathbf{A}_{\ell,2} \mathbf{y}_{t-\ell} + \boldsymbol{\Sigma}_2 \mathbf{e}_t \right)$$

Data Preparation

Data Collection: Choice of Variables

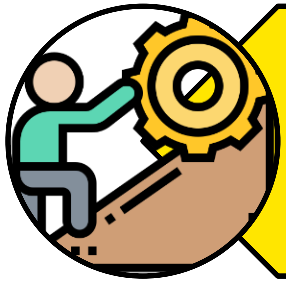
Physical risk factors as **independent variables (X)**:

Natural Resources, Waste Incineration, Air Pollutants, Water, Waste Disposal and Environmental Damage



Macro-economic indicators as **dependent variable (Y)**:

CPI (USA, national level)



Pinch of salt considerations:

Sparse environmental data at sector level

Effects of climate risk on inflation are tangible at best as it is primary driven by Fed’s monetary policy

Dataset Summary

- Summary of the variables used in our final analysis

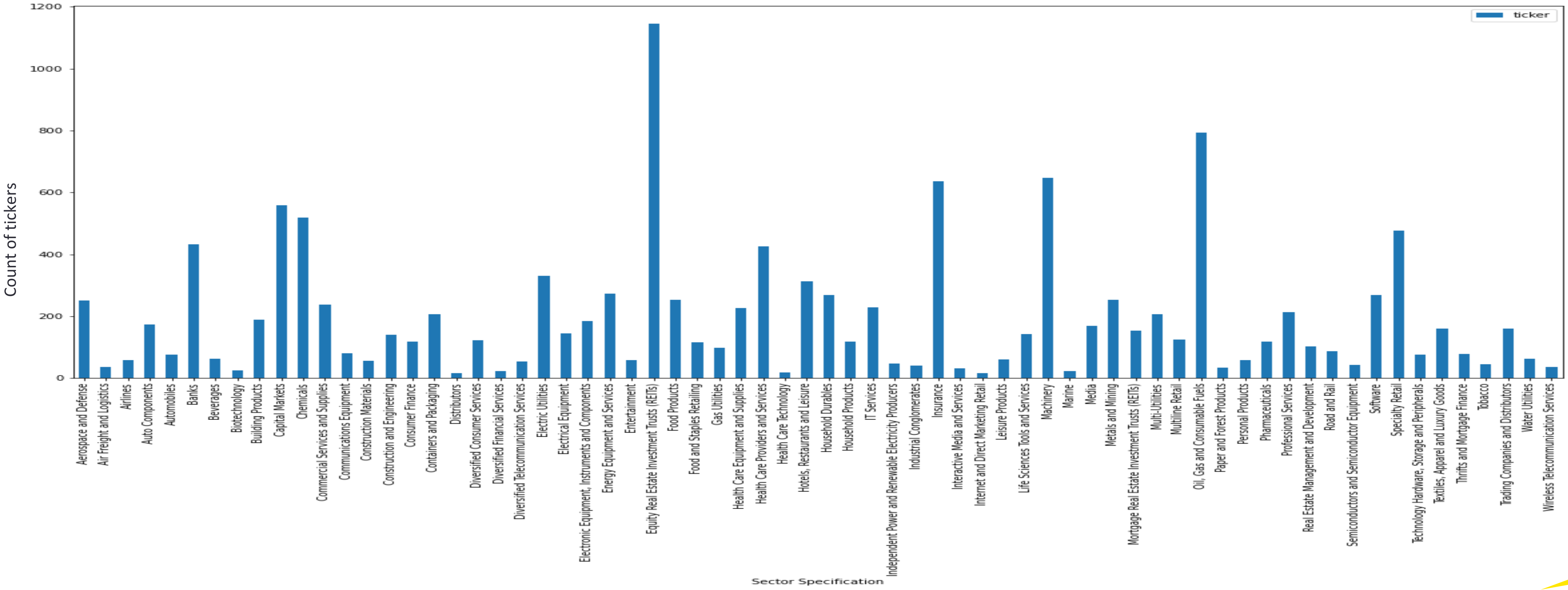
Type	Variable	Frequency	Time Range	Source
Dependent	CPI	Annual	2003-2020	Wharton Research Data Services (WRDS)
Independent	Natural Resources Expenditure			
	Direct Waste Incineration Costs			
	Air Pollutants' Emission Costs, Taxes & Penalties			
	Water Costs (acquisition, taxes, infrastructure etc.)			
	Indirect Costs Associated With Waste Disposal and Environmental Damage			

"

Exploratory Data Analysis

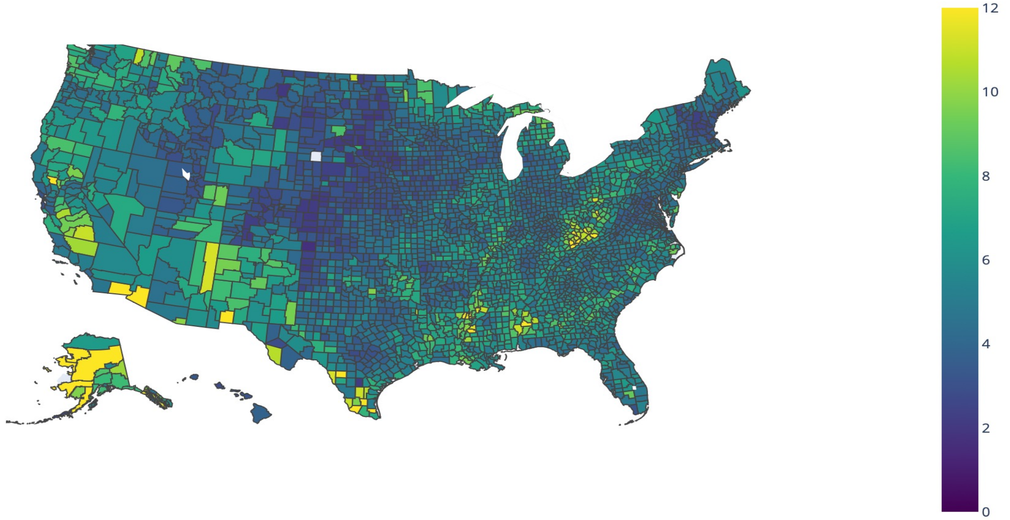
Sectoral Environmental Dataset Overview

Data Source	Timeline	Number of Tickers (Stocks)	Number of Sectors	Geographic Area of Focus	Number of Environmental Factors	Number of Factors (X-variables) in Consideration	Selection Criteria
Wharton Research Data Services (WRDS)	2003-2020	1321	68	United States	23	10	Statistical Significance



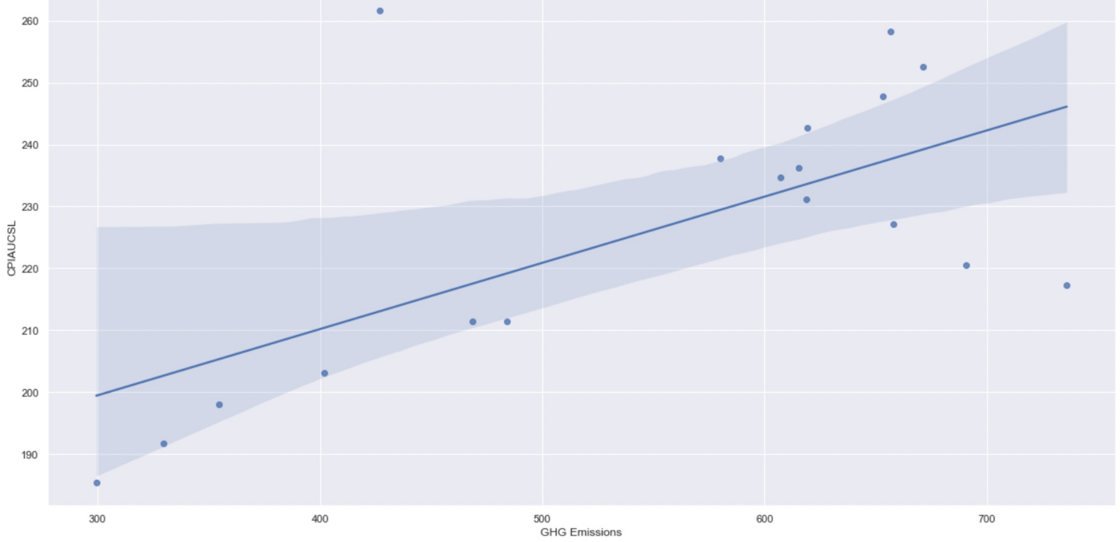
Preliminary Exploratory Analysis

- **Example: GHG Emissions**



Heat map for joint emissions (GHG Emissions & Waste Discharge) – Normalized

- **Example: CPI vs GHG Emissions (raw data) – expect linear relationship**



- **Potential genres of independent variables studied:**
  - Greenhouse Gases, Air Pollutant Emissions, Carbon Footprint
  - Natural Resources Costs, Waste Landfill, Waste Incineration, Nuclear Energy Usage Costs
  - Water Costs
- **Relation to CPI:**
  - Carbon tax gives rise to price inflation.
  - Different natural disasters could increase short-term price inflation.
  - Extreme weather conditions driven by GHG and inappropriate waste disposal mainly impact real CPI

Data Cleaning

- Frequency
  - Convert whole dataset into consistent time frequency (all data is annual)
- Time horizon
  - Start and end dates for each dataset are not the same
    - Alignment performed before running regressions in later steps
  - Targeted date range : 2003 through 2020
- Treated missing data (stochastic interpolation) & outliers using statistical methods
- Challenges
  - Sparse environmental data at sector level
    - Quality of original dataset was considerably **poor**.
  - Clustering by sectors (based on industry correlation and statistical proximity) to form 7 industries
  - Solved sparsity of sector level environmental data and gained more reliability through clustering and stacking inspired by panel regressions

Sector Clustering

Finance

- Consumer Finance
- REITS
- Insurance
- Mortgage
- …

Energy

- Oil & Gas
- Electricity
- Equipment & Services
- Gas Utilities
- …

Healthcare

- Pharmaceuticals
- Biotechnology
- Equipment & Services
- …

IT Telecom and Hardware

- Telecommunication Services
- Software
- Technology Hardware
- …

Infrastructure

- Aerospace & Defense
- Automobiles
- Construction
- Logistics
- Utilities
- …

Industrial

- Metals & Mining
- Construction Materials
- Chemicals
- Conglomerates
- …

Consumer Staples

- Beverages
- Household
- Retail
- Professional Services
- …

Data Transformations

- Transformation on the independent variables :

Preliminary Transformations
- Raw data
- Percentage change
- Single difference
- Double difference
- Raw X (Single Difference)
- Raw X (Double difference)
- Exponential - Log

Secondary Transformations
- No change
- Square
- Cube

Potential Further Transformations
- No change
- Lag 1 (VAR model)
- Lag 2 (VAR model)

From papers "Growth at Risk From Climate Change", "Impact of Climate Conditions on Economic Production", and "Climate Scenarios Database – Technical Documentation."

- Nomenclature Examples:
  - Natural Resources Direct Cost
  - Natural Resources Direct Cost Single Difference Squared
- First, for specific sector, univariate regression of CPI on all transformations of each variable
- Variables with higher R-squared and lower p-values are potentially good for multivariate regression

Univariate Analysis

Regression (Ordinary Least Squares)

Dependent Variable

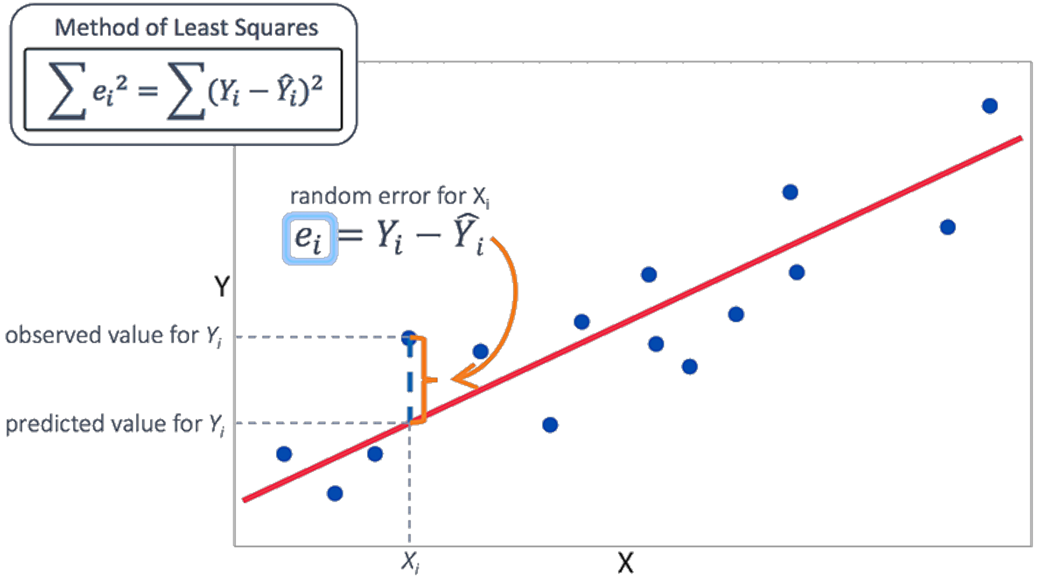
Population Y intercept

Population Slope Coefficient

Independent Variable

Random Error term

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{Linear component}} + \underbrace{\varepsilon_i}_{\text{Random Error component}}$$



- ✓ OLS Assumptions:
  - ✓ Linear relationship between X and Y
  - ✓ Residuals follows normal distribution with mean 0 and constant variance
  - ✓ No correlation among the residuals (or minimal serial autocorrelation)
  - ✓ No multicollinearity (mutual correlation in regressors)

Evaluation Metrics

R-squared	Measure of how well the regression model explains the dependent variable
p-value	The smaller the p-value, the more confident we are in the strength of linear relationship
Beta	Represents the expected change in dependent variable for a unit change in explanatory variable
RMSE (Root Mean Square Error)	Standard deviation of residuals – tells how concentrated the data is around the line of best fit
Residuals Mean	Check if errors have mean 0
Durbin Watson	Check error terms have no autocorrelation – closer to 0 shows more positive serial correlation; closer to 4 shows more negative serial correlation; value around 2 means no serial correlation.
Jarque-Bera	Check error terms have normal distribution – p-value greater than 0.10 means that we fail to reject H0 and show that error terms follow normal distribution
Breusch-Pagan	Check error terms have constant variance – p-value greater than 0.10 means that we fail to reject the null hypothesis and show that error terms have constant variance

Assumption validation for data

- Sample of results ordered by R-squared and p-value
- Restrict p-value from JB test and BP test to be less than 0.1; restrict DW stat within 1 and 3

	coef	p-value	R Sq	RMSE	Resid Mean	Durbin Watson	JB_resid_pvalue	BP-pval
Impact Ratio: LWP Direct & Indirect Cost (di_319450)_doublediff	6.089107	4.252613e-09	0.890563	7.396428	1.578984e-14	1.411724	0.832794	0.235548
Impact Ratio: LWP Direct Cost (di_319446)_doublediff_squared	0.478969	1.333433e-08	0.873889	7.939927	5.210647e-14	1.291410	0.714149	0.746034
Impact Ratio: LWP Direct Cost (di_319446)_doublediff	6.642453	2.130201e-08	0.866347	8.173893	-3.947460e-14	1.474391	0.734393	0.201228
Weighted Disclosure: Water (di_319570)	0.148546	1.007360e-06	0.784731	10.373637	1.263187e-14	1.474120	0.625861	0.781655
Weighted Disclosure: Water (di_319570)_diff	0.148546	1.007360e-06	0.784731	10.373637	1.263187e-14	1.474120	0.625861	0.781655
Intensity: Waste Direct Recycled (di_319544)	9.591437	4.979898e-05	0.652901	13.172449	1.073709e-13	1.610353	0.614942	0.897266
Intensity: Waste Direct Recycled (di_319544)_diff	9.591437	4.979898e-05	0.652901	13.172449	1.073709e-13	1.610353	0.614942	0.897266

- Top variables explain optimistic results for OLS assumption verification
  - P-value small enough to be show statistical significance, with highest R-squared of 0.89
  - Error assumptions satisfied in general, with DW stat close to 2, JB and BP p-values less than 0.1.

Variable selection post clustering

- Clustering all sectors and running univariate regression
- Select potential independent factors with low p-value, high R-squared
- Select variables across different risk categories

Best transformed variables	Beta	R-squared	P-value
Natural Resources Direct Cost_diff_squared	1.758773	0.093464	0.009013
Waste Incineration Costs_doublediff_squared	0.131316	0.140517	0.00002
Air Pollutants Indirect Cost_doublediff_squared	0.190541	0.15186	0.00009
Water Direct & Indirect Cost_squared	0.718408	0.132891	0.00005
Waste Disposal and Environmental Damage Indirect Cost_diff	9.214038	0.210050	5.156098e-05

Multivariate Analysis

Approach for Multi-variable Analysis

Pick the regressors that could best explain CPI in each sector from univariate analysis, Lasso, PCA

→

Run Multivariate regression on selected variables after adhering to assumptions and getting a model with significant p-values and highest R squared values.

Multi-regression Model

- All prior assumptions validated along with obtaining an acceptable VIF score to check for multi-collinearity
- Inflation is primarily a result of the Fed's monetary policy – the effects of these variables are tangible in nature and should be studied in conjunction to macroeconomic indicators for a better estimation of their true effects

Industry	const	Natural Resources Direct Cost_diff_squared	Waste Incineration Costs_doublediff_squared	Air Pollutants Indirect Cost_doublediff_squared	Water Direct & Indirect Cost_squared	Waste Disposal and Environmental Damage Indirect Cost_diff	R-squared
Finance	-0.131766	0.387986	0.334217	-0.575516	0.707929	-0.259776	0.118788
Energy	-0.074472	-0.40241	0.347581	0.231879	0.136221	0.552790	0.355293
IT Telecommunications and Hardware	-0.123976	-0.001386	0.006510	0.153216	0.466202	0.192805	0.150398
Infrastructure	-0.131039	-0.095410	0.178962	0.205919	0.169244	-0.044042	0.171417
Consumer Staples	-0.103378	-0.186916	0.112507	0.193338	0.398131	-0.024389	0.189913
Health	-0.083071	0.131282	-0.385048	0.593335	0.542141	-0.357696	0.521494
Industrial	-0.125443	0.405453	0.321601	0.504053	-0.264646	-0.408126	0.258223

Validation and Testing

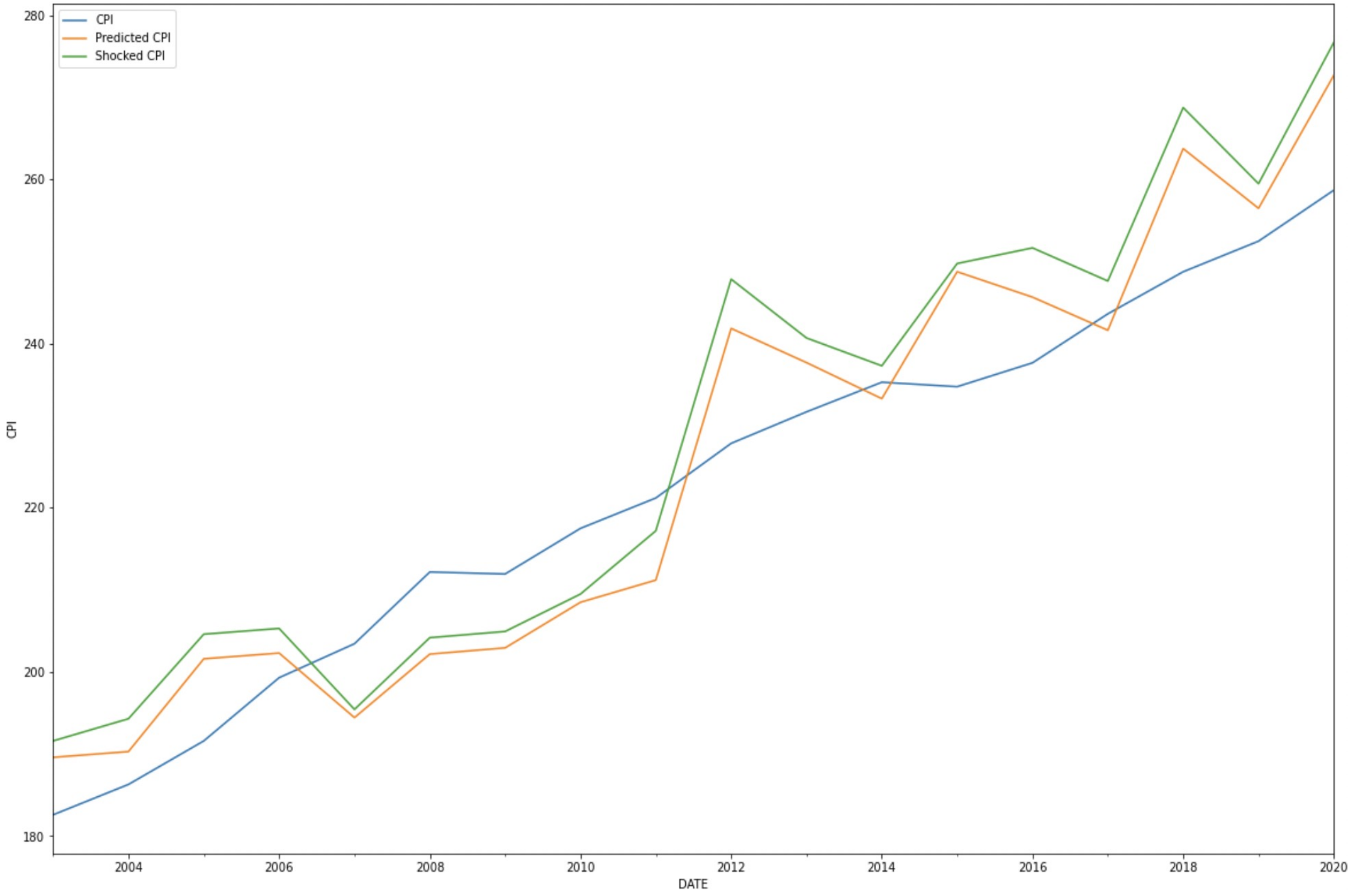
Scenario Analysis

- Issue with predicting climate change – what if our projections are too low?
- Solution – apply growth rates to raw data and find new beta coefficients

	Natural Resources Expenditure	Direct Waste Incineration Costs	Air Pollutants' Emission Costs, Taxes & Penalties	Water Costs	Indirect Costs Associated With Waste Disposal and Environmental Damage
Mean Growth Rate	10.75%	7.98%	5.61%	21.96%	1.30%

- Process:
  - Go through data points and find shocks for mean growth rate for each variable
  - Apply growth rate uniformly throughout data set
  - Perform multivariate regression again and find new beta coefficients

Results and Analysis of Scenario Testing



- **Analysis:**
  - Predicted CPI (Orange) is our original prediction, Shocked CPI (Green) is our estimate of CPI after applying the mean growth rates
  - The Shocked CPI closely follows the Predicted CPI, which is to be expected, as we propagated the growth rates evenly throughout the data
  - Next steps – find worst-case and best-case growth rates and redo analysis

Conclusion and Scope for Future Research

Concluding Remarks

This analysis recommends an approach to model the effects of environmental physical variables on macro-economic factors

Our analysis shows the scope to model the CPI levels from environmental factors

We do observe some explainability of CPI levels after transforming our input variables

We could also see the nature of hypothesized relationships across different industry sectors

However, as the reporting of environmental variables is still at a nascent stage, the models developed herein are not in equilibrium

The built model provides a general framework for attributing macroeconomic variable CPI on environmental factors

Scope for Future Research

Increase the Scope

- Our study period was limited to 2003 to 2020
- Only considered US Company Data
- Require more frequent data points to build time series models

Expand Research Methodology

- With more data points, penalized regression models can be implemented to control model overfits
- Time Series models incorporating lagged variable values can be deployed to study causal relationships over time

Back-testing Models

- The model performance can be benchmarked by undertaking back-testing

Questions?

Thank you!

Appendix

Modelling Approach and Literature Review

- Parker, M. et. Al. (2018) “Economics of Disasters and Climate Change”
- Michael (2013) "Proposed carbon tax policy in South Africa: learning from the experience of other countries and effect on consumer price index”
- Kim, Matthes, Phan (2021) "Relationship between Extreme weather index and CPI in US"

OLS Model Evaluation Metrics

R-squared	Measure of how well the regression model explains the dependent variable.	https://towardsdatascience.com/verifying-the-assumptions-of-linear-regression-in-python-and-r-f4cd2907d4c0
p-value	Smaller the p-value, the more confident we are in the strength of linear relationship.	
Beta	Represents the expected change in dependent variable for a unit change in explanatory variable.	
RMSE (Root Mean Square Error)	Standard deviation of residuals. Tells how concentrated the data is around the line of best fit.	https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/
Residuals Mean	Check errors have mean 0.	
Durbin Watson	Check error terms have no autocorrelation. Closer to 0 shows more positive serial correlation; closer to 4 shows more negative serial correlation. Value around 2 means no serial correlation.	https://medium.com/@analyttica/durbin-watson-test-fde429f79203
Jarque-Bera	Check error terms have normal distribution. P-value greater than 0.10 means that we fail to reject H0 and show that error terms follow normal distribution.	https://www.statisticshowto.com/jarque-bera-test/
Breusch-Pagan	Check error terms have constant variance. P-value greater than 0.10 means that we fail to reject H0 and show that error terms have constant variance.	https://en.wikipedia.org/wiki/Breusch–Pagan_test