Climate Change Regression

Melissa Cooper

Do a country's features affect its greenhouse gas emissions level?

Feature set

- Land Area
- Water Area
- Population
- GDP per capita
- Population below poverty line
- Electricity Production
- Electricity Consumption
- Refined petroleum Production
- Refined petroleum Consumption
- Natural gas Production
- Natural gas Consumption
- Year
- Country

Target

- CO2 Emissions

Methodology

Data

1. CIA's The World Factbook 2009-2018 - countries
2. US EIA (Energy Information Administration) - CO_2 emissions

Models

- LARS Path
- Linear, Ridge, Lasso, & Polynomial Regressions

Tools

- Web scraping: Selenium & Beautiful Soup
- Data manipulation: Numpy & Pandas
- Plotting: Matplotlib & Seaborn

Procedure

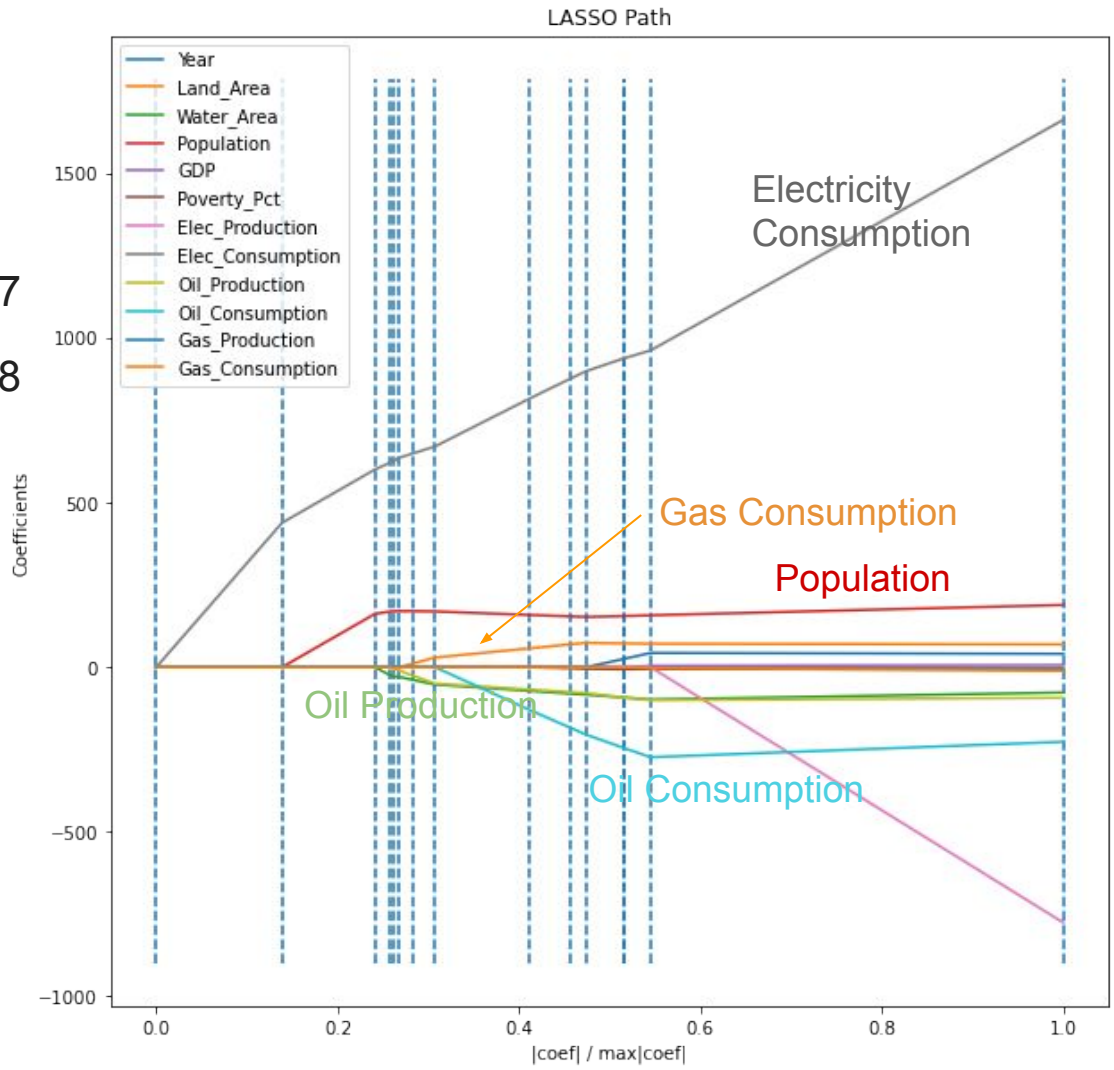
Preliminary R^2	*Val*	*Test*
Linear regression:	0.948	0.957
Ridge regression:	0.954	0.958
Polynomial regression:	0.832	------

Cross validation R^2

Linear regression:	0.9405
Ridge regression:	0.9406

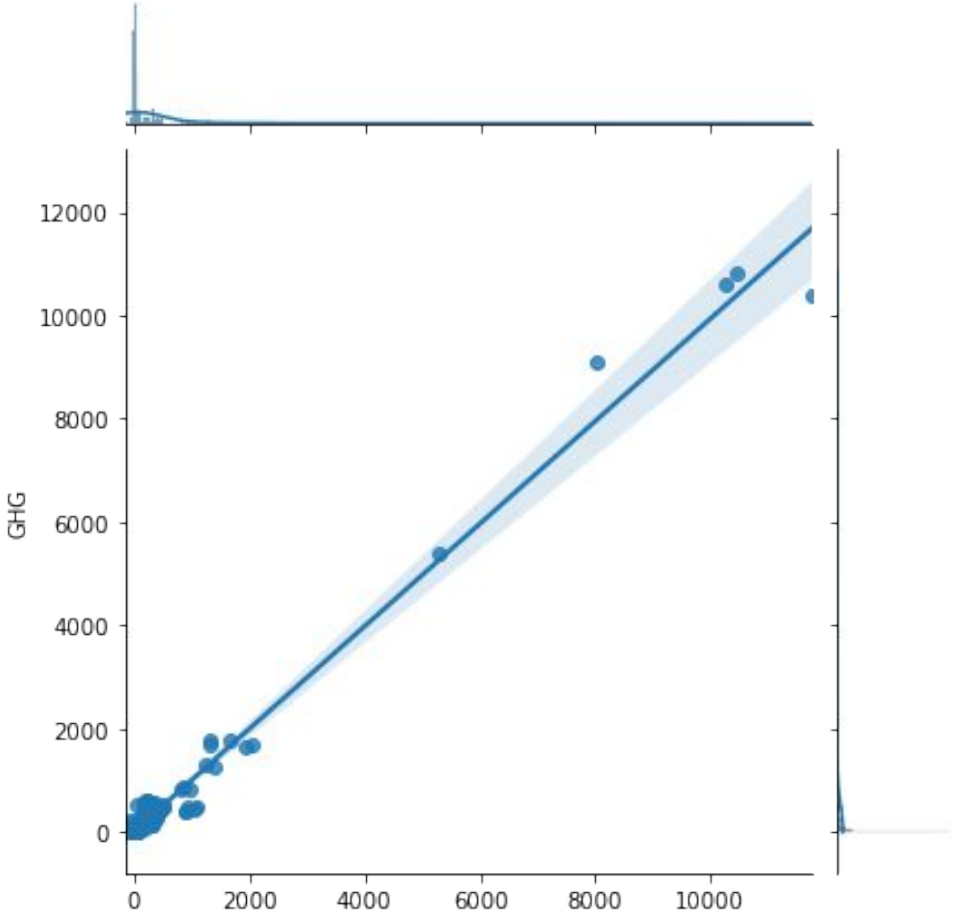
LARS Feature Importance Study

RidgeCV

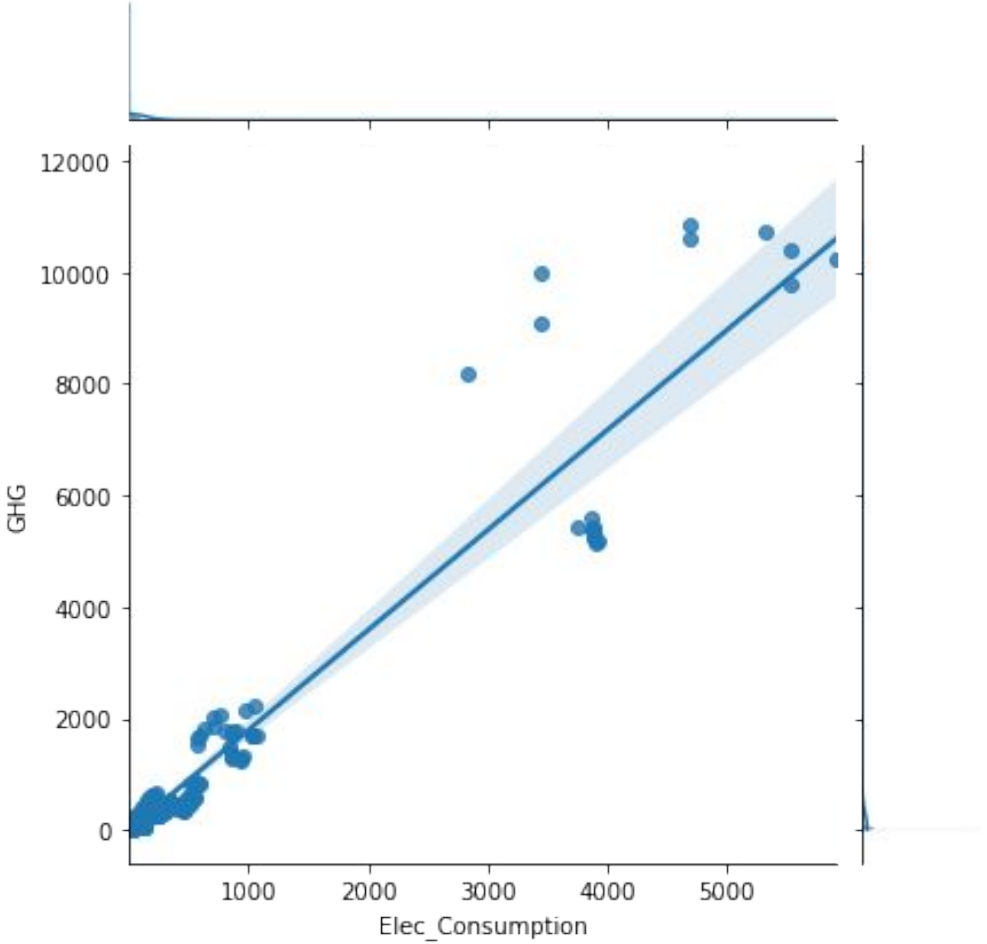


Results

Predicted values vs. test set



Electricity Consumption vs. COs Emissions



Conclusion: Ridge model predictions

Coefficients w/ best alpha:

Year	-4.89
Land_Area	81.35
Water_Area	-113.36
Population	148.75
GDP	3.36
Poverty_Pct	-6.37
Elec_Production	336.21
Elec_Consumption	613.52
Oil_Production	-104.88
Oil_Consumption	-250.09
Gas_Production	40.68
Gas_Consumption	-0.71

MAE = 69.3148

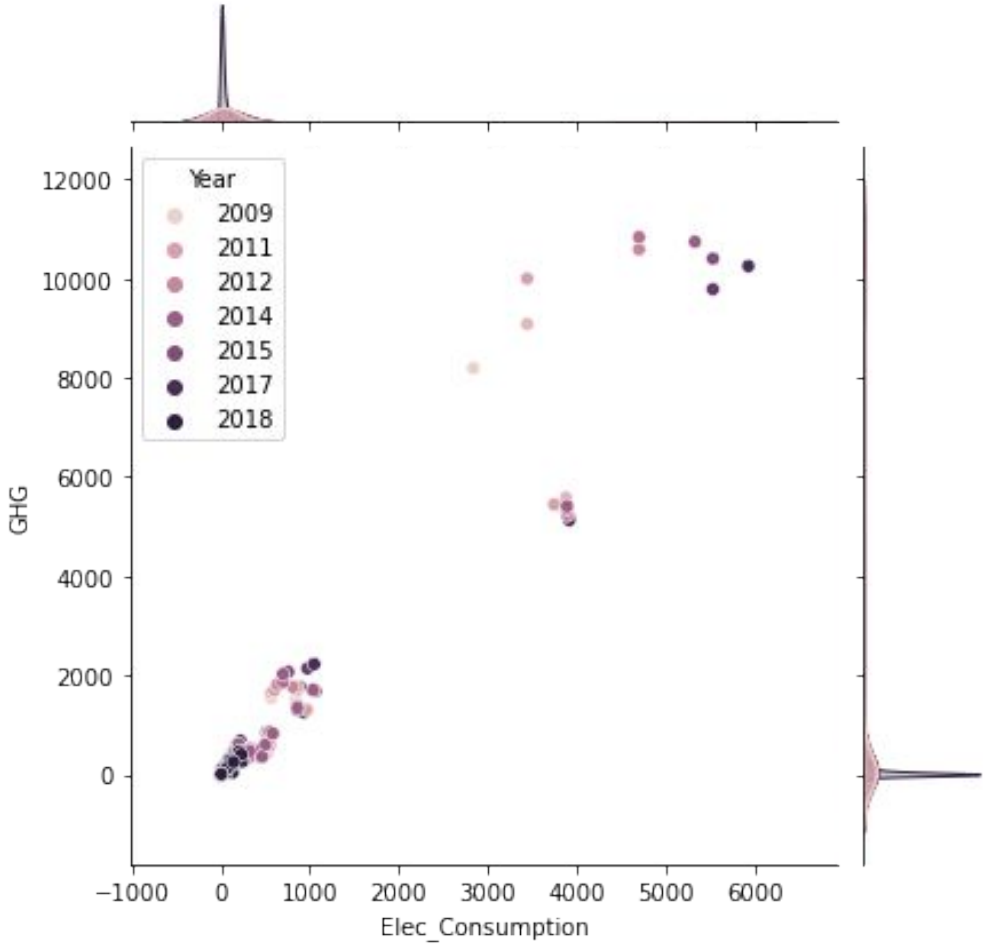
R^2 = 0.9846

Future Work

- Add countries as a feature - dummy variables by continent
- Add more years and do a time series prediction
- Divide countries into two bins: countries with emissions > 4,000 MMtonnes and the rest of the world

Appendix

Electricity Consumption vs. COs Emissions over Year



Regularization Effect