**Renewables Scenario Discovery Summary**

*Methods*

Four hundred runs from the Emissions Prediction and Policy Analysis (EPPA) model were used in this analysis. Fifty-eight original inputs, condensed into 9 salient groups, constituted the input space; the outcome of interest was renewable electricity penetration in 2050, operationalized as the percent of electricity production from renewable sources in 2050.

In line with earlier work on scenario discovery approaches (particularly the XLRM framework), policy levers can be included in the EPPA model simulations (Lempert, Popper, & Bankes, 2003). The policy used in this analysis is the Paris Agreement target – emissions constrained to achieve no more than 2 degrees Celsius of warming by the end of the century [double check this for accuracy]. Scenarios that were run with this policy lever are labeled “Policy”, while scenarios that were unconstrained (business as usual – is this correct or is something else called business as usual?) are labeled “Reference”.

Classification and regression trees (CART), a popular machine learning algorithm, was the primary technical tool used for the analysis. The CART algorithm divides the input space into regions of increasingly high purity in the output. For example, consider a simple classification problem of one input variable (call it X) and 10 samples; 5 samples belong to Group 1 and 5 samples belong to Group 2 (here the group classification is the response variable outcome of interest). Random sampling of the original data results in an equal chance of selecting a member of Group 1 or Group 2. Now suppose that, conditioned on X < 1, there are 4 members of Group 1 and 0 members of Group 2. This is a node of high purity; a member of Group 1 is selected with probability 1 conditioned on X < 1. Figure 1 visualizes the CART example given above.

In multivariable datasets, CART can be used to identify the most important input variables in the data set through the *feature importance score*, where 0 indicates no importance and 1 indicates maximal importance. The higher the feature importance score, the more effective conditioning on that variable is (or, in other words, the better that variable is at splitting the input space into regions of high purity). To add robustness to the feature importance score in this study, they were computed using bagging, where the data was bootstrapped, a classification tree was made for each bootstrapped set, and the scores from each such subtree were averaged.

**1**

**X**

Group 1

Group 2

CART splits here for maximum increase in purity

Figure 1. Diagram of a simple example suitable for analysis with CART.

*Results*

**Reference**

The CART algorithm identified the cost of wind as the most important variable in determining the global share of renewable electricity in 2050 when the outcome of interest was >80th percentile renewable penetration (see Table 1), under the Reference parameters. The importance score was robust, implicating wind as important at other nearby thresholds as well. This is reinforced visually by the parallel axis plot given in Figure 1. As Figure 2 shows, the relationship between renewable electricity penetration and the cost of wind is inverse, suggesting low costs of wind are associated with high levels of renewable penetration. Figure 2 also demonstrates that scenarios with >25% renewable electricity penetration are the main drivers of this association. Indeed, when the outcome of interest is <20th percentile renewable penetration, the cost of wind still leads, but by a much narrower margin (Table 2).

Because only a fraction of scenarios is responsible for the importance of the cost of wind, it is necessary to check that the cost of wind is not heavily correlated with another variable and therefore only indirectly associated with high renewable penetration. Table 3 shows no meaningful correlations of the cost of wind with any other variable for >25% renewable penetration in 2050, indicating that the cost of wind stands out as being singularly associated with high renewable penetration.

**Policy**

Results are similar under the scenarios with the Policy lever. As Table 4 shows, the cost of wind again leads in feature importance score. Figure 4 shows that the cost of wind continues to be inversely associated with renewable penetration under Policy.

*Discussion and Limitations*

There is a clear distinction between high and low renewable electricity scenarios in the EPPA model. Whereas the cost of wind is a highly disproportionately important variable in predicting high levels (>80th percentile renewable electricity share) of renewable penetration, there is no such clear winner when it comes to low levels (<20th percentile) of renewable penetration.

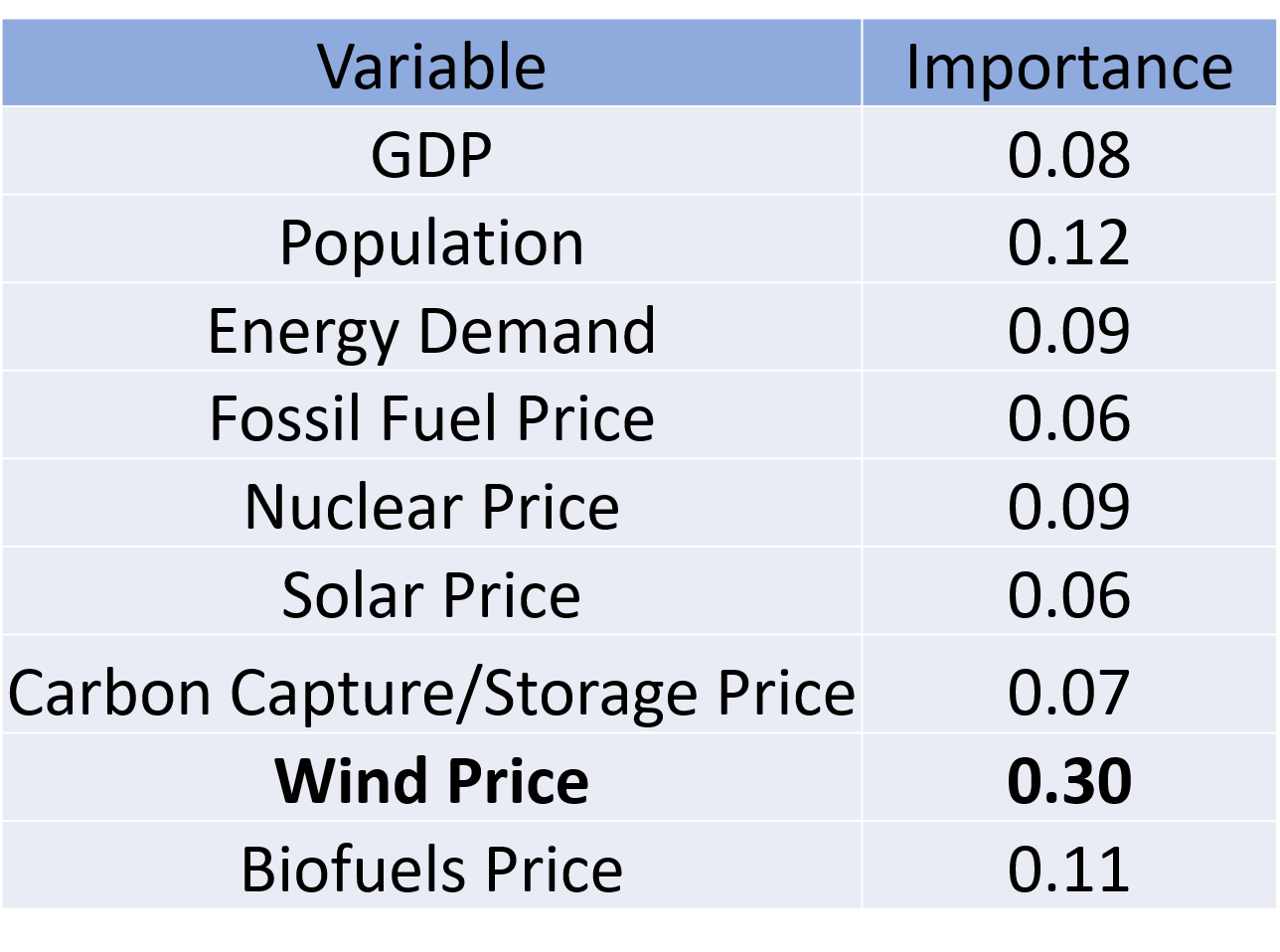


Table 1. Feature importance scores for global renewable electricity share in 2050, where the outcome of interest is >80th percentile renewable penetration and the policy lever is Reference.

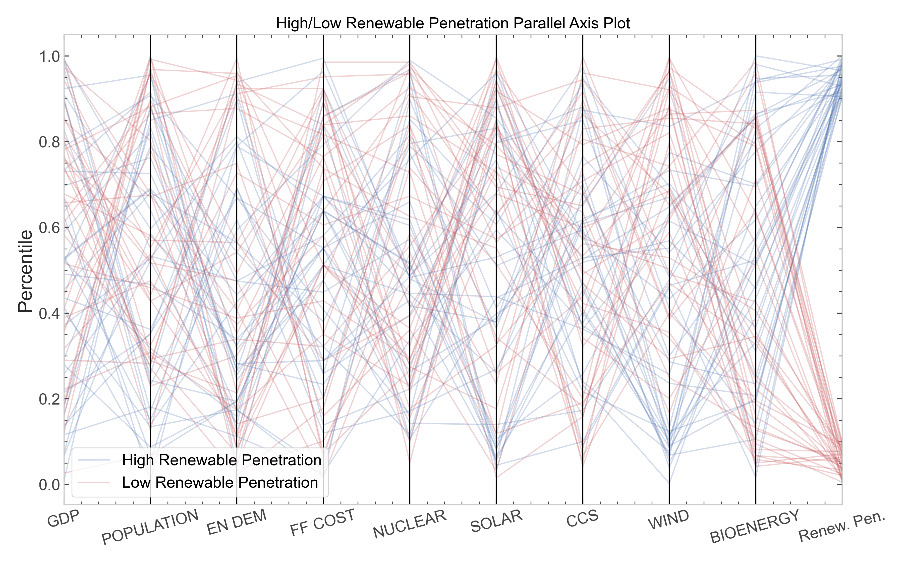


Figure 1. Parallel axis plot showing trajectories associated with high (>90th percentile) and low (<10th percentile) renewable penetration scenarios for global renewable penetration in 2050 under the Reference policy lever.

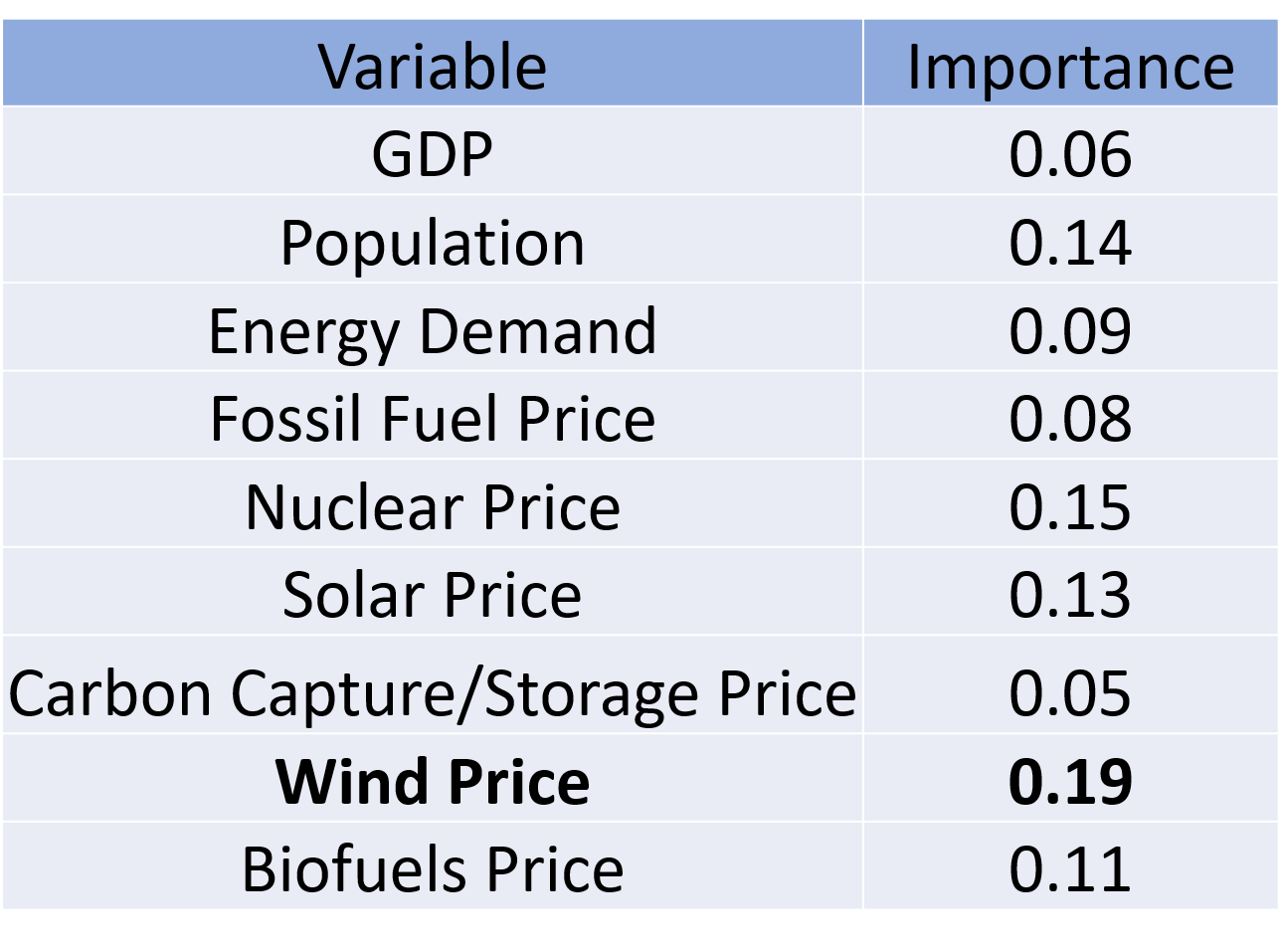


Table 2. Feature importance scores for global renewable electricity share in 2050, where the outcome of interest is <20th percentile renewable penetration.

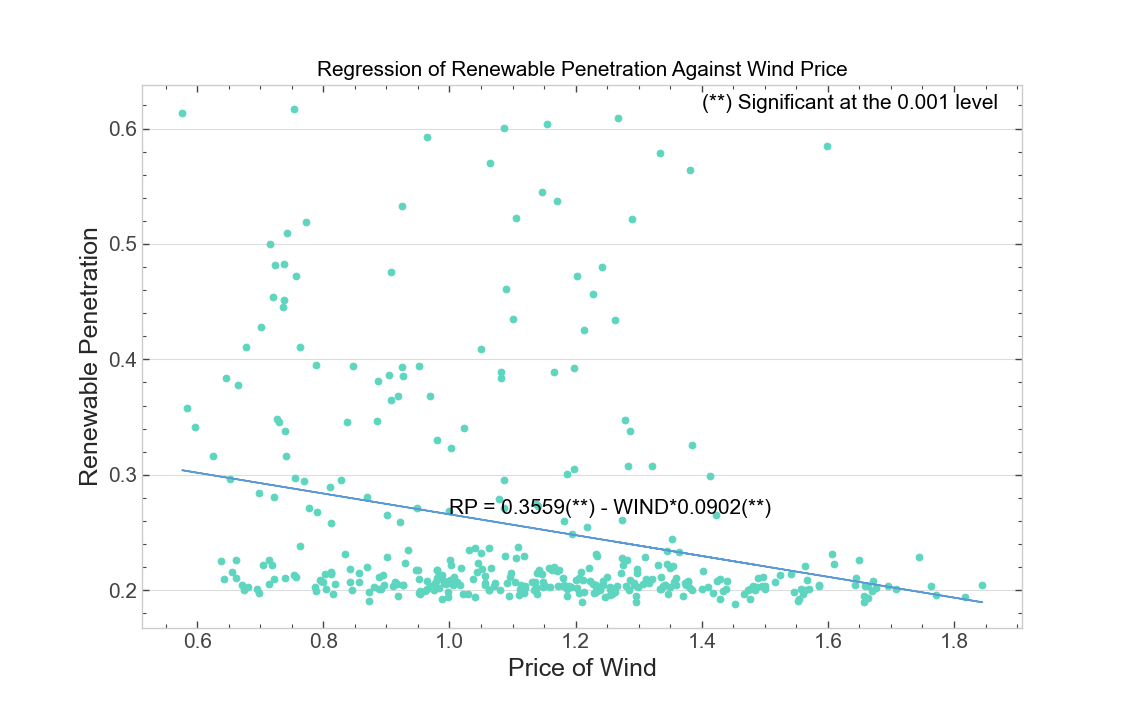


Figure 3. Scatterplot of cost of wind against renewable electricity penetration in 2050 with regression.

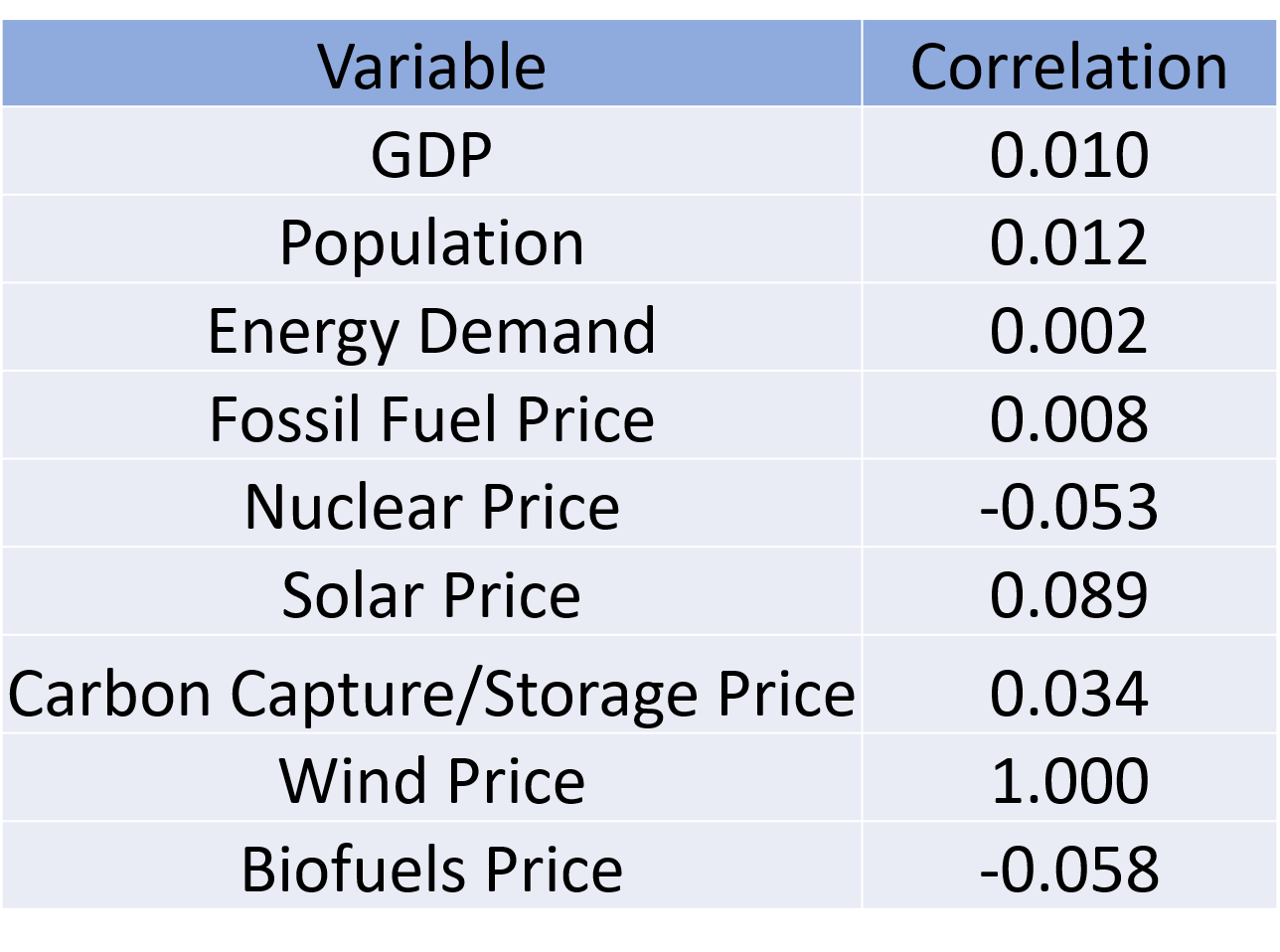


Table 3. Correlations of each variable with the cost of wind, 2050.

|  |  |
| --- | --- |
| Variable | Importance |
| GDP | 0.14 |
| Population | 0.11 |
| Energy Demand | 0.06 |
| Fossil Fuel Cost | 0.06 |
| Nuclear Cost | 0.08 |
| Solar Cost | 0.08 |
| Carbon Capture/Storage Cost | 0.12 |
| Wind Cost | 0.28 |
| Biofuels Cost | 0.08 |

Table 4. Feature importance scores under the Policy lever for global renewable penetration in 2050.

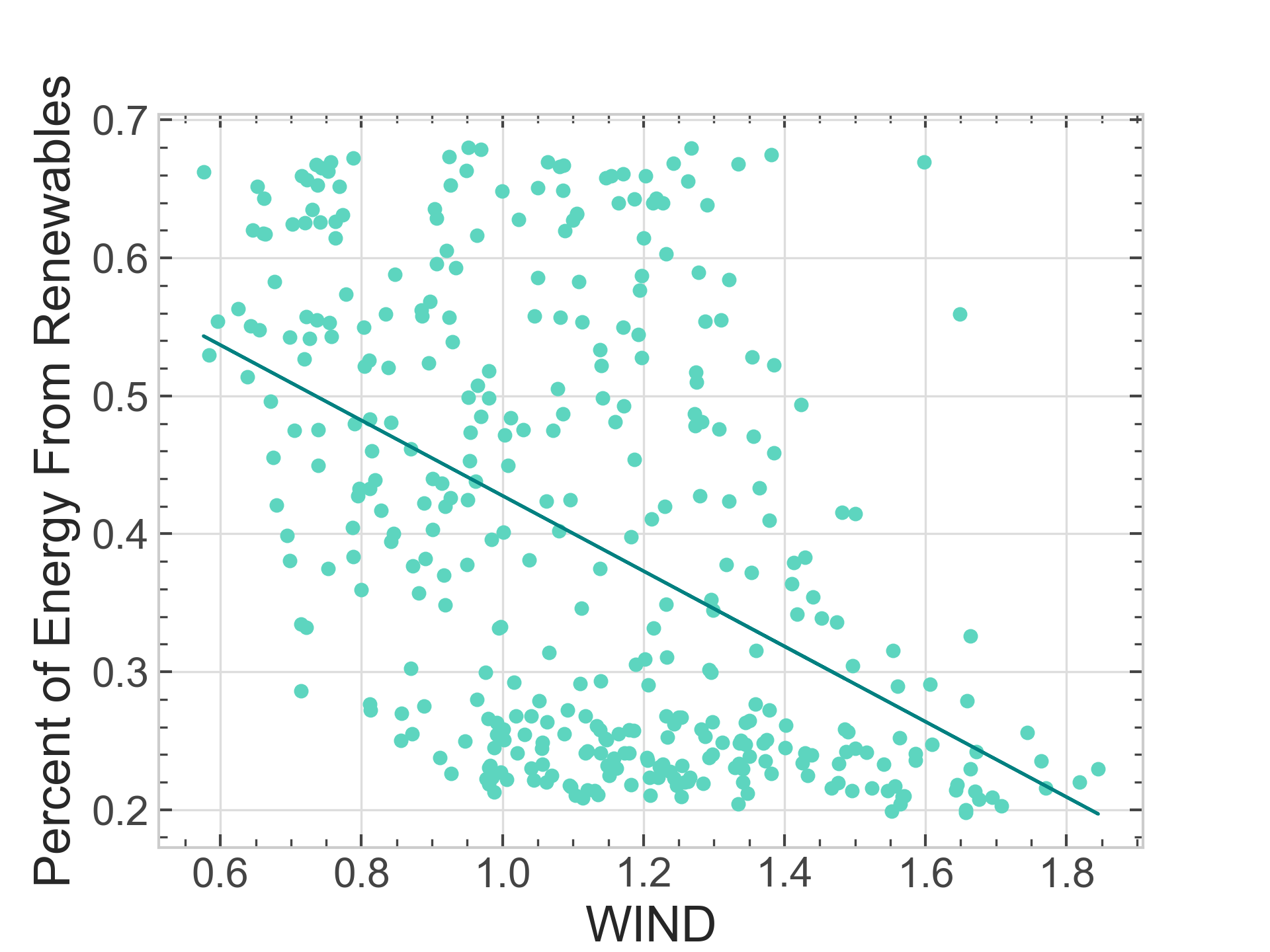


Figure 4. Regression of cost of wind against renewable penetration in 2050 under the Policy lever.

Appendix A: Importance Scores Over Time