Vertaeon Energy Performance Analysis

Montakan Thajchayapong, Wen Xu

ABSTRACT

Energy has been recognized as one of the most important determinants for economic growth and human development. Energy intensity is the ratio of energy use to output and most industries deal with several energy sources and outputs. This leads to the usual difficulties of understanding the relationship of those inputs and outputs. In this study, energy use is measured for three different products and the industry outputs are measured from each product's annual production. The data comes from two data sources (energy footprint and energy procurement) of three products (Performance, Agriculture and Specialty) from five plants during year 2000-2012. The goal of this study is to identify variables affects or identify how these variables can be used to predict energy intensity/consumption.

TABLE OF CONTENTS

INTRODUCTION

Data Source

In this project, we used two data sources provided by the company. The datasets are the energy footprint and the energy procurement of three products from five plants during year 2000-2012. Energy footprint data includes basic information of the plant, greenhouse gas production from 3 different fossil fuels, total energy used for production, production amounts for 3 kinds of products and sale price for these products. Individual plants have several types of energy source both from purchases and internal products. Energy procurement data contain the information of energy purchase amount and cost for steam, electricity, and 3 kinds of fossil fuels' as well as the amount of internal generated energies. Considering the abundance of data we have, we believe our analysis can help the company have a better evaluation for all impact factors and make the best use of energy performance.

Insight for Energy Footprint and Procurement Analysis

Energy is one of the most important indicators for economic growth and human development [1]. Governments pay close attention to the energy demand of their counties. Particularly, attention is paid to energy intensity, which is the ratio of energy use divided by output. In last 20 years, there has been significant reduction in energy intensity in the world's developed countries. OECD's energy intensity declines 26%; the Group of Seven's (G-7) declined 29% while in the U.S. there was a drop of 34% in its energy intensity [2]. Several factors contribute to how the energy intensity being calculated including several energy resources and several products. There are two common approaches used in measuring of aggregate energy use. The most popular approach is to add together the energy sources. Another approach is to rely on prices as aggregation factor.

Several techniques and methodology were used in previous studies. Studies dealing on energy demand in production sectors can be divided in two broad categories. Firstly, they focused on the demand for various types of energy, which provided information about some potential substitution between energy input (i.e., electricity and coal) [3,4]. The second category focused on substitution possibilities between energy and other factors such as labor, capital and materials [5,6]. In 1993, a study used cross-section data of 10,000 companies in Australia from 1977-85 to observe energy pattern used in its four production plants [7]. In an attempt to find out the demand for energy in Swedish manufacturing industries; Dargay et al. [8] employed a translog cost function for 12 manufacturing sub-sectors from 1952-1976. The major variables used in the study include energy consumption, capital, labor and intermediate goods. The results indicate that relative changes in energy prices have significant effects on energy consumption.

In this project, we obtained data from a company called Vertaeon -- a startup consulting firm in Atlanta that provides advisory services and implementation support for business growth strategy risk and sustainability including development of company-specific frameworks and integration to organizational goals. Our goal is to identify variables affects or identify how these variable can be used to predict energy intensity/consumption. Our analysis can help the company to have a better evaluation for all impact factors and to make the better use of energy performance.

METHODS

Description of Data

We used two data sources provided by the company. The datasets are the energy footprint and the energy procurement of three products (Performance, Agriculture, Specialty) from five plants (Plant 1-5) during year 2000-2012. For every plant, data includes basic information of the plant, overall energy intensity of all three products, energy use of each product, production amounts and unit price of each product, energy purchases and procurement cost of five external sources (Electricity, Steam, Coal, Natural gas, Oil). Energy intensity can be calculated from the variables provided in the data, but with different proportion. Equation (1) and (2) shown simple calculation of each energy intensity.

$$Total\ Energy\ Use = Sum\ procurement - \ Energy\ sold \quad (1)$$

$$Energy\ Intensity = \frac{Total\ Energy\ Use}{Production} \quad (2)$$

Our goal is to explain overall energy intensity by using as less variables as possible and also to provide clear and meaningful suggestions to decrease energy intensity. Energy intensity constitutes from individual three products are not of our interest, so we only chose overall energy intensity to be our response variable. We removed predictors that are direct products or sums of many other variables. For example, total annual production is the sum of annual production for performance product, agriculture product and specialty product. Finally, we decided to use 22 variables (Table 1) and overall energy intensity for our analysis.

Table 1. Data selected through overall intepretation

Data	Abbreviation
Overall Energy Intensity	eI
Plant	Plant
Year	Yr
Green House Gas Emissions	GHG
Production - Performance Product	productP
Production - Agriculture Product	productA
Production - Specialty Product	productS
Unit Sale Prices - Performance Product	saleP
Unit Sale Prices - Agriculture Product	saleA
Unit Sale Prices - Specialty Product	saleS
Energy Use - Performance Product	eUseP
Energy Use - Agriculture Product	eUseA
Energy Use - Specialty Product	eUseS
Procurement Cost - Coal	costC
Procurement Cost - Natural Gas	costG
Procurement Cost - Oil	costO
Procurement Cost - Electricity	costE
Procurement Cost - Steam	costS
Energy Purchase – Coal	ePC
Energy Purchase – Natural Gas	ePG
Energy Purchase – Oil	ePO
Energy Purchase – Electricity	ePE
Energy Purchase – Steam	ePS

Exploratory Data Analysis

Selection of Predictors

Our data consists of energy procurement and footprint from five plants over 13 years. These two variables have obvious relationship with overall energy intensity. Differences in energy intensity performance of each plant are shown in Figure 1. The trend of energy intensity from 2000 to 2012 (Figure 2) exhibited strong reduction except for plant 5.

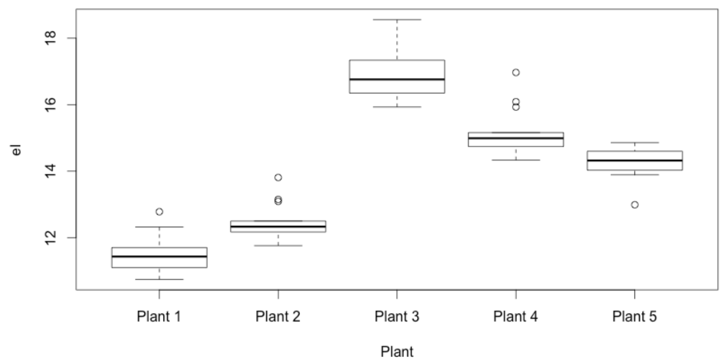

Figure 1. Energy intensity performance of each plant

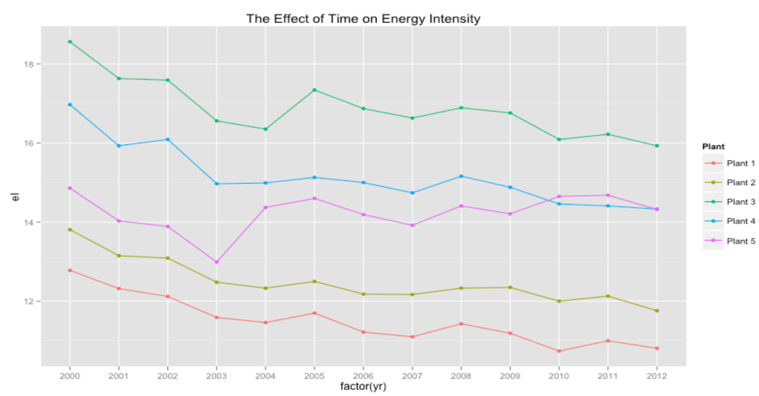

Figure 2. Energy intensity performance of each year for each plant

For all other 20 variables, scatterplots against overall energy intensity (eI) showed different trends (Figure S1). From which, we found only 5 variables (saleP, saleA, sale S, costO, ePO) shown obvious positive or negative correlation with overall energy intensity so we took these variables in consideration.
Because of missing data points, some absent data such as the internal electricity generation energy affected our observation so we factorized these variables to better understand how they contributed to overall energy intensity (Figure S2). For

example, internal electricity generation data was missing for plant 1, 4 and 5. After factorization (Figure 3), the data for all other plants spread out and become easier to characterize. We found that none of factorized variables showed any relationship with energy intensity, so we decided not to use these variables in our model selection.

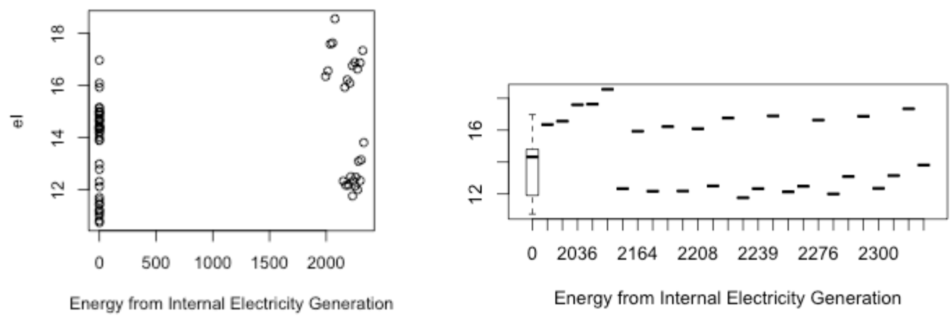

Figure 3. Energy intensity performance for different internal electricity generation energy (Before[left] and after[right] factorization)

Interaction Exploration

From the one-on-one scatterplots (Figure S1), we observed that some variables could be divided into five clusters, in which coincided with the plant number. Thus, we color-coded each plant and found obvious trend for each plant being subset of different variables (Figure S3). For example, energy use of agriculture product (eUseA) showed no overall trend against energy intensity. From the color-coded plot, we observed apparent increase of energy intensity for increase energy use of agriculture product (Figure 4). Thus we concluded that the interaction of eUseA and plant could affect energy intensity.

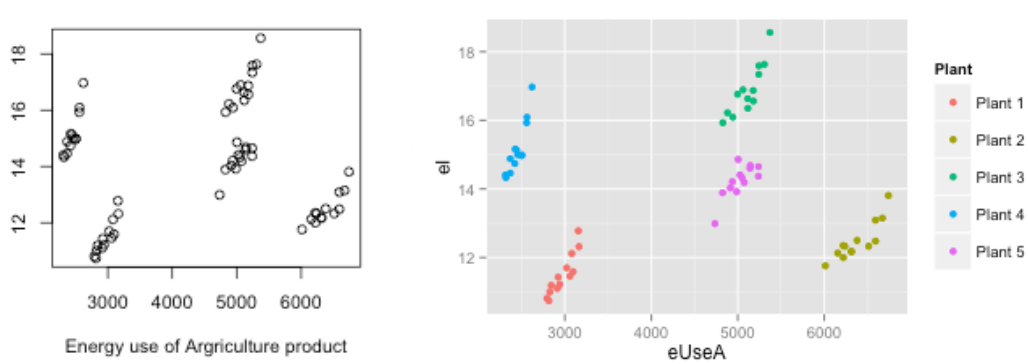

Figure 4. Energy intensity performance for different energy use of agriculture product (Before[left] and after[right] color code by plant type)

We found GHG, costC, costE, costG, costS, ePE, ePC, ePS, ePG, eUseP, eUseS and eUseA with plant interaction term have positive effect on energy intensity. On the other hand, productP, productA and productS with plant interaction term have negative effect on overall energy intensity.

Finally, we concluded our exploratory analysis of the data with examining correlations (Figure 5). The correlation results suggested a high correlated between variables Energy Purchase – Coal and Energy Use - Performance Product.

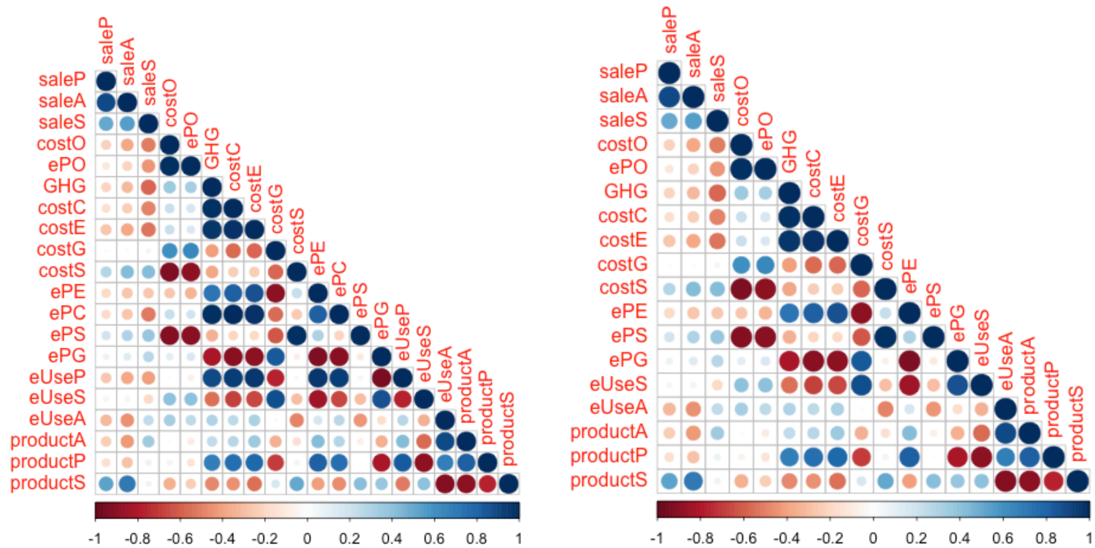

Figure 5. Correlation table of all variables
(Before[left] and after[right] removing ePC and eUseP)

From all observation and preliminary exploration, we decided to use 19 variables (Table 2 and 3) including 6 independent variables and 13 plant interaction terms for model selection.

Table 2. Data selected to have direct correlation with energy intensity

Data	Abbreviation
Green House Gas Emissions	GHG
Unit Sale Prices - Performance Product	saleP
Unit Sale Prices - Agriculture Product	saleA
Unit Sale Prices - Specialty Product	saleS
Procurement Cost - Oil	costO
Energy Purchase – Oil	ePO

Table 3. Data selected to have interaction with Plant variable

Data interact with Plant	Abbreviation
Green House Gas Emissions	GHG
Production - Performance Product	productP
Production - Agriculture Product	productA
Production - Specialty Product	productS
Energy Use - Agriculture Product	eUseA
Energy Use - Specialty Product	eUseS
Procurement Cost - Coal	costC
Procurement Cost - Natural Gas	costG
Procurement Cost - Electricity	costE
Procurement Cost - Steam	costS
Energy Purchase – Natural Gas	ePG
Energy Purchase – Electricity	ePE
Energy Purchase – Steam	ePS

Model Selection

Simple Regression

FULL MODEL

We started out with including all variables after correlation correction along with their interactions in our regression shown in Table 2 and 3. The summary of the regression result of the full model is shown in Table S1 (Appendix).

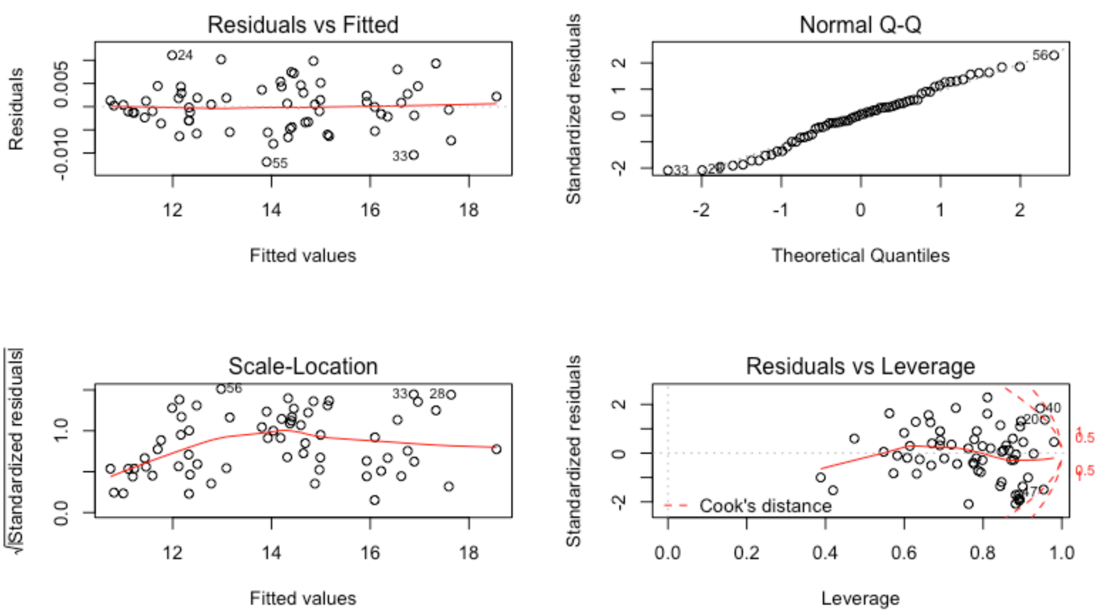

Figure 6. Standard diagnostic plots of the full model

Due to the nature of the collected data, there were some missing values in which resulted in "NA" in the regression results. Figure 6 shows the standard diagnostic plots of the full model. We noticed some unusual fit of data; therefore we chose to eliminate variables that contained missing values and not showing significant level ($P < 0.05$)

REDUCED MODEL

The reduced regression equation takes the following functional form:

$$eI \sim saleA + saleS + costE + ePE + eUseS + (productP + productA + costE) * Plant$$

The summary of the regression result of the reduced model is shown in Table S2. The diagnostic plots are shown in Figure 7.

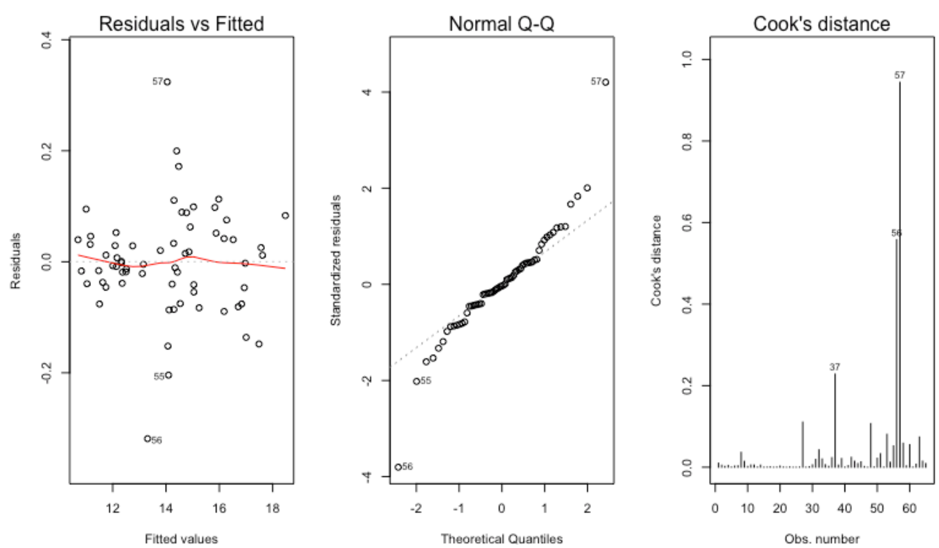

Figure 7. Standard diagnostic plots of the reduced model

We still noticed some unusual fit of data, therefore we decided to test the model using "backward stepwise" starting from the variables from the full model.

Stepwise BIC

BACKWARD STEPWISE MODEL

Using the variables from the full model, the summary of the regression result of the backward stepwise model is shown in Table S3. The diagnostic plots shown in Figure 8.

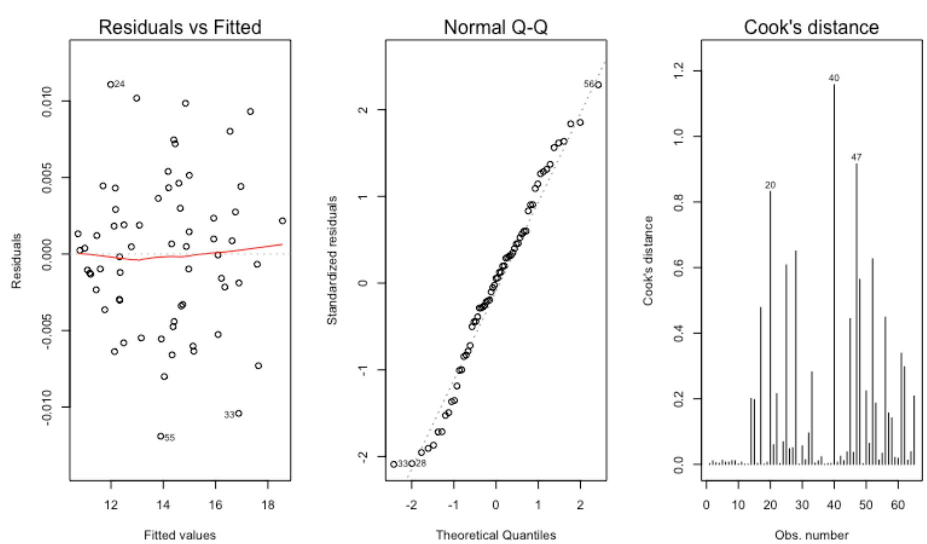

Figure 8. Standard diagnostic plots of the backward stepwise model

We found the there was an improvement on the fit of the data. However, using this backward stepwise for the variable selection resulted in quite big model as most of the variables were chosen to present in the outcome. With this reason along with a very high r-squared value of 0.9968, we suggested further select for better models using other approaches such as forward stepwise and LASSO.

FORWARD STEPWISE MODEL

Using the variables from the reduced model, the summary of the regression result of the forward stepwise model is shown in Table S4. The diagnostic plots shown in Figure 9.

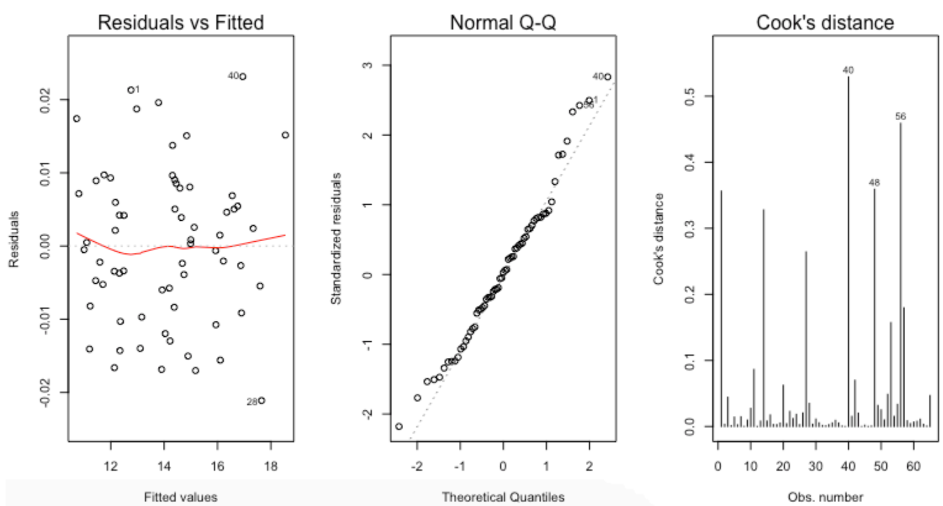

Figure 9. Standard diagnostic plots of the forward stepwise model

The final equation from backward stepwise selection is as follows:

$$eI \sim saleA + saleS + costE + ePE + eUseS + productA + productP + Plant + costE{:}Plant + productA{:}Plant + productP{:}Plant + eUseS{:}Plant$$

LASSO

The backward stepwise selection model seems to be good and clean enough to be used. However, we still wondered how each variable enters the model and wanted to see if a smaller model could be obtained. So we performed LASSO from the model we obtained from backward stepwise selection. How each variable entered the model were checked using their coefficients (Figure10 left panel).

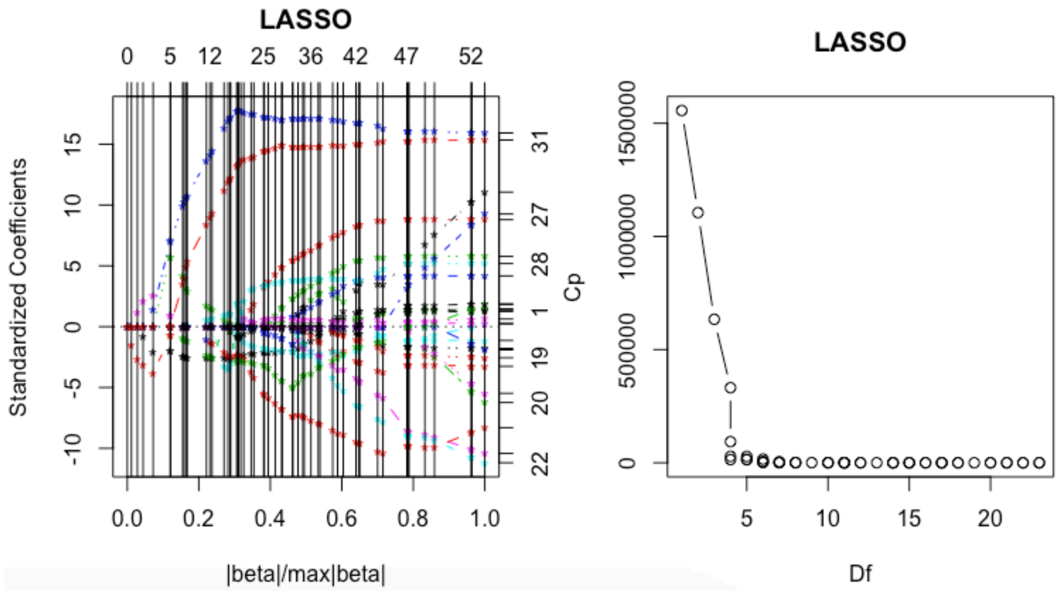

Figure 10. LASSO model

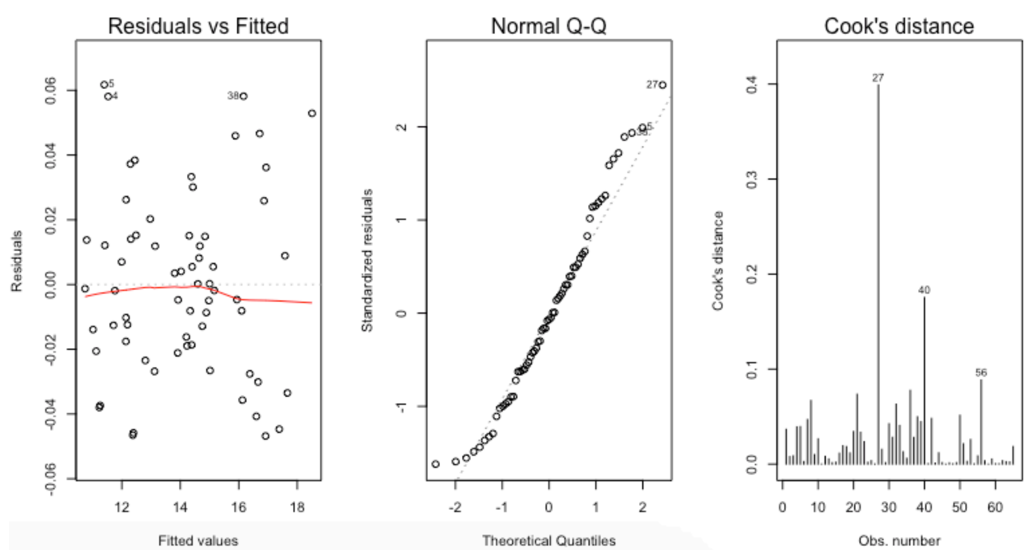

Figure 11. Standard diagnostic plots of the LASSO model

The model with the least Mallow's Cp is:

$$eI \sim saleA + saleS + costE + eUseS + productA + productP + costE{:}Plant + productA{:}Plant + productP{:}Plant + eUseS{:}Plant$$

The coefficients (Table 4 and S5) and diagnostic plots (Figure 11) showed the better fit with selected data, so further validation of the goodness of fit was performed.

Table 4. LASSO model regression summary

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.46E+01	3.71E-01	39.277	< 2e-16	***
saleA	1.20E-01	1.78E-01	0.675	0.503314	
saleS	-2.92E-01	1.21E-01	-2.422	0.019835	*
costE	2.86E-07	8.85E-08	3.225	0.002439	**
eUseS	1.82E-03	3.09E-04	5.896	5.61E-07	***
productA	5.87E-02	3.61E-02	1.624	0.111811	
productP	-2.22E-02	3.58E-02	-0.622	0.537444	
costE:PlantPlant 2	-1.75E-07	1.82E-07	-0.963	0.341014	
costE:PlantPlant 3	-2.90E-07	1.25E-07	-2.314	0.025643	*
costE:PlantPlant 4	-4.41E-07	1.71E-07	-2.582	0.013388	*
costE:PlantPlant 5	-2.81E-07	9.08E-08	-3.093	0.003511	**
productA:PlantPlant 2	-5.48E-02	4.59E-02	-1.195	0.238872	
productA:PlantPlant 3	-1.87E-02	4.00E-02	-0.469	0.641446	
productA:PlantPlant 4	-3.46E-02	3.76E-02	-0.92	0.362853	
productA:PlantPlant 5	-6.25E-02	4.70E-02	-1.33	0.190683	
productP:PlantPlant 2	4.66E-02	5.54E-02	0.842	0.40477	
productP:PlantPlant 3	-6.18E-03	4.61E-02	-0.134	0.893939	
productP:PlantPlant 4	1.92E-02	4.33E-02	0.442	0.660694	
productP:PlantPlant 5	4.02E-02	4.37E-02	0.918	0.363697	
eUseS:PlantPlant 2	1.30E-03	5.48E-04	2.363	0.022847	*
eUseS:PlantPlant 3	1.27E-03	3.57E-04	3.558	0.000942	***
eUseS:PlantPlant 4	1.28E-03	3.87E-04	3.294	0.002013	**
eUseS:PlantPlant 5	6.63E-03	3.53E-04	18.76	< 2e-16	***

Model Validation

Cross Validation

We performed leave-one-out cross-validation and 6-fold cross-validation to assess the prediction risk from LASSO model.

```
> cv.err$delta
[1] 0.001932093 0.001920391
> cv.err.6$delta[2]
[1] 0.002772158
```

From the data above, delta value reported the computed error from the model. It is clear that the error score is relatively low. So we can conclude that the LASSO model is appropriate for explaining changes in overall energy intensity.

DISCUSSION

Based on our initial exploratory of data, we found that Plant 3 has the highest energy intensity across all products while Plant 1 has the least energy intensity across all products (Figure 1). As mentioned in the introduction that the ideal energy practice is to manage the plant energy intensity level at its lowest possible while having the highest possible production output. Interestingly from our initial results, Agriculture product best performed in Plant 2 but using the most amount of energy, Performance product best performed in Plant 5 but again using the most amount of energy and Specialty product best performed in Plant 1 with an average amount of energy use (Figure S2.1).

In terms of external resources used from each plant (Figure S2.2), we would expect to have both procurement and purchase cost maintained at the lowest while using the least energy intensity. However, results from data exploratory were far from ideal situation. While some plants could maintain the low cost but at the same time the energy intensity could not be maintained at the lowest and vice versa.

In conclusion from our initial exploratory investigation, we found that an increase in energy efficiency might take place either when energy inputs are reduced for any given consumption level, or there are increased or enhanced services for a given amount of energy inputs. There are also other various reasons that could results in the inconclusive predictions of how energy intensity being explained by some variables. Therefore, we proposed an alternative approach to the client that certain variables could be included in making predictions.

Our model selection ended with the LASSO model, but this was by no means the best one. Indeed, we propose to do a separate regression with only Plant 5 since it showed some difference characters compared to other plants. We suggested that it should have its own coefficients for the variables we proposed to show any significance.

As a guideline for our clients, we conclude that by lowering the sales of agriculture products, the cost of electricity, the energy use of steam and the production for agriculture products, all plants should have lower overall energy intensity. Additionally, increase the sales for specialty product and the production for performance products should have similar effect.

To achieve lower overall energy intensity, each plant should have its unique strategy and this can be obtained from interaction terms in Table 4.

REFERENCES

[1] Dhungel, K.R. (2008), A causal relationship between energy consumption and economic growth in Nepal, Asia-Pacific Development Journal, Vol. 15 (1), pp. 137-150

[2] International energy agency (2007), Key World energy statistics, http://www.iea.org/textbase/nppdf/free/2007/key_stats_2007.pdf

[3] Halvorsen, R., (1977), Energy substitution in US manufacturing, Review of Economics and Statistics, Vol. 59 (4), pp. 381-388

[4] Pindyck, R.S. (1979), Interfuel substitution and the industrial demand for energy: an international comparison. Review of Economics and Statistics, Vol. 61 (2), pp. 169-179

[5] Berndt, E.R., and Wood, D.O. (1975), Technology, prices, and the derived demand for energy, Review of Economics and Statistics, Vol. 57 (3), pp. 259-268.

[6] Griffin, J.M., and Gregory, P.R. (1977), An inter-country translog model of energy substitution responses, American Economic Review, Vol. 66, pp. 845-858

[7] Woodland, A.D., (1993), A Micro-econometric Analysis of the Industrial Demand for Energy in NSW, Energy Journal, Vol. 14(2), pp. 57-89.

[8] Dargay, J.M. (1983), The demand for energy in Swedish manufacturing industries, The Scandinavian Journal of Economics, Vol. 85, (1), pp. 37-51

APPENDIX

Supplementary figures

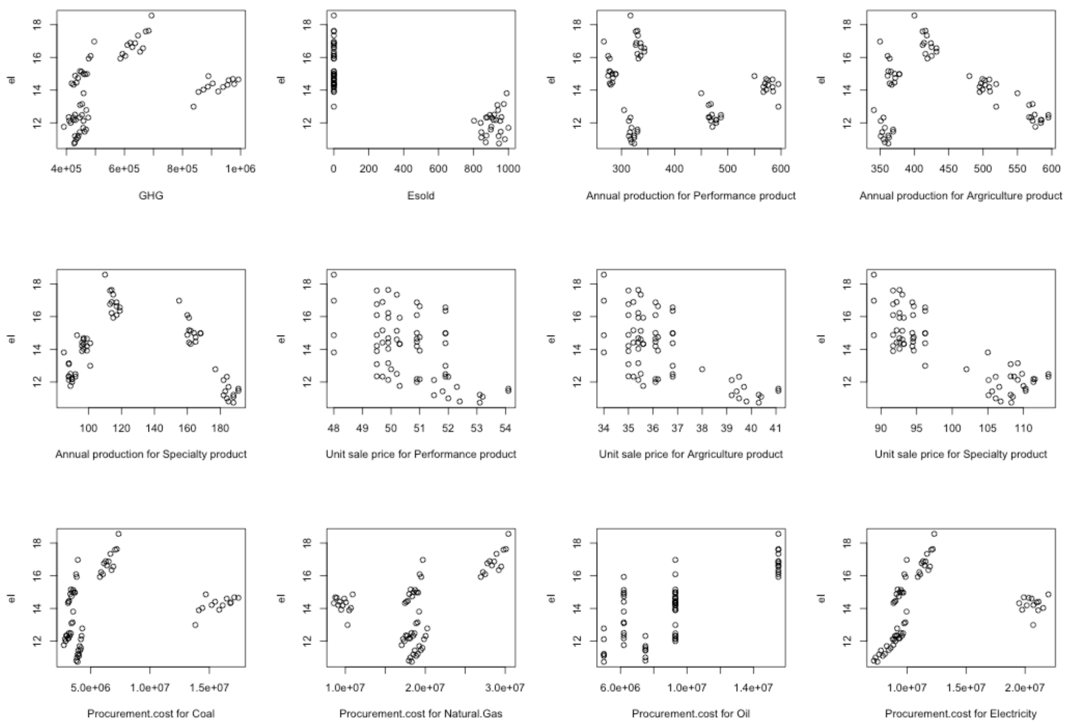

Figure S1.1. Scatterplots for different variables against overall energy intensity (eI)

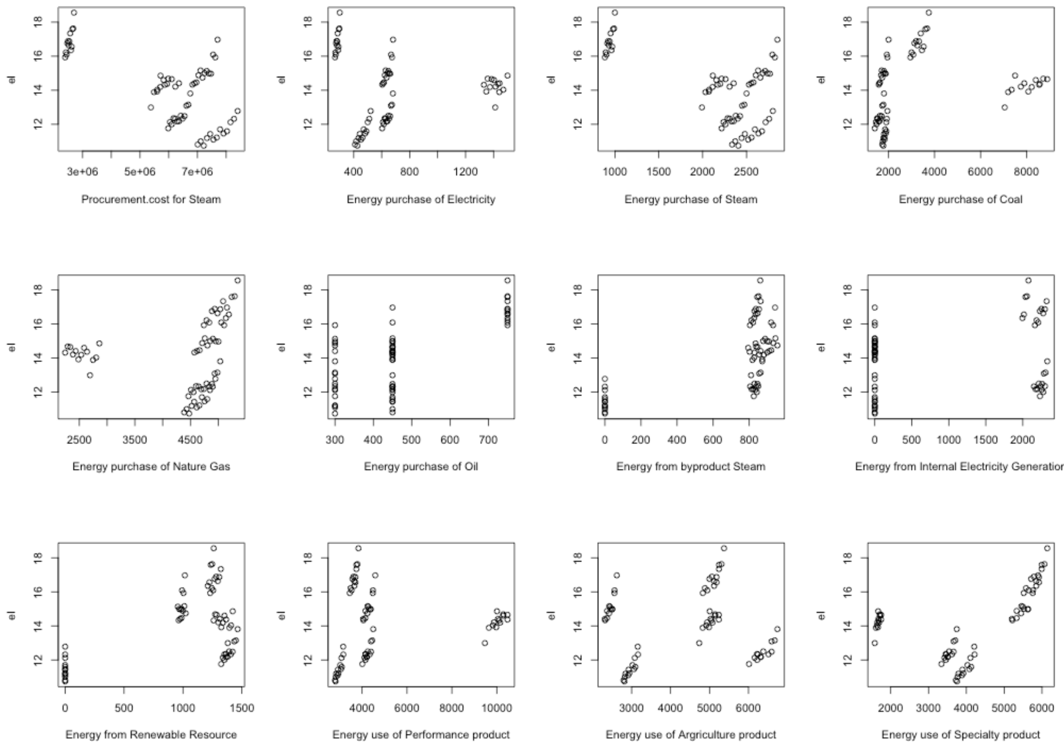

Figure S1.2. Scatterplots for different variables against overall energy intensity (eI)

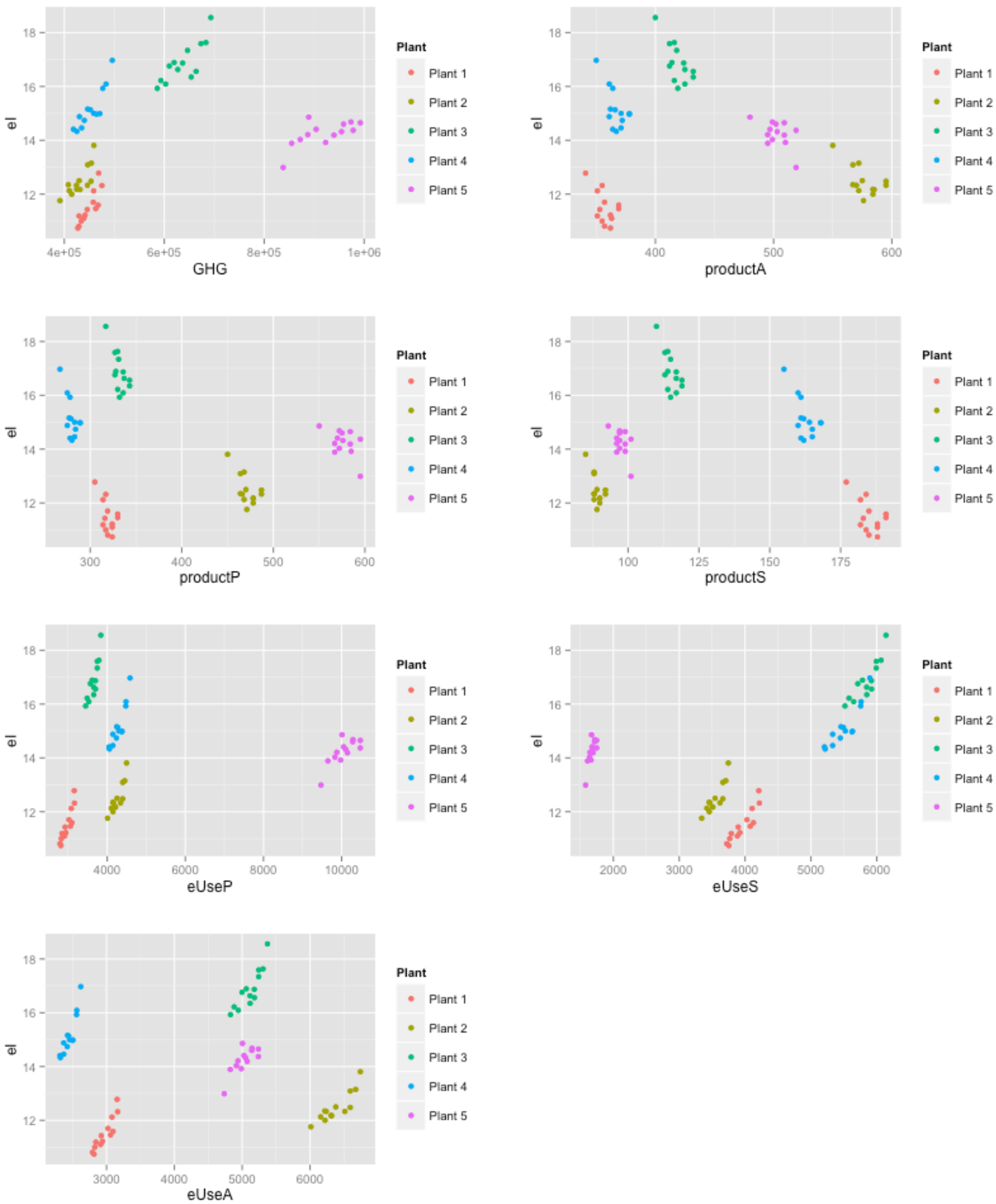

Figure S2.1. scatterplots against energy intensity color coded by plant

Figure S2.2. scatterplots against energy intensity color coded by plant

Supplementary tables

Table S1. Full model regression summary

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.60E+01	5.20E+00	3.082	0.00759	**
saleP	6.44E-02	1.06E-01	0.607	0.55262	
saleA	3.84E-01	1.07E-01	3.59	0.00268	**
saleS	-2.70E-01	7.74E-02	-3.484	0.00333	**
costO	-1.23E-06	3.01E-06	-0.409	0.68836	
ePO	2.87E-02	5.37E-02	0.535	0.60072	
GHG	-1.08E-04	1.62E-04	-0.669	0.51396	
costC	1.28E-05	2.83E-05	0.45	0.65909	
costE	9.59E-06	4.51E-06	2.126	0.05056	.
costG	1.33E-06	2.00E-06	0.666	0.5154	
costS	-2.47E-05	2.43E-05	-1.013	0.32707	
ePE	-1.40E-01	6.61E-02	-2.112	0.05184	.
ePC	-1.64E-02	5.73E-02	-0.287	0.77822	
ePS	6.70E-02	6.57E-02	1.019	0.32454	
ePG	NA	NA	NA	NA	
eUseP	7.24E-05	6.67E-04	0.109	0.915	
eUseS	2.45E-03	6.33E-04	3.876	0.00149	**
eUseA	1.46E-04	6.18E-04	0.237	0.81596	
productA	2.57E-02	2.37E-02	1.081	0.29686	
productP	-2.04E-02	2.06E-02	-0.991	0.33723	
productS	-2.33E-02	4.62E-02	-0.504	0.62133	
PlantPlant 2	-4.62E+00	5.46E+00	-0.846	0.41088	
PlantPlant 3	2.79E+00	8.87E+00	0.314	0.75772	
PlantPlant 4	1.55E-01	5.49E+00	0.028	0.97793	
PlantPlant 5	1.65E+00	6.50E+00	0.255	0.80251	
GHG:PlantPlant 2	6.42E-05	1.56E-04	0.413	0.68572	
GHG:PlantPlant 3	7.18E-06	7.54E-05	0.095	0.9254	
GHG:PlantPlant 4	6.80E-05	1.56E-04	0.435	0.66971	
GHG:PlantPlant 5	2.18E-05	7.47E-05	0.292	0.77424	
costC:PlantPlant 2	-2.88E-06	6.26E-06	-0.461	0.6517	
costC:PlantPlant 3	NA	NA	NA	NA	
costC:PlantPlant 4	-2.47E-06	6.16E-06	-0.4	0.69477	
costC:PlantPlant 5	NA	NA	NA	NA	
costE:PlantPlant 2	3.32E-06	1.28E-06	2.599	0.02013	
costE:PlantPlant 3	-6.48E-06	2.95E-06	-2.197	0.04417	

	Estimate	Std. Error	t value	Pr(>|t|)
costE:PlantPlant 4	2.78E-06	1.10E-06	2.529	0.02314
costE:PlantPlant 5	NA	NA	NA	NA
costG:PlantPlant 2	-1.74E-06	2.26E-06	-0.767	0.45518
costG:PlantPlant 3	NA	NA	NA	NA
costG:PlantPlant 4	NA	NA	NA	NA
costG:PlantPlant 5	NA	NA	NA	NA
costS:PlantPlant 2	-1.60E-06	3.86E-06	-0.414	0.68459
costS:PlantPlant 3	-3.76E-07	4.30E-06	-0.087	0.93154
costS:PlantPlant 4	-5.65E-06	4.94E-06	-1.143	0.27094
costS:PlantPlant 5	NA	NA	NA	NA
ePE:PlantPlant 2	NA	NA	NA	NA
ePE:PlantPlant 3	NA	NA	NA	NA
ePE:PlantPlant 4	NA	NA	NA	NA
ePE:PlantPlant 5	NA	NA	NA	NA
ePC:PlantPlant 2	NA	NA	NA	NA
ePC:PlantPlant 3	NA	NA	NA	NA
ePC:PlantPlant 4	NA	NA	NA	NA
ePC:PlantPlant 5	NA	NA	NA	NA
ePS:PlantPlant 2	NA	NA	NA	NA
ePS:PlantPlant 3	NA	NA	NA	NA
ePS:PlantPlant 4	NA	NA	NA	NA
ePS:PlantPlant 5	NA	NA	NA	NA
ePG:PlantPlant 2	NA	NA	NA	NA
ePG:PlantPlant 3	NA	NA	NA	NA
ePG:PlantPlant 4	NA	NA	NA	NA
ePG:PlantPlant 5	NA	NA	NA	NA
eUseP:PlantPlant 2	5.53E-04	6.29E-04	0.878	0.39369
eUseP:PlantPlant 3	NA	NA	NA	NA
eUseP:PlantPlant 4	NA	NA	NA	NA
eUseP:PlantPlant 5	NA	NA	NA	NA
eUseS:PlantPlant 2	NA	NA	NA	NA
eUseS:PlantPlant 3	NA	NA	NA	NA
eUseS:PlantPlant 4	NA	NA	NA	NA
eUseS:PlantPlant 5	NA	NA	NA	NA
eUseA:PlantPlant 2	NA	NA	NA	NA
eUseA:PlantPlant 3	NA	NA	NA	NA
eUseA:PlantPlant 4	NA	NA	NA	NA
eUseA:PlantPlant 5	NA	NA	NA	NA
productA:PlantPlant 2	-3.35E-02	2.59E-02	-1.295	0.21488

	Estimate	Std. Error	t value	Pr(>|t|)
productA:PlantPlant 3	-1.79E-02	2.93E-02	-0.61	0.5513
productA:PlantPlant 4	-5.06E-02	2.15E-02	-2.358	0.03236
productA:PlantPlant 5	-7.38E-02	3.64E-02	-2.031	0.06033
productP:PlantPlant 2	3.17E-02	3.25E-02	0.974	0.34567
productP:PlantPlant 3	-2.64E-02	2.45E-02	-1.076	0.29873
productP:PlantPlant 4	1.64E-02	2.08E-02	0.79	0.44196
productP:PlantPlant 5	5.66E-02	3.09E-02	1.83	0.08717
productS:PlantPlant 2	5.85E-03	4.96E-02	0.118	0.90766
productS:PlantPlant 3	4.31E-02	5.37E-02	0.801	0.43559
productS:PlantPlant 4	3.56E-02	4.87E-02	0.73	0.47634
productS:PlantPlant 5	-1.08E-02	4.95E-02	-0.218	0.83028

Table S2. Reduced model regression summary

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.72E+00	1.85E+00	5.266	4.46E-06 ***
saleA	-2.83E-01	5.82E-01	-0.486	0.629253
saleS	1.15E-01	3.86E-01	0.299	0.766319
ePE	3.15E-02	1.24E-02	2.548	0.014595 *
eUseS	4.53E-03	4.01E-04	11.307	2.53E-14 ***
productP	4.54E-02	1.17E-01	0.389	0.699099
productA	-7.90E-02	1.13E-01	-0.697	0.489718
costE	-2.23E-06	8.19E-07	-2.72	0.009458 **
PlantPlant 2	2.10E+00	2.79E+00	0.753	0.455715
PlantPlant 3	4.97E+00	2.94E+00	1.692	0.098145
PlantPlant 4	3.65E+00	2.94E+00	1.242	0.221165
PlantPlant 5	1.37E+01	3.29E+00	4.177	0.000146 ***
productP:PlantPlant 2	-8.82E-02	1.75E-01	-0.504	0.616996
productP:PlantPlant 3	5.92E-02	1.42E-01	0.417	0.678785
productP:PlantPlant 4	-9.91E-02	1.44E-01	-0.687	0.495616
productP:PlantPlant 5	8.71E-02	1.48E-01	0.59	0.558617
productA:PlantPlant 2	8.77E-02	1.44E-01	0.61	0.545265
productA:PlantPlant 3	-4.51E-02	1.22E-01	-0.371	0.712576
productA:PlantPlant 4	7.98E-02	1.24E-01	0.645	0.522261
productA:PlantPlant 5	-1.05E-01	1.61E-01	-0.65	0.519107
costE:PlantPlant 2	-2.26E-07	1.76E-07	-1.283	0.206526
costE:PlantPlant 3	7.91E-07	3.90E-07	2.029	0.048796 *
costE:PlantPlant 4	-8.95E-07	2.57E-07	-3.481	0.00118 **
costE:PlantPlant 5	NA	NA	NA	NA

Table S3. Backward stepwise model regression summary

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.60E+01	5.20E+00	3.082	0.00759	**
saleP	6.44E-02	1.06E-01	0.607	0.55262	
saleA	3.84E-01	1.07E-01	3.59	0.00268	**
saleS	-2.70E-01	7.74E-02	-3.484	0.00333	**
costO	-8.78E-07	2.32E-06	-0.378	0.71082	
GHG	1.85E-04	5.03E-04	0.368	0.71821	
costC	-7.31E-06	2.14E-05	-0.341	0.73746	
costE	1.63E-06	7.58E-07	2.152	0.04808	*
costG	-2.54E-06	7.09E-06	-0.359	0.72493	
costS	-2.43E-06	2.49E-06	-0.976	0.34474	
eUseS	3.33E-03	3.68E-03	0.904	0.38028	
productA	2.57E-02	2.37E-02	1.081	0.29686	
productP	-2.04E-02	2.06E-02	-0.991	0.33723	
productS	-2.33E-02	4.62E-02	-0.504	0.62133	
PlantPlant 2	-4.62E+00	5.46E+00	-0.846	0.41088	
PlantPlant 3	-5.07E+00	1.41E+01	-0.361	0.72342	
PlantPlant 4	1.55E-01	5.49E+00	0.028	0.97793	
PlantPlant 5	1.65E+00	6.50E+00	0.255	0.80251	
costC:PlantPlant 2	-4.50E-06	8.37E-06	-0.538	0.59862	
costC:PlantPlant 3	-8.11E-06	1.50E-05	-0.541	0.5967	
costC:PlantPlant 4	-4.08E-06	8.30E-06	-0.492	0.63007	
costC:PlantPlant 5	-1.62E-06	3.89E-06	-0.416	0.68348	
costE:PlantPlant 2	1.75E-06	1.48E-06	1.185	0.25428	
costE:PlantPlant 3	-1.93E-06	8.92E-07	-2.159	0.04748	*
costE:PlantPlant 4	1.21E-06	1.35E-06	0.893	0.38601	
costE:PlantPlant 5	-1.57E-06	7.59E-07	-2.074	0.05571	.
costG:PlantPlant 2	-1.74E-06	2.26E-06	-0.767	0.45518	
costG:PlantPlant 3	NA	NA	NA	NA	
costG:PlantPlant 4	NA	NA	NA	NA	
costG:PlantPlant 5	NA	NA	NA	NA	
costS:PlantPlant 2	9.76E-07	4.78E-06	0.204	0.84092	
costS:PlantPlant 3	2.20E-06	4.83E-06	0.455	0.6554	
costS:PlantPlant 4	-3.07E-06	5.61E-06	-0.548	0.59185	
costS:PlantPlant 5	2.57E-06	2.46E-06	1.046	0.31192	
eUseS:PlantPlant 2	1.40E-04	3.70E-03	0.038	0.97039	
eUseS:PlantPlant 3	-7.01E-04	3.69E-03	-0.19	0.85177	
eUseS:PlantPlant 4	-7.52E-04	3.67E-03	-0.205	0.84023	

	Estimate	Std. Error	t value	Pr(>|t|)	
eUseS:PlantPlant 5	NA	NA	NA	NA	
productA:PlantPlant 2	-3.35E-02	2.59E-02	-1.295	0.21487	
productA:PlantPlant 3	-1.79E-02	2.93E-02	-0.61	0.5513	
productA:PlantPlant 4	-5.06E-02	2.15E-02	-2.358	0.03236	*
productA:PlantPlant 5	-7.38E-02	3.64E-02	-2.031	0.06033	.
productP:PlantPlant 2	3.17E-02	3.25E-02	0.974	0.34567	
productP:PlantPlant 3	-2.64E-02	2.45E-02	-1.076	0.29873	
productP:PlantPlant 4	1.64E-02	2.08E-02	0.79	0.44196	
productP:PlantPlant 5	5.66E-02	3.09E-02	1.83	0.08717	.
productS:PlantPlant 2	5.85E-03	4.96E-02	0.118	0.90766	
productS:PlantPlant 3	4.31E-02	5.37E-02	0.801	0.43559	
productS:PlantPlant 4	3.56E-02	4.87E-02	0.731	0.47634	
productS:PlantPlant5	-1.08E-02	4.95E-02	-0.218	0.83028	

Table S4. Forward stepwise model regression summary

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.20E+01	2.90E-01	41.402	0.000097	***
saleA	1.70E-01	7.39E-02	2.304	0.026797	*
saleS	-1.05E-01	5.01E-02	-2.094	0.042966	*
costE	4.68E-07	2.45E-07	1.911	0.063532	.
ePE	-6.82E-03	3.60E-03	-1.894	0.065901	.
eUseS	2.65E-03	1.46E-04	18.086	0.000097	***
productA	7.68E-04	1.49E-02	0.051	0.959255	
productP	-2.35E-02	1.42E-02	-1.649	0.107344	
PlantPlant 2	8.65E-01	3.91E-01	2.213	0.032941	*
PlantPlant 3	5.51E+00	4.12E-01	13.363	6.14E-16	***
PlantPlant 4	3.62E+00	4.17E-01	8.674	1.52E-10	***
PlantPlant 5	2.62E+00	4.69E-01	5.571	2.21E-06	***
costE:PlantPlant 2	1.95E-09	7.10E-08	0.028	0.978204	
costE:PlantPlant 3	-2.09E-07	1.62E-07	-1.289	0.205176	
costE:PlantPlant 4	-4.19E-08	6.15E-08	-0.681	0.500076	
costE:PlantPlant 5	NA	NA	NA	NA	
productA:PlantPlant 2	-6.78E-03	1.90E-02	-0.357	0.722847	
productA:PlantPlant 3	-2.85E-03	1.60E-02	-0.179	0.859144	
productA:PlantPlant 4	-1.91E-02	1.53E-02	-1.248	0.219503	
productA:PlantPlant 5	-6.83E-03	2.00E-02	-0.341	0.734622	
productP:PlantPlant 2	1.47E-02	2.27E-02	0.648	0.520928	
productP:PlantPlant 3	-1.57E-02	1.84E-02	-0.853	0.399156	
productP:PlantPlant 4	4.05E-03	1.77E-02	0.229	0.820034	
productP:PlantPlant 5	1.01E-02	1.81E-02	0.556	0.581413	
eUseS:PlantPlant 2	9.16E-04	2.53E-04	3.621	0.000855	***
eUseS:PlantPlant 3	6.74E-05	1.70E-04	0.395	0.694787	
eUseS:PlantPlant 4	2.14E-04	1.86E-04	1.152	0.256323	
eUseS:PlantPlant 5	5.74E-03	1.72E-04	33.323	0.000097	***

Table S5. LASSO minimum model coefficients summary

	Coefficients
saleA	0.1
saleS	-0.4
costE	0.1
ePE	0
eUseS	2.1
productA	-0.1
productP	-0.7
costE.1	0
productA.1	0
productP.1	0
eUseS.1	0
PlantPlant.2	0.2
PlantPlant.3	1.1
PlantPlant.4	0.7
PlantPlant.5	0.5
costE.PlantPlant.2	-0.1
costE.PlantPlant.3	0.1
costE.PlantPlant.4	-0.2
costE.PlantPlant.5	-0.2
productA.PlantPlant.2	0
productA.PlantPlant.3	0
productA.PlantPlant.4	-0.9
productA.PlantPlant.5	0
productP.PlantPlant.2	0.4
productP.PlantPlant.3	-1.3
productP.PlantPlant.4	-0.2
productP.PlantPlant.5	0
eUseS.PlantPlant.2	0.6
eUseS.PlantPlant.3	0
eUseS.PlantPlant.4	0.1
eUseS.PlantPlant.5	1.9