Predicting Smart Electrical Grid Stability

Isabella Chittumuri
Professor Vitaly Druker
STAT 706: General Linear Models I
City University of New York, Hunter College

Abstract

The functionality of the modern world depends highly on electricity. Electrical production and consumption happen on a real time basis. The second electricity is produced it is consumed, regardless of how far the producer is from the consumer. The power grid has electrical devices and equipment to monitor and maintain stability of electrical frequency. In this project, we study statistical analysis of the smart gird to model and determine the stability of the system.

---

1. Introduction

1.1 What is a smart gird?

Previous to the 21st century, energy production and consumption were limited to a one way interaction between utilities and customers. This type of communication was vulnerable to sudden changes in electricity demand, which lead to multiple power outages. The smart grid introduces a two way dialogue where electrical information can be exchanged between utilities and customers. This new system enables both parties to effectively manage electrical costs and usage [3].

In addition, renewables have become a growing source for electrical power. However, they can have massive and unpredictable variations in capacity, providing irregular production of power to the grid. The smart grid provides the data and equipment needed to enable renewables to transport energy onto the grid and optimize its usage. Renewables also allow users, such as homeowners, to produce and supply their own energy. Smart grids can efficiently facilitate this bidirectional energy flow [2].

1.2 Grid Stability

The power grid uses alternating current (AC), where the direction of voltage and current are constantly switching between positive and negative. This oscillation is known as electrical frequency. The more cycles that occur per second, the higher the frequency. In the United States, the grid is considered stable at a frequency of 60 Hz (hertz) [2].

Changes in supply and demand for electricity can have a major effect on grid frequency. If demand for electricity is higher than supply, the frequency will decrease. Alternatively, if demand for electricity is lower than supply, frequency will increase. Frequency deviations from unmet changes in voltage can result in grid instability and lead to outages and equipment damage.

1.3 Decentral Smart Grid Control

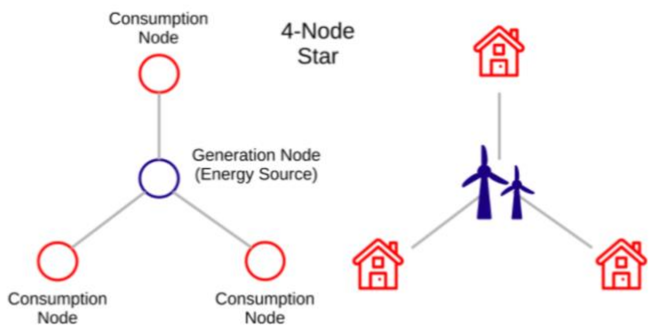
Decentral Smart Grid Control (DSGC) is a system where electrical price is directly linked to grid frequency. Grid frequency provides all the necessary information about power balance. In doing so, it is sufficient to match supply and demand without the need for a centralized infrastructure [4].

Schäfer describes DSGC as a differential equation-based mathematical model. This equation assumes that price only depends on frequency, taking into account time delays in adjusting to consumption and production. This system implements a demand response without significant changes to variable form and provides an efficient measure of ensuring grid stability [4].

1.4 Data

This project used the "Electrical Grid Stability Simulated Dataset" created by Vadim Arzamasov, donated in November 2018 to the UCI Machine Learning Repository. The simulated dataset established accuracy and validity because it was derived from current models of DSGC with its respective assumptions of what quantifies grid stability.

Figure 1. Four-node star electrical grid system



Because of the DSGC model's intrinsic simplifications, Arzamasov and his team created a simulation based on space-filling and decision trees with different DSGC parameters to process results. It simulated a four-node star electrical grid system, depicted in **Figure 1**, with centralized production where one power source node is connected to three consumer nodes.

The dataset was available in csv format and it was consistent in the measurement of 14 attributes and 10,000 observations. There were no missing values, which makes the data complete with equal probability. There were 11 independent attributes and two dependent attributes:

Independent Attributes: Electricity producer node, i=1; consumer nodes, i=2-4

1. tau[i]: reaction time in seconds, the delay between a price change and adaptation to it. A real value within the range [0.5,10].
2. p[i]: mechanical power produced/consumed. A real value within the range [-2.0,-0.5] for consumers, positive for producer. Because power produced equals power consumed, p1 = abs(p2 + p3 + p4).
3. g[i]: price elasticity. A real value within the range [0.05,1].

Price elasticity is a measure of how participants react to the prices of products and services through demand and supply. In terms of electricity, when demand is high, price is high. Likewise, when demand is low, price is low.

Dependent Attributes:

1. stab: the maximal real part of the characteristic equation root.
2. stabf: categorical stability label of the system, stable or unstable.

In this report, we explore and model attributes that identify and predict smart electrical grid stability. All of the analysis was done using R 3.6.3

2. Exploratory Analysis

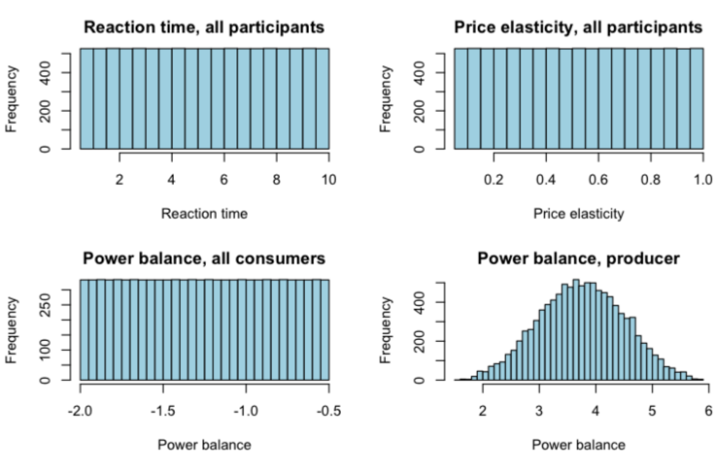
2.1 Data Observations

The first thing we did was call the summary function on the dataset. We saw that the five number summaries for all independent attributes are similar between producer and consumers, with slight deviations in value. The one exception found was in power balance, where producer values equaled the absolute sum of all consumer values. However, power balance among all three consumers was consistent in value similarities.

The dependent attribute stabf identified 3620 row vectors as stable and 6380 row vectors as unstable. Of the total number of observations in the dataset, the percentage of stability is 36.2%. We created a new column vector called num_stabf, resulting in the numeric binary output of stabf where 0 equates to unstable and 1 equates to stable.

After, we created histogram plots for all of the independent attributes. Attributes, excluding power balance, resulted in similar plots for all four participants, i of 1-4. **Figure 2** depicts four histograms. Reaction time, consumer power balance, and price elasticity depicted a uniform distribution; and producer power balance depicted a normal distribution.

Figure 2. Frequency of Reaction Time, Power Balance, and Price Elasticity



Due to the symmetric nature of the attributes, the mean was a good measure of central tendency. Variance and standard deviation measure how dispersed the data values are around the mean. After calculations, we found that the unadjusted and adjusted values for variance and standard deviation were the same. Therefore, we decided to only include the unadjusted values. **Table 1** shows the attributes' definitions and their descriptive statistics.

Table 1. Independent attribute description

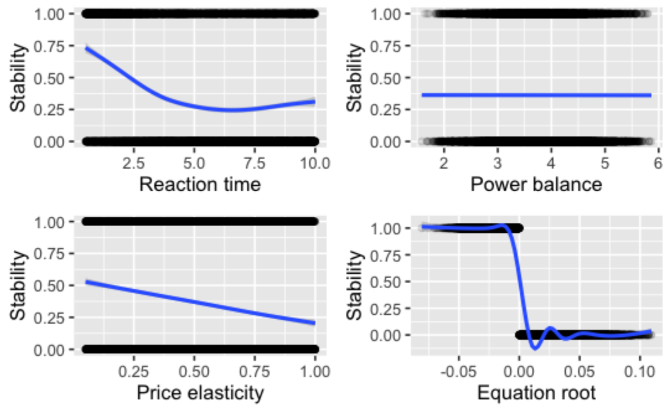
Attributes	def	μ_y	σ_Y^2	σ_Y
tau[i]	Reaction time	5.25	7.25	2.74
g[i]	Price elasticity	0.53	0.08	0.27
p[1]	Producer power balance	3.75	0.57	0.75
p[i, 2-4]:	Consumer power balance	-1.25	0.19	0.43

2.2 Correlations

Next, we plotted the independent attributes against each other and found no correlations. Then we plotted each independent attribute against the dependent numerical attribute num_stabf, the binary output of stability. All independent attributes, including power balance, resulted in similar correlation plots for all four participants. We also plotted stab, the maximal real part of the characteristic equation root, against num_stabf. **Figure 3** depicts these four correlations plots.

When stability was plotted against reaction time, it showed a negative non-linear relationship. When stability was plotted against power balance, it showed no relationship. When stability was plotted against price elasticity, it showed a negative linear relationship. Lastly, when stability was plotted against the equation root, it showed a direct relationship. Specifically, if the equation root was negative, the system was linearly stable and if the equation root was positive, the system was linearly unstable.

Figure 3. Correlation of four attributes against one attribute



3. Modeling

We wanted to predict grid stability. This referred to a binary response, a special case of the binomial which requires two response levels. In this case, our two response levels were stable and unstable.

3.1 Model Selection

To model grid stability, we first created a generalized linear model (GLM) by fitting all independent attributes as predictors with num_stabf as the response. GLM required us to specify the family, the error distribution and link function used in the model. In this case, we used the binomial family with the logit link function. We saw that reaction time and price elasticity were highly significant, whereas power balance was insignificant.

Next we looked at each predictor in respect to the model's Akaike information criterion (AIC), an estimator of the out-of-sample prediction error. We used the step function to remove one predictor at a time and refit the model until it produced the lowest AIC possible. This function dropped all attributes for power balance. The resulting model only included the predictors of reaction time and price elasticity, yielding an AIC value of 7835.6.

To further confirm this result, we ran an anova chi-square test that compared the original model to the resulting model. Seen in **Figure 4**, the p-value was 0.27, which was greater than 0.05. This suggested that we can drop all attributes for power balance as predictors during the fitting of the model.

Figure 4. Anova Chi-Square Test

```
Analysis of Deviance Table

Model 1: num_stabf ~ tau1 + tau2 + tau3 + tau4 + p1 + p2 + p3 + p4 + g1 +
    g2 + g3 + g4
Model 2: num_stabf ~ tau1 + tau2 + tau3 + tau4 + g1 + g2 + g3 + g4
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      9988     7813.6
2      9991     7817.6 -3  -3.9365   0.2684
```

3.2 Final Model

Our final model included the producer and consumers reaction times as well as their price elasticities. Seen in **Figure 5,** all of the predictors had a low p-value of 2^{-16}, which suggested high significance. β_0 was 11.78026, the value of Y when all of the X_i coefficients are zero. The β_i, for every i of 1 through 8, represented a negative slope. This meant that for every unit increase in X_i, we expect stability to go down by β_i.

Seen in **Figure 6**, we produced 95% confidence intervals (CI) for the β_i values. This meant that if we repeated the experiment multiple times, we expect that 95% of the CI to contain the true value of β_i.

Figure 5. Summary of Final Model

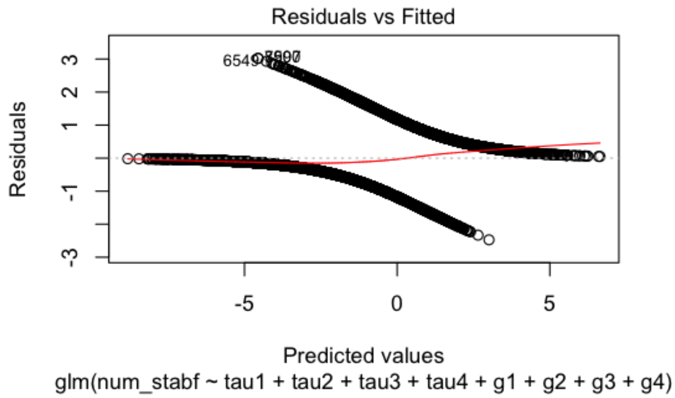
$$Y = 11.78026 - 0.31423 * tau1 - 0.32254 * tau2 - 0.31855 * tau3 - 0.33201 * tau4 - 2.68971 * g1 - 2.91493 * g2 - 3.12117 * g3 - 2.85428 * g4$$

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 11.78026    0.24821   47.46   <2e-16 ***
tau1        -0.31423    0.01129  -27.84   <2e-16 ***
tau2        -0.32254    0.01139  -28.33   <2e-16 ***
tau3        -0.31855    0.01139  -27.97   <2e-16 ***
tau4        -0.33201    0.01147  -28.95   <2e-16 ***
g1          -2.68971    0.11043  -24.36   <2e-16 ***
g2          -2.91493    0.11218  -25.98   <2e-16 ***
g3          -3.12117    0.11349  -27.50   <2e-16 ***
g4          -2.85428    0.11176  -25.54   <2e-16 ***
```

Figure 6. Confidence Intervals of Beta Values

```
                2.5 %     97.5 %
(Intercept) 11.2997112 12.2727805
tau1        -0.3364999 -0.2922514
tau2        -0.3450054 -0.3003716
tau3        -0.3410223 -0.2963747
tau4        -0.3546527 -0.3096967
g1          -2.9074408 -2.4745276
g2          -3.1362169 -2.6964274
g3          -3.3451223 -2.9002109
g4          -3.0747266 -2.6365909
```

Figure 7. Residuals vs Fitted

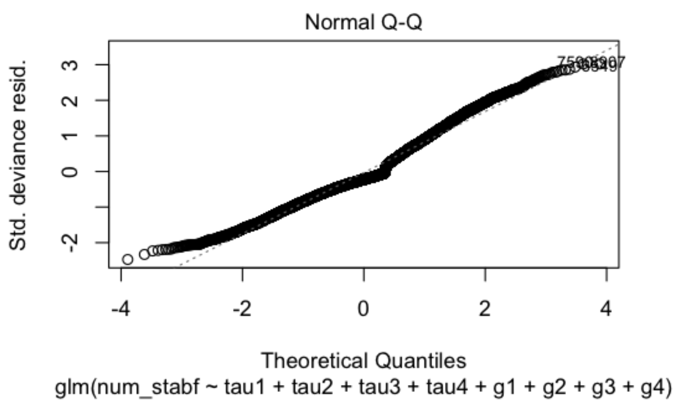


4. Diagnostics

4.1 Residuals

A residual is the difference between the observed value of Y and the estimated value of $\hat{Y}$. **Figure 7** depicts the Residuals vs. Fitted plot. Ideally, we wanted to be along the dotted line that runs across zero on the y-axis. This line means that the expected values are the same as the observed values. The red line was the result of our model's predicted values against the residuals. We had a relatively straight red line along the dotted line. This meant that our estimated values had very little deviations to the observed values.

Figure 8. Normal Q-Q



We expected the residuals to be normal. **Figure 8** looks at all the residuals and compares them according to a normal distribution, estimating the residual variance. If the errors are normally distributed, they would follow the $y = x$ dotted line. We saw that majority of the our residual variance followed the $y = x$ line. However, we saw slight deviation on the tails, which suggested abnormality in those areas.

Figure 9. Residuals vs Predictors

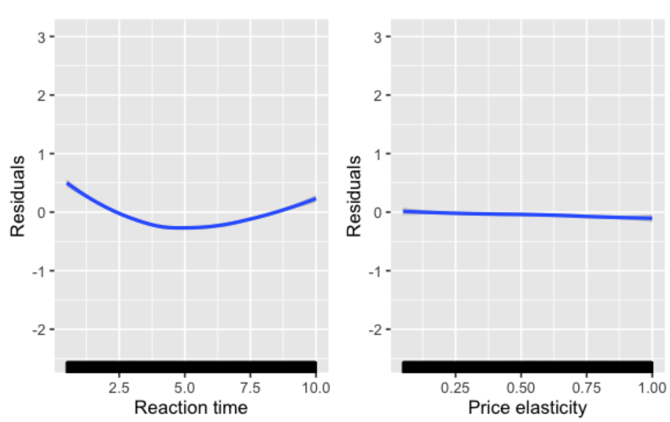


Figure 9 shows both reaction time and price elasticity against the residuals. When looking at the left plot, we saw that the residuals varied as reaction time increased. Regardless, this curve was not drastic and stayed relatively in a straight line. When looking at the right plot, we saw that the residuals were consistent as price elasticity increased, representing no relationship in variance.

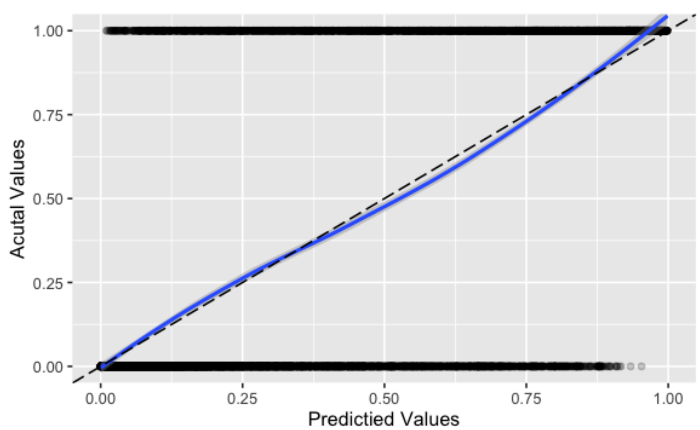
5. Predictions

5.1 Calibration Curve

A calibration curve plots predictions on the x-axis and true values on the y-axis. Just as in our residual plots, we wanted our model to follow the $y = x$ line. **Figure 10** depicts our model's calibration curve.

For the most part, our line followed the $y = x$ line, with a small underestimations in the middle and overestimations on the tails. For clarification, if we looked at 0.63 for 100 grid frequencies, ideally 63% of them would be stable. However, our model would predict only about 58% of them to be stable.

Figure 10. Calibration Curve



5.2 Confusion matrix

A confusion matrix is a table used to describe the performance of a classification model when the true values are known. It compares the true values with the model's predicted values. **Table 2** shows the results of our confusion matrix with a 0.5 cutoff.

From this confusion matrix, we were able to quantify the accuracy of our model. Accuracy is the number of correctly predicted cases divided by the total number of cases. Our model was 81.49% accurate. However, accuracy can be misleading in an imbalanced dataset and as we mentioned before, our dataset is only 36.2% stable. Because of this imbalance, we looked at other classification quantifiers.

Table 2. Confusion matrix

n = 10000	Predicted	
Actual	stable	unstable
stable	2547	1073
unstable	778	5602

In regards to specificity, of those that were actually unstable, 87.81% were correctly predicted as unstable. In regards to sensitivity, of those that were actually stable, 70.36% were correctly predicted as stable. In regards to positive predicted value (PPV), of those predicted as stable, 76.6% were actually stable. Lastly, in regards to negative predicted value (NPV), of those predicted as unstable, 83.93% were actually unstable.

Accuracy = (2547+5602)/(10000) = .8149.

Specificity = 5602/(5602 + 778) = .8781

Sensitivity = 2547/(2547 + 1073) = 0.7036

PPV= 2547/(2547 + 778) = 0.7660

NPV = 5602/(5602 + 1073) = 0.8393

5.3 Odds Ratio

An odds ratio expresses relative difference, an alternative scale for representing chance. If an odds ratio is more than 1, there is a greater likelihood of having the outcome. If an odds ratio is less than 1, there is a lesser likelihood of having the outcome. For an odd ratio that is less than 1, we have to subtract it from 1 to get the actual odds ratio. Seen in **Table 3**, we calculated the odds ratio of the coefficients by exponentiating β_i.

Table 3. Odds Ratios of model coefficients

Coef.	Odds Ratio	Coef.	Odds Ratio
Intercept	130647.1		
tau1	0.7303535	g1	0.06790042
tau2	0.7243105	g2	0.05420788
tau2	0.7272058	g3	0.04410557
tau3	0.7174766	g4	0.05759710

Looking at just the producer coefficients, we saw that both reaction time (tau1) and price elasticity (g1) had an odds ratio of less than 1. Once we subtracted this value from 1, we found the odds ratio of the grid being stable was 27% less likely with every unit increase in tau1 and 93% less likely with every unit increase in g1.

Then we looked to answer the question: what is the difference in odds for testing stable when tau1 goes from the 1st quantile to 3rd quantile? For tau1, the difference between the 1st and 3rd quantile was 4.75. We multiplied this difference to the tau1 estimate and 95% CI. And then we exponentiated the result to get the values in an odds ratio scale. The results are seen in **Table 4.**

The odds ratio estimate was 0.22 and we were 95% confident that the true value was between the CI. We subtracted the estimate from 1 to get 0.78. This meant that the odds ratio of the grid being stable was 78% less likely when tau1 went from the 1st quantile to the 3rd quantile. We executed the same procedure for g1, and found that the odds ratio of the grid being stable was 72% less likely when g1 went from the 1st quantile to the 3rd quantile.

Because all the predictors have similar statistical values, we generalized these producer odds ratios to consumers 2 through 4.

Table 4. Difference in Odds Ratios from 1st-3rd quantile

Coefficient	Conf. low	Estimate	Conf. high
Reaction Time	0.2022387	0.2248071	0.2495402
Price Elasticity	0.2513817	0.2787664	0.3087614

5.4 Probability

Probability reflects the chance or likelihood a particular event will occur. In this case, we looked at the event when all predictors were at mean value except for one. The first predictor we changed was tau1. For event 1, we assigned tau1 the value of its 1st quantile, 2.87. Using our model, we predicted the probability of stability for this event to be 42%. For event 2, we assigned tau1 the value of its 3rd quantile, 7.62. Using our model, we predicted the probability of stability for this event to be 14%.

The second predictor we changed was g1, again while all other predictors were at mean value. For event 1, we assigned g1 the value of its 1st quantile, 0.28, and found the probability of stability to be 39%. For event 2, we assigned g1 the value of its 3rd quantile, 0.76, and found the probability of stability to be 16%.

Similarly with the odds ratios, we generalized these producer probabilities to consumers 2 through 4.

6. Results and Conclusion

Overall, our analysis of smart grid stability revealed that only reaction time and price elasticity proved to be significant predictors. Our residual plots concluded that our estimated values had very little deviations from the observed values. And our model had classification scores that were between 70-84%.

This may seem like a good model, but due to the fragility of the grid, these classification deviations can lead to outages and equipment damage. Since the 21st century revolves around constant electrical usage, we would want our model classification scores to be closer to 90%.

We found that the longer the system takes to adapt to price change, the odds of the grid being stable was about 75-80% less likely when going from the 1st to the 3rd quantile. Similarly we found that the higher the price elasticity, the odds of the grid being stable was about 70-75% less likely when going from the 1st to the 3rd quantile.

When looking at a fixed event with other predictors at mean value, we found that a reaction time of 2.87 had a 42% chance of being stable. For that same fixed event, we found that a reaction time of 7.62 had a 14% chance of being stable. Similar probabilities hold for price elasticity of 0.28 and 0.76, respectively.

Because DSGC is simplified differential equation-based model, this method comes with fixed input issues when classifying grid stability. Though Arzamasov and his team made adjustments to the parameters of this model, future research can be done to explore other metrics associated with grid stability. In particular, researchers may look to include predictors that involve more traditional methods of modeling the smart grid.

Since the dataset we used was a simulated one, we can improve this model by using real data values. Another method for improvement can be linked to using data from a dynamic grid system that involves more participants than that of a four-node star system. If we can produce a highly accurate model, we can better predict smart grid stability while safely and efficiently providing for electrical needs.

References

[1] *UCI Machine Learning Repository: Electrical Grid Stability Simulated Data Data Set*, Arzamasov, Vadim. 2018. Web. https://archive.ics.uci.edu/ml/datasets/Electrical+Grid+Stability+Simulated+Data+#

[2] Hillhouse, Grady. "How Does the Power Grid Work?" *Practical Engineering*. 25 June 2019. Web. https://practical.engineering/blog/2019/6/24/how-does-the-power-grid-work

[3] Energy, Department of. "Smart Grid." *Energy.gov*. Web. https://www.energy.gov/science-innovation/electric-power/smart-grid

[4] Schäfer, Benjamin, Moritz Matthiae, Marc Timme, and Dirk Witthaut. *Decentral Smart Grid Control*. Research Paper. IOP, 2015. Web. https://iopscience.iop.org/article/10.1088/1367-2630/17/1/015002

[5] Arzamasov, Vadim, Klemens Böhm, and Patrick Jochem. *Towards Concise Models of Grid Stability*. IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, 2018. Web. http://dbis.ipd.kit.edu/english/2635.php