# Improving Day Ahead Electricity Load Forecasts with Google Trends

Cameron Mulder
James Woods

August 5, 2014

**Abstract**

Modern short term load forecasting has grown in analytically complexity and sophistication. Day ahead forecasts now commonly use neural nets, Monte Carlo simulations and a wealth of historical data. What they have not done is fully captured the sentiment and intentions of the people using the electricity. This paper introduces Google Trend data, a summary of Google searches, as a way of capturing this sentiment and refining forecasts. We show with drop all forward cross validation that this amendment decreases forecast uncertainty by approximately 5% when compared to a statistically adjusted forecast and by over 50% when compared to raw forecasts.

# 1 Introduction

1. Intro to short term load forecasting.
2. Why crowd sourced, non technical, information could be useful.
3. Google trends is the summation of Google searches.
4. Outline of paper

# 2 Data Sources

## 2.1 PJM Load Forecasts and Actuals

1. Data sources.
2. Documentation of forecasting.
3. Forecast bias
4. Statistically adjusted forecasts.
5. Note that almost all hours are biased and that co-movements are good for peak hours

## 2.2 Google Trends

1. Where to get the data
2. Limitations
3. Forming a population weighted index.
4. Other common searches that will be used as counter examples.

Figure 1: Confidence Intervals for Intercept Statistically Adjusted Models (95%)

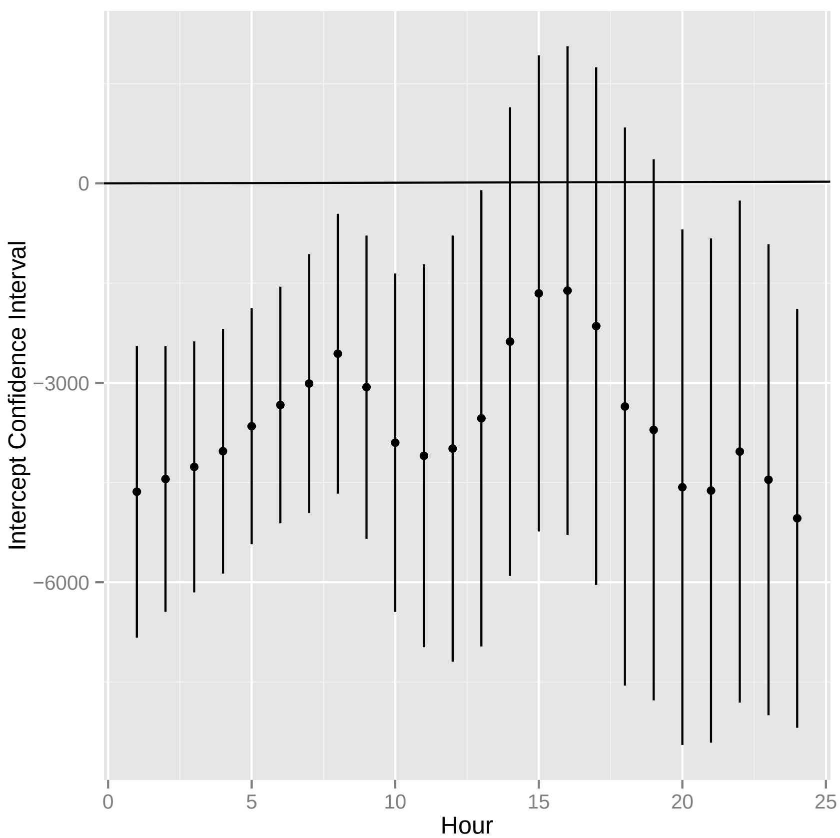


Figure 2: Confidence Intervals for Co-Movement Statistically Adjusted Models (95%)

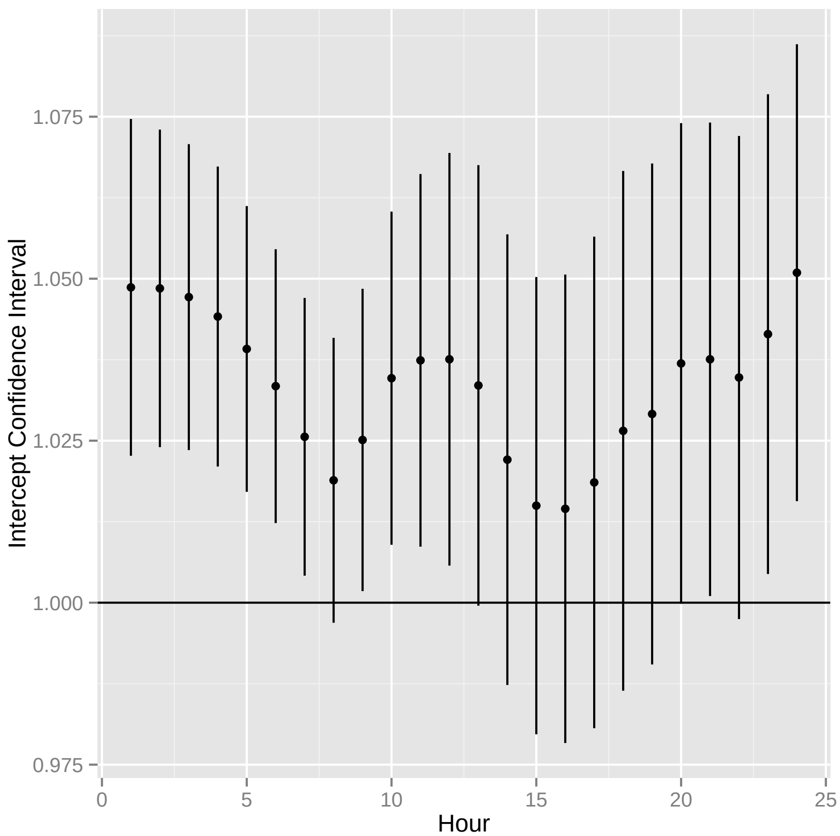


Figure 3: State Weather Trends Indexes Over Time

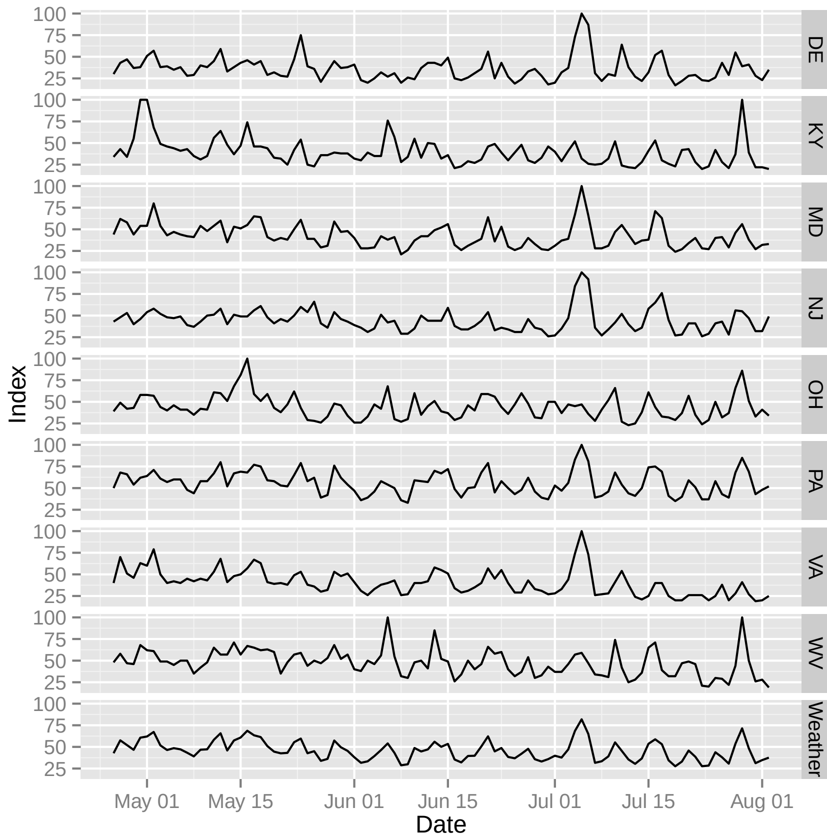
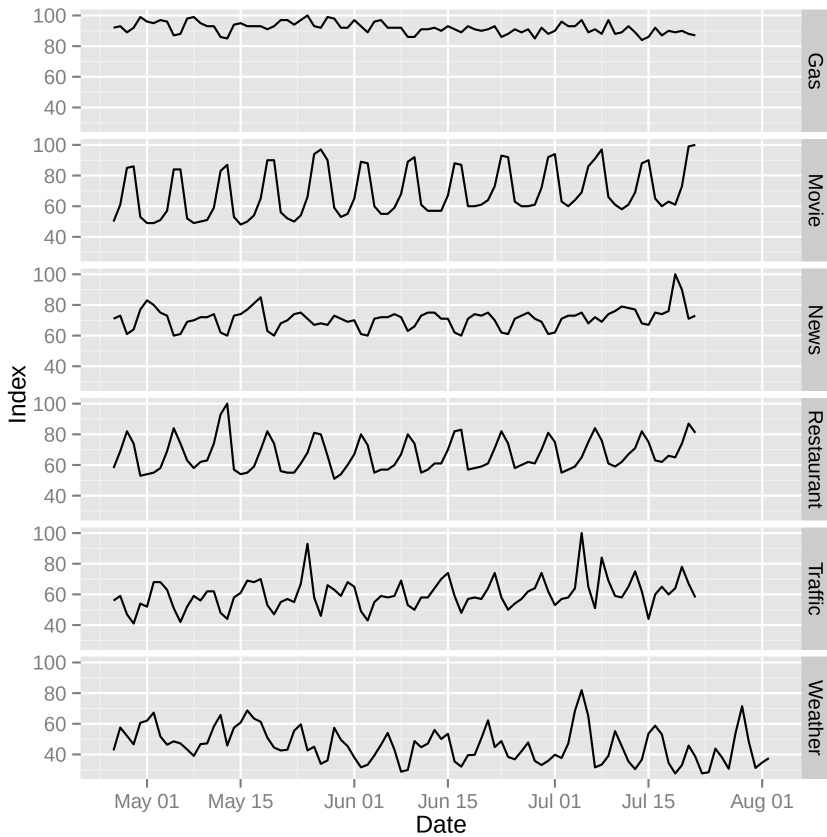


Figure 4: Trends Indexes Over Time



# 3 Post Forecast Addition of Google Trends Data

1. Simple hourly models with Trends.
2. Gross comparison with actual forecast and statistically adjusted forecasts.
3. Why this is insufficient.

## 3.1 Drop Forward Cross-validation

Table 1: Improvement in Forecasts Relative to Gross, Statistically Adjusted, Drop Forward CV (Percent)

| Hour | Direct | Statistically Adjusted (Raw) | Statistically Adjusted (CV) |
|---|---|---|---|
| 1 | | 6.832 | 4.386 |
| 2 | | 7.018 | 4.279 |
| 3 | | 6.872 | 4.672 |
| 4 | | 6.553 | 4.148 |
| 5 | | 6.630 | 3.965 |
| 6 | | 6.871 | 4.153 |
| 7 | | 7.850 | 4.299 |
| 8 | | 9.913 | 7.156 |
| 9 | | 10.652 | 11.218 |
| 10 | | 9.261 | 11.145 |
| 11 | | 7.892 | 10.028 |
| 12 | | 6.713 | 8.147 |
| 13 | | 6.403 | 6.824 |
| 14 | | 6.160 | 6.077 |
| 15 | | 6.027 | 5.374 |
| 16 | | 5.915 | 5.427 |
| 17 | | 5.617 | 4.946 |
| 18 | | 4.961 | 4.369 |
| 19 | | 4.750 | 4.365 |
| 20 | | 4.872 | 2.099 |
| 21 | | 5.694 | 1.257 |
| 22 | | 6.267 | 3.485 |
| 23 | | 5.793 | 3.631 |
| 24 | | 5.160 | 2.972 |

1. Cross validation concepts.
2. Why drop forward cross validation is the right concept.
3. Comparison of drop forward statistically adjusted and Trends adjusted with gross comparisons.
4. Reiteration that comparison with raw forecasts is a slam dunk.

## 3.2 Counter-factual Test with Other Common Google Searches

1. Comparison with: news, recipe, traffic, gas.
2. Note that some of them kinda work.

Figure 5: Confidence Intervals for “Weather” in Trends Models (95%)

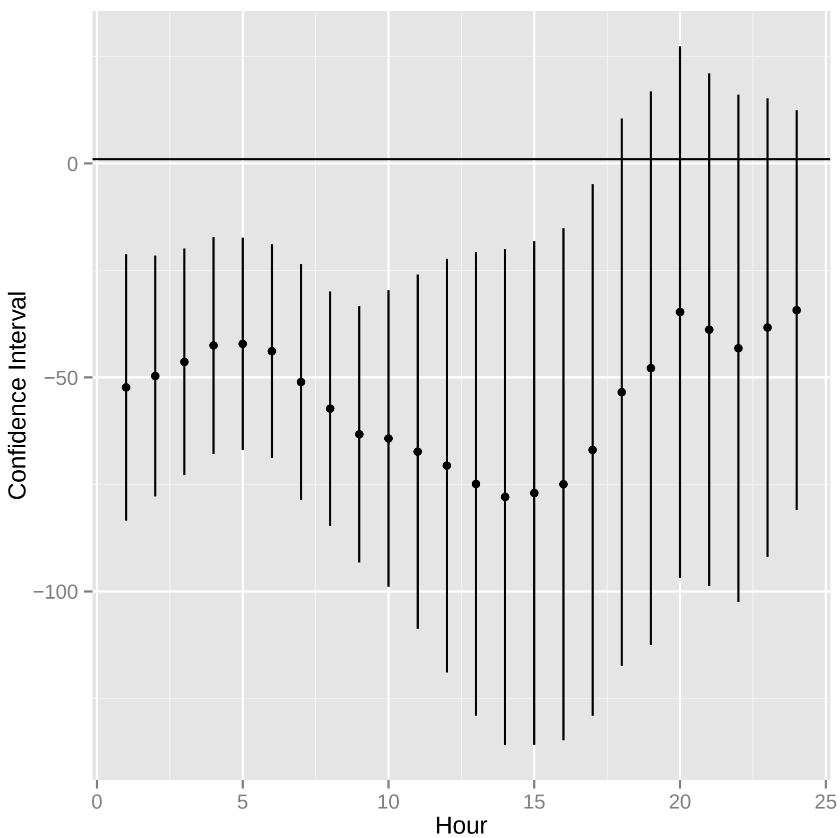


Table 2: Alternate Google Search Models for Hour 19

| | Hour 19 Load | | | | |
|---|---|---|---|---|---|
| | News | Gas | Traffic | Restaurant | Movie |
| | (1) | (2) | (3) | (4) | (5) |
| F19 | 0.987*** | 1.001*** | 1.004*** | 0.985*** | 0.974*** |
| | (0.032) | (0.032) | (0.032) | (0.031) | (0.032) |
| NewsTrends | −109.127** | | | | |
| | (54.369) | | | | |
| GasTrends | | −100.782 | | | |
| | | (88.731) | | | |
| TrafficTrends | | | −19.657 | | |
| | | | (31.495) | | |
| RestaurantTrends | | | | 84.677** | |
| | | | | (33.519) | |
| MovieTrends | | | | | 61.023** |
| | | | | | (24.515) |
| Constant | 8,950.203 | 9,010.925 | 581.895 | −4,227.396 | −1,637.760 |
| | (5,944.583) | (9,658.220) | (4,357.965) | (3,326.853) | (3,064.397) |
| Observations | 88 | 88 | 88 | 88 | 88 |
| Log Likelihood | −811.334 | −812.185 | −813.666 | −810.691 | −811.109 |
| Akaike Inf. Crit. | 1,632.669 | 1,634.370 | 1,637.332 | 1,631.383 | 1,632.219 |
| Bayesian Inf. Crit. | 1,644.882 | 1,646.584 | 1,649.546 | 1,643.596 | 1,644.432 |

*Note:* *p<0.1; **p<0.05; ***p<0.01

# 4 Summary and Conclusions

# A Hourly Models with Weather Searches

Table 3: Hour 1

| | *Dependent variable:* |
|---|---|
| | Hour 1 |
| Forecast | 0.990*** |
| | (0.023) |
| Weather | −52.323*** |
| | (15.875) |
| Constant | 2,559.469 |
| | (2,170.601) |
| Observations | 96 |
| Log Likelihood | −827.731 |
| Akaike Inf. Crit. | 1,665.461 |
| Bayesian Inf. Crit. | 1,678.124 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 4: Hour 2

| | *Dependent variable:* |
|---|---|
| | Hour 2 |
| Forecast | 0.990*** |
| | (0.023) |
| Weather | −49.664*** |
| | (14.355) |
| Constant | 2,332.603 |
| | (2,049.462) |
| Observations | 96 |
| Log Likelihood | −819.217 |
| Akaike Inf. Crit. | 1,648.434 |
| Bayesian Inf. Crit. | 1,661.097 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 5: Hour 3

| | *Dependent variable:* |
|---|---|
| | Hour 3 |
| Forecast | 0.993*** |
| | (0.024) |
| Weather | −46.363*** |
| | (13.513) |
| Constant | 1,923.066 |
| | (2,038.202) |
| Observations | 96 |
| Log Likelihood | −813.164 |
| Akaike Inf. Crit. | 1,636.327 |
| Bayesian Inf. Crit. | 1,648.990 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 6: Hour 4

| | *Dependent variable:* |
|---|---|
| | Hour 4 |
| Forecast | 0.995*** |
| | (0.025) |
| Weather | −42.525*** |
| | (12.939) |
| Constant | 1,537.579 |
| | (2,053.486) |
| Observations | 96 |
| Log Likelihood | −808.170 |
| Akaike Inf. Crit. | 1,626.340 |
| Bayesian Inf. Crit. | 1,639.003 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 7: Hour 5

| | *Dependent variable:* |
|---|---|
| | Hour 5 |
| Forecast | 0.993*** |
| | (0.025) |
| Weather | −42.147*** |
| | (12.659) |
| Constant | 1,671.668 |
| | (2,028.787) |
| Observations | 96 |
| Log Likelihood | −806.323 |
| Akaike Inf. Crit. | 1,622.647 |
| Bayesian Inf. Crit. | 1,635.310 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 8: Hour 6

| | *Dependent variable:* |
|---|---|
| | Hour 6 |
| Forecast | 0.994*** |
| | (0.022) |
| Weather | −43.865*** |
| | (12.755) |
| Constant | 1,638.583 |
| | (1,877.205) |
| Observations | 96 |
| Log Likelihood | −809.113 |
| Akaike Inf. Crit. | 1,628.226 |
| Bayesian Inf. Crit. | 1,640.889 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 9: Hour 7

| | *Dependent variable:* |
|---|---|
| | Hour 7 |
| Forecast | 0.986*** |
| | (0.019) |
| Weather | −51.069*** |
| | (14.068) |
| Constant | 2,262.272 |
| | (1,705.204) |
| Observations | 96 |
| Log Likelihood | −819.258 |
| Akaike Inf. Crit. | 1,648.516 |
| Bayesian Inf. Crit. | 1,661.179 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 10: Hour 8

| | *Dependent variable:* |
|---|---|
| | Hour 8 |
| Forecast | 0.992*** |
| | (0.016) |
| Weather | −57.280*** |
| | (13.964) |
| Constant | 2,320.879 |
| | (1,567.874) |
| Observations | 96 |
| Log Likelihood | −819.424 |
| Akaike Inf. Crit. | 1,648.849 |
| Bayesian Inf. Crit. | 1,661.512 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 11: Hour 9

| | *Dependent variable:* |
|---|---|
| | Hour 9 |
| Forecast | 1.000*** |
| | (0.017) |
| Weather | −63.284*** |
| | (15.272) |
| Constant | 2,161.022 |
| | (1,771.994) |
| Observations | 96 |
| Log Likelihood | −824.078 |
| Akaike Inf. Crit. | 1,658.157 |
| Bayesian Inf. Crit. | 1,670.820 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 12: Hour 10

| | *Dependent variable:* |
|---|---|
| | Hour 10 |
| Forecast | 1.004*** |
| | (0.019) |
| Weather | −64.260*** |
| | (17.659) |
| Constant | 1,882.086 |
| | (2,118.596) |
| Observations | 96 |
| Log Likelihood | −835.099 |
| Akaike Inf. Crit. | 1,680.197 |
| Bayesian Inf. Crit. | 1,692.860 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 13: Hour 11

| | *Dependent variable:* |
|---|---|
| | Hour 11 |
| Forecast | 1.001*** |
| | (0.021) |
| Weather | −67.335*** |
| | (21.114) |
| Constant | 2,370.312 |
| | (2,507.771) |
| Observations | 96 |
| Log Likelihood | −850.276 |
| Akaike Inf. Crit. | 1,710.552 |
| Bayesian Inf. Crit. | 1,723.215 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 14: Hour 12

| | *Dependent variable:* |
|---|---|
| | Hour 12 |
| Forecast | 0.996*** |
| | (0.023) |
| Weather | −70.596*** |
| | (24.660) |
| Constant | 3,064.845 |
| | (2,867.555) |
| Observations | 96 |
| Log Likelihood | −863.588 |
| Akaike Inf. Crit. | 1,737.175 |
| Bayesian Inf. Crit. | 1,749.838 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 15: Hour 13

| | *Dependent variable:* |
|---|---|
| | Hour 13 |
| Forecast | 0.991*** |
| | (0.024) |
| Weather | −74.887*** |
| | (27.622) |
| Constant | 3,769.514 |
| | (3,144.704) |
| Observations | 96 |
| Log Likelihood | −873.284 |
| Akaike Inf. Crit. | 1,756.568 |
| Bayesian Inf. Crit. | 1,769.231 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 16: Hour 14

| | *Dependent variable:* |
|---|---|
| | Hour 14 |
| Forecast | 0.982*** |
| | (0.024) |
| Weather | −77.924*** |
| | (29.566) |
| Constant | 4,821.008 |
| | (3,293.725) |
| Observations | 96 |
| Log Likelihood | −879.135 |
| Akaike Inf. Crit. | 1,768.270 |
| Bayesian Inf. Crit. | 1,780.933 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 17: Hour 15

| | *Dependent variable:* |
|---|---|
| | Hour 15 |
| Forecast | 0.981*** |
| | (0.025) |
| Weather | −77.012** |
| | (30.021) |
| Constant | 4,963.135 |
| | (3,352.355) |
| Observations | 96 |
| Log Likelihood | −880.334 |
| Akaike Inf. Crit. | 1,770.669 |
| Bayesian Inf. Crit. | 1,783.332 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 18: Hour 16

| | *Dependent variable:* |
|---|---|
| | Hour 16 |
| Forecast | 0.984*** |
| | (0.025) |
| Weather | −74.955** |
| | (30.520) |
| Constant | 4,599.294 |
| | (3,470.455) |
| Observations | 96 |
| Log Likelihood | −881.933 |
| Akaike Inf. Crit. | 1,773.867 |
| Bayesian Inf. Crit. | 1,786.530 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 19: Hour 17

| | *Dependent variable:* |
|---|---|
| | Hour 17 |
| Forecast | 0.986*** |
| | (0.027) |
| Weather | −66.928** |
| | (31.693) |
| Constant | 4,103.566 |
| | (3,680.228) |
| Observations | 96 |
| Log Likelihood | −885.451 |
| Akaike Inf. Crit. | 1,780.903 |
| Bayesian Inf. Crit. | 1,793.566 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 20: Hour 18

| | *Dependent variable:* |
|---|---|
| | Hour 18 |
| Forecast | 0.990*** |
| | (0.029) |
| Weather | −53.450 |
| | (32.628) |
| Constant | 3,134.080 |
| | (3,922.814) |
| Observations | 96 |
| Log Likelihood | −888.061 |
| Akaike Inf. Crit. | 1,786.122 |
| Bayesian Inf. Crit. | 1,798.785 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 21: Hour 19

| | *Dependent variable:* |
|---|---|
| | Hour 19 |
| Forecast | 0.990*** |
| | (0.031) |
| Weather | −47.827 |
| | (32.992) |
| Constant | 2,753.459 |
| | (4,094.667) |
| Observations | 96 |
| Log Likelihood | −888.663 |
| Akaike Inf. Crit. | 1,787.327 |
| Bayesian Inf. Crit. | 1,799.990 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Table 22: Hour 20

| | *Dependent variable:* |
|---|---|
| | Hour 20 |
| Forecast | 1.004*** |
| | (0.033) |
| Weather | −34.711 |
| | (31.683) |
| Constant | 312.192 |
| | (4,146.919) |
| Observations | 96 |
| Log Likelihood | −884.429 |
| Akaike Inf. Crit. | 1,778.857 |
| Bayesian Inf. Crit. | 1,791.520 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Table 23: Hour 21

| | *Dependent variable:* |
|---|---|
| | Hour 21 |
| Forecast | 0.998*** |
| | (0.034) |
| Weather | −38.839 |
| | (30.552) |
| Constant | 626.308 |
| | (4,168.857) |
| Observations | 96 |
| Log Likelihood | −879.990 |
| Akaike Inf. Crit. | 1,769.980 |
| Bayesian Inf. Crit. | 1,782.643 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 24: Hour 22

| | *Dependent variable:* |
|---|---|
| | Hour 22 |
| Forecast | 0.986*** |
| | (0.035) |
| Weather | −43.186 |
| | (30.240) |
| Constant | 2,483.505 |
| | (4,276.614) |
| Observations | 96 |
| Log Likelihood | −878.934 |
| Akaike Inf. Crit. | 1,767.868 |
| Bayesian Inf. Crit. | 1,780.531 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Table 25: Hour 23

| | *Dependent variable:* |
|---|---|
| | Hour 23 |
| Forecast | 0.987*** |
| | (0.037) |
| Weather | −38.339 |
| | (27.332) |
| Constant | 2,190.076 |
| | (4,069.434) |
| Observations | 96 |
| Log Likelihood | −869.293 |
| Akaike Inf. Crit. | 1,748.585 |
| Bayesian Inf. Crit. | 1,761.248 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Table 26: Hour 24

| | *Dependent variable:* |
|---|---|
| | Hour 24 |
| Forecast | 0.990*** |
| | (0.037) |
| Weather | −34.291 |
| | (23.840) |
| Constant | 1,551.670 |
| | (3,706.071) |
| Observations | 96 |
| Log Likelihood | −856.146 |
| Akaike Inf. Crit. | 1,722.291 |
| Bayesian Inf. Crit. | 1,734.954 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |