Improving Day Ahead Electricity Load Forecasts with Google Trends

Cameron Mulder
James Woods

July 29, 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%)

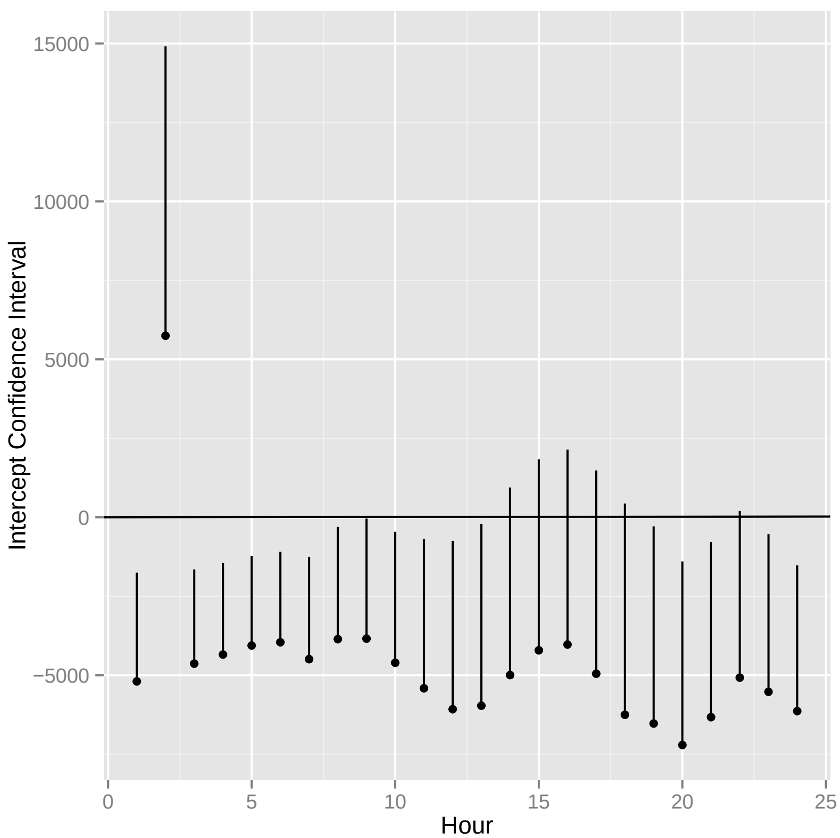


Figure 2: Confidence Intervals for Co-Movement Statistically Adjusted Models (95%)

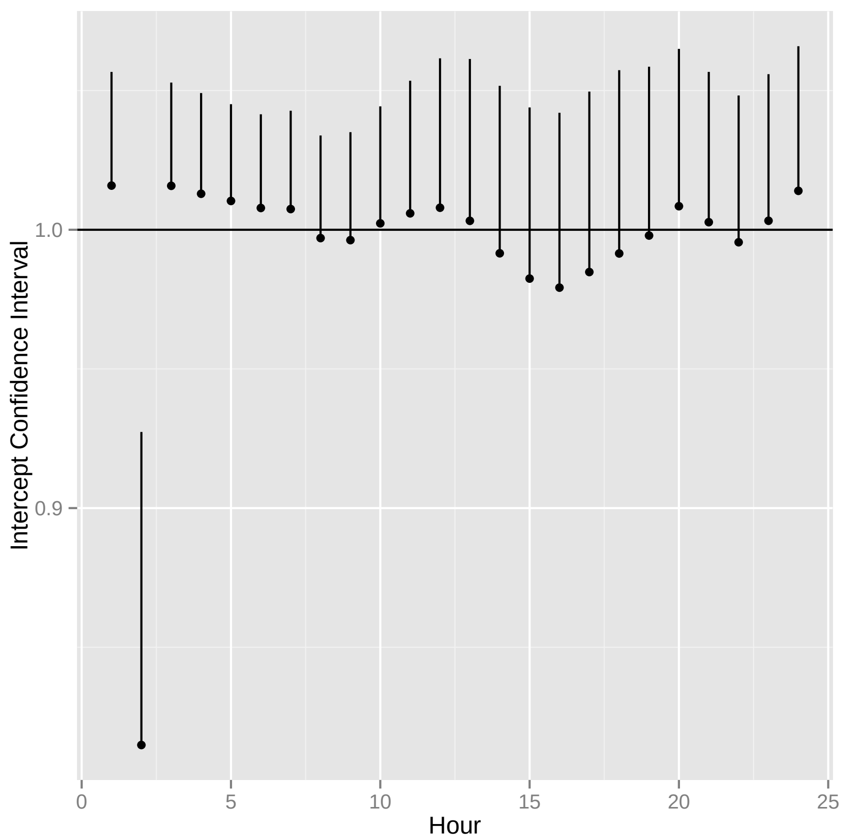


Figure 3: State Weather Trends Indexes Over Time

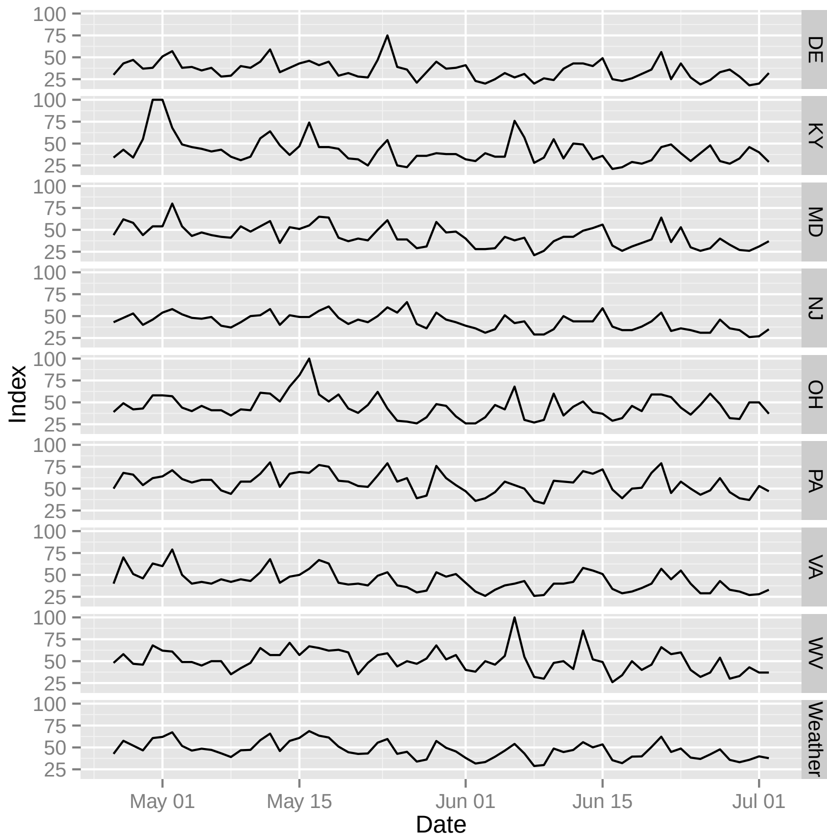
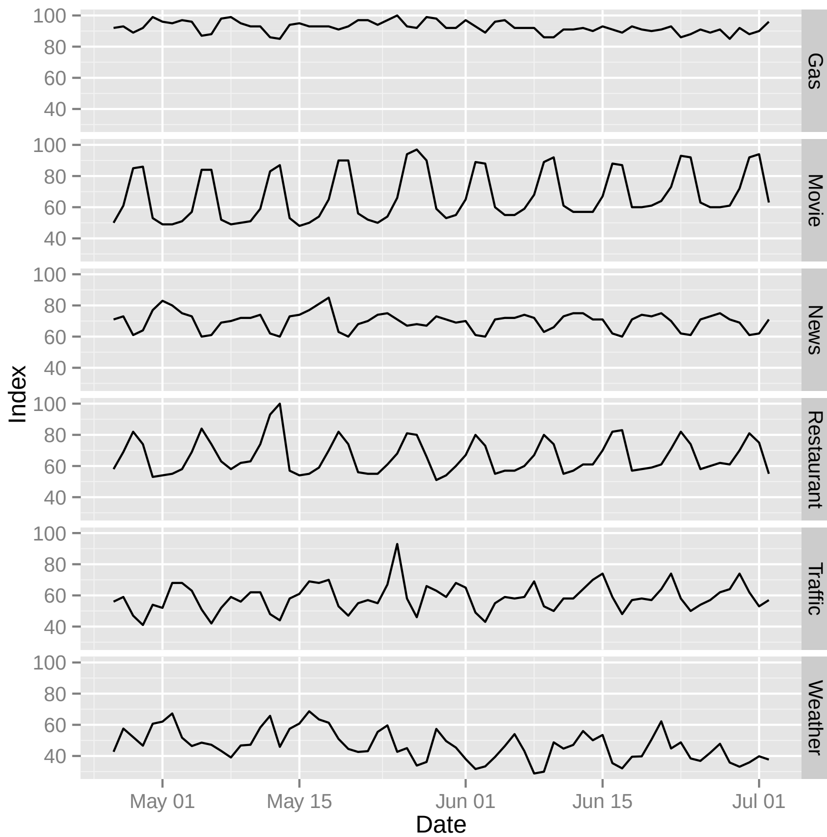


Figure 4: Trends Indexes Over Time



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	3.914	4.091	4.432
2	30.473	3.615	3.674
3	50.565	3.628	3.241
4	60.402	3.138	2.868
5	66.381	3.049	2.694
6	73.314	2.382	2.637
7	79.050	2.627	3.028
8	82.113	5.250	4.329
9	78.317	9.197	8.187
10	72.175	9.969	8.396
11	67.881	9.630	8.102
12	67.577	9.133	7.900
13	68.331	8.662	7.696
14	70.287	8.362	7.476
15	71.514	8.199	7.320
16	71.155	7.934	7.432
17	70.310	7.292	7.089
18	68.395	6.504	6.558
19	66.234	6.252	6.490
20	63.033	5.638	5.955
21	61.587	4.634	4.978
22	61.377	5.712	6.078
23	55.833	5.727	6.103
24	50.531	5.480	5.869

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%)

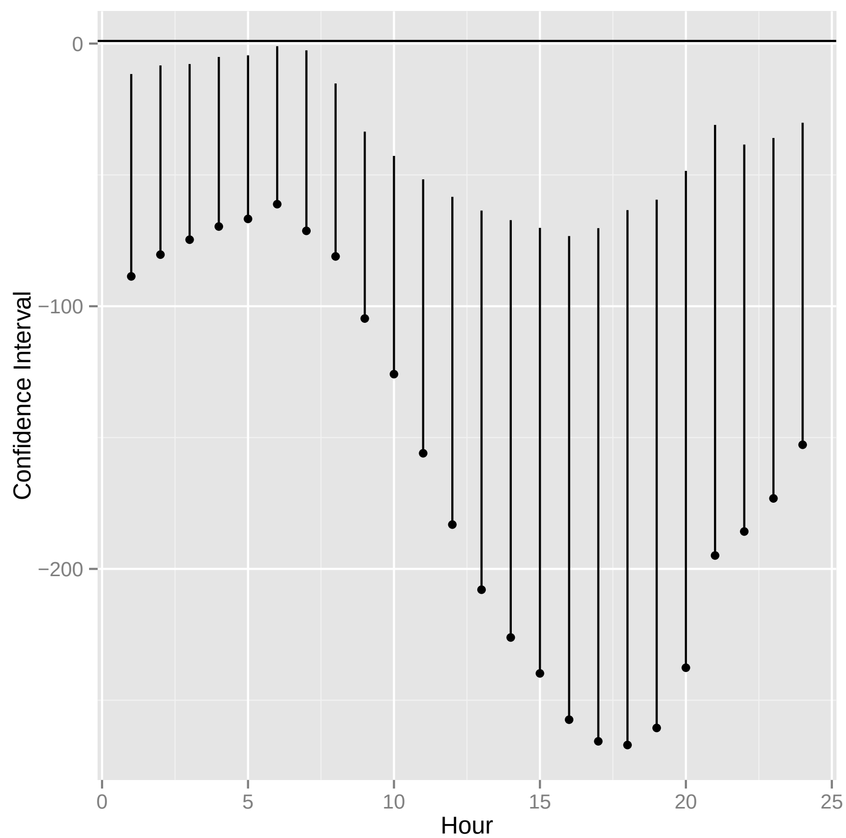


Table 2: Alternate Google Search Models for Hour 19

	Hour 19 Load				
	News	Gas	Traffic	Restaurant	Movie
	(1)	(2)	(3)	(4)	(5)
F19	0.945***	0.976***	0.969***	0.964***	0.941***
	(0.027)	(0.029)	(0.027)	(0.026)	(0.027)
NewsTrends	−206.488***				
	(77.727)				
GasTrends		43.092			
		(134.847)			
TrafficTrends			−25.655		
			(51.947)		
RestaurantTrends				66.565	
				(41.059)	
MovieTrends					87.567***
					(29.033)
Constant	20,036.530***	−1,433.478	4,752.146	−725.466	121.511
	(6,921.332)	(13,902.900)	(4,555.857)	(3,452.763)	(2,667.109)
Observations	68	68	68	68	68
R^2	0.959	0.954	0.954	0.956	0.960
Adjusted R^2	0.957	0.953	0.953	0.955	0.959
Residual Std. Error (df = 65)	3,511.070	3,693.867	3,689.852	3,624.223	3,462.409
F Statistic (df = 2; 65)	753.706***	677.818***	679.365***	705.380***	775.960***

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	0.995***
	(0.023)
Weather	−50.116**
	(19.288)
Constant	2,395.820
	(2,275.007)
Observations	68
R^2	0.973
Adjusted R^2	0.972
Residual Std. Error	1,411.073 (df = 65)
F Statistic	1,176.892*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Hour 2

	Dependent variable:
	Hour 2
Forecast	0.998***
	(0.024)
Weather	−44.331**
	(18.033)
Constant	1,796.751
	(2,239.248)
Observations	68
R^2	0.970
Adjusted R^2	0.969
Residual Std. Error	1,325.204 (df = 65)
F Statistic	1,037.589*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Hour 3

	Dependent variable:
	Hour 3
Forecast	1.000***
	(0.025)
Weather	−41.223**
	(16.742)
Constant	1,442.366
	(2,187.605)
Observations	68
R^2	0.967
Adjusted R^2	0.966
Residual Std. Error	1,237.629 (df = 65)
F Statistic	964.034*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6: Hour 4

	Dependent variable:
	Hour 4
Forecast	1.008***
	(0.026)
Weather	−37.369**
	(16.163)
Constant	684.185
	(2,206.050)
Observations	68
R^2	0.964
Adjusted R^2	0.963
Residual Std. Error	1,202.892 (df = 65)
F Statistic	880.093*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: Hour 5

	Dependent variable:
	Hour 5
Forecast	1.004***
	(0.026)
Weather	−35.611**
	(15.592)
Constant	850.957
	(2,156.112)
Observations	68
R^2	0.964
Adjusted R^2	0.963
Residual Std. Error	1,171.488 (df = 65)
F Statistic	879.703*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8: Hour 6

	Dependent variable:
	Hour 6
Forecast	1.001***
	(0.023)
Weather	−31.074**
	(15.055)
Constant	685.559
	(1,950.786)
Observations	68
R^2	0.971
Adjusted R^2	0.970
Residual Std. Error	1,152.819 (df = 65)
F Statistic	1,097.695*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9: Hour 7

	Dependent variable:
	Hour 7
Forecast	1.002***
	(0.020)
Weather	−36.938**
	(17.205)
Constant	539.285
	(1,879.510)
Observations	68
R^2	0.976
Adjusted R^2	0.976
Residual Std. Error	1,347.560 (df = 65)
F Statistic	1,341.664*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 10: Hour 8

	Dependent variable:
	Hour 8
Forecast	0.999***
	(0.016)
Weather	−48.111***
	(16.485)
Constant	1,583.648
	(1,686.644)
Observations	68
R^2	0.984
Adjusted R^2	0.984
Residual Std. Error	1,289.389 (df = 65)
F Statistic	2,000.846*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 11: Hour 9

	Dependent variable:
	Hour 9
Forecast	0.995***
	(0.017)
Weather	−69.099***
	(17.813)
Constant	3,332.168*
	(1,878.644)
Observations	68
R^2	0.983
Adjusted R^2	0.983
Residual Std. Error	1,370.582 (df = 65)
F Statistic	1,923.005*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 12: Hour 10

	Dependent variable:
	Hour 10
Forecast	0.990***
	(0.019)
Weather	−84.308***
	(20.802)
Constant	4,630.398**
	(2,204.278)
Observations	68
R^2	0.981
Adjusted R^2	0.981
Residual Std. Error	1,563.951 (df = 65)
F Statistic	1,701.187*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 13: Hour 11

	Dependent variable:
	Hour 11
Forecast	0.979***
	(0.021)
Weather	−103.833***
	(26.109)
Constant	6,646.199**
	(2,662.007)
Observations	68
R^2	0.978
Adjusted R^2	0.977
Residual Std. Error	1,921.806 (df = 65)
F Statistic	1,417.020*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 14: Hour 12

	Dependent variable:
	Hour 12
Forecast	0.966***
	(0.022)
Weather	−120.734***
	(31.242)
Constant	8,769.619***
	(3,045.475)
Observations	68
R^2	0.975
Adjusted R^2	0.974
Residual Std. Error	2,259.025 (df = 65)
F Statistic	1,273.271*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 15: Hour 13

	Dependent variable:
	Hour 13
Forecast	0.955***
	(0.023)
Weather	−135.751***
	(36.138)
Constant	10,629.560***
	(3,395.413)
Observations	68
R^2	0.973
Adjusted R^2	0.972
Residual Std. Error	2,574.630 (df = 65)
F Statistic	1,169.426*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 16: Hour 14

	Dependent variable:
	Hour 14
Forecast	0.940***
	(0.023)
Weather	−146.668***
	(39.779)
Constant	12,701.400***
	(3,609.781)
Observations	68
R^2	0.972
Adjusted R^2	0.971
Residual Std. Error	2,806.437 (df = 65)
F Statistic	1,138.708*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 17: Hour 15

	Dependent variable:
	Hour 15
Forecast	0.930***
	(0.024)
Weather	−154.972***
	(42.468)
Constant	14,170.360***
	(3,766.876)
Observations	68
R^2	0.972
Adjusted R^2	0.971
Residual Std. Error	2,972.112 (df = 65)
F Statistic	1,121.644*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 18: Hour 16

	Dependent variable:
	Hour 16
Forecast	0.923***
	(0.025)
Weather	−165.343***
	(46.098)
Constant	15,443.080***
	(4,043.314)
Observations	68
R^2	0.969
Adjusted R^2	0.969
Residual Std. Error	3,199.464 (df = 65)
F Statistic	1,031.568*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 19: Hour 17

	Dependent variable:
	Hour 17
Forecast	0.922***
	(0.026)
Weather	−167.959***
	(48.912)
Constant	15,904.810***
	(4,292.102)
Observations	68
R^2	0.967
Adjusted R^2	0.966
Residual Std. Error	3,371.576 (df = 65)
F Statistic	949.811*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 20: Hour 18

	Dependent variable:
	Hour 18
Forecast	0.923***
	(0.028)
Weather	−165.225***
	(50.984)
Constant	15,733.580***
	(4,537.302)
Observations	68
R^2	0.963
Adjusted R^2	0.962
Residual Std. Error	3,492.381 (df = 65)
F Statistic	851.267*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 21: Hour 19

	Dependent variable:
	Hour 19
Forecast	0.924***
	(0.029)
Weather	−159.993***
	(50.351)
Constant	15,212.480***
	(4,595.023)
Observations	68
R^2	0.960
Adjusted R^2	0.959
Residual Std. Error	3,439.288 (df = 65)
F Statistic	786.867*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 22: Hour 20

	Dependent variable:
	Hour 20
Forecast	0.936***
	(0.030)
Weather	−143.041***
	(47.359)
Constant	12,543.430***
	(4,550.031)
Observations	68
R^2	0.958
Adjusted R^2	0.957
Residual Std. Error	3,233.020 (df = 65)
F Statistic	739.152*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 23: Hour 21

	Dependent variable:
	Hour 21
Forecast	0.933***
	(0.030)
Weather	−112.929***
	(41.044)
Constant	10,878.270**
	(4,273.241)
Observations	68
R^2	0.958
Adjusted R^2	0.957
Residual Std. Error	2,796.714 (df = 65)
F Statistic	742.051*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 24: Hour 22

	Dependent variable:
	Hour 22
Forecast	0.925***
	(0.029)
Weather	−112.108***
	(36.882)
Constant	12,067.500***
	(3,957.362)
Observations	68
R^2	0.961
Adjusted R^2	0.960
Residual Std. Error	2,509.120 (df = 65)
F Statistic	799.347*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 25: Hour 23

	Dependent variable:
	Hour 23
Forecast	0.935***
	(0.030)
Weather	−104.553***
	(34.354)
Constant	10,486.460***
	(3,782.150)
Observations	68
R^2	0.959
Adjusted R^2	0.957
Residual Std. Error	2,347.855 (df = 65)
F Statistic	751.187*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 26: Hour 24

	Dependent variable:
	Hour 24
Forecast	0.942***
	(0.031)
Weather	−91.438***
	(30.694)
Constant	8,585.727**
	(3,508.916)
Observations	68
R^2	0.956
Adjusted R^2	0.955
Residual Std. Error	2,107.796 (df = 65)
F Statistic	708.481*** (df = 2; 65)
Note:	*p<0.1; **p<0.05; ***p<0.01