Improving Day Ahead Electricity Load Forecasts with Google Trends

Cameron Mulder
James Woods

September 16, 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%)

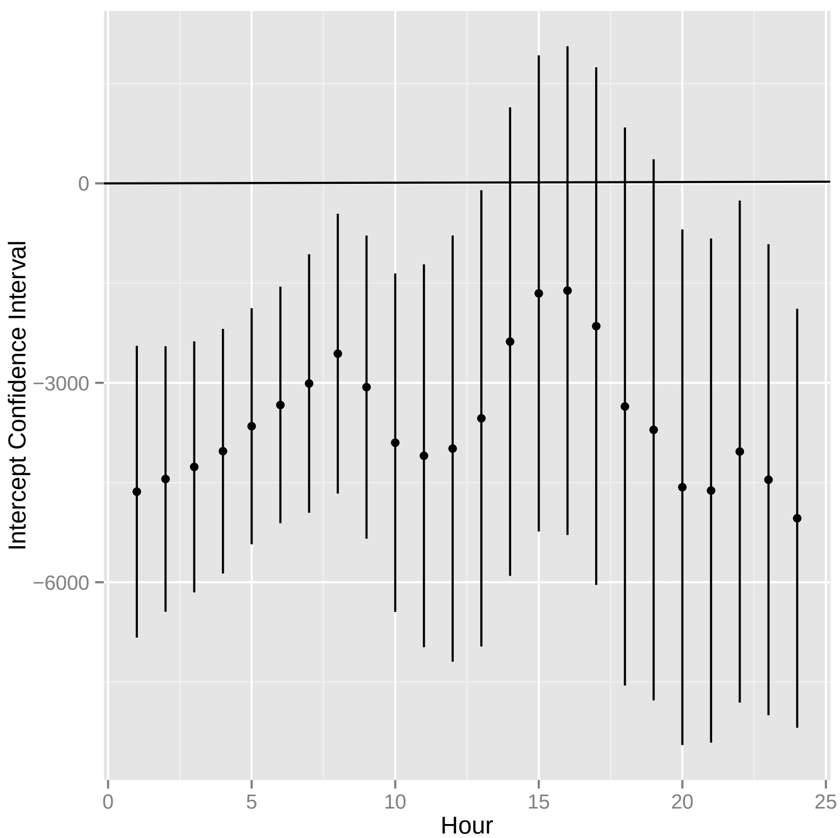


Figure 2: Confidence Intervals for Co-Movement Statistically Adjusted Models (95%)

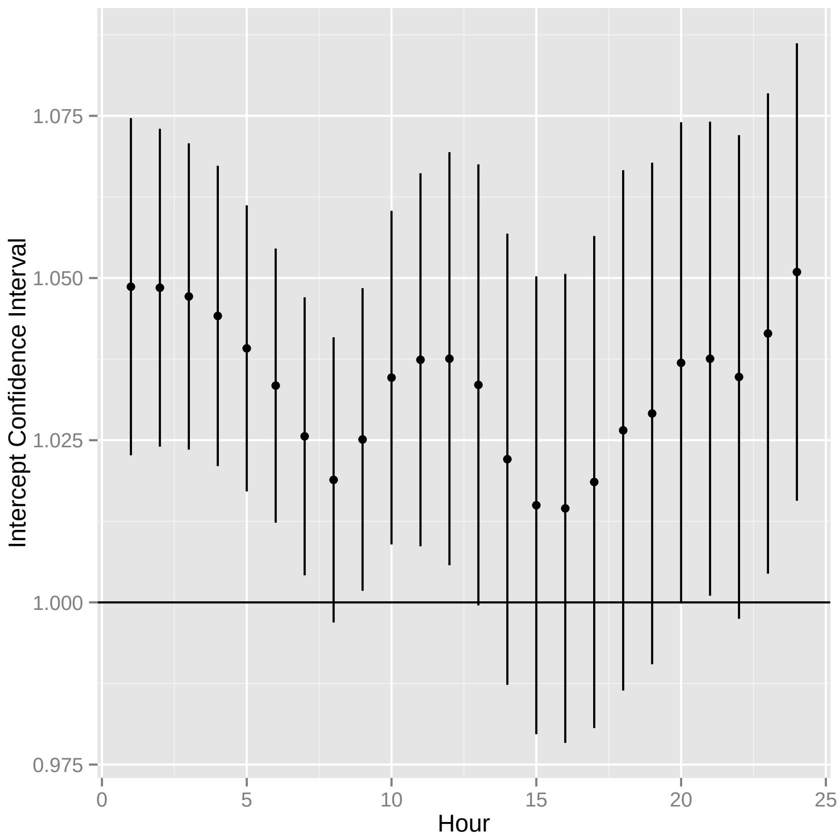


Figure 3: State Weather Trends Indexes Over Time

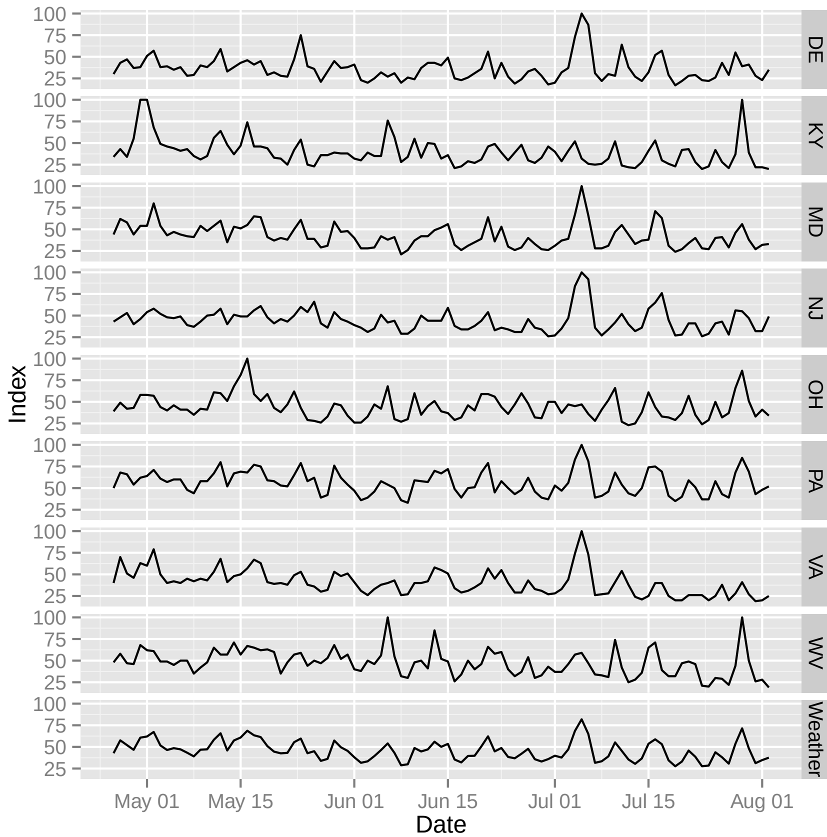
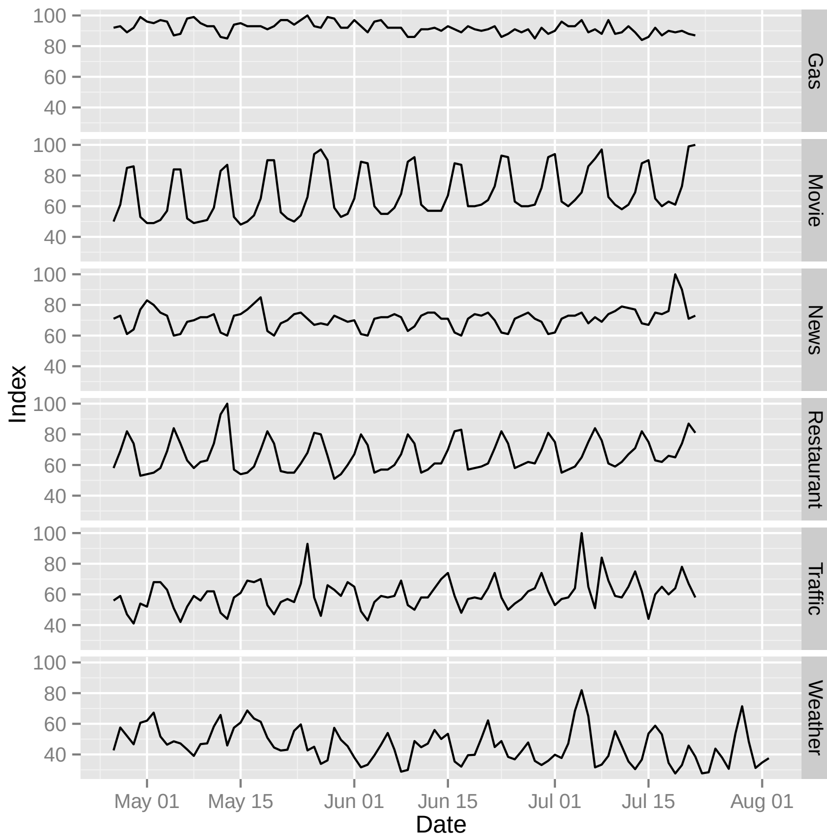


Figure 4: Trends Indexes Over Time



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		7.197	5.030
2		7.526	5.126
3		7.216	5.632
4		6.889	5.700
5		7.010	6.170
6		8.089	7.356
7		8.884	5.393
8		10.703	8.017
9		12.163	12.579
10		11.336	14.496
11		10.298	16.095
12		8.812	15.934
13		8.109	14.970
14		7.490	13.767
15		6.812	12.654
16		6.460	11.861
17		5.799	9.865
18		4.740	7.974
19		4.296	7.444
20		4.357	5.000
21		5.184	4.475
22		5.759	6.347
23		5.302	6.469
24		4.755	5.817

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%)

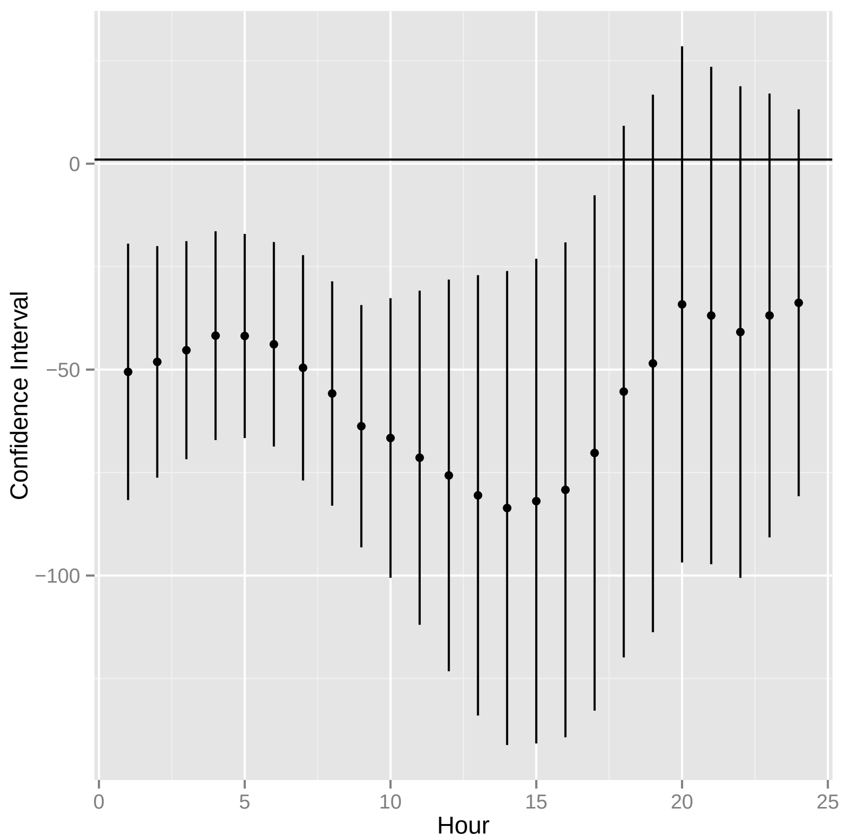


Table 2: Alternate Google Search Models for Hour 19

	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	0.981^{***}
	(0.024)
Weather	-50.544^{***}
	(15.884)
Constant	-60.958^{***}
	(17.049)
Constant	3,347.324
	(2,246.886)
Observations	96
Log Likelihood	−823.822
Akaike Inf. Crit.	1,659.644
Bayesian Inf. Crit.	1,674.774
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Hour 2

	Dependent variable:
	Hour 2
Forecast	0.979^{***}
	(0.024)
Weather	-48.120^{***}
	(14.343)
Constant	-58.330^{***}
	(15.471)
Constant	3,227.314
	(2,135.845)
Observations	96
Log Likelihood	−815.256
Akaike Inf. Crit.	1,642.512
Bayesian Inf. Crit.	1,657.642
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Hour 3

	Dependent variable:
	Hour 3
Forecast	0.980***
	(0.026)
Weather	−45.290***
	(13.511)
Constant	−54.264***
	(14.679)
Constant	2,869.489
	(2,152.385)
Observations	96
Log Likelihood	−809.418
Akaike Inf. Crit.	1,630.837
Bayesian Inf. Crit.	1,645.967
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6: Hour 4

	Dependent variable:
	Hour 4
Forecast	0.979***
	(0.027)
Weather	−41.753***
	(12.932)
Constant	−50.890***
	(14.195)
Constant	2,728.335
	(2,215.921)
Observations	96
Log Likelihood	−804.363
Akaike Inf. Crit.	1,620.725
Bayesian Inf. Crit.	1,635.856
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: Hour 5

	Dependent variable:
	Hour 5
Forecast	0.971***
	(0.029)
Weather	−41.832***
	(12.648)
Constant	−52.027***
	(14.176)
Constant	3,352.412
	(2,295.925)
Observations	96
Log Likelihood	−802.314
Akaike Inf. Crit.	1,616.627
Bayesian Inf. Crit.	1,631.758
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8: Hour 6

	Dependent variable:
	Hour 6
Forecast	0.950***
	(0.029)
Weather	−43.851***
	(12.666)
Constant	−61.109***
	(14.805)
Constant	5,030.628**
	(2,375.723)
Observations	96
Log Likelihood	−803.645
Akaike Inf. Crit.	1,619.290
Bayesian Inf. Crit.	1,634.421
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9: Hour 7

	Dependent variable:
	Hour 7
Forecast	0.931***
	(0.029)
Weather	−49.561***
	(13.959)
Constant	−73.836***
	(16.906)
Constant	6,741.116***
	(2,475.239)
Observations	96
Log Likelihood	−813.212
Akaike Inf. Crit.	1,638.425
Bayesian Inf. Crit.	1,653.556
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 10: Hour 8

	Dependent variable:
	Hour 8
Forecast	0.951***
	(0.025)
Weather	−55.806***
	(13.898)
Constant	−77.101***
	(16.996)
Constant	5,861.865**
	(2,314.052)
Observations	96
Log Likelihood	−814.086
Akaike Inf. Crit.	1,640.172
Bayesian Inf. Crit.	1,655.303
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 11: Hour 9

	Dependent variable:
	Hour 9
Forecast	0.962***
	(0.024)
Weather	−63.738***
	(15.008)
Constant	−84.619***
	(18.111)
Constant	5,737.295**
	(2,432.418)
Observations	96
Log Likelihood	−818.673
Akaike Inf. Crit.	1,649.346
Bayesian Inf. Crit.	1,664.476
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 12: Hour 10

	Dependent variable:
	Hour 10
Forecast	0.966***
	(0.025)
Weather	−66.607***
	(17.322)
Constant	−88.573***
	(20.615)
Constant	5,710.631**
	(2,721.747)
Observations	96
Log Likelihood	−829.613
Akaike Inf. Crit.	1,671.225
Bayesian Inf. Crit.	1,686.356
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 13: Hour 11

	Dependent variable:
	Hour 11
Forecast	0.962***
	(0.027)
Weather	−71.393***
	(20.696)
Constant	−96.951***
	(24.389)
Constant	6,620.051**
	(3,069.225)
Observations	96
Log Likelihood	−844.489
Akaike Inf. Crit.	1,700.979
Bayesian Inf. Crit.	1,716.109
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 14: Hour 12

	Dependent variable:
	Hour 12
Forecast	0.959***
	(0.028)
Weather	−75.690***
	(24.261)
Constant	−102.910***
	(28.373)
Constant	7,317.758**
	(3,407.924)
Observations	96
Log Likelihood	−857.897
Akaike Inf. Crit.	1,727.794
Bayesian Inf. Crit.	1,742.925
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 15: Hour 13

	Dependent variable:
	Hour 13
Forecast	0.957***
	(0.029)
Weather	−80.531***
	(27.278)
Constant	−108.107***
	(31.700)
Constant	7,842.156**
	(3,665.331)
Observations	96
Log Likelihood	−867.788
Akaike Inf. Crit.	1,747.577
Bayesian Inf. Crit.	1,762.707
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 16: Hour 14

	Dependent variable:
	Hour 14
Forecast	0.952***
	(0.029)
Weather	−83.611***
	(29.358)
Constant	−109.419***
	(34.060)
Constant	8,541.788**
	(3,834.233)
Observations	96
Log Likelihood	−874.004
Akaike Inf. Crit.	1,760.009
Bayesian Inf. Crit.	1,775.139
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 17: Hour 15

	Dependent variable:
	Hour 15
Forecast	0.956***
	(0.030)
Weather	−81.933***
	(30.024)
Constant	−102.753***
	(34.782)
Constant	8,032.069**
	(3,926.993)
Observations	96
Log Likelihood	−875.742
Akaike Inf. Crit.	1,763.484
Bayesian Inf. Crit.	1,778.615
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 18: Hour 16

	Dependent variable:
	Hour 16
Forecast	0.963***
	(0.030)
Weather	−79.197***
	(30.656)
Constant	−96.089***
	(35.378)
Constant	7,183.384*
	(4,058.602)
Observations	96
Log Likelihood	−877.680
Akaike Inf. Crit.	1,767.359
Bayesian Inf. Crit.	1,782.490
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 19: Hour 17

	Dependent variable:
	Hour 17
Forecast	0.970***
	(0.032)
Weather	−70.240**
	(31.921)
Constant	−83.057**
	(36.654)
Constant	6,100.950
	(4,281.994)
Observations	96
Log Likelihood	−881.462
Akaike Inf. Crit.	1,774.923
Bayesian Inf. Crit.	1,790.054
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 20: Hour 18

	Dependent variable:
	Hour 18
Forecast	0.981***
	(0.034)
Weather	−55.342*
	(32.930)
Constant	−62.557*
	(37.655)
Constant	4,285.160
	(4,550.374)
Observations	96
Log Likelihood	−884.298
Akaike Inf. Crit.	1,780.595
Bayesian Inf. Crit.	1,795.726
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 21: Hour 19

	Dependent variable:
	Hour 19
Forecast	0.987***
	(0.036)
Weather	−48.499
	(33.299)
Constant	−51.111
	(37.862)
Constant	3,169.498
	(4,701.715)
Observations	96
Log Likelihood	−885.001
Akaike Inf. Crit.	1,782.002
Bayesian Inf. Crit.	1,797.132
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 22: Hour 20

	Dependent variable:
	Hour 20
Forecast	1.007***
	(0.038)
Weather	−34.157
	(31.980)
Constant	−31.842
	(36.363)
Constant	−65.644
	(4,766.154)
Observations	96
Log Likelihood	−880.816
Akaike Inf. Crit.	1,773.632
Bayesian Inf. Crit.	1,788.763
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 23: Hour 21

	Dependent variable:
	Hour 21
Forecast	1.008***
	(0.039)
Weather	−36.868
	(30.814)
Constant	−30.617
	(35.065)
Constant	−482.324
	(4,781.229)
Observations	96
Log Likelihood	−876.330
Akaike Inf. Crit.	1,764.659
Bayesian Inf. Crit.	1,779.790
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 24: Hour 22

	Dependent variable:
	Hour 22
Forecast	0.997***
	(0.041)
Weather	−40.879
	(30.456)
Constant	−34.081
	(34.378)
Constant	1,229.355
	(4,835.193)
Observations	96
Log Likelihood	−875.277
Akaike Inf. Crit.	1,762.554
Bayesian Inf. Crit.	1,777.685
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 25: Hour 23

	Dependent variable:
	Hour 23
Forecast	0.996***
	(0.041)
Weather	−36.853
	(27.491)
Constant	−30.764
	(30.700)
Constant	1,149.213
	(4,505.406)
Observations	96
Log Likelihood	−865.757
Akaike Inf. Crit.	1,743.513
Bayesian Inf. Crit.	1,758.644
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 26: Hour 24

	Dependent variable:
	Hour 24
Forecast	0.996***
	(0.040)
Weather	−33.778
	(23.967)
Constant	−29.496
	(26.483)
Constant	921.069
	(4,010.567)
Observations	96
Log Likelihood	−852.807
Akaike Inf. Crit.	1,717.615
Bayesian Inf. Crit.	1,732.745
Note:	*p<0.1; **p<0.05; ***p<0.01