Improving Day Ahead Electricity Load Forecasts with Google Trends

Cameron Mulder
James Woods

July 30, 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%)

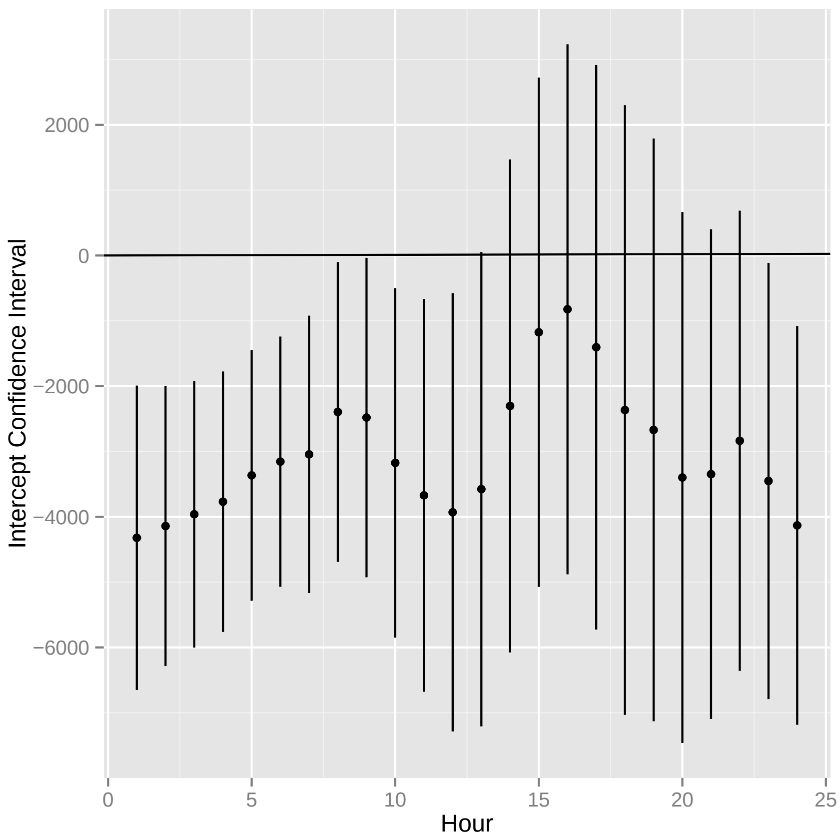


Figure 2: Confidence Intervals for Co-Movement Statistically Adjusted Models (95%)

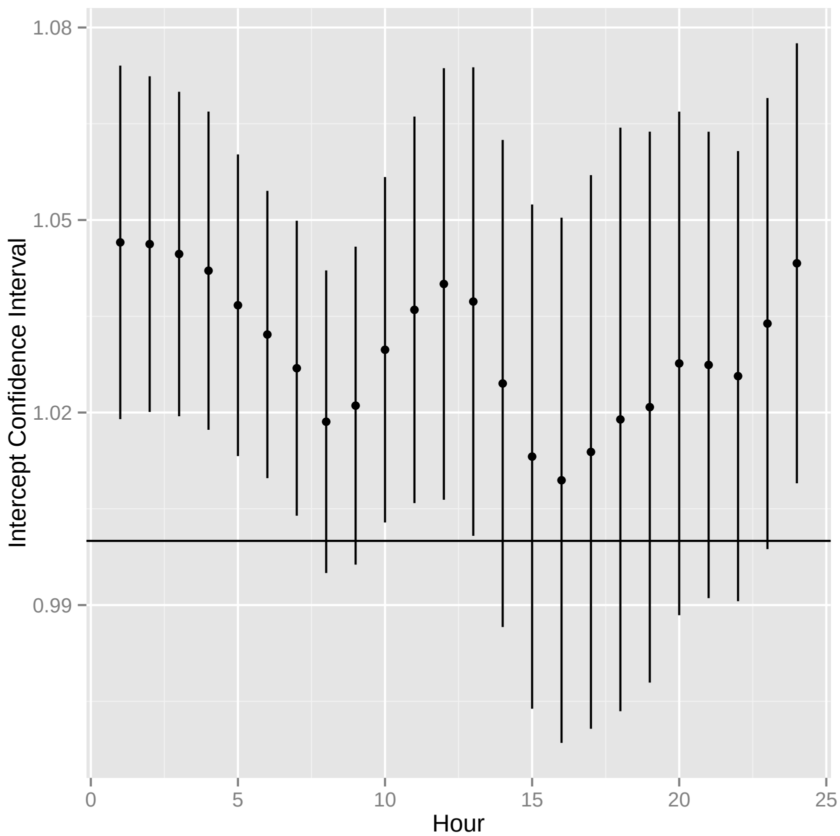


Figure 3: State Weather Trends Indexes Over Time

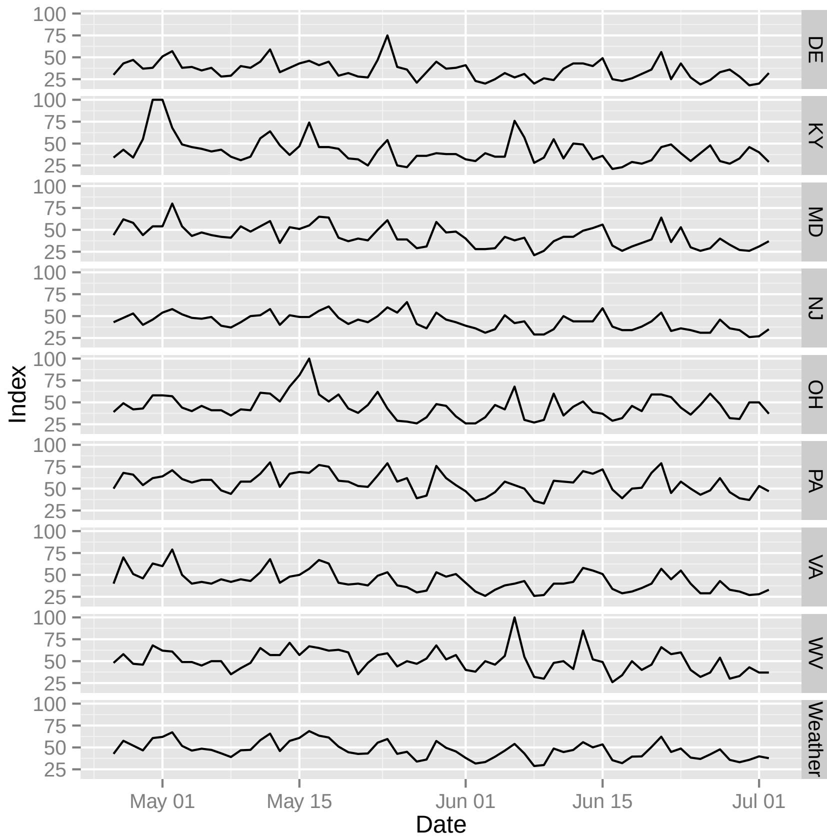
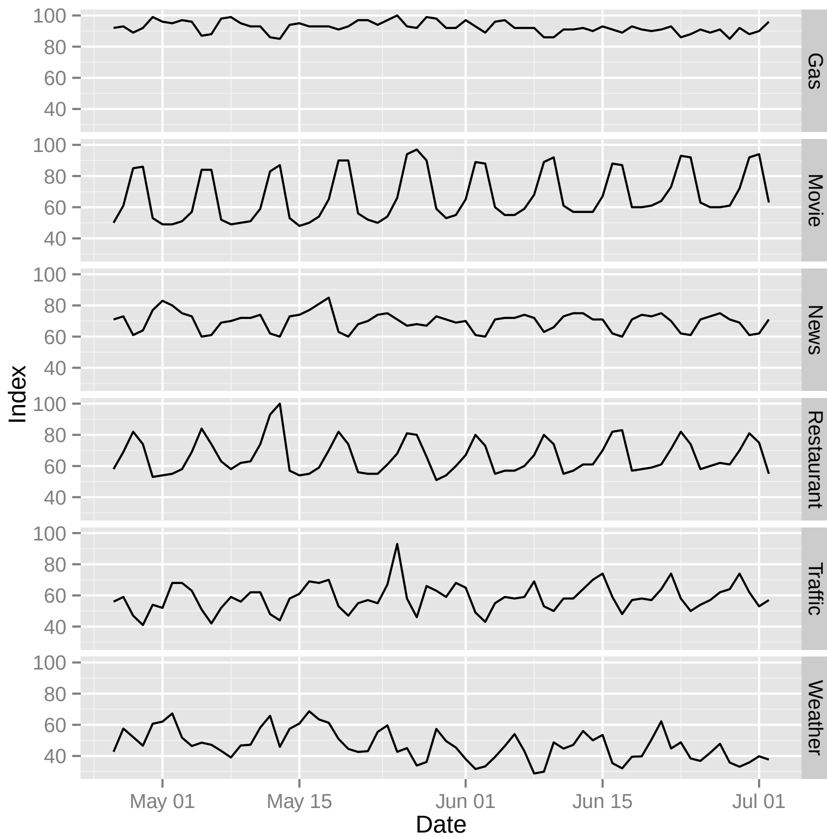


Figure 4: Trends Indexes Over Time



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	4.561
2	30.473	3.615	4.467
3	50.565	3.628	4.779
4	60.402	3.138	4.444
5	66.381	3.049	4.089
6	73.314	2.382	4.075
7	79.050	2.627	4.632
8	82.113	5.250	6.716
9	78.317	9.197	10.984
10	72.175	9.969	10.989
11	67.881	9.630	9.518
12	67.577	9.133	7.772
13	68.331	8.662	6.620
14	70.287	8.362	6.088
15	71.514	8.199	5.456
16	71.155	7.934	5.313
17	70.310	7.292	5.068
18	68.395	6.504	4.612
19	66.234	6.252	4.594
20	63.033	5.638	2.361
21	61.587	4.634	1.415
22	61.377	5.712	3.784
23	55.833	5.727	3.730
24	50.531	5.480	3.274

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%)

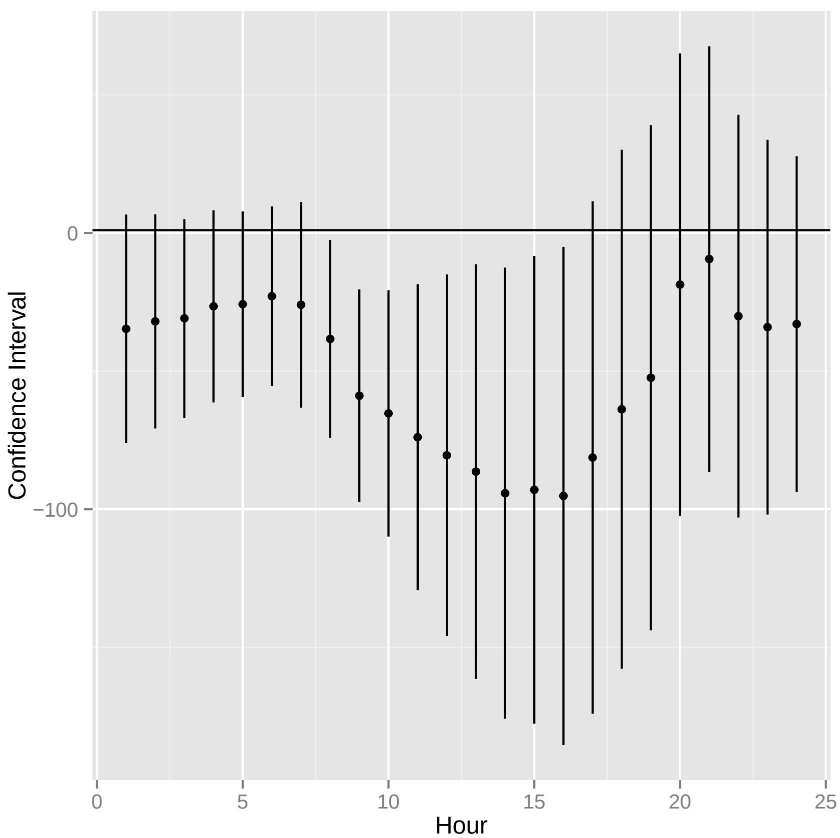


Table 2: Alternate Google Search Models for Hour 19

	Hour 19 Load				
	News	Gas	Traffic	Restaurant	Movie
	(1)	(2)	(3)	(4)	(5)
F19	0.942***	0.971***	0.952***	0.956***	0.940***
	(0.039)	(0.038)	(0.041)	(0.037)	(0.038)
NewsTrends	−165.209**				
	(69.522)				
GasTrends		−97.010			
		(106.696)			
TrafficTrends			−69.267		
			(44.882)		
RestaurantTrends				90.097**	
				(35.645)	
MovieTrends					71.976***
					(26.775)
Constant	17,443.060**	11,951.160	8,913.360	−1,400.642	1,282.578
	(7,432.784)	(11,632.000)	(5,896.900)	(3,912.481)	(3,557.924)
Observations	68	68	68	68	68
Log Likelihood	−624.411	−626.318	−626.431	−624.767	−624.639
Akaike Inf. Crit.	1,258.821	1,262.637	1,262.863	1,259.535	1,259.278
Bayesian Inf. Crit.	1,269.693	1,273.509	1,273.735	1,270.406	1,270.150
Note:					*p<0.1; **p<0.05; ***p<0.01

4 Summary and Conclusions

A Hourly Models with Weather Searches