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%)

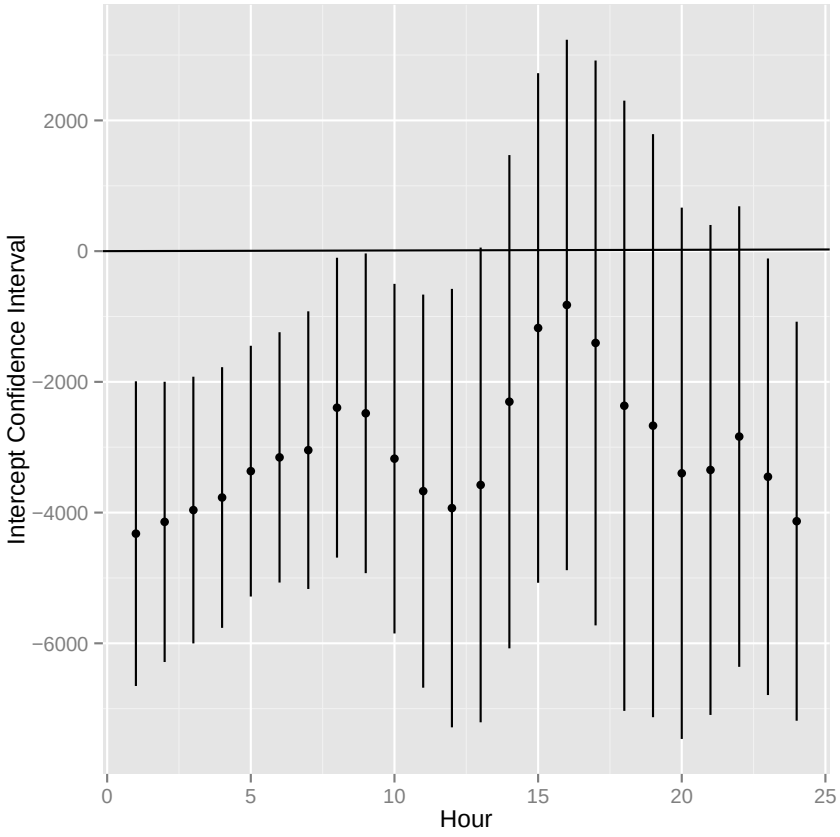


Figure 2: Confidence Intervals for Co-Movement Statistically Adjusted Models (95%)

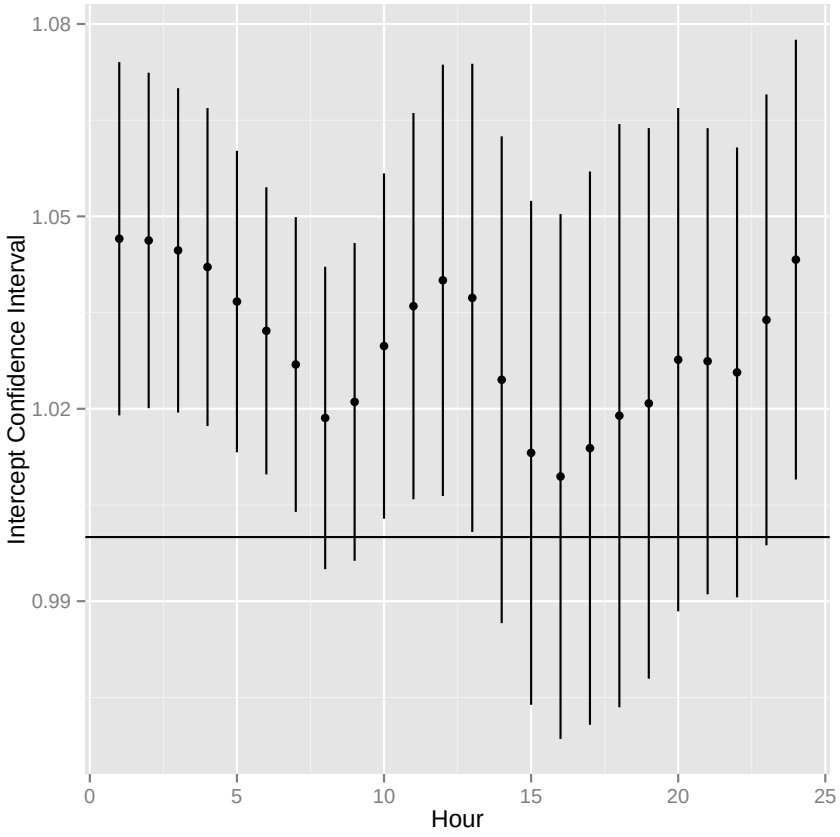


Figure 3: State Weather Trends Indexes Over Time

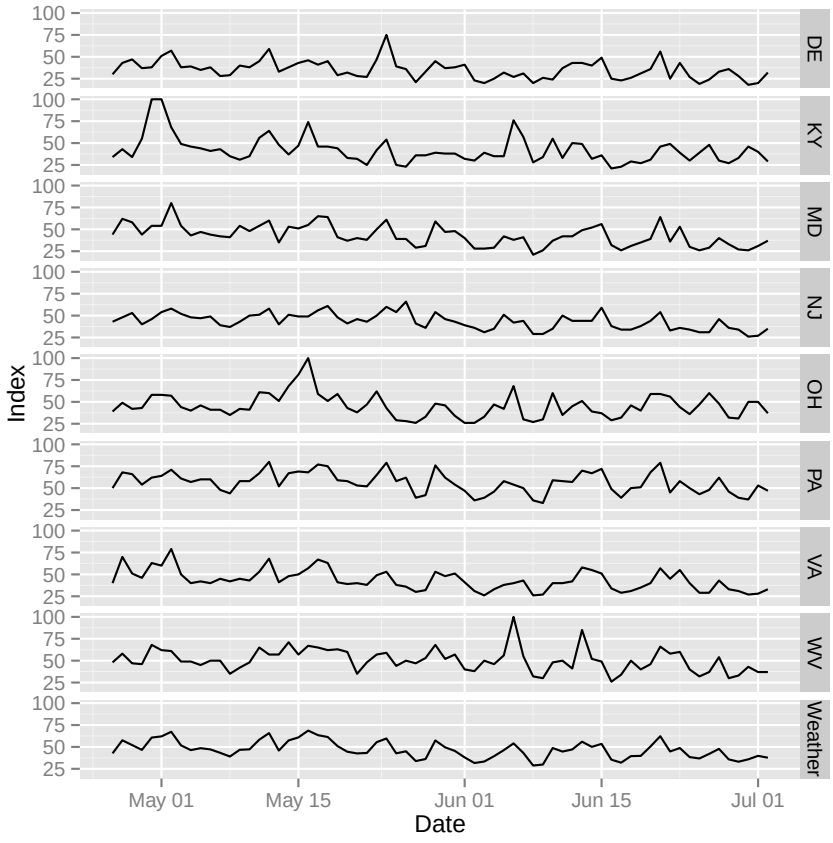
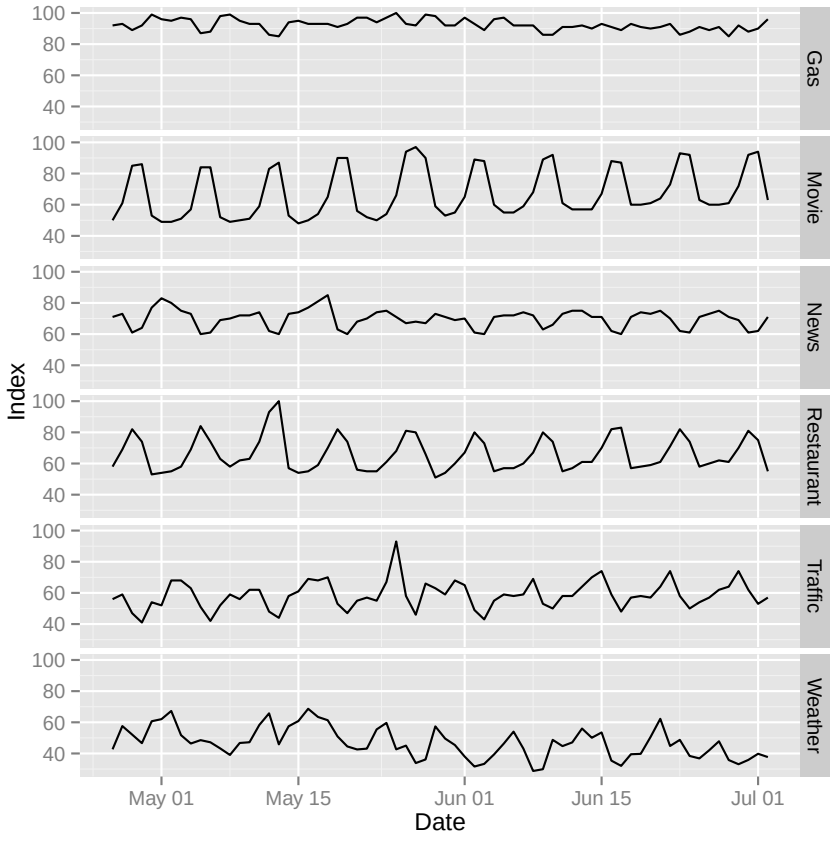


Figure 4: Trends Indexes Over Time



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%)

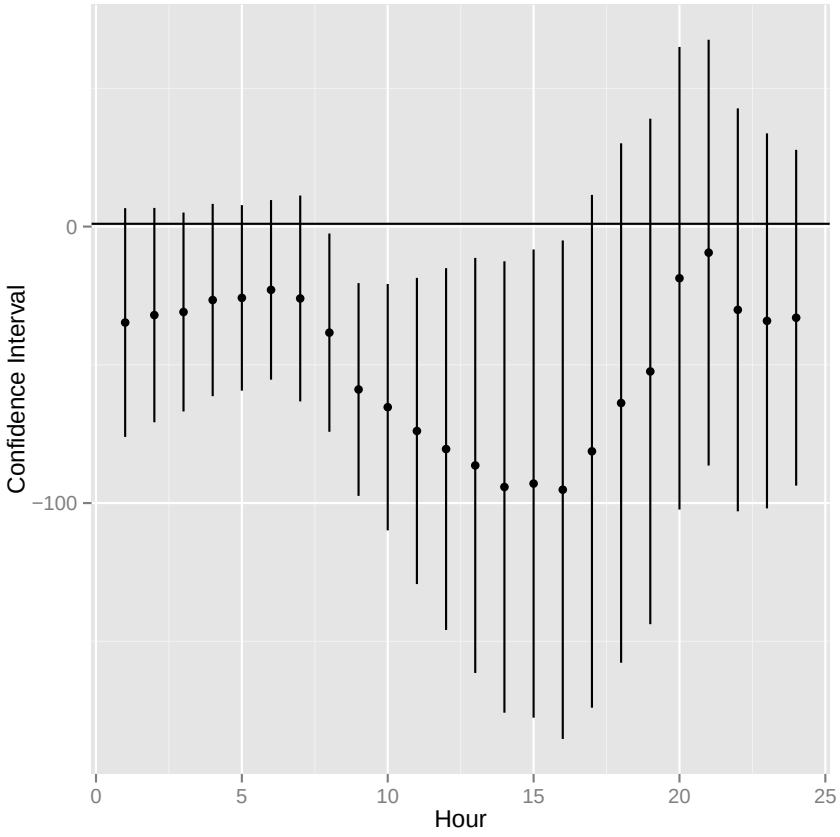


Table 2: Alternate Google Search Models for Hour 19

	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

Table 3: Hour 1

	Dependent variable:
	Hour 1
Forecast	1.003***
	(0.029)
Weather	−34.707
	(21.106)
Constant	1,054.591
	(2,721.924)
Observations	68
Log Likelihood	−577.197
Akaike Inf. Crit.	1,164.394
Bayesian Inf. Crit.	1,175.266
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Hour 2

	Dependent variable:
	Hour 2
Forecast	1.005***
	(0.031)
Weather	−32.012
	(19.775)
Constant	736.863
	(2,672.072)
Observations	68
Log Likelihood	−573.192
Akaike Inf. Crit.	1,156.385
Bayesian Inf. Crit.	1,167.256
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Hour 3

	Dependent variable:
	Hour 3
Forecast	1.007***
	(0.032)
Weather	−30.901*
	(18.352)
Constant	457.961
	(2,616.603)
Observations	68
Log Likelihood	−568.311
Akaike Inf. Crit.	1,146.622
Bayesian Inf. Crit.	1,157.494
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6: Hour 4

	Dependent variable:
	Hour 4
Forecast	1.014***
	(0.033)
Weather	−26.567
	(17.742)
Constant	−222.680
	(2,630.837)
Observations	68
Log Likelihood	−566.286
Akaike Inf. Crit.	1,142.573
Bayesian Inf. Crit.	1,153.445
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: Hour 5

	Dependent variable:
	Hour 5
Forecast	1.009***
	(0.032)
Weather	−25.792
	(17.120)
Constant	37.219
	(2,528.640)
Observations	68
Log Likelihood	−564.436
Akaike Inf. Crit.	1,138.872
Bayesian Inf. Crit.	1,149.744
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8: Hour 6

	Dependent variable:
	Hour 6
Forecast	1.009***
	(0.026)
Weather	−22.892
	(16.570)
Constant	−274.663
	(2,186.325)
Observations	68
Log Likelihood	−563.466
Akaike Inf. Crit.	1,136.932
Bayesian Inf. Crit.	1,147.804
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9: Hour 7

	Dependent variable:
	Hour 7
Forecast	1.007***
	(0.022)
Weather	−25.999
	(18.991)
Constant	−364.623
	(2,007.855)
Observations	68
Log Likelihood	−572.923
Akaike Inf. Crit.	1,155.846
Bayesian Inf. Crit.	1,166.718
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 10: Hour 8

	Dependent variable:
	Hour 8
Forecast	1.006***
	(0.018)
Weather	−38.362**
	(18.292)
Constant	593.196
	(1,811.682)
Observations	68
Log Likelihood	−572.484
Akaike Inf. Crit.	1,154.967
Bayesian Inf. Crit.	1,165.839
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 11: Hour 9

	Dependent variable:
	Hour 9
Forecast	1.004***
	(0.019)
Weather	−58.918***
	(19.630)
Constant	2,029.294
	(2,024.054)
Observations	68
Log Likelihood	−576.558
Akaike Inf. Crit.	1,163.116
Bayesian Inf. Crit.	1,173.988
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 12: Hour 10

	Dependent variable:
	Hour 10
Forecast	1.008***
	(0.021)
Weather	−65.320***
	(22.736)
Constant	2,181.328
	(2,400.470)
Observations	68
Log Likelihood	−584.082
Akaike Inf. Crit.	1,178.165
Bayesian Inf. Crit.	1,189.037
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 13: Hour 11

	Dependent variable:
	Hour 11
Forecast	1.006***
	(0.024)
Weather	−73.929***
	(28.252)
Constant	2,750.826
	(2,930.511)
Observations	68
Log Likelihood	−596.979
Akaike Inf. Crit.	1,203.958
Bayesian Inf. Crit.	1,214.830
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 14: Hour 12

	Dependent variable:
	Hour 12
Forecast	0.999***
	(0.027)
Weather	−80.469**
	(33.383)
Constant	3,790.872
	(3,378.817)
Observations	68
Log Likelihood	−606.910
Akaike Inf. Crit.	1,223.820
Bayesian Inf. Crit.	1,234.692
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 15: Hour 13

	Dependent variable:
	Hour 13
Forecast	0.991***
	(0.028)
Weather	−86.387**
	(38.294)
Constant	4,906.294
	(3,789.066)
Observations	68
Log Likelihood	−615.224
Akaike Inf. Crit.	1,240.448
Bayesian Inf. Crit.	1,251.320
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 16: Hour 14

	Dependent variable:
	Hour 14
Forecast	0.974***
	(0.029)
Weather	−94.184**
	(41.650)
Constant	6,921.458*
	(4,006.585)
Observations	68
Log Likelihood	−620.360
Akaike Inf. Crit.	1,250.721
Bayesian Inf. Crit.	1,261.592
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 17: Hour 15

	Dependent variable:
	Hour 15
Forecast	0.965***
	(0.030)
Weather	−92.953**
	(43.196)
Constant	7,807.780*
	(4,133.322)
Observations	68
Log Likelihood	−622.407
Akaike Inf. Crit.	1,254.815
Bayesian Inf. Crit.	1,265.687
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 18: Hour 16

	Dependent variable:
	Hour 16
Forecast	0.959***
	(0.031)
Weather	−95.169**
	(45.992)
Constant	8,560.330*
	(4,416.659)
Observations	68
Log Likelihood	−626.273
Akaike Inf. Crit.	1,262.546
Bayesian Inf. Crit.	1,273.418
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 19: Hour 17

	Dependent variable:
	Hour 17
Forecast	0.964***
	(0.033)
Weather	−81.275*
	(47.315)
Constant	7,610.087
	(4,654.097)
Observations	68
Log Likelihood	−628.052
Akaike Inf. Crit.	1,266.103
Bayesian Inf. Crit.	1,276.975
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 20: Hour 18

	Dependent variable:
	Hour 18
Forecast	0.967***
	(0.036)
Weather	−63.823
	(47.927)
Constant	6,525.038
	(4,893.799)
Observations	68
Log Likelihood	−628.822
Akaike Inf. Crit.	1,267.645
Bayesian Inf. Crit.	1,278.517
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 21: Hour 19

	Dependent variable:
	Hour 19
Forecast	0.967***
	(0.037)
Weather	−52.398
	(46.640)
Constant	5,840.097
	(4,967.528)
Observations	68
Log Likelihood	−626.968
Akaike Inf. Crit.	1,263.935
Bayesian Inf. Crit.	1,274.807
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 22: Hour 20

	Dependent variable:
	Hour 20
Forecast	0.982***
	(0.039)
Weather	−18.689
	(42.683)
Constant	2,166.155
	(4,875.303)
Observations	68
Log Likelihood	−621.311
Akaike Inf. Crit.	1,252.622
Bayesian Inf. Crit.	1,263.494
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 23: Hour 21

	Dependent variable:
	Hour 21
Forecast	0.979***
	(0.038)
Weather	−9.431
	(39.286)
Constant	1,495.714
	(4,698.570)
Observations	68
Log Likelihood	−615.055
Akaike Inf. Crit.	1,240.110
Bayesian Inf. Crit.	1,250.982
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 24: Hour 22

	Dependent variable:
	Hour 22
Forecast	0.968***
	(0.037)
Weather	−30.116
	(37.179)
Constant	4,126.765
	(4,515.168)
Observations	68
Log Likelihood	−611.095
Akaike Inf. Crit.	1,232.190
Bayesian Inf. Crit.	1,243.062
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 25: Hour 23

	Dependent variable:
	Hour 23
Forecast	0.968***
	(0.039)
Weather	−34.104
	(34.604)
Constant	4,255.139
	(4,392.409)
Observations	68
Log Likelihood	−606.498
Akaike Inf. Crit.	1,222.995
Bayesian Inf. Crit.	1,233.867
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 26: Hour 24

	Dependent variable:
	Hour 24
Forecast	0.967***
	(0.041)
Weather	−32.967
	(30.983)
Constant	3,830.651
	(4,141.275)
Observations	68
Log Likelihood	−599.264
Akaike Inf. Crit.	1,208.528
Bayesian Inf. Crit.	1,219.400
Note:	*p<0.1; **p<0.05; ***p<0.01