Energy Strategy Reviews 7 (2014) 9–17

Contents lists available at ScienceDirect

Energy Strategy Reviews

journal homepage: www.ees.elsevier.com/esr

ANALYSIS

Proposed coal power plants and coal-to-liquids plants in the US: Which ones survive and why?

Dean Fantazzini [a,*], Mario Maggi [b]

[a] Moscow School of Economics, Moscow State University, Leninskie Gory, Building 61/1 119992, Russia
[b] Department of Economics and Management, University of Pavia, via S. Felice, 5, Pavia 27040, Italy

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form
18 November 2014
Accepted 26 November 2014
Available online

Keywords:
Coal
Coal plants
Coal-to-liquids
Logit
Probit
Training
Validation
Forecasting
Model confidence set
Google
Google trends
Second great contraction
Global financial crisis

ABSTRACT

The increase of oil and natural gas prices since the year 2000 stimulated the planning and construction of new coal-fired electricity generating plants and coal-to-liquids (CTL) plants in the US. However, many of these projects have been canceled or abandoned since 2007. Using a set of 145 proposed coal power plants and 25 CTL plants, the determinants that influence the decision to abandon a project or to proceed with it are examined using binary data models and 20 regressors. In the case of coal power plants, the number of searches performed on Google relating to coal power plants, the project duration and the prices of alternative fuels for electricity generation are found to be statistically significant at the 5% level. As for CTL plants, the political affiliation of the state governor is the only variable significant at the 5% level across several model specifications. An out-of-sample exercise confirms these findings. These results also hold with robustness checks considering alternative Google search keywords, the potential effects of the recession between 2008 and 2009 and the inclusion of the two dimensions of the Dynamic-Weighted Nominate (DWN) database.

1. Introduction

The first decade of the 21st century witnessed a large increase in oil prices mainly due to the growing demand by China and India, as well as to a growing difficulty to increase oil production worldwide with the notable exception of North America (see Ref. [1], for a recent review). Similarly, US natural gas prices followed a rising trend reaching the level of 13 $/MMBtu in June 2008. The rise in oil and gas prices coincided with increasing power demand in the US. To counter rising fuel cost, coal was a logical choice for power generation, stimulating the planning and/or construction of almost 150 coal-fired electricity generating plants by 2007 [2]. Several coal-to-liquids (CTL hereafter) plants were also proposed (see Ref. [3] for a recent review of hydrocarbon liquefaction as a peak oil mitigation strategy). Since 2007–2008 the energy landscape has changed substantially: the advent of shale gas has reduced considerably the price of natural gas in the US reaching a low of 1.9 $/MMBtu in April 2012. Meanwhile, the construction cost for coal plants has increased considerably but US coal prices remain relatively low (see Refs. [4,5] for recent reviews). Since 2011 the US Environmental Protection Agency (EPA) has began regulating greenhouse gases from mobile and stationary sources of air pollution under the Clean Air Act. There has been an increased awareness of the health risks posed by power plant pollution (as showed by Google data, more below). All this has led to more than 100 coal plants being canceled or abandoned by 2013. This estimate is based on the Sierra Club database [6] and the Coal-Swarm database [7]). The Energy Information Administration (EIA) expects that very few new coal plants will be built through 2040 [8].

Although coal is still the main source for US electricity power production, coal plants are aging. In 2011, the capacity weighted average age of coal-fired plants was 36 years, whereas it was only 18 for natural

* Corresponding author.
E-mail addresses: fantazzini@mse-msu.ru (D. Fantazzini), maggma@eco.unipv.it (M. Maggi).

http://dx.doi.org/10.1016/j.esr.2014.11.005
2211-467X/

Table 1
Capacity weighted distribution of electricity power production plants by fuel. 2011 data from http://www.eia.gov.

Fuel type	Coal	Natural gas	Petroleum
Average size (MW)	245.54	85.65	15.39
Average age (years)	36.34	17.88	35.16
25% built before	1967	1981	1970
50% built before	1974	2001	1972
75% built before	1981	2003	1978
CO_2/capacity (Million Metric Tons/MWh)	0.9931	0.3972	0.8689

gas-fired plants (and 35 for oil-fired plants[1]), see Table 1. Refitting these coal plants to comply with the recent stricter emission standards is very expensive so many of them face retirement in the coming years [5].

Given this background, we analyze the main determinants that influenced the decision to abandon or to proceed with a coal project using a dataset of 145 coal power plants projects and 25 CTL plants, between 2004 and 2013, from the Coal-Swarm database [7], and binary data models.

Prior knowledge of the variables influencing the viability of a coal plant project is fundamental for successful strategy and policy making. To our knowledge, this is the first study that analyzes these variables after the advent of US shale gas and the global economic crisis in 2008–2009. Our findings are not limited to coal plants but also include CTL plants.

A vast body of the literature has found that the public attitude toward the location of environmentally hazardous facilities is a major determinant of siting costs, which can increase quickly when the local community agreement is missing (see Refs. [9,10] for extensive reviews). We use Google search data to measure public attitudes towards coal plants and the associated environmental issues: a tool called Google Trends provides information about users' relative interest for a particular search query in a given geographic region and at given time (the data are available on a weekly or even a daily basis). In recent years, researchers worldwide have started to use online search data for forecasting purposes (see Refs. [11–17] for some recent applications).

The predictive power of our binary data models is then tested by means of an out-of-sample comparison. The models differ along three dimensions: (i) the variables adopted; (ii) the econometric specification; and (iii) the data transformation (either in logs or in levels). A series of robustness checks is also performed to verify that our previous results hold also with alternative data setups: (i) a dataset with alternative keywords for Google search; (ii) time dummies to evaluate the effect on model estimates of the global financial crisis in 2008 and 2009; (iii) a dataset that includes as additional regressors the two dimensions of the Dynamic-Weighted Nominate (DWN) database developed by the political scientists Poole and Rosenthal in the early 1980s to analyze legislative roll-call voting behavior in the US congress, see Refs. [18,19].

The paper is organized as follows. Section 2 describes the data and methods used in our work while the empirical analysis is performed in Section 3. Robustness checks are discussed in Section 4, while Section 5 includes a brief conclusion.

2. Data and methods

2.1. Data

The National Energy Technology Laboratory (NETL), a division of the Department of Energy, maintained a database of all new projects of coal-fired electricity generating plants until May 2007. Since then, the *Coal Issues Portal* on SourceWatch (a project of CoalSwarm and the Center for Media and Democracy) has maintained a dataset of the proposed coal plants in the US and their latest status [7].[2] We separated the variable "status" into two groups: one collecting all plants that are active/upcoming/operating and another group with all plants that were canceled/abandoned or have an uncertain status.[3]

The Coal Issues portal contains some information about the coal projects, like the US state location and, in some cases, also the total capacity (in MW for power plants and bbl/day for CTL), but this information was not sufficient for the scope of our analysis and was augmented by an extensive online search for each coal project. This search was not successful for several plants, for which budget costs, capacity, carbon dioxide (CO_2) emissions, project beginning year and project duration were not available. The initial dataset was filtered and the final dataset consisted of 145 coal power plant projects and 25 CTL plant projects, observed between 2004 and2013.[4] The dataset of coal power plants projects consists of 97 plants that were canceled/abandoned and 48 plants that are active/operating/upcoming for a total of 574 yearly data samples. The dataset of CTL projects consists of 17 plants that were cancelled/abandoned and 8 plants that are active/upcoming for a total of 94 yearly data samples. The (few) projects that were either operative or cancelled before 2004 were omitted since those early projects had very different economics from subsequent ones (see Refs. [1,3,20]).

The literature has identified four main groups of variables that influence the plant location choice. First Coase [21], suggested that site-specific environmental externalities should be the main determinants of location choices: a profit-maximizing firm will try to find an agreement with the community that causes the least damage, all other things being equal. Hamilton (1993) [22], Hamilton (1995) [23] and Jenkins et al. [24] questioned this hypothesis and advanced the idea that local community's public opinion can influence externality costs: communities that show strong opposition are less likely to host a plant or any environmentally hazardous facility. Therefore, a model trying to explain the location of a (coal) plant should consider a group of "voice" indicators. A third group of variables includes traditional industrial location factors like infrastructure, construction and labor costs, see Refs. [10,25–28]. More recently, given the falling prices of renewables and natural gas, several authors have started comparing the economics of these alternative sources of electricity generation with the economics of coal plants to determine the best choice and location, see Refs. [4,5,29–31]. Table 2 illustrates the regressors that we used to explain the status of a coal plant project.

We used the state population in millions and the CO_2 output in tons to measure the external costs a state can suffer given that the larger the population and the CO_2 produced the larger the perception of the expected environmental damage (see Refs. [10,22,32]).[5]

Four indicators were used to represent the awareness of local residents and their ability to pay for environmental quality: the median household income, the labor force participation, the unemployment rate and the Google Index (GI) for the keyword "jobs" (remark that D'Amuri and Marcucci [16] found this GI to be the best predictor for the US unemployment rate). The GI is computed as the ratio of the search queries for a specific keyword (or group of keywords) relative to the total number

[1] The old age of oil-fired plants is also due to the fact that in US oil produces a small and decreasing portion of electricity production.

[2] The NETL database is no longer available but it is included in the CoalSwarm database.

[3] An online search allowed us to find that all plants with an uncertain status were either cancelled or abandoned. They had no related news for years.

[4] The names of these plants are reported in Tables 1–2 in the Technical Appendix accompanying this paper and is posted on the authors' websites.

[5] We tried population density in place of the population data, as done by Garrone and Groppi [10], but this resulted in worse in-sample results, models' residuals and out-of-sample results; we used the population data instead.

of searches performed in the selected region at a given point of time. The result was then standardized between 0 and 100 (where the standardization is done over the whole time period and all searches).

To measure awareness and the ability to organize protests against coal plant projects as well as to ask for compensation (the so-called "voice" factors), we used the GI for the keyword "coal," the GI for the keywords "coal plant + coal power," the GI for the keywords "coal-to-liquids + ctl coal" and the GI for the keyword "pollution." The analysis of Google data showed that several searches for the previous keywords included and/or were related also to "legal action," "protest," "stop," etc., which shows that it is not easy to separate awareness from voice factors.[6] Following Ansolabehere and Konisky [9], we also used the political affiliation of the state governor as a voice factor.

Five indicators were used to consider traditional industrial location factors: the plant cost estimate, the plant capacity, the coal price, the available rail miles (which is important for coal transportation) and the average electricity price. The latter can also be interpreted as a measure of (past) profitability. The plant cost estimates were updated each year using the Chemical Engineering Plant Cost Index (CEPCI), which is a dimensionless number used to update the capital cost required to build a chemical plant from a past date to a later time. This index is widely accepted and consists of subcomponents dealing with equipment, labor costs, buildings, engineering, supervision and other parameters affecting costs. Kreutz et al. [33] provide a comparison of the CEPCI with the Marshall and Swift index, the US GDP deflator and the Handy-Whitman Total Plant-All Steam Generation Index, while Höök et al. [3] used this index to compute the economics of coal-to-liquids and gas-to-liquids plants.

The competition with alternative energy sources was measured by using the average levelized long-term wind price, the average price of residential and commercial solar photovoltaics, and the Henry Hub natural gas price.

We also considered the number of years since the coal plant project had started: we noted that the more time the project spends in its planning phase the less probable will be its full development. We found two reasons for this phenomenon: strong cost escalations and a prolonged legal battle between the local communities and the plant developers. These two reasons were interconnected: the legal battle delayed the coal project to such an extent that the new price environment was no longer profitable due to cost escalations and falling prices of energy alternatives (see the Coal Issues Portal and the history of each coal plant reported there). This phenomenon also confirms again that separating the different indicators in clear-cut categories is not always possible.

All data were collected for each US state for the period January 2004 through December 2013. The data had yearly frequency or were converted to a yearly frequency to match the coal plants data. All data were transformed into logs, except for the *duration* indicator and the binary variable *governor*. In Section 3.2, devoted to the out-of-sample forecasting analysis, we considered a wide set of models also including models with data in levels that is without any transformation.

2.2. Methods: collinearity, stationarity and econometric analysis

We computed the correlation among regressors (see Figs. 1–2 in the Technical Appendix) as well as the Variance Inflation Factors[7] for each regressor (see Tables 3–4 in the Technical Appendix), where we differentiated between coal power plants and CTL plants.[8] Classical "rules of thumbs" to get rid of collinearity are to eliminate those variables with a VIF higher than 10 or to eliminate one of the two variables with a correlation higher than 0.7–0.8 (in absolute value). Given that simply eliminating variables may not be a good solution, as shown by O'Brien [34] and Dormann et al. [35], we followed a less aggressive approach: when two variables had a correlation coefficient (in absolute value) higher than 0.8 we took the first one and the ratio between the first and the second one. The following ratios were considered: CO_2/capacity in place of CO_2 output (for coal power plants only); cost/capacity in the place of cost; rail/population in place of rail miles; solar price/natural gas price in place of solar price (for coal power plants only); solar price/wind price in place of solar price (for CTL plants only); the GI for the keywords "coal-to-liquids + ctl" divided by the natural gas price, in place of the initial GI (for CTL plants only). The last one is the only ratio without an immediate economic meaning. However, considering that hydrocarbon liquefaction can be implemented either using coal or natural gas, this ratio can be roughly interpreted as a ratio between the interest for coal-to-liquids plants and gas-to-liquids plants.

The next step was to check whether our data are stationary. Given the moderate size of our dataset in case of coal power plants and the small size for CTL plants, we employed a battery of panel unit root tests: the test by Levin et al. [36], the test by Im et al. [37], the Fisher-type tests using ADF and PP tests by Maddala and Wu [38] and Choi [39], and the Hadri test [40]. The results in Table 5 in the Technical Appendix show that our data are stationary.

Our dataset consists of the binary dependent variable "status" Y_{it}, which indicates whether a coal project is active/upcoming ($Y_{it} = 0$) or abandoned/canceled ($Y_{it} = 1$), for observation i, ($i = 1, \ldots, n$) and time t, ($t = 1, \ldots, T$), and of the $p \times 1$ vector of regressors X_{it}. We are interested in predicting the expectation of the response variable as a function of the regressors. The expectation of a simple binary response is just the probability that the response is 1:

$$E(Y_{it}|X_{it}) = \pi(Y_{it} = 1|X_{it}). \tag{1}$$

To model this expectation we use logit and probit models:

$$g\{\pi(Y_{it} = 1|X_{it})\} = \beta X_{it} = \nu_i, \tag{2}$$

where ν_i is referred to as the *linear predictor*, $g(\cdot)$ is the logit or probit *link function*, while the distribution of the response given the regressors is always specified as a Bernoulli distribution (see Ref. [41] for more details)[9].

2.3. Model evaluation

To compare the alternative models, we will report the standard Akaike and Schwartz information criteria (AIC and SIC, respectively). We will also compute the Ljung–Box [42] test statistic for testing the absence of autocorrelation up to order k in the models' standardized residuals and squared residuals, as well as the BDS test by Brock et al. [43], to test whether the standardized residuals are independent and identically distributed. This test is robust against a variety of possible deviations from independence, including linear dependence, nonlinear dependence, or chaos.

We will also report the Area Under the Receiver Operating Characteristic (ROC) curve (AUC) by Metz and Kronman [44], Goin [45] and

[6] Alternative keywords for Google search with smaller search volumes will be analyzed in Section 4, Robustness Checks.

[7] Variance Inflation Factors (VIF) are used to measure the degree of collinearity among the regressors in a linear equation. They can be computed by dividing the variance of a coefficient estimate with all the other regressors included by the variance of the same coefficient estimated from an equation with only that regressor and a constant.

[8] The CO_2 output was not considered for CTL plants because this data was available for very few plants.

[9] Random intercepts and random coefficients binary models were not employed because either they did not converge numerically or their random component showed variances that were not statistically different from zero. In the following discussion, we will only consider simple (pooled) logit and probit models.

Hanley and McNeil [46] for all competing models. The ROC curve is obtained by plotting, for any probability threshold, the proportion of correct predictions that a project is abandoned/canceled (*y*-axis), with respect to the proportion of incorrect predictions that a project is active/upcoming (*x*-axis). In terms of model comparison, the best curve is the one that is leftmost, the ideal one coinciding with the *y*-axis (see Refs. [47,48] for a recent application). The AUC lies between zero and one and the closer it is to one the more accurate the classifier is.

Although the AUC is one of the most common tools to measure classifier predictive performance, it has some drawbacks, as recently reviewed by Figini and Maggi [49] and references therein. We also computed the *Model Confidence Set* (MCS), proposed by Hansen et al. [50] and extended by Figini and Maggi [49] to binary models, to assess the prediction power of the competing models (see also Refs. [17,51] for recent applications in financial forecasting). The MCS is a set containing the best forecasting models at a given confidence level. Following Hansen et al. [50], the MCS procedure selects the best model and computes the probability that other models are undistinguishable from the best one using an evaluation rule based on a *loss function*. The more the data are informative the smaller the MCS will be. We computed the MCS following the procedure set up by Hansen et al. [50], adopting the χ^2 test, at 10% confidence level, for the model elimination rule and using different loss functions.[10] The full technical details and the loss functions used in our analysis are reported in Technical Appendix.

3. Results

3.1. In-sample analysis

Tables 3–4 report the results for coal power plants and CTL plants respectively, where the left columns show the results with all the regressors, while the right columns the restricted models with only the regressors that were significant at the 5% level (for coal power plants) and at the 10% level (for CTL plants).[11]

	Logit		Probit		Logit restricted		Probit restricted	
	Coef.	*p-value*	*Coef.*	*p-value*	*Coef.*	*p-value*	*Coef.*	*p-value*
CO_2/Capacity	0.95	0.21	0.48	0.26				
GI (coal)	0.09	0.62	0.06	0.57				
Coal price	**−1.32**	**0.04**	**−0.71**	**0.04**	**−1.59**	**0.00**	**−0.86**	**0.00**
GI (coal plant + coal power)	**13.17**	**0.00**	**7.35**	**0.00**	**14.18**	**0.00**	**7.85**	**0.00**
Cost/Capacity	−0.53	0.55	−0.28	0.59				
Duration	**0.22**	**0.03**	**0.13**	**0.03**	**0.22**	**0.02**	**0.12**	**0.02**
Electricity	−0.01	0.99	−0.01	0.98				
Governor	−0.29	0.26	−0.17	0.23				
Income	1.36	0.40	0.84	0.37				
GI (jobs)	1.06	0.38	0.63	0.36				
LFP	−3.22	0.37	−1.96	0.34				
NG price	**−10.32**	**0.00**	**−5.72**	**0.00**	**−10.22**	**0.00**	**−5.72**	**0.00**
GI (pollution)	0.02	0.94	0.02	0.88				
Population	−0.25	0.15	−0.14	0.16				
Rail/Population	−4.19	0.16	−2.25	0.16				
Solar price/NG price	**−14.78**	**0.00**	**−8.13**	**0.00**	**−15.34**	**0.00**	**−8.53**	**0.00**
Capacity (MW)	0.51	0.08	0.26	0.11				
UR	−1.26	0.08	−0.74	0.06				
Wind price	0.24	0.80	0.03	0.95				
Constant	−16.63	0.22	−9.51	0.21	**−14.77**	**0.00**	**−8.20**	**0.00**
Information criteria and AUC								
AIC		516.52		515.44		501.22		**500.12**
SIC		603.58		602.50		527.33		**526.23**
AUC		71.02%		71.08%		67.08%		67.11%
Residual tests								
Ljung–Box (50) res. [*p*-val]		0.21		**0.03**		0.36		0.51
Ljung–Box (50) res.sq. [*p*-val]		0.29		0.34		**0.00**		0.50
BDS (dim = 2) [*p*-val]		0.09		0.33		0.14		0.09
BDS (dim = 6) [*p*-val]		0.56		0.91		0.56		0.57

In case of coal power plants, as expected, the longer the planning period the higher the probability that the project will be abandoned/cancelled. The lower the price for natural gas and the lower the price for solar photovoltaics with respect to natural gas the higher the probability that the project will be abandoned. The higher the Google search volumes about coal plants and/or coal power the higher the probability the coal project will be abandoned: an increasing number of people looking for information about coal plants on the web highlights a growing opposition to coal projects. An unexpected result is that a higher coal price will increase the probability that a coal plant will be fully developed. A possible explanation of this result could be that the strong commercial relationships between coal mining companies and coal power companies allow high coal prices to be economically viable. Given the very sketchy information about the business structure of the companies involved in coal projects (particularly for abandoned projects), we leave this issue as an interesting avenue for further research.

As for CTL plants only two regressors were found to be significant at the 10% level: the political affiliation of the state governor and the ratio between solar and wind prices. A republican governor will increase the likelihood that the coal plant will be built, whereas a lower price for solar photovoltaics with respect to wind price will increase the probability that the project will be abandoned. The importance of the governor political affiliation is not a surprise given the greater technical complexity, strong environmental impacts and the higher costs of CTL plants. These plants have very high CO_2 emissions, more than double the amount produced by the oil industry (see Refs. [52,53]), they are extremely water-intensive (see Ref. [3] and references therein), and the discharged water must be treated to avoid environmental harm (see Ref. [54]). The recent analysis in 3 highlights very poor economics and "*a strong risk for CTL plants to become financial black holes*" and helps explain why China has strongly slowed down the development of its CTL program (see Ref. [55]). A CTL project will likely succeed only with strong political support at the level of the local state government otherwise it will be better not to proceed further.

Probit models fared better than logit models showing lower information criteria and better residuals properties. Restricted models

[10] Other tests can be applied: for instance, the *F* statistic or other statistics built on the *t*-statistic that do not require the computation of the model covariance matrix. In our applications, the *F* statistic and other *t*-statistics delivered similar results to the χ^2. However, the *t*-statistics are much more demanding in terms of computing time and are convenient when the number of models is large relative to the sample, which is not our case.

[11] We used a higher probability level for CTL plants due to the small size of the dataset.

Table 2
Regressors: description and source.

Variables	Description	Sources
Externalities costs		
CO_2 (tons)	Carbon dioxide output in tons	Carbon Monitoring for Action (CARMA) database
Population	Population by US state in millions	U.S. Department of Commerce: Census Bureau
Awareness and ability to pay for environmental quality		
Income	Median Household Income by US state	U.S. Department of Commerce: Census Bureau
LFP	Labor Force Participation by US state	U.S. Department of Labor: Bureau of Labor Statistics
UR	Unemployment Rate by US state	U.S. Department of Labor: Bureau of Labor Statistics
GI (JOBS)	Google index for the keyword "jobs"	Google trends
Awareness and voice factors		
GI (coal)	Google index for the keyword "coal"	Google trends
GI (coal power + coal plant)	Google index for the keywords "coal power + coal plant"	Google trends
GI (coal-to-liquids + ctl coal)	Google index for the keywords "coal-to-liquids + ctl coal"	Google trends
GI (pollution)	Google index for the keyword "pollution"	Google trends
Governor	Binary variable that is 1 if Republican and 0 otherwise	www.rulers.org
Traditional industrial location factors		
cost	Plant cost estimate (billion $)	CMD/Google search
Coal price	US Central Appalachian coal spot price ($/ton)	BP Statistical Review of World Energy 2013/US EIA
Rail	Rail miles by US state	Association of American Railroads
Capacity (MW)	Plant capacity expressed in MW for coal power	NETL-US DOE/CMD/Google search
Capacity (BBL/day)	Plant capacity expressed in bbl/day for CTL plants	
Electricity	Average electricity price by US state ($/Kwh)	US Energy Information Administration (EIA)
Economics of alternative energy sources		
Wind price	Average levelized long-term wind power purchase agreement prices ($/Mwh)	US Department of Energy/Energy Analysis and Environmental Impacts Department – Lawrence Berkeley National Laboratory
Solar price	Installed price of residential and commercial solar photovoltaics ($/W)	US Department of Energy (DOE)/Lawrence Berkeley National Laboratory
NG price	US Henry Hub natural gas price ($/MmBtu)	BP Statistical Review of World Energy 2013/US EIA
Additional indicators		
Duration	The number of years that has passed at time *t* since the project started	The National Energy Technology Laboratory (NETL) The Center for Media and Democracy (CMD) Google search

Table 3
Coal power plants: model estimation results. Smallest information criteria and *p*-values smaller than 5% are reported in bold font.

	Logit		Probit		Logit restricted		Probit restricted	
	Coef.	*p-value*	*Coef.*	*p-value*	*Coef.*	*p-value*	*Coef.*	*p-value*
CO_2/Capacity	0.95	0.21	0.48	0.26				
GI (coal)	0.09	0.62	0.06	0.57				
Coal price	**−1.32**	**0.04**	**−0.71**	**0.04**	**−1.59**	**0.00**	**−0.86**	**0.00**
GI (coal plant + coal power)	**13.17**	**0.00**	**7.35**	**0.00**	**14.18**	**0.00**	**7.85**	**0.00**
Cost/Capacity	−0.53	0.55	−0.28	0.59				
Duration	**0.22**	**0.03**	**0.13**	**0.03**	**0.22**	**0.02**	**0.12**	**0.02**
Electricity	−0.01	0.99	−0.01	0.98				
Governor	−0.29	0.26	−0.17	0.23				
Income	1.36	0.40	0.84	0.37				
GI (jobs)	1.06	0.38	0.63	0.36				
LFP	−3.22	0.37	−1.96	0.34				
NG price	**−10.32**	**0.00**	**−5.72**	**0.00**	**−10.22**	**0.00**	**−5.72**	**0.00**
GI (pollution)	0.02	0.94	0.02	0.88				
Population	−0.25	0.15	−0.14	0.16				
Rail/Population	−4.19	0.16	−2.25	0.16				
Solar price/NG price	**−14.78**	**0.00**	**−8.13**	**0.00**	**−15.34**	**0.00**	**−8.53**	**0.00**
Capacity (MW)	0.51	0.08	0.26	0.11				
UR	−1.26	0.08	−0.74	0.06				
Wind price	0.24	0.80	0.03	0.95				
Constant	−16.63	0.22	−9.51	0.21	**−14.77**	**0.00**	**−8.20**	**0.00**
Information criteria and AUC								
AIC		516.52		515.44		501.22		**500.12**
SIC		603.58		602.50		527.33		**526.23**
AUC		71.02%		71.08%		67.08%		67.11%
Residual tests								
Ljung–Box (50) res. [*p*-val]		0.21		**0.03**		0.36		0.51
Ljung–Box (50) res.sq. [*p*-val]		0.29		0.34		**0.00**		0.50
BDS (dim = 2) [*p*-val]		0.09		0.33		0.14		0.09
BDS (dim = 6) [*p*-val]		0.56		0.91		0.56		0.57

Table 4
Coal-to-liquids plants: model estimation results. Smallest information criteria and *p*-values smaller than 5% are reported in bold font.

	Logit		Probit		Logit restricted		Probit restricted	
	Coef.	*p-value*	*Coef.*	*p-value*	*Coef.*	*p-value*	*Coef.*	*p-value*
GI (coal)	1.59	0.31	0.92	0.30				
Coal price	0.31	0.88	0.18	0.87				
GI (coal-to-liquids + ctl)/ng price)	0.57	0.28	0.32	0.29				
Cost/Capacity	−1.83	0.40	−1.08	0.37				
Duration	0.45	0.25	0.26	0.18				
Electricity	3.20	0.28	1.86	0.18				
Governor	**−2.28**	**0.02**	**−1.32**	**0.01**	**−1.27**	**0.03**	**−0.76**	**0.02**
Income	0.81	0.87	0.43	0.86				
GI (jobs)	−4.70	0.38	−2.74	0.31				
LFP	−8.11	0.41	−4.69	0.36				
NG price	−0.22	0.92	0.00	1.00				
GI (pollution)	−0.54	0.37	−0.34	0.32				
Population	1.32	0.14	0.74	0.10				
Rail/Population	2.20	0.86	1.12	0.85				
Solar price/Wind price	*−54.47*	*0.08*	*−31.10*	*0.07*	**−20.32**	**0.03**	**−11.80**	**0.02**
Capacity (BBL/day)	0.06	0.88	0.04	0.85				
UR	−3.91	0.12	−2.21	0.12				
Wind price	1.83	0.49	0.86	0.54				
Constant	42.34	0.28	25.55	0.21	*8.60*	*0.06*	**4.98**	**0.04**
Information criteria and AUC								
AIC		109.51		108.99		87.44		**87.07**
SIC		157.84		157.31		95.07		**94.70**
AUC		79.83%		80.21%		70.51%		70.51%
Residual tests								
Ljung–Box (50) res. [*p*-val]		0.42		0.62		0.20		0.47
Ljung–Box (50) res.sq. [*p*-val]		0.21		0.91		0.06		0.44
BDS (dim = 2) [*p*-val]		0.41		0.53		**0.00**		0.85
BDS (dim = 6) [*p*-val]		0.54		0.95		0.07		0.93

p-values between 5% and 10% are reported in italics.

showed lower information criteria but full models had higher AUC values.

3.2. Out-of-sample forecasting analysis

To better evaluate the predictive performance for each model we also implemented a cross-validation procedure. We divided our dataset into two parts of equal size: the first one was used as the training set while the second one as the validation set. Similarly to Ref. [56], we compared a set of alternative models whose characteristics are reported in Table 5. We considered both logit and probit models, models with all the regressors, as well as restricted models with only significant parameters at the 5% level; models with data in logs and models with data in levels; models without GIs and models with only GIs. In case of CTL plants, due to the very small sample size of the training and validation sets, we only considered restricted models with GIs only and without GIs.

The estimated AUC for all previous models are reported in Table 6. The restricted probit model with data in logs was the best for coal power plants projects, while the probit model with data in logs and no GIs was the best for CTL projects, thus confirming previous in-sample results discussed in Section 3.1.

Table 5
List of forecasting models.

Coal power plants				Coal-to-liquids plants			
Model	*Data transformation*	*All regressors/restricted model*	*Google data*	*Model*	*Data transformation*	*All regressors/restricted model*	*Google data*
Logit	log	All	Yes	Logit	log	Restricted	No
Probit	log	All	Yes	Probit	log	Restricted	No
Logit	log	Restricted	No	Logit	log	Restricted	Only
Probit	log	Restricted	No	Probit	log	Restricted	Only
Logit	log	Restricted	No	Logit	Levels	Restricted	No
Probit	log	Restricted	No	Probit	Levels	Restricted	No
Logit	log	Restricted	Yes	Logit	Levels	Restricted	Only
Probit	log	Restricted	Yes	Probit	Levels	Restricted	Only
Logit	log	Restricted	Only				
Probit	log	Restricted	Only				
Logit	Levels	All	Yes				
Probit	Levels	All	Yes				
Logit	Levels	Restricted	Yes				
Probit	Levels	Restricted	Yes				
Logit	Levels	Restricted	No				
Probit	Levels	Restricted	No				
Logit	Levels	Restricted	No				
Probit	Levels	Restricted	No				
Logit	Levels	Restricted	Only				
Probit	Levels	Restricted	Only				

Table 6
A.U.C. for each forecasting model. The best model is reported in bold font.

Models: coal power plants	AUC	Models: coal-to-liquids	AUC
Logit log	59.48%	Logit log (no Google)	60.74%
Probit log	60.13%	**Probit log (no Google)**	**61.85%**
Logit log (no Google)	57.14%	Logit log (only Google)	52.04%
Probit log (no Google)	57.74%	Probit log (only Google)	52.78%
Logit log (no Google) restricted	58.23%	Logit levels (no Google)	60.37%
Probit log (no Google) restricted	58.25%	Probit levels (no Google)	59.07%
Logit log restricted	63.91%	Logit levels (only Google)	53.89%
Probit log restricted	**64.13%**	Probit levels (only Google)	55.74%
Logit log (only Google)	48.18%		
Probit log (only Google)	48.08%		
Logit levels	60.13%		
Probit levels	60.00%		
Logit levels restricted	62.35%		
Probit levels restricted	63.87%		
Logit levels (no Google)	57.46%		
Probit levels (no Google)	58.01%		
Logit levels (no Google) restricted	60.17%		
Probit levels (no Google) restricted	60.17%		
Logit levels (only Google)	48.11%		
Probit levels (only Google)	48.33%		

We then employed the MCS approach developed by Hansen et al. [50] and discussed in Section 2.2 to test for statistically significant differences in the forecast performances among the competing models. We recall from Section 2.2 that the MCS procedure will yield a set containing the best forecasting models at a given confidence level. The full results of the MCS procedure are reported in Table 6 in the Technical Appendix.

In case of coal power plants, the restricted probit model with data in logs is the model with the lowest loss for almost all loss functions considered, thus confirming the previous results. Moreover, models with Google data represent the majority of models included in the MCS, while models without Google data are seldom included, thus confirming the important information that this type of data can provide. As for CTL plants, the logit models with data in levels without Google data is the one that has the lowest loss across a spectrum of loss functions. Almost all models are now included in the MCS, which highlights that the validation set is not very informative (which was expected given its small size).

4. Robustness checks

To verify that our previous results hold also with alternative data setups, we performed a series of robustness checks: we considered alternative keywords for Google search and we evaluated the effect on our estimates of the global financial crisis in 2008 and 2009. Finally, we included as additional regressors the two dimensions of the Dynamic-Weighted Nominate (DWN) database developed by the political scientists Poole and Rosenthal in the early 1980s to analyze legislative roll-call voting behavior in the US congress, see Refs. [18,19].

4.1. Alternative keywords

One of the regressors in our analysis was the GI for the search term "pollution." While this keyword is very general and should include all possible searches related to environmental hazards, it may be too way general and not related to coal plants: for example, two of the top rising searches for this term in the US were "*pollution in china*" and "*china pollution*". Google Trends provides also the search trends for specific categories, which include all searches related to the chosen category according to some internal selection algorithms. The closest category related to pollution and environmental hazards is *Business and Industrial/Energy and utilities/Waste Management*. Similarly, we also downloaded the GI related to the keywords "*coal power + coal plant*" and "*coal-to-liquids + ctl coal*," but restricted to the category *Business and Industrial/Energy and utilities*. The estimated coefficients for the models including these two alternative GIs in the place of the initial ones are reported in Tables 7–8 in the Technical Appendix for coal power plants and CTL plants, respectively.

In the case of coal power plants the results do not change much in terms of signs and significance with respect to the baseline case in Table 3. The only difference is that now the *coal price* is not statistically significant. The information criteria (AIC and SIC) are higher than in the baseline case in Table 3 and the AUCs are lower. The residuals tests highlights some small misspecification in the squared residuals. In general, this robustness check confirms the results discussed in Section 3.1 but with a worse fit than the baseline case.

In the case of CTL plants, the main findings of the baseline case are also confirmed with some interesting differences. The indicator *governor* continues to be a strong significant variable (now even at the 1% level) and with the same sign as in the baseline case. Similarly, lower solar prices (with respect to wind prices) increase the probability that the project will be abandoned. The GI for "waste management" is not statistically significant as was the case for the GI for the keyword "pollution." Differently from the baseline case in Table 4, the ratio of the GI for the keywords "*coal-to-liquids + ctl coal*" and the natural gas price is now significant at the 5% level with a positive coefficient: the more people look for information about CTL plants (with respect to natural gas prices) the higher the probability the project will be abandoned. The lower the unemployment rate and the lower the number people looking for "jobs," the higher is the probability that the coal project will be abandoned/canceled. Furthermore, a higher population will decrease the odds that the coal plant will be built. The information criteria are now slightly lower than those in Table 4 for the baseline case, the AUCs are slightly higher, while the residuals tests do not highlight any particular misspecification. In general, restricting the selection criteria for Google data seems to be beneficial for the analysis of CTL plants by eliminating unrelated searches and highlighting additional significant factors beyond the political affiliation of the state governor and renewable prices, which still remain the most important factors.[12]

4.2. The recession in the years 2008–2009

The second robustness check was to evaluate the effect on our estimates of the global financial crisis in 2008 and 2009. Given the small temporal dimension of our dataset, we used a dummy variable for the years 2008 and 2009 in correspondence to the official NBER recession for the US. The estimated models including this dummy variable are reported in Tables 9–10 in the Technical Appendix for coal power plants and CTL plants[13], respectively.

In the case of coal power plants, the dummy variable is not statistically significant across all model specifications and the coefficients of all other parameters are very close to the baseline case reported in Table 3. Similar results can be observed also for CTL plants: the coefficient for the dummy variable is not significant and the results remain basically the same as those reported in Table 8 in the Technical Appendix and discussed in Section 4.1.

[12] Eliminating these variables results in steep increases of information criteria and a worse AUC, much more than the other regressors.

[13] We used restricted GIs for CTL plants, given the better results found in Section 4.1.

The global financial crisis seems not to have influenced the fate of coal projects significantly.

4.3. The DW-Nominate database

Poole and Rosenthal [18] showed that roll-call voting in both the US House and the Senate can be organized and explained by no more than two dimensions throughout the whole of American history: the first dimension is the familiar left-right (or liberal-conservative) spectrum on economic matters. The second dimension explains attitudes on cross-cutting, salient issues of the day (which include or have included slavery, bimetallism, civil rights, regional, and social/lifestyle issues), see Ref. [57] for an informal discussion. The DW-Nominate database can allow to more carefully disentangle ideology (which somehow reflects voters-at-large perception) and political affiliation of the politician in charge (possibly more affected by lobbying). We included as additional regressors the two dimensions for the US senators whose jurisdiction covered the location of the coal plants. The estimated models are reported in Tables 11—12 in the Technical Appendix.

The estimates for coal power plants show that the two dimensions are not significant while the other coefficients are very close to the baseline case in Table 3. As for CTL plants, the second dimension turns out to be significant at the 10% level with a positive coefficient: since this dimension is also interpreted as the North/South divide, this may indicate that the lobbies for coal plants are much stronger in socially conservative southern states than in northern socially liberal states. This result and the governor affiliation confirm again that CTL plants can be built only in socially conservative states with a strong lobby. Otherwise, the complex engineering and high environmental impacts make the building of these plants extremely difficult, if not impossible.

5. Conclusions

The construction of new coal plants has become an issue of great relevance in the US given the large fleet of old coal power plants that should be replaced. The analysis of the determinants that influence the success or failure of coal plants projects may be relevant for both energy policy making and project planning. We analyzed 145 coal plants and 25 coal-to-liquid plants that have been proposed in US in the period 2004—2013 and we investigated the decision to complete the plant or abandon the project using several variables and binary data models. Beside common industrial explanatory variables (size, input, output and labor costs, substitute costs, infrastructure), we also considered measures of social and environmental awareness using Google search data. After controlling for collinearity, stationarity and robustness, we performed an extensive model specification, comparison and selection.

We found that the project duration, the prices of energy substitutes for electricity generation and the awareness about the coal projects and its hazards are the main factors for coal power plants. The longer the planning period the less likely the project will be implemented: expensive legal disputes and costly project modifications to meet new requirements can make plant profitability vanish. The lower the price for natural gas and the lower the price for solar photovoltaics with respect to natural gas price the higher is the probability that the project will be abandoned. Awareness by local communities as measured by the Google search volumes about coal plants and/or coal power plants increases the probability that the project will be abandoned.

As for CTL plants, we found that the state governor's political affiliation, the ratio between solar and wind prices, the population size, the unemployment rate and the job searches as measured by Google data are the main drivers (however, the latter three are only weakly significant). CTL plants are more likely to be completed in conservative states where we presume that there is stronger political support for heavy industry projects. The lower price of solar photovoltaics with respect to wind price the higher the probability that the project will be abandoned. Larger state populations make these projects less likely, as expected, while higher unemployment rates and job searches increase the probability of successful implementation.

Appendix A. Supplementary data

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.esr.2014.11.005.

References

[1] D. Fantazzini, M. Höök, A. Angelantoni, Global oil risks in the early 21st century, Energy Policy 39 (12) (2011) 7865—7873.
[2] The National Energy Technology Laboratory Department of Energy. www.netl.-doe.gov, 2007.
[3] M. Höök, D. Fantazzini, A. Angelantoni, S. Snowden, Hydrocarbon liquefaction: viability as a peak oil mitigation strategy, Philos. Trans. R. Soc. A (2014) (forthcoming).
[4] B. Freese, S. Clemmer, C. Martinez, A. Nogee, A Risky Proposition — The Financial Hazards of New Investments in Coal Plants, Working Paper, Union Of Concerned Scientists, March 2011.
[5] L. Fleischman, R. Cleetus, S. Clemmer, J. Deyette, S. Frenkel, Ripe for retirement: an economic analysis of the U.S. coal fleet, Electr. J. 26 (10) (2013) 51—63.
[6] Sierra Club, Proposed Coal Plant Tracker, 2014. http://content.sierraclub.org/coal/environmentallaw/plant-tracker.
[7] CoalSwarm Center for Media and Democracy, Proposed Coal Plants in the United States, 2014. http://www.sourcewatch.org/index.php/Category:Proposed_coal_plants_in_the_United_States.
[8] EIA, EIA Annual Energy Outlook 2014, 2014. http://www.eia.gov/forecasts/aeo/.
[9] S. Ansolabehere, D. Konisky, Public attitudes toward construction of new power plants, Public Opin. Q. (2009). http://dx.doi.org/10.1093/poq/nfp041.
[10] P. Garrone, A. Groppi, Siting locally-unwanted facilities: what can be learnt from the location of Italian power plants, Energy Policy 45 (2012) 176—186.
[11] H. Choi, H. Varian, Predicting the Present with Google Trends, Working Paper, Google Inc., 06 2009.
[12] N. Askitas, K.F. Zimmermann, Google econometrics and unemployment forecasting, Appl. Econ. Q. 55 (2009) 107—120.
[13] T. Suhoy, Query Indices and a 2008 Downturn, Discussion Paper 6, Bank of Israel, 2009.
[14] J. Ginsberg, M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski, L. Brilliant, Detecting influenza epidemics using search engine query data, Nature 457 (2009) 1012—1014.
[15] Z. Da, J. Engelberg, G. Pengjie, In search of attention, J. Finance 5 (2011) 1461—1499.
[16] F. D'Amuri, J. Marcucci, The Predictive Power of Google Searches in Forecasting Unemployment. Temi di discussione (Economic Working Papers) 891, vol. 11, Bank of Italy, 2013.
[17] D. Fantazzini, N. Fomichev, Forecasting the real price of oil using online search data, Int. J. Comput. Econ. Econ. 4 (1—2) (2014) 4—31.
[18] K.T. Poole, H. Rosenthal, A spatial model for legislative roll call analysis, Am. J. Political Sci. 29 (2) (1985) 357—384.
[19] K.T. Poole, Spatial Models of Parliamentary Voting, Cambridge University Press, 2005.
[20] M. Höök, K. Aleklett, A review on coal to liquid fuels and its coal consumption, Int. J. Energy Res. 34 (10) (2014) 848—864.
[21] R. Coase, The problem of social cost, J. Law Econ. 3 (1960) 1—23.
[22] J.T. Hamilton, Politics and social costs: estimating the impact of collective action on hazardous waste facilities, RAND J. Econ. 24 (1) (1993) 101—125.
[23] J.T. Hamilton, Testing for environmental racism: prejudice, profits, political power? J. Policy Anal. Manag. 14 (1) (1995) 107—132.
[24] R.R. Jenkins, K.B. Maguire, C.L. Morgan, Host community compensation and municipal solid waste landfills, Land Econ. 80 (4) (2004) 513—528.
[25] D.L. Anderton, A.B. Anderson, J.M. Oakes, M.R. Fraser, Environmental equity: the demographics of dumping, Demography 31 (2) (1994) 171—192.
[26] V. Been, F. Gupta, Coming to the nuisance or going to the barrios? A longitudinal analysis of environmental justice claims, Ecol. Law Q. 24 (1) (1997) 1—56.
[27] S. Arora, T.N. Cason, Do community characteristics influence environmental outcomes?: evidence from the toxics release inventory, J. Appl. Econ. 1 (2) (1998) 413—453.
[28] A. Wolverton, Effects of socio-economic and input-related factors on polluting plants' location decisions, B.E. J. Econ. Anal. Policy 9 (1) (2009). Article 14, Available at: http://www.bepress.com/bejeap/vol9/iss1/.
[29] R. Cleetus, S. Clemmer, E. Davis, J. Deyette, J. Downing, S. Frenkel, Ripe for Retirement: the Case for Closing America's Costliest Coal Plants, Working Paper, Union of Concerned Scientists, Cambridge, MA, 2012.
[30] S. Tierney, Why Coal Plants Retire: Power Market Fundamentals as of 2012, 2012. http://www.analysisgroup.com/uploadedFiles/News_and_Events/News/2012_Tierney_WhyCoalPlantsRetire.pdf.

[31] L.F. Pratson, D. Haerer, D. Patino-Echeverri, Fuel prices, emission standards, and generation costs for coal vs natural gas power plants, Environ. Sci. Technol. 47 (9) (2013) 4926—4933.
[32] J.T. Boer, M. Pastor, J.S. Sadd, L.D. Snyder, Is there environmental racism? The demographics of hazardous waste in los angeles county, Soc. Sci. Q. 78 (4) (1997) 793—809.
[33] T. Kreutz, E. Larson, G. Liu, R. Williams, Fischer-Tropsch fuels from coal and biomass, Pittsburgh, USA, in: Proceedings of the 25th Annual International Pittsburgh Coal Conference. 29 September to 2 October, 2008. Available from: http://web.mit.edu/mitei/docs/reports/kreutz-fischer-tropsch.pdf.
[34] R. O'Brien, A caution regarding rules of thumb for variance inflation factors, Qual. Quantity 41 (5) (2007) 673—690.
[35] C. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. García Marquéz, B. Gruber, B. Lafourcade, P. Leitao, T. Münkemüller, C. McClean, P. Osborne, B. Reineking, B. Schröder, A. Skidmore, D. Zurell, S. Lautenbach, Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, Ecography 36 (2013) 27—46.
[36] A. Levin, C. Lin, C. Chu, Unit root tests in panel data: asymptotic and finite-sample properties, J. Econ. 108 (2002) 1—24.
[37] K. Im, M. Pesaran, Y. Shin, Testing for unit roots in heterogeneous panels, J. Econ. 115 (2003) 53—74.
[38] G. Maddala, S. Wu, A comparative study of unit root tests with panel data and a new simple test, Oxf. Bull. Econ. Statistics 61 (1999) 631—652.
[39] I. Choi, Unit root tests for panel data, J. Int. Money Finance 20 (2001) 249—272.
[40] K. Hadri, Testing for stationarity in heterogeneous panel data, Econ. J. 3 (2000) 148—161.
[41] P. Giudici, S. Figini, Applied Data Mining for Business and Industry, second ed., Wiley, 2009.
[42] G. Ljung, G. Box, On a measure of lack of fit in time series models, Biometrika 66 (1979) 265—270.
[43] W. Brock, D. Dechert, J. Sheinkman, B. LeBaron, A test for independence based on the correlation dimension, Econ. Rev. 175 (3) (1996) 197—235.
[44] C. Metz, H. Kronman, Statistical significance tests for binormal ROC curves, J. Math. Psychol. 22 (1980) 218—243.
[45] J.E. Goin, ROC curve estimation and hypothesis testing: applications to breast cancer detection, J. Pattern Recognit. Soc. 15 (1982) 263—269.
[46] A. Hanley, B. McNeil, The meaning and use of the area under a receiver operating characteristics (ROC) curve, Diagn. Radiol. 143 (1982) 29—36.
[47] W.J. Krzanowski, D.J. Hand, ROC Curves for Continuous Data, CRC/Chapman Hall, 2009.
[48] D. Fantazzini, M. De Giuli, S. Figini, P. Giudici, Enhanced credit default models for heterogeneous SME segments, J. Financial Transformation 25 (2009) 31—39.
[49] S. Figini, M. Maggi, Performance of Credit Risk Prediction Models via Proper Loss Functions, Working Paper 64, Department of Economics and Management, University of Pavia, 2014.
[50] P. Hansen, A. Lunde, J.M. Nason, The model confidence set, Econometrica 79 (2) (2011) 453—497.
[51] E. Rossi, D. Fantazzini, Long memory and periodicity in intraday volatility, J. Financial Econ. (2014) (forthcoming).
[52] H. Mulder, Coal-To-Liquids: Can CO_2 Emissions be Significantly Reduced?, 2009. http://www.engineering.pitt.edu/Coal_Conference/Mulder_Presentation.aspx.
[53] S. Mui, L. Tonachel, B. McEnaney, E. Shope, Ghg Emission Factors for High Carbon Intensity Crude Oils. Technical Report 2, Natural Resources Defense Council, September 2010.
[54] S. Lei, J. Zhang, An analysis of coal-to-oil industrys impact on environment, Technology 7 (2009) 84—88.
[55] F. Rong, D.G. Victor, Coal liquefaction policy in china: explaining the policy reversal since 2006, Energy Policy 39 (2011) 8175—8184. http://dx.doi.org/10.1016/j.enpol.2011.10.017.
[56] T. Young, R. Zaretski, J. Perdue, F. Guess, X. Liu, Logistic regression models of factors influencing the location of bioenergy and biofuels plants, BioResources 6 (1) (2011) 329—343.
[57] Wikipedia, Nominate (Scaling Method), 2014. http://en.wikipedia.org/wiki/NOMINATE_(scaling_methodhttp://en.wikipedia.org/wiki/NOMINATE_(scaling_method).