Overview of Cost Estimation Methods

James Woods

Prelude

- Costs are always measured with some error.
- Not all cost forecasts are good.
- The incentives to reduce costs vary widely.

How do you mismeasure costs?

- Misclassify regulated vs non-regulated
  - Incentives to do this.
- Cost categories (Account codes)
  - Insurance moves from selling expenses to administrative
- Reallocation from one customer type to another.
  - Industrial to residential, for example if a new subdivision or two springs up.
- To and from Rate Base
  - Costs flow through but Rate base gets a return.

Random events.

- Cost increasing random events are more frequent than cost saving.
- Extraordinary items are on income statement.
- Random events are more likely to increase costs than lower them.

General Methods

- Base or Test year
- Ideal system
- Regression
- Corrected Regression
- Frontier Regression
- Data Envelopment
- Yardstick

In almost all cases, the idea is to find out what the revenue requirement should be and after that setting prices, tariffs, to meet that requirement given other objectives.

Test Year

- Use the costs and sales volume from this year to develop revenue requirements and tariff structure for the next few years.
  - One data point is used (under one interpretation)
- *Adversarial system*: Utilities on one side and rate payer advocate (Often commissioners staff) on the other and commissioners acting as judge.
  - Fight over cost, rate base, share holder allocation of costs.
- *Accounting system*: Auditable with receipts. It is what really happened.

Ideal System

- Observe the needs of the system, the location of the customers and what they consume.
- Design a system that would satisfy these customers.
- Price that system.
- Example: Oregon Quality Education Model (https://www.oregon.gov/ode/reports-and-data/taskcomm/Pages/QEMReports.aspx)
- You can see other examples of this in the EIA cost estimates.

Ideal System Comments

- Very hard to estimate costs. Common numbers
  - Rough: within 30-60% and biased low
  - Semi-detailed: within 20-30% and biased low
  - Detailed (with blueprints and bids): 3-5% and biased low
- They run low because all cost estimates run low. You don't accidentally add in something expensive but you do forget to add in something expensive.
- Ideal systems don't have legacy systems or history.
  - Example of SMUD and Rancho Seco in 1966, which closed in 1989.
- All these cost estimates are very expensive.

Multi Year (Period) Methods

- The next group:
  - Regression
  - Corrected Regression
  - Frontier Regression
  - Data Envelopment Analysis (DEA)
- Use comparable data from many time periods.
- I will demonstrate with artificial data.

Please note that even in the test year methodology you still use these techniques behind the scenes.

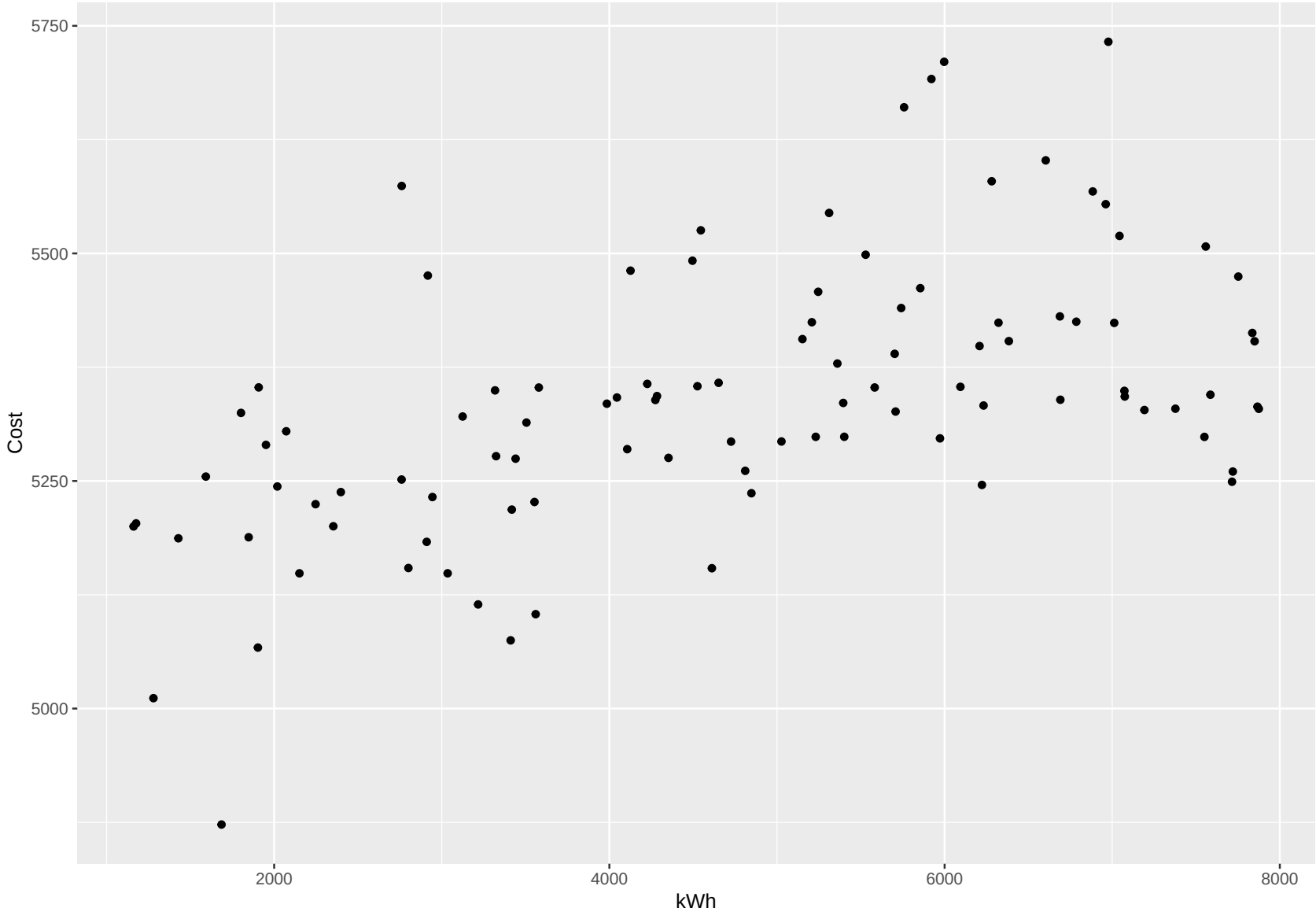
Fake Data

- I will create data of the form $Cost = F + \beta kWh_t + e_t + \nu_t$
  - F is intended to be a fixed cost ($ 5000)
  - β is a constant marginal cost ($ 0.05)
  - kWh_t is the kWh generated in time t.
  - e_t is a mean zero uncertainty in time t.
  - ν_t is uncertainty that is positive and bias up. Accidents.
- More could be done with increasing cost of generation and different costs of producing at different times of day and by different generation units.

Data Summary

kWh	e	nu	Cost
Min. :1161	Min. :-240.025	Min. : 0.924	Min. :4873
1st Qu.:3194	1st Qu.: -57.923	1st Qu.: 26.885	1st Qu.:5251
Median :4829	Median : 3.270	Median : 85.185	Median :5334
Mean :4757	Mean : 7.697	Mean : 93.945	Mean :5339
3rd Qu.:6291	3rd Qu.: 82.394	3rd Qu.:131.896	3rd Qu.:5424
Max. :7875	Max. : 261.776	Max. :525.089	Max. :5732

The Data

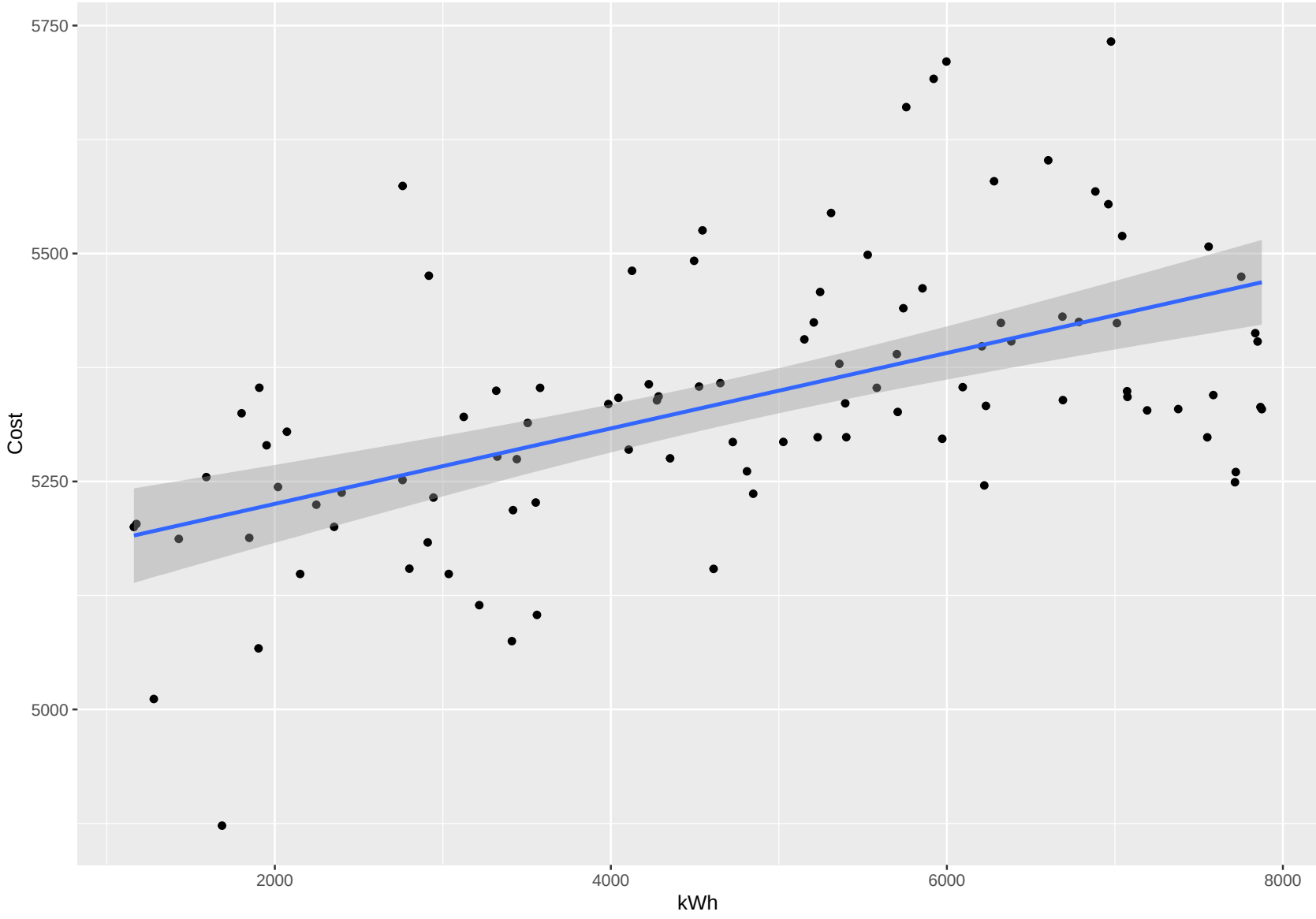


Regression

- The idea of a regression is to run a function through the dots such that the sum of squares of the vertical distances is as small as possible.

Table 2

	Dependent variable:
	Cost
kWh	0.041^{***}
	(0.006)
Constant	$5{,}142.709^{***}$
	(32.751)
Observations	100
R^2	0.301
Adjusted R^2	0.293
Residual Std. Error	123.834 (df = 98)
F Statistic	42.119^{***} (df = 1; 98)
Note:	*p<0.1; **p<0.05; ***p<0.01

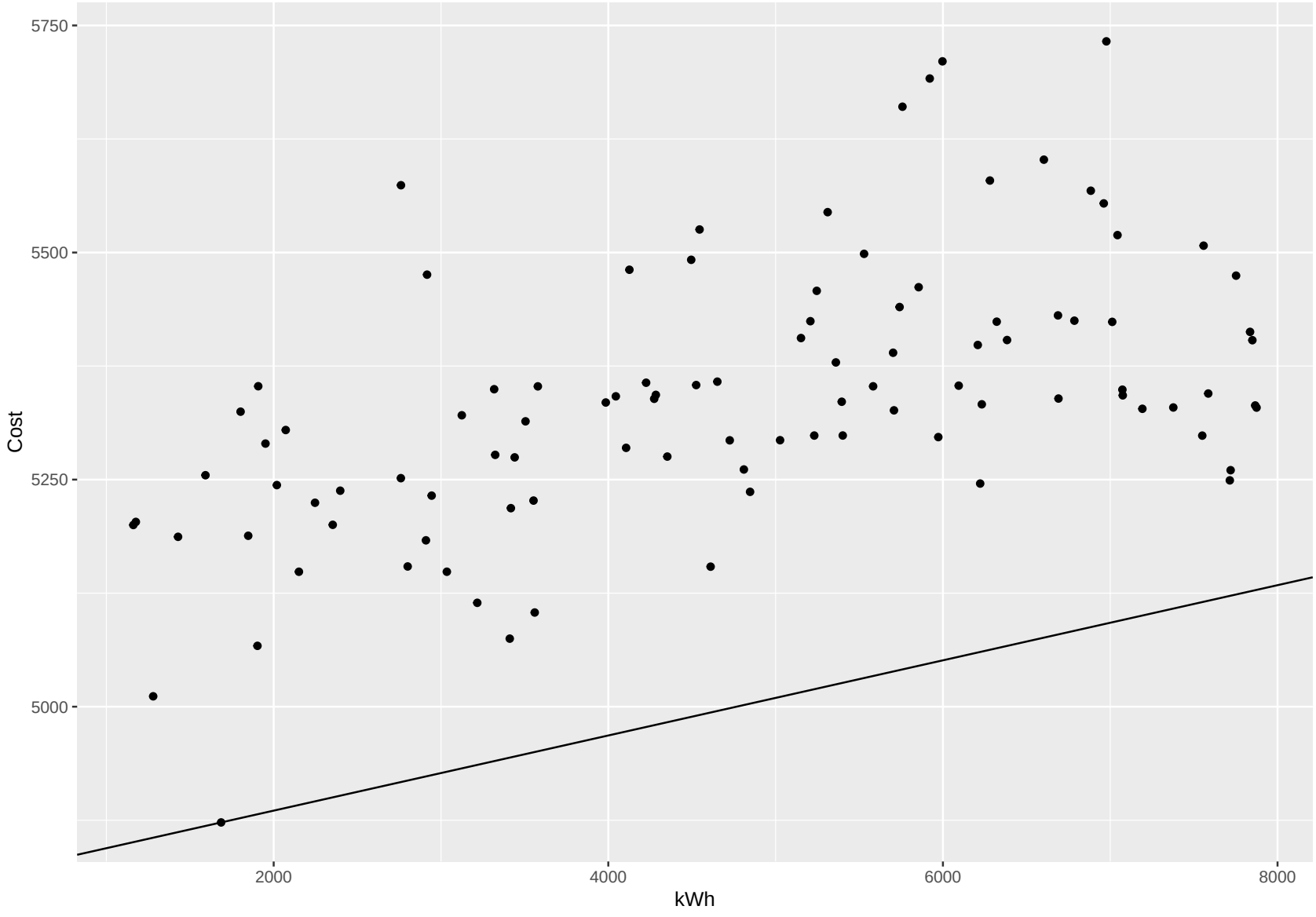

5750
5500
5250
5000
Cost
2000
4000
6000
8000
kWh

Comments on Regression

- Will generally hit the true slope parameters (If the volume is measured without noise) but not the intercept.
  - When it is correctly specified, right parameters and shape.
- Yes, there is uncertainty.
- One observation can't do much but you can measure the influence and see the effect easily.

Corrected Regression

- Slides the regression line so that all the dots are above the line.
- Intercept term is $intercept + min(e + \nu)$
- Intended to be the ideal, what can be accomplished.


5750
5500
5250
5000
Cost
2000
4000
6000
8000
kWh

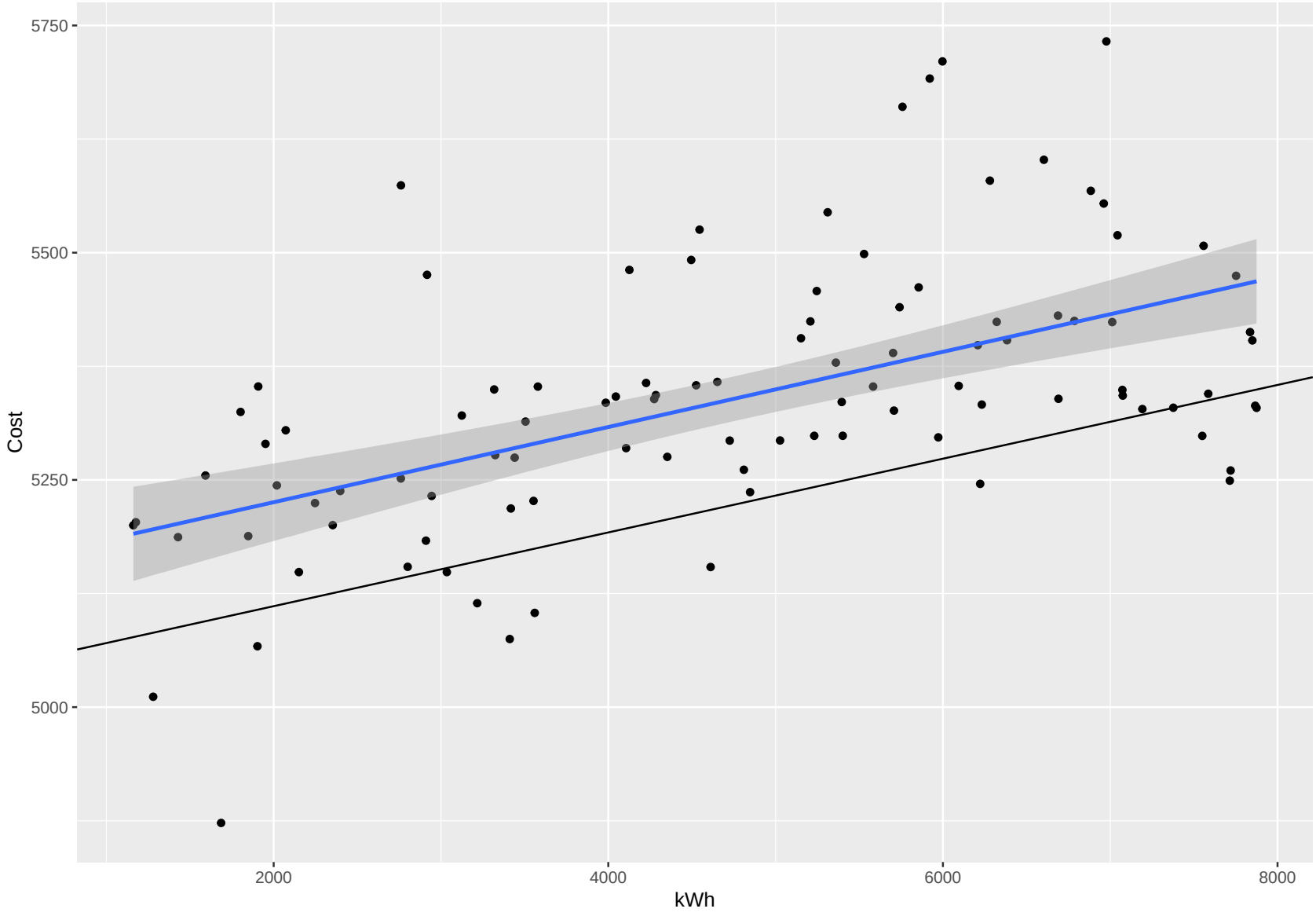
Corrected Regression

- One observation moves the line.
- Always understates fixed cost.
- Makes statisticians sigh.

Frontier Regression

- Regression assumes that the errors, deviations from the line, are symmetric, same amount above as below.
- Frontier says the errors can by asymmetric
- Commonly assume a non-normal distribution like the gamma. We used normal-exponential (Special kind of gamma) mixture

```
## Error Components Frontier (see Battese & Coelli 1992)
## Inefficiency increases the endogenous variable (as in a
## The dependent variable is logged
## Iterative ML estimation terminated after 7 iterations:
## log likelihood values and parameters of two successive i
## are within the tolerance limit
##
## final maximum likelihood estimates
##               Estimate Std. Error    z value  Pr(>|z|)
## (Intercept) 5.0299e+03 3.1315e+01   160.6259 < 2.2e-16 *
## kWh         4.0567e-02 5.8366e-03     6.9505 3.641e-12 *
## sigmaSq     2.8760e+04 1.0002e+00 28753.1323 < 2.2e-16 *
## gamma       7.5223e-01 7.6767e-02     9.7989 < 2.2e-16 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
## log likelihood value: -621.1869
##
## cross-sectional data
## total number of observations = 100
##
```

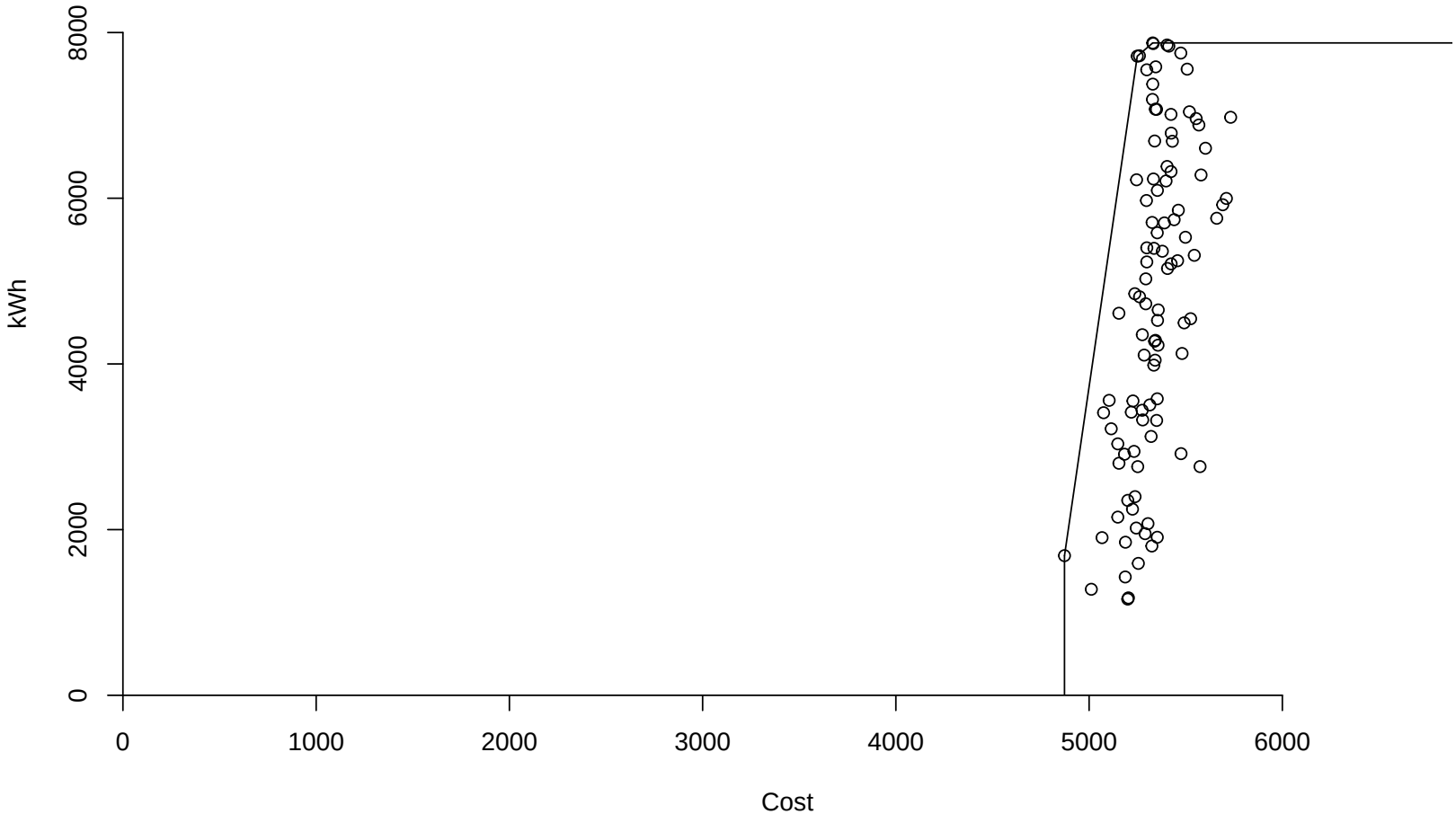

5750
5500
5250
5000
Cost
2000
4000
6000
8000
kWh

Comments on Frontier

- This is how we generated the data
- Not as sensitive to one observation point
- Gives an estimate of the potential efficiency gains.

Data Envelopment

- Create the smallest shape around the bottom of the dots drawing from dot to dot.
- Convex hull is the technical term.
- Efficiency is measured as distance from the frontier
- This one looks stupid. The axes are reversed because of technical limitations. Shows the highest kWh per cost level.


kWh
0
2000
4000
6000
8000
0
1000
2000
3000
4000
5000
6000
Cost

Comments on DEA

- Preferred by engineers
- Some variations, bootstrap, give uncertainty.
- See how it depends on only a few observations.
- You can also use it for isoquants.

Yardstick

- This means look at the costs of other firms.
- The focus is not on accuracy but on incentives to tell the truth and to reduce costs.
  - If every cost reduction you make results in lower prices and no benefit to shareholders, why do it?
  - With yardstick, if you make a cost reduction before the other guy does, your shareholders gain.
- Can be combined with the other techniques.