GERNER IV
Efficiency analyses for network regulation:
Efficiency Analyses GAS

Per AGRELL, pja@sumicsid.com
Peter BOGETOFT, pb@sumicsid.com

BONN
2008-10-15

Team

SUMICSID Group

- Prof. Per AGRELL (Univ Cath de Louvain)
- Prof. Peter BOGETOFT (Copenhagen Business School)
- Mathias Lorenz

EE2 (TU Dresden)

- Prof Christian von HIRSCHHAUSEN
- Astrid Cullman
- Mathias Walter
- Anne Neumann

Outline

Task

Methods

Process

Average model specfication

Frontier model specification

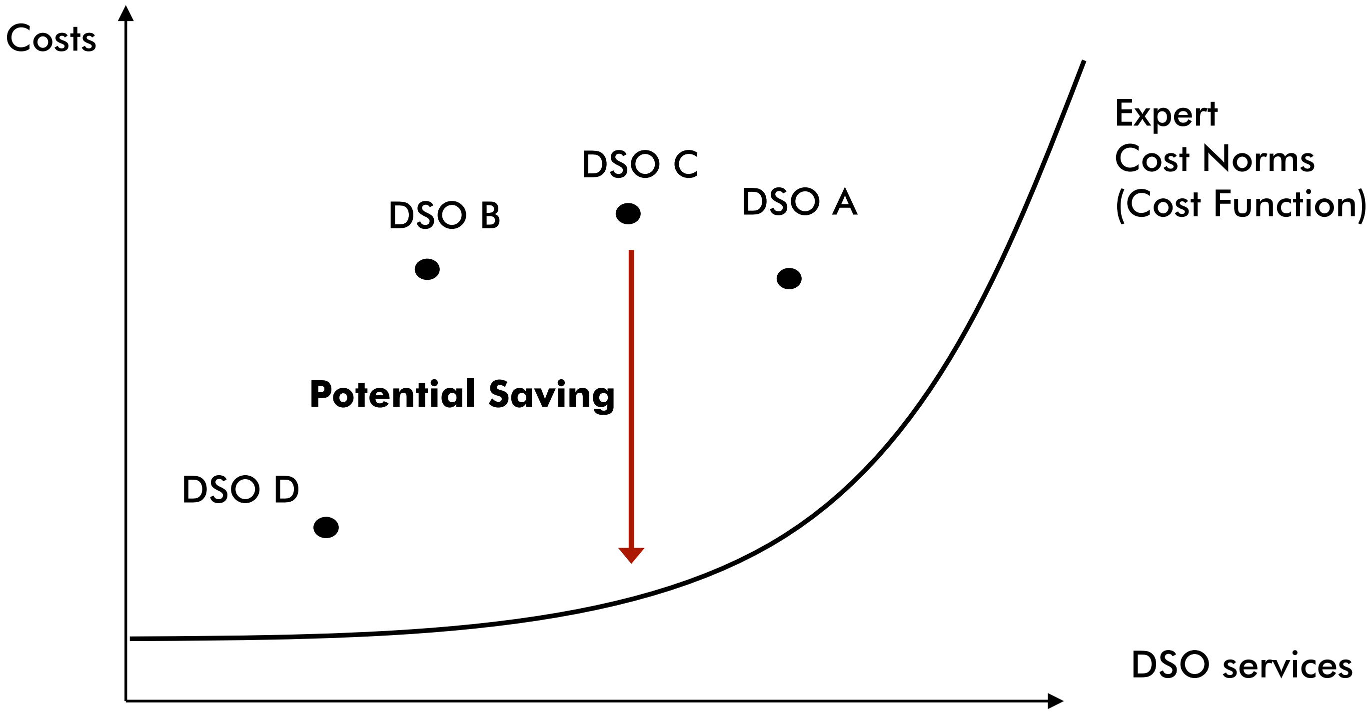
Best-of models

Second stage analysis

Conclusions

Task

Regulator's Ideal World



Real Evaluations

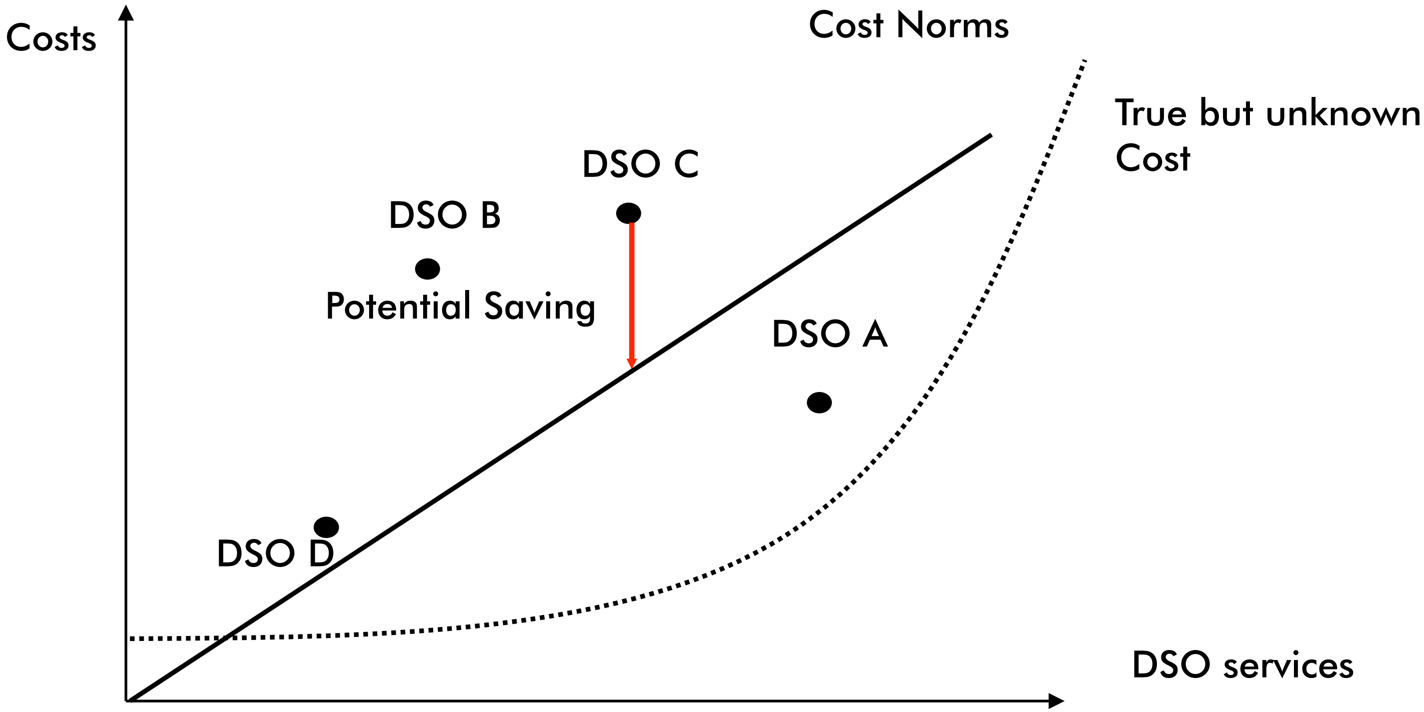
Real evaluations complicated by

- Unknown true underlying cost function
- Multiple inputs (=cost types) and outputs (=grid elements
- Different environments

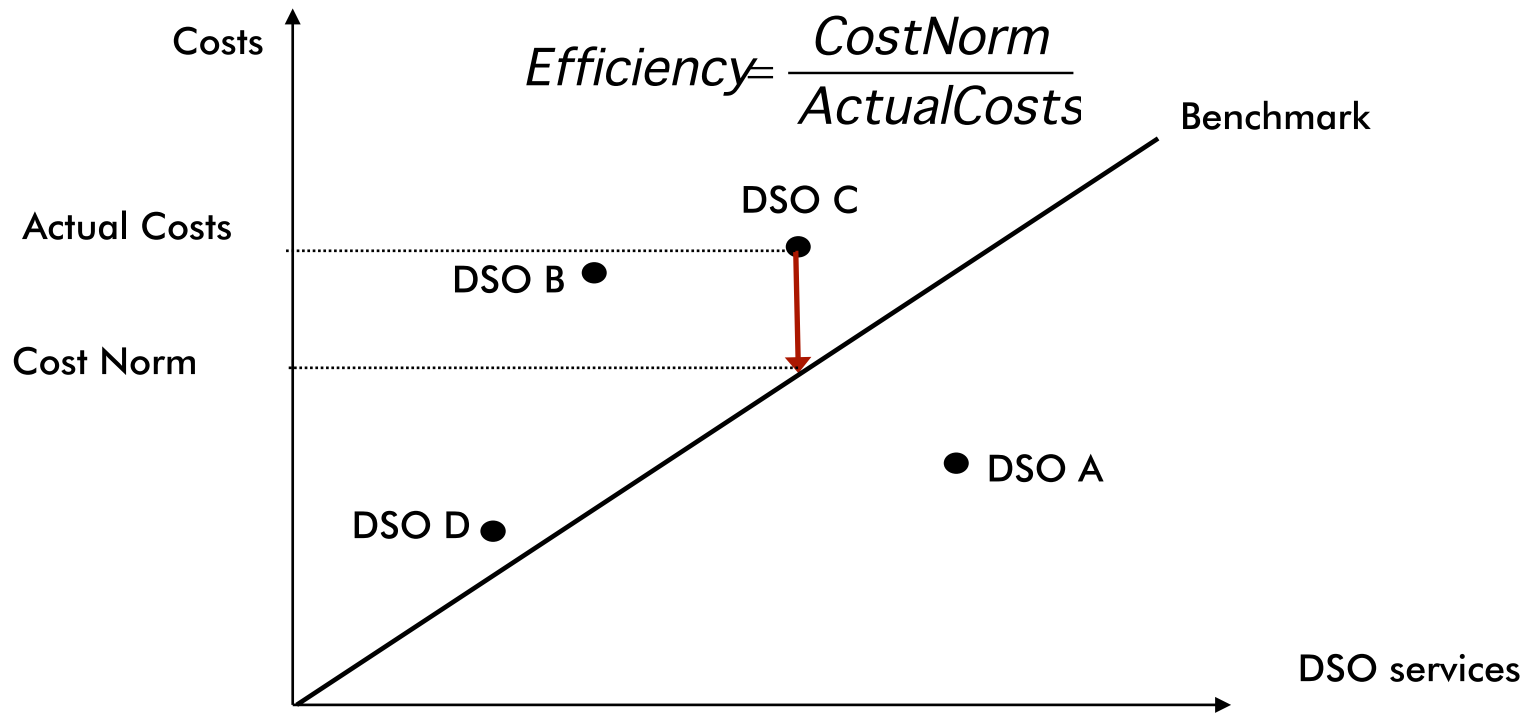
Solutions

- Empirical cost norms /relative performance eval.
- Aggregations
- Sensitivity and robustness analyses

Modelled cost norms



Efficiency



Efficiency = 0.7 suggests: potentially save 30% of present costs

Ordinance ARegV

Find model to determine

- Efficient costs

Certain cost drivers mandatory

- Connections
- Service area
- Pipeline length
- Annual Peakload

Scientific methods

- State of the art in DEA and SFA methods

Specific outlier detection rules

- Two elimination principles in DEA context

Cautious combination

- Best-of-four-models and >= 60%

Best of four (five) approach

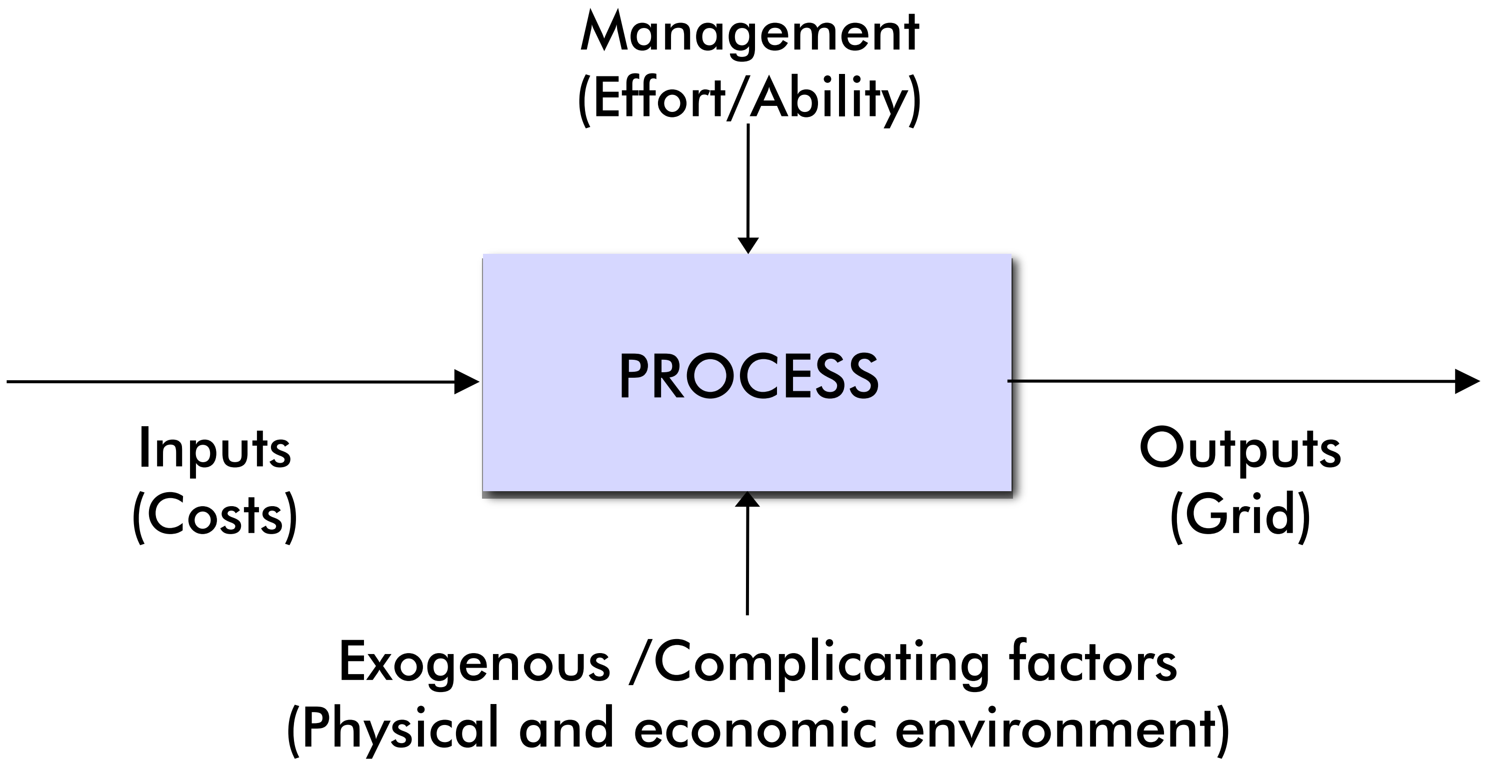
Ordinance

- Best of DEA-NDRS and SFA
- Best of Totex (B) and Standardized Totex (S)
- Minimal efficiency = 0.6

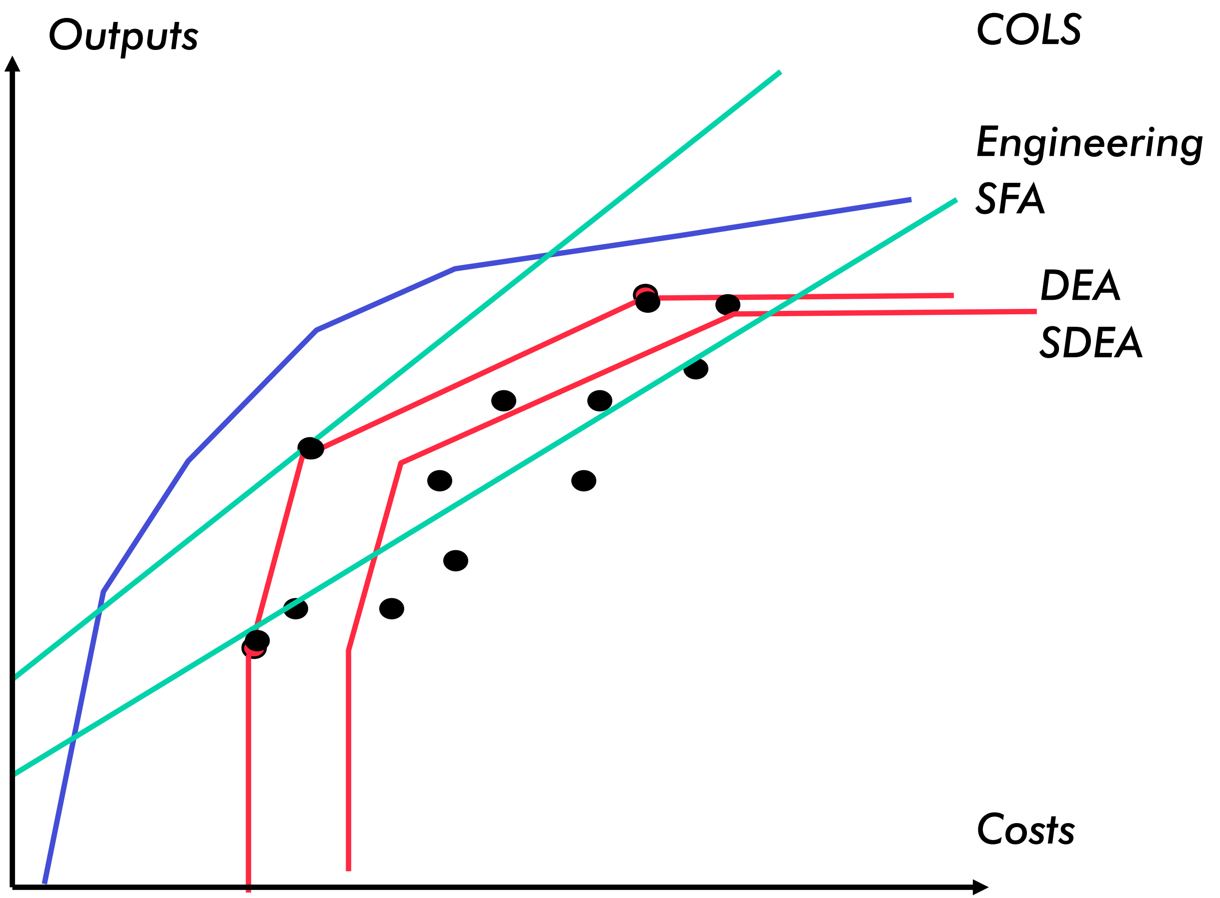
Max[Edea(B),Edea(S),Esfa(B),Esfa(S),0.6]

Methods

Benchmarking system model



Empirical models



Model taxonomy

	Deterministic	*Stochastic*
Parametric	Corrected Ordinary Least Square (COLS) Greene(1997), Lovell(1993), Aigner and Chu(1968)	Stochastic Frontier Analysis (SFA) Aigner, Lovel and Schmidt (1977), Batesee and Coelli (1992), Coelli, Rao and Battsee (1998)
Non-Parametric	Data Envelopment Analysis (DEA) Charnes, Cooper and Rhodes(1978), Deprins, Simar and Tulkens(1984)	Stochastic Data Envelop-ment Analysis (SDEA) Land, Lovell and Thore(1993), Petersen and Olsen(19xx), Weyman-Jones(2001)

Data Envelopment Analysis DEA

NUMEROUS STUDIES

Charnes, Cooper and Rhodes (1978)
1340 studies (-1999?)

PUBLIC SECTOR

Gas, water, heat, roads, police, firefighters, schools, universities, hospitals, courts, …

PRIVATE SECTOR

Banks, insurance, mail, railroads, industries, farms, service providers, consultancy firms, sports…

DEA Basic Principle

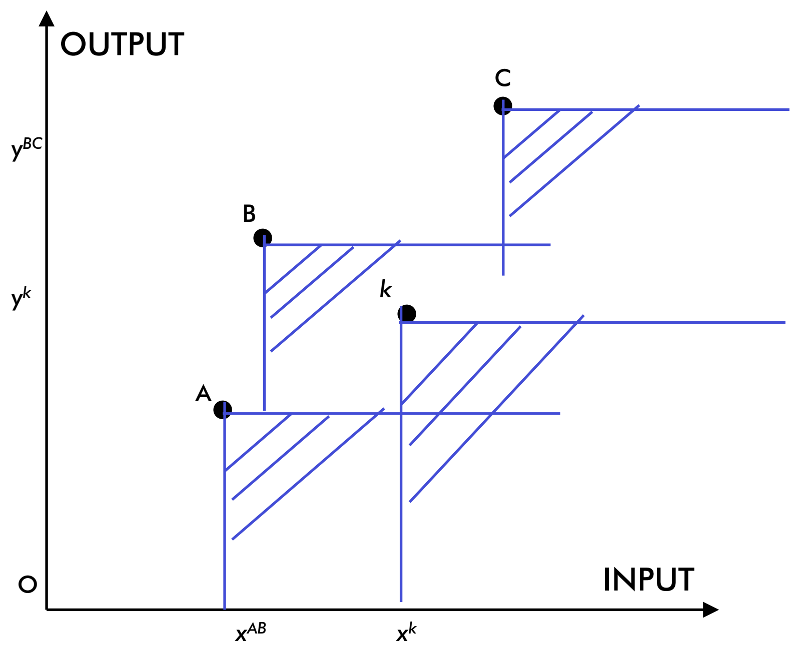
MINIMAL EXTRAPOLATION PRINCIPLE:

T* is <u>smallest</u> subset of that contains data and satisfy certain regularity assumptions

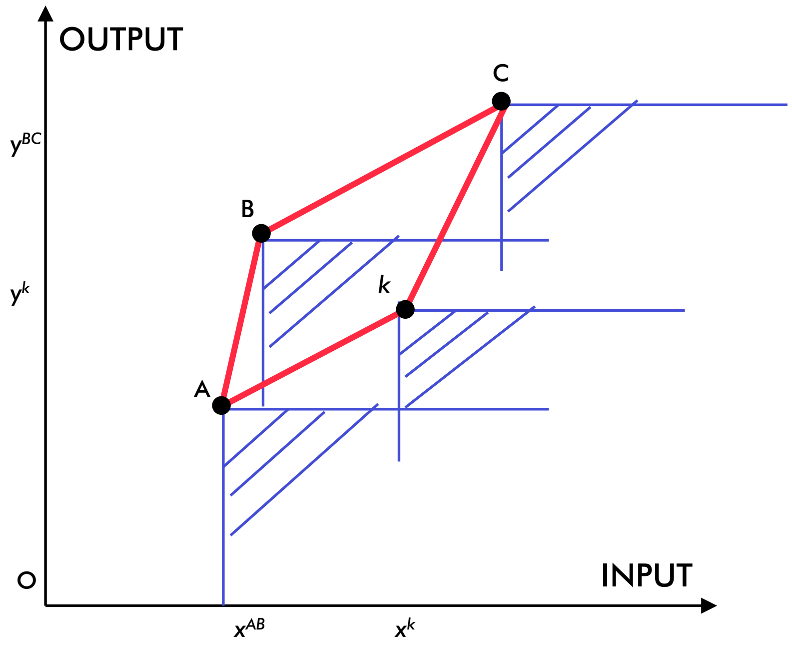
Leads to

- Inner approximation
- Cautious evaluations
- Conservative estimates of improvement potentials

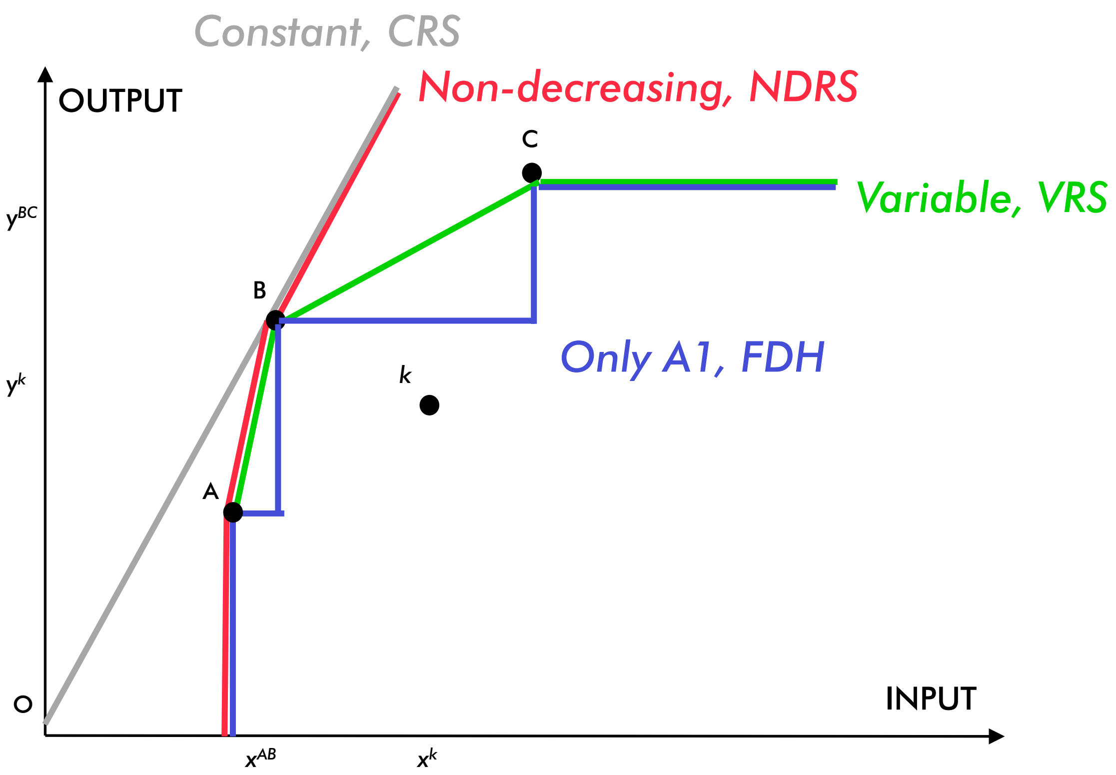
Strong Disposability



Convexity



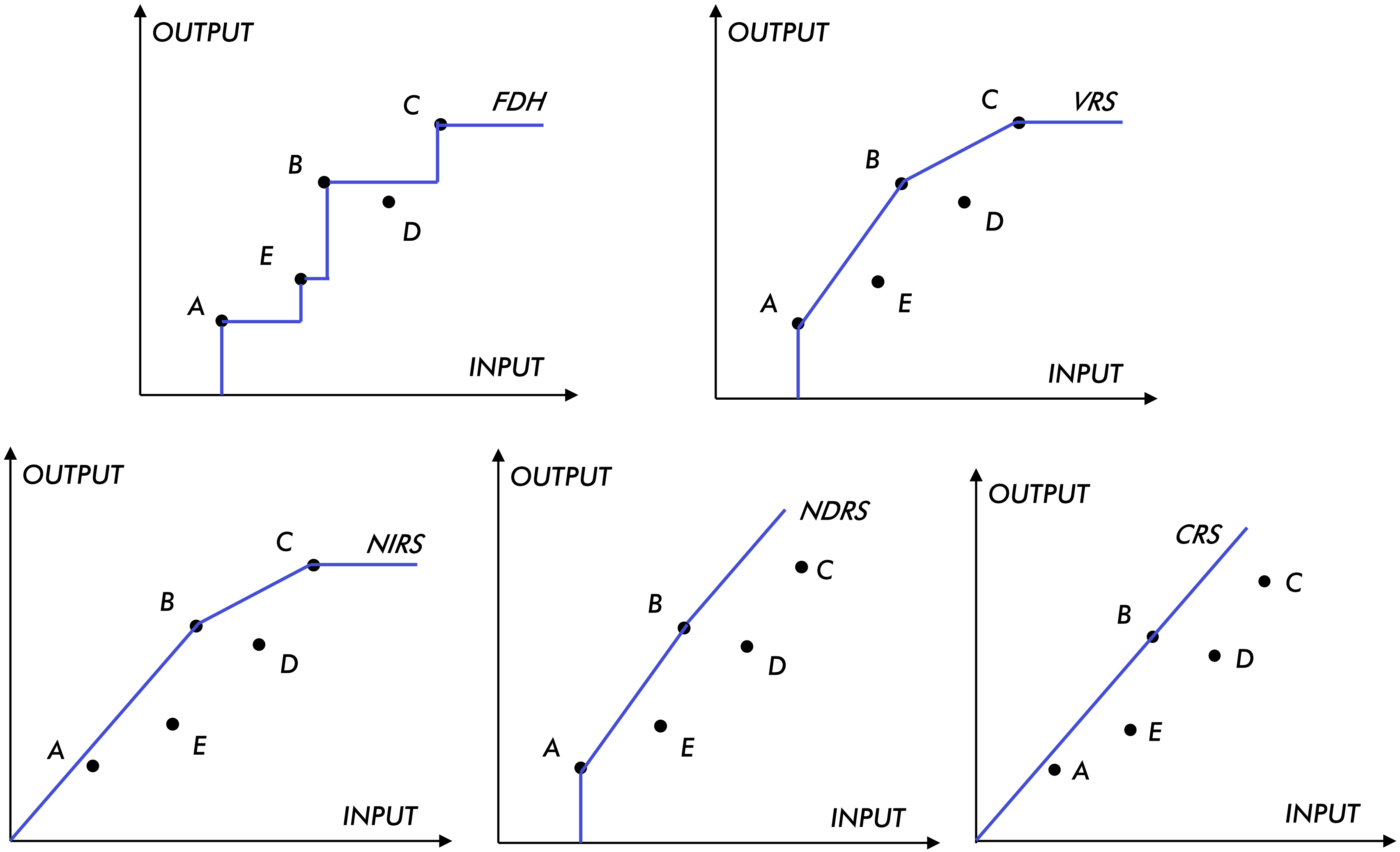
Economies of scale



r returns to scale

$r = \{fdh, crs, vrs, ndrs\}$

Classical DEA models



Pros and Cons of DEA

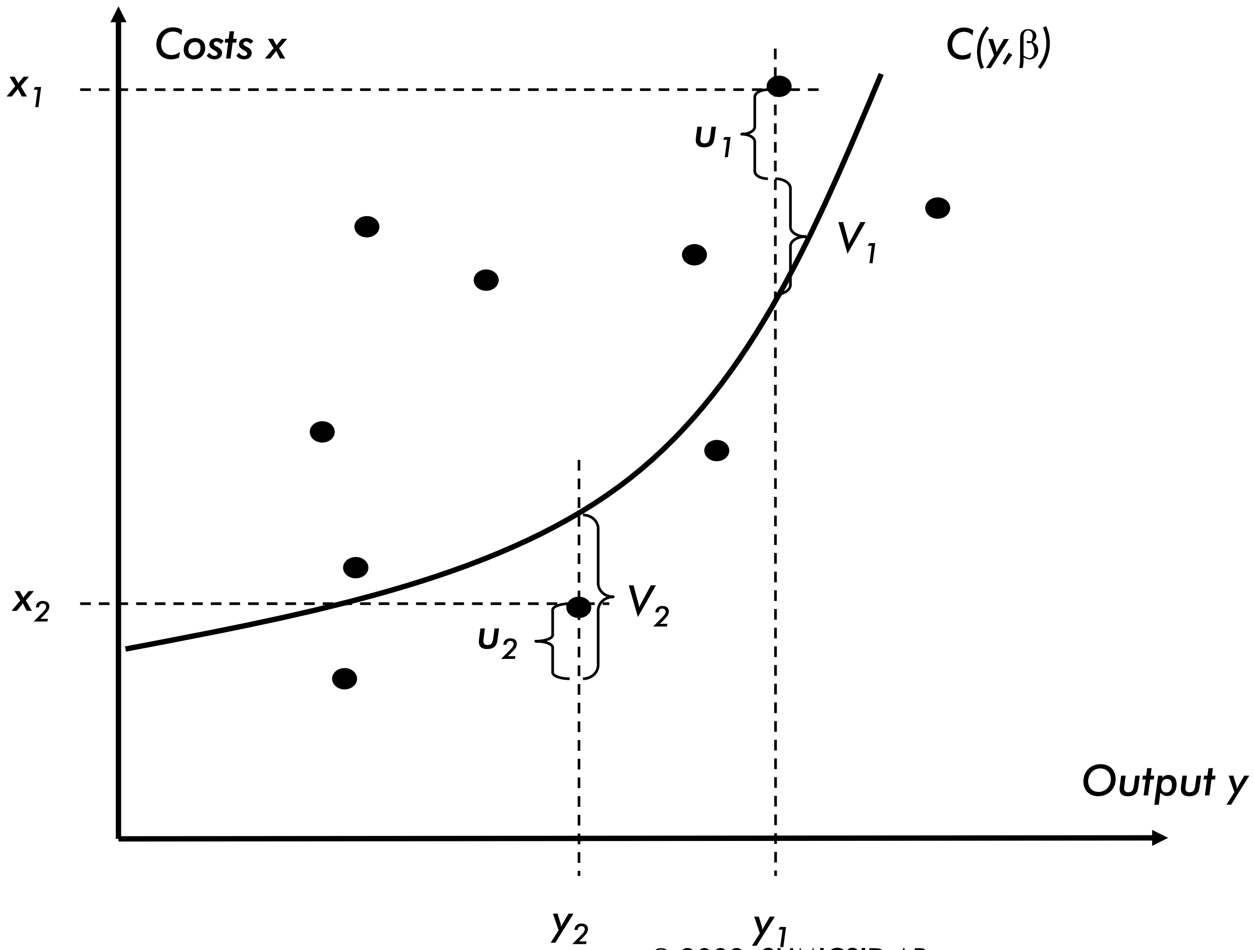
PROS

- Requires no or little price or priority information
- Requires no or little technological information
- Makes weak a priori assumptions
- Handles multiple inputs and multiple outputs
- Provides reel peers
- Identifies best practice
- Cautious or conservative evaluations (minima extrapolation)
- Supports learning, planning, motivation

CONS

- Weaker theory of significance testing (sensitivity, resampling, bootstrapping, asymp theory)

SFA cost function



Basic idea of SFA

Aigner, Lovell and Smith(1977), Meuser and Van den Broeck(1977)

Specification e.g.

$$x_i=\beta_0 + \beta_1 y_{i1} + \beta_2 y_{i2} + .. + \beta_n y_{in} + v_i + u_i$$

where

v_i is normal $N(0, \sigma_v^2)$

u_i is truncated normal $N_+(\mu, \sigma_u^2)$

$\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\gamma = \sigma_u^2 / \sigma^2$

Estimation

– Maximum likelihood estimation

How does it work ?

MLE

- Find the parameter values (beta's and variances) to maximize the likelihood of observing what we observed.

How can we split up deviations in two ?

- Due to asymmetry in distribution of sum of u and v

Pros

- Nice idea: Separate noise and inefficiency
- MLE is consistent (|Estimate-True|->0)
- MLE is asymptotic efficient (Smallest variance)
- Testing etc

Drawbacks

- How to find functional form
- How to justify distribution for inefficiency a priori

Present solutions

Functional form

- Use flexible forms
- Here: Loglinear, Normed linear crs, and Normed linear vrs give almost identical results

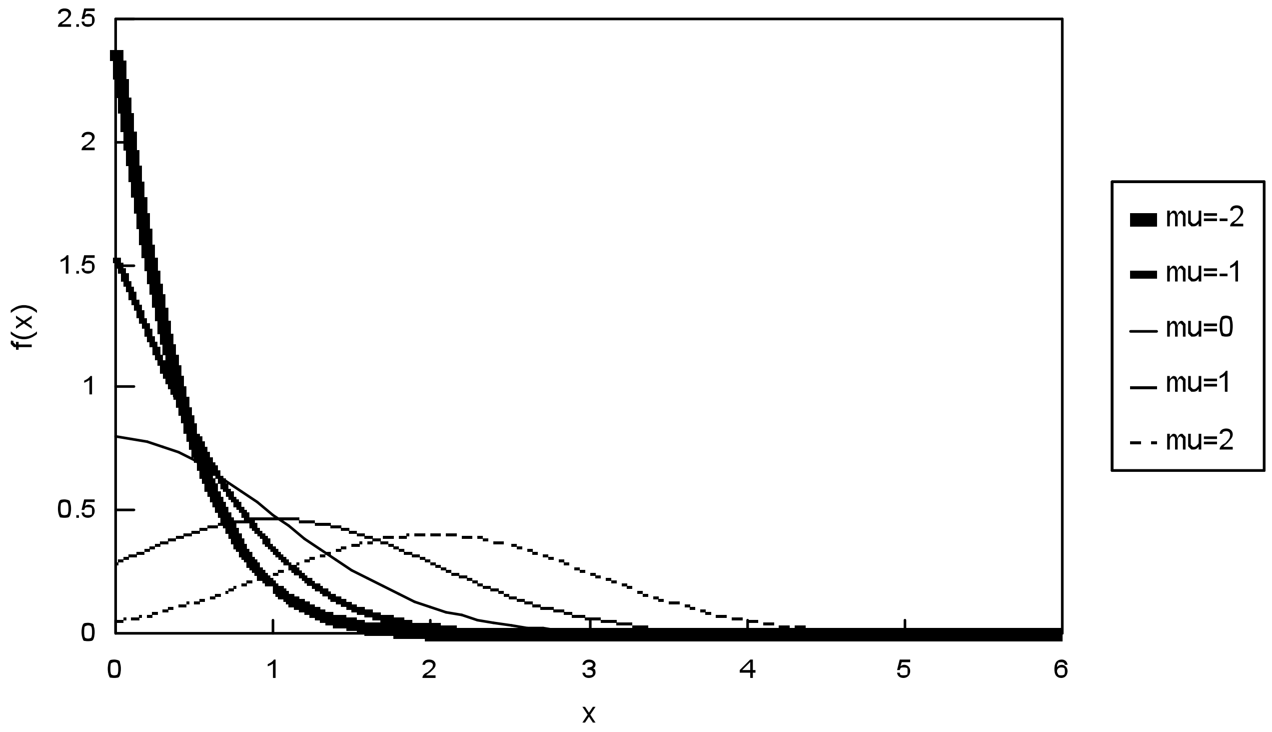
Distribution of inefficiency

- Truncated normal is very flexible

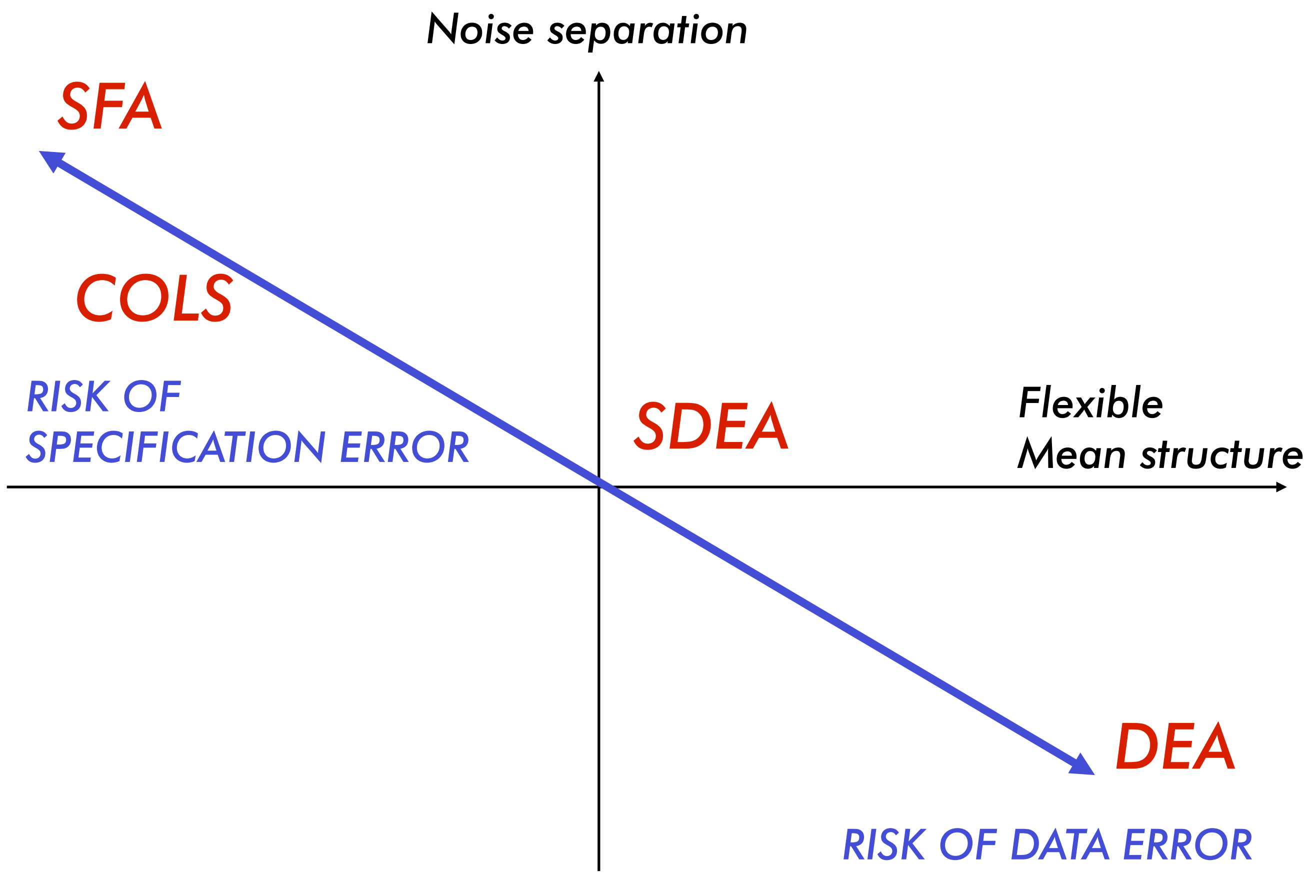
Different normal distributions

Trucated normal in error component model

$$u_i \sim N_+(\mu, \sigma^2)$$



Tradeoffs



Noise separation
SFA
COLS
RISK OF
SPECIFICATION ERROR
SDEA
Flexible
Mean structure
DEA
RISK OF DATA ERROR

Interpretation of the residual

Statistical model

data	=	*Model*	+	*residual*
data	=	*OLS norm*	+	*noise*
data	=	*DEA norm*	+	*inefficiency*
data	=	*SFA norm*	+	*inefficiency* + *noise*

Process

Many elements

Choice of variable standardization

- Accounting standards, cost-allocations, in/out of scope, assets definitions, operating standards etc etc
- Guides, DSO specifics etc

Choice of variable aggregation

- Interest rate, inflation etc
- Expert information, economic theory, index theory etc

Data cleaning and validation

- Ratios etc
- Outlier detection using Cook distance etc, C(X) models etc

Choice of model (variables, parameters, cost-drivers)

- Inputs, outputs, exogenous variables
  Functional form
- Statistical, conceptual reasoning

Model validation

- Second stage analyses

Remaining steps

Outlier detection

- Data cleaning

Average model specification analyses

- Significant cost drivers

Frontier model estimations

- DEA, SFA (form and distributions)

Second stage analyses

- Test for significant omitted variables

Not a linear process

- e.g. average and frontier models -> outliers
- e.g.,average -> frontier -> average -> frontier
- e.g second stage -> frontier

Criteria (1)

The choice of model is a multiple criteria problem!

Conceptual

- The interpretation should be easy
- The properties should be natural (e.g. Normed linear better than loglinear)

Statistical

- Unbiased, consistent and efficient estimates
- Significance testing, hierarchical model development etc
- The parameters with correct sign and significant

Pragmatic / Regulatory

- Access to data
- Robustness
- Adhere to regulatory rules
- Cautious

Criteria (2)

Frontier models

- Statistical theory younger
- Testing and estimation more difficult
  - Data generation process depends on management
  - DEA models non-parametric
- Understand the logic of the models
- Parameters of right sign and preferably significant (frontier versus average models) and important size
- Robust results
- Correlation and levels of efficiency comparable

Basically

- What does it take to get the results published in best international journals
- Cautious – as also in the ordinance

Trade-offs

Best-of-four

- Four different estimations – not necessarily same significant cost drivers
- E.g. pipeline configuration significant in some specifications – include an aggregate.
- Consequence: Bias in favor of companies and at risk of over-specification

Requested variables

- Certain variable groups shall be included
- E.g. pipelines and connections strongly correlated
- Consequence: Signs and significance criteria less strict

Outlier conditons

- Ordinance rather specific
- Stronger than international practice
- Consequence: Bias in favor of companies

Average model specification

Model structure

Controllable resources

X Inputs → DSO → Y Outputs →

Exogenous demand (task)

Totex = Opex + Capex

Transport work
Capacity provision
Service provision

Z Environment

Structural factors

Proxies for
- Geography, climate, soil type,
- Complexity, density
- …

Steps

Large data set

- 200 DMUs
- Approx 250 base variables (parameters X,Y,Z)

Steps towards average model specification

- Make conceptual groups
- Identify most important indicators in different group
- Size of good model
- Good models provide starting point for frontier modeling
- Reduced model – start frontier model, reduce to improve information criterion (AIC, BIC)
  - gives ideas for possible reductions of frontier model
- Extended model – start reduced model, use marginal extensions, reduce to increase significance
  - gives ideas for relevant extensions

Model specification

Selecting too many variables leads to overfitting of the data!

Stepwise procedure

- Forward (start with minimum, select until $p < t$)
- Backward (start with large model, deselect until all sign)
- Sequential (forward, but deselect non-significant var)

Criterion-based procedure

- Penalize models for high number of parameters
- AIC (Akaike Information Criterion: -2loglikelihood+2p, small attractive
- BIC (Bayes Information Criterion: -2loglikelihood+p*log(n), small attractive
- Adjusted R^2 ($1-((n-1)/(n-p))(1-R^2)$)
- Etc

ARegV 13(3-4)

"Output parameters are parameters used to determine supply services and regional features, in particular geographic, geological and topographic characteristics and special structural features of supply brought about by demographic change in the area served. The parameters must be suitable to support robust benchmarking. This is assumed, in particular, when they are measurable or quantifiable, when they cannot be determined by decisions of the operator and are not wholly or partly similar in effect and, in particular, cannot be replicated by other parameters"

"In the first and second regulatory period the regulatory authority must use the following output parameters–

1. the number of connection points in electricity supply networks and the number of offtake points in gas supply networks,

2. the area supplied,

2a. (pipe)line length (system length), and

3. annual peak load. "

ARegV 13(3-4), excerpt

Parameter groups

C(X,Z)	X(Y)	Y	Z
Totex	Pipelines	Peakload	Location
Totex.standard		Energy.del	Coverage
	Net.structure		
	Storage	Area	Population
	Compressors		
	Buffers	Meters	Utilities
	Pressure	Connections	
	Gradtagzahlen		
	Brennwerte		
	Flowcontrol		

Example 2: Compressors

Compressor.out

yCompressor.out.1.hd
yCompressor.out.16.hd
yCompressor.out.70.hd
[4] yCompressor.out.hd
yCompressor.out.SQmax.R.1.hd
yCompressor.out.SQmax.R.16.hd
yCompressor.out.SQmax.R.70.hd
yCompressor.out.SQmax.R.hd
yCompressor.out.SQmax.A.1.hd
yCompressor.out.SQmax.A.16.hd
yCompressor.out.SQmax.A.70.hd
yCompressor.out.SQmax.A.hd
yCompressor.out.SQact.A.1.hd
yCompressor.out.SQact.A.16.hd
yCompressor.out.SQact.A.70.hd
yCompressor.out.SQact.A.hd

Other groups

Compressor.in

Compressor.transport

Compressor.net

Compressor.SQmax

Compressor.SAmax

Example 1: Pipelines

```
yLength.cus.nd
yLength.cus.md
yLength.cus.class.E.16.hd
yLength.cus.class.F.16.hd
yLength.cus.class.G.16.hd
yLength.cus.16.hd
yLength.cus.class.E.70.hd
yLength.cus.class.F.70.hd
yLength.cus.class.G.70.hd
yLength.cus.70.hd
yLength.cus.hd
```

Particular challenges GAS

Coverage potential:

- Potential peakload
- Potential number of connection points
- Potential pipeline length ND

Population changes etc

- Working with multiple measures of growth

Correlation analysis of groups

	zArea.geographic	zArea.supplied
yCompressor.SAmax.R.1.hd	0.922	0.982
yCompressor.SAmax.R.16.hd	0.996	0.952
yCompressor.SAmax.R.70.hd	0.653	0.798
yCompressor.SAmax.R.hd	0.973	0.997

Modelnetwork data

	zPopulation.1995	zPopulation.2000	zPopulation.2005	zPopulation.2006
yEnergy.in	0.819	0.825	0.825	0.825
yEnergy.out	0.813	0.819	0.820	0.820
yEnergy.out.2008	0.798	0.805	0.806	0.806
yMeters.nd	0.849	0.847	0.846	0.846
yMeters.md	0.205	0.206	0.213	0.214
yMeters.hd	0.311	0.302	0.302	0.303

Data material

Rev25

– 187 observations (31/156)

– 184 potential output parameters (Y)
– 26 environmental variables (Z)
– 19 weights/shares/ratios

Relevant indicators in groups

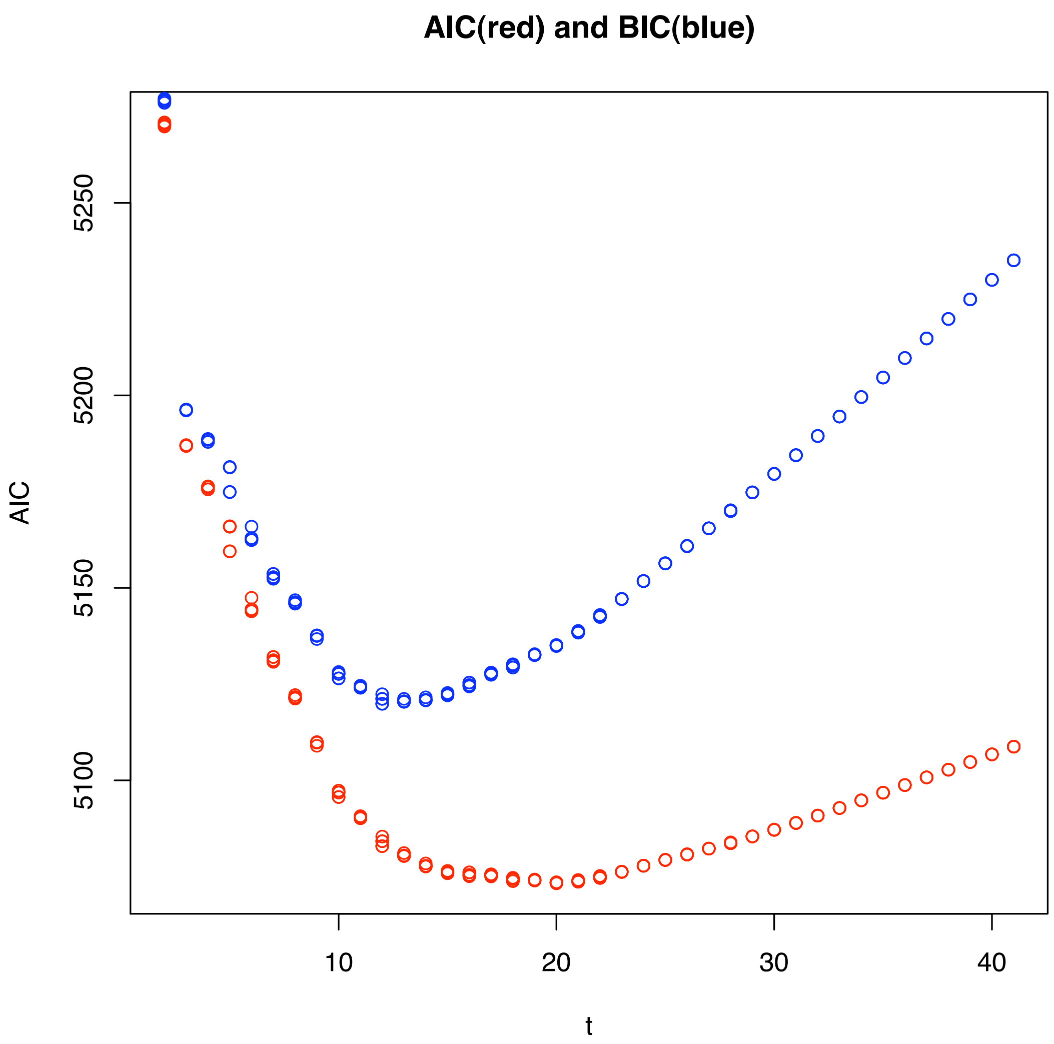
General finding

- Groups have good correlation with costs
- Weak explanatory power for (excluded from primary analysis)
  - Utility (Sewer, Telecom, District heating, Electricity)
  - Compressors (Compressors,,in, Compressor.out, Compressor.net and Compressor.transport)
  - Energy delivered
  - Temperature days
- Many (output) variables are correlated, direct interpretation of coefficients requires analysis of collinearity

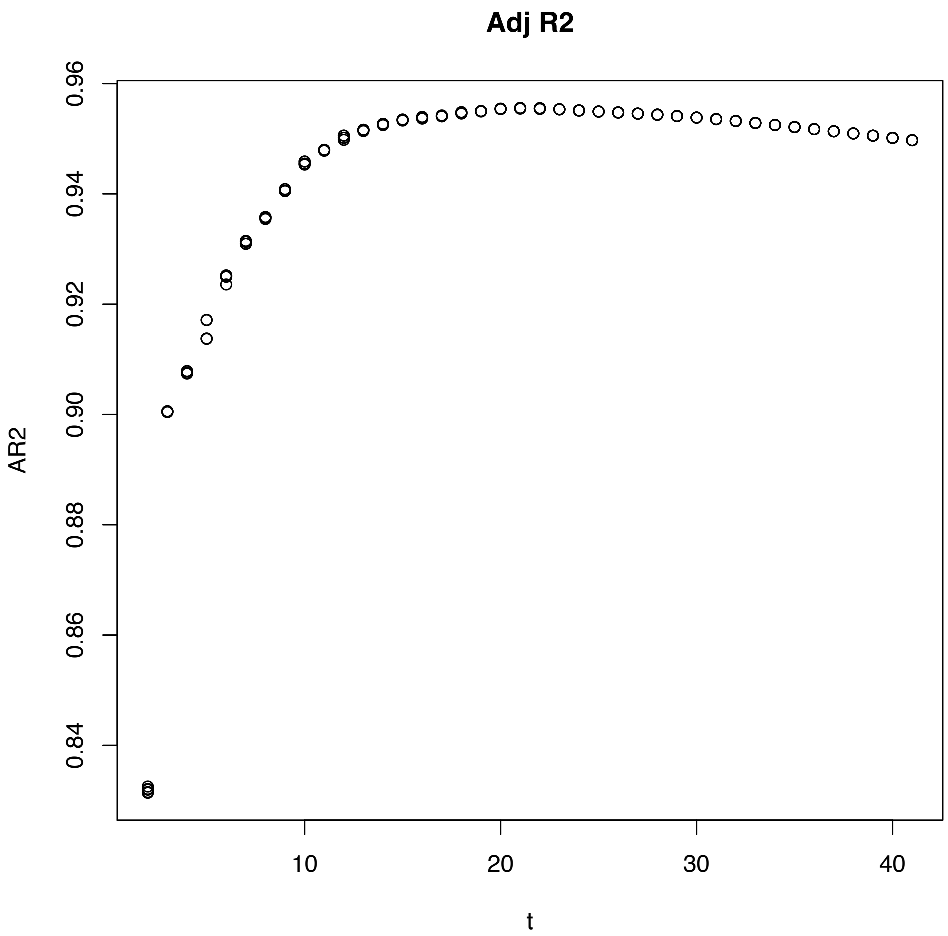
Example

```
Start:  AIC=5897.29
                         Df  Sum of Sq        RSS        AIC F value     Pr(F)
- yPotential.area         1 9.2346e+14 9.8220e+15 5.9140e+03  18.991 2.186e-05 ***
- yPotential.connections  1 8.8073e+15 1.7706e+16 6.0240e+03 181.124 < 2.2e-16 ***
- yPotential.peakload     1 1.0720e+16 1.9618e+16 6.0430e+03 220.458 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

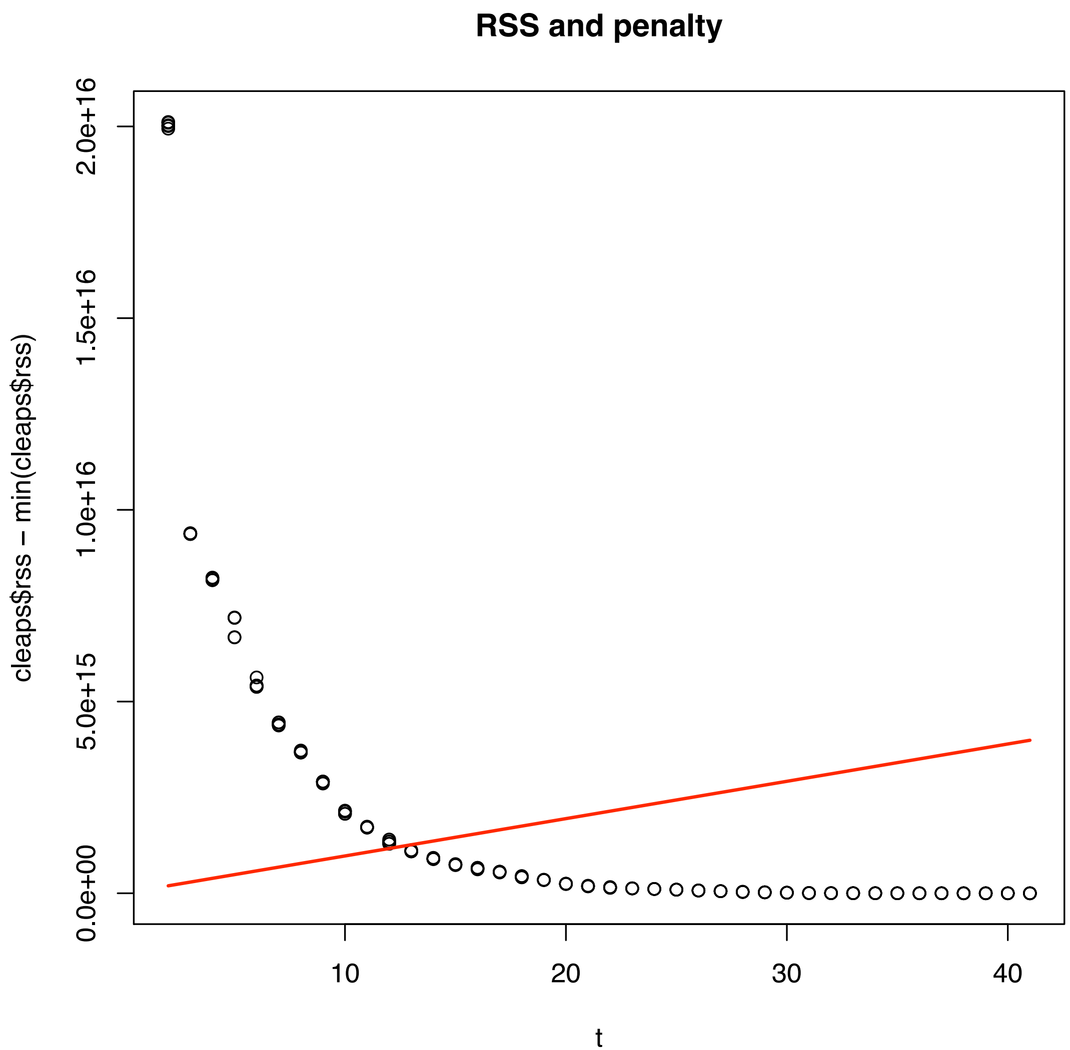
Size of good GAS models



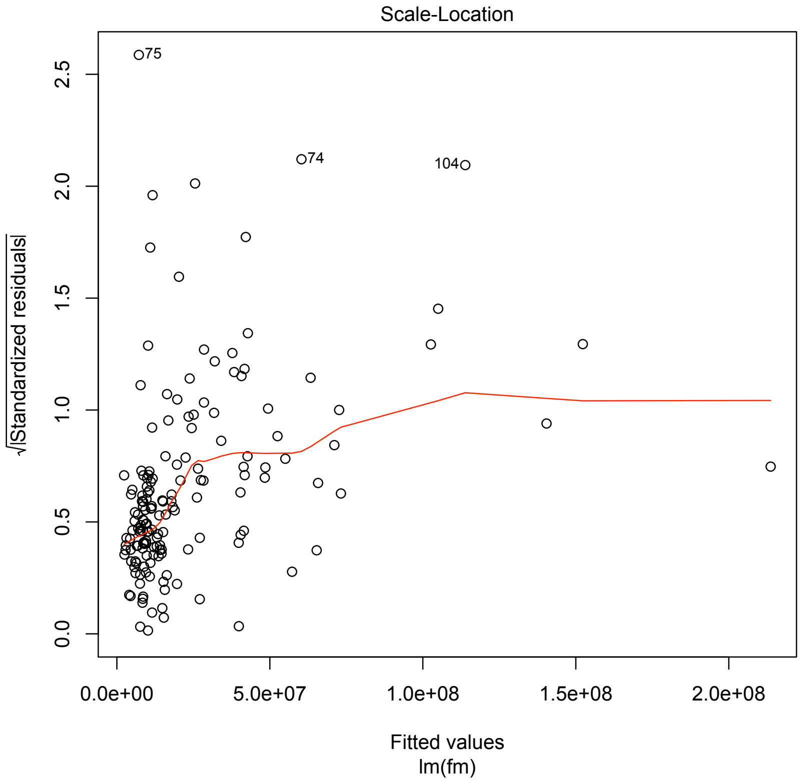
Modelsize vs Adj R2

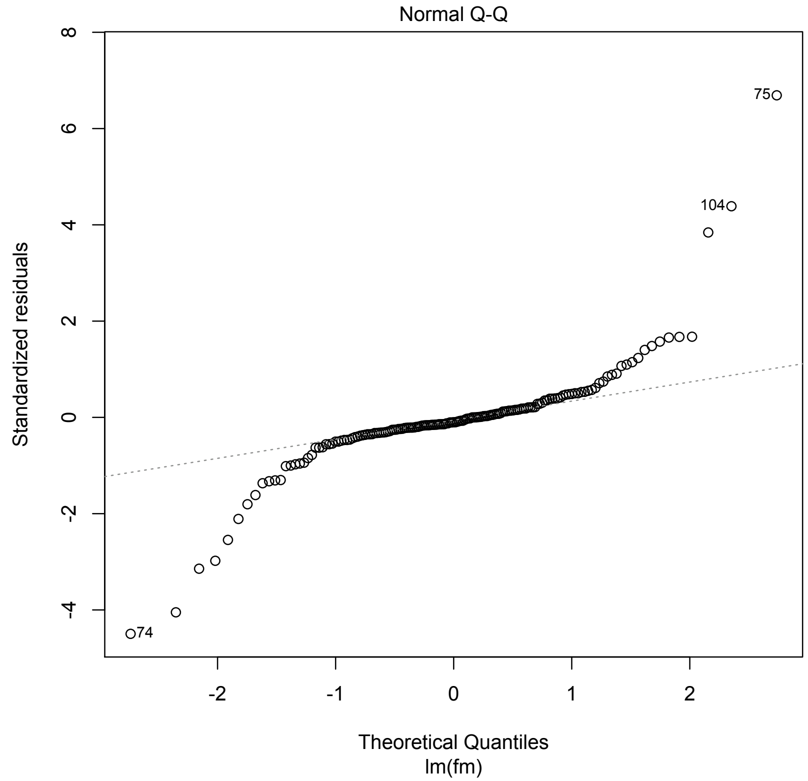


Modelsize: RSS

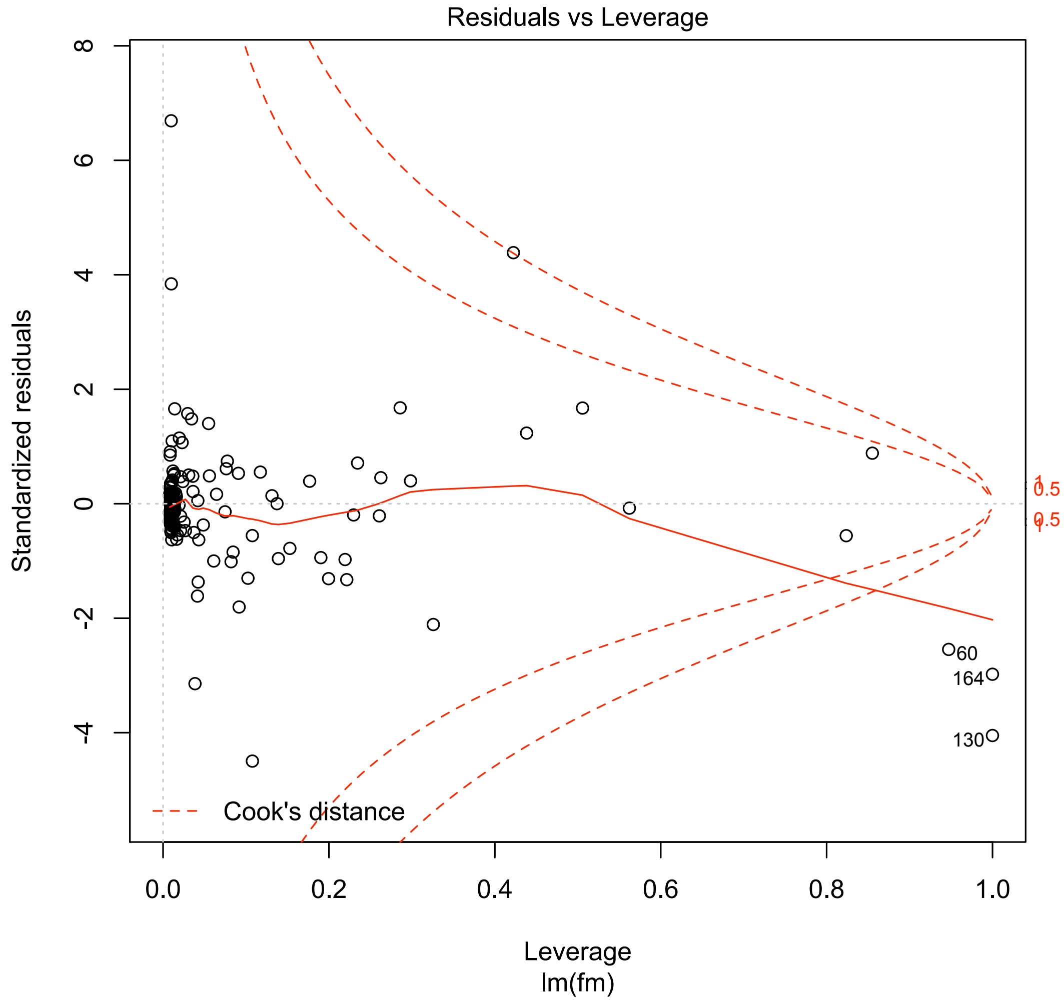


Model fit and outliers





Cooks distance



Steps in extending base models

Procedure

– Base models are selected from groups imposed by ARegV
– Significance is assured for base model
– First stepwise extension of minimal model using BIC
– Then stepwise reduction using significance

Base model

```
Call:
lm(formula = lo)

Residuals:
      Min        1Q    Median        3Q       Max
-18850146  -1884601   -648676   1450513  25344033

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    1.328e+06  5.047e+05   2.631  0.00925 **
yConnection.out.tot            4.147e+02  6.941e+01   5.974 1.23e-08 ***
yPeakload.out                  2.103e+01  1.911e+01   1.101  0.27249
yArea.supplied                 3.638e+03  2.319e+03   1.569  0.11853
yPipelength.lower5bar.inclHAL -5.059e+03  5.008e+02 -10.102  < 2e-16 ***
yPipelength.abover5bar        -5.449e+02  1.571e+03  -0.347  0.72908
yPotential.connections        -6.822e-01  1.322e+01  -0.052  0.95889
yPotential.peakload            2.717e+01  1.359e+01   1.999  0.04710 *
yVolume.tot.cus                1.869e+02  1.730e+01  10.801  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5048000 on 178 degrees of freedom
Multiple R-squared: 0.9731,             Adjusted R-squared: 0.9719
F-statistic: 806.3 on 8 and 178 DF,  p-value: < 2.2e-16
BIC = 6346
```

Minimal Base model

```
Call:
lm(formula = lo)

Residuals:
      Min        1Q    Median        3Q       Max 
-18733436  -1881994   -648692   1415629  25521894 

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)    
(Intercept)                    1.369e+06  4.866e+05   2.814  0.00544 ** 
yConnection.out.tot            4.086e+02  3.182e+01  12.843  < 2e-16 ***
yPeakload.out                  2.175e+01  9.617e+00   2.261  0.02495 *  
yArea.supplied                 3.337e+03  2.152e+03   1.551  0.12267    
yPipelength.lower5bar.inclHAL -4.997e+03  4.476e+02 -11.164  < 2e-16 ***
yPipelength.abover5bar
yPotential.connections
yPotential.peakload            2.616e+01  5.366e+00   4.875 2.38e-06 ***
yVolume.tot.cus                1.868e+02  1.711e+01  10.914  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 5021000 on 180 degrees of freedom
Multiple R-squared: 0.9731,            Adjusted R-squared: 0.9722 
F-statistic:  1086 on 6 and 180 DF,  p-value: < 2.2e-16 
BIC = 6336
```

Extension of minimal model

```
B.5 Maximum significant model (backward stepwise reduction from Bplus)

                                            Estimate Std. Error t value Pr(>|t|)
Df   Sum of Sq         RSS         AIC     F Value      Pr(F)
<none>                                   4.5351e+15 5.7810e+03
yMeters.tot                 1 1.7068e+15 2.8283e+15 5.6950e+03 1.0700e+02 < 2.2e-16 ***
yPeakload.in                1 1.1085e+12 4.5340e+15 5.7830e+03   0.043273    0.8355
zEast_west                  1 4.6731e+13 4.4883e+15 5.7810e+03 2.0000e+00    0.1763
zCoverage.conn.fraction     1 1.8531e+13 4.5165e+15 5.7820e+03 1.0000e+00    0.3953
zCoverage.actual            1 1.5153e+12 4.5336e+15 5.7830e+03   0.059162    0.8081
zCoverage.potential         1 4.4152e+13 4.4909e+15 5.7810e+03 2.0000e+00    0.1888
zElectricity.nets           1 7.3933e+11 4.5343e+15 5.7830e+03   0.028860    0.8653
zSewer.nets                 1 1.0080e+11 4.5350e+15 5.7830e+03   0.003934    0.9501
zDistrictheating.nets       1 2.1709e+11 4.5349e+15 5.7830e+03   0.008473    0.9268
zTelecom.net                1 2.6062e+13 4.5090e+15 5.7820e+03 1.0000e+00    0.3132
zTemp.avg.day               1 3.0891e+13 4.5042e+15 5.7820e+03 1.0000e+00    0.2721
yVolume.nd                  1 8.6522e+14 3.6698e+15 5.7440e+03 4.2000e+01 9.794e-10 ***
yVolume.md                  1 5.5499e+14 3.9801e+15 5.7590e+03 2.5000e+01 1.586e-06 ***
yVolume.hd                  1 4.1878e+12 4.5309e+15 5.7830e+03   0.163597    0.6864
yVolume.cus                 1 3.4476e+13 4.5006e+15 5.7820e+03 1.0000e+00    0.2458
yVolume.tot                 1 3.4476e+13 4.5006e+15 5.7820e+03 1.0000e+00    0.2458
yPotential.area             1 3.8282e+12 4.5312e+15 5.7830e+03   0.149538    0.6994
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Maximal OLS model

Vergleichsparameter	Vergleichsparameter (deutsche Übersetzung)
yArea.geographic	Geographische Fläche in km
yArea.supplied	Versorgte Fläche in km
yPeakload.out	Zeitgleiche Jahreshöchstlast aller Ausspeisungen in mn^3/h
yPipelength.value.steered.lower5bar	Netzlänge Wert unter 5 bar
yPipelength.value.steered.above5bar	Netzlänge Wert über 5 bar
yPipelength.pressure.steered.lower5bar	Netzlänge Druck unter 5 bar
yPipelength.pressure.steered.above5bar	Netzlänge Druck über 5 bar
yPipelength.abover5bar	Netzlänge über 5 bar
yPop.2006	Bevölkerung im Jahre 2006
yMeters.nd	Anzahl der Messstellen ND
yMeters.md	Anzahl der Messstellen MD
yMeters.5.hd	Anzahl der Messstellen HD ≤ 5 bar
yMeters.16.hd	Anzahl der Messstellen HD ≤ 16 bar
yMeters.70.hd	Anzahl der Messstellen HD ≤ 70 bar
yVolume.nd	Ausgespeiste Jahresarbeit ND in mn3
yVolume.hd	Ausgespeiste Jahresarbeit HD mn3
yPotential.connections	Potentielle Ausspeisepunkte
yPotential.peakload	Potentielle zeitgleiche Jahreshöchstlast
yConnection.out.1.hd	Anzahl Ausspeisepunkte ≤ 5 bar
yConnection.out.16.hd	Anzahl Ausspeisepunkte ≤ 16 bar
yConnection.out.70.hd	Anzahl Ausspeisepunkte ≤ 70 bar
yConnection.out.hd	Anzahl Ausspeisepunkte HD

Application

Econometric cost function can be found with other parameters than those in ARegV

The imposed parameter groups will lead to slightly overparametrized models, in particular with respect to area and pipelines

However, the variable selection phase here is not intended to result in an empirical cost function, as the sample is assumed to have inefficiency in the observations.

It provides guidance to the frontier modeling

Frontier model specification

Final GAS model

Input

- Total costs (xTotex or xTotex.standard)

Output

- yArea.supplied
- yConnections.out.tot
- yPipelength.lower5bar.inclHAL
- yPipelength.above5bar
- yVolume.tot.cus
- yPotential.connections
- yPeakload.out
- yPotential.peakload
- yPop.1995
- yPop.2006

Scale

- yConnections.out.tot (if used)

Cost drivers

Bezeichnung in den Modellen	Deutsche Übersetzung	Aufgabenbereich
Input		
xTotex	Aufwandsparameter	
Exogene Output- und Vergleichsparameter		
yConnection.out.tot	Anzahl aller Ausspeisepunkte an Letztverbraucher	Versorgung der Kunden
yArea.supplied	Versorgte Fläche	Versorgung der Kunden
yPeakload.out	Zeitgleiche Jahreshöchstlast aller Ausspeisungen	Versorgung der Kunden
yPotential.connections	Potentielle Anzahl aller Ausspeisepunkte	Versorgung der Kunden
yPotential.peakload	Potentielle Zeitgleiche Jahreshöchstlast aller Ausspeisungen	Versorgung der Kunden
yPipelength.lower5bar.inclHAL	Netzlänge (Druckstufe kleiner als 5 bar)	Versorgung der Kunden
yPipelength.abover5bar	Netzlänge (Druckstufe größer als 5 bar)	Versorgung der Kunden
yVolume.tot.cus	Volumen	Versorgung der Kunden
yPop.1995	Bevölkerung 1995	Bereitstellung von Kapazität
yPop.2006	Bevölkerung 2006	Bereitstellung von Kapazität
Normierungsvariable		
yConnection.out.tot	Anzahl aller Ausspeisepunkte an Letztverbraucher	Versorgung der Kunden

Main models

Modell	Beschreibung	Bezeichnung in den Ergebnistabellen
DEA Modell 1	DEA Modell mit NDRS mit Ausreißern	D_dea_far_ndrs
DEA Modell 2	DEA Modell mit NDRS sowie Annahme der Supereffizienz zur Ermittlung der Ausreißer	D_dea_sup_far_ndrs
DEA Modell 3	DEA Modell NDRS unter Ausschluss der Ausreißer	D_dea_far_ndrs_ex_out
DEA Modell 4	DEA Modell mit NDRS korrigiert mit dem durch Bootstrapping ermittelten Bias	D_dea_far_ndrs_biascorr
DEA Modell 5	DEA Modell mit NDRS untere Grenze des durch Bootstrapping ermittelten Konfidenzintervalls	D_dea_far_ndrs_biascorr_c1
DEA Modell 6	DEA Modell mit NDRS obere Grenze des durch Bootstrapping ermittelten Konfidenzintervalls	D_dea_far_ndrs_biascorr_c2
DEA Modell 7	Skaleneffizienz (Verhältnis der Effizienzwerte unter Annahme von CRS und VRS)	D_dea_far_se
SFA Modell 1	SFA Modell linearer funktionaler Zusammenhang mit normalisierten Daten (CRS Annahme)	D_sfa_normedlinear_crs_far

Efficiency scores Totex

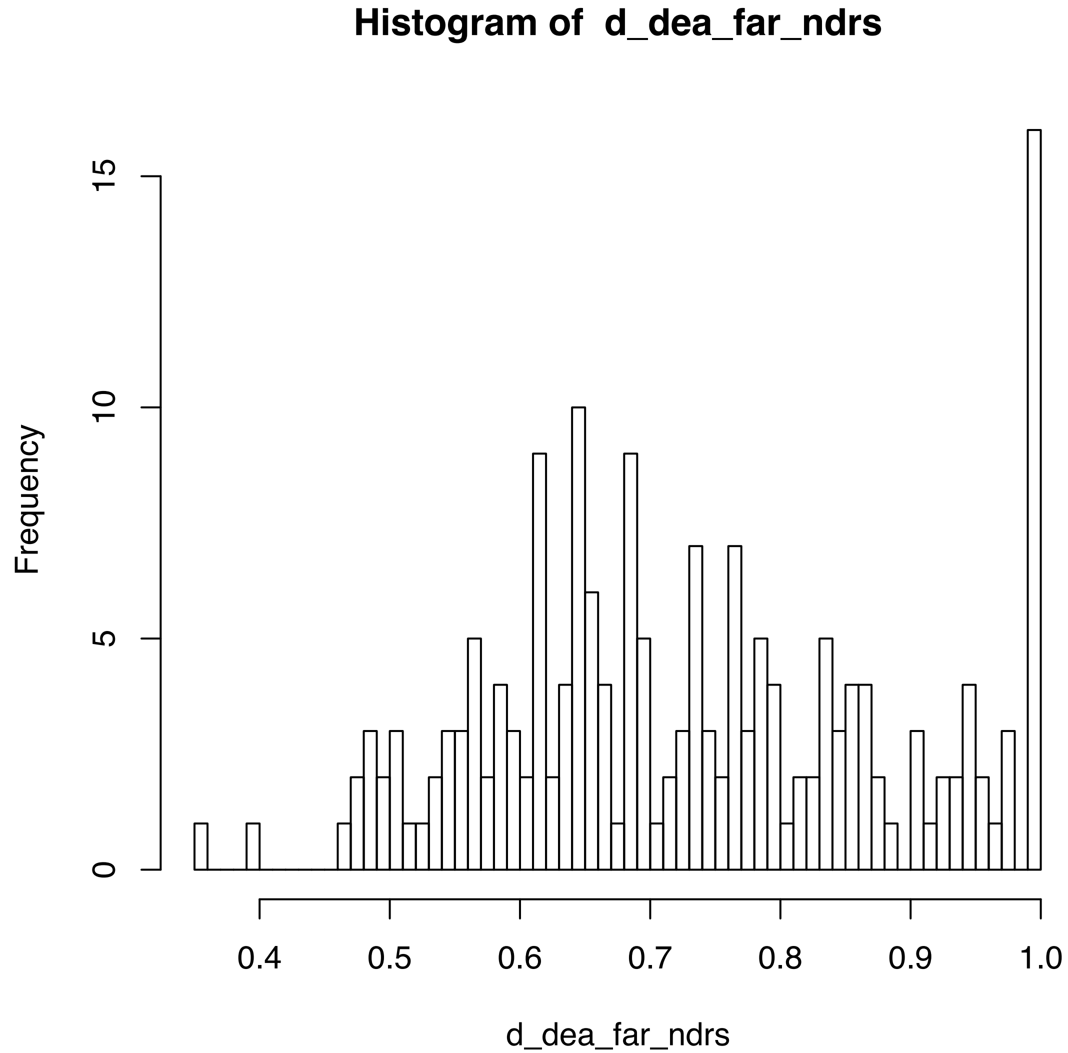
Modelle mit Aufwandsparametern	Mittelwert	Standard-abweichung	Anzahl der Unternehmen, die einen Effizienzwert kleiner als 0.6 aufweisen	Minimum	Anzahl der Unternehmen, die auf der Effizienzgrenze liegen (≥ 1)
DEA Modell 1	0.732	0.153	37	0.352	16
DEA Modell 2	0.755	0.233	37	0.352	16
DEA Modell 3	**0.778**	**0.151**	**22**	**0.382**	**25**
DEA Modell 4	0.665	0.127	61	0.335	0
DEA Modell 5	0.625	0.116	77	0.326	0
DEA Modell 6	0.721	0.150	39	0.346	0
DEA Modell 7	0.932	0.065	0	0.675	16
SFA Modell 1	**0.836**	**0.079**	**4**	**0.536**	**0**

Efficiency summary Totex.standard

Modelle mit Aufwandsparametern nach Vergleichbarkeits-prüfung	Mittelwert	Standard-abweichung	Anzahl der Unternehmen, die einen Effizienzwert kleiner als 0.6 aufweisen	Minimum	Anzahl der Unternehmen, die auf der Effizienzgrenze liegen (≥ 1)
DEA Modell 1	0.724	0.158	45	0.370	15
DEA Modell 2	0.745	0.229	45	0.370	15
DEA Modell 3	**0.771**	**0.157**	**29**	**0.393**	**28**
DEA Modell 4	0.658	0.131	64	0.353	0
DEA Modell 5	0.618	0.119	90	0.343	0
DEA Modell 6	0.713	0.156	50	0.365	0
DEA Modell 7	0.937	0.070	0	0.639	16
SFA Modell 1	**0.824**	**0.085**	**4**	**0.530**	**0**

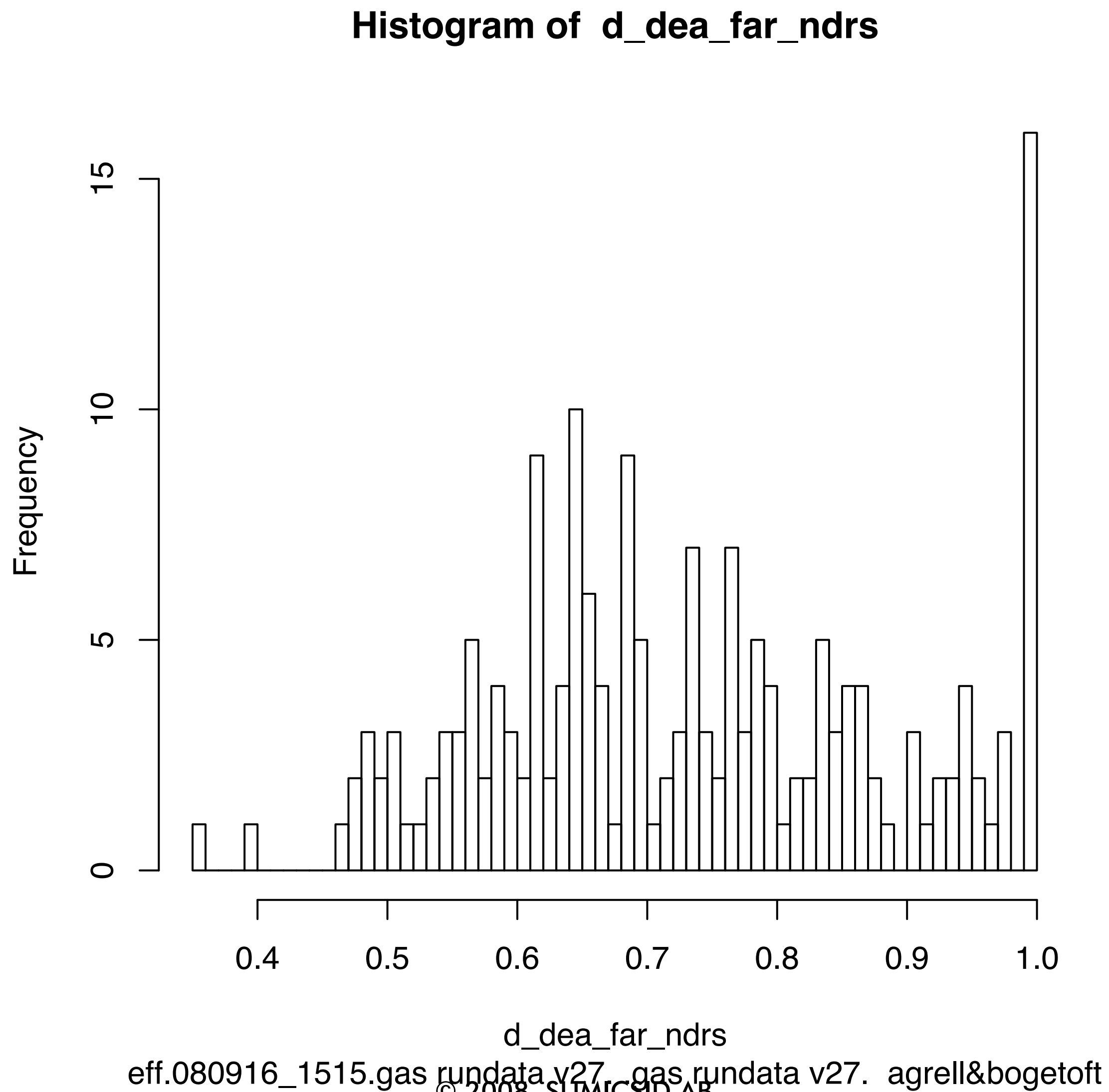
Efficiency distributions
(Illustrated with xTotex.standard version)

Distribution DEA NDRS

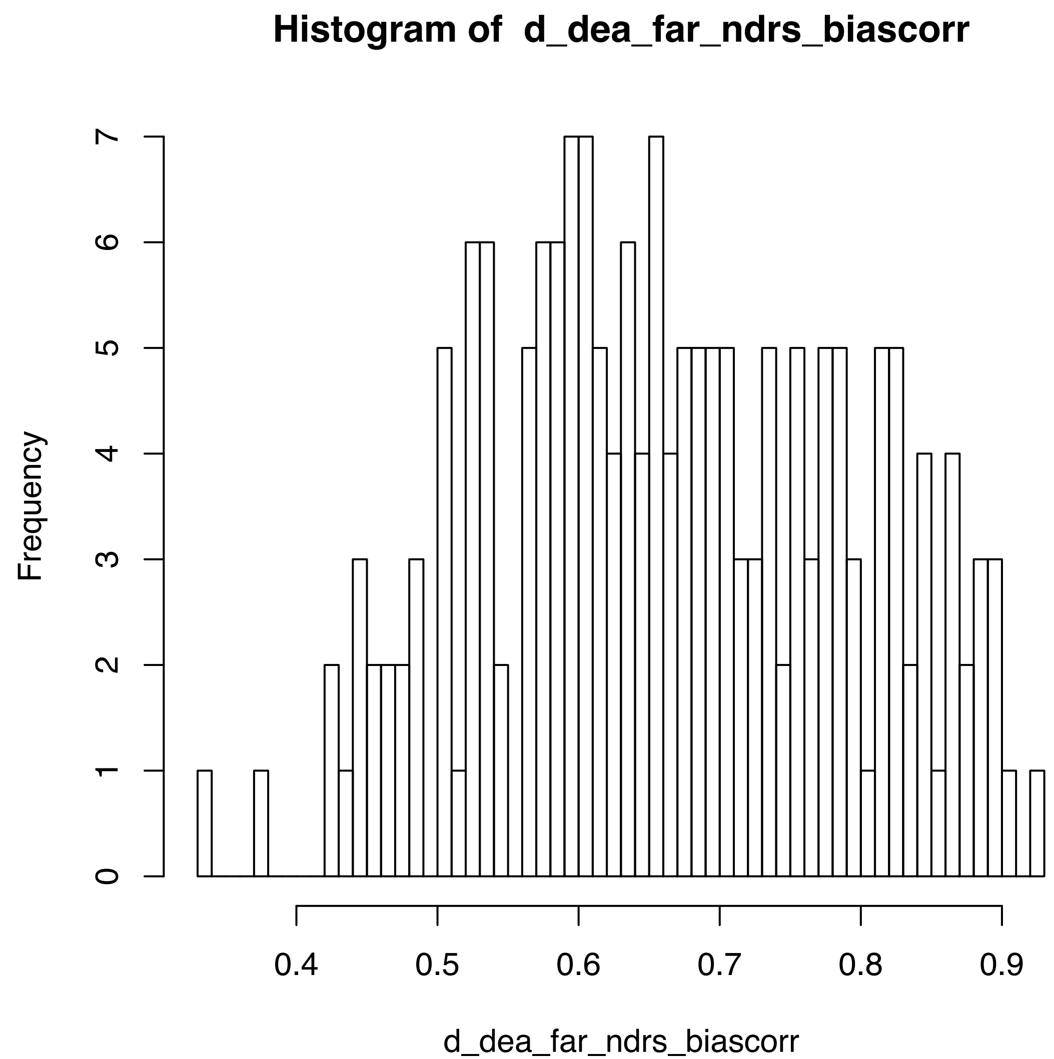


eff.080916_1515.gas rundata v27. .gas rundata v27. agrell&bogetoft

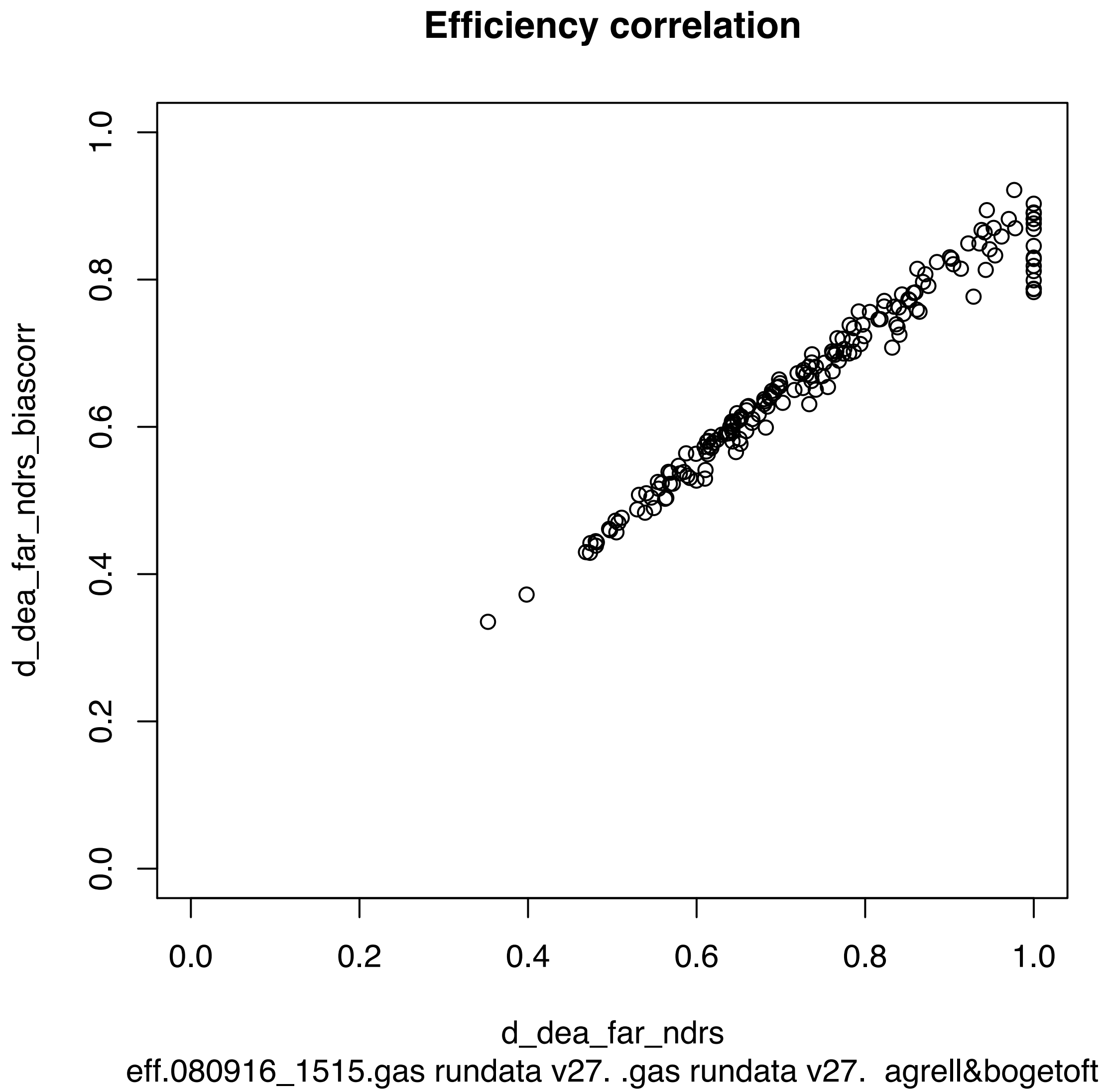
Distribution DEA NDRS



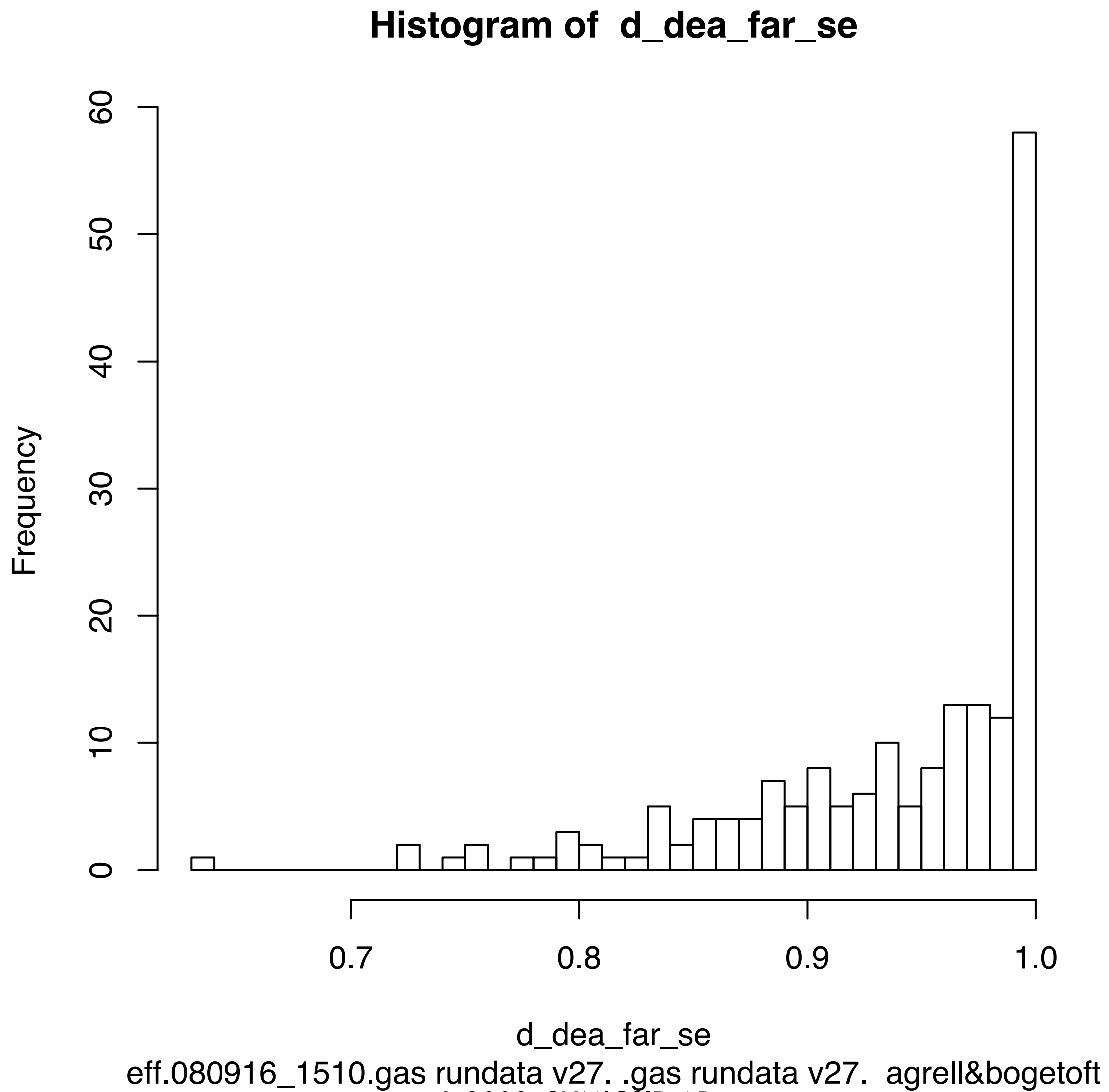
Bias corrected DEA NDRS



Role of bias correction



Distribution SE



Outliers a la Ordinance

No single unit have a large general impact

$$\sum_{k \in I \setminus i} (E(k; I \setminus i) - 1))^2 / \sum_{k \in I \setminus i} (E(k; I) - 1)^2$$

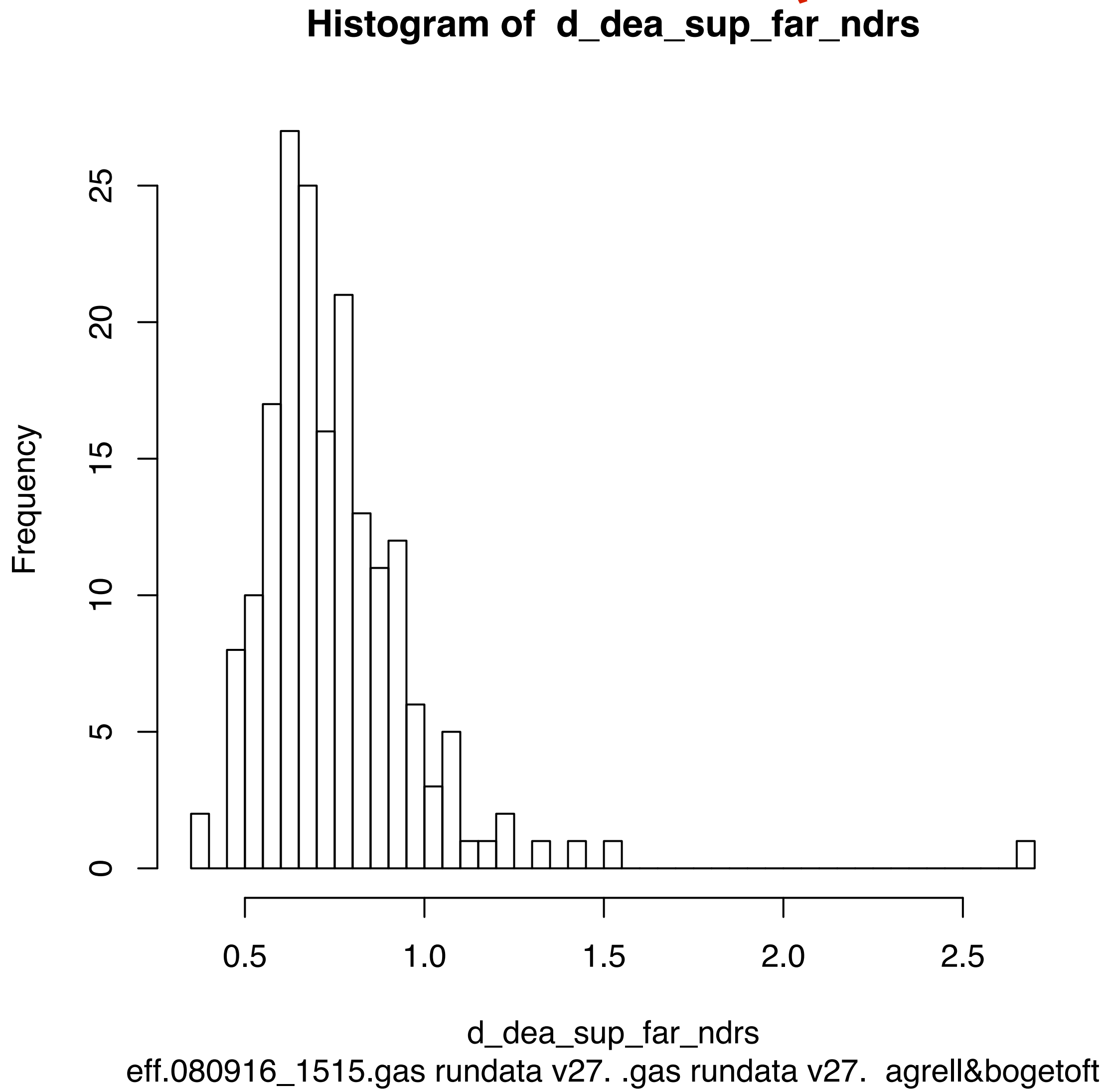
Not too large

No units are extremely super-efficient

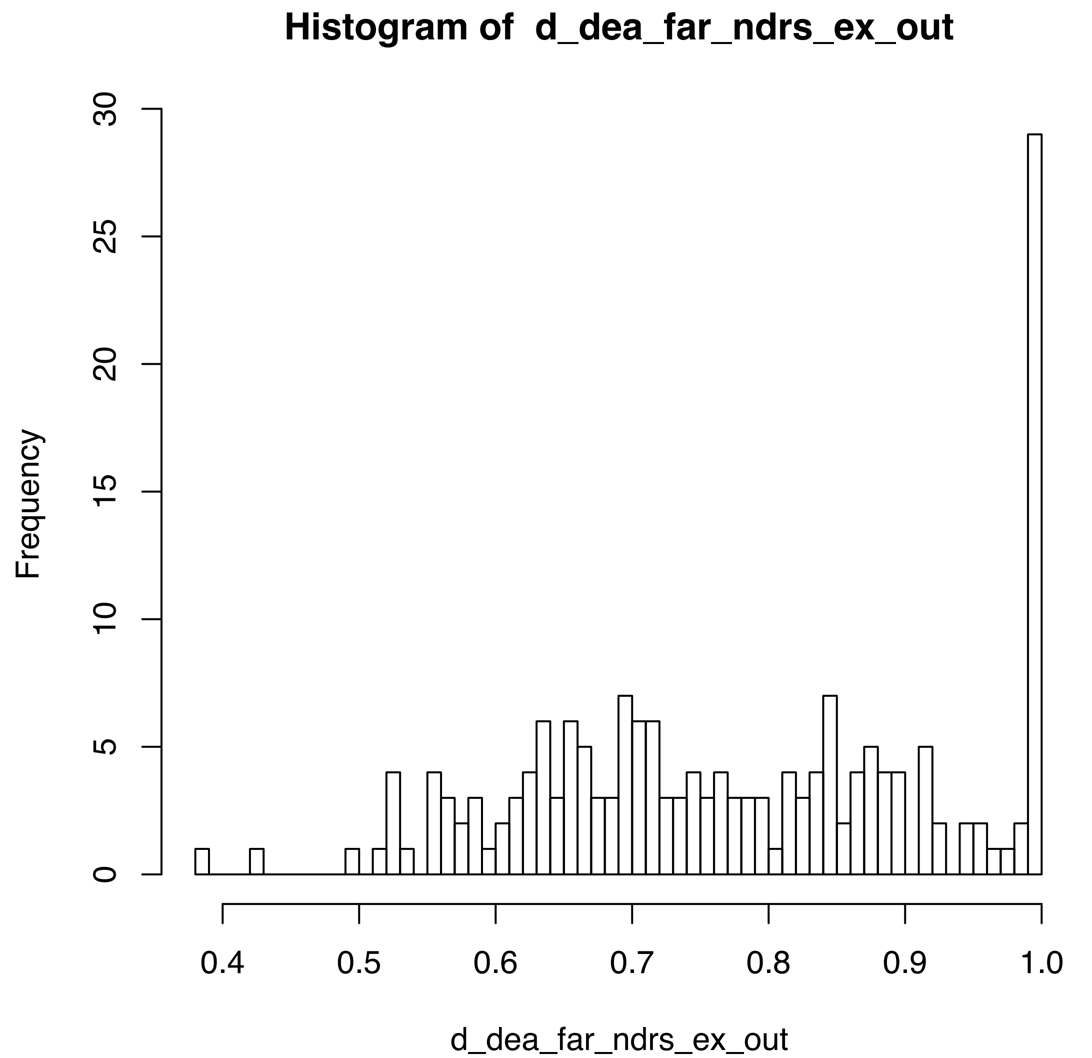
$$E(i; I \setminus i) > q(0.75) + 1.5 * (q(0.75) - q(0.25))$$

3.6% outliers

DEA outliers (7 DSOs)

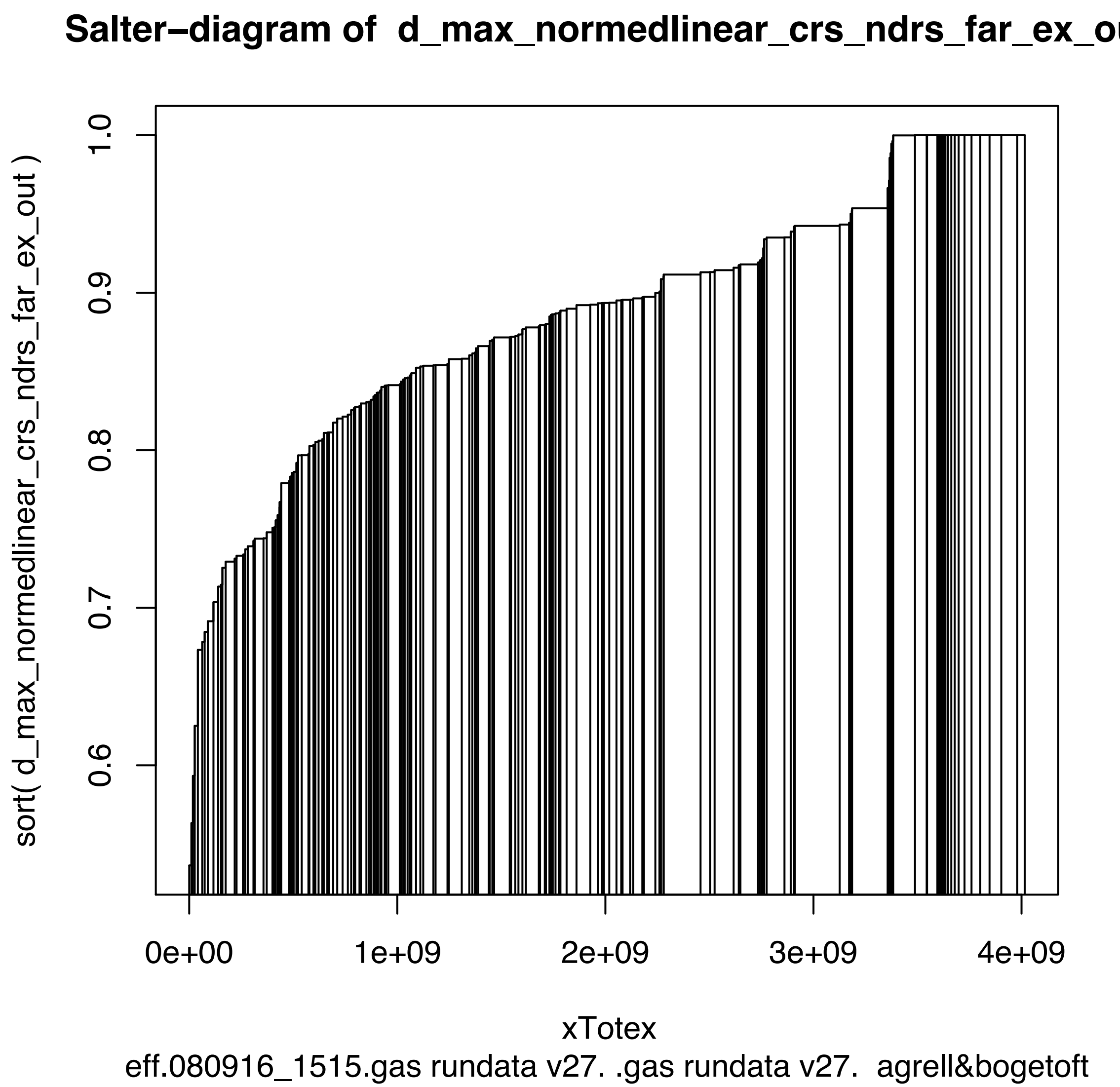


Distribution DEA NDRS ex out

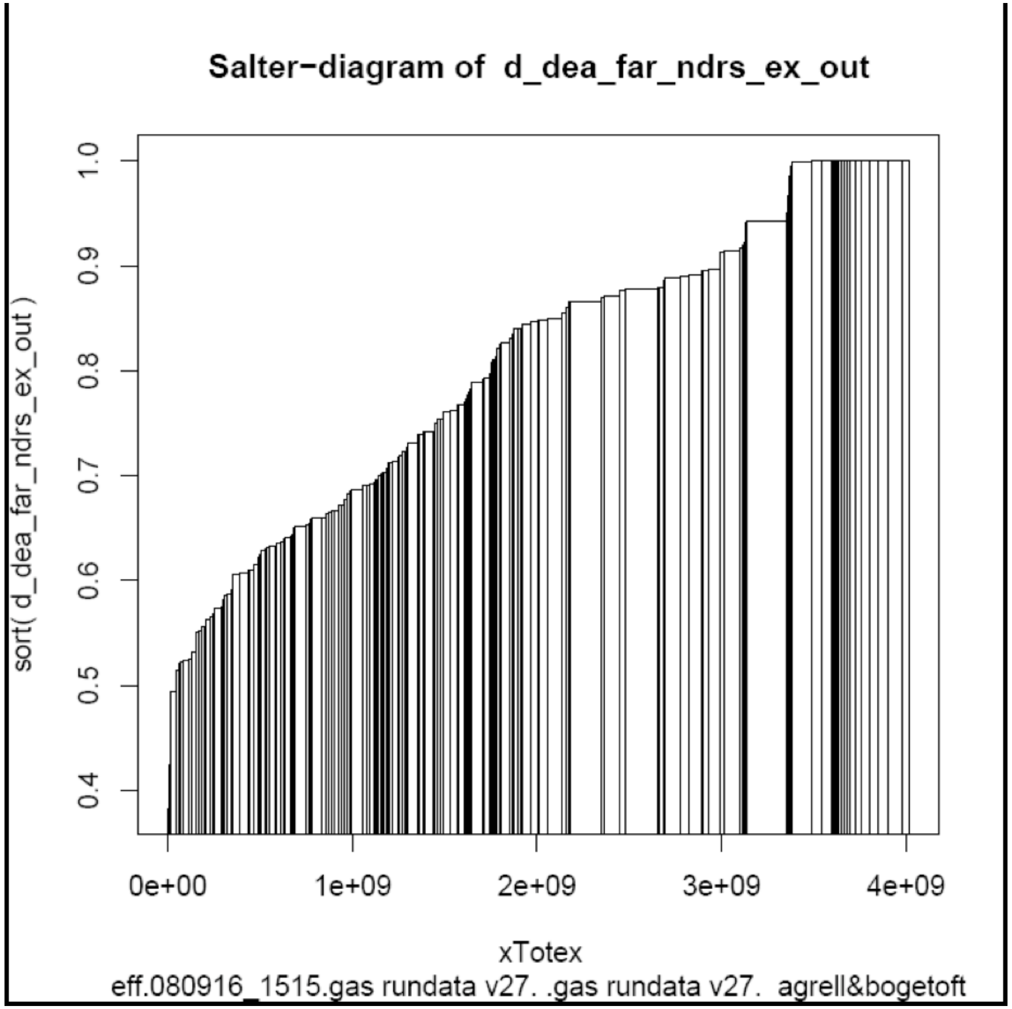


eff.080916_1515.gas rundata v27. .gas rundata v27. agrell&bogetoft

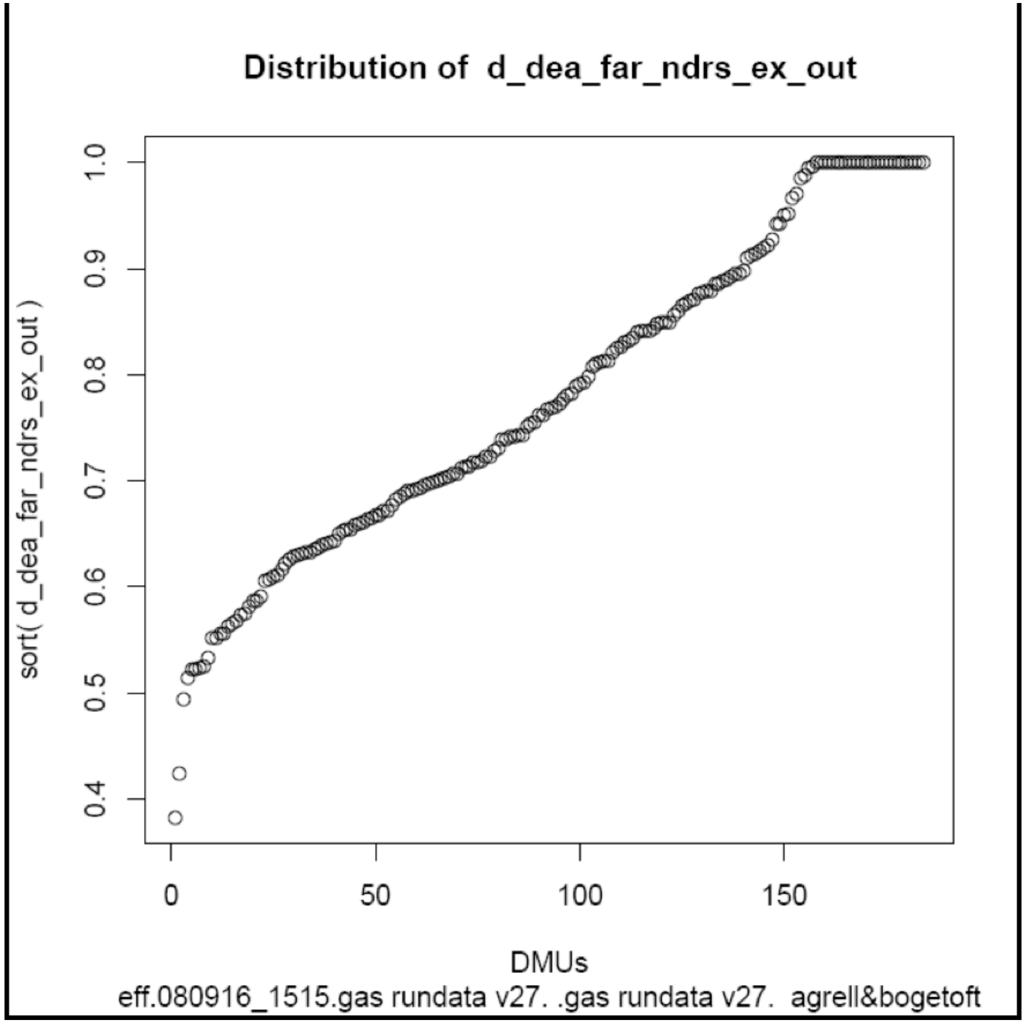
Cumulative DEA NDRS ex out



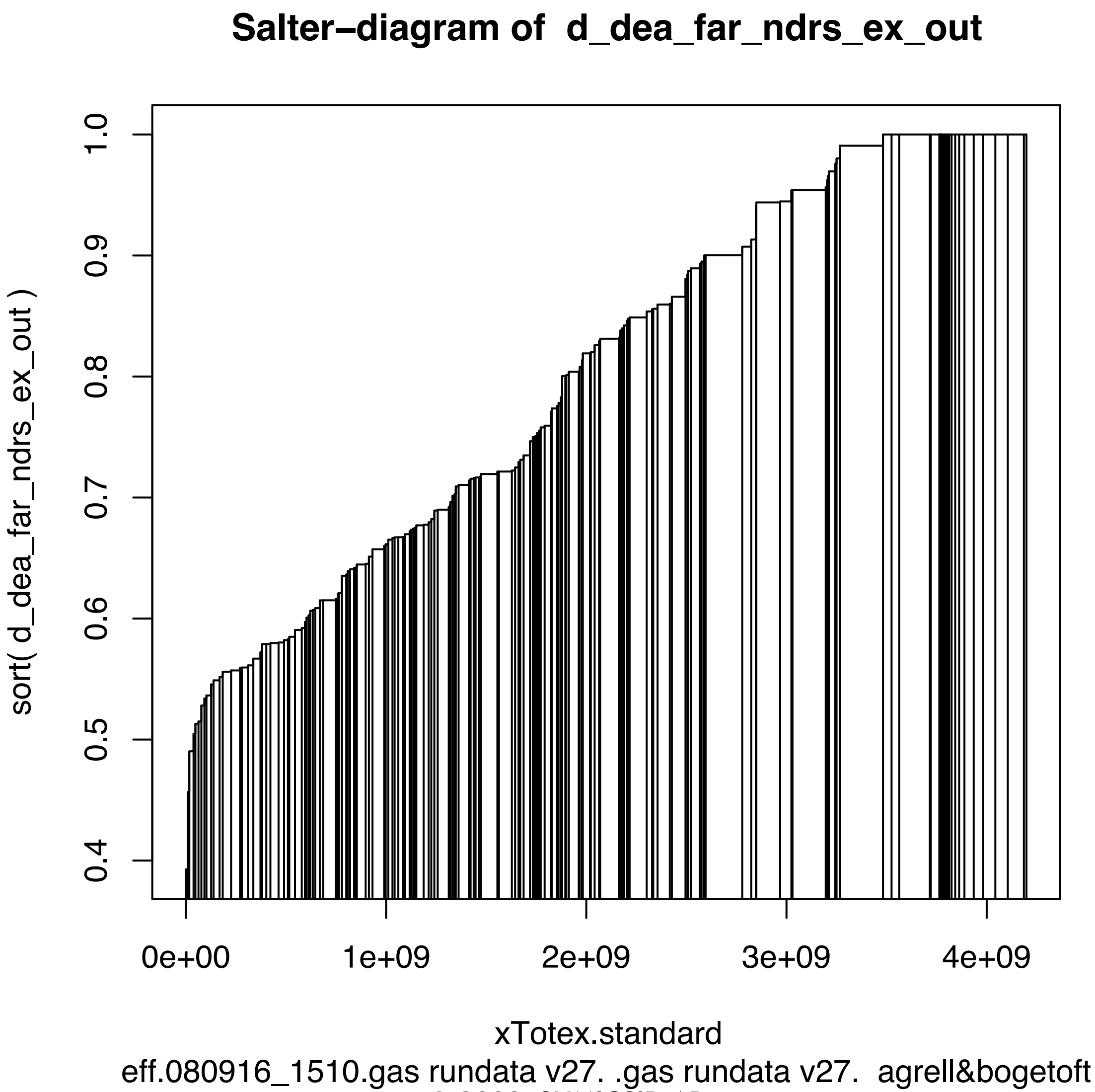
Salter DEA NDRS



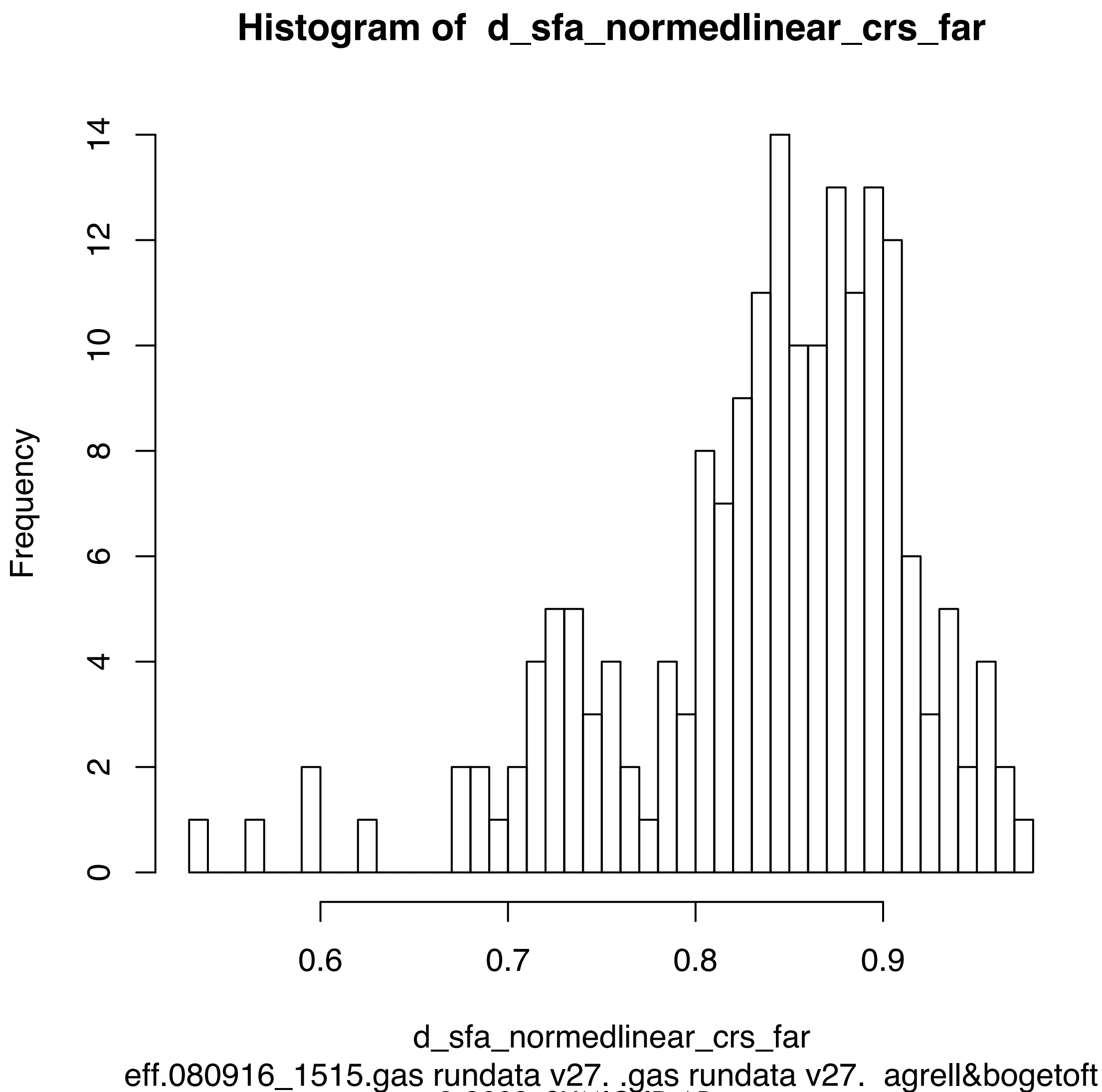
Cumulative DEA NDRS



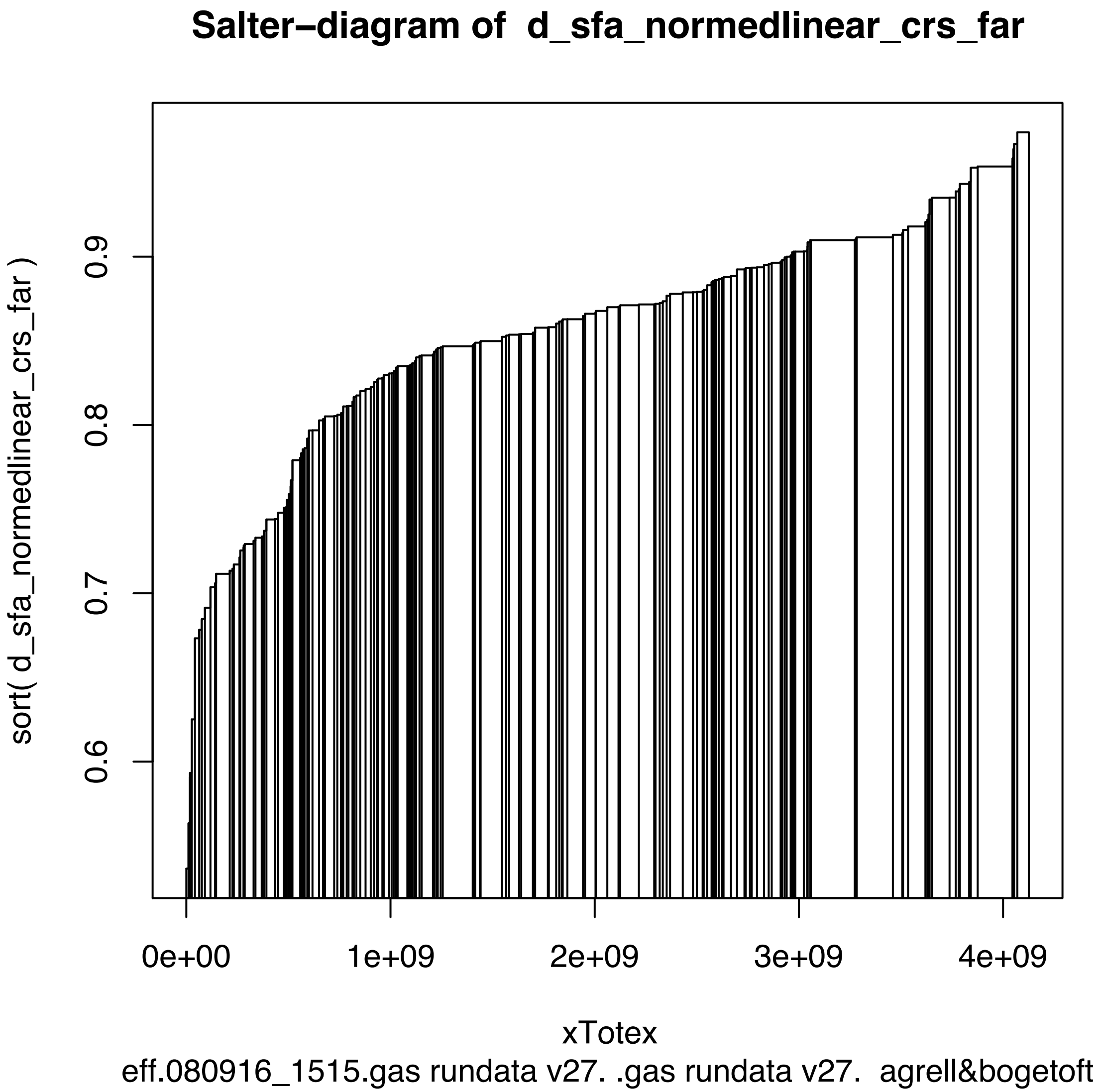
Salter DEA NDRS ex out



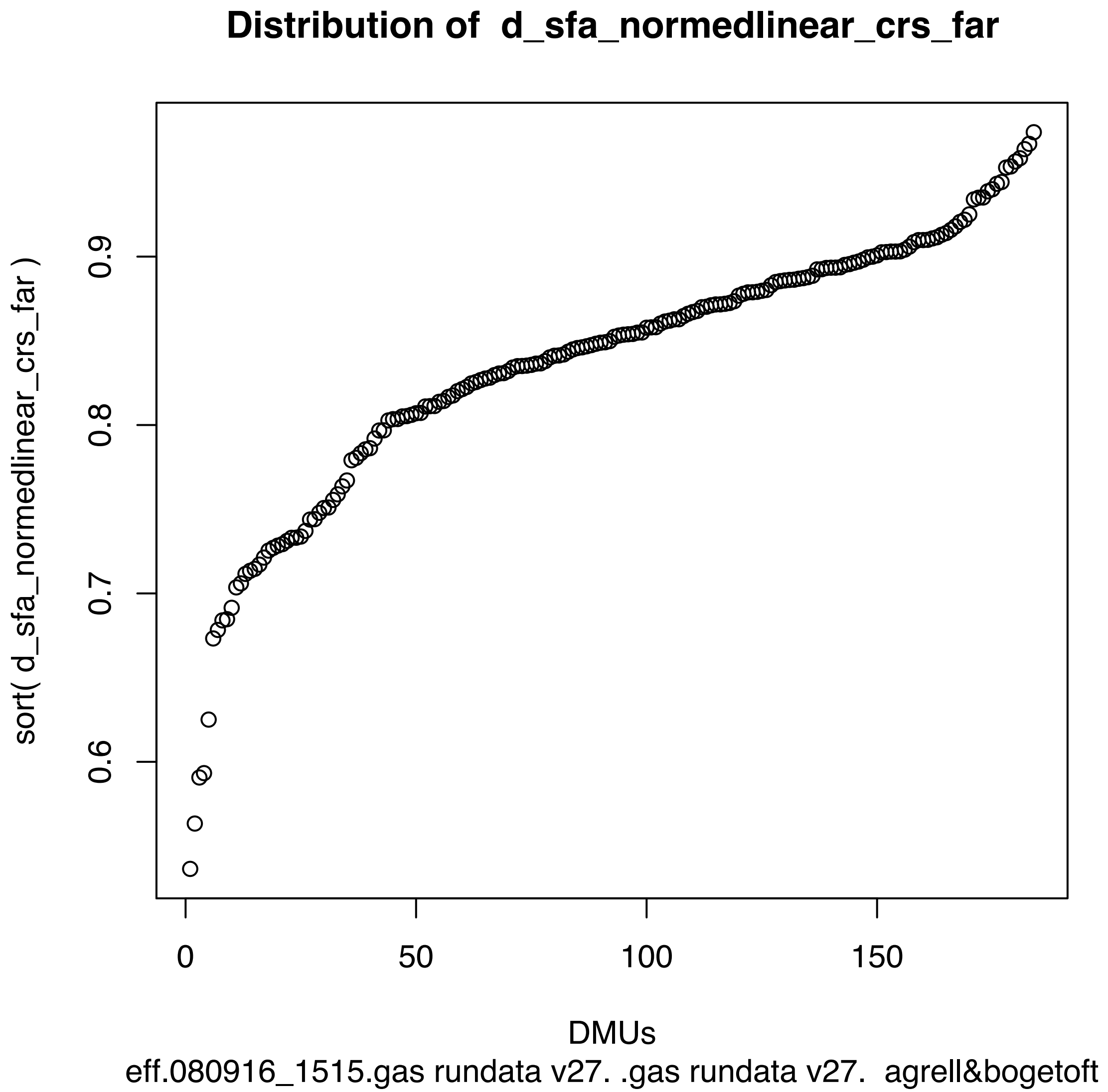
Distribution SFA normedlinear crs



Salter SFA normedlinear crs



Cumulative SFA normedlinear crs



SFA normedlinear crs Totex

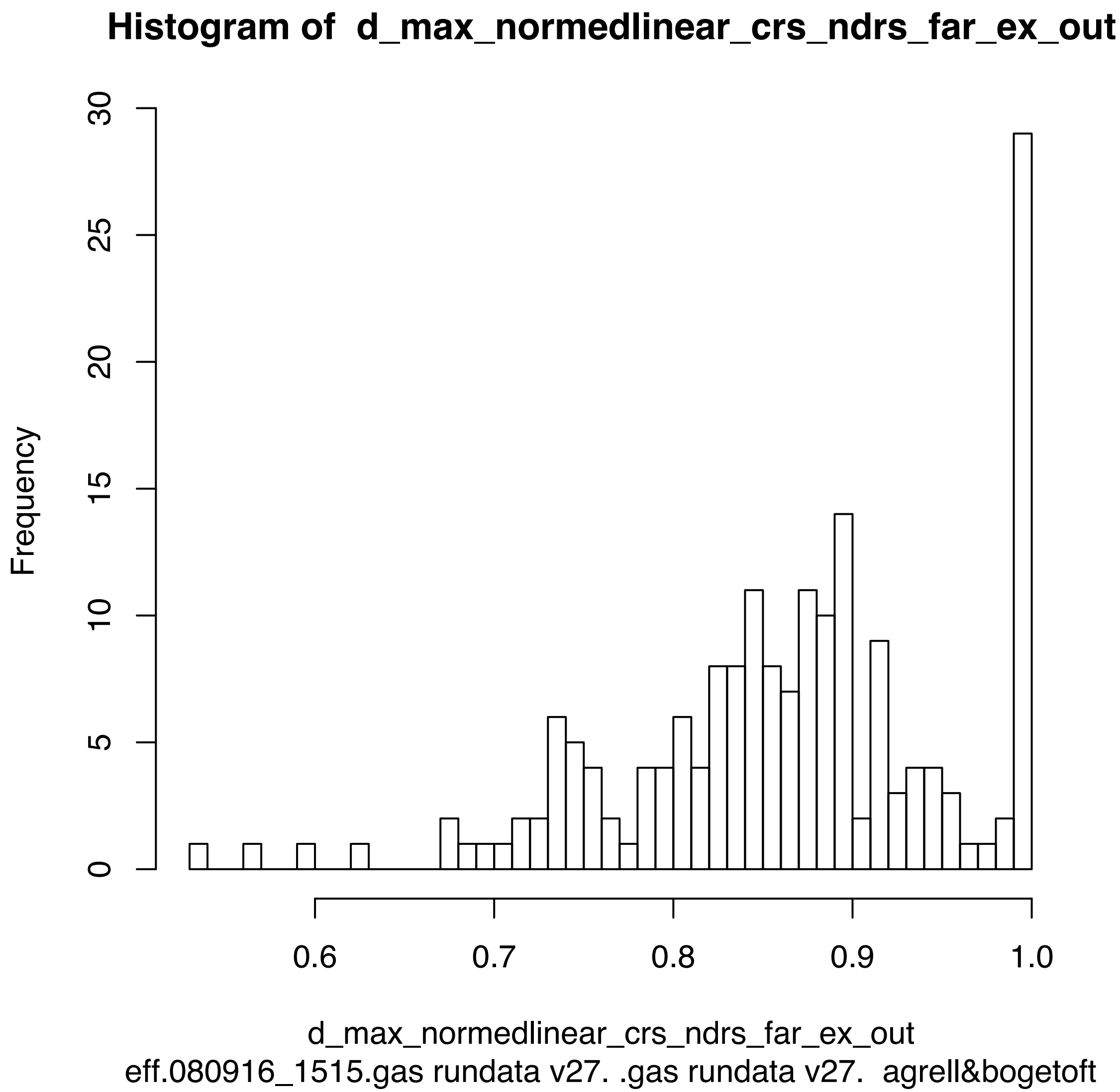
sfa.normedlinear.crs			
Aufwandsparameter als abhängige Variable	Koeffizient	Standardfehler	T-Ratio
yConnection.out.tot	-376.43	102.42	-3.68
yArea.supplied	-9386.00	72.49	-129.49
yPeakload.out	260.82	40.04	6.51
yPotential.connections	377.31	67.75	5.57
yPotential.peakload	-145.41	28.20	-5.16
yPipelength.lower5bar	-997.55	41.30	-24.15
yPipelength.abover5bar	1854.91	320.37	5.79
yVolume.tot.cus	131.73	20.67	6.37
yPop.1995	18.83	12.04	1.56
yPop.2006	-0.94	12.21	-0.08
sigma-squared	21391.75	249.46	85.75
gamma	0.78	0.06	13.55
mu	-35.16	52.80	-0.67

SFA normed linear crs Totex.stand

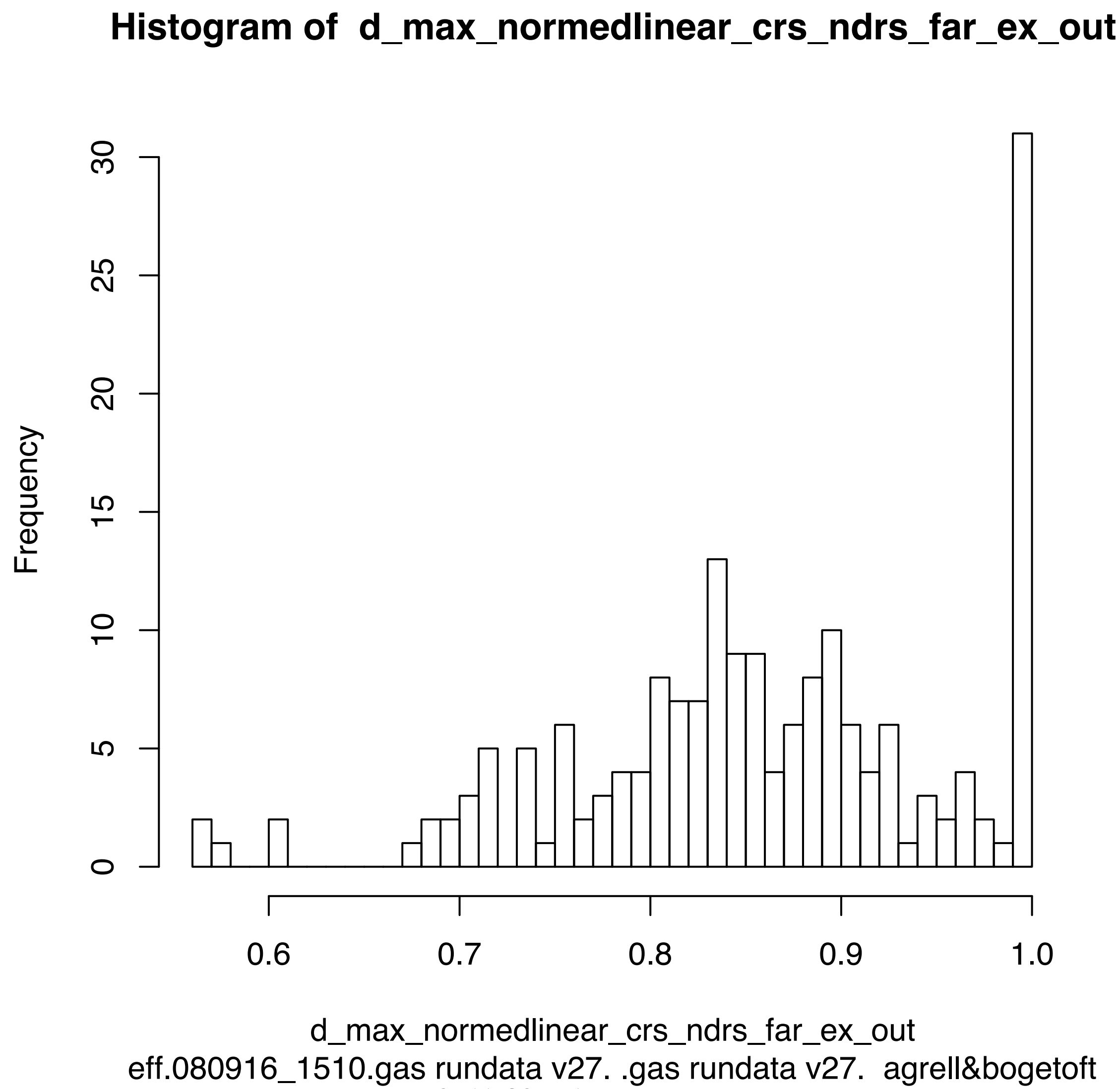
sfa.normedlinear.crs			
Aufwandsparameter als abhängige Variable	Koeffizient	Standardfehler	T-Ratio
yConnection.out.tot	-251.12	104.43	-2.40
yArea.supplied	-8929.16	4.20	-2124.85
yPeakload.out	236.44	41.94	5.64
yPotential.connections	319.27	69.27	4.61
yPotential.peakload	-122.12	29.28	-4.17
yPipelength.lower5bar	-2413.45	6.84	-352.99
yPipelength.abover5bar	1005.31	14.89	67.52
yVolume.tot.cus	154.92	20.55	7.54
yPop.1995	15.07	12.01	1.26
yPop.2006	1.15	12.20	0.09
sigma-squared	20867.61	9.80	2128.87
gamma	0.79	0.07	10.62
mu	-1.89	52.37	-0.04

Best-Of Models

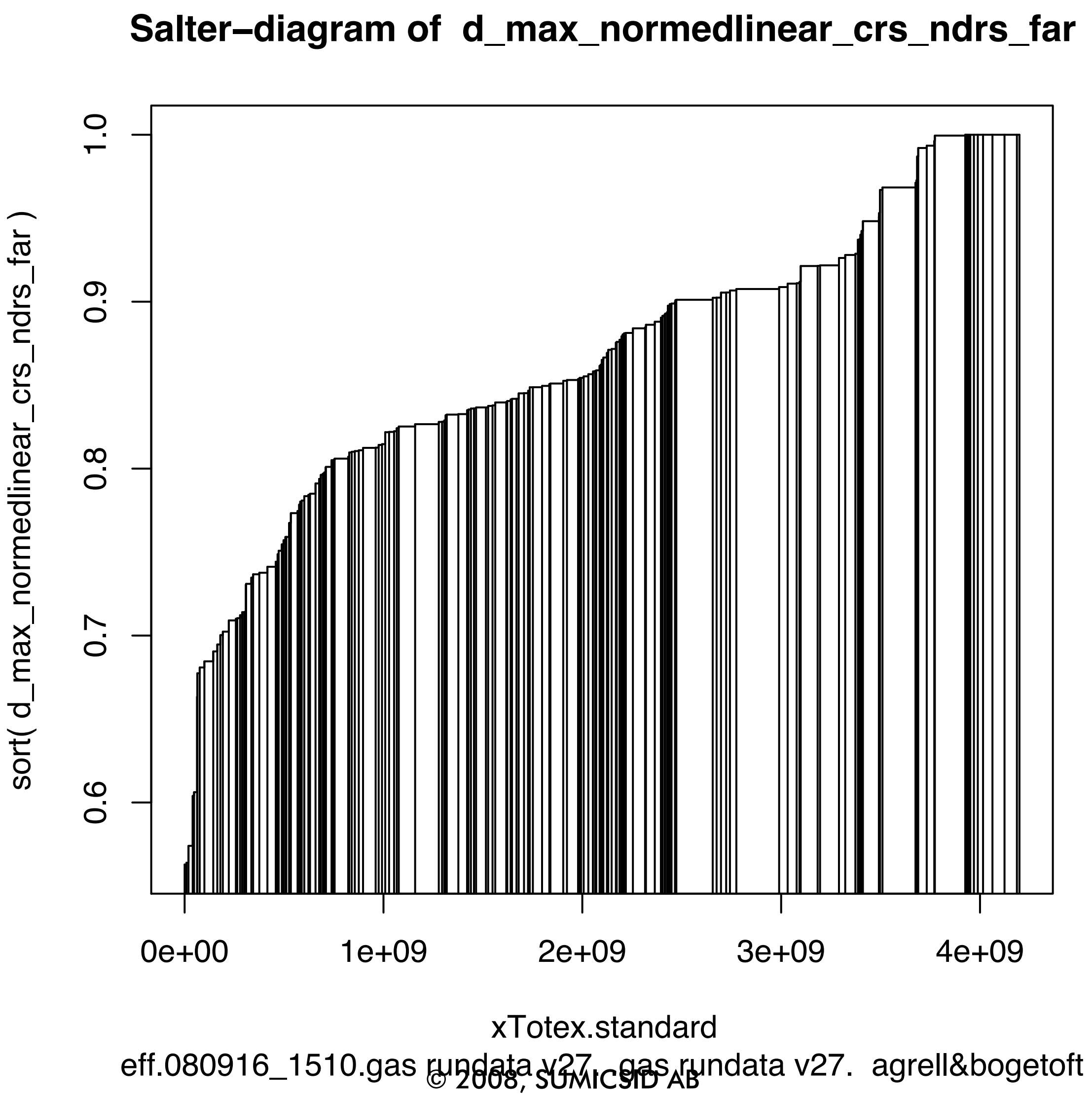
Distribution Best-Of-2 Totex



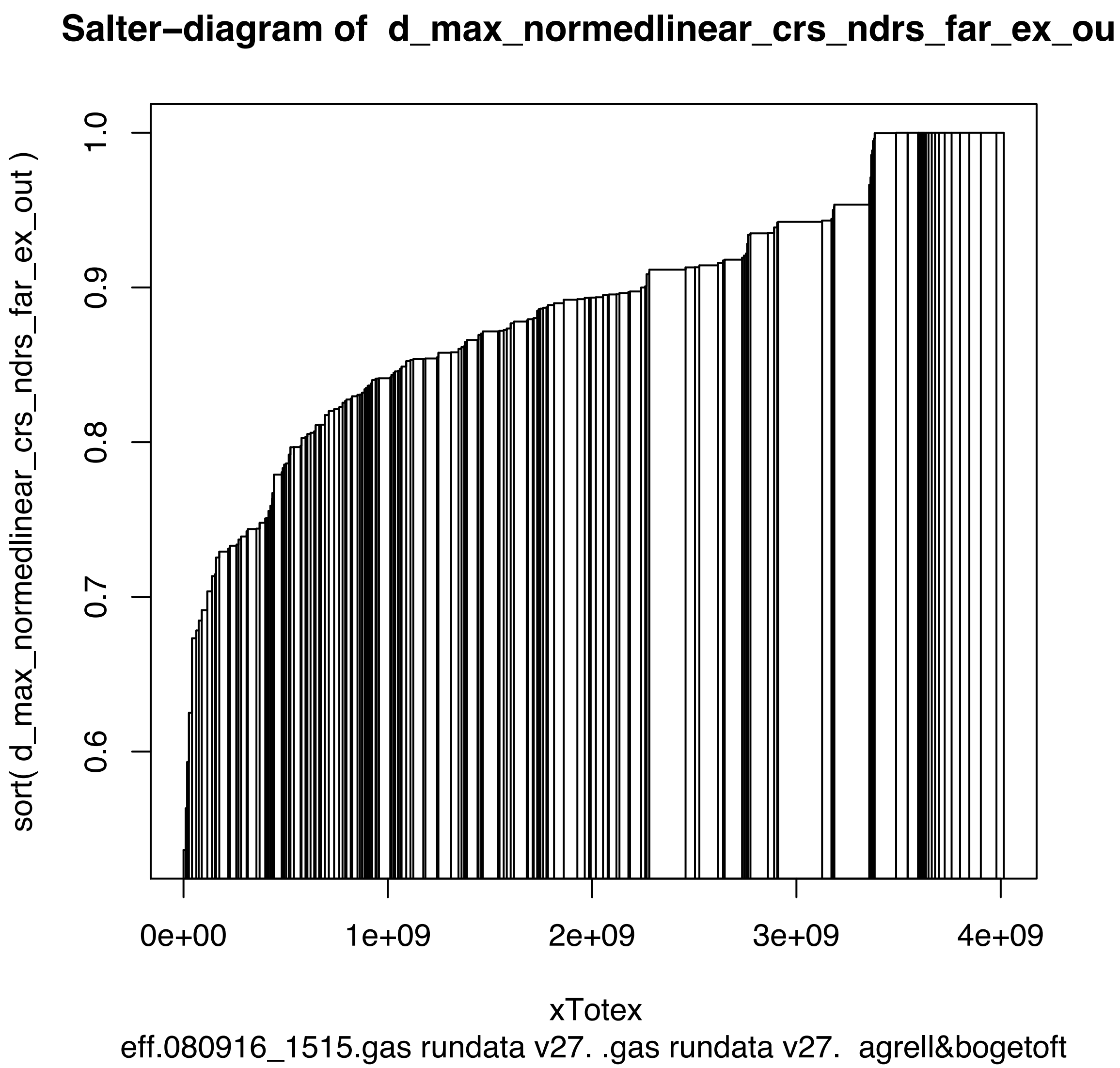
Distribution Best-Of-2 Totex.stand



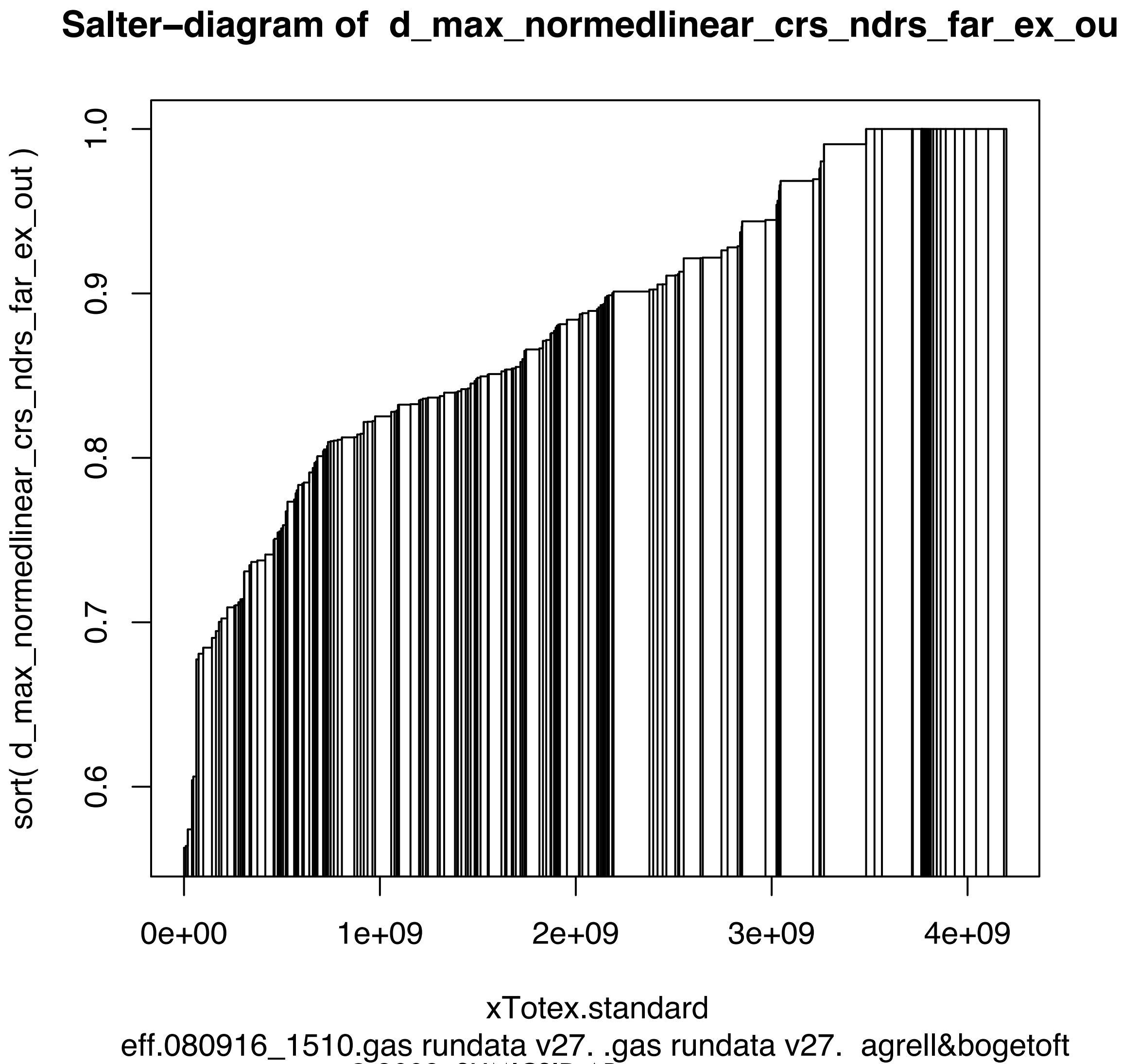
Salter Best-Of-2 (incl outliers)



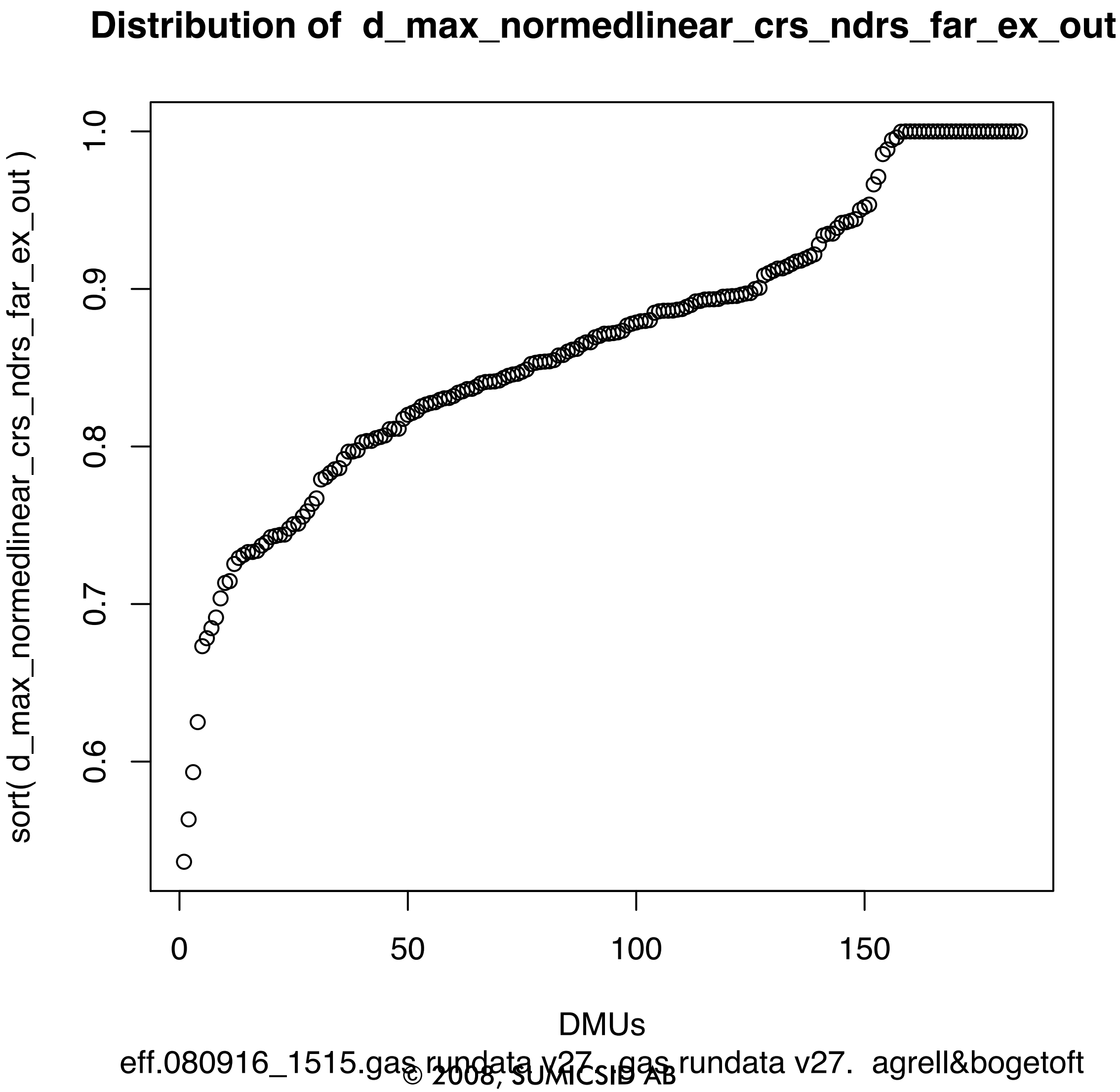
Salter Best-Of-2 Totex



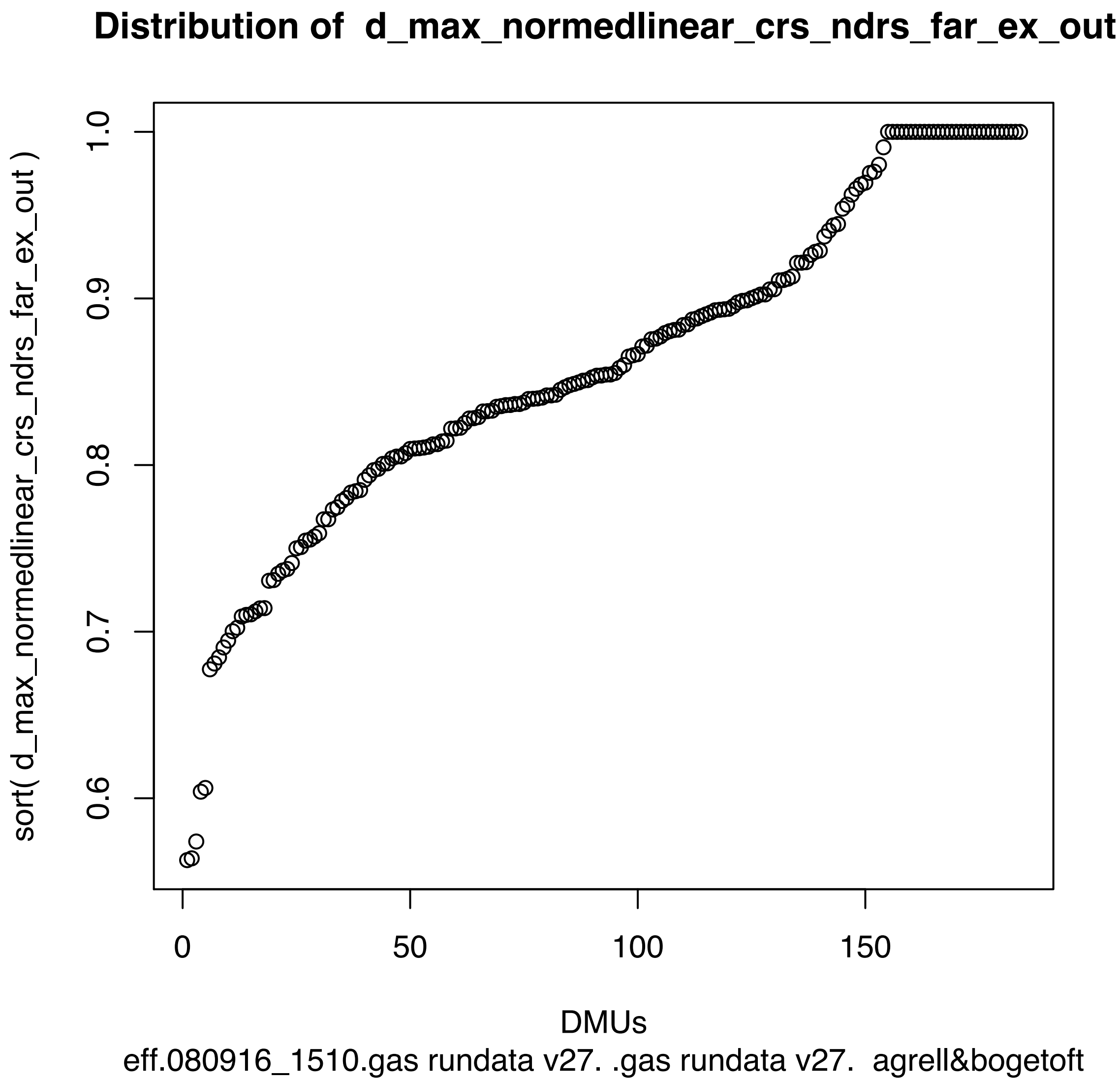
Salter Best-Of-2 Totex.stand



Cumulative Best-Of-2 Totex



Cumulative Best-Of-2 Totex.stand



Robust results

	Mittelwert	Standard-Abweichung	Minimum	Anzahl der Unternehmen, die einen Effizienzwert kleiner als 0.6 aufweisen	Anzahl der Unternehmen, die auf der Effizienzgrenze liegen (inkl. supereffiziente Unternehmen)
Maximum aus DEA Modell 3 und SFA Modell 1 (Aufwandsparameter als Input)	0.864	0.094	0.536	3	25
Maximum aus DEA Modell 3 und SFA Modell 1 (Aufwandsparameter mit standardisierten Kapitalkosten als Input)	0.859	0.100	0.563	3	28
„Best-of-Four": Maximum aus DEA Modell 3 und SFA Modell 1 (jeweils mit Aufwandsparametern und Aufwandsparametern mit standardisierten Kapitalkosten)	0.873	0.094	0.496	2	31

Efficiency of econometric outliers

Efficiency of outliers

6 DEA outlier

$$E(i;I\backslash i) > q(0.75)+1.5*(q(0.75)-q(0.25))$$

– E=1 by construction

3 econometric outliers

– Cooks distance

– E determined against fixed DEA and SFA model

Procedure

Against DEA_ndrs_ex_out

– Super-efficiency evaluation

Against SFA_normedlinear_crs

– Calculate residual, i.e. x-f(y) = v+u

– Determine efficiency as $E(u|v+u)$

– Hereby we estimate the efficiency that would have resulted if the outliers did not affect mean structure nor efficiency and noise distributions

Second stage analyses

Steps

Look for missing cost drivers etc

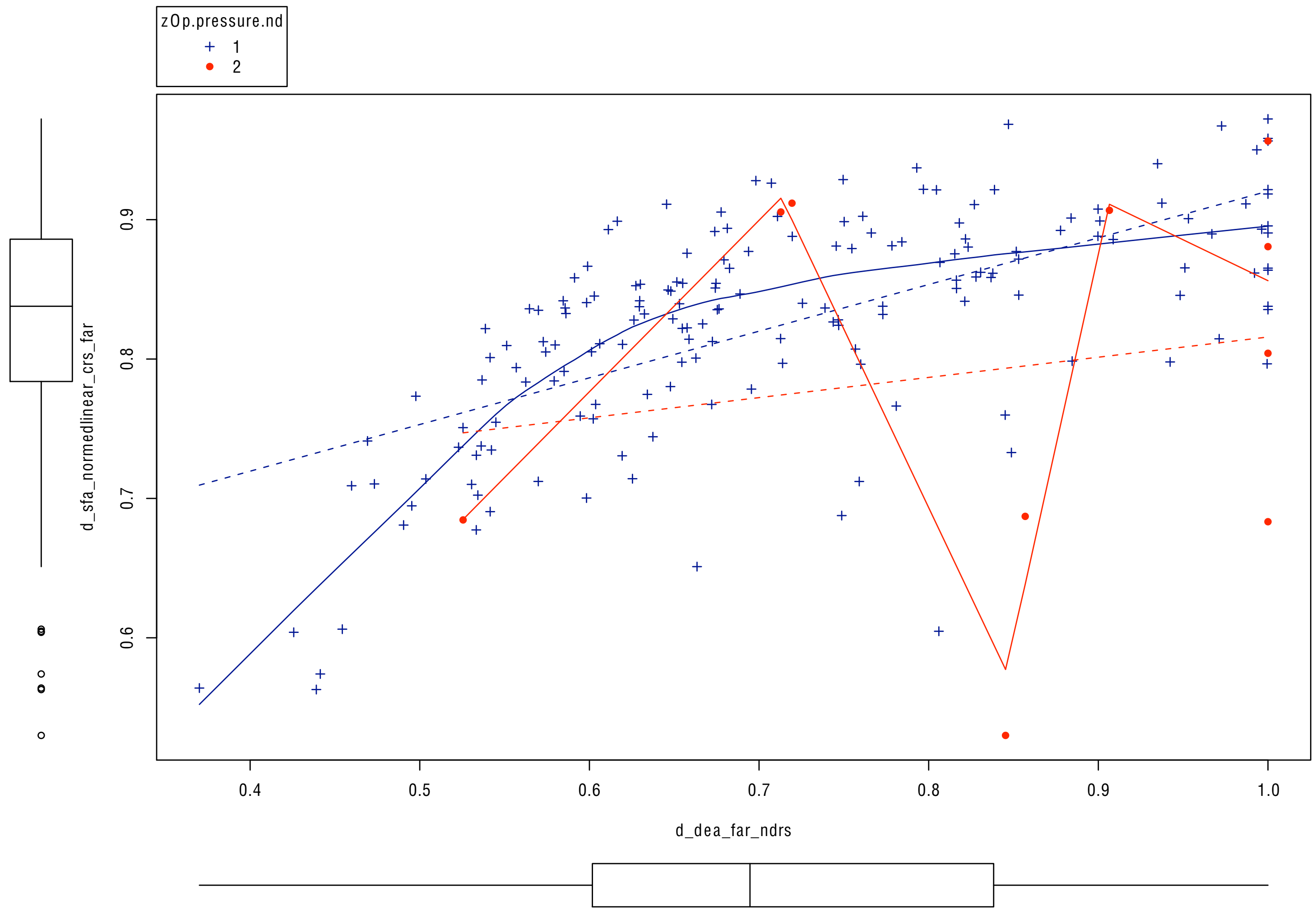
- Graphical inspection
- Kruskal Wallis test for ordinal (0-1) differences
- Tobit regressions for cardinal variables

Kruskal-Wallis

Test	Beschreibung der Variablen die anhand des Kruskal-Wallis-Tests überprüft werden
Test 1	Belegenheit der Netze (Ost vs. West, Trennkriterium: Rohrleitungen liegen ihı Länge nach zu mehr als 50% auf dem Gebiet der neuen Bundesländer.
Test 2	Betreiber von Wassernetzen (ja vs. nein)
Test 3	Betreiber von Fernwärmenetzen (ja vs. nein)
Test 4	Betreiber von Abwassernetzen (ja vs. nein)
Test 5	Betreiber von Gasspeichern (ja vs. nein)
Test 6	Betreiber von Telekommunikationsnetzen (ja vs. nein)
Test 7	Betreiber von Stromnetzen (ja vs. nein)
Test 8	Netzpuffer (vorhanden vs. nicht vorhanden; nicht unterteilt nach nutzbarem Volumen)
Test 9	Speichernutzung (vorhanden, nicht vorhanden, nicht getrennt nach geometrischem Volumen)

Best-Of-Two and Best-Of-Four models

Operation at ND/not ND



Kruskal-Wallis (App C)

<table>
<tr><td>Test 1</td><td>Lage der Netzes (Ost vs. West, Trennkriterium ist wenn die Rohrleitungen ihrer Länge nach zu mehr als 50% auf dem Gebiet der neuen Bundesländer liegen.</td></tr>
<tr><td></td><td><pre>Test: Equality of populations (Kruskal-Wallis test)

  +----------------------+
  | z12 | Obs | Rank Sum |
  |-----+-----+----------|
  | 1 | 30 | 3278.50 |
  | 2 | 153 | 13557.50 |
  +----------------------+
chi-squared =   3.820 with 1 d.f.
probability =   0.0506
chi-squared with ties =   3.854 with 1 d.f.
probability =   0.0496</pre></td></tr>
</table>

Ex: Location (East/West) is significant w r t scores.

Tobit regressions 1

Variablen Beschreibung	Bezeichnung in Tobit-Schätzungen
zOp.pressure.nd	z3
zOp.pressure.md	z4
zOp.pressure.1.hd	z5
zOp.pressure.5.hd	z6
zOp.pressure.16.hd	z7
zOp.pressure.70.hd	z8
zCoverage.conn.fraction	z9
zCoverage.actual	z10
zCoverage.potential	z11
zShare.gas.H	z14
zShare.gas.L	z15
zShare.gas.other	z16
yCalorificvalue.grade.H	y3
yCalorificvalue.grade.L	y4
yCalorificvalue.grade.other	y5
yCapacity.out	y6
yPeakload.in	y7
yPeakload.out	y8
yEnergy.in	y9
yEnergy.out	y10
yEnergy.out.2005	y11
yEnergy.out.2004	y12
yEnergy.out.2003	y13
yEnergy.out.2002	y14
yEnergy.out.2001	y15
yConnection.out.nd	y16
yConnections.other.nd	y17
yConnections.own.nd	y18
yMeters.nd	y19
yLength.diameter.class.A.nd	y20
yLength.diameter.class.B.nd	y21
yLength.diameter.class.C.nd	y22
yLength.diameter.class.D.nd	y23
yLength.diameter.class.E.nd	y24
yLength.diameter.class.F.nd	y25
yLength.diameter.class.G.nd	y26
yLength.diameter.nd	y27
yLength.cus.nd	y28
yShare.value.steered.nd	y29
yShare.parallel.laying.nd	y30
yNet.structure.ring.nd	y31
yNet.structure.mashed.nd	y32
yNet.structure.stream.nd	y33
yConnection.out.md	y34
yConnections.other.md	y35
yConnections.own.md	y36
yMeters.md	y37
yLength.diameter.class.A.md	y38
yLength.diameter.class.B.md	y39
yLength.diameter.class.C.md	y40
yLength.diameter.class.D.md	y41
yLength.diameter.class.E.md	y42
yLength.diameter.class.F.md	y43
yLength.diameter.class.G.md	y44
yLength.diameter.md	y45
yLength.cus.md	y46
yShare.value.steered.md	y47
yShare.parallel.laying.md	y48
yNet.structure.ring.md	y49
yNet.structure.mashed.md	y50
yNet.structure.stream.md	y51

Tobit regressions 2

yConnection.out.1.hd	y52
yConnections.other.1.hd	y53
yConnections.own.1.hd	y54
yMeters.1.hd	y55
yLength.diameter.class.A.1.hd	y56
yLength.diameter.class.B.1.hd	y57
yLength.diameter.class.C.1.hd	y58
yLength.diameter.class.D.1.hd	y59
yLength.diameter.class.E.1.hd	y60
yLength.diameter.class.F.1.hd	y61
yLength.diameter.class.G.1.hd	y62
yLength.diameter.1.hd	y63
yLength.cus.1.hd	y64
yShare.value.steered.1.hd	y65
yConnection.out.5.hd	y66
yConnections.other.5.hd	y67
yConnections.own.5.hd	y68
yMeters.5.hd	y69
yLength.diameter.class.A.5.hd	y70
yLength.diameter.class.B.5.hd	y71
yLength.diameter.class.C.5.hd	y72
yLength.diameter.class.D.5.hd	y73
yLength.diameter.class.E.5.hd	y74
yLength.diameter.class.F.5.hd	y75
yLength.diameter.class.G.5.hd	y76
yLength.diameter.5.hd	y77
yLength.cus.5.hd	y78
yShare.value.steered.5.hd	y79

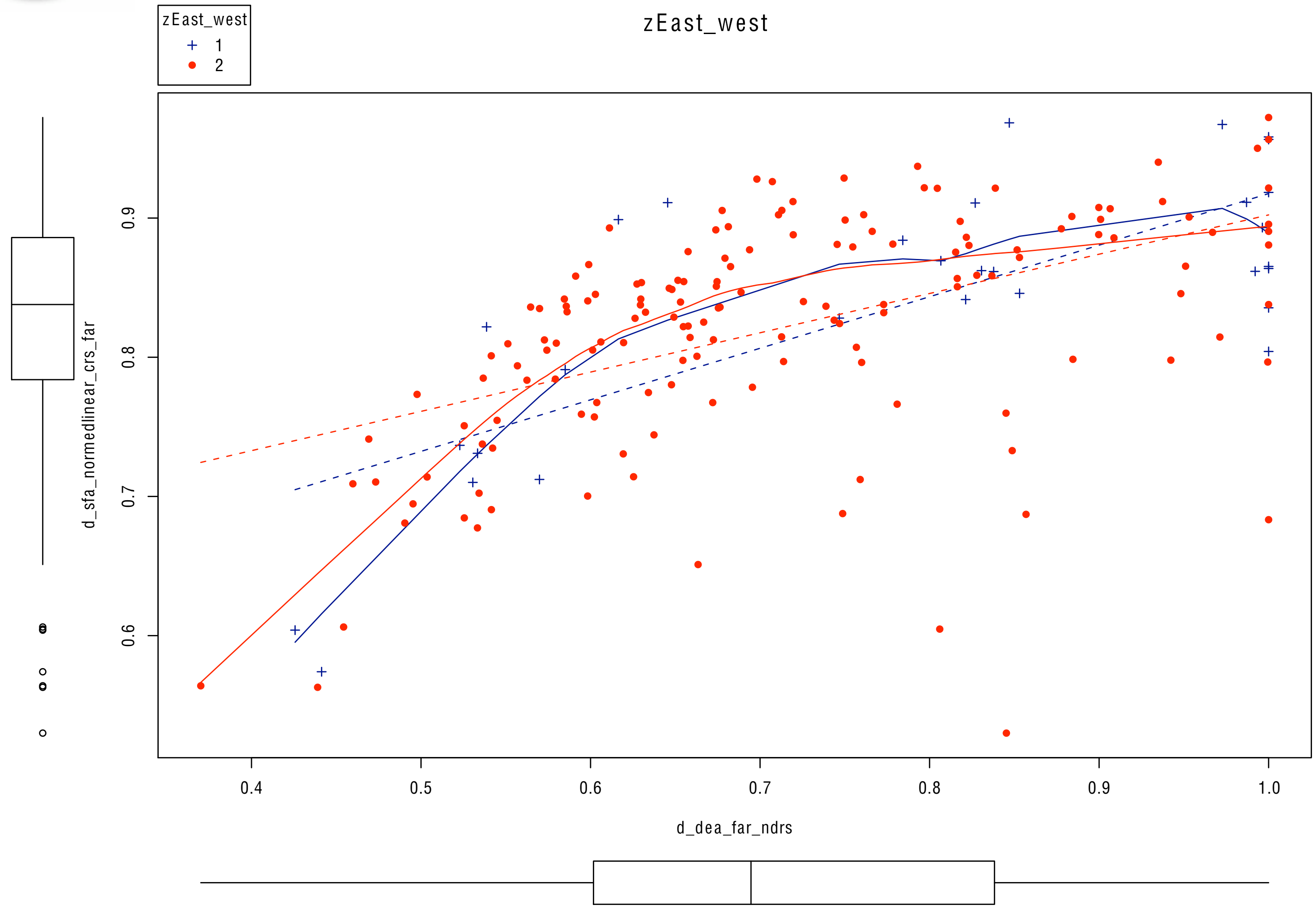
yConnection.out.16.hd	y80
yConnections.other.16.hd	y81
yConnections.own.16.hd	y82
yMeters.16.hd	y83
yLength.diameter.class.A.16.hd	y84
wx.share.ownership.A.16.hd	y85
yLength.diameter.class.B.16.hd	y86
wx.share.ownership.B.16.hd	y87
yLength.diameter.class.C.16.hd	y88
wx.share.ownership.C.16.hd	y89
yLength.diameter.class.D.16.hd	y90
wx.share.ownership.D.16.hd	y91
yLength.diameter.class.E.16.hd	y92
wx.share.ownership.E.16.hd	y93
yLength.diameter.class.F.16.hd	y94
wx.share.ownership.F.16.hd	y95
yLength.diameter.class.G.16.hd	y96
wx.share.ownership.G.16.hd	y97
yLength.class.16.hd	y98
wx.share.ownership.16.hd	y99
yLength.cus.16.hd	y100
yShare.value.steered.16.hd	y101
yConnection.out.70.hd	y102
yConnections.other.70.hd	y103
yConnections.own.70.hd	y104
yMeters.70.hd	y105
yLength.diameter.class.A.70.hd	y106
wx.share.ownership.A.70.hd	y107
yLength.diameter.class.B.70.hd	y108

Tobit regressions 3

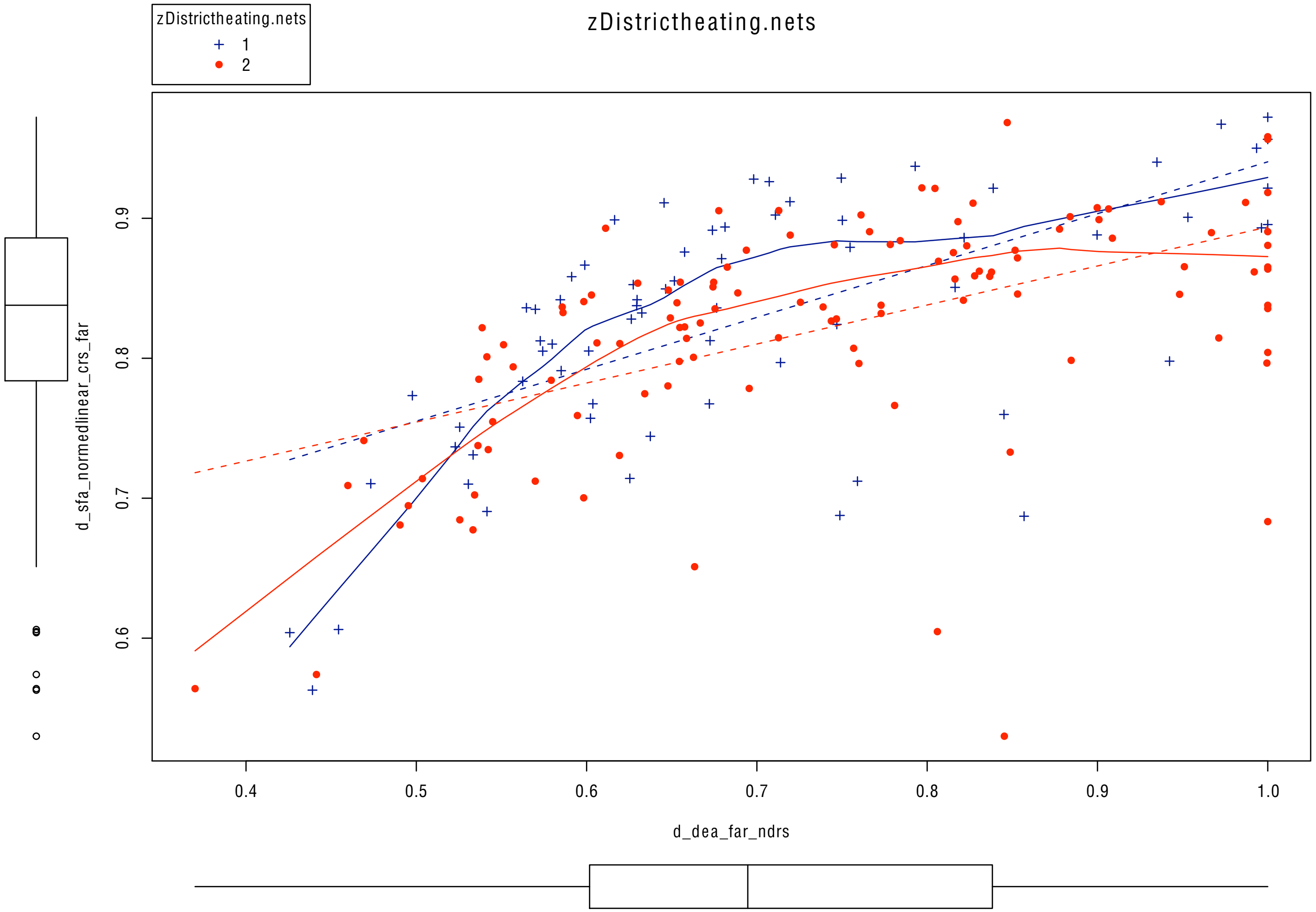
wx.share.ownership.B.70.hd	y109
yLength.diameter.class.C.70.hd	y110
wx.share.ownership.C.70.hd	y111
yLength.diameter.class.D.70.hd	y112
wx.share.ownership.D.70.hd	y113
yLength.diameter.class.E.70.hd	y114
wx.share.ownership.E.70.hd	y115
yLength.diameter.class.F.70.hd	y116
wx.share.ownership.F.70.hd	y117
yLength.diameter.class.G.70.hd	y118
wx.share.ownership.G.70.hd	y119
yLength.diameter.70.hd	y120
wx.share.ownership.70.hd	y121
yLength.cus.70.hd	y122
yShare.value.steered.70.hd	y123
yShare.parallel.laying.hd	y124
yNet.structure.ring.hd	y125
yNet.structure.mashed.hd	y126
yNet.structure.stream.hd	y127
yCompressor.hd	y128
yCompressor.SQmax.R.hd	y129
yCompressor.SAmax.R.hd	y130
yCompressor.in.hd	y131
yCompressor.in.SQmax.R.hd	y132
yCompressor.in.SQmax.A.hd	y133
yCompressor.in.SQact.A.hd	y134
yCompressor.out.hd	y135
yCompressor.out.SQmax.R.hd	y136
yCompressor.out.SQmax.A.hd	y137
yCompressor.out.SQact.A.hd	y138

yAssets.hd	y139
yAssets.SQmaxR.hd	y140
yAssets.SQmax.A.hd	y141
yAssets.SQact.A.hd	y142
wxAssets.share.hd	y143
yCompressor.transport.hd	y144
yCompressor.transport.SQmax.TV.hd	y145
yCompressor.transport.SQact.TV.hd	y146
yCompressor.net.hd	y147
yCompressor.net.SQmax.NSV.hd	y148
yCompressor.net.SQact.NSV.hd	y149
wxCompressor.share.hd	y150
yConnection.out.hd	y171
yConnection.points.hd	y172
yConnection.out.tot	y173
yConnection.points.tot	y174
yPotential.connections	y175
yPotential.area	y176
yPotential.peakload	y177
yMeters.hd	y178
yMeters.tot	y179
yVolume.nd	y180
yVolume.md	y181
yVolume.hd	y182
yVolume.cus	y183
yVolume.tot	y184
yVolume.tot.cus	y185
yPop.1995	y200
yPop.2006	y201

East West location



District Heating operation



Tobit Density DEA/NDRS

```
d_dea_far_ndrs
Call:
tobit(formula = S[, s] ~ dens, left = 0, right = 1)

Observations:
         Total  Left-censored     Uncensored Right-censored
           184              0            169             15

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.87705    0.04194  20.911  < 2e-16 ***
dens        -0.09744    0.02654  -3.672 0.000241 ***
Log(scale)  -1.80930    0.05573 -32.465  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Scale: 0.1638

Gaussian distribution
Number of Newton-Raphson Iterations: 3
Log-likelihood: 46.73 on 3 Df
Wald-statistic: 13.48 on 1 Df, p-value: 0.00024079
```

Tobit
Density DEA/NDRS/ex.out

```
d_dea_far_ndrs_ex_out
Call:
tobit(formula = S[, s] ~ dens, left = 0, right = 1)

Observations:
         Total  Left-censored     Uncensored Right-censored
           184              0            156             28

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.94748    0.04487  21.115  < 2e-16 ***
dens        -0.10816    0.02830  -3.822 0.000132 ***
Log(scale)  -1.75535    0.05904 -29.731  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Scale: 0.1728

Gaussian distribution
Number of Newton-Raphson Iterations: 3
Log-likelihood: 20.69 on 3 Df
Wald-statistic: 14.61 on 1 Df, p-value: 0.00013215
```

Tobit Density SFA

```
d_sfa_normedlinear_crs_far
Call:
tobit(formula = S[, s] ~ dens, left = 0, right = 1)

Observations:
         Total  Left-censored     Uncensored Right-censored
           184              0            184              0

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.80582    0.02152  37.443   <2e-16 ***
dens         0.01194    0.01363   0.876    0.381
Log(scale)  -2.46958    0.05213 -47.375   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Scale: 0.08462

Gaussian distribution
Number of Newton-Raphson Iterations: 2
Log-likelihood: 193.3 on 3 Df
Wald-statistic: 0.7676 on 1 Df, p-value: 0.38095
```

Not significant

Tobit Density /Best-of-Two

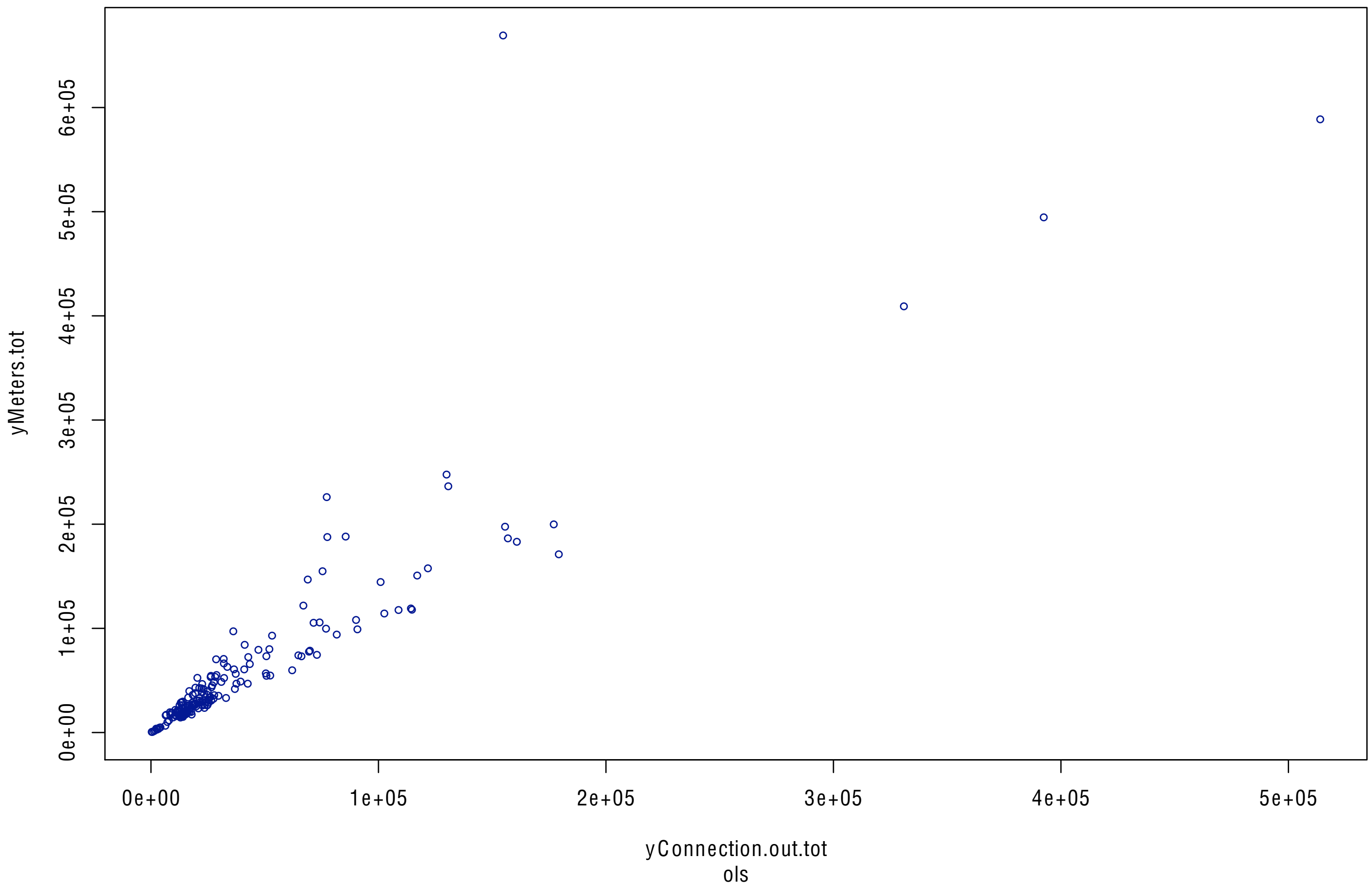
```
d_max_normedlinear_crs_ndrs_far
Call:
tobit(formula = S[, s] ~ dens, left = 0, right = 1)

Observations:
         Total  Left-censored     Uncensored Right-censored
           184              0            169             15

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.87438    0.02562  34.131   <2e-16 ***
dens        -0.01625    0.01621  -1.002    0.316
Log(scale)  -2.30272    0.05568 -41.360   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Scale: 0.09999

Gaussian distribution
Number of Newton-Raphson Iterations: 3
Log-likelihood: 130.6 on 3 Df
Wald-statistic: 1.004 on 1 Df, p-value: 0.31631
```

Not significant

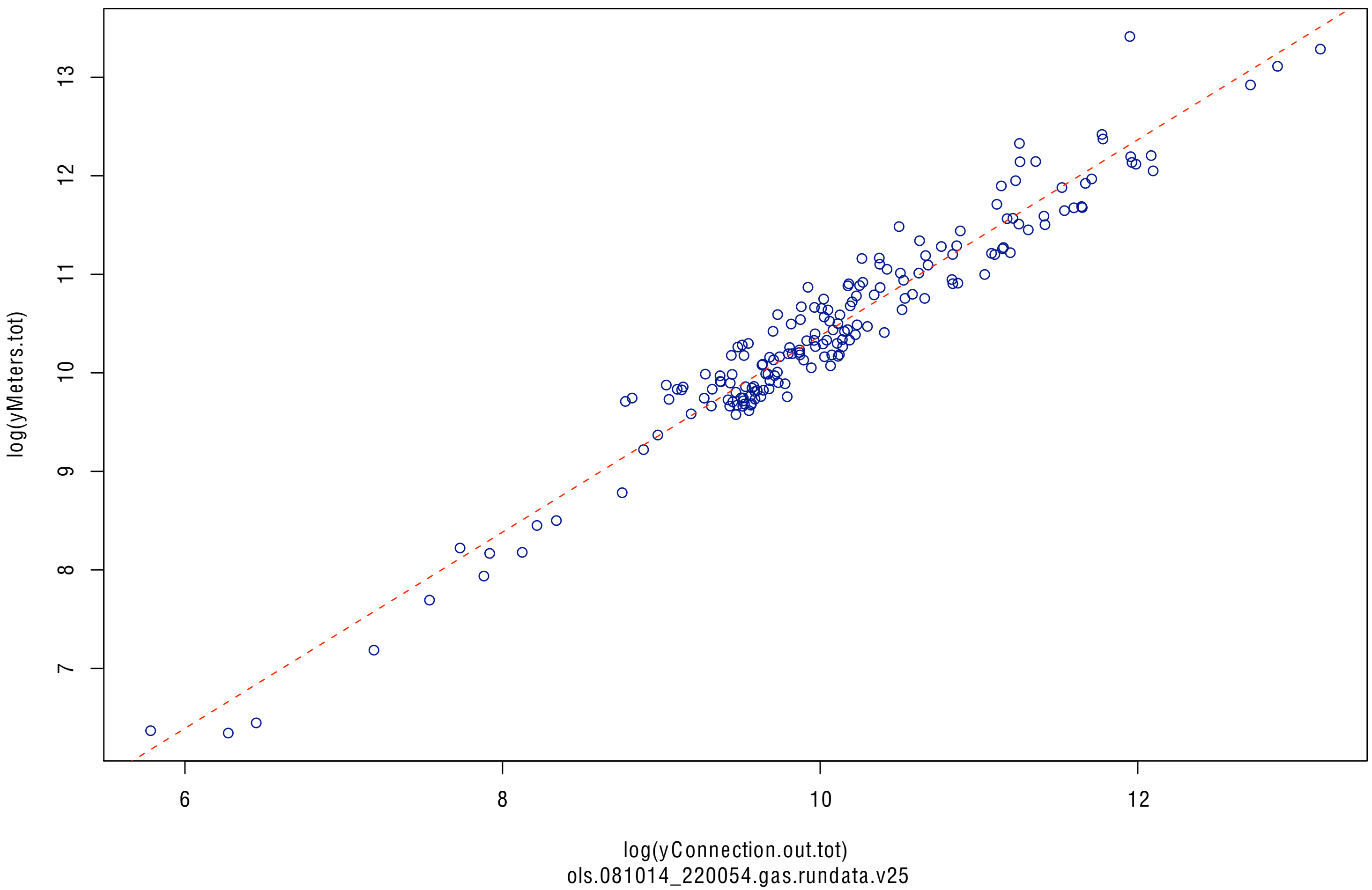
Tobit Density

```
d_max_normedlinear_crs_ndrs_far
Call:
tobit(formula = S[, s] ~ dens, left = 0, right = 1)

Observations:
         Total  Left-censored     Uncensored Right-censored 
           184              0            169             15 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  0.87438    0.02562  34.131   <2e-16 ***
dens        -0.01625    0.01621  -1.002    0.316    
Log(scale)  -2.30272    0.05568 -41.360   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Scale: 0.09999 

Gaussian distribution
Number of Newton-Raphson Iterations: 3 
Log-likelihood: 130.6 on 3 Df
Wald-statistic: 1.004 on 1 Df, p-value: 0.31631
```

Meters and Connections 1



Meters and Connections 2



Tobit testing of final model

Exogene Regressoren in der Tobit-Schätzung
yConnection.out.tot
yArea.supplied
yPeakload.out
yPotential.connections
yPotential.peakload
yPipelength.lower5bar
yPipelength.above5bar
yVolume.tot.cus
yPop.1995
yPop.2006
zOp.pressure.nd
zOp.pressure.70.hd
zEastWest
zDistrict.heating.nets

„Best-of-Four“-Modell

```
Tobit regression                                  Number of obs   =        183
                                                  LR chi2(14)     =      41.43
                                                  Prob > chi2     =     0.0002
Log likelihood =   60.69451                       Pseudo R2       =    -0.5182

------------------------------------------------------------------------------
          r3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        y173 |   3.62e-06   3.83e-06     0.94   0.347    -3.95e-06    .0000112
          y2 |   .0002039     .00007     2.91   0.004     .0000656    .0003421
          y8 |  -1.72e-06   1.24e-06    -1.39   0.167    -4.16e-06    7.26e-07
        y175 |  -3.14e-06   2.29e-06    -1.37   0.172    -7.67e-06    1.38e-06
        y177 |   1.42e-06   8.78e-07     1.62   0.108    -3.15e-07    3.15e-06
        y190 |   4.60e-06   .0000253     0.18   0.856    -.0000453    .0000545
        y191 |   -.000027   .0000584    -0.46   0.645    -.0001423    .0000884
        y185 |   1.75e-08   5.68e-07     0.03   0.975    -1.10e-06    1.14e-06
        y200 |  -3.90e-07   5.47e-07    -0.71   0.477    -1.47e-06    6.89e-07
        y201 |   3.03e-07   5.76e-07     0.53   0.599    -8.33e-07    1.44e-06
      dummy1 |   .1360636    .050456     2.70   0.008     .0364584    .2356688
      dummy2 |   .0181056   .0294912     0.61   0.540    -.0401131    .0763243
      dummy3 |  -.0475361    .030699    -1.55   0.123    -.1081391    .0130668
      dummy4 |   .0275427   .0207005     1.33   0.185    -.0133222    .0684075
       _cons |   .8720524   .0353197    24.69   0.000     .8023279     .941777
-------------+----------------------------------------------------------------
      /sigma |   .1247833   .0077173                      .1095486     .140018
------------------------------------------------------------------------------
  Obs. summary:          1  left-censored observation  at r3<=.49592751
                       144     uncensored observations
                        38 right-censored observations at r3>=1

------------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+----------------------------------------------------------------
           . |    183    39.97721    60.69451     16    -89.38902   -38.03724
------------------------------------------------------------------------------
```

Final second stage analysis

Potentially missing variables (App E)

– zOp.pressure.nd

Final Tobit regression with base plus five extra

– The potentially missing variables are either nonsignificant or impact magnitude is very small

– Model must not necessarily handle all special conditions (ARegV § 15(1))

Conclusions

Conclusions

Robust model

- Conceptually (model specification and OLS validation)
- Statistically (hedge to both data errors and errors in functional form)
- Intuitive and international experience

Good data quality

- Multiple validation steps, double outlier detection

Inefficiency

- 12.7%

Cautious

- Many cost drivers
- Outlier criteria harsh
- Best-of-four

We feel comfortable about the results

SUMICSID

SWEDEN–BELGIUM–DENMARK