ECONOMETRIC ANALYSIS

Benchmarking Parameters

AGRELL

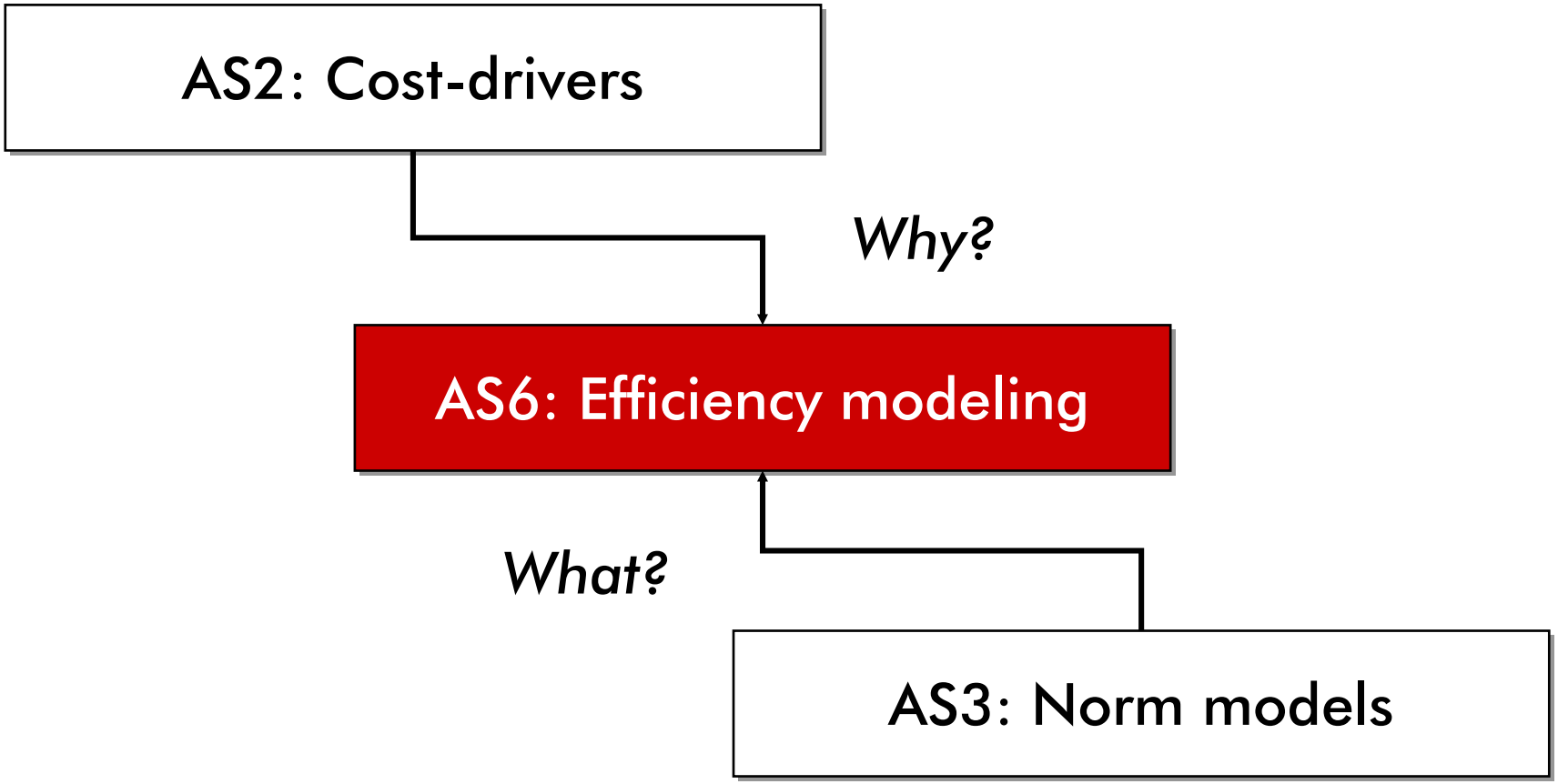
BOGETOFT

2006-02-28

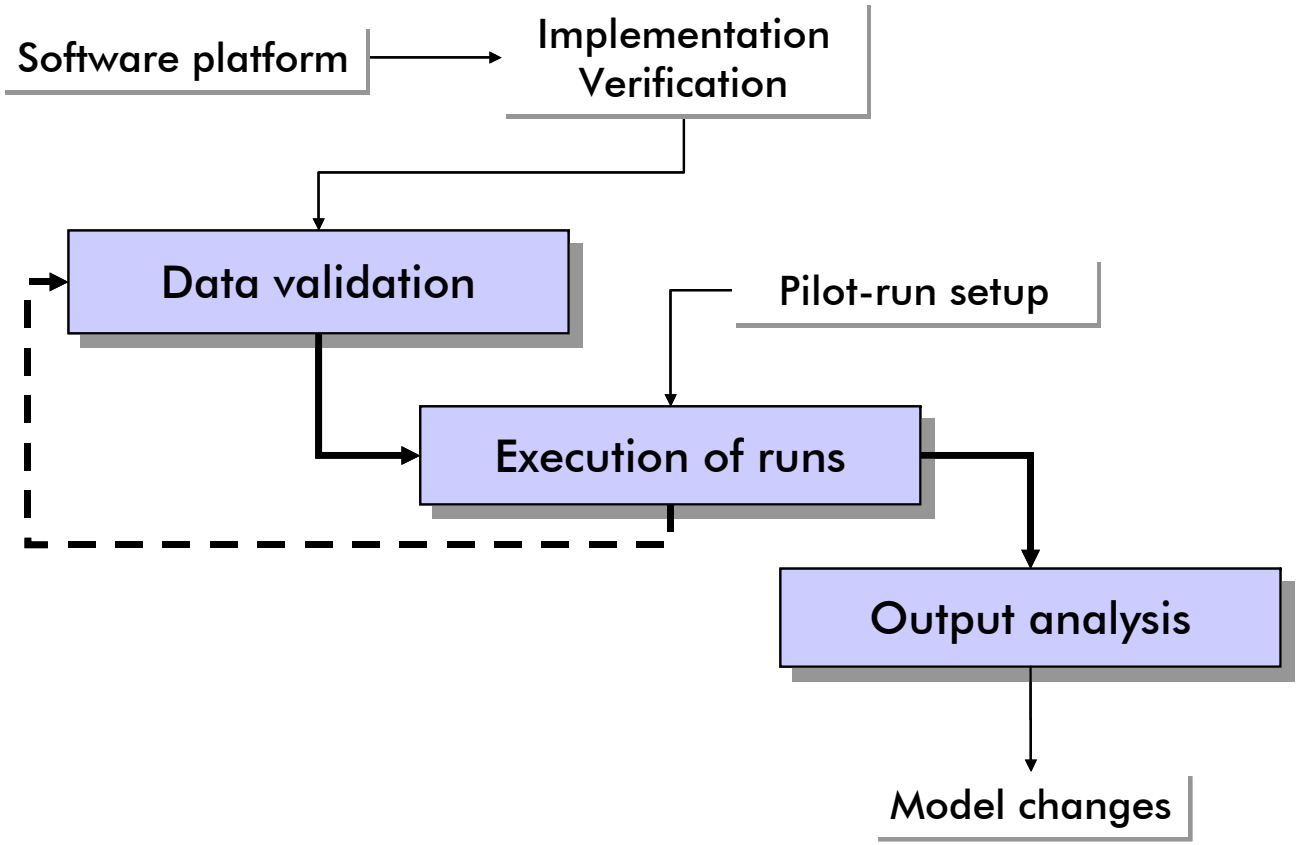
Outline

- Process
  - Project integration
  - Implementation
- Data validation
- Parameter significance tests
- Benchmarking methods
  - Model structure
  - DEA models
  - SFA models
- Regulatory Robustness
- Summary

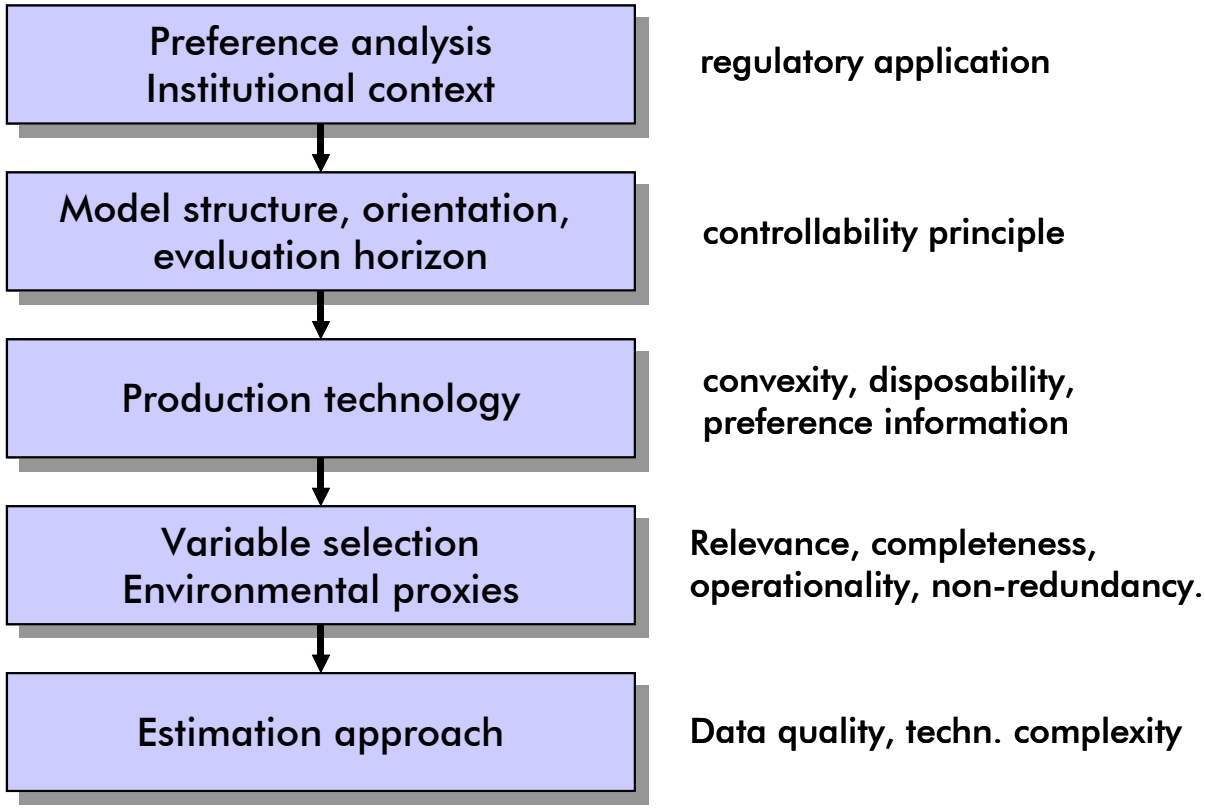
Project integration



Implementation



Process


Preference analysis
Institutional context
regulatory application
Model structure, orientation,
evaluation horizon
controllability principle
Production technology
convexity, disposability,
preference information
Variable selection
Environmental proxies
Relevance, completeness,
operationality, non-redundancy.
Estimation approach
Data quality, techn. complexity

Preference analysis

Some distinctions

Descriptive statistical models

- Significance test for technically and empirically relevant variables on BNA data

Benchmarking models (ex post)

- Determine best practice performance for the past period using hindsight

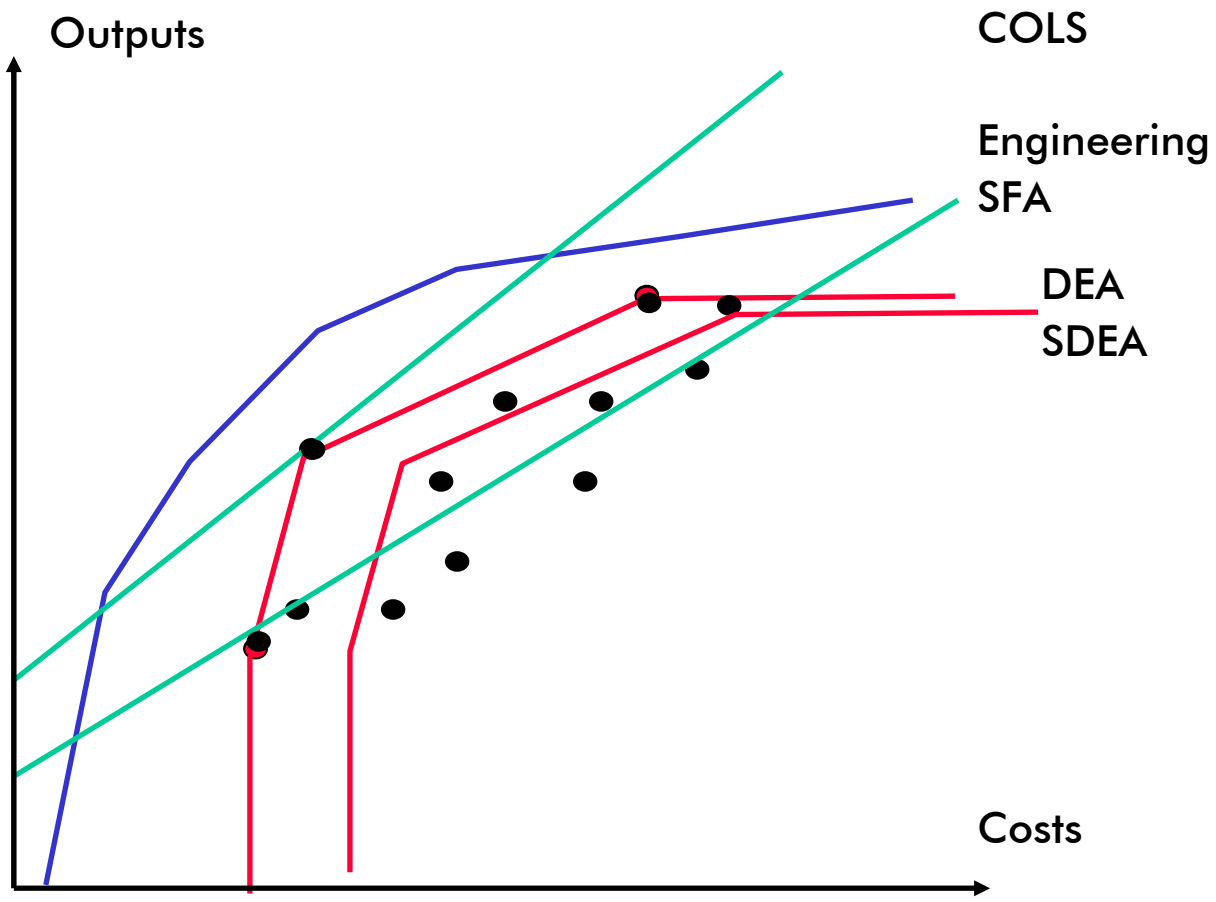
Benchmarking models (ex ante)

- Provide improvement targets through best practice based on a long-term robust specification.

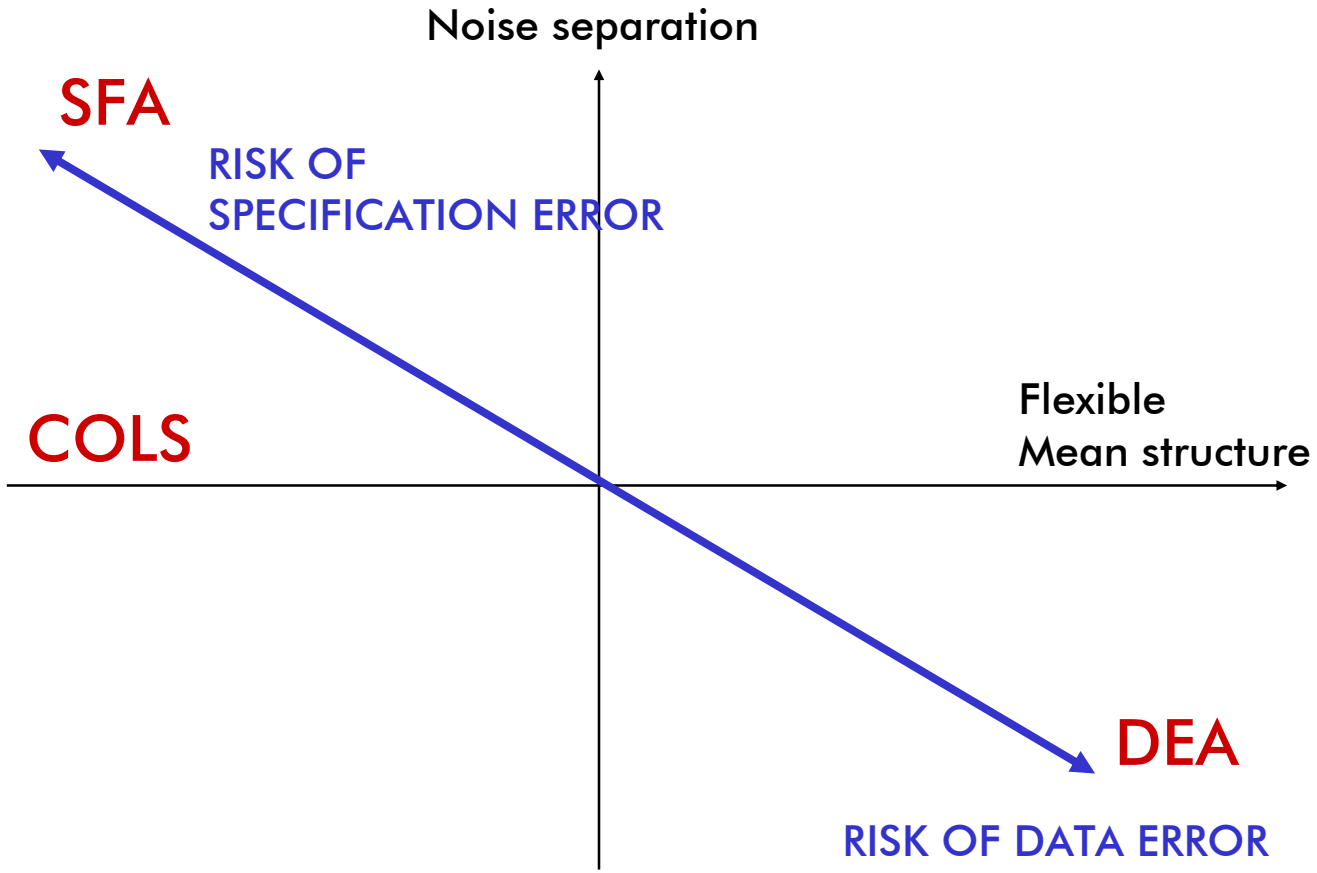
Model taxonomy

	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 Battesee (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)

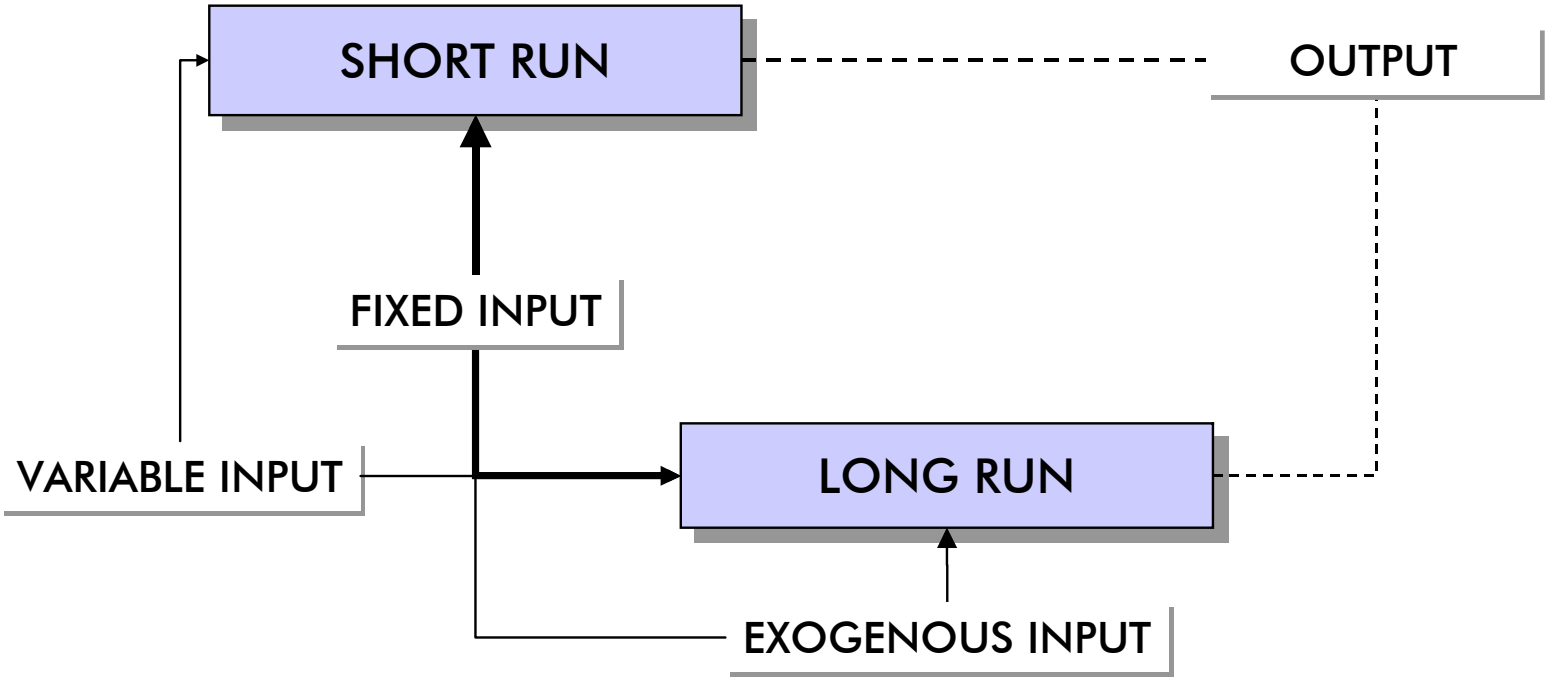
Benchmarking methods



Tradeoffs


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

Time horizon



Hybrid models

Multiple models

- Right info for right question
- Putting everybody in the best light

Combined models

- DEA-SFA: hedging for both data error and specification errors
- SR-LR: giving information about catch-up possibilities (see also T3)

Data validation notes

Methodology

AS5 Data validation tool

Outlier detection of errors

- Cascading errors
- Errors in magnitude
- Arbitrary decomposition of joint costs

Control of sign and magnitude in OLS

Comparative international studies

NS,MS,HS,UHS: Input variables

xCablesCircuit	Stromkreislänge Kabel (Km)
xCablesRoute	Trassenlänge Kabel (Km)
xCostDIR	Direkte Kosten
xCostIN	Gewälzte Kosten
xLinesCircuit	Stromkreislänge Freileitung (Km)
xLinesRoute	Trassenlänge Freileitung (Km)
xRevenue	Erlöse
xTOTEX	Kosten gesamt
xTowers	Masten

Tr NS/MS, Tr MS/HS, Tr HS/UHS: Input variables

xCablesRoute	Trassenlänge Kabel
xCostDIR	Direkte Kosten
xCostIN	Gewälzte Kosten
xInstalledPower	Installierte Leistung (MVA)
xLinesRoute	Trassenlänge Freileitung
xRevenues	Erlöse
xSubstation	Anzahl Umspannstationen
xTOTEX	Gesamtkosten

NS,MS,HS,UHS: Output variables

yArea1	SummevonFläche1
yArea2	SummevonFläche2
yArea3	SummevonFläche3
yArea4	SummevonFläche4
yArea5	SummevonFläche5
yAreaS	Versorgte Fläche (Km2)
yCustomers	Entnahmestellen
yEnergy	Jahresarbeit (MWh)
yMeters	Zählpunkte
yPeakLoad	zgl JHL (MW)
yPeakLoad_Levels	zgl JHL durch nachgelagerte Ebene

Tr NS/MS, Tr MS/HS, Tr HS/UHS: Output variables

yArea	Versorgte Fläche
yArea1	SummevonFläche1
yArea2	SummevonFläche2
yArea3	SummevonFläche3
yArea4	SummevonFläche4
yArea5	SummevonFläche5
yCustomers	Entnahmestellen
yEnergy	Jahresarbeit
yMeters	Zählpunkte
yPeakLoad	zgl Jahreshöchstlast
yPop	SummevonEinw

NS,MS,HS,UHS: Proxy variables

yPop	SummevonEinw
zCabelprop	Verkabelungsgrad
zCircuitTot	Stromkreislänge Gesamt (Km)
zCircuitTot_AreaS	Stromkreislänge/ Versorgte Fläche
zCostD_CircuitTot	Direkte Kosten/Stromkreislänge Gesamt
zEnergy_CircuitTot	Jahresarbeit/Stromkreislänge
zRoute_CircuitTot	Trassenlänge/Stromkreislänge
zRouteTot	Trassenlänge Gesamt (Km)
zTow_km	Masten pro Km

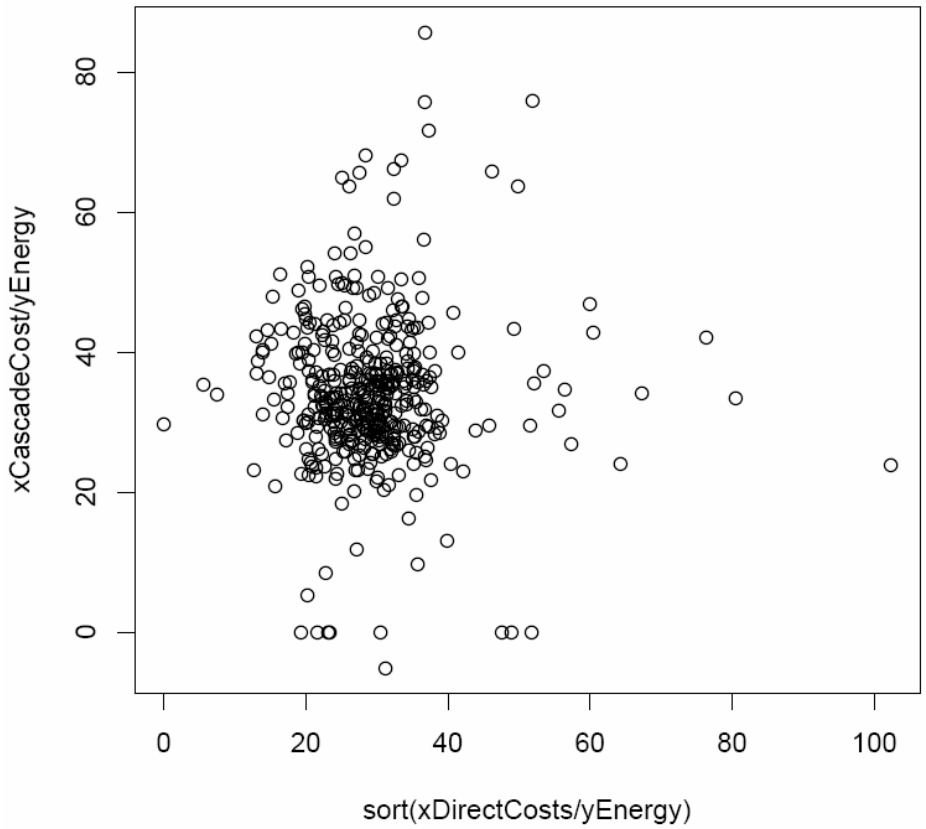
Tr NS/MS, Tr MS/HS, Tr HS/UHS: Proxy variables

zCostD_MWh	Kosten/MWh
zEnergy_PeakLoad	Jahresarbeit/zgl JHL
zInstPower_PeakLoad	Installierte Leistung/zgl JHL
zInstPower_Substations	Instllierte Leistung/Anzahl Stationen
zMWh_Customers	MWh/Zählpunkt
zRevenue_MWh	Erlöse/MWh

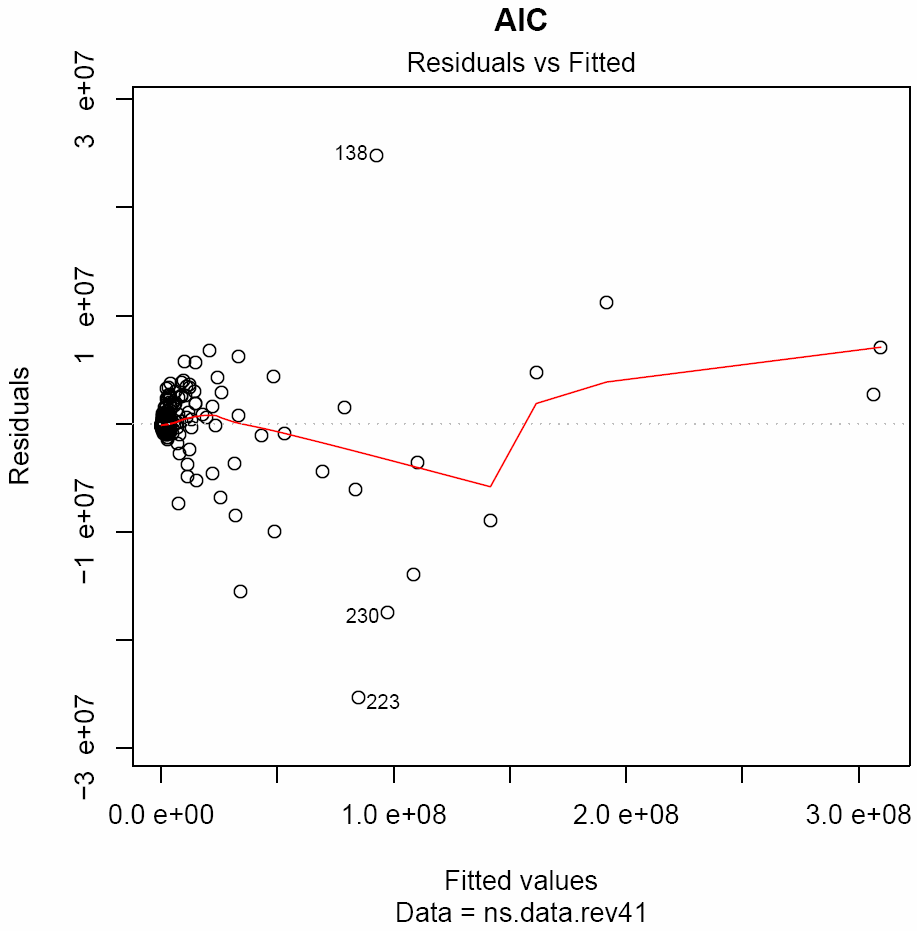
Data set

	Available DMU	Validated DMU
Ultra high- voltage level (UHS)	5	5
High-voltage level (HS)	93	39
Medium-voltage level (MS)	804	519
Low-voltage level (NS)	835	440
Ultra high-voltage level/high-voltage level (TrUHS/HS)	19	11
High-voltage level/medium-voltage level (Tr HS/MS)	183	69
Medium-voltage level/low-voltage level (Tr MS/NS)	814	222

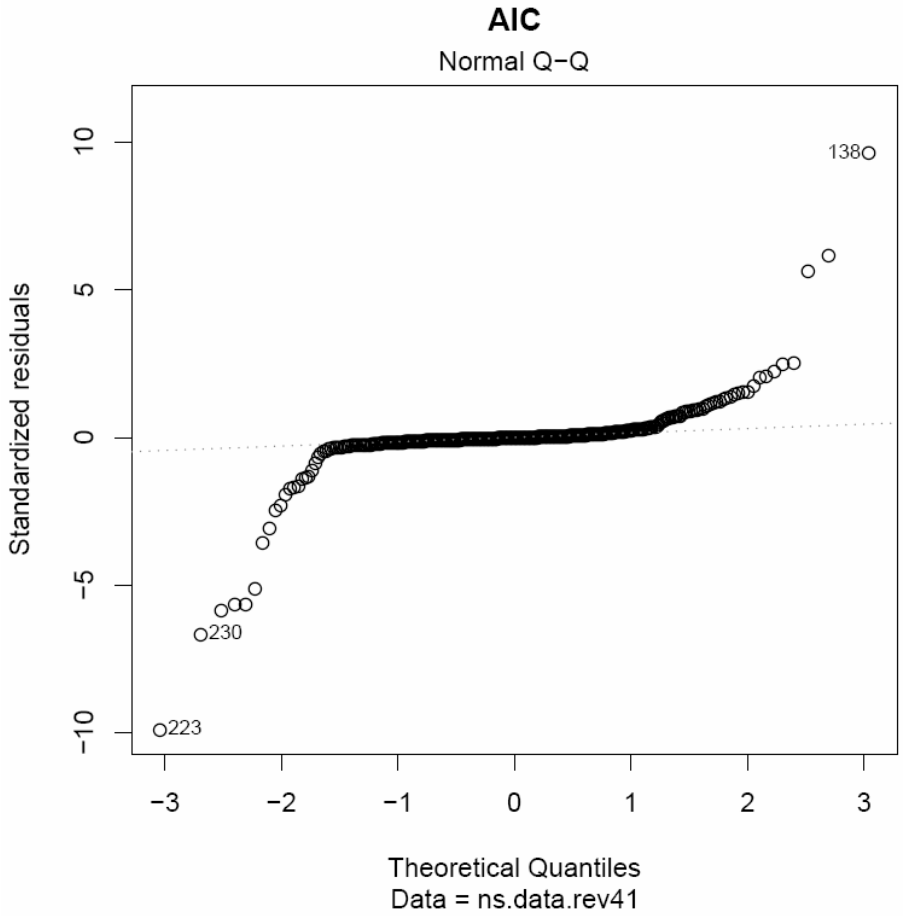
Data dispersion (1)



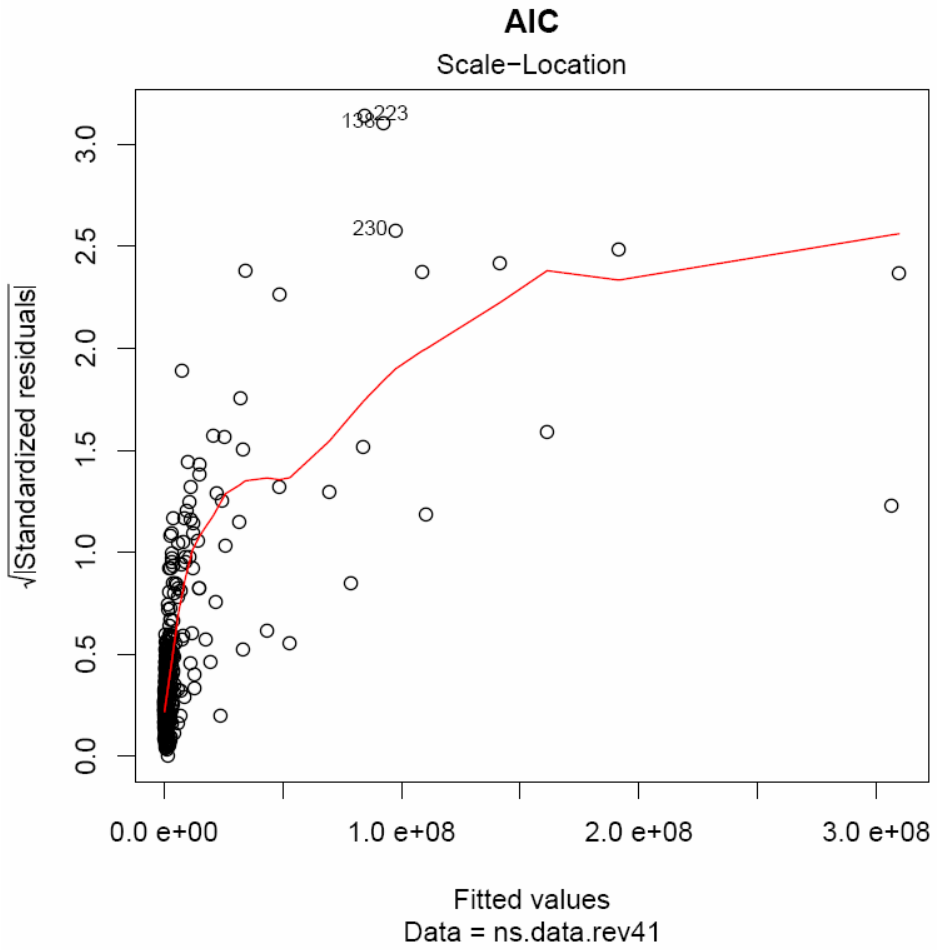
Outlier analysis OLS (1)



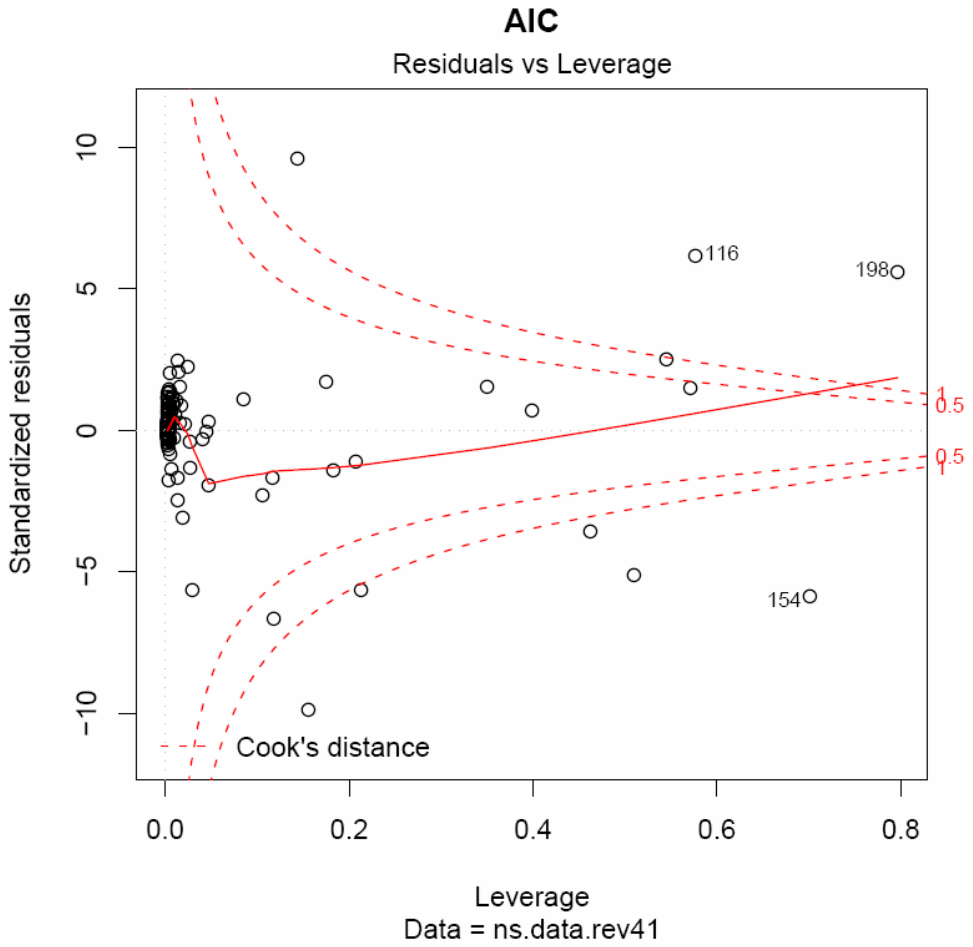
Outlier analysis OLS (2)



Outlier analysis OLS (3)

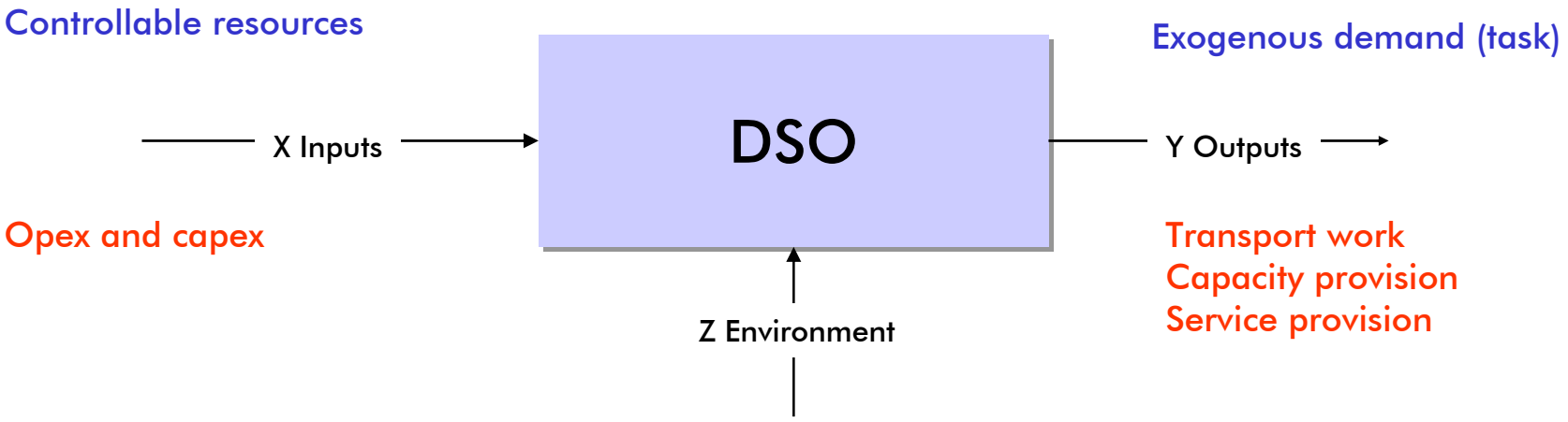


Outlier analysis OLS (4)



Model structure

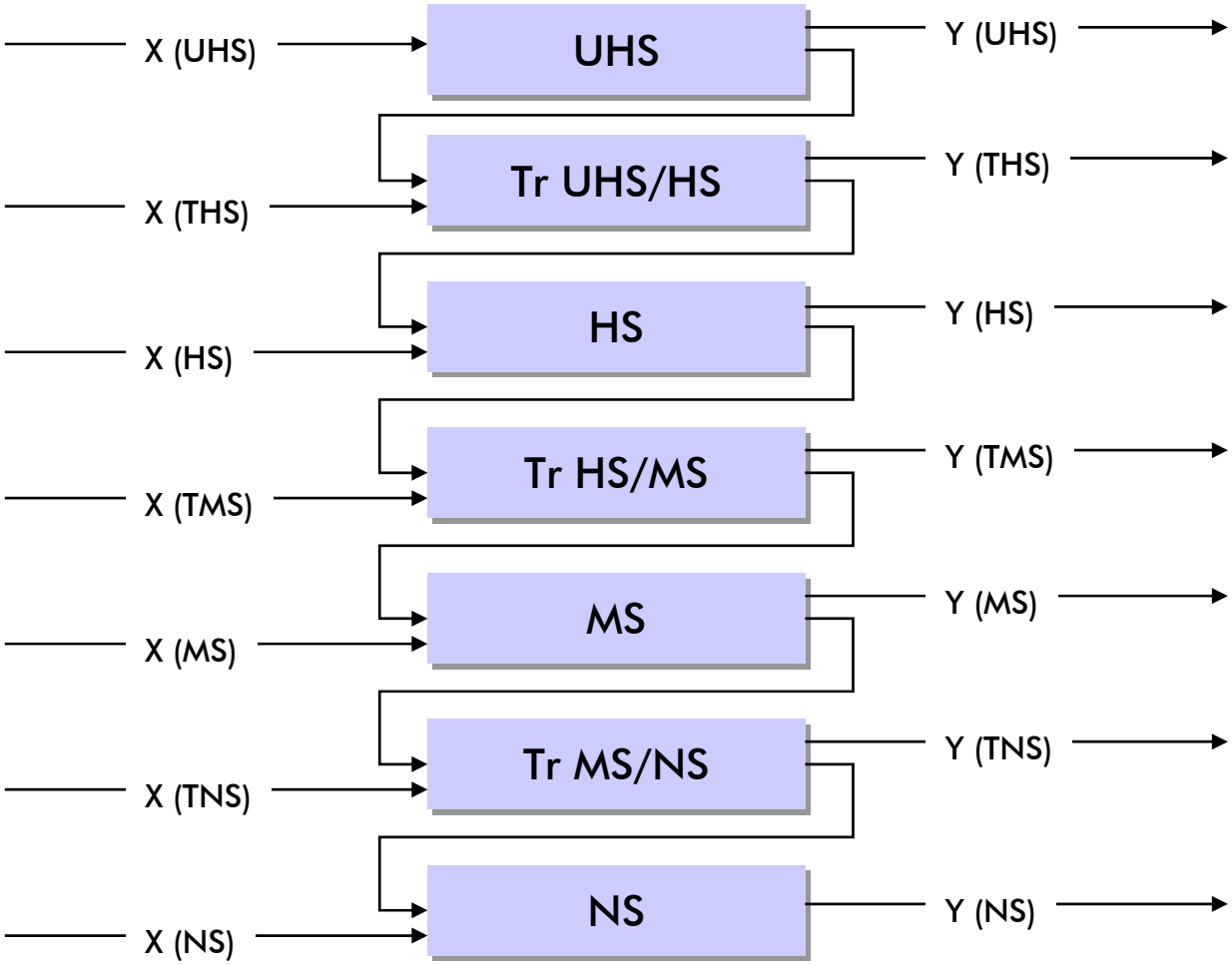
Model structure



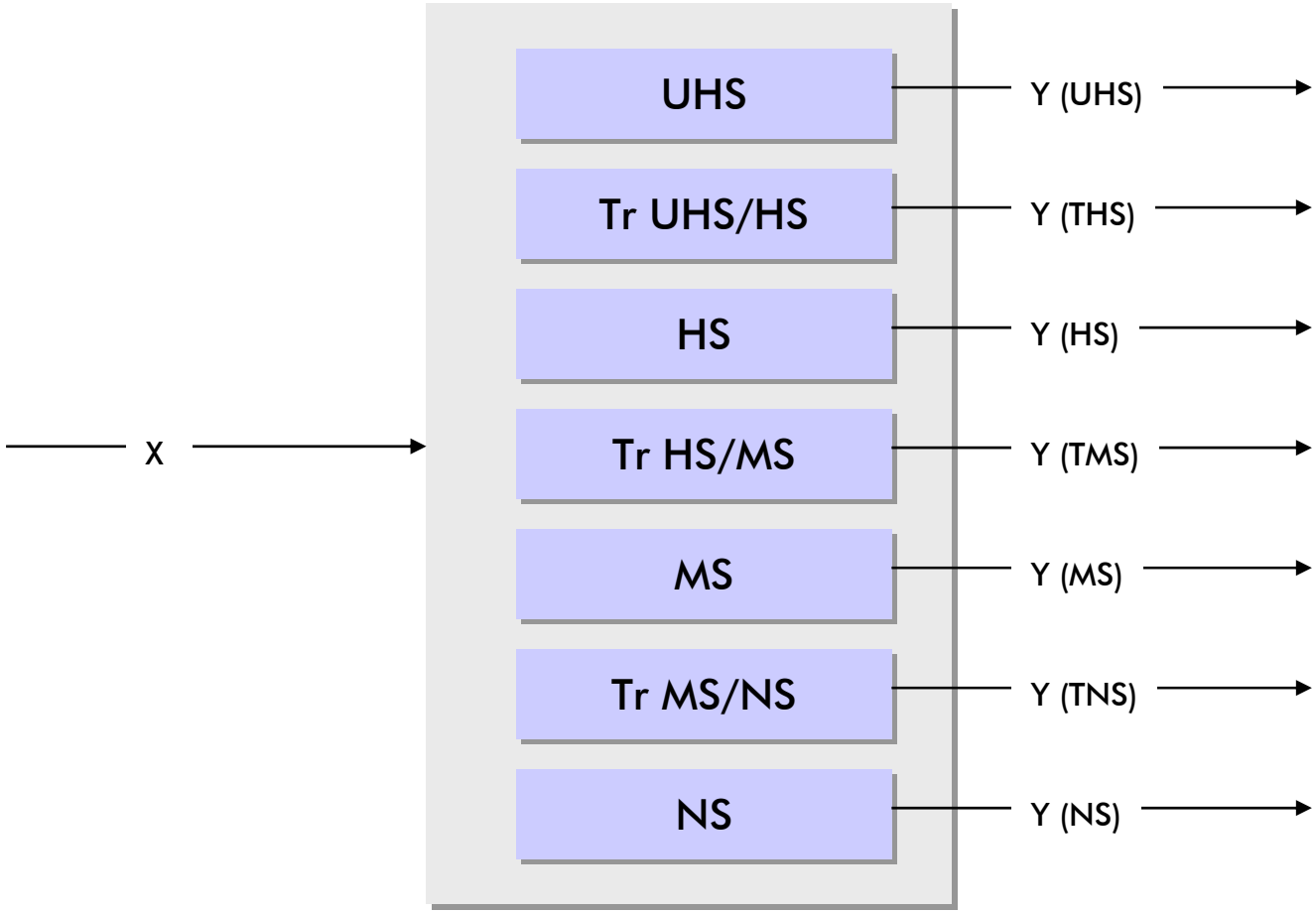
Estimation approach

1. Costs C(X) | Resources X
2. Resources X(Y) | Delivery task Y
3. Costs C(Y) | Delivery task Y
4. Costs C(Y|Z) | Delivery task Y | Environment Z

Separable model



Integrated model



T1: Subadditivity test

Choice of separable model

- $C(x,y_1,y_2,\ldots)$ = Integrated cost function
- $C_1(x_1,y_1,w_1)$ = Voltage level cost function, intermediate output w_1

Subadditivity

- $C(x,y_1,y_2,\ldots) \leq C_1(x_1,y_1,w_1) + C_2(x_2,y_2,w_2) + \ldots$

T1: Testing integrated and separable sets

- If subadditivity is not rejected, run integrated model

T2: Age effect test

Problem:

- Accounting measures (investment cycle) give bias in favour of older networks in capex
- Operating cost may have an age bias in favour of newer grids

T2: Testing C(y|z) with age proxies

- If age hypothesis is not rejected, add age to Z

T3: Decomposition tests

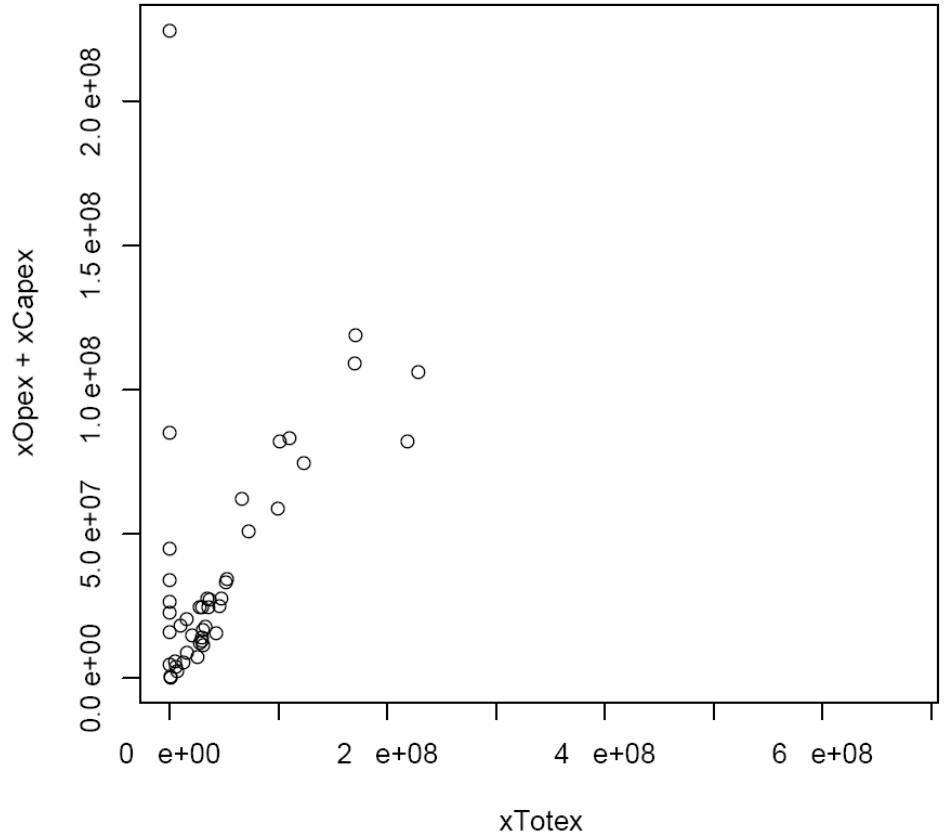
Problem:

- The incumbent inefficiency in grid asset valuation (capex) gives bounds for feasible productivity improvement targets
- Depends on imposed accounting practices

T3: Testing models EOpex(y|z), ECapex(y|z)

- If EOpex(y|z) >> ECapex(y|z) then inefficiency is primarily a question of sunk investments

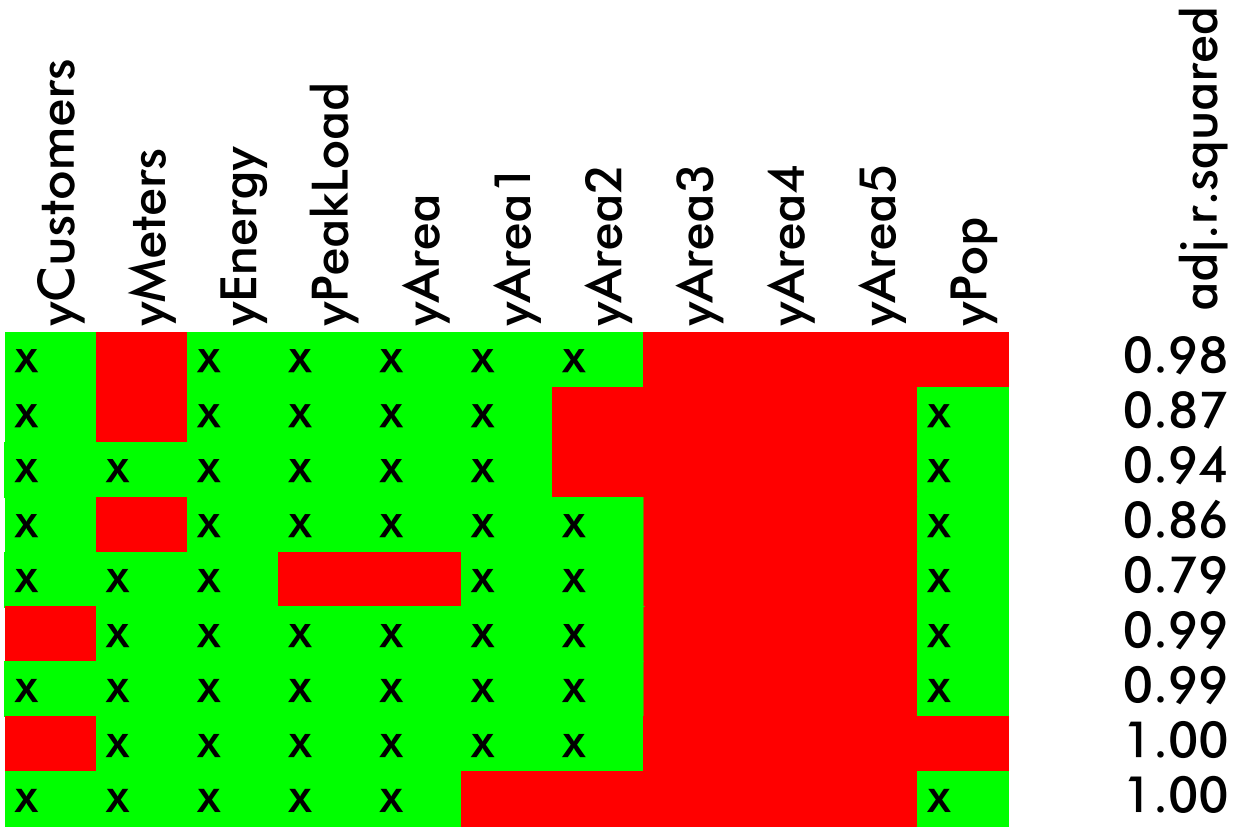
Capex – Opex decomposition



Regression results

Variables NS

low-voltage



Low voltage – best model – xCostDIR

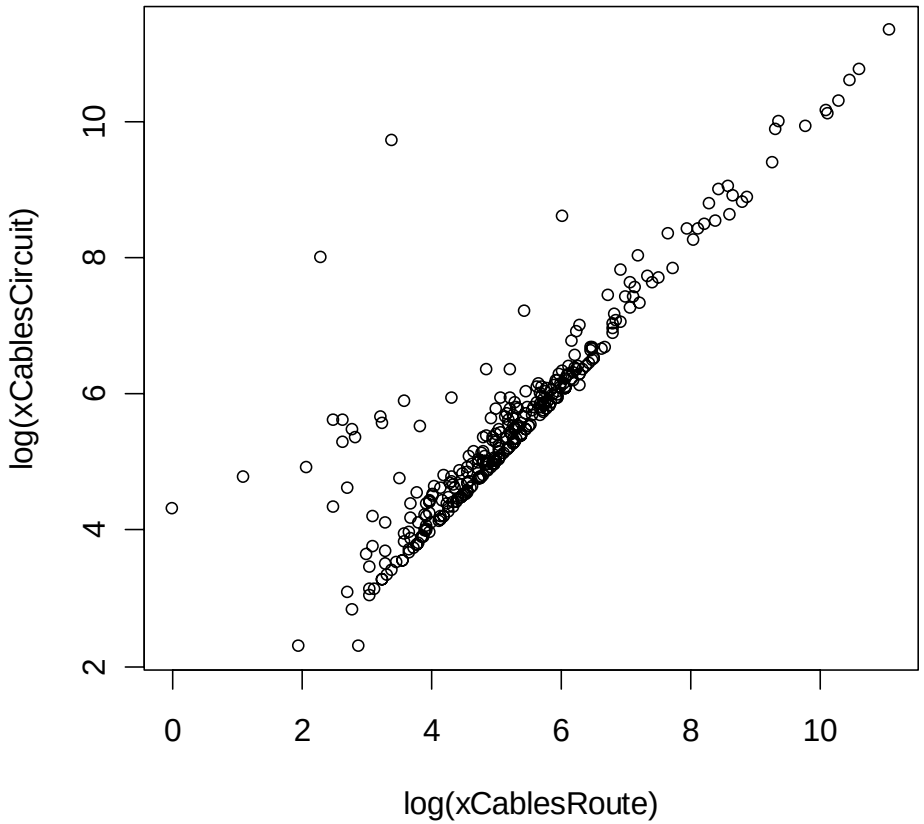
Voltage low-voltage
x-variable xCostDIR
Best x-y model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.415E+05	143380.15	0.99	0.32
yCustomers				
yMeters	-3.187E+01	5.59	-5.70	0.00
yEnergy	3.706E+01	1.65	22.41	0.00
yPeakLoad	-3.589E+04	3753.36	-9.56	0.00
yArea	-2.357E+03	681.88	-3.46	0.00
yArea1	2.304E+03	774.55	2.97	0.00
yArea2	1.109E+04	5077.29	2.18	0.03
yArea3				
yArea4				
yArea5				
yPop	5.511E+00	3.17	1.74	0.08
	adj.r.squarec	0.99		

Correlation, low voltage, input vs. input variables

cor.elec.ns.data.rev4.060226 234137.csv									
	xCircuitCable	xCircuitLines	xCablesRoute	xLinesRoute	xTowers	xDirectCosts	xCascadeCost	xrevenues	xTotex
xCircuitCable	**1.00**	0.71	**0.98**	0.68	0.53	**0.96**	0.93	0.95	0.95
xCircuitLines	0.71	**1.00**	0.72	**0.98**	0.89	0.82	0.80	0.81	0.81
xCablesRoute	**0.98**	0.72	**1.00**	0.72	0.56	0.92	0.90	0.92	0.92
xLinesRoute	0.68	**0.98**	0.72	**1.00**	0.90	0.78	0.77	0.78	0.78
xTowers	0.53	0.89	0.56	0.90	**1.00**	0.61	0.65	0.64	0.64
xDirectCosts	**0.96**	0.82	0.92	0.78	0.61	**1.00**	**0.97**	**0.99**	**0.99**
xCascadeCost	0.93	0.80	0.90	0.77	0.65	**0.97**	**1.00**	**0.99**	**0.99**
xrevenues	0.95	0.81	0.92	0.78	0.64	**0.99**	**0.99**	**1.00**	**1.00**

Cables (NS)



Correlation, low voltage, input vs. output variables

	yCustomers	yMeters	yEnergy	yPeakLoad	yArea	yArea1	yArea2	yArea3	yArea4	yArea5	yPop
xCircuitCable	0.68	0.90	0.94	0.91	0.91	0.91	0.94	0.92	0.92	0.91	0.93
xCircuitLines	0.71	0.77	0.78	0.78	0.85	0.86	0.87	0.86	0.86	0.86	0.82
xCablesRoute	0.65	0.86	0.90	0.88	0.93	0.92	0.94	0.93	0.93	0.92	0.90
xLinesRoute	0.69	0.75	0.75	0.74	0.83	0.84	0.84	0.84	0.84	0.84	0.79
xTowers	0.57	0.60	0.57	0.54	0.74	0.73	0.71	0.72	0.73	0.73	0.64
xDirectCosts	0.83	**0.97**	**0.99**	**0.97**	0.87	0.88	0.94	0.90	0.89	0.88	**0.98**
xCascadeCost	0.87	**0.99**	**0.98**	**0.96**	0.86	0.86	0.91	0.88	0.86	0.86	**0.99**
xrevenues	0.86	**0.98**	**0.99**	**0.97**	0.87	0.87	0.93	0.89	0.88	0.88	**0.99**
xTotex	0.86	**0.99**	**0.99**	**0.97**	0.87	0.87	0.93	0.89	0.88	0.88	**0.99**

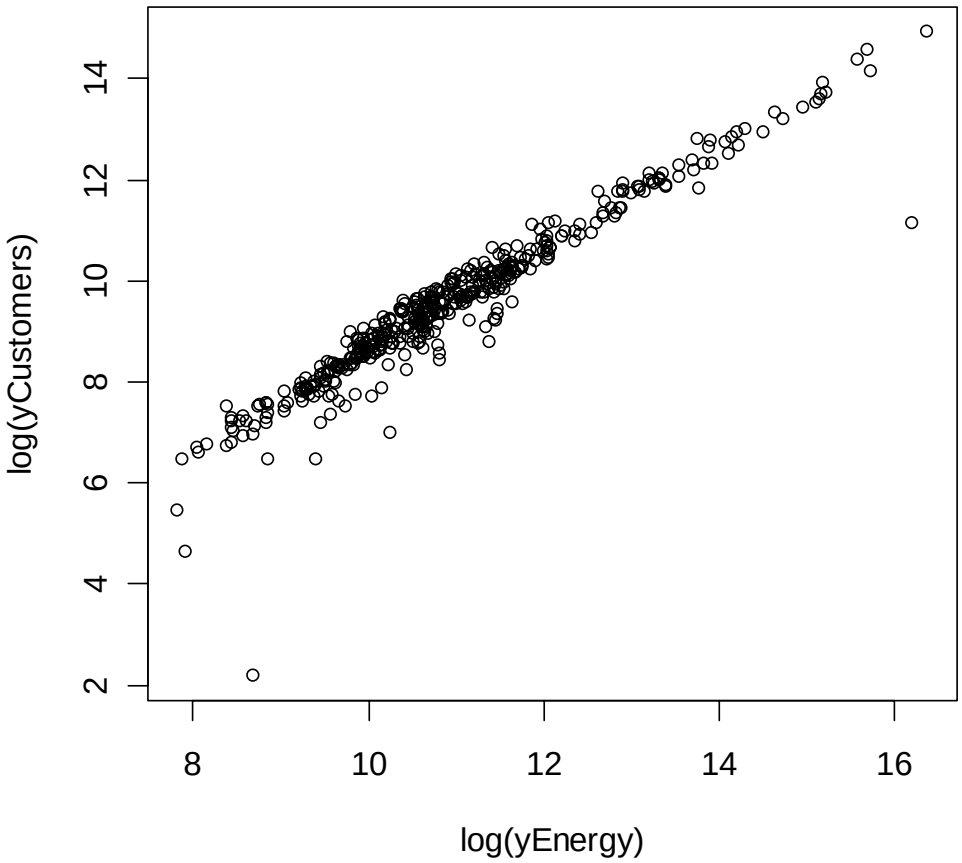
Correlation, low voltage, input vs. proxy variables

	zLengthkm	zRoutekm	zCableprop	zTowerkm	zCircuit_Area	zCostD_Circuit	zRoute_Circuit	zEnergyRoute	zPeakHours
xCircuitCable	**0.99**	**0.96**	-0.04	0.03	-0.02	-0.07	0.01	-0.04	0.00
xCircuitLines	0.81	0.83	-0.23	0.01	-0.09	-0.07	0.03	-0.07	0.00
xCablesRoute	**0.97**	**0.98**	-0.05	0.03	-0.05	-0.07	0.05	-0.05	-0.01
xLinesRoute	0.78	0.83	-0.22	0.02	-0.09	-0.07	0.05	-0.07	0.00
xTowers	0.64	0.68	-0.24	0.08	-0.08	-0.07	0.04	-0.07	0.01
xDirectCosts	**0.97**	0.94	-0.09	0.04	-0.01	-0.04	0.01	0.00	0.00
xCascadeCost	0.95	0.92	-0.08	0.05	0.01	-0.05	0.01	0.01	0.00
xrevenues	**0.97**	0.93	-0.08	0.05	0.01	-0.05	0.01	0.01	0.00
xTotex	**0.97**	0.94	-0.08	0.05	0.00	-0.04	0.01	0.00	0.00

Correlation, low voltage, output vs. output variables

	yCustomers	yMeters	yEnergy	yPeakLoad	yArea	yArea1	yArea2	yArea3	yArea4	yArea5	yPop
yCustomers	**1.00**	0.91	0.87	0.86	0.58	0.58	0.68	0.61	0.59	0.58	0.86
yMeters	0.91	**1.00**	**0.98**	**0.97**	0.79	0.80	0.87	0.82	0.81	0.80	**0.98**
yEnergy	0.87	**0.98**	**1.00**	**0.99**	0.83	0.84	0.91	0.86	0.85	0.84	**0.98**
yPeakLoad	0.86	**0.97**	**0.99**	**1.00**	0.81	0.82	0.89	0.84	0.83	0.82	**0.97**
yArea	0.58	0.79	0.83	0.81	**1.00**	**1.00**	**0.98**	**0.99**	**1.00**	**1.00**	0.87
yArea1	0.58	0.80	0.84	0.82	**1.00**	**1.00**	**0.98**	**1.00**	**1.00**	**1.00**	0.87
yArea2	0.68	0.87	0.91	0.89	**0.98**	**0.98**	**1.00**	**0.99**	**0.99**	**0.98**	0.93
yArea3	0.61	0.82	0.86	0.84	**0.99**	**1.00**	**0.99**	**1.00**	**1.00**	**1.00**	0.89
yArea4	0.59	0.81	0.85	0.83	**1.00**	**1.00**	**0.99**	**1.00**	**1.00**	**1.00**	0.88
yArea5	0.58	0.80	0.84	0.82	**1.00**	**1.00**	**0.98**	**1.00**	**1.00**	**1.00**	0.88
yPop	0.86	**0.98**	**0.98**	**0.97**	0.87	0.87	0.93	0.89	0.88	0.88	**1.00**

Customers vs Energy (NS)



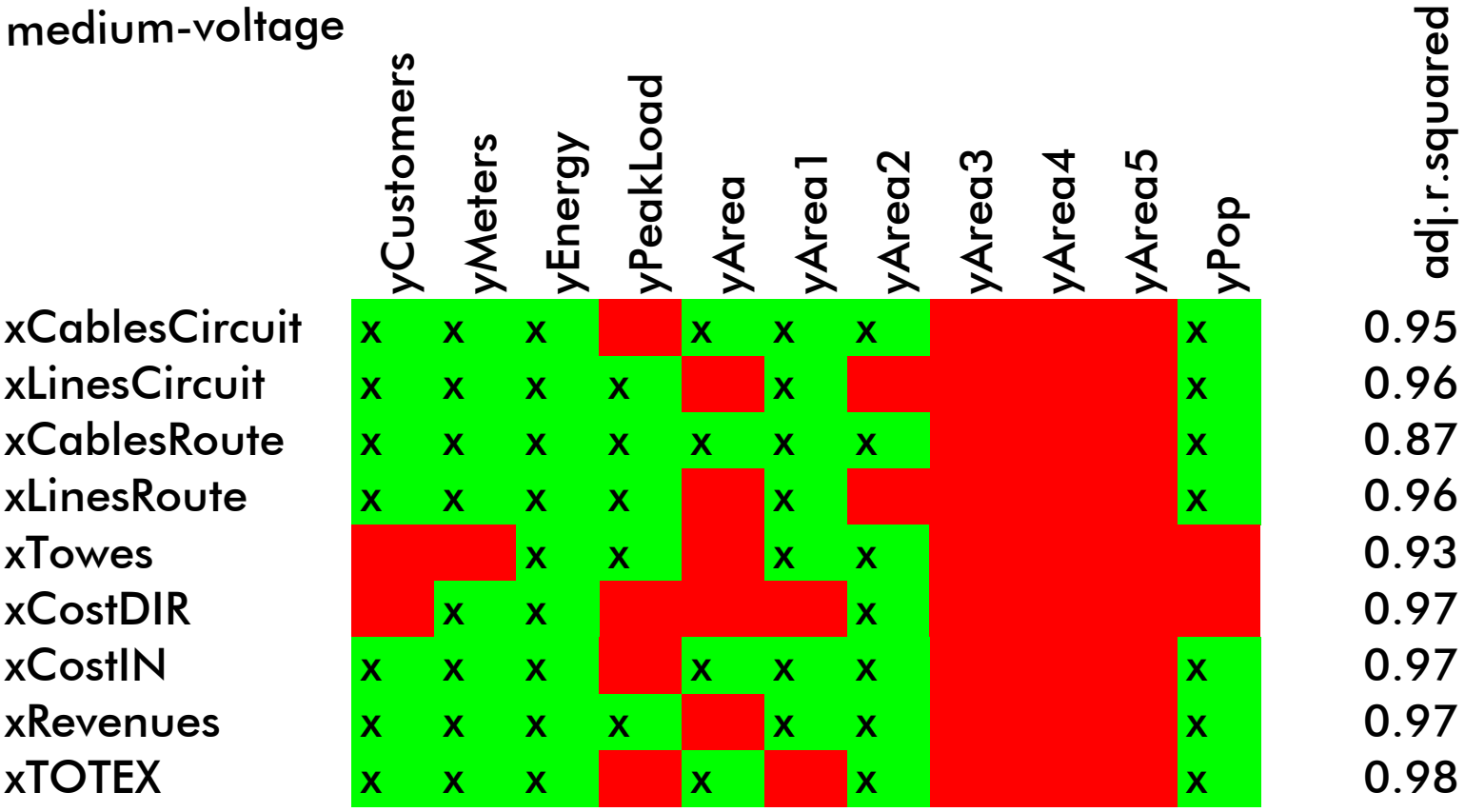
Low voltage – marginal extensions – xCostDIR

Voltage low-voltage
x-variable xCostDIR
Marginal extensions

	RSS	AIC	F Value	Pr(F)
<none>	3.225E+15	12563.18	NA	NA
yCustomer	3.217E+15	12564.08	1.08	0.30
yMeters				
yEnergy				
yPeakLoad				
yArea				
yArea1				
yArea2				
yArea3	3.225E+15	12563.18	NA	NA
yArea4	3.225E+15	12563.18	NA	NA
yArea5	3.225E+15	12563.18	NA	NA
yPop				
zLengthkm	2.548E+15	12465.57	109.92	0.00
zRoutekm	3.120E+15	12551.20	13.91	0.00
zCablepro	3.214E+15	12563.73	1.43	0.23
zTowerkm	3.223E+15	12564.93	0.25	0.62
zCircuit_Ar	3.219E+15	12564.37	0.80	0.37
zCostD_Ci	3.175E+15	12558.52	6.57	0.01
zRoute_Cir	3.224E+15	12565.09	0.09	0.77
zEnergyRo	3.208E+15	12562.89	2.25	0.13
zPeakHour	3.222E+15	12564.78	0.39	0.53

Variables MS

medium-voltage



	yCustomers	yMeters	yEnergy	yPeakLoad	yArea	yArea1	yArea2	yArea3	yArea4	yArea5	yPop	adj.r.squared
xCablesCircuit	x	x	x		x	x	x				x	0.95
xLinesCircuit	x	x	x	x		x					x	0.96
xCablesRoute	x	x	x	x	x	x	x				x	0.87
xLinesRoute	x	x	x	x		x					x	0.96
xTowes			x	x		x	x					0.93
xCostDIR		x	x				x					0.97
xCostIN	x	x	x		x	x	x				x	0.97
xRevenues	x	x	x	x		x	x				x	0.97
xTOTEX	x	x	x		x		x				x	0.98

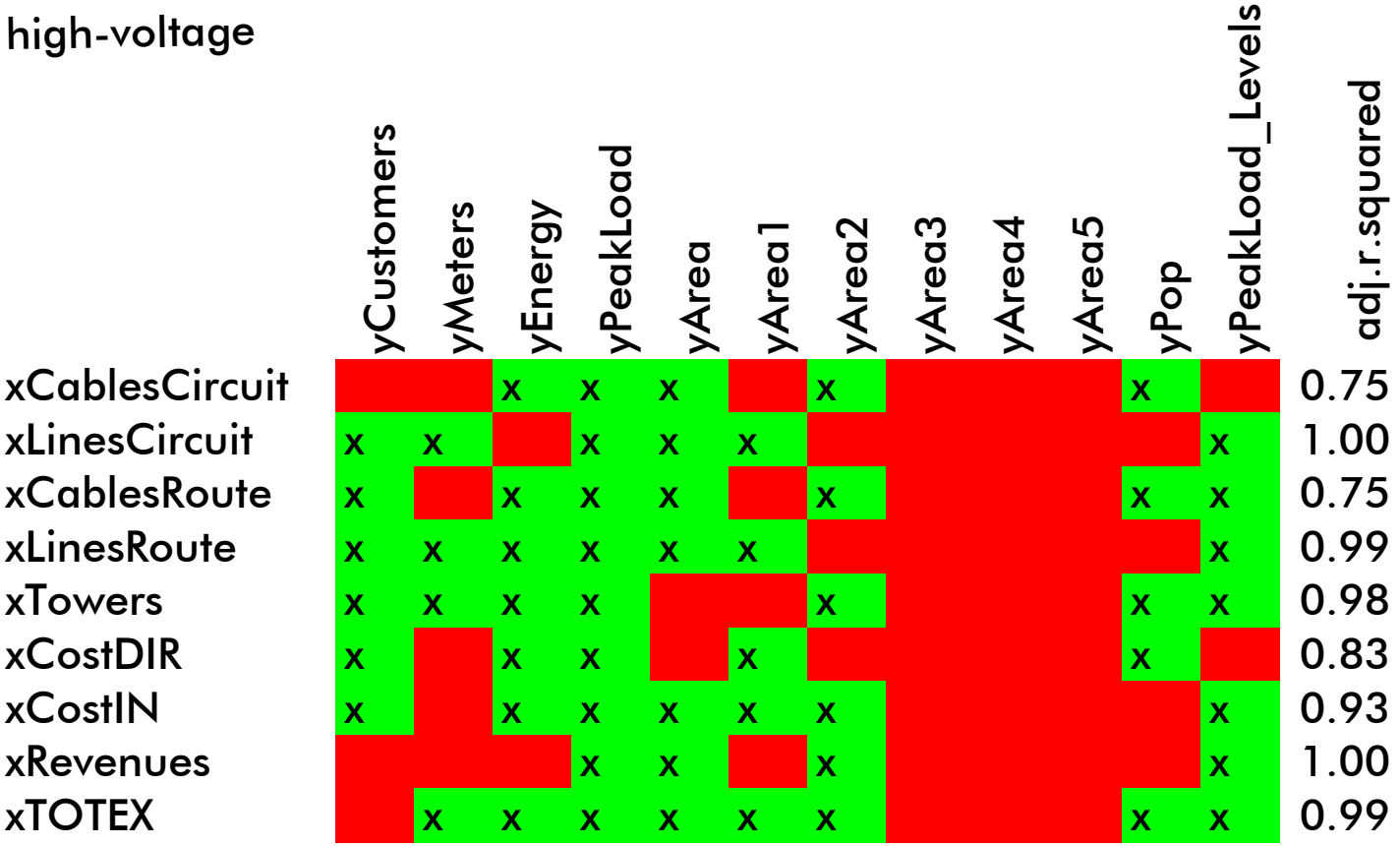
Medium voltage – best model – xCostDIR

Voltage medium-voltage
x-variable xCostDIR
Best x-y model

	Std. Error	t value	Pr(>|t|)	
(Intercept)	3.338E+05	161810.89	2.06	0.04
yCustomers				
yMeters	1.455E+02	57.38	2.54	0.01
yEnergy	1.086E+01	0.29	37.55	0.00
yPeakLoad				
yArea				
yArea1				
yArea2	2.554E+04	1278.47	19.98	0.00
yArea3				
yArea4				
yArea5				
yPop				
	adj.r.squared	0.97		

Variables HS

high-voltage



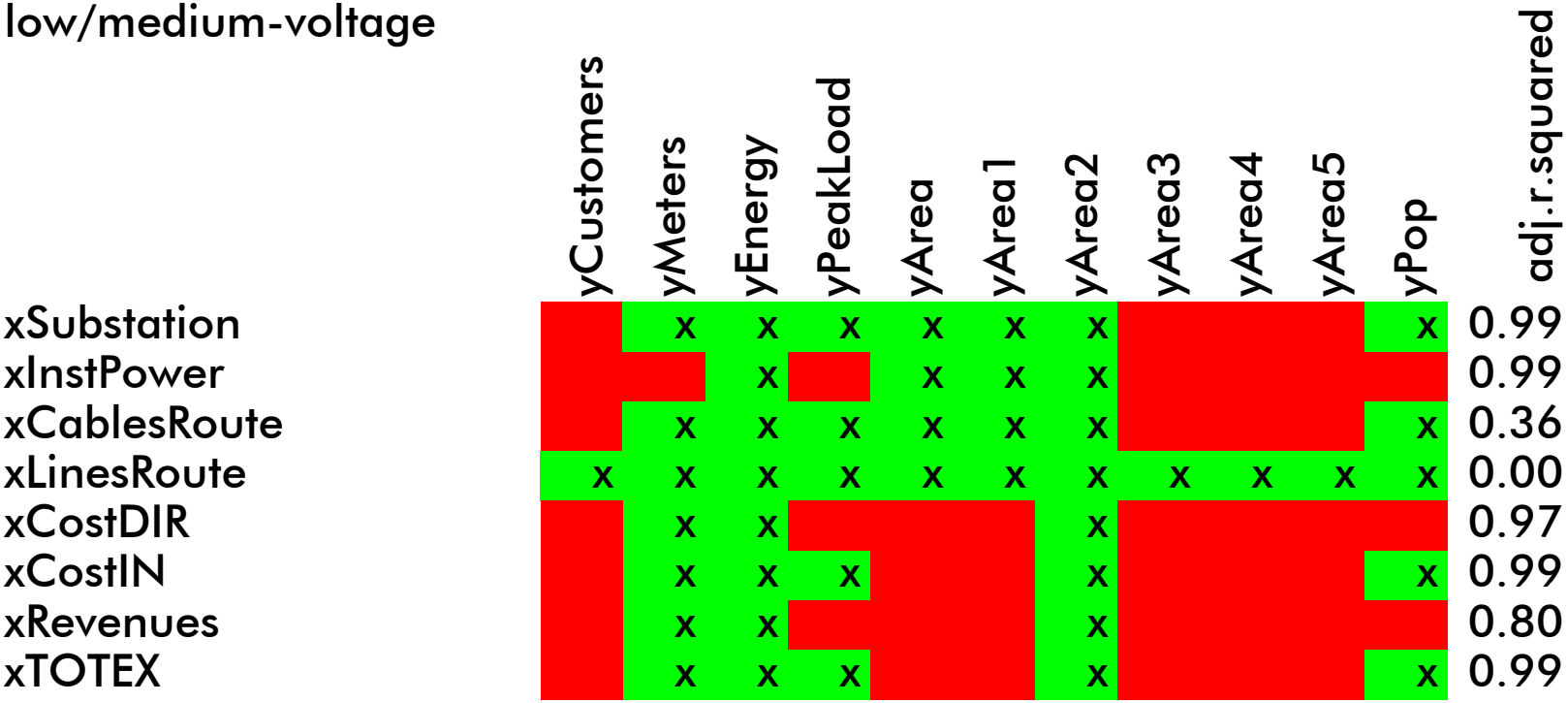
High voltage – best model – xCostDIR

Voltage high-voltage
x-variable xCostDIR
Best x-y model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-3.352E+06	7926695.92	-0.42	0.68
yCustomers	1.663E+05	100915.58	1.65	0.11
yMeters				
yEnergy	7.370E+01	15.68	4.70	0.00
yPeakLoad	-4.250E+05	91998.39	-4.62	0.00
yPeakLoad_Levels				
yArea				
yArea1	-4.590E+03	1734.26	-2.65	0.01
yArea2				
yArea3				
yArea4				
yArea5				
yPop	5.880E+01	20.21	2.91	0.01
	adj.r.squared	0.83		

Variables Transf NS/MS

low/medium-voltage



	yCustomers	yMeters	yEnergy	yPeakLoad	yArea	yArea1	yArea2	yArea3	yArea4	yArea5	yPop	adj.r.squared
xSubstation		x	x	x	x	x	x				x	0.99
xInstPower			x		x	x	x					0.99
xCablesRoute		x	x	x	x	x	x				x	0.36
xLinesRoute	x	x	x	x	x	x	x	x	x	x	x	0.00
xCostDIR		x	x				x					0.97
xCostIN		x	x	x			x				x	0.99
xRevenues		x	x				x					0.80
xTOTEX		x	x	x			x				x	0.99

Low/medium-voltage – best model – xSubstation

Voltage low/medium-voltage
x-variable xSubstation
Best x-y model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-6.228E+01	33.39	-1.87	0.06
yCustomers				
yMeters	-7.423E-02	0.02	-3.51	0.00
yEnergy	3.644E-03	0.00	15.15	0.00
yPeakLoad	-3.352E+00	0.62	-5.38	0.00
yArea	1.835E-01	0.07	2.67	0.01
yArea1	2.425E-01	0.09	2.80	0.01
yArea2	2.896E+00	0.80	3.63	0.00
yArea3				
yArea4				
yArea5				
yPop	-5.368E-03	0.00	-14.09	0.00
	adj.r.squared	0.99		

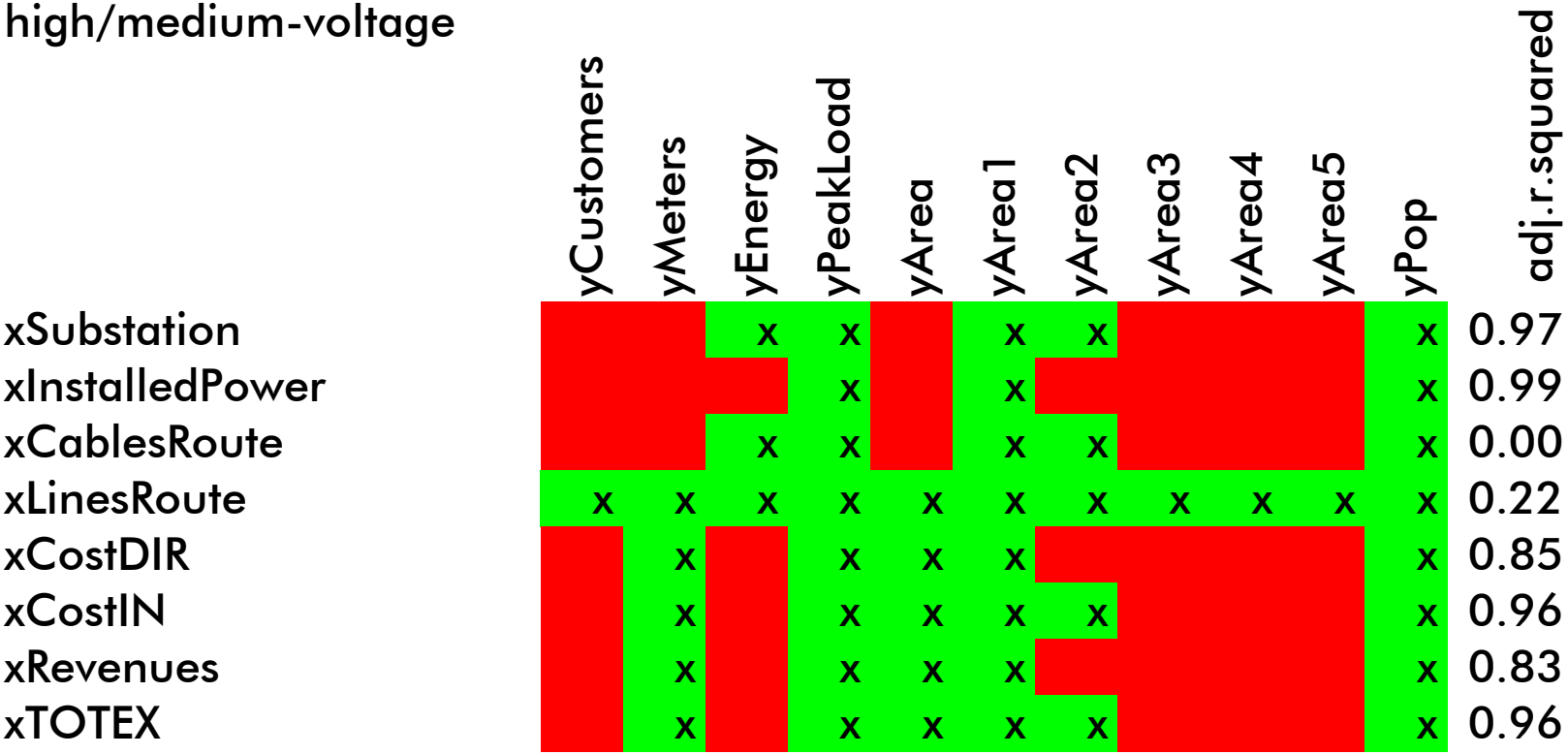
Low/medium-voltage – best model – xInstPower

Voltage low/medium-voltage
x-variable xInstPower
Best x-y model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.789E+01	9.40	1.90	0.06
yCustomers				
yMeters	2.047E-02	0.01	3.40	0.00
yEnergy	1.035E-03	0.00	15.26	0.00
yPeakLoad	-1.499E+00	0.18	-8.54	0.00
yArea				
yArea1				
yArea2	8.382E-01	0.06	14.96	0.00
yArea3				
yArea4				
yArea5				
yPop	-3.588E-04	0.00	-3.55	0.00
	adj.r.squared	0.99		

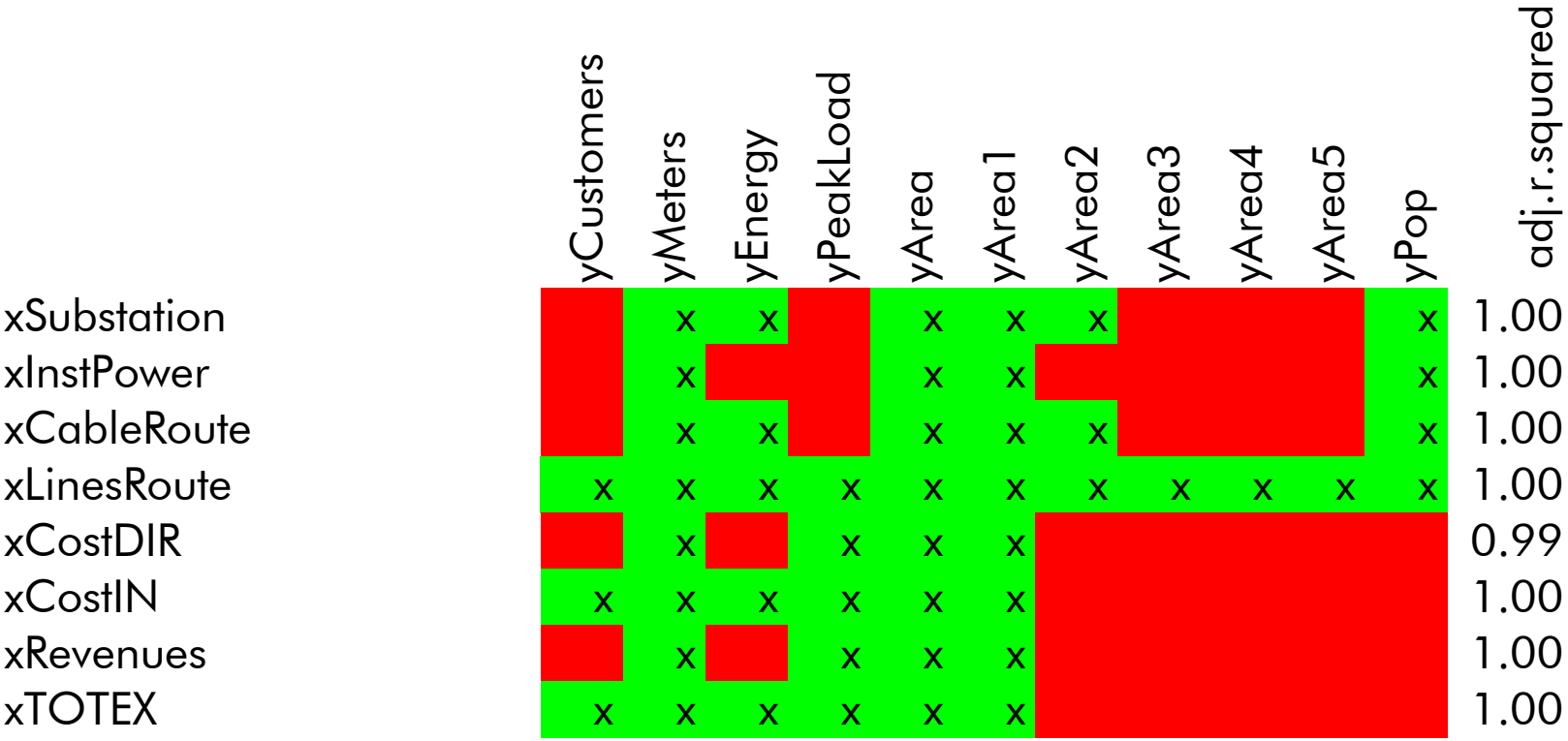
Variables Transf MS/HS

high/medium-voltage



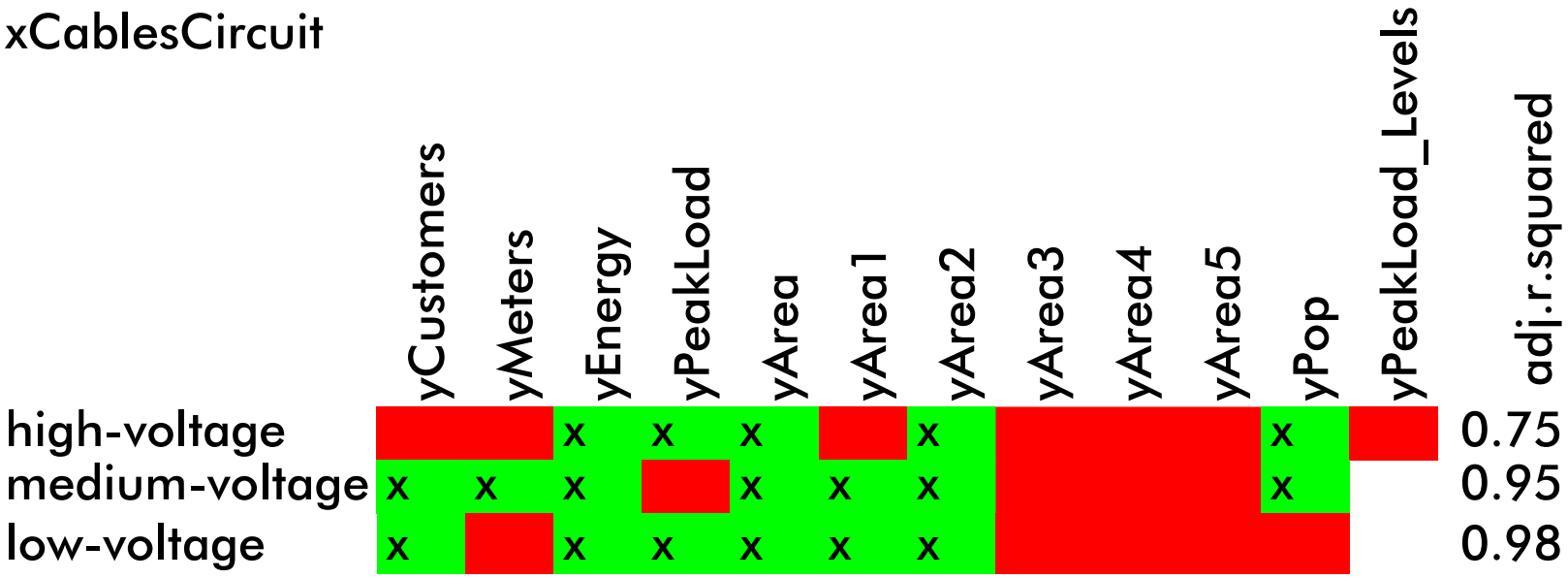
Variables Transf HS/UHS

ultra-high/high-voltage

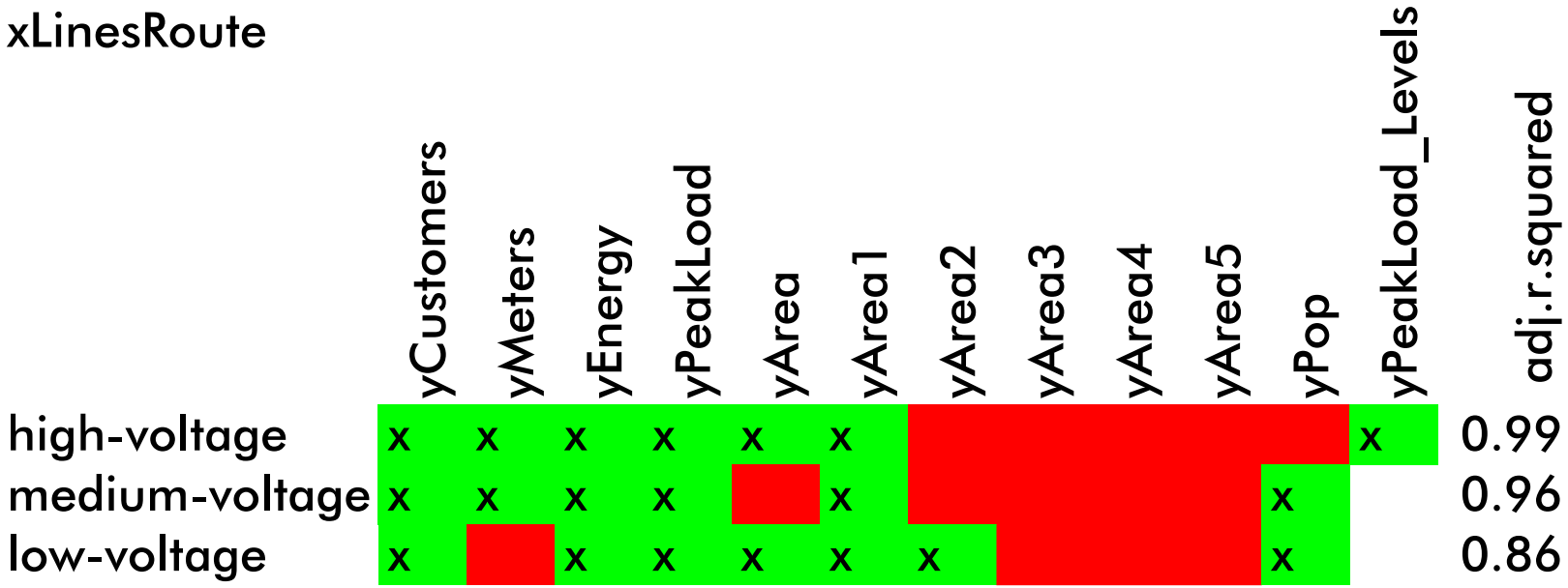


	yCustomers	yMeters	yEnergy	yPeakLoad	yArea	yArea1	yArea2	yArea3	yArea4	yArea5	yPop	adj.r.squared
xSubstation		x	x		x	x	x				x	1.00
xInstPower		x			x	x					x	1.00
xCableRoute		x	x		x	x	x				x	1.00
xLinesRoute	x	x	x	x	x	x	x	x	x	x	x	1.00
xCostDIR		x		x	x	x						0.99
xCostIN	x	x	x	x	x	x						1.00
xRevenues		x		x	x	x						1.00
xTOTEX	x	x	x	x	x	x						1.00

Voltage level - model comparison: Cables Circuits

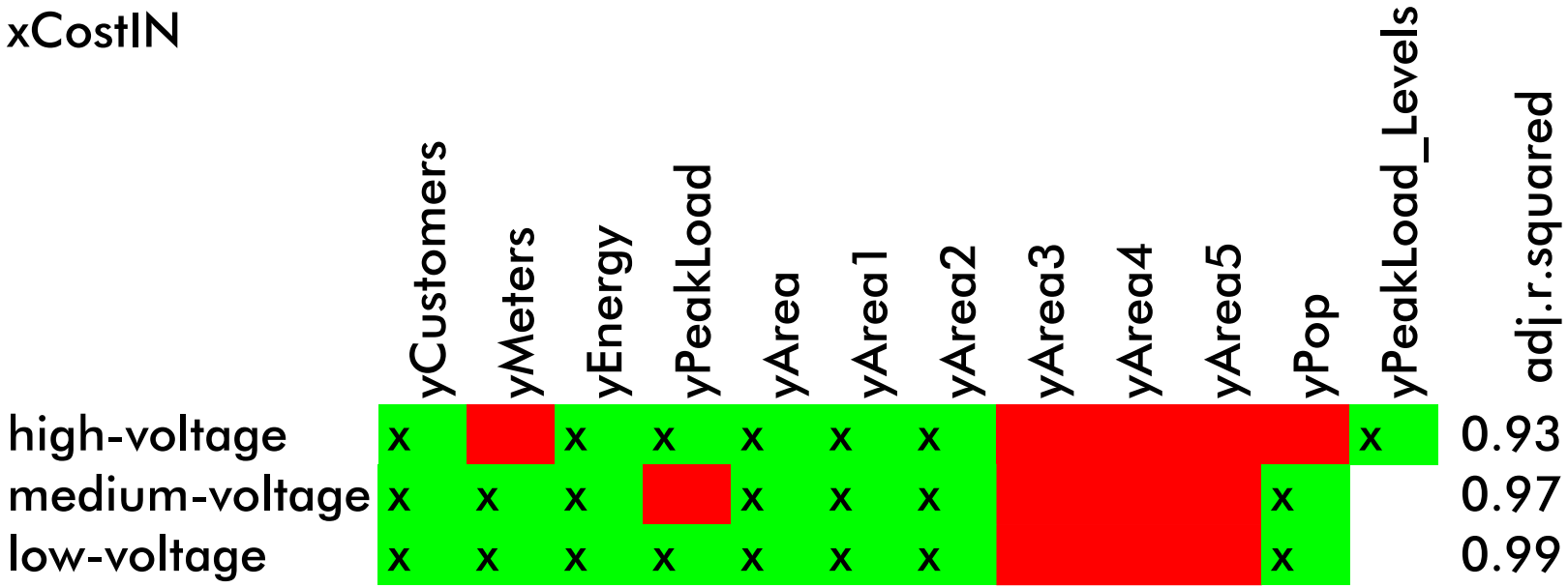


Voltage level - model comparison: Lines Route

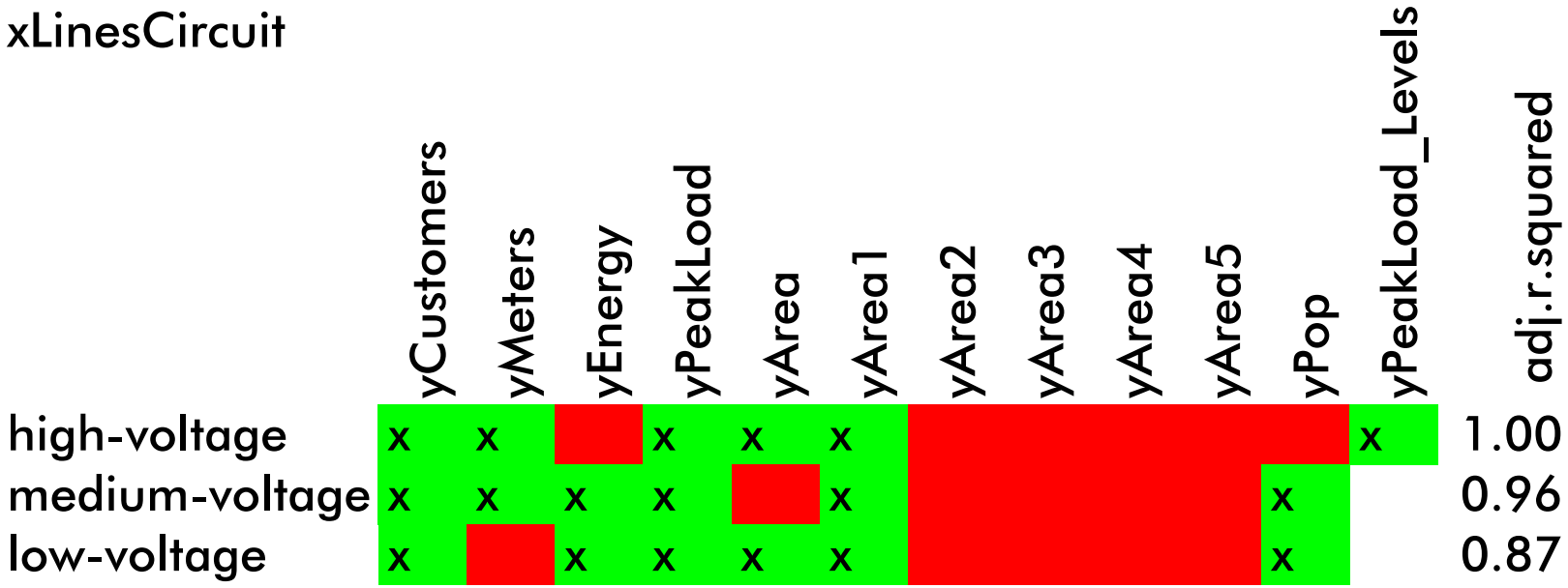


Voltage level - model comparison: Cost IN

xCostIN

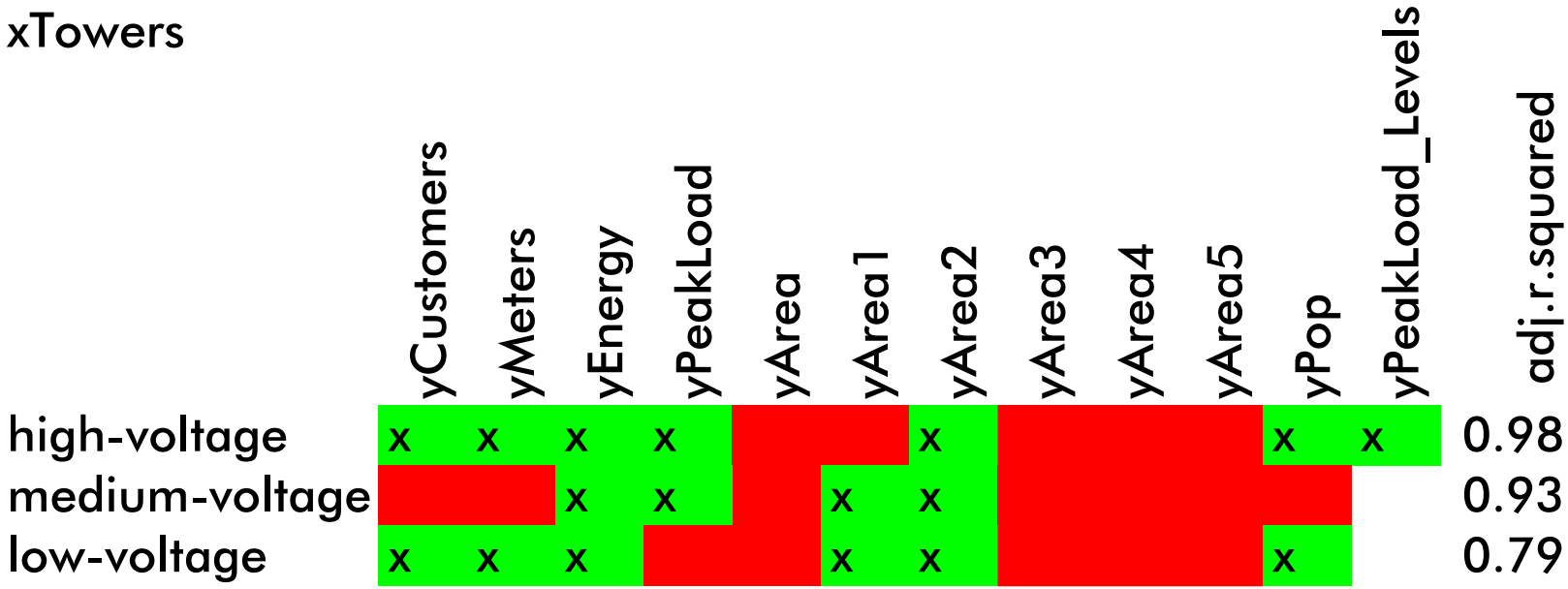


Voltage level - model comparison: Lines Circuits

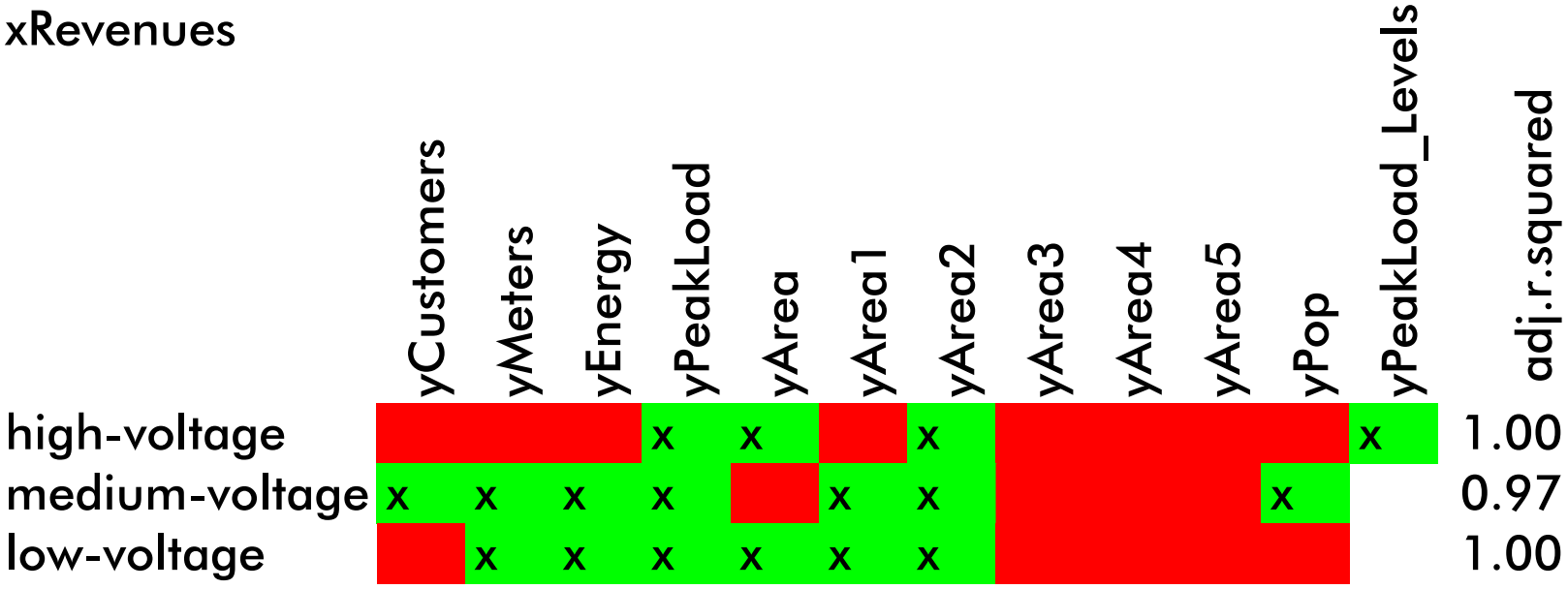


Voltage level - model comparison: Towers

xTowers

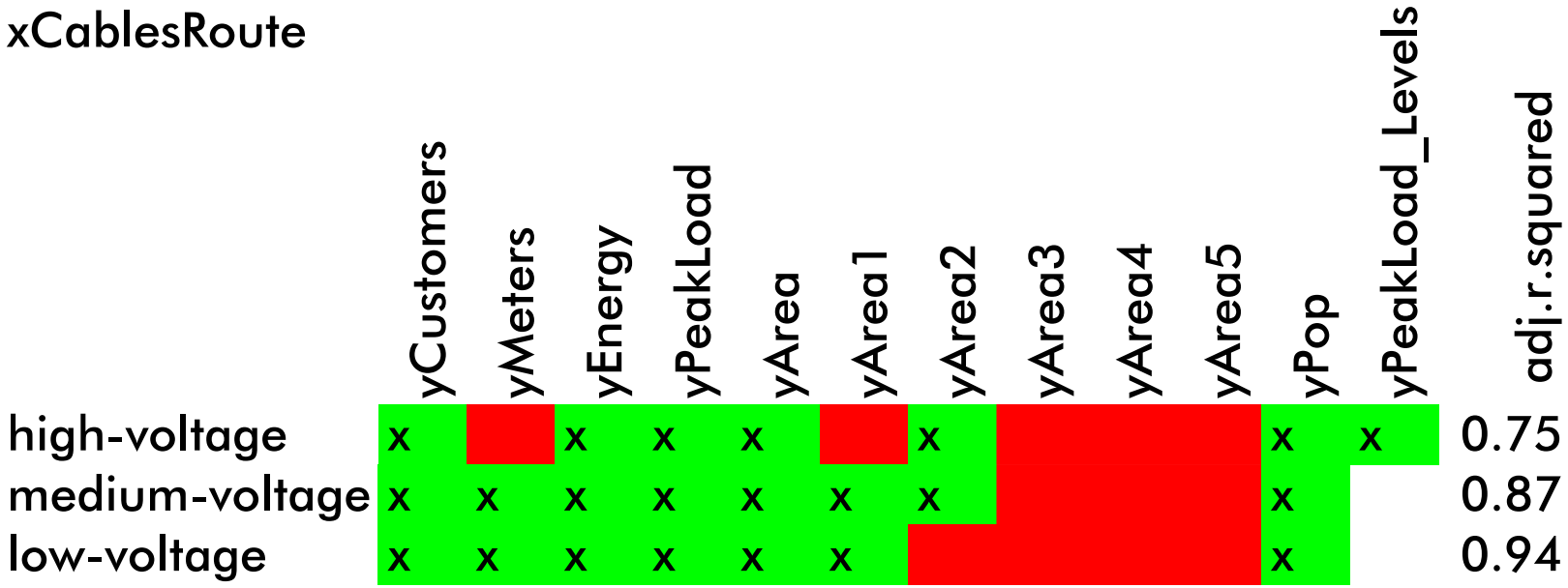


Voltage level - model comparison: Revenues



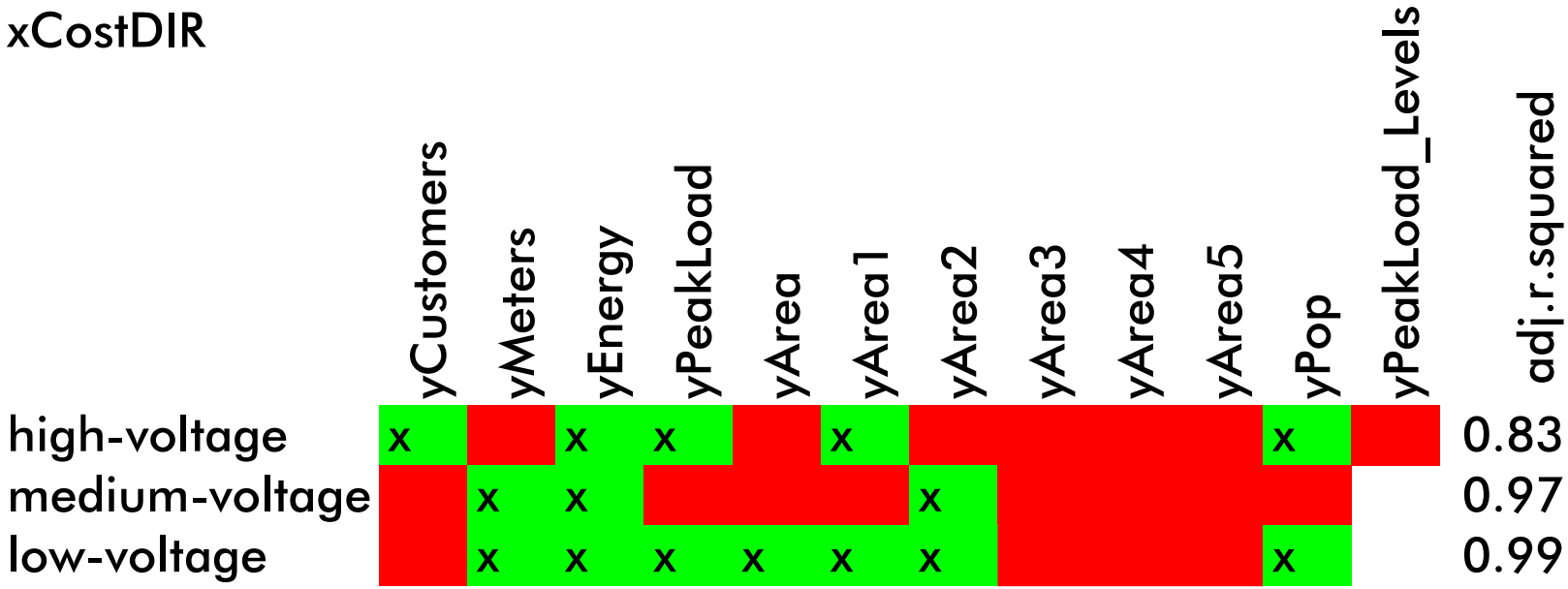
Voltage level - model comparison: Cables Route

xCablesRoute



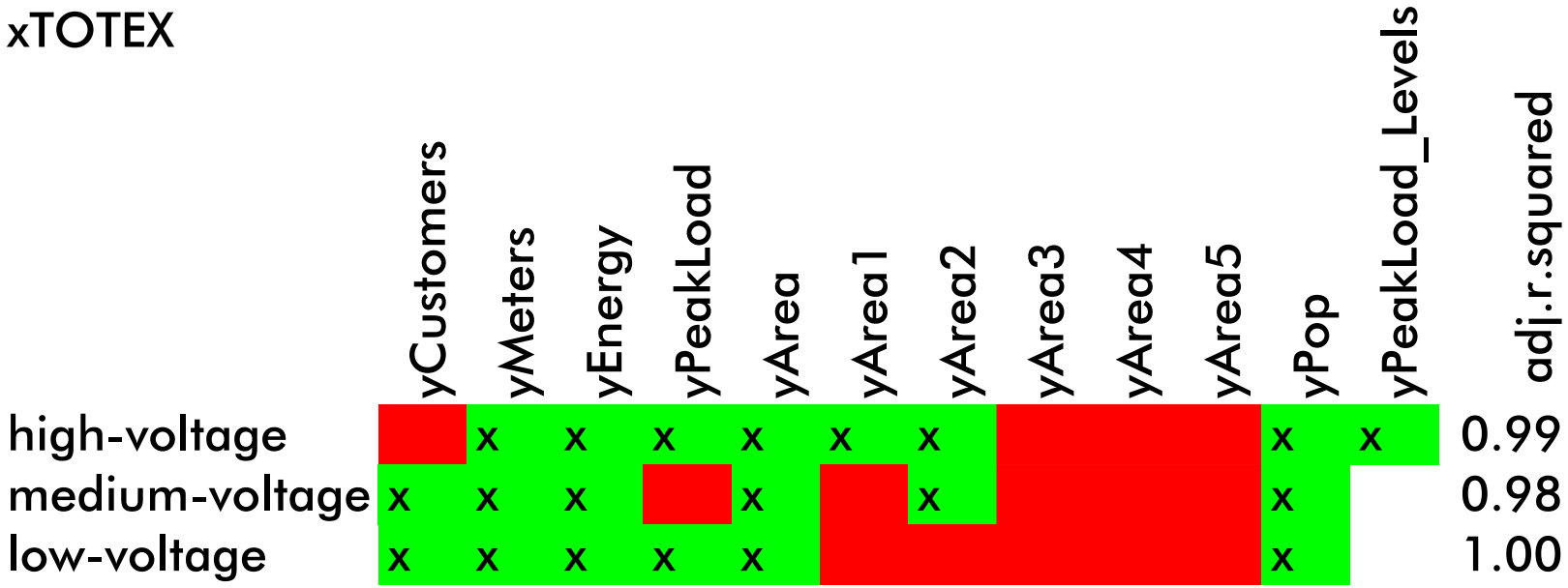
Voltage level - model comparison: CostDIR

xCostDIR



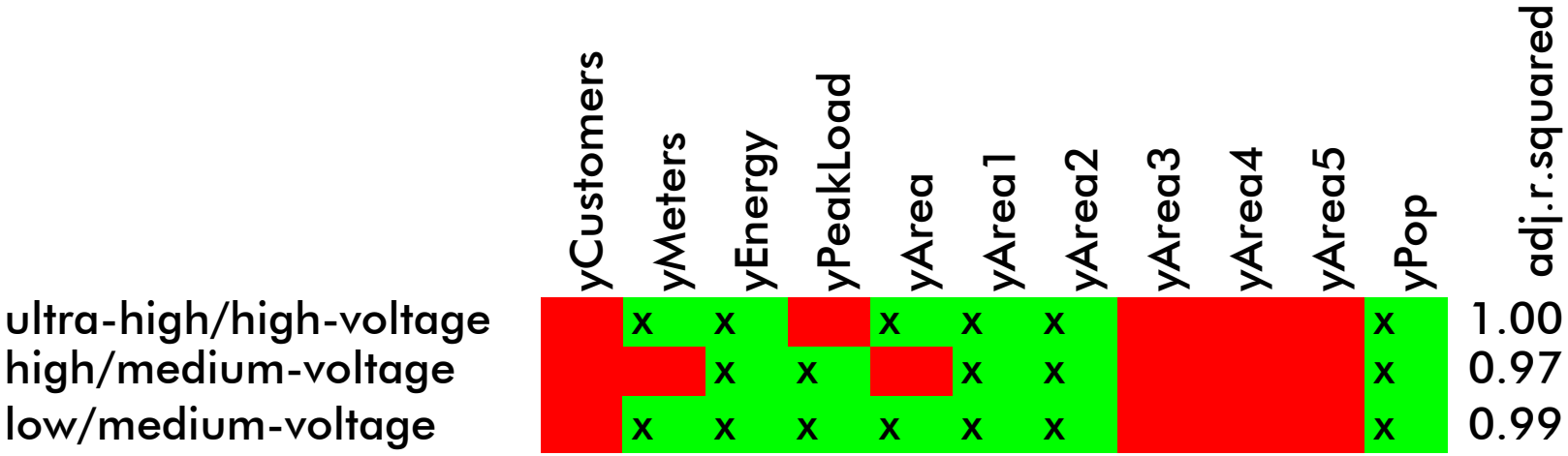
Voltage level - model comparison: TOTEX

xTOTEX



Transmission level - model comparison: Substation

xSubstation



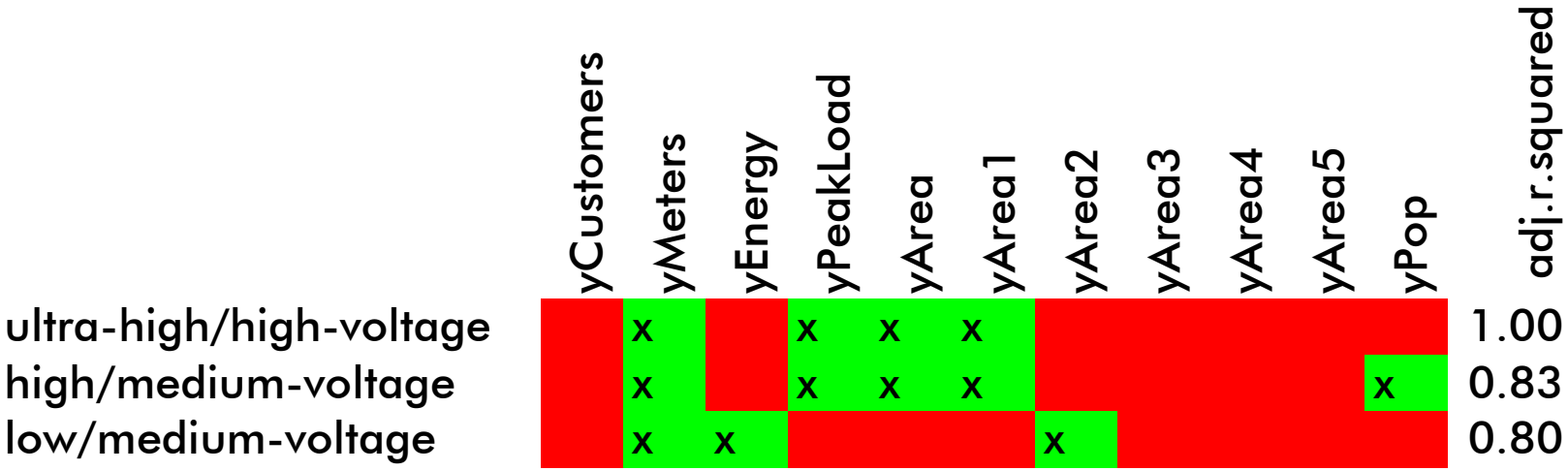
Transmission level - model comparison: Lines Route

xLinesRoute

	yCustomers	yMeters	yEnergy	yPeakLoad	yArea	yArea1	yArea2	yArea3	yArea4	yArea5	yPop	adj.r.squared
ultra-high/high-voltage	x	x	x	x	x	x	x	x	x	x	x	1.00
high/medium-voltage	x	x	x	x	x	x	x	x	x	x	x	0.22
low/medium-voltage	x	x	x	x	x	x	x	x	x	x	x	0.00

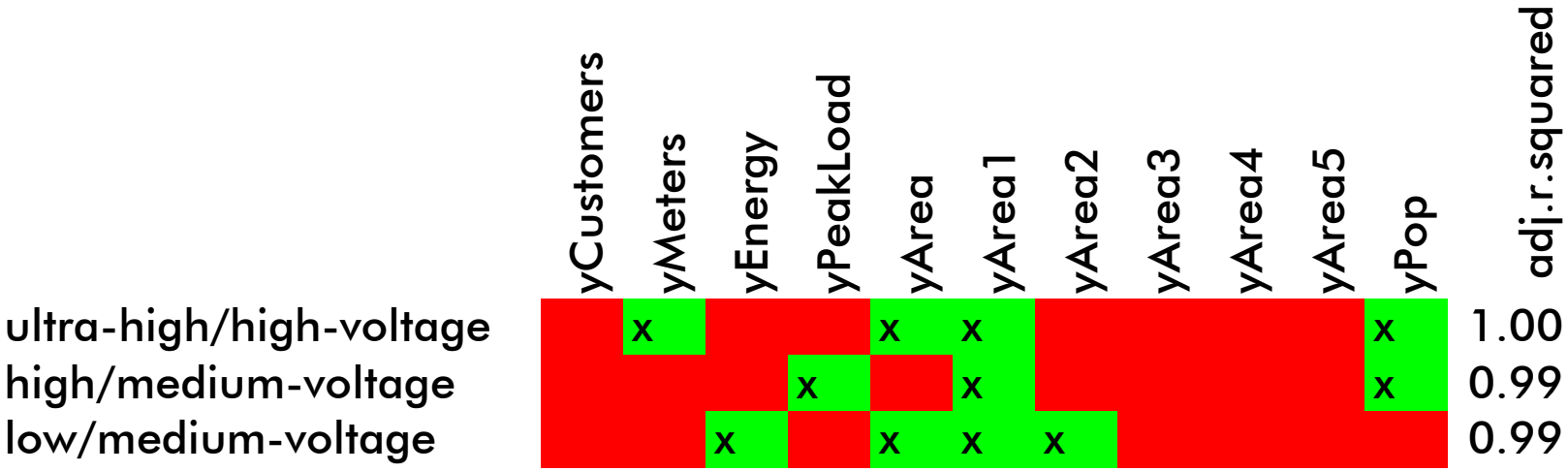
Transmission level - model comparison: Revenues

xRevenues



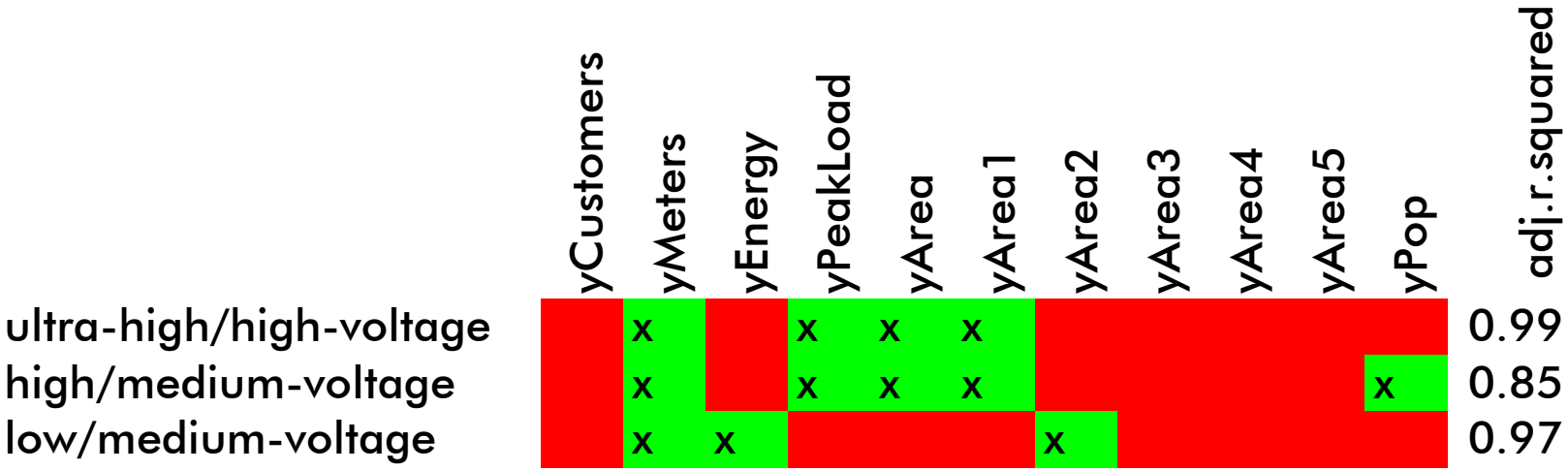
Transmission level - model comparison: Installed Power

xInstPower



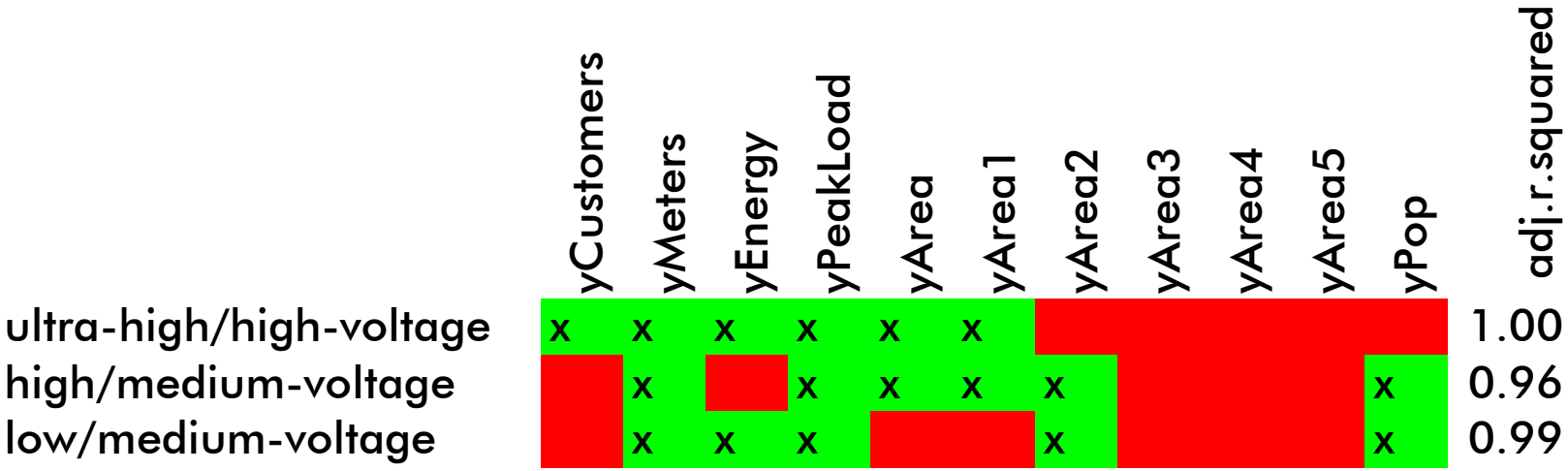
Transmission level - model comparison: CostDIR

xCostDIR



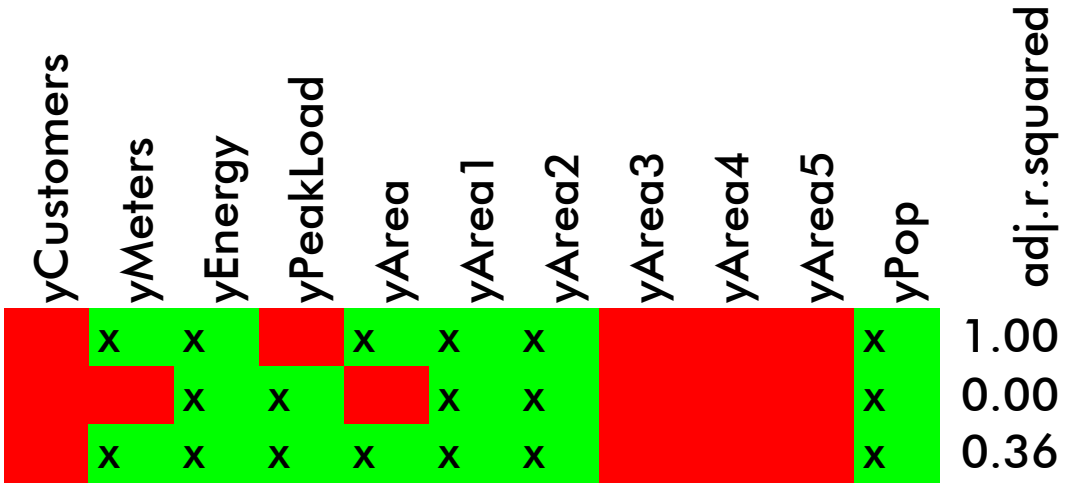
Transmission level - model comparison: TOTEX

xTOTEX



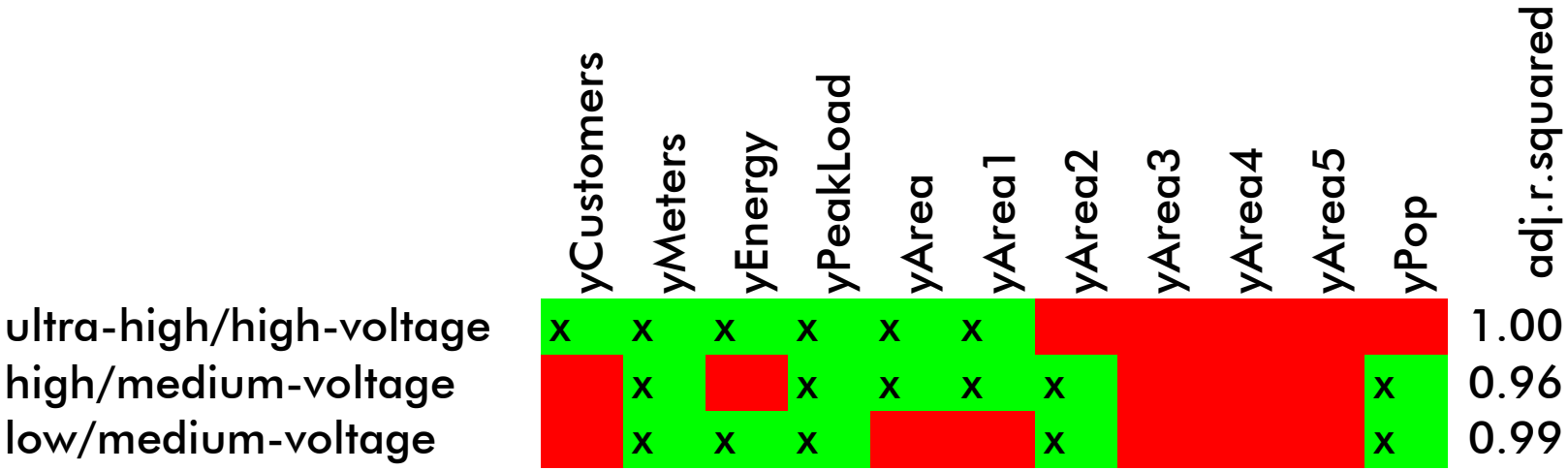
Transmission level - model comparison: Cables Route

xCableRoute



Transmission level - model comparison: CostIN

xCostIN



Typical output dimensions

Transport work

- Delivered energy (MWh)

Capacity provision

- Max peak coincidental power (MW)

Service provision

- Number of clients
- Number of metering points
- Service area

DEA models

Cost efficiency defined

$$C^{DEA}\left(y|z,w\right)=\min_{x,\lambda} wx$$

s.t. $$\sum_{i=1}^{n} x^i \lambda^i \le x$$

$$z\lambda^i \le z^i \lambda^i \qquad i=1,\ldots,n$$

$$\sum_{i=1}^{n} y^i \lambda^i \ge y,$$

$$\lambda \in \Gamma(r)$$

Structural choices

Disposability

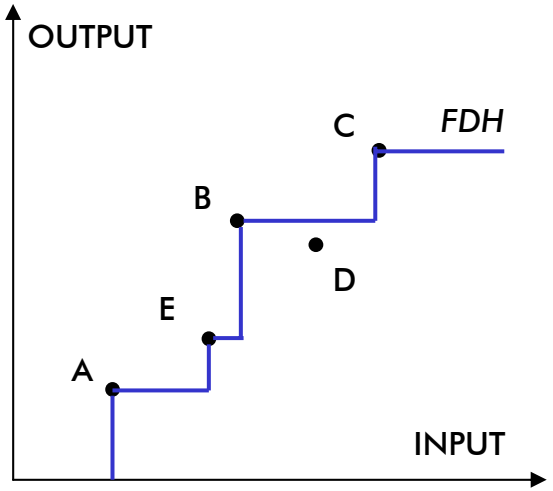
Convexity

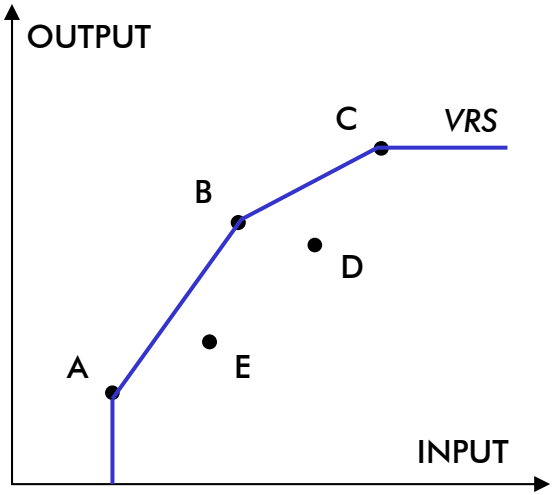
Scale

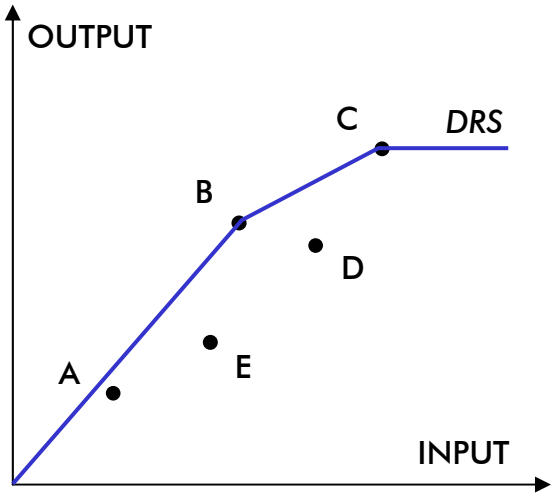
Partial pricing

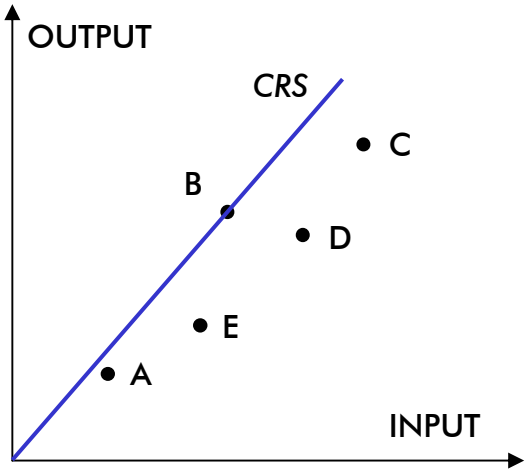
Orientation

Returns to scale

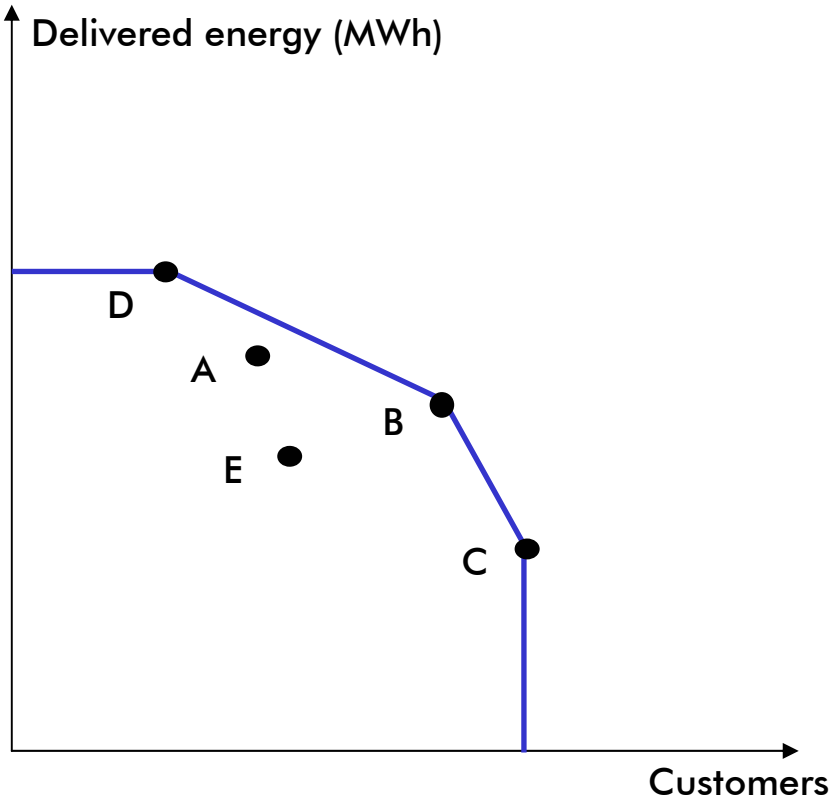




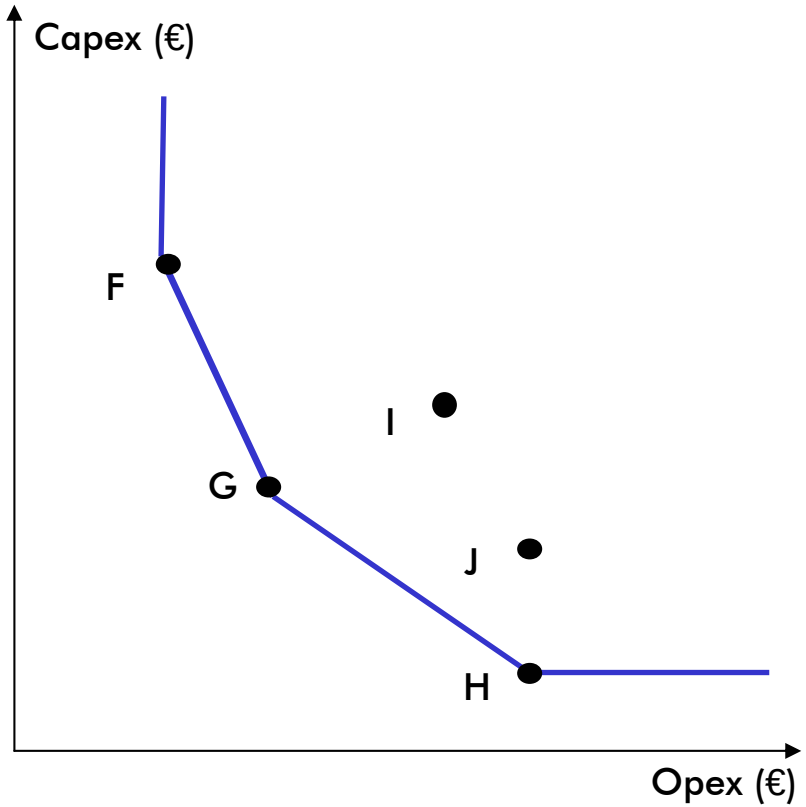




Imperfect substitution

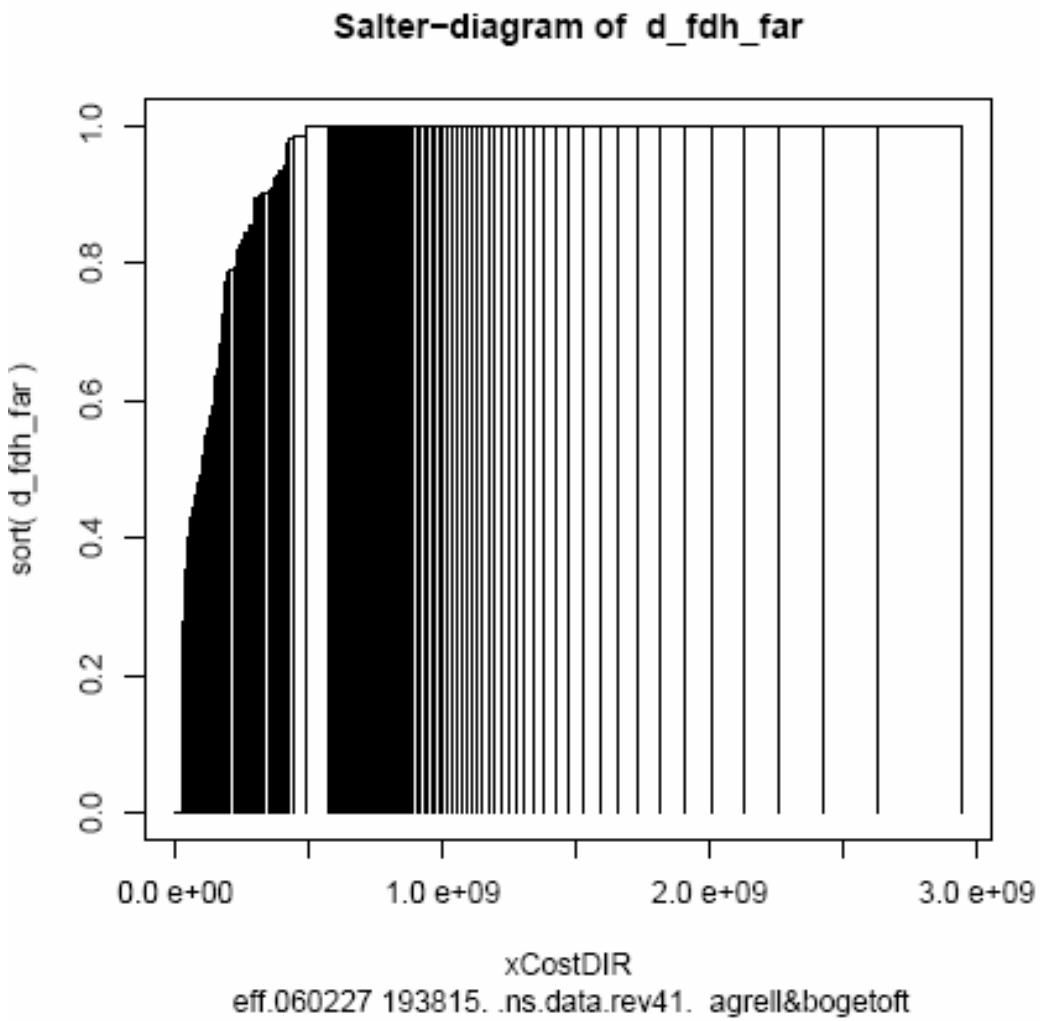


Fixed input (x)

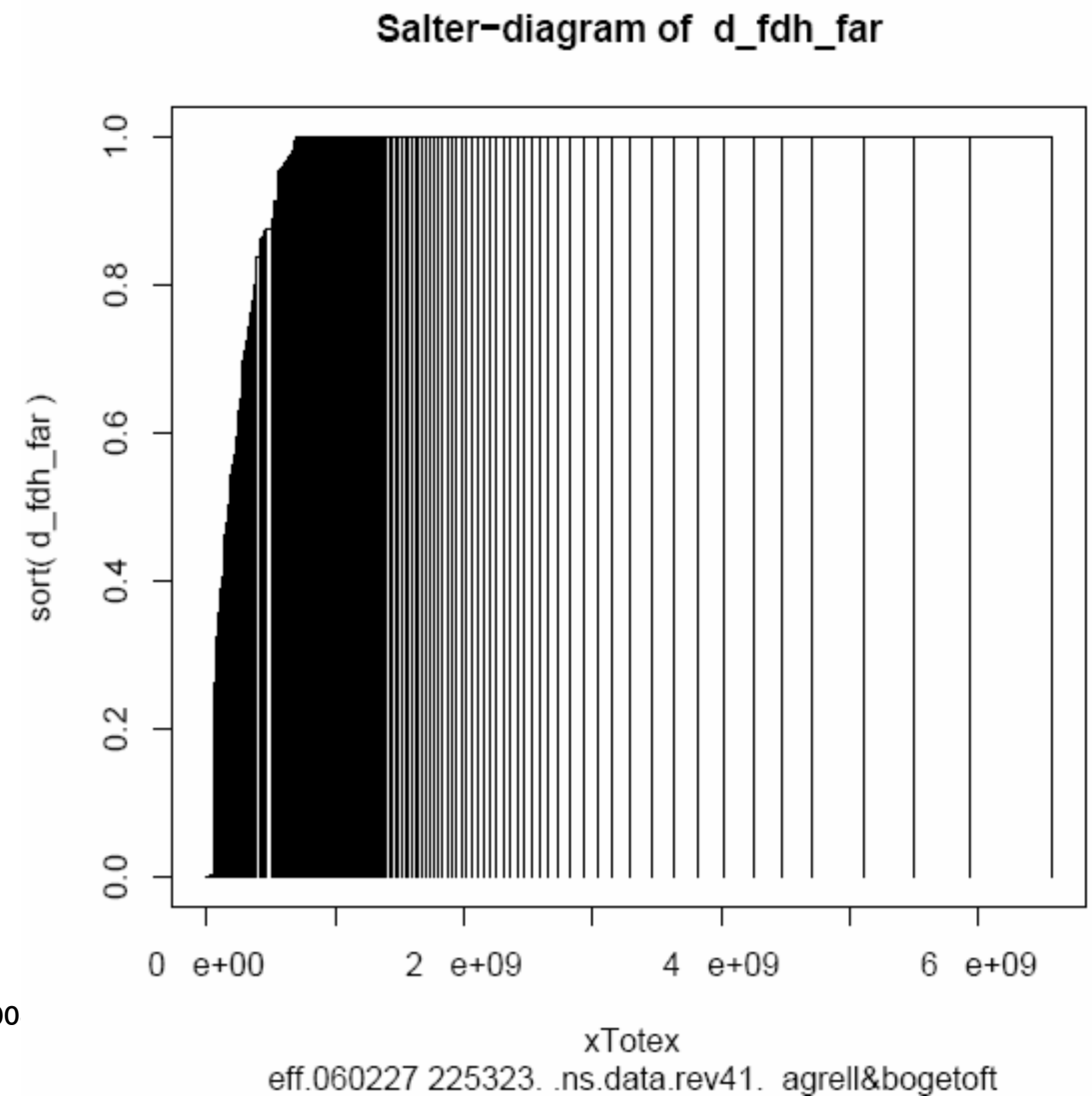


Fixed output (y)

FDH / xCostDIR (NS)

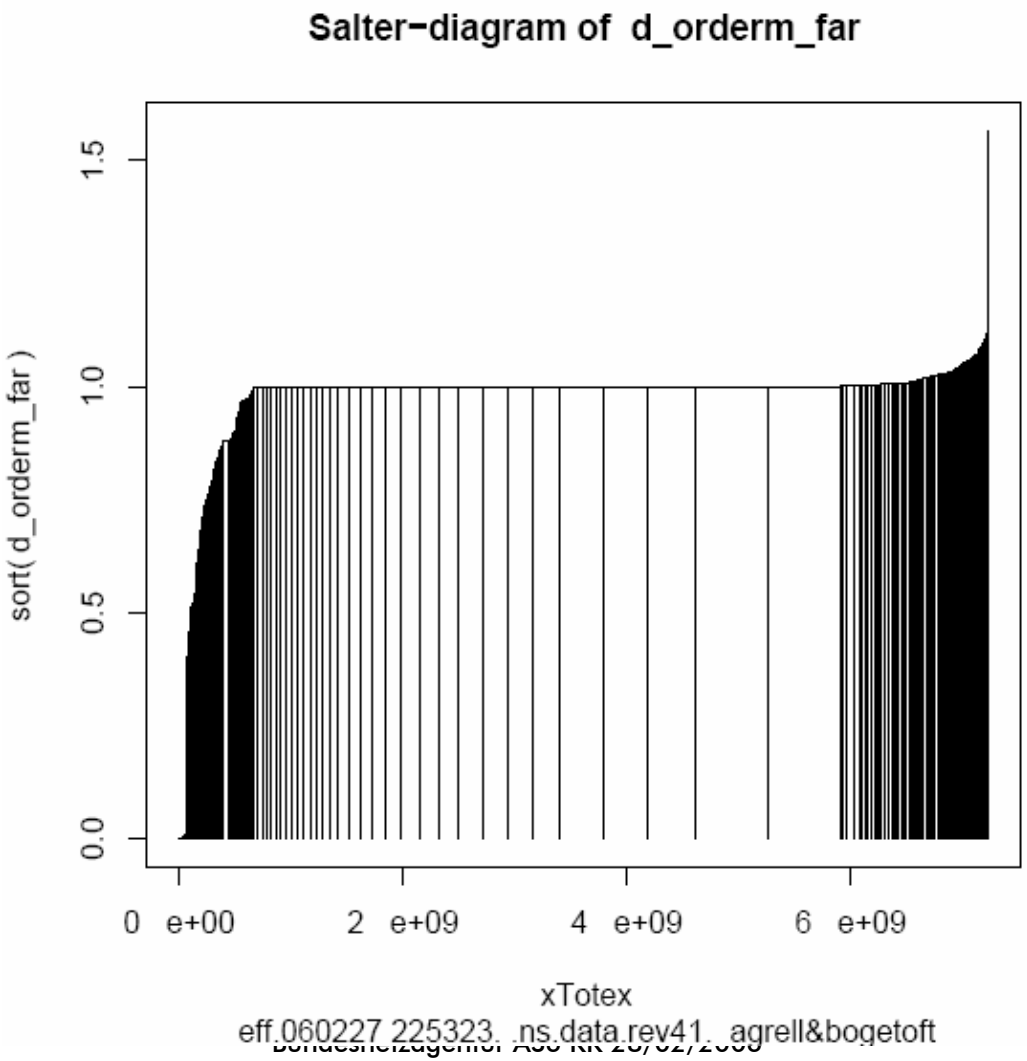


FDH / xTotex (NS)

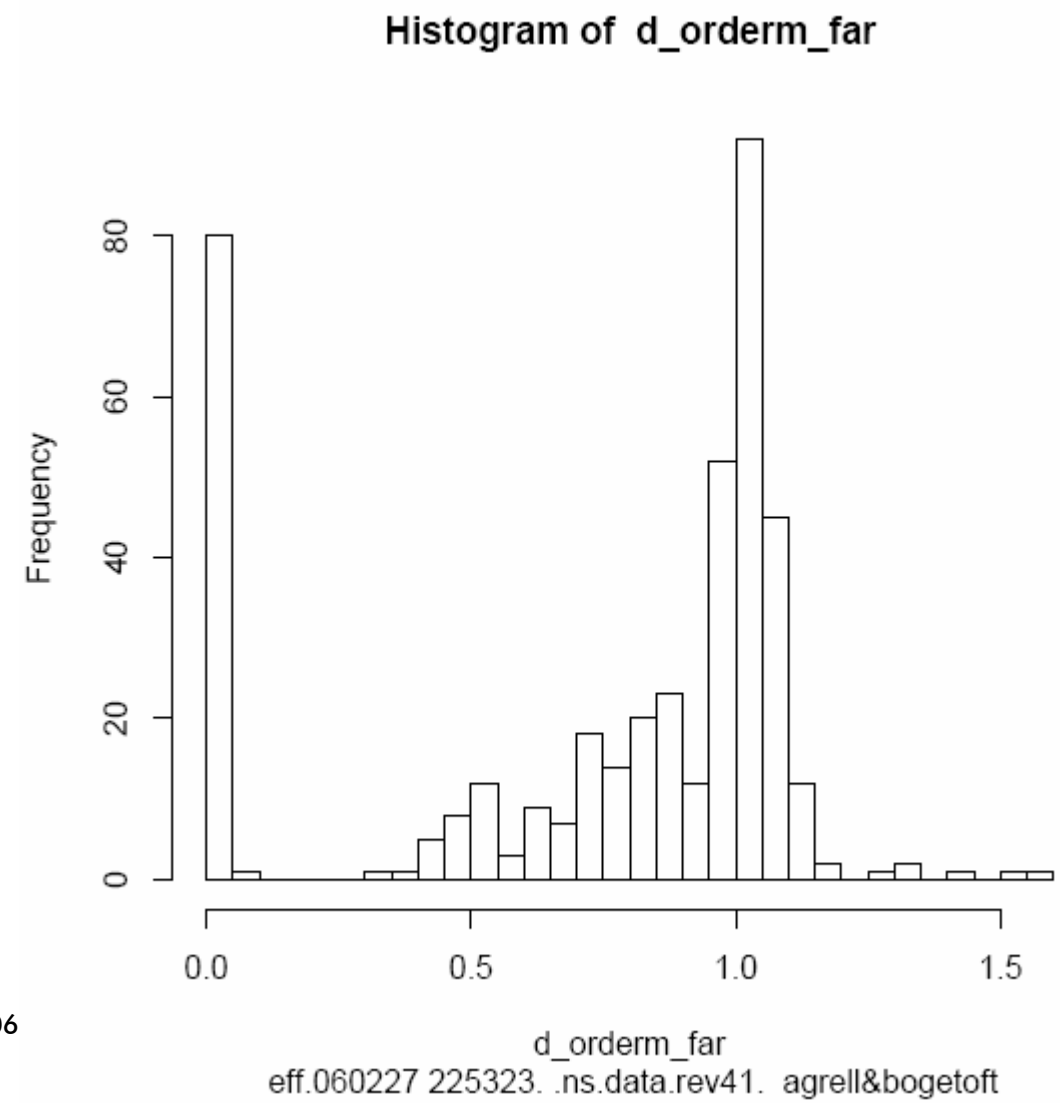


© SUMICSID, 200

Order-m / xTotex (NS)



Order-m / xTotex (NS)



© SUMICSID, 2006

SFA models

SFA technicality

Multiple inputs and outputs

- Stronger than OLS assumptions

Explain inefficiency

- Use $U_i \sim N_+(z_i\delta, \sigma^2)$
  instead of two stage
- Impose time varying structure $U_{it} = Exp[-\eta(t-T)]U_i$
  in panels

SFA technicality

STRUCTURES

$y_i = f(x_i;\beta) + V_i - U_i$

$x_i = C(y_i, w_i;\beta) + V_i + U_i$

f and C typically Cobb-Douglas or Translog

DISTRIBUTIONS

V_i □ *iid* $N(0,\sigma_V^2)$ Noise

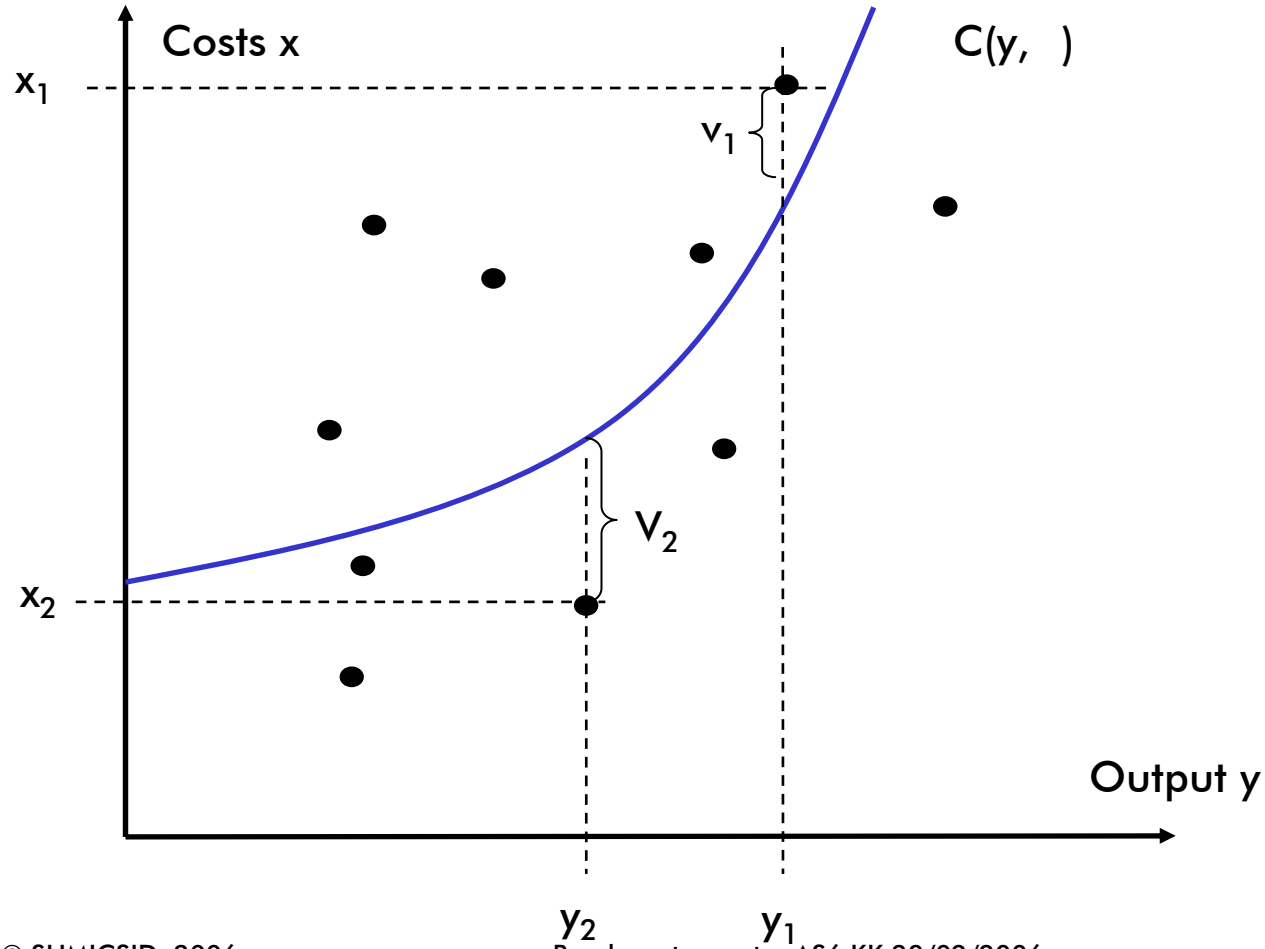
U_i □ *id* $N_+(\mu_i,\sigma^2)$ Inefficiency

ESTIMATES

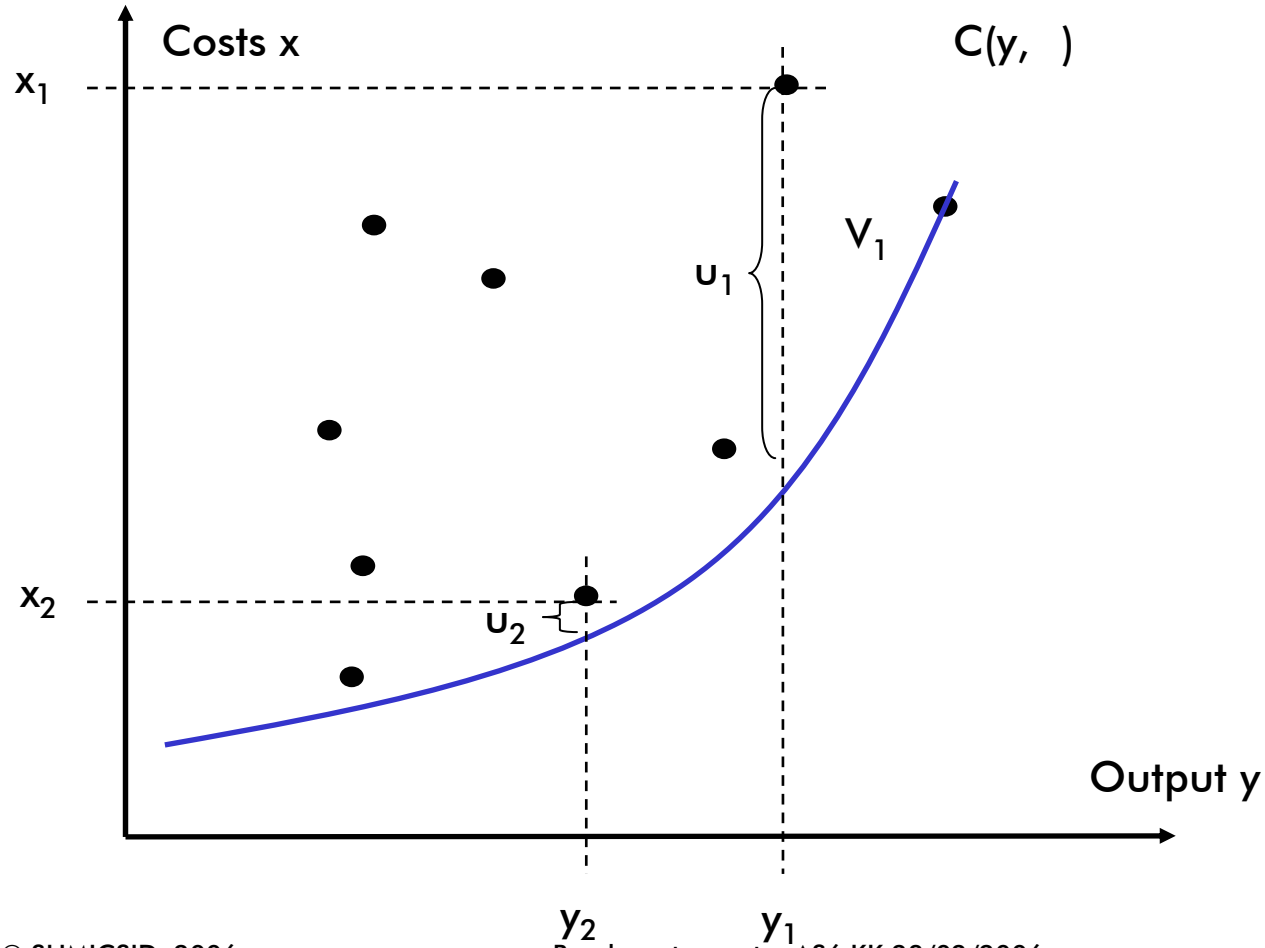
$TE_i = [f(x_i;\beta) - U_i] / f(x_i;\beta)$ with y as output

$TE_i = E(-U_i \mid v_i - u_i)$ with log y as output

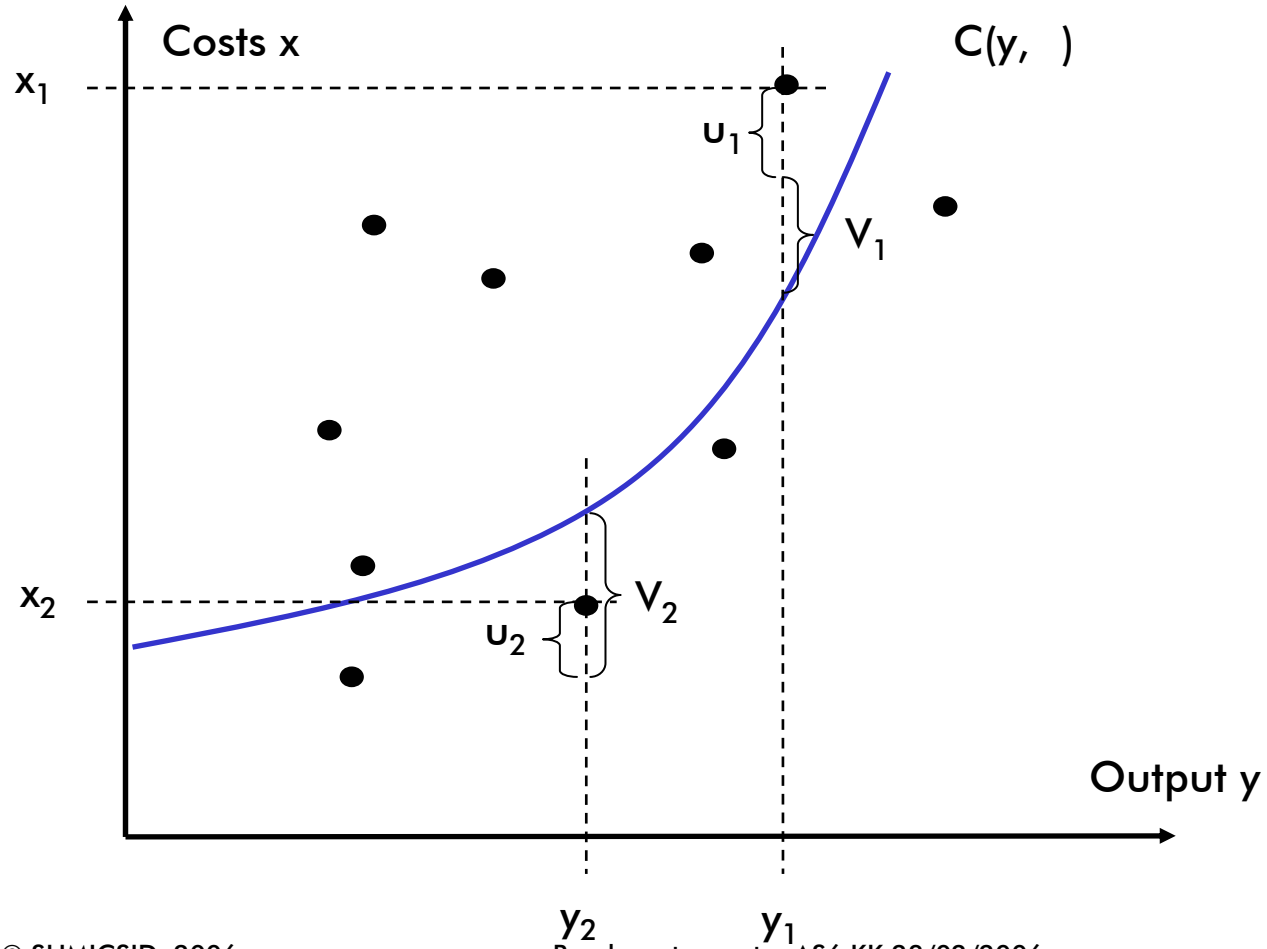
Average (OLS)



Deterministic frontier (COLS)



SFA



DEA versus SFA

DEA

- Ignores noise in data and model
- Compensates with a very flexible (non-par) strucure
- Invokes limited assumptions about distr. of inefficiencies
- Makes testing a little complicated (bootstrapping)

SFA

- Allows noise in data
- Impose restirctions (par) on structure
- Invokes distributional assumptions
- Makes testing easy (likelihood ratio testing)

Regulatory Robustness

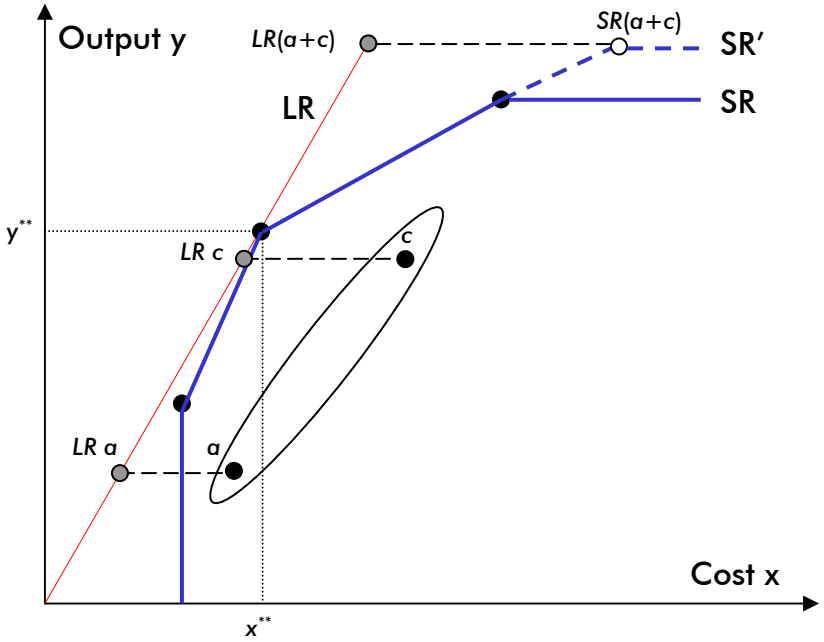
Methodology

Robustness against data error in DEA
- Wilson-Simar test
- Wilcoxon test

Robustness in estimation
- Bootstrapping
- Peeling
- Order-m scores

Robustness against bias
- Agrell-Bogetoft bias correction

Sample size and bias



Summary

Summary

Project components

- Parameter selection by significance testing
- Model structure selection
- Model estimation
- Model validation

Ongoing tasks

- Complementary data for test T1, T2 and T3
- Development of proxies for environmental complexity (AS7)
- Final data delivery for model estimation

SUMICSID
SWEDEN–BELGIUM–DENMARK