# Appendix

## A1 Adjustments of the Revenue cap - Swedish regulatory scheme

Three adjustments of the revenue cap were conducted by the Swedish regulator. All the explanations provided below have been taken from the paper of Wallnerström and al.(2016).

#### A1.0.1 Quality

The quality adjustments is calculated from the information yearly provided by every operator concerning their outages planned or unplanned faced in the past year. Because customers have different outage costs (no electricity for a company will have a greater impact than for a household for instance), every customer belongs to one of the five categories facing different outage costs determined in advance by the EI. The annual quality adjustments are computed for each operator as follows:

$$Q = \sum_{k=1}^{5} \sum_{j=1}^{2} \big((SAIDI_{b,j,k} - SAIDI_{o,j,k})Ke_{j,k} + (SAIFI_{b,j,k} - SAIFI_{o,j,k})Kp_{j,k}\big)P_{av}$$

where SAIDI is the system average interruption duration index and SAIFI the system average interruption frequency index, k represents the customer group concerned and j the nature of the interruption (planned or unplanned), b is the norm level (that is to say, the SAIDI and SAIFI levels tolerated by the EI) and o is the outcome during the period of regulation, Ke is the cost parameter given in SEK/kWh, Kp is the cost parameter given in SEK/kW and $P_{av}$ is the average yearly power usage defined as a fraction between the total energy consumption for a customer and the number of hours within a year.

### A1.0.2 Energy losses

The energy losses are considered as non-controllable costs. However, on the long run, lines can be changed. For this reason, the following incentive mechanism has been imagined.

$$L = \frac{(N_{\text{norm}} - N_{\text{turn-out}})pE_{\text{turn-out}}}{2}$$

where $N_{norm}$ and $N_{turn-out}$ are respectively the historical average share of electricity network losses in the previous regulatory period and the share of network losses for the operator during the year in percentage.The parameter p is the corresponding price per megawatt hour for network losses calculated as an average price during the regulatory period and $E_{turn-out}$ represents the amount of distributed energy during the regulatory period.

### A1.0.3 System utilization

The cost for feeding grid paid to the superior operator is comprised in the non-controllable cost and therefore the consumer has to pay for it. Nevertheless, this tariff often depends on the highest load. Since the network is designed to consider the peak loads, the capacity of the system will rarely be fully utilized and a better utilization leads to lower tariffs. To account for this, the following equation depicts the incentive measure set up the EI:

$$U = \left\{ \begin{array}{c} |\text{Lf}_{turn-out} * B_{\text{diff}} * E_{\text{turn-out}} \quad if \quad B_{\text{diff}} < 0 \\ 0 \quad \text{if} \quad \text{B}_{\text{diff}} \leq 0 \end{array} \right\}$$

where s the sum of all daily load factors $Lf_i$ divided by the number of days during the regulatory period. $B_{\text{diff}} = B_{\text{norm}} - B_{\text{turn-out}}$ is the saving for the cost that DSOs pay to the feeding grid (kSEK/MWh). This can be computed as the difference between the cost paid to the superior grid during the previous period and during the current regulatory period divided by the amount of distributed energy during the period (SEK/MWh). $E_turnout$ is the distributed energy during the regulatory period (MWh).

The term $Lf_{\text{turn-out}}$ can be found as follows:

$$Lf_{\text{turn-out}} = \frac{\sum_{i=1}^{D} Lf_i}{D}$$

where $Lf_i$ is the average load divided by the maximum load in day i and D is the amount of day for the regulatory period.

### A1.1 Consequences on the revenue cap

Based on the incentives presented previously, the total adjustment of the revenue cap (RC) can be computed as shown below:

$$\text{Total adjustments} = \left\{ \begin{array}{c} \text{-0.05[RC]} \quad \text{if} \quad \text{(Q+L+U)} \leq -0.05\text{[RC]} \\ +0.05\text{[RC]} \quad \text{if} \quad \text{(Q+L+U)} \geq +0.05\text{[RC]} \\ \text{Q+L+U} \quad \text{otherwise} \end{array} \right\}$$

## A2 Outliers

We can distinguish two different definitions according to Agrell and Niknazar (2014):

"According to a typical intrinsic definition, cf. Barnett and Lewis (1994), an outlier is an observation which appears to be inconsistent with the rest of the data set [...] According to an influence perspective, a unit acquires the quality of outlier for a given method and reference set through the (undue) impact that its inclusion exerts on the quality of the estimation."

In view of those two different definitions, two different methods has been used.

### A2.1 Outliers via data cloud method

For outliers such as presented in the first definition, we will use what is commonly called the data cloud method (Bogetoft and Otto, 2011). This method first consists of computing the determinant of the matrix representing the dataset. Since this determinant is directly proportional to the volume of the data cloud, removing outliers would greatly reduce the data cloud and therefore the determinant. Therefore, a determinant being far lower than the initial determinant after the removal of an observation underlines that it is part of the outliers.

In light of what has been said, let us consider the following ratio:

$$R^i = \frac{D^i}{D} \tag{.1}$$

Where $D^i$ is the determinant without the observation(s) i and D is the initial determinant.

The method proposed relies on an iterative process that aims to deduce which observation are minimizing the ratio $R^i$ when deleted. Since the process requires an exponential number of combination as the number of observations deleted increases, the reasonable number of three observations removal is chosen as the best trade-off between the time consumed and the accuracy of the answer.

Results for the dataset 2002-2006 can be observed in Table A2.1.

| Additional firm deleted | Firms deleted (DSO number) | $R^i$ |
|---|---|---|
| 2002-2006 | | |
| Vattenfall Eldistribution AB | 583 | 0.0822 |
| E.ON Elnaet Sverige AB | 583,593 | 0.0094 |
| Fortum Distribution | 583,593,176 | 0.0013 |
| 2011-2015 | | |
| Jukkasjärvi Sockens | 83 | 0.0081 |
| E.ON Energidistribution | 93, 957 | 0.0009 |
| Ellevio AB | 93,957,3008 | 0.0000 |
| 2014-2017 | | |
| Jukkasjärvi Sockens | 93 | 0.0015 |
| E.ON Elnät Sverige | 93,957 | 0.0002 |
| Vattenfall Eldistribution | 93,957,572 | 0.0000 |

**Table A2.1:** Outliers according to the data cloud method

## A2.2 Outliers via super-efficiency

Another way of determining the outliers proposed by Banker and Chang (2006) is to make use of the super-efficiency. The main idea is to compute the super-efficiency of each firm. If the super-efficiency is higher than a predetermined threshold, the observation will be seen as greatly influencing the best-practice frontier. The method suggested to compute the threshold according to the data is to use the following formula:

$$T = q(0.25) + 2 * [q(0.75) - q(0.25)] \tag{.2}$$

Based on this method, outliers for our different dataset are displayed in Table A2.2

| 2002-2006 | Super-efficiency |
|---|---|
| Larvs Elektriska | 1.48 |
| LJW Naet HB | 2.50 |
| NVSH Energi AB | 1.53 |
| 2011-2015 | |
| Carlfors Bruk | 99.66 |
| Sturefors Eldistribution | 10.10 |
| Vaggeryds kommun | 1.56 |
| 2014-2017 | |
| Carlfors Bruk El. Björklund | 32.40 |
| Jukkarsjärvi Belysnings. | 3.32 |
| Vaggeryds kommun Elverket | 1.72 |

**Table A2.2:** Outliers according to the super-efficiency method

### A2.3 Discussion

In view of the table presented, it can first be concluded that the two methods lead to sensible different results when classifying observations as outliers, confirming our intuition that several kinds of outliers exist. Another conclusion regarding the data cloud methods is that it tends to systematically detect biggest operators as outliers. Based on this, only outliers detected thanks to the super-efficiency method are removed.

## A3 Variables significance

### A3.1 Inputs and outputs

It is important to verify that our models fits correctly to the data. On the one hand, a simplified model would offer less accurate performance. On the other hand, it would increase the discriminatory power of the model and allows a clearer ranking of the firms. In order to determine the relevance of a variable, we can compare the efficiency scores which would be obtained with and without the variable. The intuition is that if efficiency scores deduced from the two models are divergent above a certain threshold, the initial model offers a better representation of the performance of the firms.

The Kolmogorov-Smirnov (KS) test proposed by Banker (1993) can be seen as a convenient test with free assumption on the distribution in order to test the significance of the variables. The test statistic can be defined as:

$$T_{KS} = max\{|G_1(E^k) - G_2(E^k)|\} \tag{.3}$$

Where $G_1$ and $G_2$ are the empirical cumulative distributions of the model with and without the variable such as $T_{KS}$ is the largest vertical distance between the cumulative distribution. A large value for $T_{KS}$ indicates that the distribution differ and therefore, that the variable is significant.

The results of the test on our set below.

| Variable | Statistical value | P-value |
|---|---|---|
| 2002-2006 | | |
| CAPEX | 0.567 | 0.0000 |
| OPEX | 0.998 | 0.0000 |
| Connections | 0.062 | 0.0427 |
| Power sub-transformers | 0.017 | 0.7885 |
| Peak load | 0.0121 | 0.8858 |
| Energy delivered LV | 0.383 | 0.0000 |
| Energy delivered HV | 0.094 | 0.0006 |
| 2011-2015 | | |
| CAPEX | 0.583 | 0.0000 |
| Controllable OPEX | 0.280 | 0.0000 |
| Connections | 0.096 | 0.0038 |
| Network stations | 0.371 | 0.0000 |
| Peak load | 0.005 | 0.9852 |
| Energy delivered LV | 0.238 | 0.0000 |
| Energy delivered HV | 0.149 | 0.0000 |
| 2014-2017 | | |
| CAPEX | 0.673 | 0.0000 |
| Controllable OPEX | 0.280 | 0.0000 |
| Connections | 0.096 | 0.0038 |
| Network stations | 0.371 | 0.0000 |
| Peak load | 0.005 | 0.9852 |
| Energy delivered LV | 0.238 | 0.0000 |
| Energy delivered HV | 0.149 | 0.0000 |

**Table A3.1:** Kolmogorov-Smirnov test

The application of the Kolmogorov-Smirnov test shows interesting results. A first conclusion is that the number of power sub transformers does not seem to be a relevant variable for Model 1 and can be removed. Being highly insignificant, the variable referring to the peak load is also removed from Model 2.

## A3.2 Non-discretionary variables

As you may have noticed, the non-discretionary variables of Model 1 have not been tested through the Kolmogorov-Smirnov test. Instead, Tobit regressions such as explained in Section 6.3.1 seem to be more adequate. To conduct these regressions, efficiency scores without the non-discretionary variables have been computed for each year. Then, the Tobit regression has been performed to see if there is any link between them and the non-discretionary variables.

| Year | **Cable proportion** | | **LV proportion** | | **Network size** | |
|---|---|---|---|---|---|---|
| | Intercept | Coefficient | Intercept | Coefficient | Intercept | Coefficient |
| 2002 | 0.459 (0.0000) | 0.260 (0.0000) | 0.485 (0.0000) | 0.222 (0.1830) | 0.636 (0.0000) | -0.0000 (0.679) |
| 2003 | 0.453 (0.0000) | 0.301 (0.0000) | 0.316 (0.0030) | 0.515 (0.0012) | 0.661 (0.0000) | -0.0000 (0.463) |
| 2004 | 0.463 (0.0000) | 0.251 (0.0002) | 0.471 (0.0000) | 0.248 (0.1340) | 0.636 (0.0000) | -0.0000 (0.937) |
| 2005 | 0.479 (0.0000) | 0.197 (0.0057) | 0.531 (0.0000) | 0.131 (0.445) | 0.618 (0.0000) | -0.0000 (0.934) |
| 2006 | 0.382 (0.0000) | 0.347 (0.0000) | 0.282 (0.0884) | 0.521 (0.0012) | 0.630 (0.0000) | -0.0000 (0.704) |

**Table A3.2:** Tobit regression for 2002-2006

Obviously, a relation between the cable proportion and the efficiency score is detected. This legitimates its presence in the model.

A boxplot chart depicting the efficiency scores obtained with and without the non-discretionary variable is depicted below.

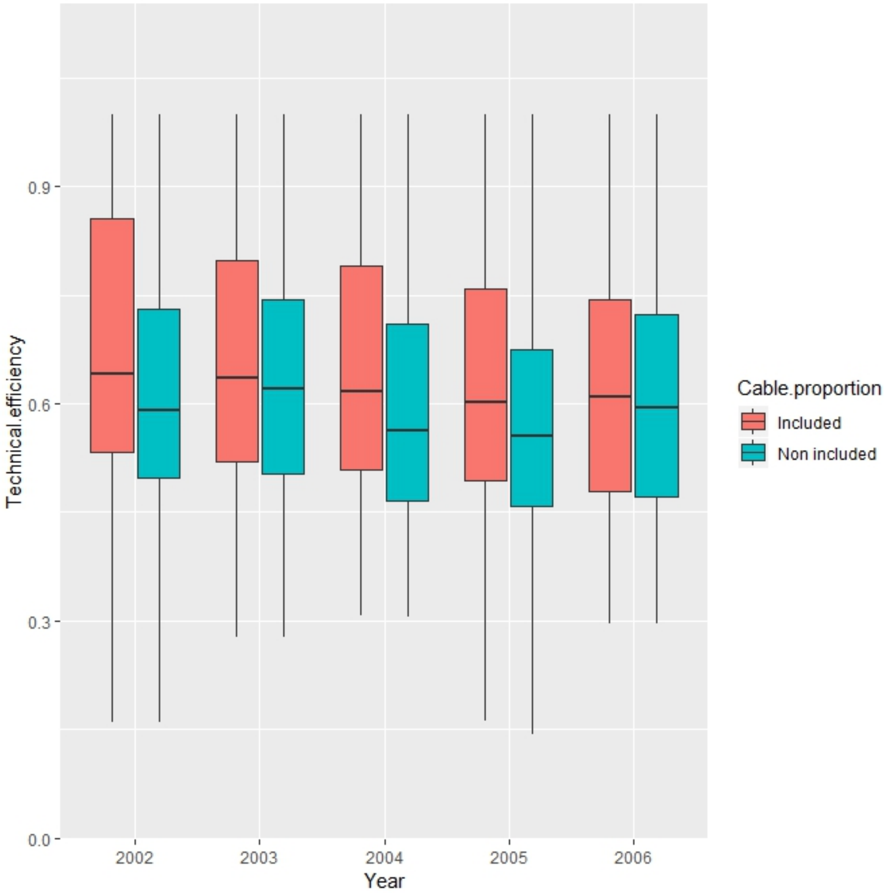


**Figure A3.1:** Efficiency scores with and without non-discretionary variables 2002-2006

# A4 Effect on average productivity change when removing biggest operators

| Year | MPI |
|---|---|
| 2002-2006 | -0.0090 |
| 2002-2003 | -0.0032 |
| 2003-2004 | -0.0028 |
| 2004-2005 | -0.0092 |
| 2005-2006 | 0.0038 |
| 2011-2015 | 0.0001 |
| 2011-2012 | 0.0007 |
| 2012-2013 | -0.0002 |
| 2013-2014 | 0.0000 |
| 2014-2015 | -0.0005 |
| 2014-2017 | -0.0003 |
| 2014-2015 | -0.0011 |
| 2015-2016 | 0.0016 |
| 2016-2017 | -0.0005 |

**Table A4.1:** Difference in the Malmquist with and without the biggest operators 2011-2017

# A5 Histograms of the Malmquist index and its components 2002-2006

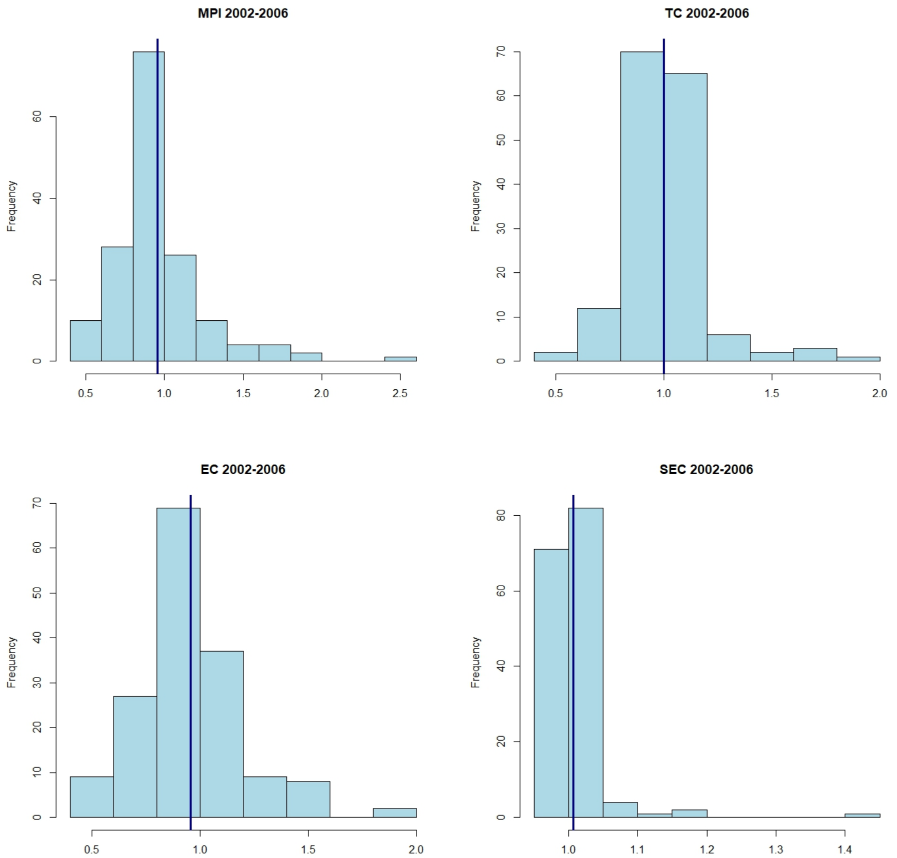


**Figure A5.1:** Histograms of the Malmquist index and components 2002-2006