Scale effects in Ofgem's DPCR5 cost benchmarking analysis

Andrew Chesher[1]

Draft: August 16th 2011

1 Introduction

In DPCR5 Ofgem estimated equations to predict total, indirect and direct costs of network operation and used the predictions for 2008 to assess the efficiency of 14 UK Distribution Network Operators (DNOs).

Ofgem's DCPR5 cost benchmarking analysis employed equations for components of costs which have an "economies of scale" property. Cost equations with this property predict that average costs per cost driver unit fall as cost drivers increase. Ofgem's results suggest that a proportionate increase, say of 10%, in a cost driver is predicted to result in a smaller proportionate increase in costs with predicted amounts ranging from 5% to 9%. This is the case for total costs, for indirect costs in total and disaggregated into three categories and for all the major components of direct costs of network operation studied by Ofgem.

UK Power Networks EPN DNO fared badly in the DPCR5 cost benchmarking exercise. EPN has the largest or close to the largest values of most of the cost drivers employed by Ofgem and may be seriously disadvantaged if the "economies of scale" embodied in Ofgem's cost functions do not apply at the largest scales of operation observed. In many cases EPN's costs in 2008 were higher than those predicted by the Ofgem equations.

This document reports on research that examines this question. The focus in this report is on the scale effects. The suitability of the choice of cost drivers has not been addressed

Ofgem plans to use cost benchmarking in the next price control review, RIIO-ED1, and early indications are that the approach planned by Ofgem

[1] Andrew Chesher is Professor of Economics at University College London, Director of the ESRC Centre for Microdata Methods and Practice and Research Fellow of the Institute of Fiscal Studies. The views expressed here are his alone expressed in a private professional capacity. During the period March - December 2009 he advised EDF Energy on aspects of the cost benchmarking process undertaken for DPCR5 by Ofgem. Subsequently he provided a review of the DPCR5 process [7].

may not vary greatly from that employed in DPCR5. The results given here will inform UK Power Networks during the run-up to RIIO-ED1 when influence can be brought to bear on Ofgem concerning the conduct of cost benchmarking.

The report starts with a brief account of Ofgem's cost analysis. The data employed in DPCR5, augmented with 2009 data giving 5 years of data on 14 DNOs, is briefly described and EPN's position relative to the other DNOs is set out.

The international literature on scale effects in electricity distribution is reviewed and implications for the type of cost benchmarking conducted in DPCR5 by Ofgem are considered.

The cost equations employed by Ofgem are re-estimated using the new 5 year data set and the scale effects are summarised. Analysis is conducted to determine the robustness of the scale effects with attention to the impact on EPN's position of moving to alternative specifications.

Large scale utilities with large service areas experiencing diseconomies of scale may improve efficiency by running separate regional operations. In earlier work we analysed a 5 year data set in which EPN's operations were split between two new operators, EPN North and EPN South. This work is reviewed here.

2 Summary and recommendations

1. *Economies of scale.* The empirical economics literature on economies of scale in electricity distribution suggests that in operations at the scale of the DNOs that operate in the UK there are likely at most small economies of scale when one considers increasing service region area with customers per unit area and electricity delivered per customer held constant. There is probably more scope for economies of scale as customer numbers increase with service region area and electricity supplied per customer held fixed. There may be significant economies of scale as electricity supplied per customer is increased with customers per unit area and service region area held fixed.

2. *Ofgem cost equations.* The cost equations employed by Ofgem in DPCR5 have no provenance in economics. They relate costs at different levels of aggregation variously to measures of the capital stock embodied in the DNOs and aspects of the operations of the DNOs.

As such they are purely empirical relationships without a theory basis and should be assessed on the basis of their goodness of fit to the data. I find that the fit is inadequate in some cases and that this depresses efficiency scores for UK Power Networks' distribution network operator EPN.

3. *Scale effects in Ofgem's equations.* Many of the Ofgem cost drivers are correlated with network size. The empirical results in the economics literature suggest that scale effects should be small or absent for changes in network size. However some quite large scale effects are present in the Ofgem cost equations for some elements of direct costs. There is no economic reason for such effects. In some cases these effects may arise because of features of particular DNOs. Most of the cost driver variation that delivers the results in the Ofgem analysis occurs at the DNO level and with only 14 DNOs aberrant data from just one or two can have a substantial effect on the position and form of the estimated cost equations.

4. *Alternative forms for cost equations.* The empirical economics literature employs a variety of forms of cost functions but these relate costs to outputs and have a basis in economic theory unlike the Ofgem cost equations, and so cannot be used in this context. The alternative forms for cost equations considered here are simply more flexible versions of the models employed by Ofgem which add quadratic terms in log cost drivers to the Ofgem specification. Experiments with alternative approaches, for example using piecewise linear approximations produced very similar results.

5. *Misspecification of indirect cost equations.* The Ofgem cost equations for total indirect costs, indirect costs group 2 and indirect costs group 3 are misspecified. There is evidence in the data that economies of scale are smaller for large scale operations like EPN than for small to medium scale operations. EPN's position is much improved in all three cases on moving to a more flexible cost equation which captures this effect.

6. *Misspecification of direct cost equations.* Tests indicate rejection of the Ofgem equation for underground faults with the data suggesting increasingly strong economies of scale for operators with large numbers of faults. There seems no good economic reason for this. The result is probably influenced by data from a small number of DNOs. Using a

more flexible cost function improves fit and leaves the positions of UK Power Networks' DNOs almost unchanged. In other cases increasing the flexibility of the fitted cost equations does no harm to the position of UK Power Networks' DNOs and in the case of inspection and maintenance results in substantial improvement.

7. *Totex benchmarking and the specification of cost functions.* In RIIO-ED1 there will be increased emphasis on total expenditure benchmarking. This may bring Ofgem back to the sort of analysis done in DPCR4 in which there was consideration of economic cost functions which relate costs to aspects of output produced by DNOs, for example electricity delivered, customer numbers and network size. It will be prudent for UK Power Networks to examine its likely performance under this sort of benchmarking regime.

8. *Data quality and improvements.* The five years of data covering 14 DNOs available at present is insufficient to allow complex modelling of costs. Much of the cost and cost driver variation occurs across rather than within DNOs. This situation may change by the time cost benchmarking is done in RIIO-ED1 when 8 years of data will be available. It will be prudent to revisit DNO cost data during the run-up to RIIO-ED1 and investigate the potential for using within DNO cost and driver variation to develop more robust cost equations.

9. *Splitting EPN.* Network operators with large geographically dispersed networks may face diseconomies of scale and experience lower average costs by splitting their business into two or more stand-alone entities. Faced with a regulator that persists in employing constant economies of scale cost functions better efficiency scores may be achieved if a large operator like EPN is assessed as two distinct entities. Analysis of data in which EPN's costs and cost drivers are allocated to new DNOs, EPN North and EPN South, shows that, using Ofgem's favoured DPCR5 models, the new EPN South DNO achieves much improved efficiency scores and rankings and there is some significant improvement for the new EPN North.

3 Ofgem's cost analysis

Ofgem's approach to cost benchmarking in DPCR5 involved estimation of equations designed to predict costs defined at a variety of levels of aggregation.[2] Data covering the period 2005-2008 and 14 DNOs was employed. DNO data for 2008 were compared with the predicted costs delivered by the equations for that year and efficiency scores were derived based on the deviations of actual DNO costs from predicted costs.

3.1 Aggregation

The Ofgem cost analysis was done at three levels of aggregation.

1. *Topdown analysis*. In the "topdown" analysis there is a single cost equation for total costs, the sum of: network operating costs whose components are set out below in 2, plus aggregate indirect costs whose components are set out below in 3.

2. *Single group analysis*. In the "single group" analysis there is one cost equation for aggregate indirect costs and one equation for each of four components of network operating costs as follows:

   (a) LV and HV underground faults
   (b) LV and HV overhead faults
   (c) inspection and maintenance
   (d) tree cutting.

3. *Groups analysis*. In the "groups" analysis there are equations for the four components of network operating costs that appear in the single group analysis, as above, and additionally equations for three components of indirect costs as follows.

   (a) Group 1. Network Design, Project Management, System Mapping.
   (b) Group 2. Engineering Management & Clerical Support, Control Centre, Customer Call Centre, Stores, Health & safety.
   (c) Group 3. Network Policy, HR & Non-operational Training, Finance & Regulation, CEO, IT & property.

[2] [OFG1] gives a high level summary of Orgem's procedures and there are details in [OFG2] and [OFG3]. I provided a review of the methods employed in [7].

3.2 Ofgem's cost equations

The cost equations employed by Ofgem in DPCR5 relate costs, expressed in logs, measured over a one year period, to a single explanatory variable which appears in logarithmic form.[3] With costs denoted by C and the explanatory variable denoted by X there is:

$$\log C = a + b \times \log X + U \tag{1}$$

where a and b are unknown parameters to be estimated and U represents variation in log costs not attributable to variation in the explanatory variable.

In some cases the explanatory variable is a single cost driver in logarithmic form. For example in the equation for LV and HV overhead fault costs there is a single cost driver, namely the log of the volume of LV and HV overhead faults.

In other cases the explanatory variable is a weighted average of two cost drivers expressed in logarithmic form.[4] For example in the equation for LV and HV underground fault costs the explanatory variable is a weighted average of the log of the volume of LV and HV underground faults and the log of the length of line replaced.

Table 1 lists the cost drivers that feature in Ofgem's final proposals [OFG1]. Table 2 shows how these cost drivers are used in constructing explanatory variables used in each of the estimated cost equations.

Where two drivers are used let the two drivers be denoted by X_1 and X_2. The single explanatory variable, denoted by X, employed in an Ofgem equation is given by

$$\log X = w_1 \log X_1 + w_2 \log X_2$$

with $w_1 + w_2 = 1$ and the cost equation (1) can be written as

$$\log C = a + b \times (w_1 \log X_1 + w_2 \log X_2) + U$$

equivalently as

$$\log C = a + b_1 \times \log X_1 + b_2 \times \log X_2 + U \tag{2}$$

[3] I use logarithms to base e throughout.

[4] In the final proposals analysis Ofgem used a driver which was a product of log cost drivers each raised to a power. This was clearly an error and there is no doubt that at the next price control review would not persist. The problem is set out in [7].

where $b_1 = bw_1$ and $b_2 = bw_2$ and the weights are given as follows.

$$w_1 = \frac{b_1}{b_1 + b_2} \qquad w_1 = \frac{b_2}{b_1 + b_2} \tag{3}$$

Ofgem originally selected two cost drivers for each of the cost components. In initial analyses they estimated separate coefficients on the cost drivers as in equation (2). In some cases the estimates were in accord with Ofgem's expectations and in those cases the weights are calculated as in (3) using estimates of b_1 and b_2. In some cases one of the two coefficient estimates was close to zero and then the explanatory variable is a single cost driver. In some cases initial estimates gave results using two drivers which Ofgem regarded as implausible and then Ofgem used a single cost driver.

Ofgem reported that some of the cost equations failed specification tests which indicated deficiencies in the functional form employed in certain cases. These failures were dismissed by Ofgem as unimportant. It was stated that *"...the results of our analysis remain robust and fit for purpose."* [5] Some degree of error in functional specification may be acceptable in some circumstances but in cost benchmarking the fit of the equations should be good at all values of cost drivers at which predictions are made or some DNOs may be disadvantaged. It is shown in Section 5 that in some cases EPN was disadvantaged by poor choice of functional form.

Ofgem defended their choice of model stating that the log-log cost equations made economic sense. In fact there is no result in economics to suggest that such a functional form is appropriate - indeed there is a large body of economics research that employs alternative forms for cost functions including the flexible translog cost functions which feature in the literature review in Section 4.

3.3 Ofgem's scale effects

The cost equation (1) written directly in terms of costs rather than log costs takes the form

$$C = X^b e^{a+bU}.$$

and average costs per unit of the cost driver are

$$\frac{C}{X} = X^{b-1} e^{a+bU}.$$

[5] See the last two bullet points in paragraph 1.114 of [OFG3].

In all the cost equations employed by Ofgem in the Final Proposals the value of b is less than one. In this case average costs fall as values of cost drivers increase.

Comparing two cases identical except that in one case the cost driver takes a value $100s\%$ higher than in the other case, costs will be $100\left(s^b\right)\%$ higher. Thus according to the Ofgem cost equation each doubling of the value of a driver delivers costs that are higher by $100(2^b)\%$. For small values of s since $(1+s)^b \simeq 1+bs$ a $100s\%$ increase in a driver is associated with approximately a $100bs\%$ increase in costs.

Values of b substantially below one suggest considerably lower average costs in situations when cost drivers are large. The estimates reported below in Table 5 suggest that a 10% increase in the value of a cost driver is associated with an increase of from 5% to 9% in costs depending on the cost item considered.

3.4 The 5 year data set and the scale of operation of UK Power Networks' DNOs

If substantial economies of scale persist at all scales of operation then large scale operations can be expected to have significantly lower average costs then small scale operations. Eastern Power Networks PLC (EPN) is one of the largest scale DNOs in the UK but its average costs relative to the cost drivers employed by Ofgem are not as low as the Ofgem models suggest they should be. As a result EPN fared poorly in DPCR5. The robustness of the Ofgem models will shortly be examined.

The relative scale of EPN's operation can be seen in Tables 3 and 4 which show the size rankings in 2009 of the DNO's owned by UK Power Networks for each of the cost categories and cost drivers employed by Ofgem. Table 3 gives rankings of costs (1 is the highest) and 4 gives rankings by magnitude of cost drivers. In all but two cases EPN has the largest value of the cost driver and in all cases but one the highest costs.

4 Economic cost functions and economies of scale

In an economic analysis of costs of production interest centres on the relationship between costs of production and the outputs produced. Remarkably outputs figure nowhere in Ofgem's analysis. This was a major change in DPCR5. In DPCR4 costs were related to a composite scale variable which depended on number of customers, electricity delivered and network length.

An economic cost function describes the dependence of costs on the volumes and characteristics of the outputs that a firm produces. The outputs produced by an electricity distribution network operator, say in a year, comprise the amounts of electricity, say KWH delivered to each of its customers in that year. There are quality considerations too, an important aspect of the service provided being its reliability.

DNOs distribute electricity in different amounts to different numbers of customers over service regions of different areas and topographies. There is a substantial empirical economics literature dealing with the economies of scale of various sorts that can arise in such network operations. This is briefly reviewed now.

Following [1] let Q denote the volume of electricity supplied in some period, say KWH, to N customers distributed over a service region with area A, say square miles and define elasticities:

1. the elasticity of cost with respect to volume of *output*
$$\varepsilon_Q = \frac{\partial \log(C)}{\partial \log(Q)} = \frac{\partial C}{\partial Q} / \frac{C}{Q}$$
2. the elasticity of cost with respect to *customer numbers*
$$\varepsilon_N = \frac{\partial \log(C)}{\partial \log(N)} = \frac{\partial C}{\partial N} / \frac{C}{N}$$
3. the elasticity of cost with respect to *service area*
$$\varepsilon_A = \frac{\partial \log(C)}{\partial \log(A)} = \frac{\partial C}{\partial A} / \frac{C}{A}.$$

These are *partial* derivatives measuring the rate of proportionate change in costs that arises for proportionate changes in respectively output, customer numbers and area serviced, Q, N and A, in each case with *all other factors fixed*. So, for example, the output elasticity of cost above shows how costs increase as more electricity is delivered to an unchanging number of customers over a fixed supply area.

In [1] the following measures of economies of scale are defined.[6]

[6] Derivations are given in Annex 1. These are not essential for understanding what follows.

1. R_{OD} captures economies of output density

$$R_{OD} = \frac{1}{\varepsilon_Q}$$

measuring how cost changes when output is increased while customer numbers and service area remain fixed,

2. R_{CD} captures economies of customer density

$$R_{CD} = \frac{1}{\varepsilon_Q + \varepsilon_N}$$

measuring how cost changes when customer numbers increase while output per customer and service area remain fixed.

3. R_S captures economies of network size

$$R_S = \frac{1}{\varepsilon_Q + \varepsilon_N + \varepsilon_A}$$

measuring how cost changes when service area increases while customer density and output per customer remain fixed.

In each case a value of a measure exceeding 1 indicates a degree of economies of scale with average cost falling as the appropriate output measure increases, other factors held fixed as described above.

There is some empirical evidence from a number of countries concerning the magnitudes of these various scale effects. This is now briefly reviewed.

US data from 65 privately owned utilities in the USA 1978 are studied in [1] which finds that average values of R_S and R_{CD} are very close to 1 and that average R_{OD} is significantly in excess of 1 - a value of 1.2 is estimated. So [1] concludes that there is no evidence for falling average costs with size of service area (customer density and output per customer held fixed) or with customer numbers (output per customer and service area held fixed). There is evidence suggesting moderate economies of scale as the amount of electricity delivered is increased to a constant number of customers over an unchanging service area.

[2] studies data on 39 Swiss municipal electricity distribution utilities covering 1988-1991. As in [1] and many of the papers discussed below, an

approximate translog cost function[7] is used which allows the possibility that the various scale measures vary by size of operator. Focussing on the results in [2] for large operators (which are most relevant for UK experience) the estimates suggest R_S is close to 1 and R_{OD} significantly exceeds 1 - a value of 1.15 is reported. The agreement with [1] is quite close in this respect. There is some evidence for small economies of customer density - a value of 1.12 is reported for R_{CD} indicating a small fall in average costs as customer numbers increase in a fixed service area with constant output per customer.

[3] studies around 100 Norwegian electricity distribution companies observed in 1988 and finds no economies of customer density and small economies of output density.

[4] studies 81 Canadian municipal electricity distribution utilities observed in 1993-1995 using state–of-the-art econometric methods and finds little evidence for economies of scale amongst the largest utilities in the study which have around 20,000 customers. It is only these results in [4] that are relevant when considering the large scale DNOs that operate in the UK.

Around 500 US electricity distribution utilities observed in 1989 are studied in [5] which finds evidence for economies of customer density only for relatively small utilities. Amongst larger utilities (>1.5m customers) average distribution costs per KWH actually rise with the number of customers.

Just over 90 electricity distribution units operating in the French distribution network, Electricité Réseau Distribution France (ERDF), in 2003-2005 are studied in [6]. A Cobb-Douglas functional form[8] is used and, using a latent class approach, four types of network are identified. There is evidence of economies of size and customer density in all but the most populous and concentrated of the network types, types which typically arise in urban areas. The values of the scale measures vary from around 1.1 to around 1.4. The utilities in the study had number of customers ranging from 109,000 to 1,596,000 with an average number of 338,000. The service area of the utilities ranged from 107 km^2 to 13,000 km^2 and averaged 5,400 km^2. The results obtained for the large customer numbers network type may be a better guide to the experience of the relatively large UK DNOs.

[7] A translog cost function writes log cost as a general quadratic function of log outputs and factor prices with interactions. It can be thought of as an approximation to a more general functional form. Scale effects can vary with the size of firm allowing the possibility that there are economies of scale amongst small scale enterprises and diseconomies of scale amongst large scale enterprises. See Annex 2.

[8] The Cobb-Douglas form has an equation linear in the various log explanatory variables rather than quadratic as in the translog model used in many of the other studies.

In summary, the evidence in the literature suggests (i) economies of scale (output density) as volumes of electricity delivered are increased, number of customers and service area held fixed and (ii) small or insignificant economies of scale (customer density) amongst larger utilities as customer numbers are increased, output per customer and service area held fixed.

Scale effects seem to be smaller amongst larger utilities. In the UK DNOs typically serve much larger numbers of customers (from 1 million to over 3.5 million customers) over much larger networks than do the utilities covered in most of the studies reviewed here. It seems likely that economies of network size are quite small at these levels of activity.

4.1 Implications for Ofgem's cost analysis

It is not straightforward to relate the empirical evidence on scale effects in electricity distribution to the Ofgem analysis. The reason is that the cost equations used by Ofgem in DPCR5 are not cost functions in the usual economic sense because they do not involve any measures of outputs delivered by DNOs. Instead they involve measures of the size of the asset base and expenditure on additions to it, total costs and measures of volumes of actual or expected inputs to the conduct of DNO's business. Details are given in Table 2. The abbreviations used in that table are given in Table 1.

All of Ofgem's cost drivers are somewhat within the control of the DNOs who can choose their values, at least in the medium to long term, to alter costs of operation. They are what in econometrics are called endogenous variables - that is, variables whose values are determined to some extent by the firms being studied. As such they should not be included in a cost benchmarking analysis based on sound economic principles.

The DPCR4 approach was superior in this regard because it used measures of output as explanatory variables in the cost equations. It is possible that in the next cost benchmarking exercise that approach may be revived since the DPCR5 approach is difficult to defend. In the context of the total expenditure benchmarking proposed in RIIO-ED1 an approach like that taken in DPCR4 based on assessing costs relative to outputs is likely to have some appeal.

All of Ofgem's cost drivers are positively related to scale of operation which recall involves consideration of output (KWH) per customer, number of customers and service area. One would not expect changes in output density (electricity delivered per customer) to have a substantial effect on these cost related items. They are likely to be affected more by changes in customer density (customers per unit service area) and changes in service area. The review of the empirical economics literature set out above suggests that economies of customer density and of size may be rather small at the scale of operation found amongst UK DNOs.

5 An assessment of the scale effects in the Ofgem models

The results reported in this Section are obtained using the 5 year data set covering the period 2005-2009 provided by UK Power Networks May 17th 2011.

5.1 Ofgem equations

The cost equations employed by Ofgem are estimated using the new 5 year data set. The equations take the form

$$\log C = a + b \times \log X + a_{05} \times D_{05} + a_{06} \times D_{06} + a_{07} \times D_{07} + a_{08} \times D_{08} + U \quad (4)$$

where C is the value of a cost item, X is the value of a cost driver (simple or composite as appropriate, as shown in Table 2). The variables D_{05}, D_{06}, D_{07} and D_{08} are binary variables taking the value 1 in the indicated year and zero otherwise.(e.g. D_{05} is 1 for all data in year 2005 and zero for all data in years $2006 - 2009$). The effect of these variables is to shift the cost equations proportionally from year to year. The unobserved term U captures variation in costs not attributable to variation in the explanatory variables.

Following Ofgem the estimates are obtained using the ordinary least squares procedure. Accuracy of estimates can be assessed by considering estimated standard errors. An approximate 95% confidence interval for the value of a coefficient, say b , is given by the estimated value of the coefficient plus and minus twice the estimated standard error. In the analysis reported here standard errors are calculated allowing for the correlation amongst costs that is likely to arise across years within DNOs.[9]

Table 5 shows the estimated values of the coefficients on the cost drivers, simple or composite as indicated.

[9] This allowance for "clustering" in the data causes estimated standard errors to rise quite substantially, in many cases by a factor of around two.

In each case the estimated coefficient is less than one and in the cases starred (total costs, total indirect costs, indirect costs group 3, trees) the estimated coefficient is more than two standard errors below 1 indicating a statistically significant deviation. However in many cases the estimates are quite inaccurate and quite large departures from values of 1 are not judged significantly different from 1. The inaccuracy in these estimates feeds through to the efficiency scores.

Ofgem noted that a number of the DPCR5 cost equations failed specification tests which detect incorrect specification of functional form. The possibility of departures from Ofgem's simple log linear form is now investigated.

5.2 Functional form

The simple Cobb-Douglas form used by Ofgem is extended by introducing a quadratic term, adding the term $c \times (\log X)^2$ to the cost function (4). This creates a model like the translog model referred to earlier and the term "translog" will be used in this Section to describe the model in which there is a linear and a quadratic term in the log cost driver.

The final column of Table 5 shows the value of the t-statistic for testing the hypothesis $c = 0$.[10] A positive value indicates a less concave cost-cost driver relationship than Ofgem found with economies of scale being attenuated at high values of the cost driver. A negative value indicates stronger economies of scale than Ofgem found. Only values of the test statistic with magnitudes close to or exceeding 2 indicate a statistically significant departure from Ofgem's form and such magnitudes are found in only two cases - underground faults (negative) and indirect costs group 2 (positive).

The problem here is that the cost data are quite dispersed and the data is not very extensive. There are 70 data points in the data set but the variation in the data within DNOs across time is not very informative, being rather loosely related to within DNO variations in cost drivers. Most of the variation in costs and cost drivers arises at the DNO level and there are only 14 of these. So there is limited scope for accurate estimation of cost equations with a more complex structure than the linear in logs Cobb-Douglas form used by Ofgem. By the time of the cost benchmarking exercise in RIIO-ED1 there will be 8 years of data and then there may be scope for more complex modelling.

[10] The computation is done taking account of clustering within DNOs.

Further insight into the performance of the Ofgem equations and the impact of alternative forms on the position of UN Power Networks' DNOs can be obtained by examining graphs of the data. The 8 cost items are considered in turn.

5.2.1 Topdown analysis: Total costs

Figure 1 shows the 70 total cost data points. Here and in the graphs that follow the vertical axis measures log costs and the horizontal axis measures the cost driver which is in logarithmic form, in this case a weighted average of log MEAV and log LDNL. Data points for 2009 are shown as filled circles. Data points for EPN, LPN and SPN are coloured respectively in red, blue and green. Data points for the other DNOs are plotted in grey. The cost equation estimated using the Ofgem model is superimposed. Year effects have been removed from the costs data prior to plotting by adjusting the costs for all years except 2009 using the estimated coefficients on the year indicator variables, $a_{05}, \ldots, a_{08}$ (see equation (4)).

The estimated total cost equation fits the data well and there is no indication of nonlinearity. The 2009 result for EPN lies a little above the fitted cost equation which leads to a good efficiency score outcome. The topdown results were given a low weighting by Ofgem in their Final Proposals. That situation may change with the increased emphasis on total cost benchmarking proposed for RIIO-ED1.

5.2.2 Single group analysis: total indirect costs

Figure 2 shows the total indirect costs data plotted against the cost driver in logarithmic form. The statistical test for nonlinearity is at the margins of significance suggesting some degree of misspecification and the reason can be seen in Figure 2a where the fitted quadratic relationship is plotted. According to this estimate the economies of scale parameter[11] estimated for operations on the scale of EPN is 1.12 compared with the estimate of 1.44 obtained using the linear model. The graph shows that the position of EPN is much improved on moving to the quadratic model while the position of LPN and SPN is virtually unchanged.

[11] The economies of scale parameter is the inverse of the elasticity $\left(\frac{\partial \log C}{\partial \log X}\right)^{-1}$ where C is cost and X is the cost driver. In the translog model the measure varies with C and X. In the Cobb-Douglas model it is constant.

5.2.3 Groups analysis: indirect costs group 1

Figure 3 shows the group 1 indirect cost data plotted against the cost driver in logarithmic form. The cost data are quite dispersed and there is little sign of nonlinearity which is in accordance with the small value of the nonlinearity test statistic shown in Table 5. Figure 3a show the fitted quadratic relationship which has hardly any curvature. EPN's cost data lie way above the fitted cost functions in 2008 and 2009. This is in contrast to the 2005-2007 data for EPN and occurs because group 1 costs for EPN increased by 25% between 2007 and 2008.

5.2.4 Groups analysis: indirect costs group 2

The group 2 indirect cost data are plotted in Figure 4. In this case the test statistic given in Table 5 indicates decisive rejection of the Ofgem model. Figure 4a shows the fitted quadratic relationship and there is clearly substantial curvature. The economies of scale parameter for operations at the scale of EPN is estimated to be 0.76 using the extended model compared with 1.10 using the Ofgem model. The quadratic model suggests significant *diseconomies* of scale for large scale operations.

The graphs show that EPN's position is greatly improved on moving to the quadratic model, LPN's position is slightly improved and SPN's position is unchanged.

5.2.5 Groups analysis: indirect costs group 3.

Ofgem's analysis was done at the DNO group level. Group 3 costs and the cost driver (MEAV) were aggregated across DNOs within 7 groups. The driver does not vary within DNO groups across years. As a result there is rather limited data but nevertheless there is clear evidence of misspecification.

The test statistic for nonlinearity in Table 5 is highly significant. Figure 5 shows the group 3 cost data. UKPN has costs well above predicted costs resulting in a poor efficiency score. Figure 5a shows the fitted quadratic relationship which seems to capture some essential nonlinearity and results in a major improvement in the position of UK Power Networks. The Ofgem model has constant economies of scale at all scales of operation with a scale measure equal to 1.3. The extended model predicts a scale measure around 1.5 for moderate sized operations but 0.6 (diseconomies of scale) for operations at the scale of UKPN.

5.2.6 Single group and groups analysis: underground faults

In this case the Ofgem model is decisively rejected. The underground faults cost data are plotted in Figure 6. Data from three of the DNOs make the main contributions to the rejection of the Ofgem form. SSE Hydro has very low underground faults costs even given the low value for its cost driver. CE YEDL and SSE Southern have high values for the cost driver but relatively low costs. Removing these three DNOs results in less curvature in the fitted cost equation but some remains.

Figure 6a shows the fitted quadratic (translog) cost function using the full data set. Here EPN is not the DNO with the largest value of the cost driver and its position is almost unchanged on moving to the translog form. There is no obvious economic argument for the nonlinearity in the log cost relationship for underground faults. It seems likely that it arises because of particular conditions in a few DNOs. One possibility, perhaps not a very likely one, is that there is reverse causation here. DNO's who make low quality, low cost repairs experiencing larger numbers of faults in consequence.

5.2.7 Single group and groups analysis: overhead faults

Figure 7 show the cost data for overhead faults costs. LPN does not have costs in this category and is excluded. The cost data are quite dispersed. There is no evidence in the plot of nonlinearity which accords with the statistical test in Table 5. EPN's cost data lies well above the fitted cost equation resulting in a poor efficiency score for this cost component. EPN's cost driver value is lower in 2009 than in 2007 by around 20% but its costs were around 50% higher comparing 2009 with 2007. If the 2007 level of performance had continued into 2009 the position of EPN would be much improved.

5.2.8 Single group and groups analysis: tree cutting costs

Figure 8 shows the tree cutting cost data. LPN does not have costs in this category. The statistical test for nonlinearity is on the margins of rejection and the fitted translog cost function has some slight curvature - see Figure 8a. This seems to be partly due to data from ENW which had very low values for tree cutting costs and the cost driver "spans cut" at the start of the period studied. ENW's driver values increased 5-fold between 2005 and 2009 while its tree cutting costs increased only 3 fold. Removing ENW from the estimation removes the nonlinearity but EPN's position is not greatly improved.

5.2.9 Single group and groups analysis: inspection and maintenance costs

Figure 9 show the inspection and maintenance cost data. There is a lot of dispersion in the cost data with some of the medium sized DNOs having relatively low costs in this category. The statistical test does not reject the Ofgem form but on fitting the translog form there is some sign of curvature in the log cost relationship - see Figure 9a. EPN's position is much improved on moving to the translog form.

5.3 Variation within DNOs

Confidence in the form of the cost equation employed by Ofgem would be increased if the estimated magnitudes of the effect of cost drivers on cost outcomes were similar when only variation in drivers and costs within DNOs is considered. We have attempted to estimate cost equations using only within-DNO cost and driver variation but the results are in the main uninformative. A contributory factor is the relatively limited variation in driver values within DNOs for some cost categories. This situation may have changed by the time the cost benchmarking is done in RIIO-ED1 by when there will be 8 years of data.

6 Splitting EPN

Companies that operate over a large geographically dispersed network at a scale at which they experience diseconomies of scale may do better to split their operations into two or more free standing organisations. Even if there are not in fact diseconomies of scale, faced with assessment by a regulator that employs models that exhibit economies of scale splitting operations may lead to better performance in cost benchmarking.

In order to assess the benefits of such a manoeuvre the Ofgem efficiency analysis of DPCR5 was re-done on a data set in which EPN was split into two operations, EPN North and EPN South.[12]

[12] The results reported here follow on from two earlier sets of results. Initially the data were split in a mechanical fashion and various ratios of split between EPN South and EPN North were tried. Those data were received May 15th 2011. Then followed a data set in which the data split was done with attention to geographical differences in activities,

Table 6 shows the rankings of the UK Power Networks DNOs with EPN as a single entity, (i) using the 2005-2008 data - as in DPCR5 itself - and (ii) using the 5 year data set 2005-2009. The final column in Table 6 shows (iii) the rankings obtained when EPN is split into two entities.

Table 29 show the efficiency scores and rankings of all the DNOs when EPN is treated as two entities. Results are presented for a topdown analysis, a single group analysis and a groups analysis. In DPCR5 Ofgem constructed weighted averages of scores from such results bringing into the average scores obtained with some alternative specifications as well. This operation has not been reproduced.

The splitting operation brings a great improvement for EPN South relative to EPN and, except in the topdown analysis, some improvement too for EPN North. Ofgem placed only a small weight on the topdown analysis in DCPR5 so the single group and groups analysis are of more interest at least historically. The position of SPN and LPN are slightly damaged by the EPN split, but mainly because the new EPN DNOs come before them in the rankings shifting their ranking down by one or two points.

7 Conclusions

The results of this research are summarised in Section 2.

References

1.Ofgem documents[13]

[OFG1] EDPCR Final Proposals. Ref: 144/09. 7th December 2009.

[OFG2] EDPCR Final Proposals - Allowed Revenue - Cost Assessment. Ref: 146/09. 7th December 2009.

[OFG3] EDPCR Final Proposals - Allowed Revenue - Cost Assessment. Appendix 8. Operational Cost Assessment Further Details. Ref: 146a/09. 7th December 2009.

[OFG4] EDPCR Final Proposals, November 2004, Ofgem report reference 265/04.

network features and so forth. These data were received June 6th 2011. The results presented here were obtained using a revised version of that split data delivered on July 11th 2011..

[13] EDPCR: Electricity Distribution Price Control Review.

2. Other documents

[1] "Economies of density and size in the production and delivery of electric power," Mark Roberts, 1986, *Land Economics*, 62(4), 378-387.

[2] "Are municipal electricity distribution utilities natural monopolies?" Massimo Filippini, 1998, *Annals of Public and Cooperative Economics*, 69(2), 157-174.

[3] "Productivity differences in multi-output industries: an empirical application to electricity distribution," K. Selvanes and S Tjøta, 1994, *The Journal of Productivity Analysis*, 5, 23-43.

[4] "Scale economies in electricity distribution: a semiparametric analysis," A. Yatchew, 2000, *Journal of Applied Econometrics*, 15, 187-210.

[5] "Electric power distribution costs: analysis and implications for restructuring," J. Kwoka, 2001, Report prepared for the American Public Power Association.

[6] "The economies of scale in French power distribution utilities," M. Farsi, M. Filippini, M-A. Plagnet and R. Saplacan, 2010, CEPE Working Paper No. 73, Swiss Federal Institute of Technology.

[7] "A Review of the DPCR5 Cost Benchmarking Exercise," Andrew Chesher, January 18th 2010, reported prepared for EDF Energy.

[8] *Economies of Scale and the Form of Production Function*, Z. Griliches and V. Ringstad, 1971, North-Holland Publishing Co., Amsterdam.

[9] "A Theory of Production," C. Cobb, and P. Douglas, 1928, *American Economic Review*, 18, 139–65.

[10] "The Translog Function and the Substitution of Equipment, Structures and Labor in U.S. Manufacturing, 1929-1968", E. Berndt and L. Christensen, 1973, *Journal of Econometrics*, 1, 81-114.

[11] "Transcendental logarithmic production functions," L. Christensen, D. Jorgenson and L. Lau, 1973, *The Review of Economics and Statistics*, 55(1), 28-45.

Annex 1: Measures of economies of scale

In this Annex the expressions for the various measures of economies of scale set out in Section 4 and employed in [1] are derived. It is assumed that costs over some period, C, are a function of the volume of electricity supplied, Q, number of customers, N, service region area, A, and other factors, Z.

$$C = f(Q, N, A, Z)$$

In terms of differentials the equation

$$dC = \frac{\partial C}{\partial Q} dQ + \frac{\partial C}{\partial N} dN + \frac{\partial C}{\partial A} dA.$$

shows how small changes, dQ, dN and dA, in respectively electricity supplied, customer numbers and service region area lead to a change, dC, in costs. It is assumed that the variables can be regarded as continuously varying and that the cost function is smooth. The derivatives above are conventional partial derivatives that give the rate of change of cost with respect to arguments of the cost function, in each case with all other arguments held constant.

In the economics and productivity analysis literature much of the discussion proceeds in terms of elasticities, log derivatives, defined as follows.[14]

$$\varepsilon_Q \equiv \frac{\partial \log C}{\partial \log Q} = \frac{\partial C}{\partial Q} / \frac{C}{Q}$$

$$\varepsilon_N = \frac{\partial \log C}{\partial \log N} = \frac{\partial C}{\partial N} / \frac{C}{N}$$

$$\varepsilon_A = \frac{\partial \log C}{\partial \log A} = \frac{\partial C}{\partial A} / \frac{C}{A}.$$

The following three measures of economies of scale are commonly considered. In each case a value exceeding 1 indicates that economies of scale (as opposed to diseconomies of scale) are present.

1. *Economies of scale with respect to output density* - R_{OD}. This measures how cost changes when output is increased while customer numbers and service area remain fixed. There is:

$$R_{OD} \equiv \left(\left. \frac{d \log C}{d \log Q} \right|_{N=n, A=a} \right)^{-1} = \frac{1}{\varepsilon_Q}$$

because:

$$\left. \frac{d \log C}{d \log Q} \right|_{N=n, A=a} = \frac{\partial \log C}{\partial \log Q} = \varepsilon_Q.$$

The derivative involved here is the conventional partial derivative.

[14] Throughout natural logarithms are used.

2. *Economies of scale with respect to customer density* - R_{CD}. This measures how cost changes when customer numbers increase while output per customer and service area remain fixed. There is:

$$R_{CD} \equiv \left(\left. \frac{d \log C}{d \log N} \right|_{Q/N=q, A=a} \right)^{-1} = \frac{1}{\varepsilon_Q + \varepsilon_N}$$

and now the derivative is *not* the conventional partial derivative because Q and N are allowed to vary, but always such that Q/N is constant. The derivative is calculated as follows:

$$\begin{aligned} \left. \frac{dC}{dN} \right|_{Q/N=q, A=a} &= \frac{\partial C}{\partial Q} \times \left. \frac{dQ}{dN} \right|_{Q/N=c} + \frac{\partial C}{\partial N} \\ &= \frac{\partial C}{\partial Q} \times \frac{Q}{N} + \frac{\partial C}{\partial N} \end{aligned}$$

where

$$\frac{dQ}{dN}|_{Q/N=c} = c = \frac{Q}{N}$$

is employed, and finally there is the following.

$$\begin{aligned} \left. \frac{d \log C}{d \log N} \right|_{Q/N=q, A=a} &= \left. \frac{dC}{dN} \right|_{Q/N=q, A=a} \times \left(\frac{N}{C} \right) \\ &= \frac{\partial C}{\partial Q} \times \frac{Q}{C} + \frac{\partial C}{\partial N} \times \frac{N}{C} \\ &= \varepsilon_Q + \varepsilon_N \end{aligned}$$

3. *Economies of scale with respect to size* - R_S. This measures how cost changes when service area increases while customer density and output per customer remain fixed. Here one is essentially replicating the network over a larger service area. There is:

$$R_S \equiv \left(\left. \frac{d \log C}{d \log A} \right|_{Q/N=q, N/A=a} \right)^{-1} = \frac{1}{\varepsilon_Q + \varepsilon_N + \varepsilon_A}$$

and again this is not a conventional partial derivative because Q and N are allowed to vary but subject to the restriction that $Q/N = q$ and

$N/A = a$. The derivative is calculated as follows:

$$\begin{aligned}
\left.\frac{dC}{dA}\right|_{Q/N=q,N/A=a} &= \frac{\partial C}{\partial Q} \times \left.\frac{dQ}{dA}\right|_{Q/N=q,N/A=a} + \frac{\partial C}{\partial N} \times \left.\frac{dN}{dA}\right|_{N/A=a} + \frac{\partial C}{\partial A} \\
&= \frac{\partial C}{\partial Q} \times \left.\frac{dQ}{dA}\right|_{Q/A=qa} + \frac{\partial C}{\partial N} \times \left.\frac{dN}{dA}\right|_{N/A=s} + \frac{\partial C}{\partial A} \\
&= \frac{\partial C}{\partial Q} \times \frac{Q}{A} + \frac{\partial C}{\partial N} \times \frac{N}{A} + \frac{\partial C}{\partial A}
\end{aligned}$$

and so:

$$\begin{aligned}
\left.\frac{d\log C}{d\log A}\right|_{Q/N=q,N/A=a} &= \left.\frac{dC}{dA}\right|_{Q/N=q,N/A=a} \times \left(\frac{A}{C}\right) \\
&= \left(\frac{\partial C}{\partial Q} \times \frac{Q}{C} + \frac{\partial C}{\partial N} \times \frac{N}{C} + \frac{\partial C}{\partial A} \times \frac{A}{C}\right) \times \left(\frac{A}{C}\right) \\
&= \frac{\partial C}{\partial Q} \times \frac{Q}{C} + \frac{\partial C}{\partial N} \times \frac{N}{C} + \frac{\partial C}{\partial A} \times \frac{A}{C} \\
&= \varepsilon_Q + \varepsilon_N + \varepsilon_A.
\end{aligned}$$

Annex 2: Cobb-Douglas and translog cost functions

In a simple cost model in which a Cobb-Douglas form [ref] is used the cost function is specified as

$$\log C = \alpha + \beta_Q \log Q + \beta_N \log N + \beta_A \log A + g(Z) + U$$

where g is some function bringing other factors(for example factor prices) into the model and U is an unobservable term capturing cost variations not attributable to the observed cost drivers. In many applications the function g will be linear and then the coefficients can be estimated by the Ordinary Least Squares (OLS) method or some variant of this. The coefficients β_Q, β_N and β_A are equal to the elasticities of Annex 1, respectively ε_Q, ε_N and ε_A.

A more sophisticated approach frequently adopted specifies a quadratic function on the right hand side allowing not only squared terms $(\log Q)^2$, $(\log N)^2$ and $(\log A)^2$ but also interactions which may involve elements of Z. In this "translog" model, [9], [10], [11], the elasticities depend on the values of Q, N and A at which derivatives are evaluated. So the translog model allows the possibility that the various economies of scale measures can take different values at different scales and type of operation. Some of the empirical evidence suggests variations in scale effects. The model is usually regarded as an approximation to a unknown nonlinear function.

The translog model is commonly estimated using OLS or some variant. To obtain usable accurate estimates requires substantial amounts of data with a good representation of various combinations of output measures.

Abbreviation in:		Comments
data files	this document	
MEAV	MEAV	modern equivalent asset value
LDNL	LDNL	network investment (labour and contractors' costs only)
directlabandcont	TDC	total direct costs (less nonoperational capex)
lvhvugfaultvols	UGFV	number of low and high voltage underground faults
lvhvohfaultvols	OHFV	number of low and high voltage overhead faults
lengthlinereplaced	LLR	length of line replaced
trees	TSC	tree cutting: number of spans cut
iandm	AMH	hours of inspection and maintenance for asset base

Table 1: Abbreviated names of cost drivers

Cost item	Analysis	Cost drivers and weights
total direct and indirect costs	topdown	MEAV (.73), LDNL (.27)
total indirect costs	single group	MEAV (.50), TDC (.50)
indirect costs, group 1	groups	MEAV (.50), LDNL (.50)
indirect costs, group 2	groups	MEAV (.54), TDC (.46)
indirect costs, group 3	groups	MEAV
lv & hv underground faults	single group, groups	UGFV (.82), LLR (.18)
lv & hv overhead faults	single group, groups	OHFV
trees	single group, groups	TSC
inspection and maintenance	single group, groups	AMH

Table 2: Cost drivers in Ofgem's final proposals

Cost item	Cost drivers and weights	EPN	LPN	SPN
total direct and indirect costs	MEAV (.73), LDNL (.27)	1	9	6
total indirect costs	MEAV (.50), TDC (.50)	1	8	5
indirect costs, group 1	MEAV (.50), LDNL (.50)	1	7	5
indirect costs, group 2	MEAV (.54), TDC (.46)	1	7	5
indirect costs, group 3	MEAV	na	na	na
lv & hv underground faults	UGFV (.82), LLR (.18)	5	9	11
lv & hv overhead faults	OHFV	1	na	8
trees	TSC	1	na	4
inspection and maintenance	AMH	2	14	5

Table 3: Ranking by size of cost driver of UK Power Network's DNO's, "na" = not applicable.

Cost item	Cost drivers and weights	EPN	LPN	SPN
total direct and indirect costs	MEAV (.73), LDNL (.27)	1	10	6
total indirect costs	MEAV (.50), TDC (.50)	1	9	4
indirect costs, group 1	MEAV (.50), LDNL (.50)	1	7	3
indirect costs, group 2	MEAV (.54), TDC (.46)	1	9	4
indirect costs, group 3	MEAV	na	na	na
lv & hv underground faults	UGFV (.82), LLR (.18)	2	11	9
lv & hv overhead faults	OHFV	1	na	5
trees	TSC	1	na	7
inspection and maintenance	AMH	1	9	6

Table 4: Ranking by magnitude of costs of UK Power Network's DNO's

Cost item	**Analysis**	**Cost drivers**	Estimated b (standard error)	Nonlinearity test
total direct and indirect costs	topdown	MEAV, LDNL	0.81 * (.05)	0.26
total indirect costs	single group	MEAV, TDC	0.69 * (.08)	1.60
indirect costs, group 1	groups	MEAV, LDNL	0.85 (.21)	-0.61
indirect costs, group 2	groups	MEAV, TDC	0.91 (.12)	2.72**
indirect costs, group 3	groups	MEAV	0.79 (.19)	2.72**
lv & hv underground faults	single group, groups	UGFV, LLR	0.86 (.22)	-2.49**
lv & hv overhead faults	single group, groups	OHFV	0.95 (.09)	0.78
trees	single group, groups	TSC	0.50 * (.12)	-1.78
inspection and maintenance	single group, groups	AMH	0.73 (.21)	1.21

Table 5: Cost drivers in Ofgem's final proposals and estimated value of b. Estimates marked "*" are significantly below 1. Tests statistics marked "**" indicate rejection of linearity.

Analysis	DNO	2005-2008	2005-2009	2005-2009 EPN split
Topdown	LPN	4	3	3
	SPN	14	8	9
	EPN	13	12	
	EPN North			14
	EPN South			6
Single Group	LPN	8	7	9
	SPN	12	9	10
	EPN	14	14	
	EPN North			8
	EPN South			6
Groups	LPN	10	11	12
	SPN	14	10	13
	EPN	13	13	
	EPN North			10
	EPN South			8

Table 6: Rankings of the UK Power Networks DNOs with and without EPN split into EPN North and EPN South

DNO	**Topdown**		**Single Group**		**Groups**	
	Score	**Rank**	**Score**	**Rank**	**Score**	**Rank**
CN West	112	12=	117	15	116	15
CN East	102	8	107	10	102	6
ENW	112	12=	113	13=	104	7
CE NEDL	88	4	83	1=	85	3
CEYEDL	86	2	83	1=	82	2
WPD S Wales	109	11	100	7	106	9
WPD S West	108	10	93	5	93	4
UKPN LPN	**87**	**3**	**102**	**9**	**109**	**12**
UKPN SPN	**104**	**9**	**107**	**10=**	**110**	**13**
UKPN EPNN	**113**	**14**	**101**	**8**	**107**	**10**
UKPN EPNS	**94**	**6**	**96**	**6**	**105**	**8**
SP Distribution	100	7	113	13=	108	11
SP Manweb	116	15	112	12	112	14
SSE Hydro	91	5	89	4	96	5
SSE Southern	85	1	85	3	79	1

Table 7: Efficiency scores and rankings of all DNOs with EPN split into EPN North and EPN South, 2005-2009 data

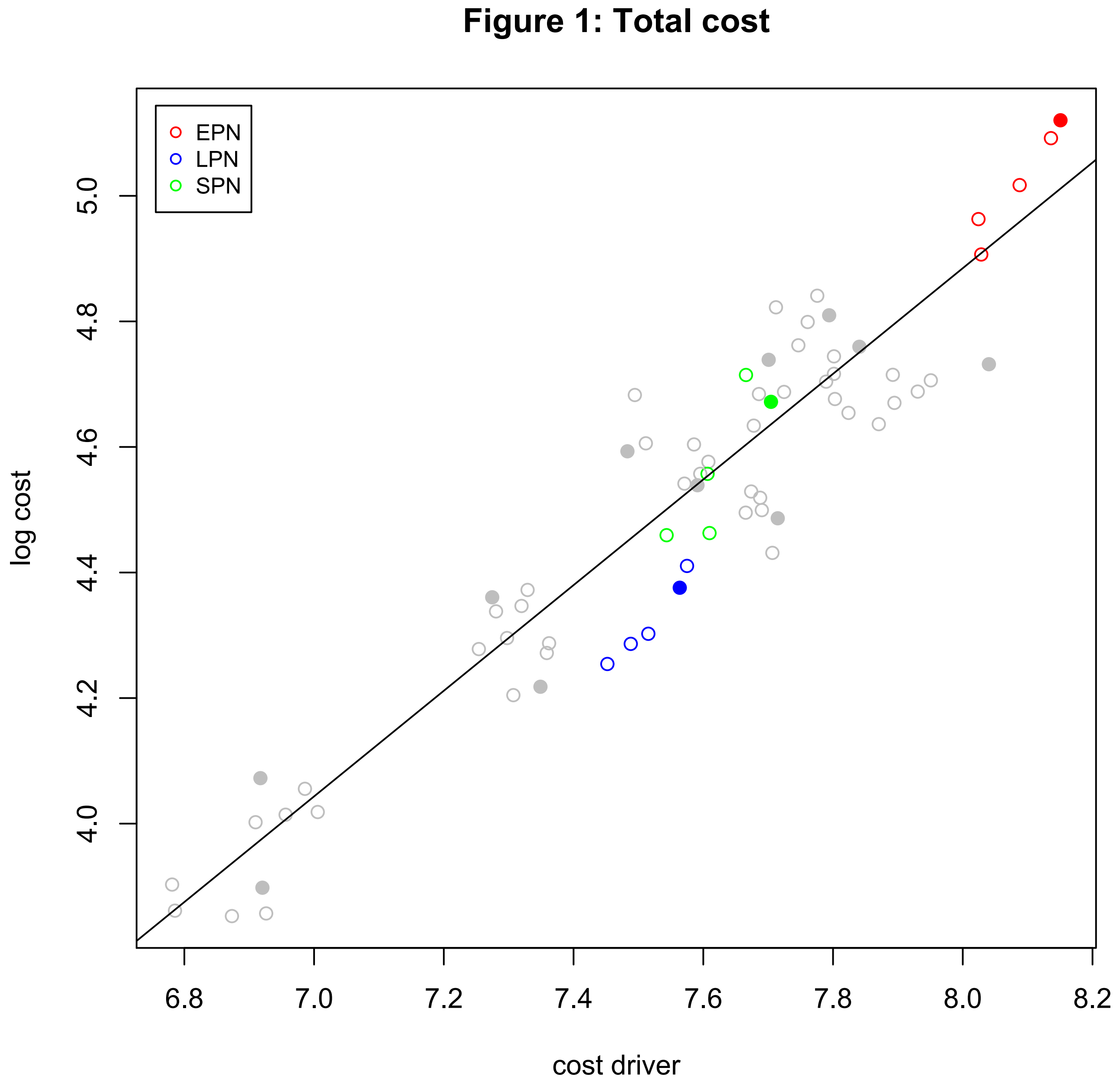


Figure 1: Total cost
EPN
LPN
SPN
log cost
cost driver
6.8
7.0
7.2
7.4
7.6
7.8
8.0
8.2
4.0
4.2
4.4
4.6
4.8
5.0

Figure 2: Total Indirect Costs

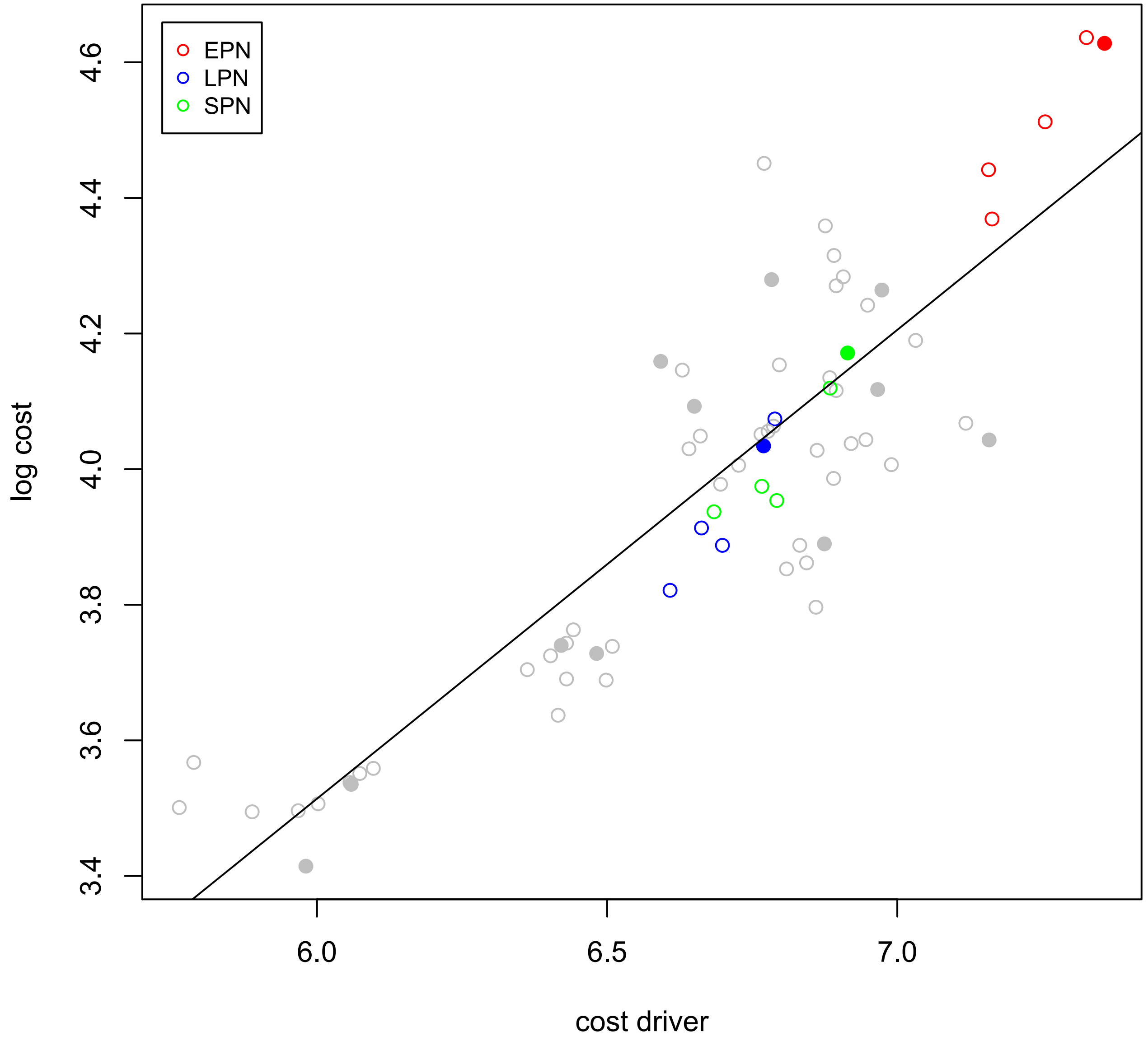


Figure 2a: Total Indirect Costs, quadratic

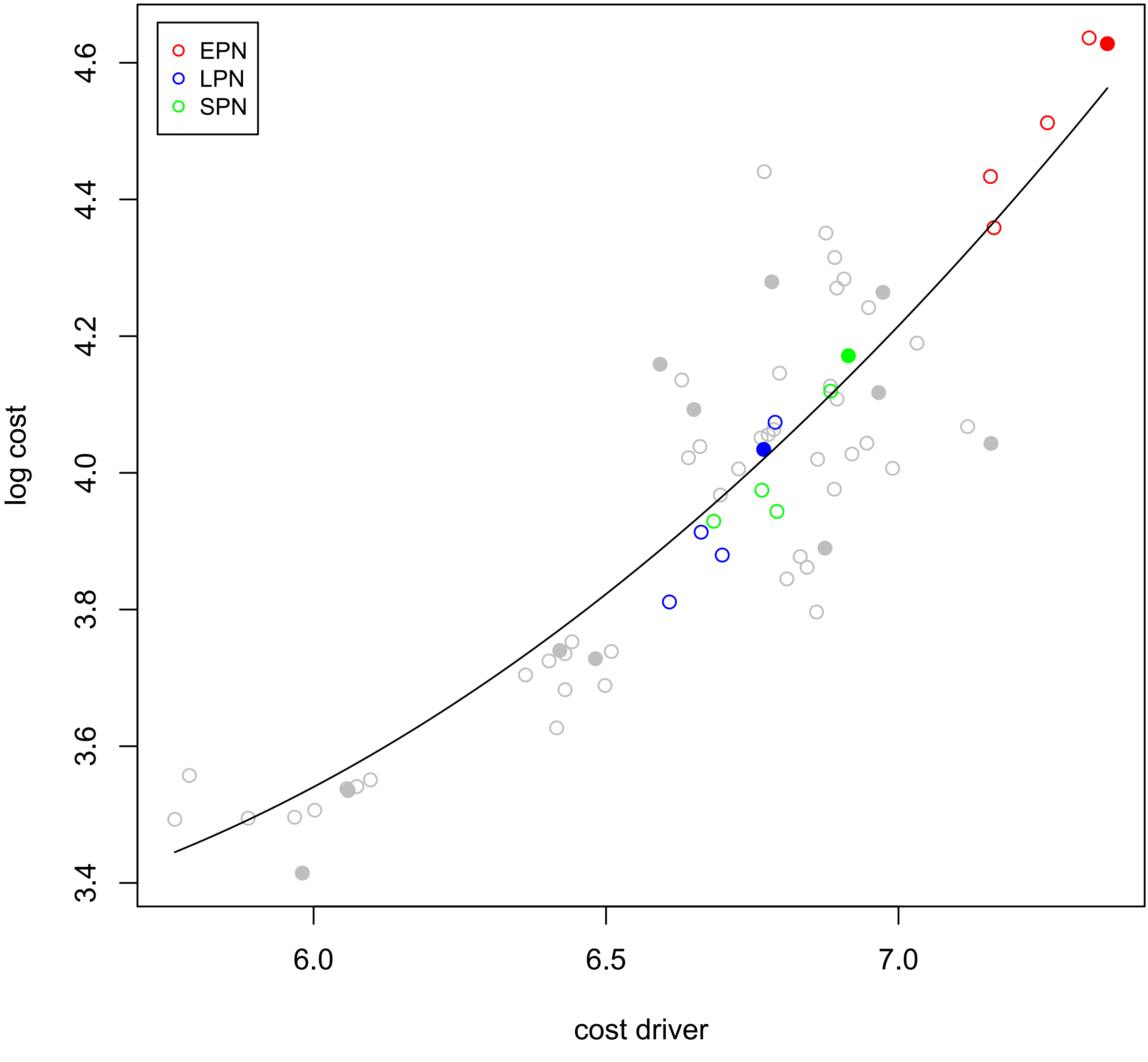


Figure 3: Indirect Costs Group 1

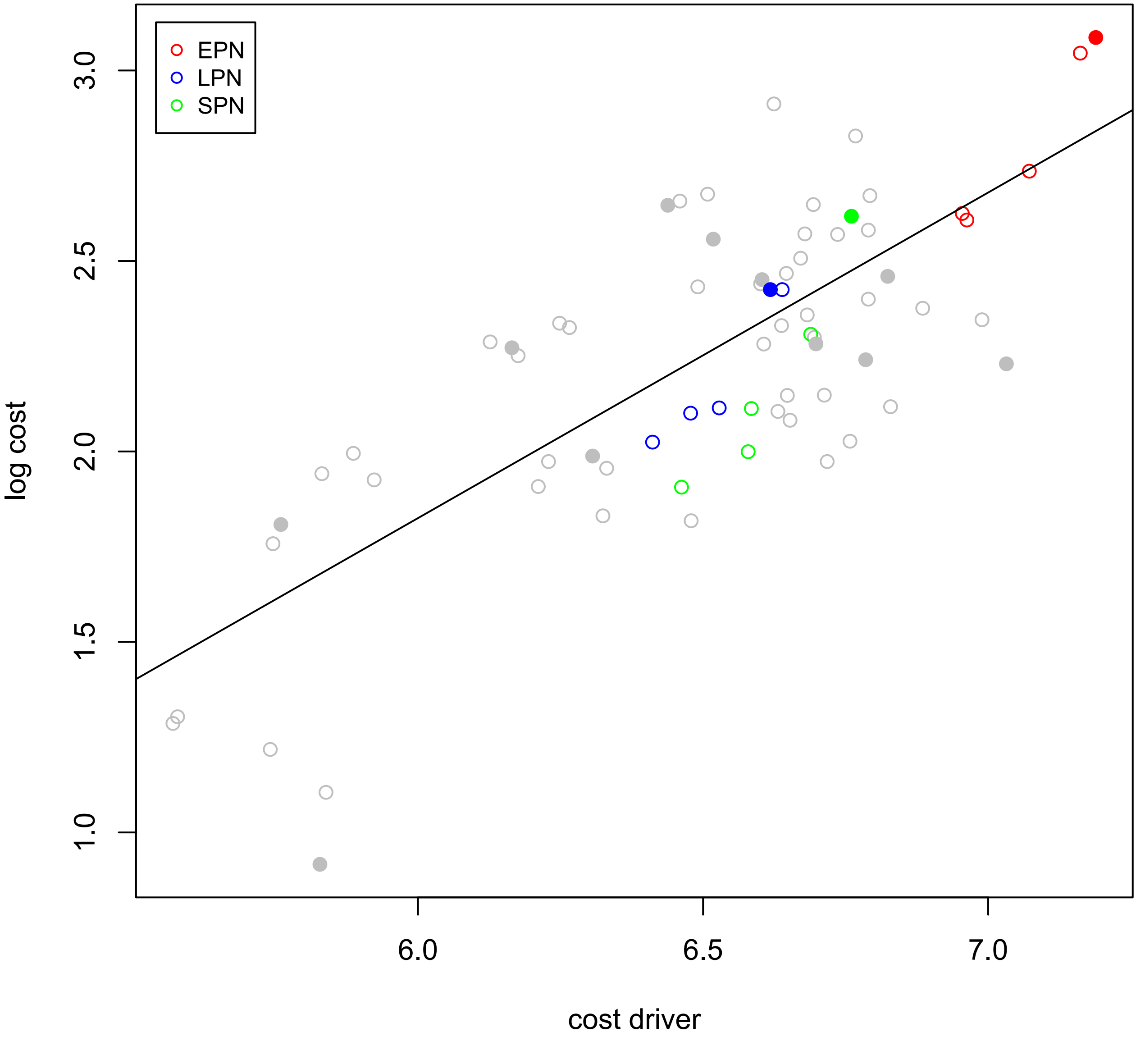


Figure 3a: Indirect Costs Group 1, quadratic

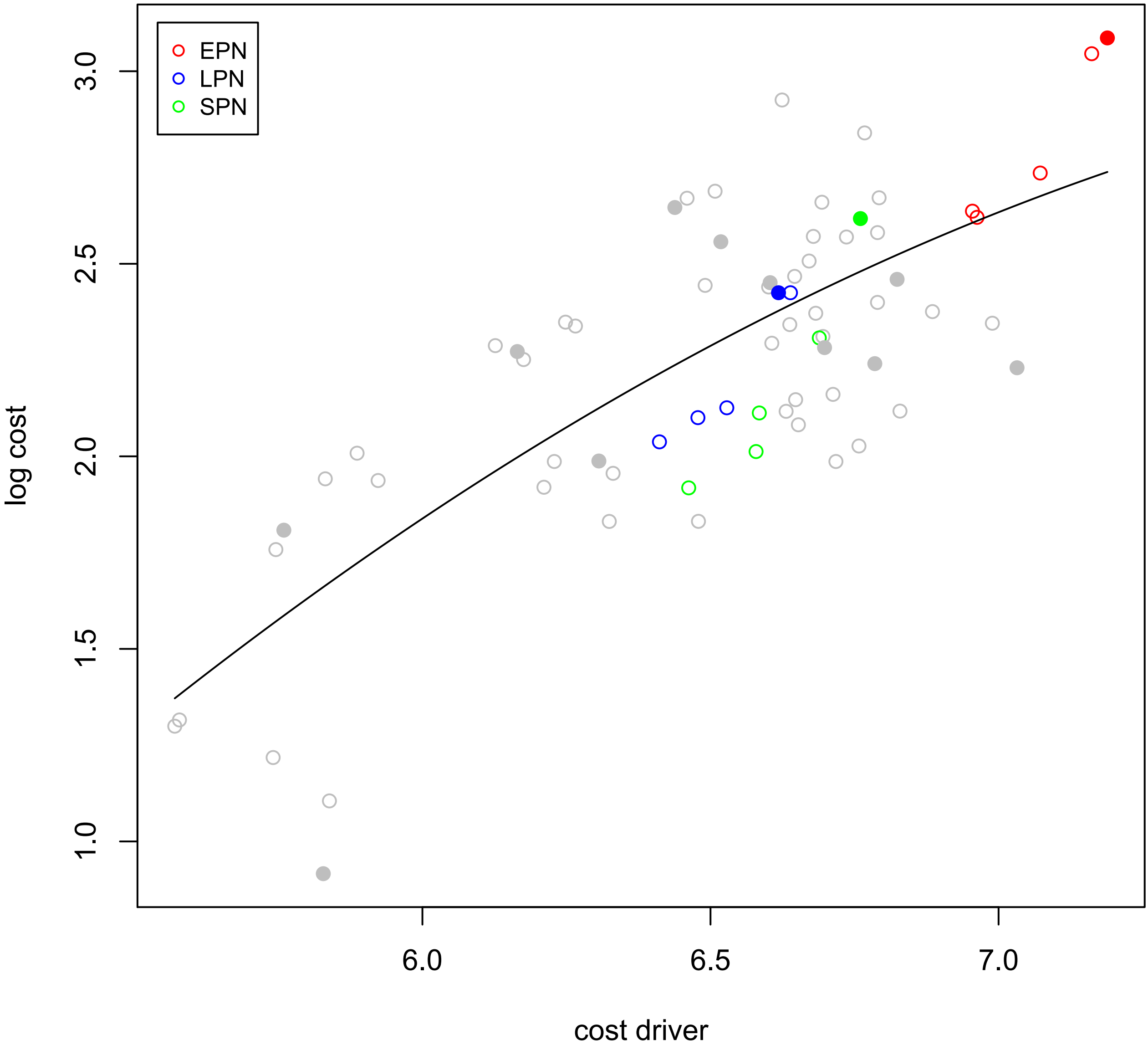


Figure 4: Indirect Costs Group 2

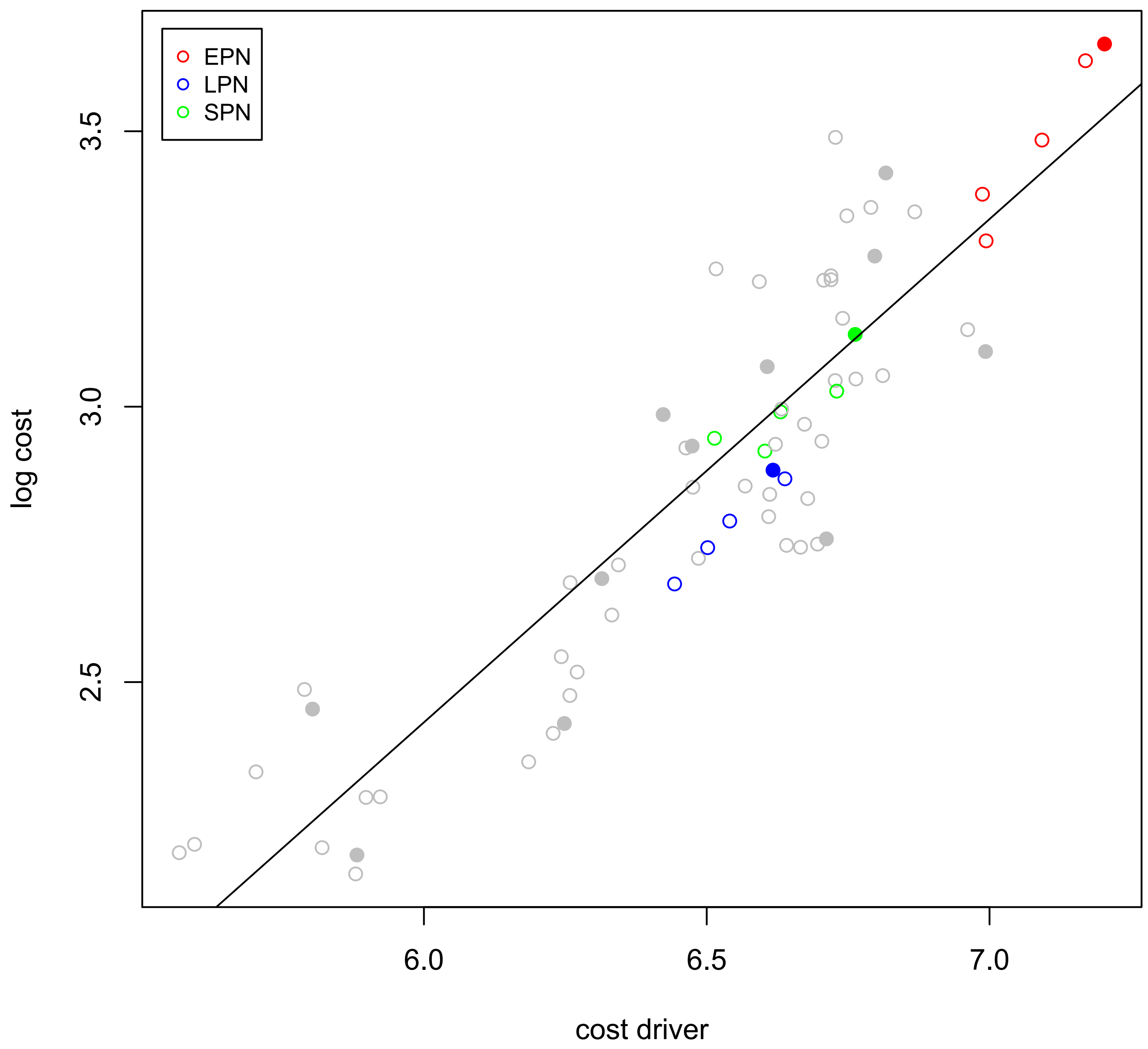


Figure 4a: Indirect Costs Group 2, quadratic

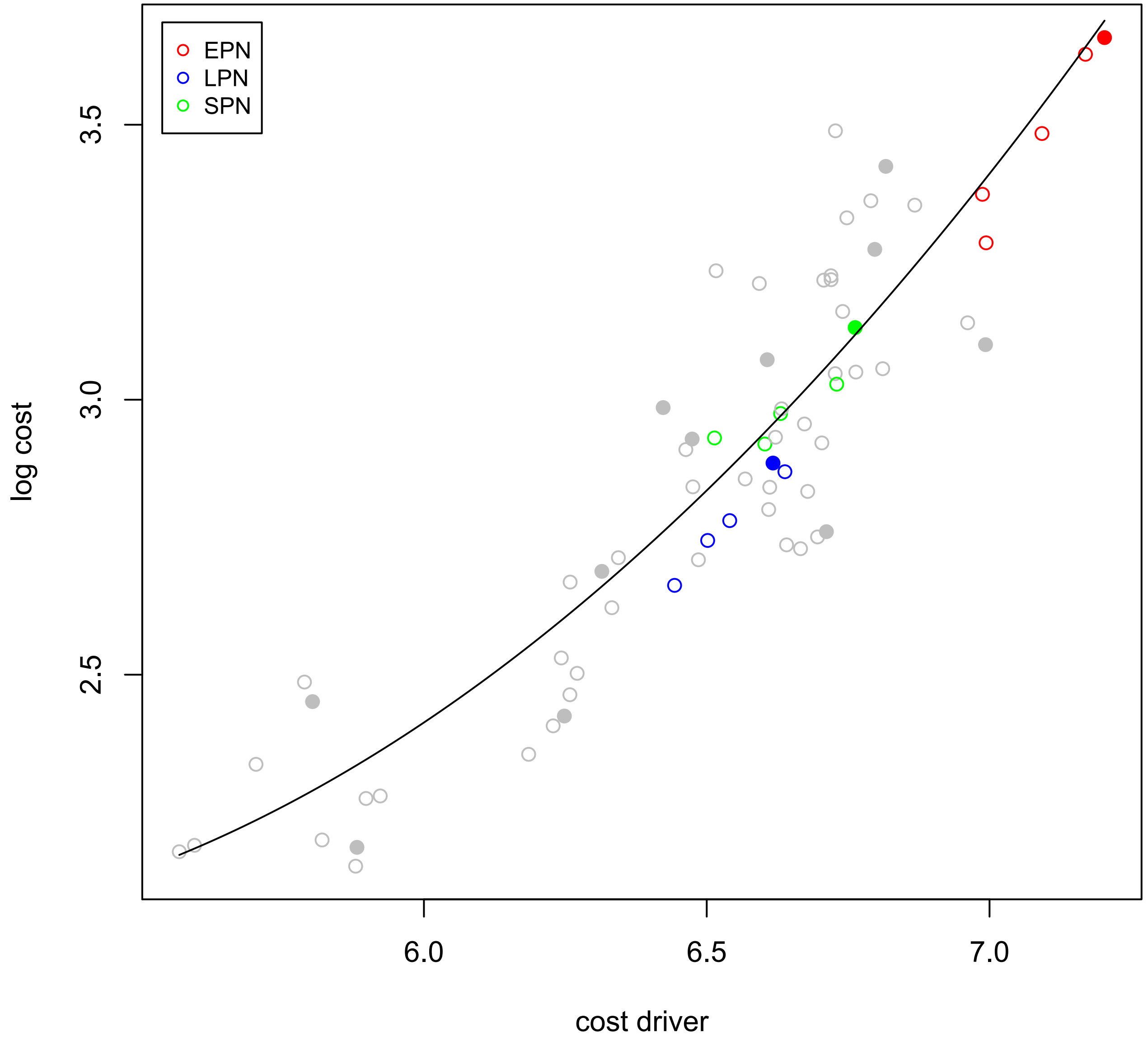


Figure 5: Indirect costs group 3

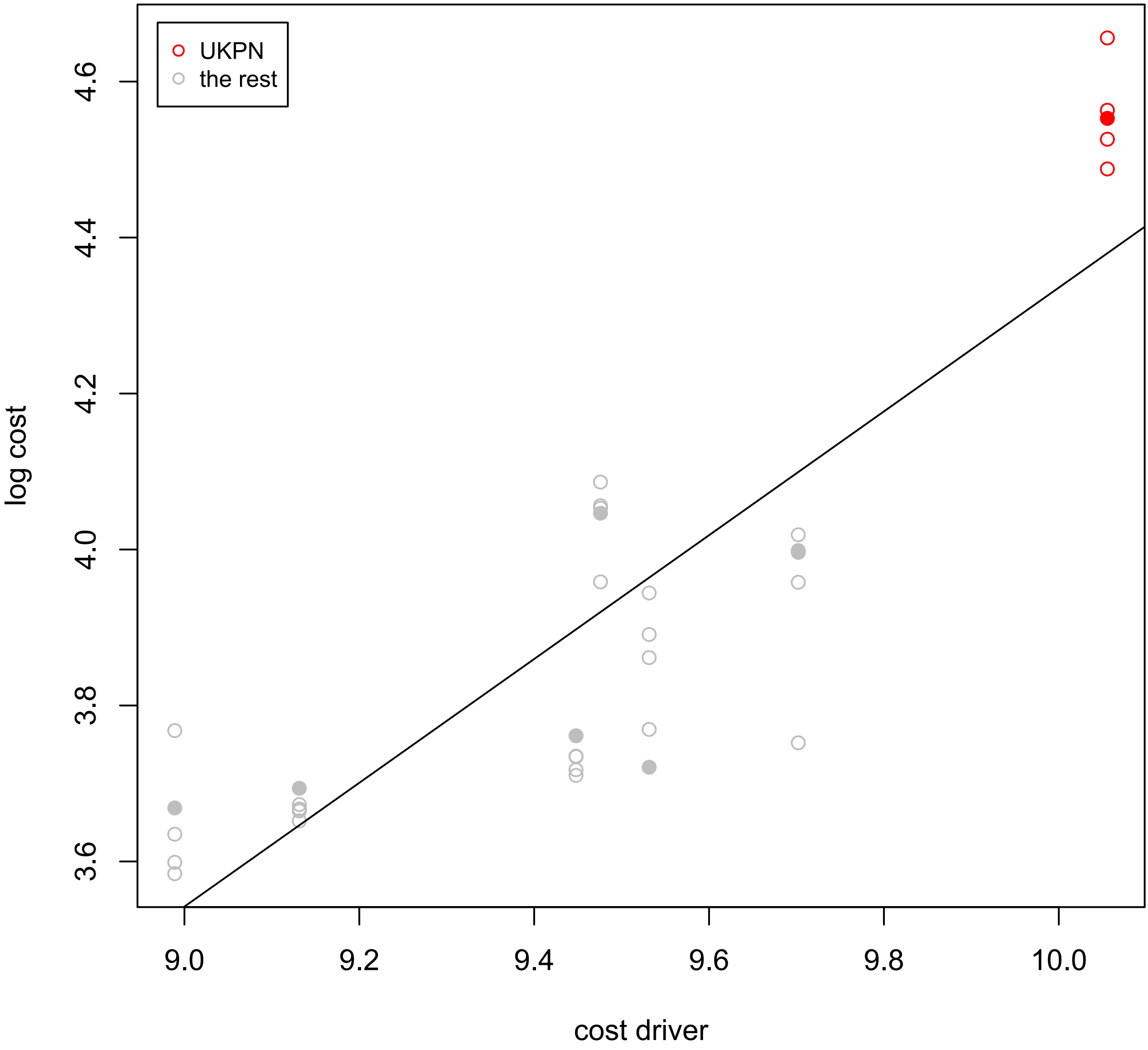


Figure 5a: Indirect costs group 3

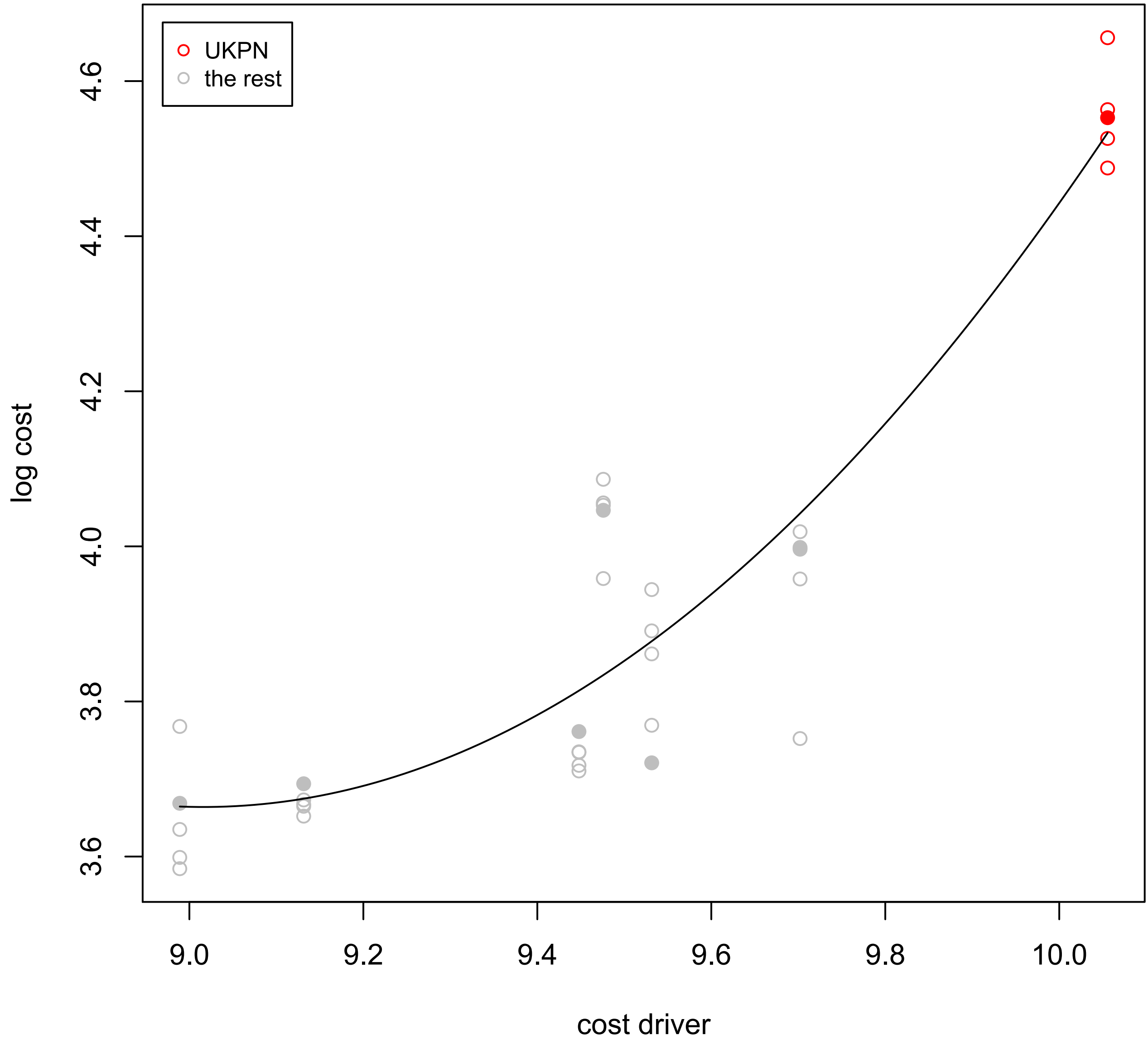
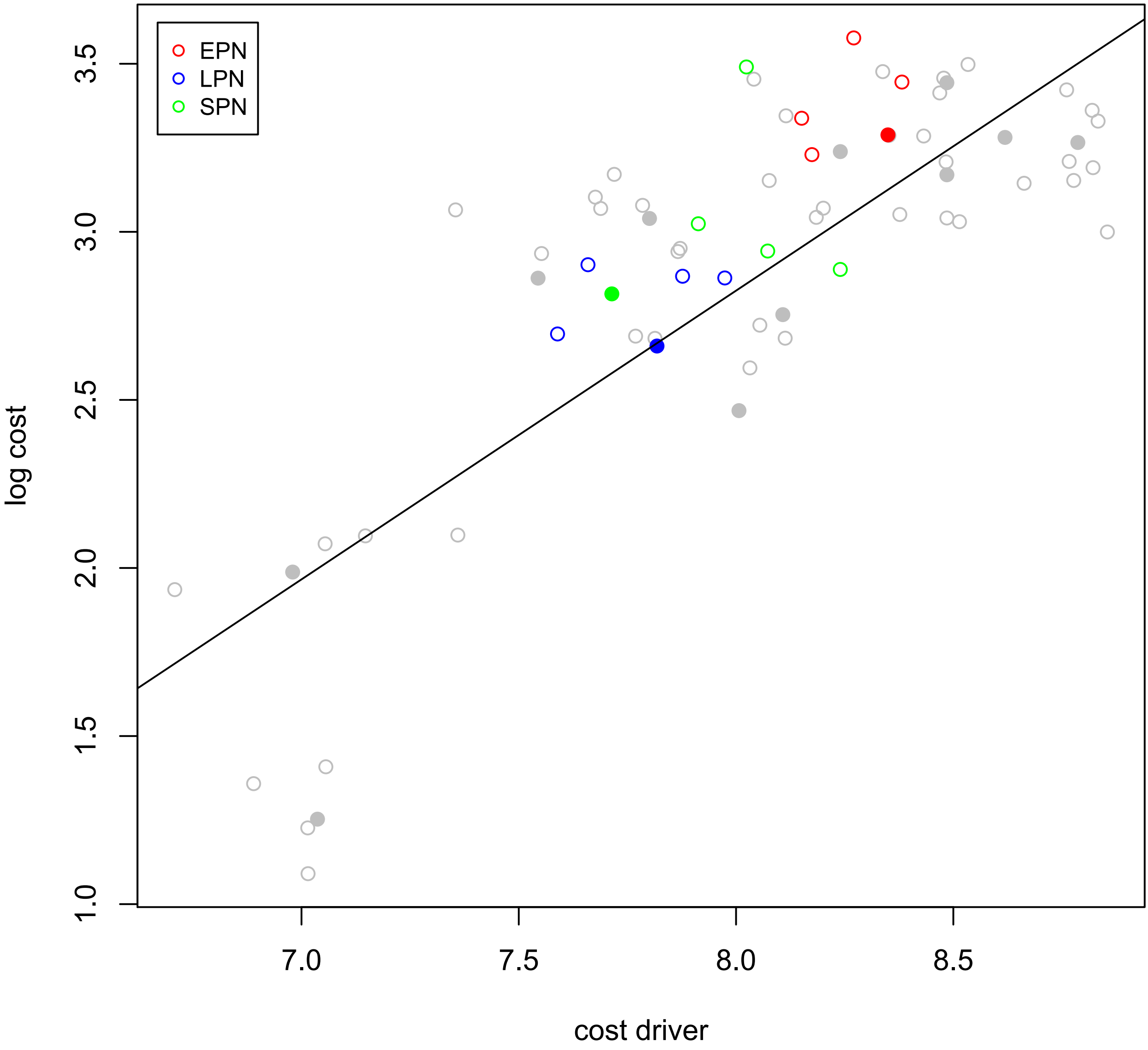


Figure 6: Underground Faults Costs



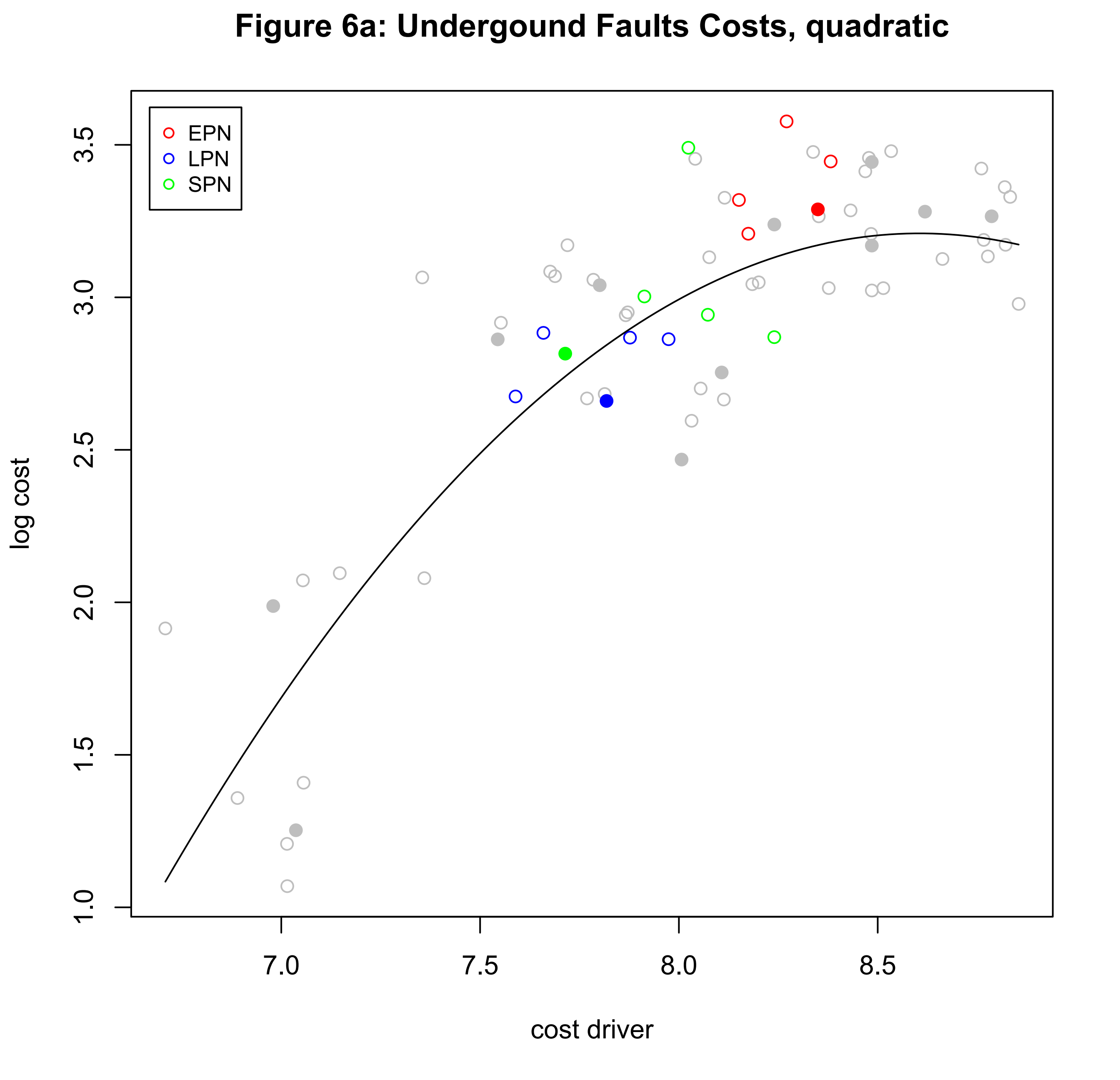


Figure 6a: Undergound Faults Costs, quadratic
EPN
LPN
SPN
log cost
cost driver
1.0
1.5
2.0
2.5
3.0
3.5
7.0
7.5
8.0
8.5

Figure 7: Overhead Faults Costs

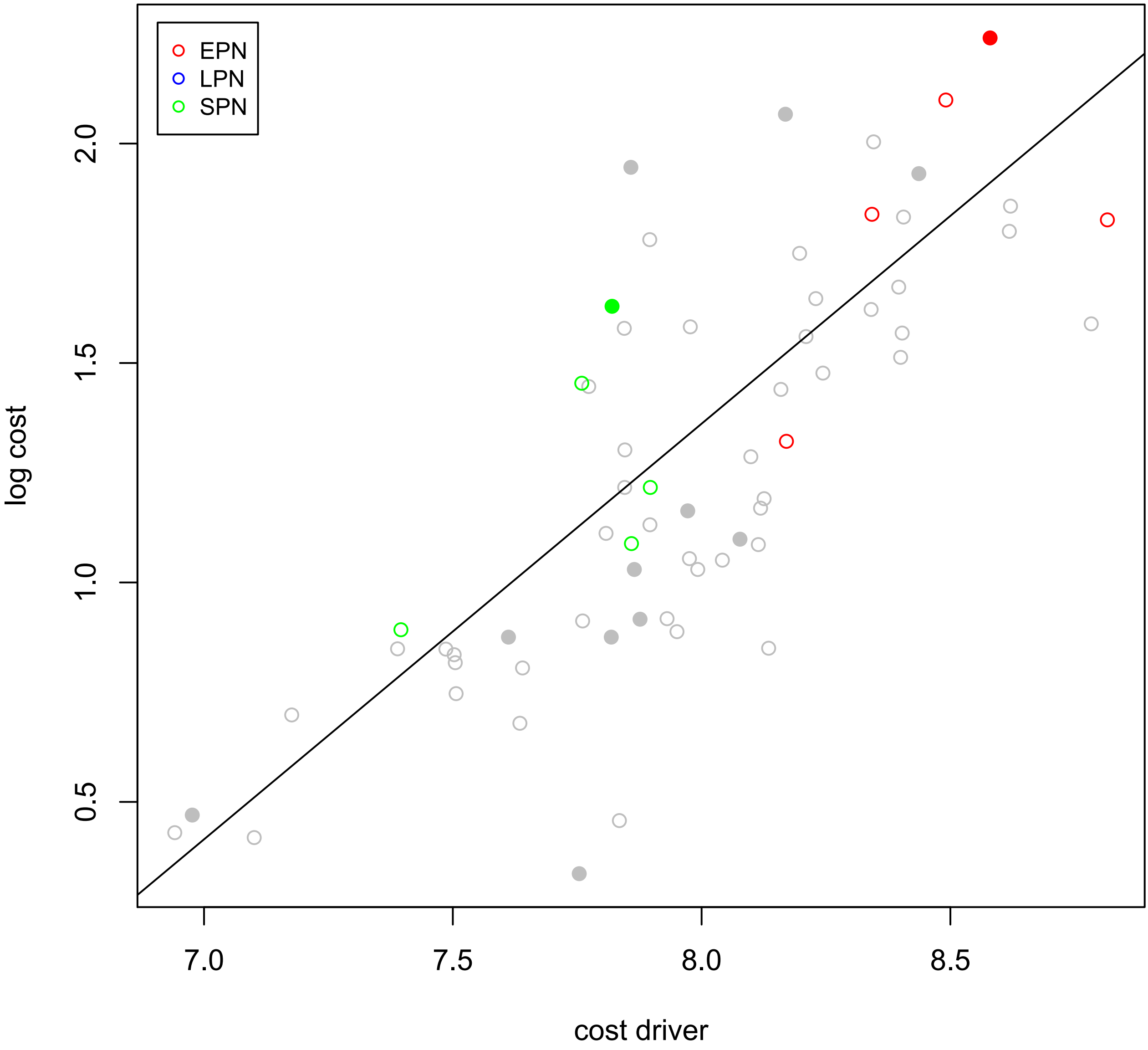


Figure 8: Tree Cutting Costs

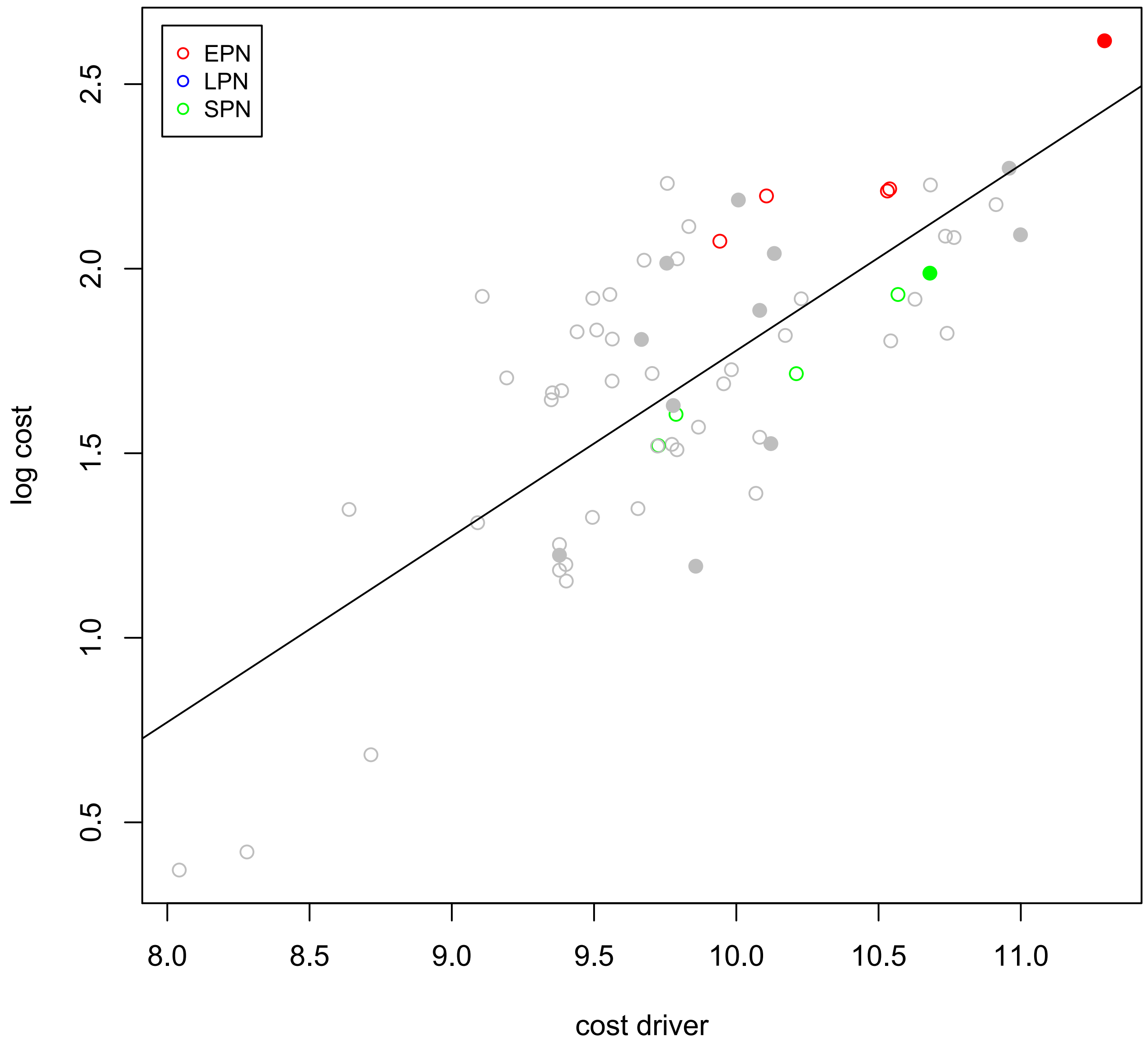


Figure 8a: Tree Cutting Costs, quadratic

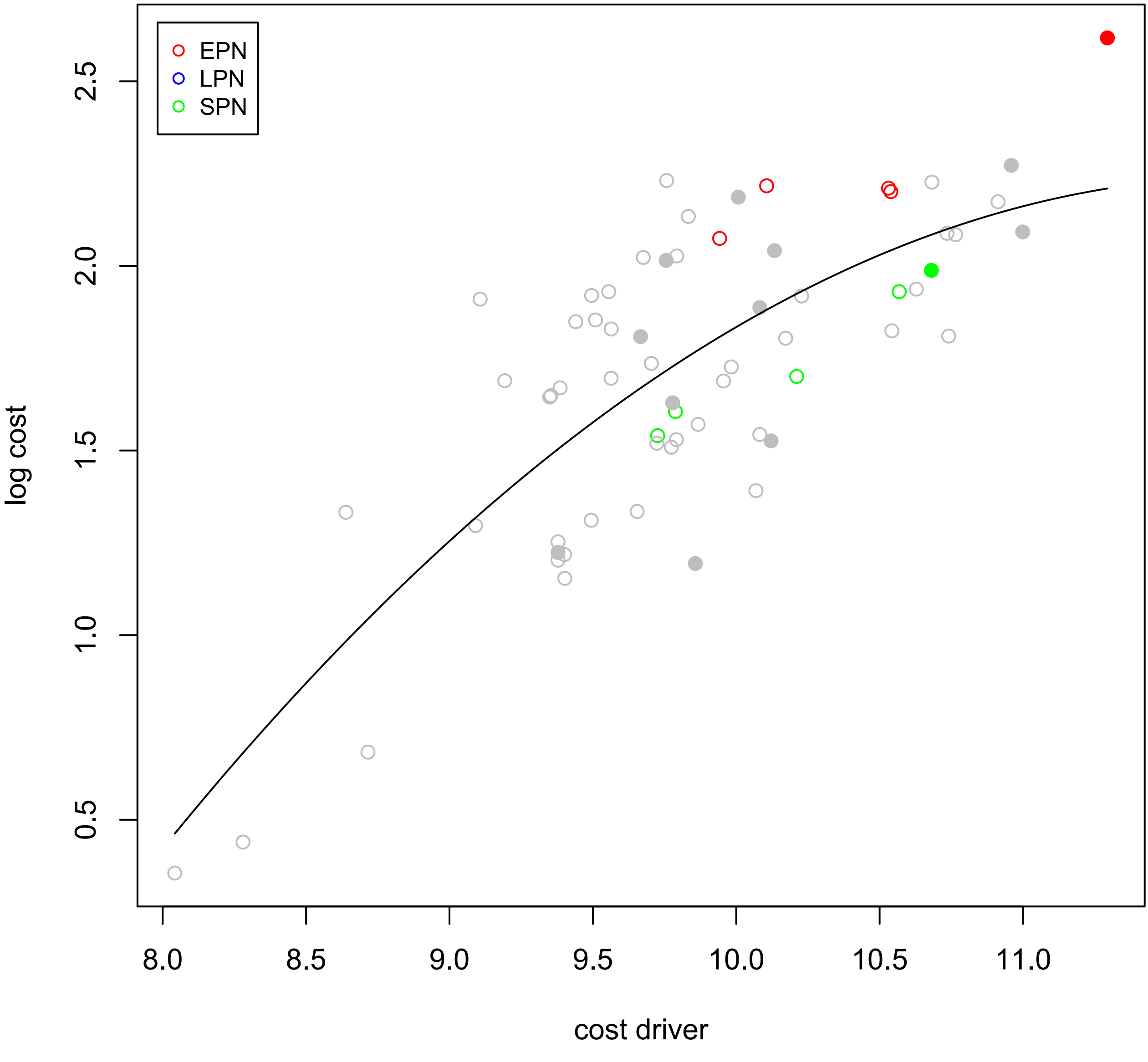


Figure 9: Inspection and Maintenance Costs

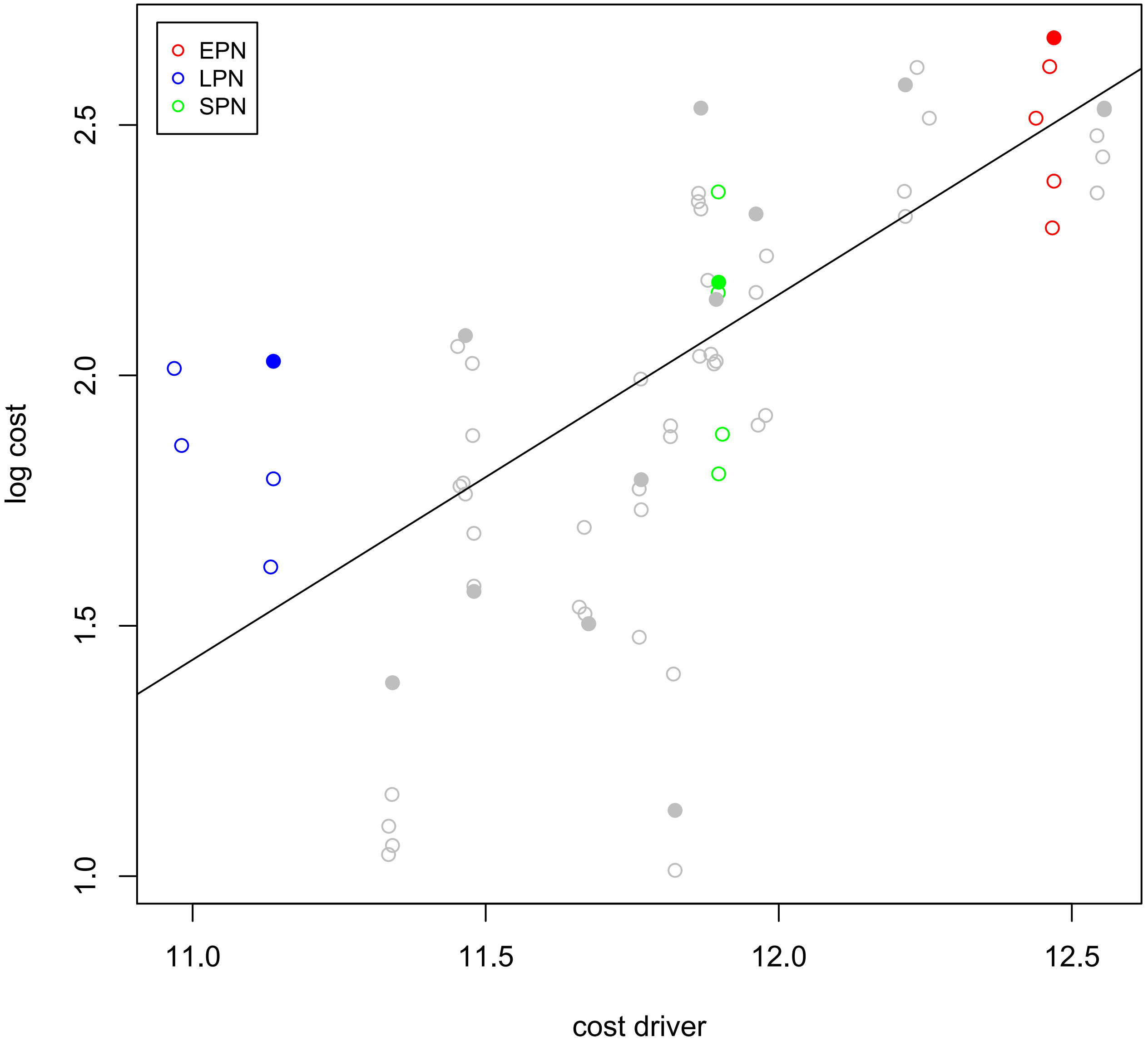


Figure 9a: Inspection and Maintenance Costs, quadratic