Appendix C

Operating Costs for the Access Network in Ireland: an Econometric Approach

Europe Economics
Chancery House
53-64 Chancery Lane
London WC2A 1QU
Tel: (+44) (0) 20 7831 4717
Fax: (+44) (0) 20 7831 4515
www.europe-economics.com

27 February 2004

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Europe Economics has been commissioned by ComReg to provide expert economic advice on specific issues within the process of setting wholesale access charges to eircom's local loop.

1.2 This paper considers how evidence from the US Local Exchange Carriers (LECs) can be used to help estimate an efficient level of operating costs for the provision of LLU services in the Republic of Ireland.

1.3 The aim of the exercise described in this paper is to make the best use of the available data from other operators (in particular the US LECs) in estimating operating costs. This is done by analysing separately the cost categories included in the eircom Bottom-up LRIC model in the broader class Direct Opex, i.e. the operating costs associated with network capital (Direct Network operating costs) and the operating costs associated with non-network capital (Direct Non-network operating costs), as well as Indirect operating costs and non-specific operating costs.

1.4 Section 2 of this paper describes the data that have been used to estimate the econometric model. Section 3 explains the model and the econometrics behind the model. Section 4 explains how the model has been applied to Ireland. Annex 1 describes the model specification and some technical aspects of the analysis. Annex 2 describes the derivation of the exchange rate used to convert the dollar values to euros.

2 THE DATA

Selected Comparators

2.1 The main data source for this exercise is the Automated Reporting Management Information System (ARMIS), a database initiated in 1987 by the Federal Communications Commission (FCC), the US Telecommunications Regulator, with the aim of collecting financial and operational data from the largest US carriers.[1] Today the ARMIS database consists of ten public reports.

2.2 Within this information system, detailed data on operating costs are available for all the 31 "Large Size" LECs. However, some of these companies operate in more than one State and some of the ARMIS reports do not list data on a State-by-State basis, whereas others do. For example, BellSouth Corporation operates in nine different States. In some ARMIS Reports, e.g. Report 43-05 (which we have used as a source for the variable *Metropolitan*, see below), data for BellSouth Corporation are provided for each sub-company operating in each separate State; in other Reports, e.g. Report 43-08 (which we have used as a source for the variable *Number of Lines*, see below), data are aggregated and provided for the whole Corporation.

2.3 We have therefore split the 31 "Large Size" LECs into two groups:

(a) The companies that operate in one State only. For these companies, the data are provided consistently over all the ARMIS reports.

(b) The companies that operate in more than one State. For these companies, weights are needed to average the values taken by some of the variables (e.g. share of customers in residential areas as opposed to metropolitan areas, see below) over the sub-company operating in different States.

2.4 However not all 31 companies report data for the whole sample period from 1991 to 2002. For this reason, seven companies were excluded from the sample. Moreover, we have also excluded Verizon Southwest, GTSW, on the basis of inconsistent accounting reports due to the merger activity between the telecommunication operators that led to the creation of this company. The 23 companies included in the sample are listed in Table 2.1.

---

[1] Additional ARMIS reports were added in 1991 to collect service quality and network infrastructure information from local exchange carriers subject to price cap regulations, in 1992 for the collection of statistical data formerly included in Form M, and in 1995 for monitoring video dial tone investment, expense and revenue data.

Table 2.1: List of the 23 US LECs included in our sample

No.	COSA index*	Name of the Company	State of operation
1	SWTR	Southwestern Bell Telephone	Several
2	PTCA	Pacific Bell – California**	California
3	PTNV	Nevada Bell	Nevada
5	LBIL	Illinois Bell	Illinois
6	NBIN	Indiana Bell	Indiana
7	MBMI	Michigan Bell	Michigan
8	OBOH	Ohio Bell	Ohio
9	WTWI	Wisconsin Bell	Wisconsin
10	CDDC	Verizon Washington D.C.	Washington D.C.
11	CMMD	Verizon Maryland	Maryland
12	CVVA	Verizon Virginia	Virginia
13	CWWV	Verizon West Virginia	West Virginia
14	DSDE	Verizon Delaware	Delaware
15	PAPA	Verizon Pennsylvania	Pennsylvania
16	NJNJ	Verizon New Jersey	New Jersey
17	NETC	Verizon New England	Several
18	NYNY	Verizon New York Telephone**	New York
19	GTFL	Verizon Florida	Florida
20	GTHI	Verizon Hawaii	Hawaii
21	GTMW	Verizon North	Several
22	GTNW	Verizon Northwest	Several
23	GTSO	Verizon South	Several

* *Index used to identify these companies in the ARMIS database*

** *These two companies also operate in other States, but at least 98 per cent of the lines they operate are in the State shown in the table.*

2.5 For those companies that operate in more than one State, in order to average the values taken by some of the variables, we have used as weights the number of all lines over different States (as of in year 2000).[2] Table 2.2 shows the States of operation and the weights used for the companies with substantial operations in more than one state.

2.6 We have reviewed whether the merger activity between telecommunication operators that has taken place in the US during the period under scrutiny in this exercise has affected the consistency of the reported accounts over the years for the companies selected in our sample and concluded that, apart from Verizon Southwest, it has not, as the merged entities remained separate filing companies for the purposes of the ARMIS database.

---

[2] Table III of the FCC Report 43-08 (the ARMIS Operating Data report)

Table 2.2: The weights attached to each state for multi-state LECs

State	SWTR	NETC	GTMW	GTNW	GTSO
Alabama	0	0	0	0	0.0835
Arkansas	0.0670	0	0	0	0
California	0	0	0	0.0101	0
Idaho	0	0	0	0.0882	0
Illinois	0	0	0.1782	0	0.0184
Indiana	0	0	0.2188	0	0
Kansas	0.0888	0	0	0	0
Kentucky	0	0	0	0	0.2836
Maine	0	0.1035	0	0	0
Massachusetts	0	0.6403	0	0	0
Michigan	0	0	0.1614	0	0
Missouri	0.1665	0	0	0	0
New Hampshire	0	0.1128	0	0	0
North Carolina	0	0	0	0	0.1856
Ohio	0	0	0.2048	0	0
Oklahoma	0.1061	0	0	0	0
Oregon	0	0	0	0.3167	0
Pennsylvania	0	0	0.1492	0	0
Rhode Island	0	0.0926	0	0	0
South Carolina	0	0	0	0	0.1101
Texas	0.5716	0	0	0	0
Vermont	0	0.0509	0	0	0
Virginia	0	0	0	0	0.3189
Washington	0	0	0	0.5849	0
Wisconsin	0	0	0.0874	0	0
Total	1	1	1	1	1

Source: FCC Report 43-08, Table II, column cj (total switched access lines)

Operating Expenditure

2.7 For direct network operating costs, in each of these companies we have selected as dependent variable for this exercise (i.e. the variable whose behaviour we are trying to explain) the costs incurred in operating what, in the LECs' accounts, is defined as *Cable and Wire Facilities Expenses*.[3] We have retrieved these variables from the ARMIS database, specifically from Table I-1 (Income Statement Accounts) of the ARMIS Report

---

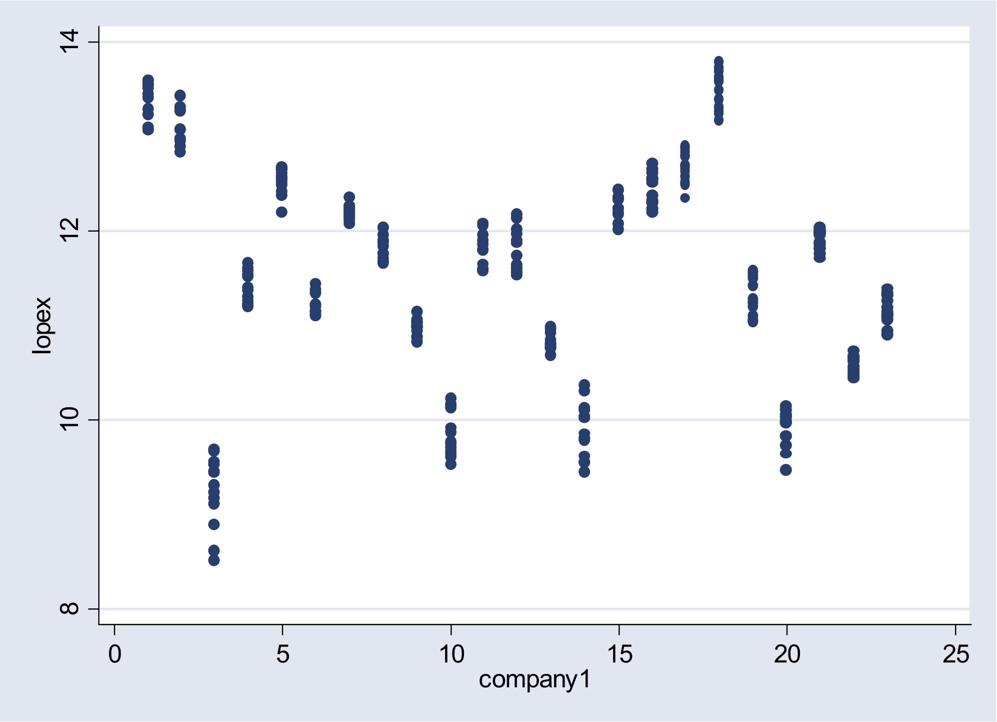
[3] This corresponds to the account category number 6410 of the Statistics of Communications Common Carriers, Federal Communications Commission, Washington, DC.

43-02 (the USOA Report).[4] The period for which these data are available is 1990 – 2002. The data refer to annual operating costs.

2.8 This variable is given by the sum of the operating costs, excluding depreciation,[5] directly associated to the asset categories that make up the infrastructure of the access network (i.e. underground and overhead drop cable, underground and overhead distribution and feeder cable, poles, duct and manholes). These categories constitute the bulk of the network assets in an access network.

2.9 The following figure shows the natural logarithms of the operating costs (in thousands US dollars) for the different companies. The 23 companies are on the horizontal axis and for each company the different dots refer to the value the variable takes over the years included in the sample.[6]

Figure 1: Direct Network Operating Costs (in logs) by Company[7]



---

[4] http://svartifoss2.fcc.gov/eafs/PresetMenu.cfm

[5] Some accounting practices include asset depreciation as part of their operating expenditure. In the ARMIS database, depreciation is not treated as such.

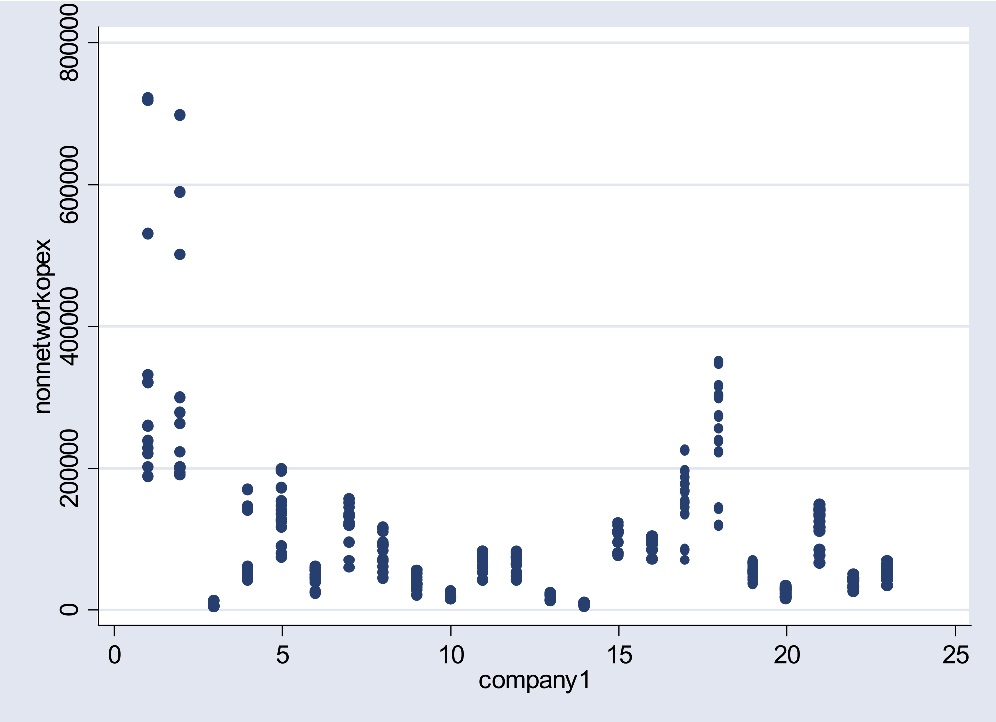
[6] The usage of the natural logarithm function is common in econometric exercises of this nature not only because logarithms smooth the difference between the values of the variable under consideration but also because they allow for an easier interpretation of the results (i.e. percentage changes of the dependent variable, see below).

[7] In order to work out for each Company the level of operating costs in US dollars, the natural number (i.e. 2.718) should be powered to the figure indicated on the vertical axis of the chart and the results multiplied by 1,000 (this is because the variable is originally expressed in US $000).

2.10 As can be seen from the chart, the amount of direct network operating costs varies substantially among companies, as one would expect since this variable does not take into account the size of the companies under consideration (see below).

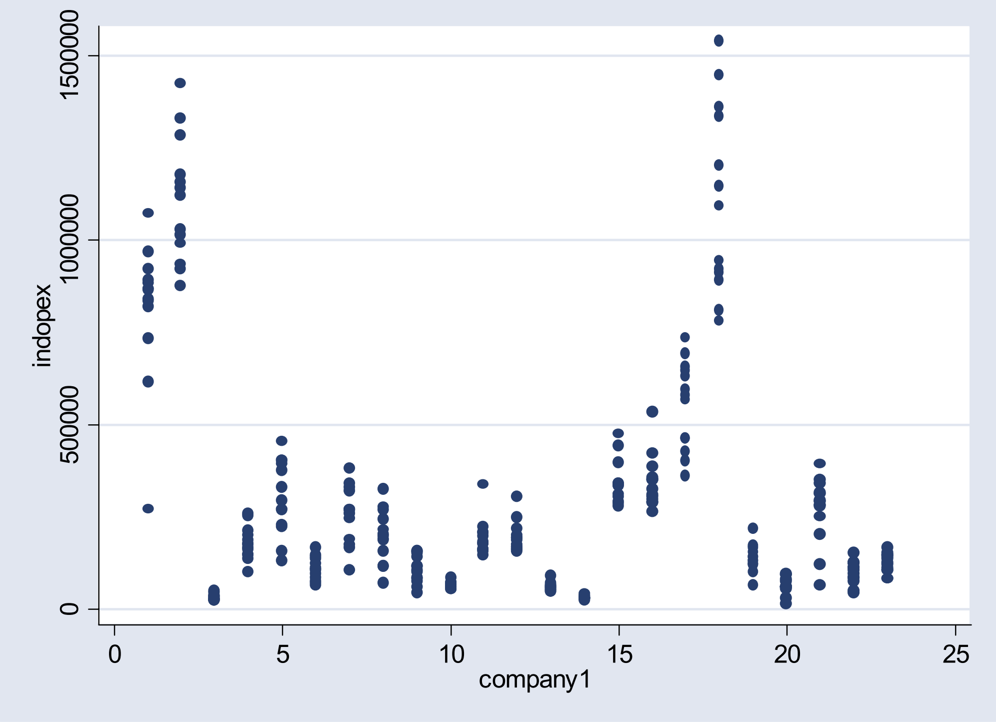
2.11 The cost category used for the Direct Non-network Operating expenditure refers to the LECs account numbers 6110 and 6120: i.e. Network support expenses and General support expenses. This cost category includes the rental cost of buildings and land, LEC account number 6121, which are estimated from Irish data. The non-network operating costs considered for this exercise correspond therefore to the following LEC account categories: 6110 + 6120 – 6121.

Figure 2: Direct Non-network Operating Costs by Company



2.12 The indirect operating costs correspond to the LEC account number 710: Total Corporate Operations expenses. This includes, but is not limited to, cost categories like executive pay, accounting and finance, human resources and external relations.

Figure 3: Indirect Operating Costs by Company

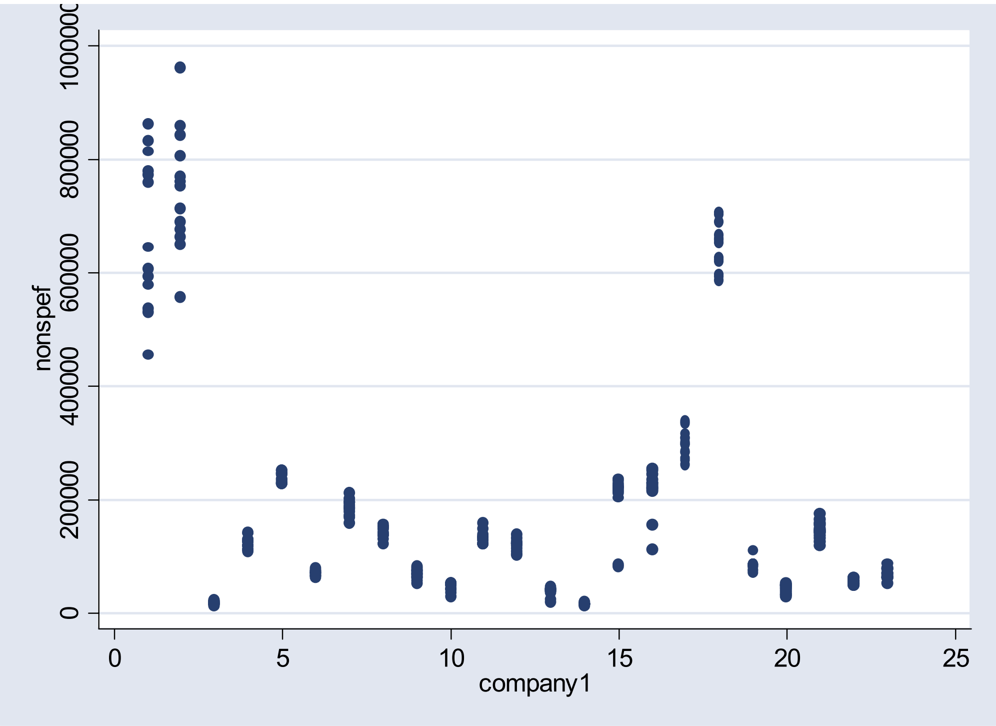


2.13 The non-specific operating costs refer to the LEC accounts 6530, which include testing, engineering and plant operations administration expenses.[8]

---

[8] Category 6510: "Other property plant and equipment expense" was not included due reporting anomalies by various companies. It is on average less than 1 percent of the amount of category 6530.

Figure 4: Non-specific Operating Costs by Company



2.14 Table 2.3 presents a summary of the cost categories and the LEC account numbers to which they correspond.

Table 2.3: Selected Operating Costs Categories for Large LECs

Operating Cost Category	Account number	Operating expenses for Year ending 31 December 2000. (US$m)
Direct Access Network Operating Costs	6410	7,236
Of which: operating costs associated with overhead, buried and underground (i.e. ducted) cable	*6421-6423*	*6,642*
Direct Non-network Operating Costs	6110, 6120 (excluding Land and Building), and 6530: this covers operating costs of assets used by different increments	11,843
Of which testing, engineering and plant operations admin. Expenses	*6530*	*6,347*
Of which land and buildings	*6121*	*2,029*
Of which general purpose computers and other assets	*6110+6120-6121*	*3,467*
Indirect operating costs	710: this is for the companies as a whole	7,978

Source: Table 2.10 of the Statistics of Communications Common Carriers for year 2000.

Explanatory Variables

2.15 The operating activities/interventions that give rise to the operating expenditure of a telecom operator could be affected by a number of different variables. We have classified these possible explanatory variables/cost drivers into three groups:

(a) **plant** variables: technical characteristics of the network under consideration; for example, cable strung over poles might imply an higher level of these operating activities than cable buried underground;

(b) **demographic** variables: demand features of the network under consideration; for example, customers in urban areas might give rise to more of these interventions than customers in metropolitan areas;

(c) **meteorological** variables: forces of nature to which the network under consideration is subject; for example, more extreme environmental conditions might give rise to more of these operating activities than a network running in a milder environment.

2.16 In the next sections we describe, for each group, which specific variables we have considered as possible explanatory variables. Although these variables might be considered as having an impact mostly on the direct network operating costs, they have been tested on all the operating cost categories identified in the Consultation Document.

Plant Variables

2.17 An important explanatory variable is likely to be the total number of lines served by the operator in question.

2.18 We have also tried a number of other plant variables that could be relevant in explaining the characteristics of the network in question and the level of operating expenditure. Such variables include length of cable, length of trench, number of poles, size of cable (number of copper pairs) and duct (number of bores). Among these, the ones that turn out, from our analysis, to be able to better explain the level of operating expenditure are cable length per line, trench length per line, aerial sheath share and ducted cable share. We have also included as plant variables a quality index and a time trend. We now describe each of these variables in more detail.

Total number of lines

2.19 For the US LECs there are three main types of lines:

(a) switched access lines (including payphones lines);

(b) special access lines (these are dedicated lines from the customer to the inter-exchange carrier point of presence); and

(c) leased lines.

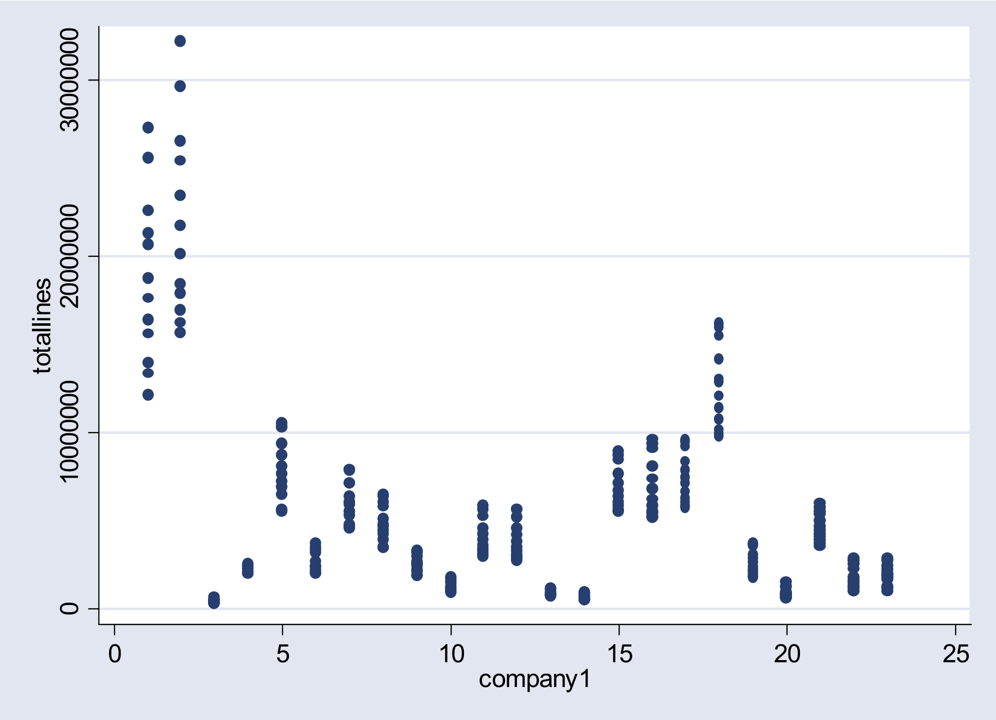
2.20 We believe that it would be appropriate to use volume measures of all three types of lines. However, volume figures for leased lines are only available for Year 2002 (column *fm* of Table III of the FCC Report 43-08). We have therefore decided not to include the number of leased lines as an explanatory variable (i.e. the variable Total Number of Lines has been constructed adding up the number of Switched and Special Access Lines).[9]

2.21 Apart from the impact on the accuracy of the results deriving from the omission of a potentially explanatory variable (see below), the exclusion of leased lines might have an impact when the results of this exercise are applied to the Irish case. This issue is discussed in Section 4.

---

[9] A possible proxy for a pure volume based measure of leased lines could have been obtained by dividing revenues by product over the corresponding price. However, in the ARMIS database, revenue data are provided at a level which is too aggregated to make any sensible inference about the corresponding volumes.

Figure 5: Total Number Lines (Switched and Special Access) by Company



2.22 As is clear from this chart, the size of the different companies included in the sample varies considerably. Clearly, the sign of the coefficient that this variable (i.e. total number of lines, switched and special) is expected to take is positive; the more the number of lines, the more the costs that are likely to be incurred.

2.23 As to switched access lines, data are available for residential and business lines and metropolitan and non-metropolitan lines. We have considered and tested for the significance of both ratios (i.e. residential over total and metropolitan over total) as possible explanatory variables (see section on demographic variables below).

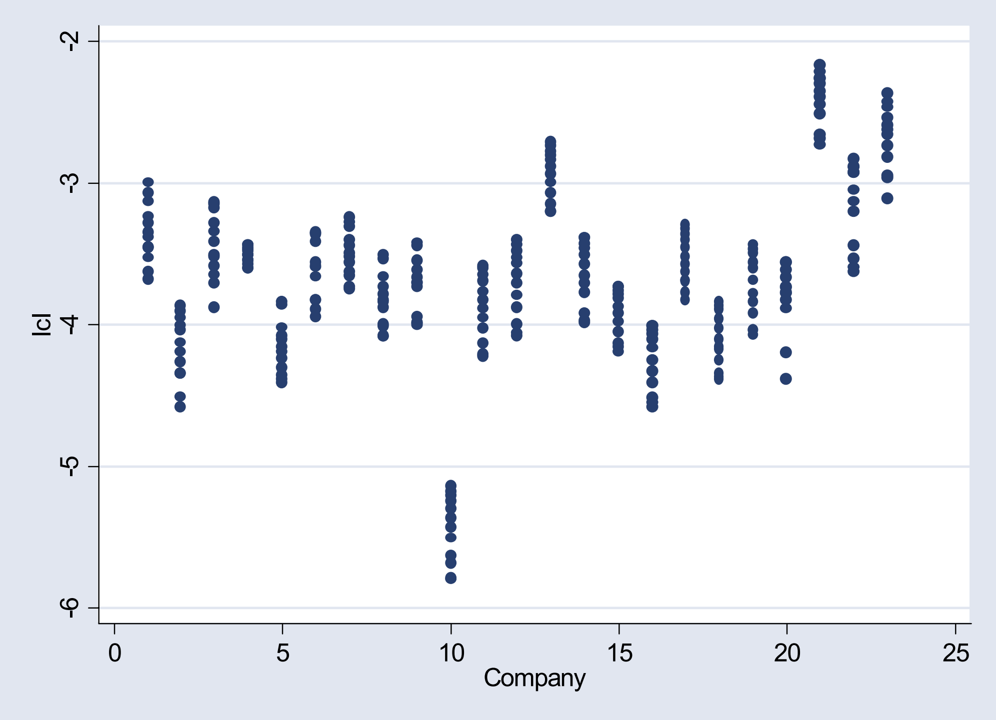
2.24 The source for number of lines is the FCC Report 43-08 (the ARMIS Operating Data Report), specifically column *fl* of Table III (Access Lines in Service by Customer). The data are available from the period 1991-2002.

Cable length per line

2.25 This variable has been constructed as total length of copper cable (kilometres) divided by the number of lines. Total copper cable is the sum of aerial, underground (split in ducted and buried) and intra-building network cable. The source of this variable is the FCC Report 43-08, specifically column *p* of Table I.A (Outside Plant Statistics - Cable and Wire Facilities).

2.26 This variable captures an important aspect of the configuration of the access network, with higher values indicating bigger distances between the customer premises and the concentrator (which would in turn imply a higher likelihood of line faults) or less amount of cable sharing between different lines (implying a lower level of customer density); the sign of the corresponding coefficient is therefore expected to be positive. The period for which these data are available is 1991 – 2002.

Figure 6: Length of cable per line (in logs) by Company



2.27 Company 10, the company operating in Washington, District of Columbia, displays the shortest cable length per line, as might be expected.

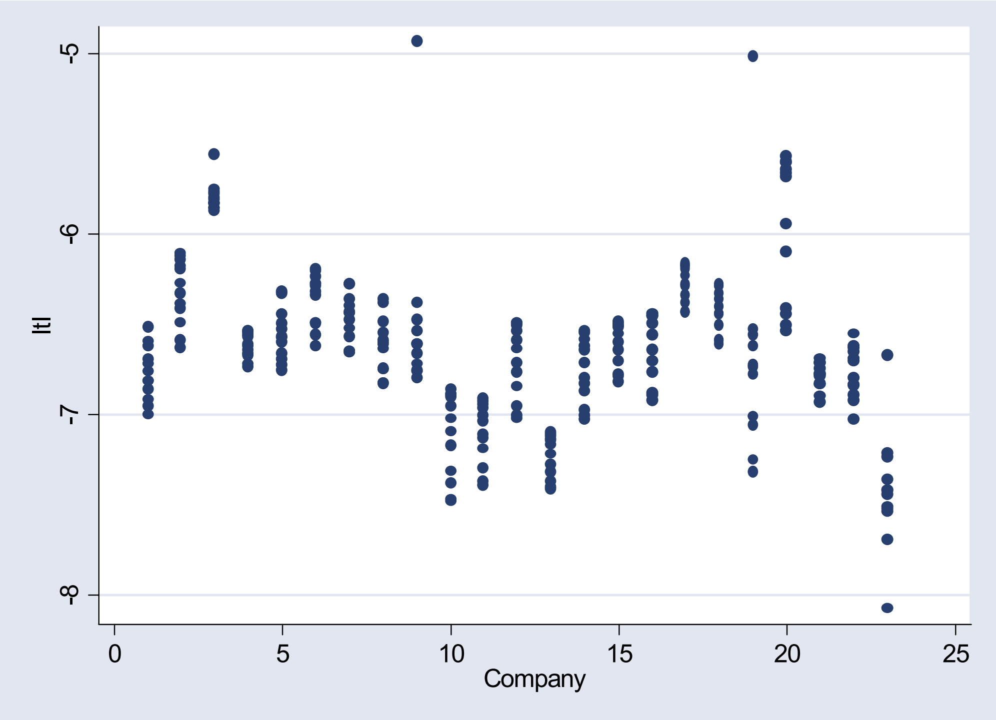
Trench length per line

2.28 This variable has been constructed as total length of trench (kilometres) divided by the number of lines. The source of this variable is the FCC Report 43-08, specifically column *w* of Table I.A (Outside Plant Statistics - Cable and Wire Facilities).

2.29 The bigger this variable, the more the dispersion of the customers around the country. The sign of the coefficient of this variable is therefore expected to be positive.

2.30 Whereas the previous variable would be influenced by the sharing of cable sheath by different lines this variable is more likely to capture sharing of trench by different cable sheath. The period for which these data are available is 1991 – 2002.

Figure 7: Length of trench per line (in logs) by Company



2.31 Although the outliers for Companies 9 and 19 could be considered errors in the dataset and therefore reasonably excluded from it, subject to any further information becoming available we have however decided not to correct for any of these observations.

Aerial share

2.32 This variable refers to the share of length of cable that is strung on poles, as opposed to being underground. Whereas the capital cost of laying cable in the ground is bigger than the cost incurred in hanging the equivalent length of cable on poles, maintenance costs are likely to be lower for underground cable.

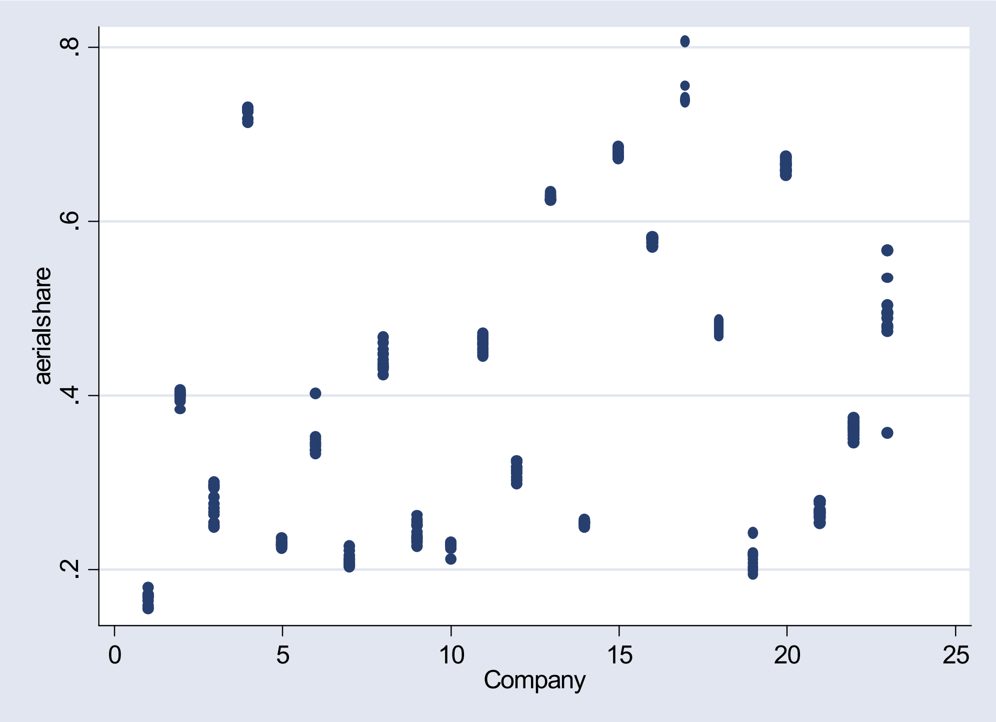
2.33 This variable has been constructed by dividing the length of cable on poles (aerial sheath kilometres of metallic) by total length of cable in the network (total sheath kilometres of metallic). These two variables correspond respectively to columns *d* and *p* of Table I.A. (Outside Plant Statistics - Cable and Wire Facilities) of the FCC Report 43-08 (the ARMIS Operating Data Report).[10]

---

[10] Column *p,* i.e. Sheath kilometres of Total metallic Cable, is defined as the sum of the following columns: column *d,* i.e. Sheath Kilometres of Metallic Aerial Cable, column *f,* i.e. Sheath Kilometres of Metallic Underground Cable, column *h,* i.e. Sheath Kilometres of Metallic Buried Cable, column *n,* i.e. Sheath Kilometres of Metallic Intrabuilding Network Cable.

2.34 The higher this variable the higher the expected level of direct operating costs. The annual data are available from the period 1991-2002.

Figure 8: Proportion of length of cable sheath on poles by Company



2.35 As can clearly be seen from this chart, this variable is a network characteristic of the Company under consideration, but it does not vary much over time.

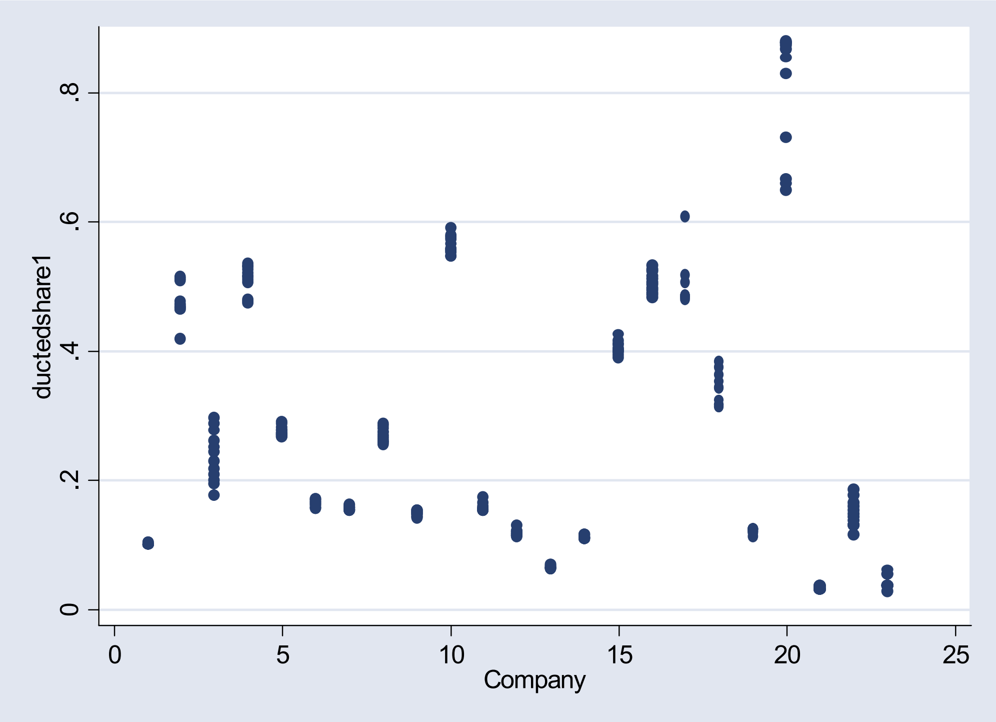
Duct share

2.36 This variable refers to the share of length of underground copper cable (constructed as the sum of ducted and buried copper cable length) that is ducted, as opposed to being simply buried. Whereas the capital costs of putting cable in ducts is bigger than the cost incurred in simply burying cables in terrain, maintenance costs are expected to be lower for ducted cable.

2.37 This variable has been constructed by dividing the length of ducted cable (sheath kilometres of metallic "underground" cable, as this is called in the ARMIS database) by total length of underground cable (constructed adding up "underground" cable length and buried cable length). The data are available in Columns d, f, h and p of Row 910 of Table I.A of the FCC Report 43-08, the ARMIS Operating Data Report.

2.38 The higher this variable the lower the expected level of direct operating costs (the sign of the coefficient is expected to be negative). The annual data are available from the period 1991-2002.

Figure 9: Share of ducted cable of total cable underground



2.39 Similarly to *Aerialshare*, this chart indicates that *Ductedshare* is a network characteristic of the Company under consideration, which for most companies does not vary much over time.

Quality

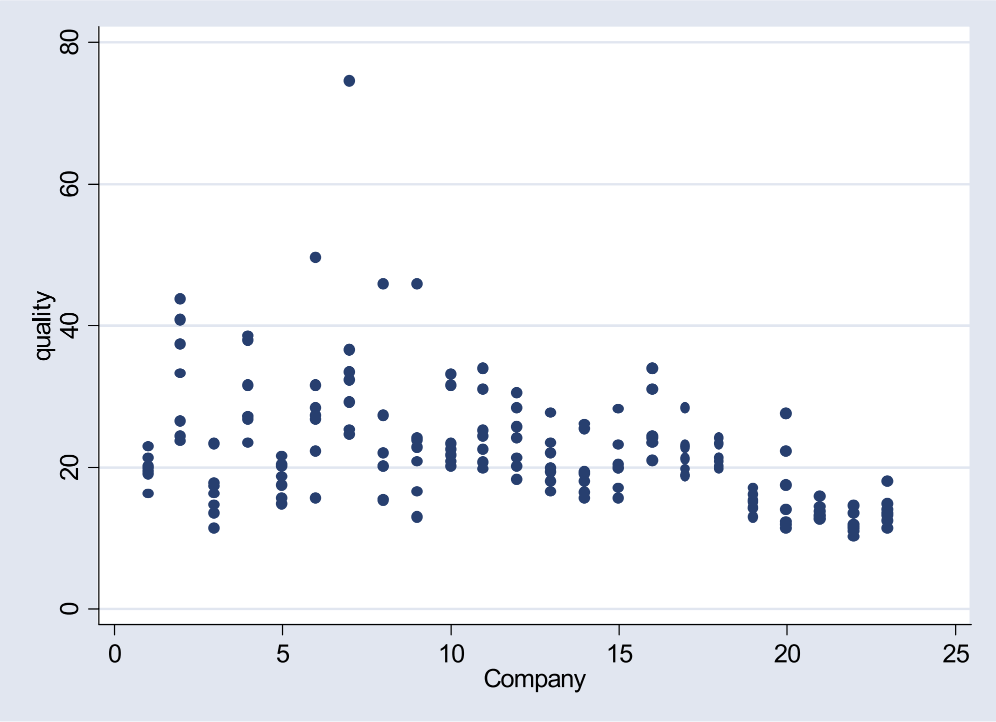
2.40 This variable refers to the quality of the service offered to end-users in terms of repair times when a fault is reported. The variable measures the interval (in hours) between the time a trouble report is received and the time the trouble report is cleared.

2.41 The source for this variable is Row 0145 of Table II (Installation and Repair Intervals, Local Service) of the FCC Report 43-05 (the ARMIS Service Quality Report).

2.42 Maintenance costs might be directly affected by this variable (sending an engineer to repair a fault in the least possible time might have an implication on the number of lines per employee). The sign of the coefficient of this variable is expected to be negative.

2.43 Since data for this variable are available on an annual basis only for the period 1996 – 2002, it was only possible to estimate models between those dates when the quality variable was included.[11]

Figure 10: Repair intervals (in hours) by Company



Time

2.44 We have also tested for the significance of a variable built as cardinal numbers increasing with the year under considerations (starting with 1 for 1990 and finishing with 13 for 2002) in order to take account of a possible time trend.

2.45 Given that possible inflationary pressures on operating costs would be picked up through *Wages*, if this variable is included as one of the explanatory variables, *Time* is meant to pick up any cost savings accruing on a yearly basis that are common across all companies due to, for example, possible efficiency improvements. The sign the variable *Time* would take in this case is expected to be negative.

2.46 If the variable *wages* were not to be included as a potentially explanatory variable, then the variable *Time* would pick up both an inflationary trend and the annual cost savings due to possible efficiency improvements. The sign of the coefficient of this variable would

---

[11] The variable was found consistently insignificant, therefore the issue of restricting the time span of the final model did not arise.

take in this case depends on the relative impact of inflationary pressure and efficiency improvements.

Demographic Variables

Wages

2.47 The level of wages is a variable directly relevant in explaining the level of operating costs. *Ceteris paribus* (including assuming the same level of efficiency), the higher the relevant level of wages in a given State, the higher the cost of operating the network in that State.

2.48 However, in order to find any significant relationship between labour and operating costs, it is important that wage costs refer to the right professional category. Fortunately, the national accounts in the US provide relevant data for the specific category "Telecommunications Equipment Installers and Repairers".[12] The source of these data is the Bureau of Labor Statistics, US Department of Labor.[13]

2.49 However, the wage rates for the category "Telecommunications Equipment Installers and Repairers" are only available, on an annual basis, for the period 1999 – 2001. We have, therefore, decided to use an appropriate index to have a proxy for the same variable for the period 1990 – 1998.

2.50 The index used for these purposes has been taken from Table 10 of the publication *Employment Cost Indexes, 1975-1999* of the Bureau of Labor Statistics and refers to wages and salaries of the *Professional specialty and technical occupations*.[14]

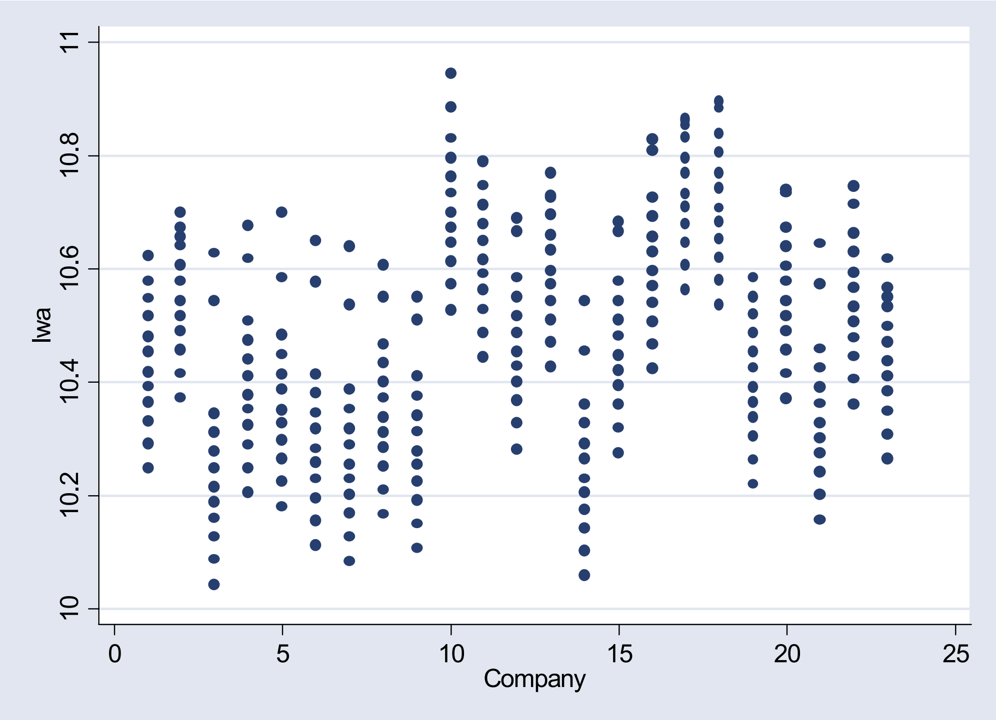
---

[12] This corresponds to category 49-2022 of the Occupational Employment Statistics (OES) Survey.

[13] Website: http://www.bls.gov/oes/2001/oes_dl.htm

[14] Website: http://www.bls.gov/ncs/ect/sp/ecbl0014.pdf.

Figure 11: Wages (in logs) by Company



2.51 This chart shows that, apart from the first few observations, the time pattern of this variable is very similar for all companies. This is hardly surprising, given the lack of data and the methodology used to proxy the missing values (the index that has been used to spread the data over the missing years is the same over all companies). This will have an implication for the results of the regression since this variable will turn out to be highly correlated with the variable *Time*, aimed to pick up both the inflationary trend and possible productivity improvements.

Metropolitan

2.52 The level of capital expenditure in telecoms access networks partly depends on the dispersion of customers. For a given number of lines, the more dispersed the end-users the greater the investment needed to connect these customers to the network. In order to increase the accuracy of costing exercises, therefore, the surface served by the network operator is usually split into geo-types, defined in terms of line density (i.e. number of lines per square kilometre).

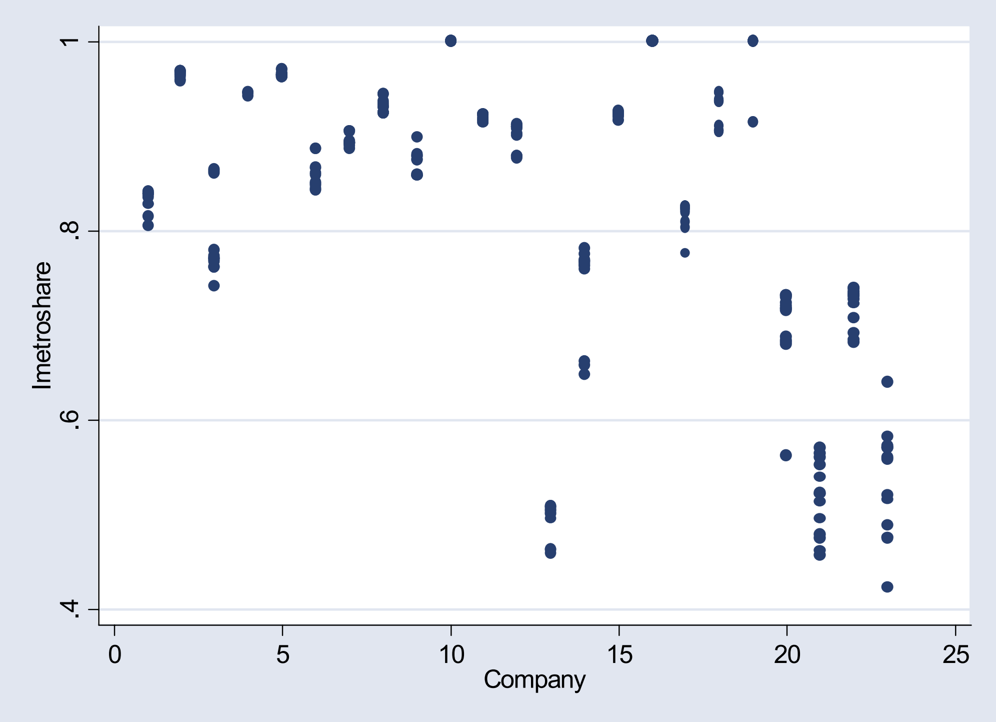
2.53 The dispersion of customers, however, might also play a role in the costs incurred in operating the network in question. Maintaining a given number of lines may be cheaper if lines are aggregated in urban areas, where the length of the loop is shorter, than in the countryside. Moreover, in urban areas connections to the network are likely to be quite stabilised, with customers moving around dwellings, as opposed to the countryside where new connections may create more troubles.

2.54 For all these reasons, we have considered the number of lines provided in metropolitan conurbations in relation to the total number of lines as a potential explanatory variable for the level of operating costs. Our expectation was that the higher this variable (the more the share of customers living in metropolitan areas) the lower the level of operating costs is likely to be.

2.55 The source for this information is the FCC Report 43-07. Specifically, column *b* (Within MSA) of Row 120 of Table I have used to give the total number of switched lines served in MSAs. This figure has then been divided by the total number of switched lines served by the operator in question, column *a* (Total Study Area), to obtain the variable under scrutiny.

2.56 Metropolitan Statistical Areas are here defined as including at least one city with a minimum population of 50,000 and its surrounding areas.

Figure 12: Proportion of customers living in metropolitan areas by Company



2.57 As turned out to be the case for the variable *Aerialshare* (see Figure 8), this variable characterises the Company under consideration, but it does not vary much over time.

Residential

2.58 The proportion of customers that are residential could be a relevant explanatory variable in so far as:

(a) business customers are usually located in more densely populated areas; and

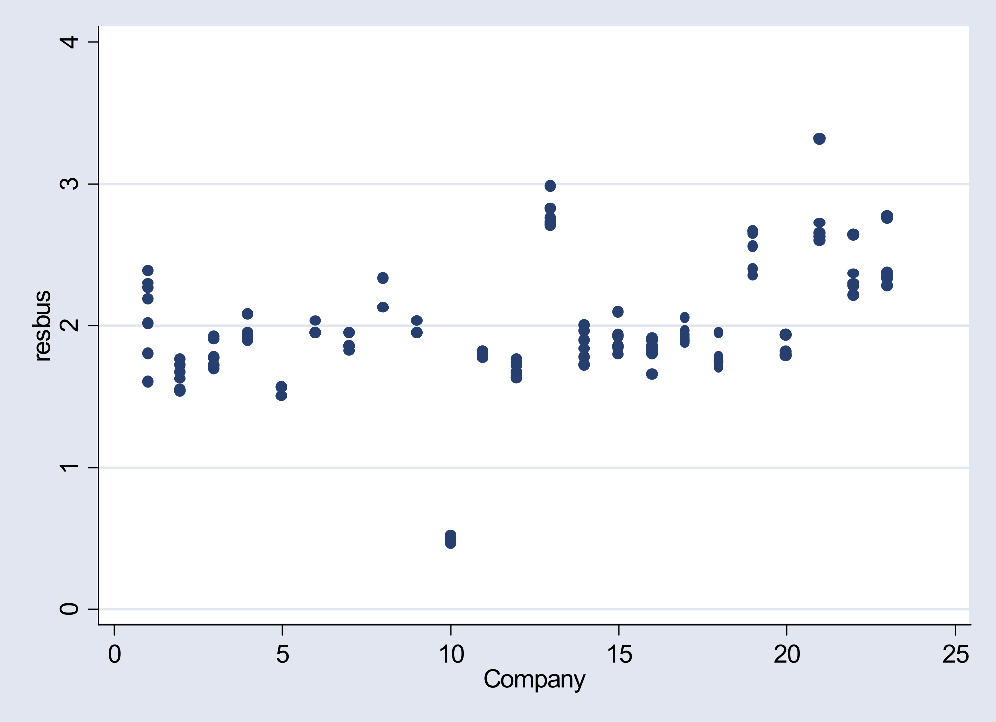
(b) business customers might require an higher level of assistance and maintenance than residential customers.

2.59 These two factors work in opposite directions. The first factor on its own implies a positive sign of the coefficient (the higher the share of residential customers, the lower the number of customers located in more densely populated areas, the higher the operating costs). The second factor implies a negative one (the higher the share of residential customers, the lower the number of high maintenance customers, the lower the operating costs). The econometric analysis would clarify which of the two effects, if either, dominates.

2.60 The source for this information is the FCC Report 43-05. Specifically, column *af* (total number of residential lines) of Row 140 of Table II has been divided by total number of business lines served by the operator in question, column *ai*, to obtain the variable under scrutiny.

2.61 Since data for this variable are available on an annual basis only for the period 1996 – 2002, the models were estimated only over these years when this variable was included. Although the outlier for Company number 10 might be considered to be an error in the dataset, and therefore excluded from it, we decided not to correct for it.

Figure 13: Proportion of residential customers by Company



Meteorological Variables

2.62 We believe that meteorological data might be relevant in explaining the level of direct operating expenditure. The access network is exposed to the force of the elements and metallic components may be subject to expansions and contractions due to warm and cold temperatures.

2.63 Moreover, factors like high winds, rain, snow and ice can have a major impact not only on the overhead/aerial share of lines that are more exposed, but also, though to a lesser extent, on underground cables. Excessive water can find poor joints and corroded cables, mainly in the distribution part of the loop generally not pressurised against water ingress. This would affect operating costs, including repair costs.

2.64 We have tried to capture these factors testing two variables here called: *Rain* and *Temperatures*.

2.65 We have calculated the average yearly amounts for these variables for each company including these either as time varying explanatory variables (to account for factors like especially cold or rainy years with more floods or ice breakage) or as the average of the values that these variables take over the period 1990 - 2000 and attached it as dummy variables for each company. The state shares in Table 3 were used for the companies with operations in more than one state.

2.66 We expect a positive correlation between these variables and the level of operating costs.

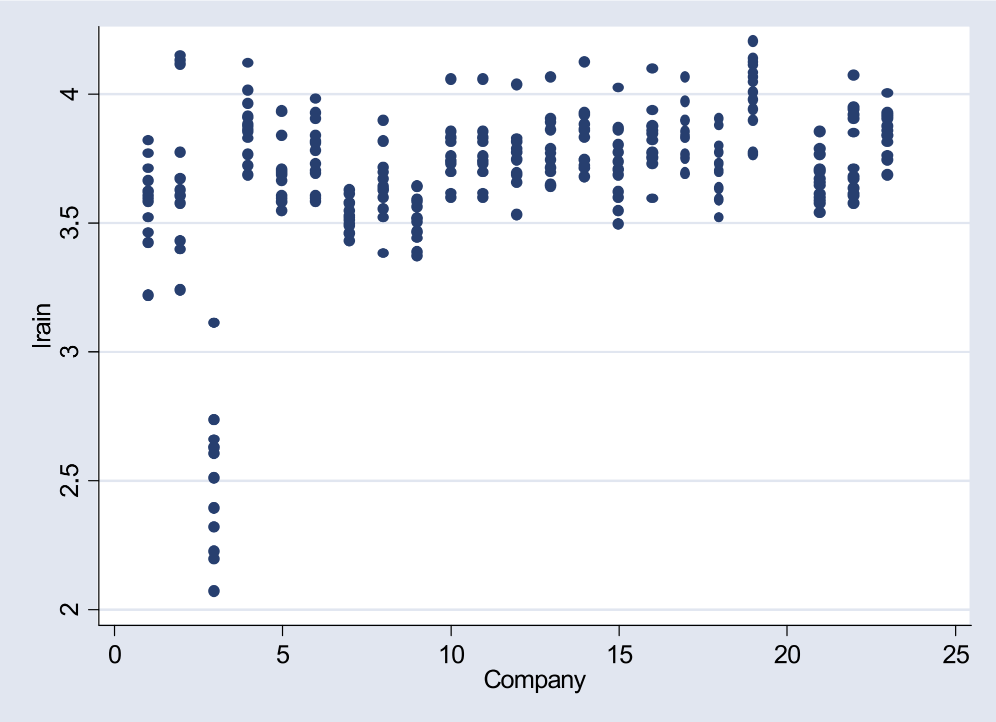
Rain

2.67 The precipitation data were obtained from the United States Historical Climatology Network (USHCN) website.[15] The data have been collected by the 1,221 weather stations comprising the USHCN for the 48 contiguous states (excluding Alaska and Hawaii), giving the average monthly precipitation in a year from the late 1800s to 2002 adjusted for the Maximum/Minimum Temperature System (MMTS) bias, station moves/changes bias and containing estimated values for missing and/or outlier data. We extracted the data for years between 1990 and 2002, calculating the average (over individual weather stations) amount of rainfall in each state and year, and used these to calculate the values the variable takes for the different companies.

---

[15] ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/

Figure 14: Variable *Rain* by Company



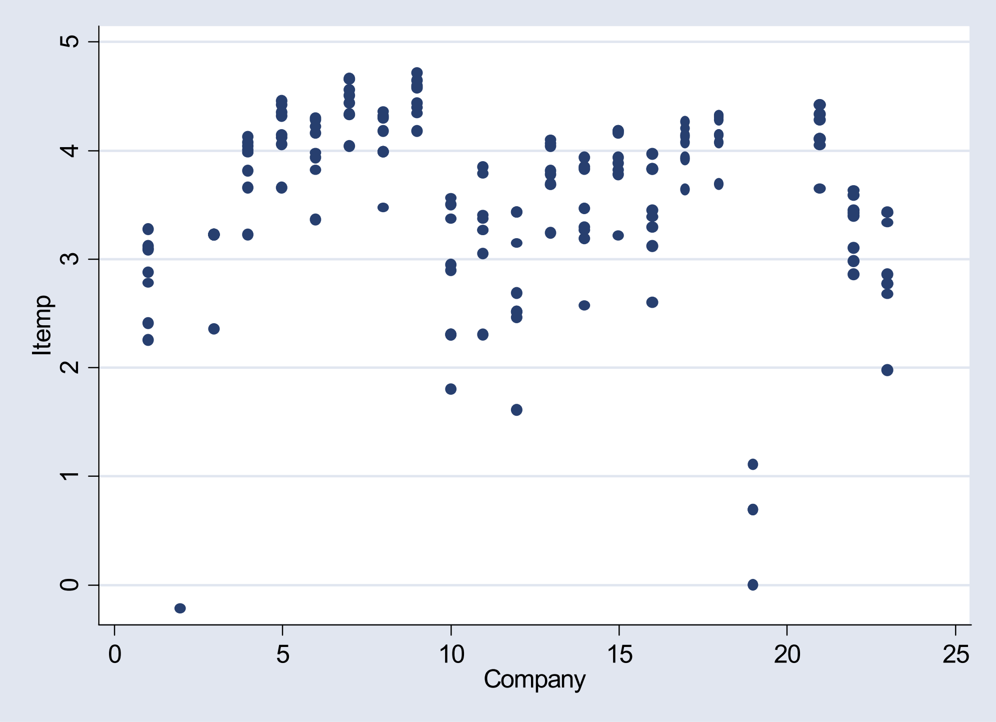
Temperature

2.68 The temperature data were downloaded from the Average Daily Temperature Archive at the University of Dayton, United States.[16] The archive gives the average daily temperatures for 157 U.S. cities covering every State. The data currently run from 1st January 1995 to 4th October 2003. We used the data between 1995 and 2002, calculating for each State the average (over cities) number of days in a year in which the mean temperature was below freezing.

---

[16] http://www.engr.udayton.edu/weather/default.htm

Figure 15: Variable *Temperatures* by Company



3 THE MODEL

3.1 This section describes the methodology and process that has been used for modelling the direct network operating expenditure. More detail on the methodology is provided in the technical annex. Models of the same nature were also investigated for the other cost categories (Direct Non-network, Indirect and Non-specific operating expenditure), but this did not produce satisfactory results.

3.2 Data specific to the Irish situation for the operating (and capital) costs associated with Land and buildings and averages of cost per line over the sample of U.S. LECs in 2002 are used for the other operating cost categories identified in the Consultation Document.

The Dependent Variable

3.3 A preliminary analysis of the data showed that the level of direct network operating expenditure for each company is strongly correlated with the number of lines. This is hardly surprising, given the characteristics of operating an access network and the costs arising from this activity.

3.4 In fact, the variable total number of lines is able to explain almost 80 per cent of the variation of operating costs around its mean.

3.5 However, simply treating number of lines as one of the explanatory variables, among a set of other regressors, would induce a heteroskedasticity problem. Heteroskedasticity arises when the variance of the error is not constant among the companies making up the panel. This turns out to be the case if operating costs are regressed over a number of variables, without adjusting for the size of the network under consideration. Heteroskedasticity would result in inefficiencies in the estimates (i.e. other estimators would be more efficient) as well as in wrong standard errors (i.e. tests on the significance of the coefficients might not be correct).

3.6 Using the logarithm of the dependent variable reduces potential for the model to suffer from a heteroskedasticity problem.

3.7 For the modelling that follows, the dependent variable would therefore be equal to:

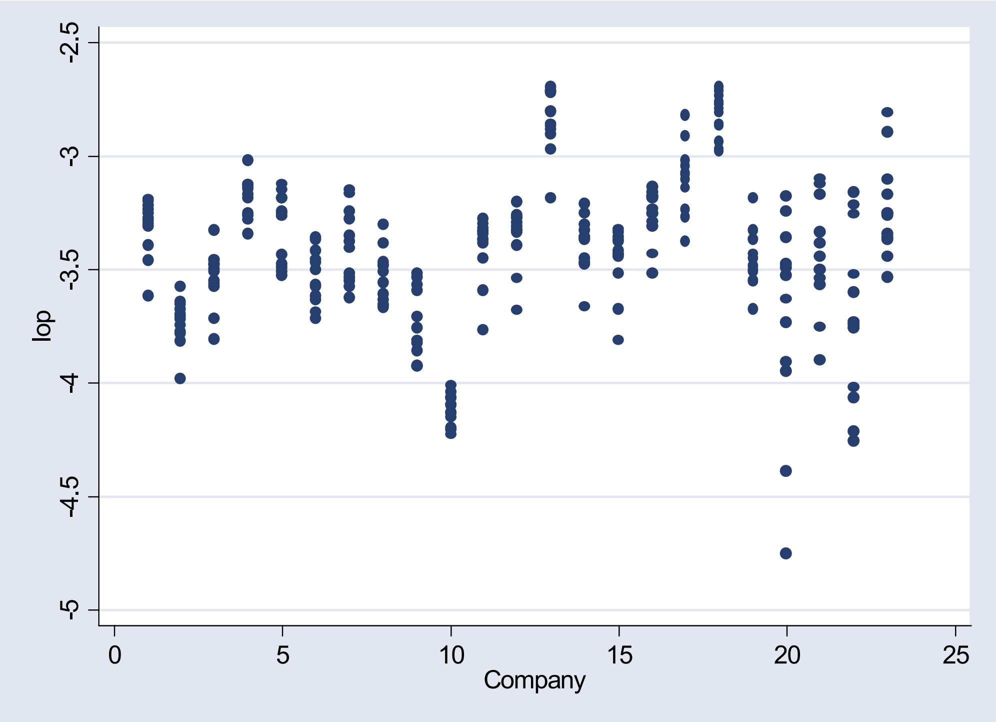
$$y_{i,t} = \ln(opex_{i,t} / totallines_{i,t})$$

3.8 We have taken the natural logarithms of the variable “operating cost per line” to ensure that the noise is expressed as a proportion of unit costs, also reducing the risk of heteroskedasticity.

3.9 Taking $y_{i,t}$ as the model dependent variable would correspond to constraining the coefficient of the natural logarithm of the number of lines, as an explanatory variable of operating expenditure, to 1. This would correspond to assuming constant returns of scale of operating expenditure to number of lines. This hypothesis is not rejected when regressing the natural logarithm of operating expenditure on the full model (see below) including natural logarithm of number of lines as the explanatory variable.

3.10 Taking the natural logarithms also allows useful and constructive observations on the percentage change of the dependent variables induced by corresponding changes of the independent variables.

Figure 16: The Dependent Variable by Company



3.11 As can be seen comparing this chart with Figure 1, the band within which the values of this variable lie is much narrower than the corresponding one for the logarithm of operating costs. The standard deviation of this variable (0.34) is much lower than the standard deviation of the logarithm of Operating costs (1.16). This shows that the total number of lines is a (very) significant explanatory factor of variation of operating costs between companies and is consistent with the use of the chosen dependent variable.

The Model

3.12 The model we have estimated takes the following form:

$$y_{i,t} = \alpha + \beta_1 X_{1i,t} + \beta_2 X_{2i,t} + ... + \beta_N X_{Ni,t} + u_i + \varepsilon_{i,t}$$

where $i = 1 - 23$ and $t = 1991 - 2002$

3.13 Where $y_{i,t}$ is the dependent variable for company i and time *t*, α a constant, $X_{i,t}$ a set of N independent variables, u_i the error term related to each company and $\varepsilon_{i,t}$ the error specific related to each observation.

3.14 $\varepsilon_{i,t}$ are assumed to be independent and identically distributed. The following assumptions on the errors are also made:

- $E[\varepsilon_{i,t}|X] = 0$: the expected value of the error, for any company at any time, is assumed to be zero.
- $E[\varepsilon_{i,t}^2|X] = \sigma^2$: the variance of the error is assumed to be constant over time and over Companies (the assumed scale of the statistical noise is taken to be constant).
- $E[\varepsilon_{i,t}\varepsilon_{j,s}|X] = 0$ if $t \neq s$ or $i \neq j$: the matrix of the covariance of the errors is assumed to be zero (statistical noise is assumed to be uncorrelated over time and between companies).

where X is the set of explanatory variables.

3.15 The same assumptions as above apply to u_i, with the exception that u_i varies only across companies and is constant over time. That is, all the correlation in the overall disturbance term is attributed to the individual random effects u_i. For estimation purposes it is usually assumed to follow a normal distribution.

3.16 The dataset at hand can either be considered as time-series cross-sectional (TSCS) data or panel data. These two classifications correspond to different characteristics of the dataset. Different estimation techniques are associated to the different classifications.

3.17 For this exercise, we have treated the dataset as a panel data and the model has been assumed to be a random effect model. This is because the time period is relatively short and hence not considered a result of any kind of sampling, whereas the choice of companies can be considered a random sample from a population of all operators. Appendix 1 deals with this issue and provides the theoretical background and justification for the technique that we have chosen.

The Results for Direct Network Operating Expenditure

3.18 The results of the model for Direct Network Operating expenditure are shown in the following table.

3.19 The model was developed sequentially, first by judging the potential variables on economic criteria for identifying cost drivers. The dependent variable seems to be trended with time, hence a time variable was included from the beginning together with other variables thought to be the key cost drivers. The significance of the other possible explanatory variables described above was then tested by including them in the model. Firm inference was drawn only from statistically valid models.

Table 3.1: The Model

```
. xtreg ldop aerialshare ductedshare1 time

Random-effects GLS regression                   Number of obs      =       276
Group variable (i): company                     Number of groups   =        23

R-sq:  within  = 0.4337                         Obs per group: min =        12
       between = 0.4439                                        avg =      12.0
       overall = 0.4368                                        max =        12

Random effects u_i ~ Gaussian                   Wald chi2(3)       =    207.24
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        ldop |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 aerialshare |   1.261013   .2652898     4.75   0.000     .7410545    1.780971
ductedshare1 |  -1.252722   .2162927    -5.79   0.000    -1.676648    -.828796
        time |  -.0279105   .0025134   -11.10   0.000    -.0328367   -.0229844
       _cons |  -3.384916   .1114262   -30.38   0.000    -3.603308   -3.166525
-------------+----------------------------------------------------------------
     sigma_u |  .21793932
     sigma_e |  .13912725
         rho |  .71046743   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

3.20 The first line of the results shows the specification of the model. The variable *ldop* is the dependent variable and, as explained above, has been constructed as the natural logarithm of operating costs per number of lines. The variables *aerialshare*, *ductedshare1* and *time* are the independent variables and, as explained above, are respectively the share of sheath cable on poles, the share of underground sheath cable that is ducted and a time trend.

3.21 Above the table with the estimates of the coefficients and the corresponding statistics, information is provided on the overall model and results.

3.22 There are 276 observations corresponding to annual data for 23 companies (the ones listed in Table 2) over a 12-year period (1991-2002).[17] Within a model with four independent variables this corresponds to 272 degrees of freedom.

3.23 The R-squared statistic tells us what percentage of the variation around the mean of the dependent variable is explained by the regressors, the explanatory variables. We believe that being able to explain 44 per cent of the variation of our dependent variable around its mean is a good result.

3.24 The “sigma_u” and “sigma_e” statistics provide us information on the “between firms” and “within firms” standard errors of prediction respectively. The within firm standard error is 14 per cent (assuming normality two thirds of the time the predicted value will be within 14 percent of the actual) and the between firm error is about 22 per cent.

3.25 The Wald test gives us useful information on the general “goodness of fit” of the regression. Under the null hypothesis, all the coefficients are equal to zero (i.e. all the included variables considered together have no explanatory power over the dependent variables). The Wald statistic indicates the confidence with which we can reject this hypothesis. Here the statistic takes the value of 207, a value that enables us to reject the null hypothesis (with more than 99.9 per cent level of confidence).

3.26 In order to verify the significance of the explanatory variables we have referred to the *t*-values. The greater the t-value (corresponding to a smaller standard error), the greater the level of confidence in rejecting the hypothesis that the true coefficient is in fact zero. In this case, the t-statistics for all the variables in the table above indicate that, under the model assumptions, for each variable we can confidently reject the hypothesis that the true coefficient is equal to zero and we can therefore reject the hypothesis that each variable has no power in explaining the level of operating costs.[18]

3.27 The fourth column of the table indicates the likelihood of making a mistake in rejecting the hypothesis that the variable under consideration is insignificant. As shown above, this likelihood is very small for all the regressors: it is actually zero (rounded to the third decimal point) for all variables. We can hence be confident in stating that all the considered variables have explanatory power.

3.28 By the same token, inserting in this model the total number of lines as an additional explanatory variable indicates that this variable is statistically not significantly different from zero. This shows that operating costs do not exhibit economies of scale with respect to number of lines, i.e. operating costs per line do not decrease with the number of lines (if this variable had turned out to be significant and with a negative coefficient, this would have indicated the presence of economies of scale).

---

[17] The year 1990 has been excluded from the sample because of missing observations for total number of lines.

[18] The *t*-values indicated in the table above would allow us to reject the null hypothesis relative to each coefficient with more than 99 per cent level of confidence.

3.29 The last column provides an interval for the value of the estimated coefficient within which the "true" coefficient lies with a likelihood of 95 per cent. The middle point of this interval is the estimated value, as reported in the second column of the table and listed in the table below.

Table 3.2: Estimated coefficients

Independent variables	Estimated coefficient
Aerial share	1.261
Duct share	-1.253
Time	-0.028
Constant	- 3.385

3.30 The estimated coefficients can be interpreted as follows:

(a) A 1 per cent increase in the share of aerial cable would induce a 1.261 per cent increase in operating expenditure.

(b) A 1 per cent increase in the share of underground cable that is put in duct (as opposed to buried) would induce a 1.253 per cent decrease in operating expenditure.

(c) *Ceteris paribus,* operating expenditure decreases by 2.8 per cent per year.

(d) A 1 per cent increase in the number of lines would induce a 1 per cent increase in operating expenditure.

3.31 None of these coefficients is out of line with what we might have expected and this increases our confidence in the robustness of the results.

3.32 We have also found out an acceptable level of significance for the variables Metropolitan and Residential. However, we have decided not to include these variables in the specification of the model for the following reasons:

(a) their high correlation with the variable Aerial Share: negative between Metropolitan and Aerial Share (i.e. the higher the share of lines in metropolitan areas the lower the share of cable on poles) and positive between Residential and Aerial Share (i.e. the higher the share of residential lines the higher the share of cable on poles);

(b) with regard to the variable Metropolitan: when this variable is included, the diagnostic tests performed on the errors show a clear-cut presence of heteroskedasticity and non-normality;

(c) with regard to the variable Residential, the estimated value of the coefficient is such that it does not provide any confidence in the obtained results; the estimated value of around 2 would imply, *ceteris paribus*, that operating cost per line in a network comprising only residential lines is seven times bigger than operating cost per line in a network comprising only business lines.

3.33 We have also initially found an acceptable level of significance for the variable length of cable per line. The model leads close to the same prediction on overall cost as the one without cable length per line. However, the inclusion of the cable length per line does not improve the explanatory power of the model, and also leads to slightly higher standard error of the model. Further, it (clearly) does not pass the Hausman test. This mean that the coefficients estimated from this model by random effects are not consistent. Therefore this model would have to be estimated by fixed effects, but this in turn would mean that the results are conditional on the sample and hence not applicable to Ireland (out of sample prediction). It is therefore not judged a suitable model for this study.

3.34 It is also worth noting that we have considered appropriate the use of nominal operating cost per line (as dependent variable) rather than any real measure of it, because we believe that the time trend of this variable is due to both an inflationary trend and a productivity improvement. Trying to get rid of either of them separately would not solve the problem of eliminating the time trend and therefore would not allow to check for the significance of any explanatory variable that would increase/decrease with time. On the other hand, trying to get rid of both of them at the same time would seem to be artificial and would not make much economic sense.

Testing the Model

3.35 We have checked for multicollinearity, heteroskedasticity, normality of the errors and omitted variables. The results of these tests are reported below.

Multicollinearity

3.36 In general there is nothing wrong with including in a regression model variables that are correlated. Multicollinearity arises from high correlation between independent variables and *might* lead to unreliable estimates with high standard errors and of unexpected sign or magnitude. The problem arises as it is hard for the model to separate the individual impact of each correlated variable.

3.37 For prediction purposes however, multicollinearity has typically little impact as the total impact of all explanatory variables is accurately identified.[19] Therefore the fact that two of the explanatory variables are correlated, as shown in the table below, does not invalidate the predictions based on the model. It could mean that we should not read too much into the individual values of the coefficients on the variables *Aerialshare* and *Ductedshare1*.

---

[19] Verbeek (2000): “A Guide to Modern Econometrics”, page 40.

Table 3.3: Correlation between independent variables

	Aerial Share	Duct share	Time
Aerial Share	1		
Duct share	0.5324	1	
Time	- 0.0276	0.0256	1

Heteroskedasticity

3.38 In order to test our model for heteroskedasticity, we have run a variant of the Breusch-Pagan test as described in the Appendix. The results are shown below.

Table 3.4: Test for heteroskedasticity

```
. reg e2 aerialshare ductedshare1 time

      Source |       SS       df       MS              Number of obs =     276
-------------+------------------------------           F(  3,   272) =    3.25
       Model |  .0183277     3  .006109233           Prob > F      =  0.0223
    Residual | .51137032   272  .001880038           R-squared     =  0.0346
-------------+------------------------------           Adj R-squared =  0.0240
       Total | .52969802   275  .001926175           Root MSE      =  .04336

------------------------------------------------------------------------------
          e2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 aerialshare |  .0016456   .0170348     0.10   0.923    -.0318912    .0351824
ductedshare1 |   .037801   .0153477     2.46   0.014     .0075857    .0680164
        time |  .0006638   .0007572     0.88   0.381    -.0008269    .0021545
       _cons |  .0016365    .008685     0.19   0.851    -.0154618    .0187348
------------------------------------------------------------------------------
```

3.39 The null hypothesis assumes that the variance of the errors is constant. The test is based on the R-squared of an auxiliary regression of the squared residuals from fixed effects estimation of the original model, on the set of explanatory variables in the original model. The R-squared of the auxiliary regression is 0.0346. The test statistic is given by: N(T-1)*R-squared = 23*11*0.0346 = 8.754, and follows a chi-squared distribution with degrees of freedom equal to the number of explanatory variables (minus the constant). Chi-squared (3) critical value at 5 per cent significance level is 7.82. The null hypothesis of homoskedasticity is rejected at 5 per cent level. The critical value at 2.5 per cent level however is 9.35, hence at 2.5 per cent level the null of homoskedasticity would (just) not be rejected. The results of this estimation indicate that the heteroskedasticity problem in our model is a mild one. Therefore using heteroskedasticity robust “White” standard errors, an option in STATA, is not needed here.

3.40 The coefficient estimates would still be consistent even in the presence of heteroskedastic error term. However, the standard errors of the estimates are wrong, therefore the estimates are not efficient and the tests, and confidence intervals, based on the standard errors can be misleading. The most common source of apparent heteroskedasticity in regression models is omitted relevant variable(s).

Normality of the errors

3.41 Under the assumptions set up at the beginning of this section, the random effect estimators are unbiased and efficient. This result does not specifically depend on the assumption that the disturbance terms are normally distributed. However, the confidence intervals computed and significance tests carried out do rely on this assumption of normality – without some such assumption we do not know the critical values against which to compare our test statistics.

3.42 In order to test our model for the normality of the distribution of the errors, we have run the Skewness Kurtosis test on the overall residual term of the model. The results are shown below.

Table 3.5: Test for normality of errors

```
Skewness/Kurtosis tests for Normality
                                                 ------- joint ------
    Variable |  Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)    Prob>chi2
-------------+-------------------------------------------------------
          ue |      0.280         0.124            3.56        0.1685
```

3.43 This test presents a test of normality based on the "skewness" of the distribution and another based on its "kurtosis" and then combines the two results into an overall test statistic. The null hypothesis that the residuals are normally distributed is not rejected at 95 or 90 percent confidence level. The overall residual can therefore be assumed normally distributed.

Omitted variables

3.44 There may be a mismatch between the set of variables included in a regression and those that do in fact explain the movement of the dependent variable. This mismatch may reveal itself when relevant explanatory variables are omitted from the estimated regression or when irrelevant variables are included. The implications of the two mistakes differ.

3.45 If relevant variables are omitted the estimators of the parameters of the remaining variables will be biased unless, fortuitously, the omitted variable is uncorrelated with the other independent variables that have been included. Further, the estimator of the variance of the estimated coefficients will be biased upwards leading to inaccurate inferences that the relevant variables may not be significant.

3.46 Where irrelevant variables are included in the regression, the random effect estimators of the parameters remain unbiased. However, the estimated variance of these parameters becomes larger, so that they may be found to be non-significant (this should not be the case for the variables included in our model).

3.47 The Ramsey test forms an auxiliary regression which includes a (rather arbitrary) number of powers of the fitted values from the model to the set of explanatory variables in the original model. It is not available as a standard option in STATA for random effects models, but specified manually for powers from two to four the results of this test, when modelling the original dependent variable, indicate that the null hypothesis that there are no omitted variables is not rejected at 5 per cent level. This shows as jointly statistically insignificant coefficients on the fitted value parameters (yhat2 to yhat4).

Table 3.6: Ramsey test

```
Random-effects GLS regression                   Number of obs      =       276
Group variable (i): company                     Number of groups   =        23

R-sq:  within  = 0.4397                         Obs per group: min =        12
       between = 0.5001                                        avg =      12.0
       overall = 0.4824                                        max =        12

Random effects u_i ~ Gaussian                   Wald chi2(6)       =    216.16
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        ldop |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 aerialshare |   249.5404   129.3753     1.93   0.054    -4.030588    503.1114
ductedshare1 |  -247.7642   128.5048    -1.93   0.054     -499.629     4.10061
        time |  -5.525133    2.86401    -1.93   0.054    -11.13849    .0882232
       yhat2 |   84.72583    45.0587     1.88   0.060    -3.587606    173.0393
       yhat3 |   16.07012   8.750239     1.84   0.066     -1.08003    33.22028
       yhat4 |   1.133748   .6340583     1.79   0.074     -.108983     2.37648
       _cons |  -499.7351   260.2042    -1.92   0.055    -1009.726    10.25586
-------------+----------------------------------------------------------------
     sigma_u |  .18603375
     sigma_e |  .13911624
         rho |  .64135188   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. test yhat2=yhat3=yhat4=0

 ( 1)  yhat2 - yhat3 = 0
 ( 2)  yhat2 - yhat4 = 0
 ( 3)  yhat2 = 0

           chi2(  3) =    6.41
         Prob > chi2 =    0.0934
```

3.48 Also, none of the variables for which we had data and thought of as possible meaningful explanatory variables were proven to be significant and we have therefore ground to believe that the model is robust, given the limitation of the available data and the inherently complicated task of estimating operating costs only on the basis of cost drivers and assuming the same level of efficiency for the companies included in the sample.

The Other Operating Cost Categories

3.49 This section briefly describes the attempts to find econometric models to explain the other operating cost categories than direct network operating expenditure.

Direct non-network operating costs

3.50 The direct non-network operating costs considered for this exercise corresponds to the following LEC account categories: 6110 + 6120 – 6121, which are the network and general support expenditures less land and building costs.

3.51 This variable was modelled as logarithm of cost per line, using random effects estimation, on several of the explanatory variables described above, including variables measuring the physical size and dispersion of the network, division of customers and switches between rural and metropolitan areas, wages and other variables. Some progress was made toward a satisfactory model with generally the variables measuring density and share of the network in metropolitan areas being significant.

3.52 However, the statistical properties of these models were consistently unsatisfactory for the results to be used in the current exercise. Particularly the model did not pass the tests of heteroskedasticity of residuals and the Ramsey test for omitted relevant variables. The model also (narrowly) failed the Hausman test. All of these failures could reflect the same problem: omitted relevant variables that cannot be treated as random. In combination the failures mean that the coefficient estimates from the econometric model are biased, inconsistent and not applicable to out of sample prediction.

3.53 The direct non-network operating costs were therefore estimated as an average of cost per line over the companies in the sample in year 2002, shown in Table 3.7.

Non-specific direct operating costs

3.54 The non-specific operating costs refer to the LEC accounts 6530, which include testing, engineering and plant operations administration expenses.

3.55 This variable was modelled as logarithm of cost per line, using random effects estimation, on the set of explanatory variables described above. Again, the variables measuring the dispersion of the network and proportion of the network in metropolitan areas were initially found to be significant.

3.56 However, the models suffered from near zero “between” R-squared values. This means the model could only explain movement of the dependent variable through time within companies, but none of the differences in costs between companies.

3.57 Further, the model performed poorly on the diagnostic tests, failing the heteroskedasticity, omitted relevant variables and Hausman tests. As above, these failures could reflect the same factor: omitted relevant variables that cannot be treated as random. The effect of these failures is that the estimated coefficients from these models are biased, inconsistent and not applicable to out of sample prediction.

3.58 This category of cost has therefore been estimated as an average of cost per line over the companies in the sample in year 2002, shown in Table 3.7.

Indirect operating costs

3.59 The indirect operating costs correspond to the LEC account number 710: Total Corporate Operations expenses. This includes, but is not limited to, cost categories like executive pay, accounting and finance, human resources and external relations. They are incurred to run the business of a telecommunications operator as a whole. They do not refer to the costs incurred in running assets, but are rather expenses associated with the administration of the business.

3.60 There are few variables in the ARMIS data set that could be thought to explain the levels of indirect cost. It has not proved possible to develop a statistically satisfactory and economically meaningful model to explain indirect operating costs with regard to network variables. This is hardly surprising given these particular costs have little in common with the design of a telecommunications network.

3.61 Again, this category has been estimated on a simple basis of taking the average of cost per line over the companies in the sample in year 2002, shown in Table 3.7.

Summary of Results

3.62 The data used did not contain variables to build valid econometric models for these cost categories. As the econometric models do not satisfy the assumptions they are based on, no inference should be drawn from them.

3.63 The following table summarises the results in terms of operating cost per line for the categories for which an econometric model was not found to be satisfactory. These costs have to be further adjusted for the year 2004, converted into euros and allocated to the local loop part of the access network.

Table 3.7: Operating cost per line of the selected US LECs for year 2002

Operating cost category	$ per line in 2002 (on monthly basis for the operator as a whole)
Direct Non-network Opex	1.41
Indirect Opex	4.61
Non-specific costs (6530)	1.97

4 APPLICATION TO THE IRISH CASE

4.1 In this section we apply the results of our exercise to the Irish case.

Data Sources

4.2 The values used for Ireland for the reference year 2004 are listed in the table below.

Table 4.1 Values for Ireland for Year 2004

Independent variables	Values
Aerial Share	0.53
Duct Share	1
Time	15

4.3 As explained in Section 2, the exclusion of leased lines from the explanatory variable *Number of lines* might have an impact on the results.

4.4 Apart from the problems arising from the omission of a potentially explanatory variable, including leased lines in *Number of lines* might imply a different level of operating costs per line. In this case, if the number of leased lines relative to the number of other lines (switched and special in the US case and only switched in the Irish case) differs between in-sample and out of sample observations, then the estimate of overall operating costs would be affected.

4.5 As explained above, we have ARMIS data available for leased lines only for 2002. However, the available data seem to suggest that the relative number of leased lines do not differ substantially between US and Ireland (i.e. in sample and out of sample). The US data show that, for 2002, the number of leased lines is 7 per cent the number of other lines (including special lines). For Ireland the corresponding figure is similar.

The Results

4.6 The values of the explanatory variables for Ireland, given in Table 4.1, are inserted in to the model using the estimated coefficients, in Table 3.2. This gives a prediction for the logarithm direct network operating cost per line in dollars per annum. Taking the exponential of this number and multiplying by the number of lines in Ireland we obtain the following result for Ireland for 2004: US $ 1.04 per line per month. This assumes that the Irish operator has average U.S. efficiency.

4.7 Table 3.7 above summarises the results obtained for the other cost categories for the network as a whole (before being allocated to the local loop part of the network) for 2002, the most recent year for which US LECs data are available.

4.8 The estimates of operating cost per line for the categories Direct Non-network Opex, Indirect Opex and Non-specific costs have therefore to be transformed into 2004 values. (This is not needed for the category Direct Network operating cost, since an estimate of the time trend is available from the model.)

4.9 For this purpose, a simple time line was fitted to the "cost-per-line" data over companies and years from 1991 to 2002. The implied time trends for the operating cost categories, are shown in the Table below.

Table 4.2 Annual time trends

Operating cost category	Annual Time Trend (%)
Direct Non-network Opex	- 6.06
Indirect Opex	- 6.66
Non-specific costs (6530)	- 11.33

4.10 These are then applied to the 2002 estimates to obtain the 2004 figures.

Table 4.3 Operating cost per line in US $ of the selected US LECs for years 2002 and 2004

Operating cost category	$ per line in 2002 (on monthly basis for the operator as a whole)	$ per line in 2004 (on monthly basis for the operator as a whole)
Direct Network Opex		1.04
Direct Non-network Opex	1.41	1.24
Indirect Opex	4.61	4.02
Non-specific costs (6530)	1.97	1.55

Conversion to euros

4.11 Once the dollar estimate of the operating costs is available, it will have to be converted into euros using a meaningful exchange rate.[20]

4.12 The exchange rate used for this exercise is constructed as a weighted combination of two factors:

(a) One to be applied to the wage part of the operating expenditure (which we estimate to be around 75 per cent of the total). This would correspond to the ratio between the average Irish and US nominal employment cost in the telecoms sector. This corresponds to the relative compensations of the labour category "Telecommunications equipment installers and repairers" in Ireland and in the U.S.,

---

[20] The conversion is discussed in more detail in Annex 2.

taking into account national insurance contributions and other compulsory employer costs. The ratio of compensations in Ireland and the U.S. in 2002 was 0.755 to 1.

(b) One to be applied to the non-wage part of the direct operating expenditure (which relates to tools and other materials and we estimate to cover the remaining 25 per cent). This would correspond to a nominal exchange rate adjusted for PPP. According to the latest OECD statistics, the PPP for Ireland relative to U.S. dollars in 2002 was 1.01 (pushing the conversion from dollars to euros 1 per cent up for that share of the operating expenditure).

4.13 The overall effective exchange rate is estimated as approximately 0.82 euros to a dollar.

4.14 Using this exchange rate the results of the exercise when applied to Ireland are shown in the following table.

Table 4.4 Summary of the results for 2004

Operating cost category	€ per line per month in 2004 (for the operator as a whole)
Direct Network Opex	0.85
Direct Non-network Opex	1.02
Indirect Opex	3.29
Non-specific costs (6530)	1.27
Total	**6.42**

4.15 These figures have now to be allocated to the local loop part of the access network. In order to do this the allocation keys shown in the table below have been used.

Allocation to the local loop product

Table 4.5 Summary of the results for 2004

Operating cost category	Allocation key to the Access network	Allocation key to the Core network	Allocation key to the Retail Business	Allocation key to Other Businesses
Direct and non-specific Operating expenditure	62.41%	37.59%		
Indirect Operating expenditure	38.40%	23.13%	20.03%	18.44%

4.16 The appropriate level of these costs for Ireland will then need to be allocated between services because only part of the cost should properly lie with the local loop network. As US LECS data is not split between access network and other categories ComReg proposes to use the ratio of eircom's access network costs to total cost from the eircom Separated Accounts.

4.17 The allocation key for the overall category Direct and non-specific Opex has been used to allocate the sum of the relevant three operating cost categories shown in Table 4.4: Direct Network Opex, Direct Non-network Opex and Non specific costs. The results are shown in the table below.

Table 4.7 Summary of the results for 2004

Operating cost category	€ per line per month in 2004
Direct and non-specific Opex	1.96
Indirect Opex	1.26
Total	**3.22**

ANNEX 1: ECONOMETRIC METHODOLOGY

Key Elements in Model Design

A1.1 This technical annex provides a brief review of the econometric methodology used in the Europe Economics study for ComReg.

A1.2 There are four key elements in the design of a statistical models:

A1.2.1 the choice of the dependent variable;

A1.2.2 the choice of the explanatory variables;

A1.2.3 the choice of the functional form;

A1.2.4 the assumptions on the model disturbances.

A1.3 The choice of the dependent variable and of the explanatory variables is set out and explained in the main report. In short, we have separately modelled direct network and non-network operating costs and indirect operating cost as functions of a number of explanatory variables.

A1.4 The selected explanatory variables have been classified into three broad categories: plant, demographic and meteorological. The plant variables refer to the physical characteristics of the network infrastructure (e.g. proportion of underground cable in ducts rather than directly buried); the demographic variables refer to demand characteristics of the network (e.g. proportion of lines in metropolitan areas rather than in non-metropolitan areas); the meteorological variables refer to the potential impact of weather on operating expenditure (e.g. monthly precipitation figures). These are discussed in more detail in the main report.

A1.5 The choice of the functional form is also explained in the report. We have assumed a logarithmic functional form, i.e. the logarithm of cost or unit cost (i.e. the dependent variable) is fitted to a linear combination of the logarithms of the chosen explanatory variables.

A1.6 The main advantages of the logarithmic form are that it provides a robust specification of a range of cost structures, including fixed and variable returns to scale, whilst avoiding the heteroskedasticity problems associated with simpler models (e.g. linear functions). In a logarithmic model, the disturbances are expressed as proportion of costs, whereas in a linear form the disturbances are expressed in levels and the distribution that they follow will depend on the size of the company within the panel.

A1.7 In this note we focus on the assumptions regarding the model disturbances.

Disturbances

A1.8 The total disturbance term can be considered to be made up of two components:

(a) a term capturing any factors, like managerial skill and natural environment, affecting the dependent variable that we have not been able to measure through the selected set of available explanatory variables; we define this term as the “unmeasured disturbance term”; and

(b) a residual, truly random, term that contains no information on the dependent variable; we define this part of the disturbance “random error term”.

A1.9 In all the models presented in this paper, the random error term is assumed to be independently and identically distributed over companies and time (IID). This means the error is considered to follow identical distribution pattern over companies and time, and be independent of every other information, including its own past values and the present values for other companies (if this were not assumed to be the case, then the random error term of a company at any time would contain useful information on the same or other companies at the same or different times).

We express this assumption in the following way: $\varepsilon_{\mathrm{it}} \sim IID(0, \sigma_{\varepsilon}^{2})$.

A1.10 The difference in the models comes from the different assumptions about the behaviour, and existence, of the unmeasured disturbance term. The following two general considerations, among others, will inform our choice on the model specification:

(c) **The nature of the data**.[21] Depending on the exact set of variables under consideration, the sample comprises data for between 20 and 25 companies, and between 7 and 12 years. The data are here considered to be “panel data” in that, conceptually, the companies in the sample are drawn from a notional sample of all possible telecommunications operations, and the results are used for out-of-sample projections (to provide cost benchmarks for eircom), whereas the number of years is inherently limited and does not represent any form of sampling.

This is in contrast to “time-series cross-sectional data” in which a limited number of units are observed over a large number of periods with a view to predicting behaviour in future periods. This means that we are most interested in aspects of the modelling that have implications for across companies comparisons, rather than comparisons over time.

(d) **The purpose of the exercise**. For the purposes of this exercise, the estimated coefficients from the model (calibrated on US data) are used on out of sample data

---

[21] The source of the data on the dependent and the explanatory variables are fully described in the main report. We here focus on the main characteristics of the dataset.

(data that refer to a network operating in Ireland). This, as will made clear below, has a bearing on the selection of econometric methodology used.

Overview

A1.11 There are many possible models, depending on the exact assumptions made about the distribution of the unmeasured disturbance term. In the remainder of this note, we consider the following models (which provide a reasonably comprehensive cross-section of possible panel data models):

(a) ordinary least squares without fixed effects;

(b) fixed effects;

(c) random effects

(d) stochastic frontier model (random effects / panel data version); and

(e) a customised model to take account of different types of random effects and LECs' ownership by wider groups of companies.

A1.12 Whilst statistical tests may sometime enable some models to be rejected as unable to provide a reasonable fit with empirical data, or implying unreasonable levels of supposed purely random error, it is always likely that several different types of models will not be rejected by such tests.

A1.13 The choice of model is therefore mainly a question of principle about the appropriate assumptions to be made about the data sample, and not a purely experimental or statistical matter.

A1.14 We give below a brief description of the different models and the assumptions they are based on, describing what they imply for the unmeasured disturbance term. The last section of this paper describes our informed choice of the methodology applicable to this exercise.

Ordinary Least Squares

A1.15 Equation (2) presents a simple panel data model where operating costs are explained by k explanatory factors (x), a constant (α) and an i.i.d. random error term (ε). The "i" subscript indexes companies and the subscript "t" indexes time.

- $$y_{it} = \alpha + \sum_k \beta_k x_{it}^k + \varepsilon_{it}, \qquad \varepsilon_{\mathrm{it}} \sim IID(0, \sigma_\varepsilon^2) \tag{2}$$

A1.16 This model can be estimated by ordinary least squares (OLS) and it is therefore here called the OLS model.

A1.17 The model implies that the total disturbances, including the unmeasured factors, are independent of each other over companies and time, and that the disturbances have a zero mean. This implies that the disturbances are not "correlated" with each other or with time or company or environment variables, which may not be expected to be the case if total disturbance is made of two parts, as discussed above, one part containing unmeasured company-specific factors (e.g. company-specific efficiency levels, nature of ground, typical wind speeds). In effect, it is here assumed that there are no unmeasured factors that contain information. OLS is therefore the same as a random effects model (see below) where the variance of the random unmeasured effects is zero (they do not exist).

A1.18 It is therefore difficult to suppose that this hypothesis holds true for the dataset considered in this exercise, for which it is quite likely that the error terms, as specified in this model, are not uncorrelated over time for different companies.

A1.19 Some statistical tests to assess the credibility of OLS against more complicated models are discussed in the section below. However, one possible rule of thumb to judge the credibility of an OLS model "on its own" is to consider the estimate standard deviation of the errors. In a logarithmic model, if the errors (ε) were, say, of the order to 0.3-0.5 then this would indicate a degree of noise which does not seem credibly attributable to measurement and accounting errors: instead it would probably show that there are significant effects from unmeasured explanatory factors or efficiency differences, for which the OLS model is not appropriate as outlined above.

A1.20 By contrast if the OLS error is only a few percent, then the model may be considered credible on its own (even though it might still be rejected by statistical tests in favour of another credible model).

Fixed effects

A1.21 If the view is taken that the unmeasured disturbance term is company specific, fixed, and constant through time, the model can be estimated using fixed effects estimation (the model is here called the fixed effect model).

A1.22 The fixed effect model adds a company specific intercept variable into equation (2), i.e. the OLS model. The advantage of this model specification is that no assumptions about the form of the distribution of the unmeasured term are needed, as these are controlled for by the direct estimation of the company specific intercept term α_i, a constant rather than a stochastic variable.

A1.23 The disturbance term contains only the random error term, as for the OLS model.

- $$y_{it} = \alpha_i + \sum_k \beta_k x_{it}^k + \varepsilon_{it}, \qquad \varepsilon_{\mathrm{it}} \sim IID(0, \sigma_\varepsilon^2) \qquad (3)$$

A1.24 For the estimates of the coefficients to be unbiased, all the influence coming from the unmeasured, hence omitted, factors has to be captured by the inclusion of the company-specific intercept terms, in terms of jargon the x has to be independent of the ε. This assumption reinforces the underlying hypothesis that the error term is assumed to be a purely random, i.e. it does not contain any information.

A1.25 For the estimates of the coefficients to be consistent (i.e. converging to the true value in large samples), a slightly laxer condition must hold: the observed explanatory variables (x) have to be strictly exogenous (not dependent on current, future or past values or the disturbance term).[22]

A1.26 However, consistency does not require any restrictions on the relationship between α_i and x_{it}. In other words, the unobserved characteristics, now included explicitly in the model (if they are fully captured by the company specific intercept terms), can be correlated with the observed explanatory variables. An example of this would be if the set of explanatory variables included production input variables such as labour hours, with which the unobserved managerial quality are likely to be correlated (a task under good management can be expected to be completed with fewer inputs or in shorter time than under inefficient management).

A1.27 The fixed effect model provides the "within companies" estimator for the coefficients, useful for in sample inferences, i.e. when we are interested in some specific company within the sample. However, this model is not particularly useful when our objective is to find out characteristics and coefficients common for all companies, including those outside the sample, as in our case.

A1.28 Other drawbacks of this model specification are:

(f) it removes the effect of all explanatory factors that are constant through time (they have zero deviation from their time means); and

(g) it reduces degrees of freedom in the model as a new set of coefficients have to be estimated. The reduction of the degrees of freedom is generally not desirable.

Random effects

A1.29 The random effects model specification assumes that the unmeasured disturbance terms are not fixed, but random and that they are independently and identically distributed (IID) over companies. Unlike the SFA analysis however, no strong restrictions on the shape of the distribution in terms of cut off points are needed for estimation.[23]

---

[22] To reiterate, this assumption is not likely to hold if the company specific intercepts do not sufficiently control for the omitted variables problem.

[23] The default distribution used by STATA is a symmetric normal, which can be a strong restriction, but random effects estimation does not inherently require this, and a different distribution can be used if it is so desired.

A1.30 The random effect model specification corresponds to equation (4).

- $$y_{it} = \alpha + \sum_k \beta_k x_{it}^k + u_i + \varepsilon_{it}, \qquad \varepsilon_{\mathrm{it}} \sim IID(0, \sigma_\varepsilon^2),\ u_{\mathrm{i}} \sim IID(0, \sigma_u^2) \qquad (4)$$

A1.31 In the random effect model, the unmeasured disturbance term, u_i, is assumed to be company specific and constant over time.

A1.32 For the estimate of the coefficients of this model to be consistent, it is assumed that u_i and ε_{it} are mutually independent and independent of all x_{js} (for all j and s).[24] This corresponds to assuming that the set of explanatory variables are independent, over time and companies, of the unmeasured disturbance term (which includes an “efficiency factor”), an assumption that crucially depends on the variables included in the model specification and that needs careful scrutiny. So, for example, if the set of explanatory variables included production input variables such as labour hours, with which the unobserved managerial quality are likely to be correlated (differences in labour hours between companies can depend on the unobserved managerial quality), then the consistency of the estimates would not be assured by this kind of model (a fixed effect model would be, *ceteris paribus*, more apt in this case, see above).

A1.33 The model in (4) can also be extended to take into account of the presence of a time trend in the random error term (see section below) or more than one kind of random error terms.

A1.34 For example, the random effects model specified in equation (4) assumes that the data error term (ε) is homoscedastic (its variation is constant over companies and time). One way to deal with heteroskedasticity, without having to assume its form, is to use the OLS estimator for the coefficients, while adjusting their standard errors for a general form of heteroskedasticity: this however does not produce efficient estimates.

A1.35 Alternatively, assumptions about the specific form of heteroskedasticity in the model can be made, allowing the exploitation of the structure of the error covariance matrix using a feasible GLS or a maximum likelihood approach. These estimators, however, are typically computationally cumbersome.

A1.36 The major attractiveness of the random effect model is that it is not conditional on the company specific effects, and would hence allow us to make inference with respect to the characteristics of companies outside the sample.

Stochastic Frontier Analysis

A1.37 Stochastic frontier analysis aims to estimate an efficiency frontier by assuming that the unmeasured disturbance term is distributed through a known function. The panel data

---

[24] This means that the overall disturbance term has a specific from of autocorrelation, and the model is estimated by Generalised Least Squares.

version of the stochastic frontier model can be viewed as a modification of the general panel data “random effects” model. It would take the following form:

- $$\begin{aligned} & y_{it} = \alpha + \sum_k \beta_k x_{it}^k + w_{it} \\ & w_{it} = u_i + \varepsilon_{it}, \qquad \mathrm{u_i} \geq 0 \; and \sim IID\left(\eta_u \geq 0, \sigma_u^2\right), \; \varepsilon_{\mathrm{it}} \sim IID(0, \sigma_\varepsilon^2) \end{aligned} \qquad (5)$$

A1.38 ε is here assumed to be a white noise error term (for each company) and u is usually assumed to be half-normal or exponentially distributed, with mean greater than zero. The model is estimated by maximum likelihood. In the random effect model, the assumed distribution of the unmeasured disturbance term is generally normal (with zero mean and symmetric around its mean). Other assumptions about the properties of the two components of the disturbance term are the same as in the symmetric random effects model.

A1.39 The difference between this model and a fixed effect model boils down to the characteristic of the variable for the unmeasured company specific effects: in the fixed effect model, that would be a deterministic variable (whose estimate would correspond to the average of the constant α), whereas in the SFA and in random effect models it is a stochastic variable.

A1.40 Also, the assumed distribution (normal or exponential) of the inefficiency term can have a significant effect on the results. For example, the half-normal distribution constricts most of the companies to be relatively close to the efficiency frontier, whereas in reality there could be one clear efficiency leader. Further, as each distribution gives a different likelihood function to be maximised, it is not straightforward to compare the models and understand differences in the results under alternative distributional assumptions.

Customised model

A1.41 It is possible to extend and/or combine the above models in several different ways. It could be desirable, for example, to allow for efficiency to vary by time in a way that is not captured by the measured explanatory variables. This would capture technical change in the industry, affecting all the companies, or company specific efficiency improvement, due, for example, to managerial experience.

A1.42 In equation (6) below, technological improvement common to all the LECs is captured by the time trend variable, and also the company specific random efficiency component is allowed to vary by time.[25]

- $$y_{it} = \alpha + \sum_k \beta_k x_{it}^k + \gamma t + u_{it} + \varepsilon_{it}, \qquad u_{it} = f(t) * u_i \qquad (6)$$

---

[25] f(t) could be a set of time dummy variables.

A1.43 Whether to allow the company specific effect, u_i, to vary by time depends on the investigator's view on how significant company specific (i.e. not shared by the whole industry) efficiency improvements are.

A1.44 It is also possible to include more than one type of random inefficiency term. For example, if business or physical environment was thought to differ between the companies, in a different way to management quality, a second inefficiency term can be included in (6).

A1.45 It could be thought that whereas managerial quality varies between companies according to the uniform distribution, which would allow for a wider variation in the underlying efficiency of the management, unobserved environment effects are more likely to follow some other distribution, like the normal distribution. Along the same lines, the management quality can be set to vary by company, and the environmental factors across geographical or operation regions or state. Under these assumptions, the model would look like equation (7), where "j" indexes regions and "i" indexes companies and the error is now made of three components.[26]

- $$y_{it} = \alpha + \sum_k \beta_k x_{it}^k + \gamma t + u_i + \varepsilon_{it} + v_j \qquad (7)$$

A1.46 All the above customised random effect models, however, would still require the disturbance term, including the unmeasured factors, to be uncorrelated with the measured explanatory variables for the estimates of the coefficients to be consistent, as described above.

A1.47 This assumption, common to all the random effect models described in this section, is tested through the Hausman test, described in the section below.

Choice of the model for estimation

A1.48 As discussed above, the choice of the econometric methodology can be seen to proceed in two steps.

Step 1: First, a decision has to be made whether to treat the individual effects as fixed or random.

Step 2: Second, if the random effects are chosen, one has to decide how to 'best' represent these through the distributional assumptions of the unmeasured disturbance terms.

---

[26] These additions would require programming and estimation of some non-standard functions for the statistical package used, Stata 8.

A1.49 For both steps, statistical methods can give at best only partial advice on the way to proceed, thus a significant amount of informed judgement must be exerted – concerning both how the world is viewed to work, and what the final application of the model is.

Step 1: fixed or random effects?

A1.50 It is not straightforward to decide whether to treat the company specific effects as fixed or random. The particular modelling environment in each case, the proposed use of the estimates, as well as how the world is viewed to work all influence the decision. There is always an element of judgement involved, as the few statistical tests available (described in the section below) can only reject a hypothesis in favour of another one.

A1.51 The fixed effect specification has some desirable qualities:

(h) there is no need for distributional assumptions for the unmeasured inefficiency terms as these are estimated directly through the company specific intercept term.

(i) No restriction is placed on the relationship between the unobserved inefficiency and included explanatory variables; and

(j) it can be extended to include a general time trend, as shown with the random effect models above.

A1.52 However, for the purposes of this study fixed effects are not to be used for two main reasons.

1 First, we use the estimated coefficients on out of sample data. This largely rules out the usefulness of fixed effects estimation for this particular exercise, because we have no way of knowing what the individual effect would be for Ireland. We have no interest in the estimated in-sample fixed effects.

2 Second, some of the explanatory variables do not vary much over time, which dummy variables fixed effects estimation would remove from the sample.

A1.53 The estimates from the random effects estimation are not conditional on the individual effects, and hence allow us to make inference with respect to the population characteristics: they lend themselves more readily to out of sample analysis than with fixed effects estimation.

A1.54 The random effect specification also exploits both dimensions of variation — within and between companies — estimating an efficient combination of these.[27] Random effects estimation is therefore the chosen estimation platform for this study.

---

[27] The simple OLS also estimates a combination of within and between effects. However, the random effects estimator gives the efficient (minimum variance) combination of the two (OLS places same weight on the two types of variation, random effects uses the variances of the two error terms to construct the weighting), and is hence preferred.

Step 2: which specification of the random disturbance term?

A1.55 The SFA models differ from symmetric random effect models in that the unmeasured company specific factors are all assumed to be positive. That is, there is some ideal level of efficiency and the unmeasured factors (inefficient management, harsher than the best environmental factors) move companies away from that frontier, increasing cost. This also means that the estimated coefficient from SFA models should refer to the ideally efficient operator. This is not viewed as desirable in reality, when the results of the estimation are applied to *eircom*.

A1.56 Further, there is no strong reason to expect asymmetric, half-normal type distribution of the inefficiency error terms for the dependent variables used in the study. Hence we do not expect SFA to produce very much different results from symmetric random effect models.

A1.57 Further, the cost data for the U.S. LECs come in such aggregated form that it would be meaningless to include two kinds of random inefficiency terms — a possibility which was described by equation (8). It is likely, however, that there is some shared efficiency improvement over time in the communications industry. Therefore a time trend is included to the model.

The Choice

A1.58 For the various reasons discussed above, a symmetric random effects estimation is chosen as the estimation platform for this study, with the basic theoretical form shown below:

$$y_{it} = \alpha + \sum_{k} \beta_k x_{it}^k + \gamma t + u_i + \varepsilon_{it}, \quad \varepsilon_{\mathrm{it}} \sim IID(0, \sigma_\varepsilon^2),\ u_{\mathrm{i}} \sim IID(0, \sigma_u^2) \qquad (9)$$

A1.59 This choice is broadly supported by the results of the statistical tests described in the following section.

Statistical tests

A1.60 This section briefly outlines some of the statistical tests available to assist in the various choices over the models described above. It is not an exhaustive set of statistical tests and it is not our intention to present it as such.

1.1.1.1 Testing OLS versus random effects

A1.61 One available statistical test for the validity of the random effects against OLS is a Breusch-Pagan test, which tests the hypothesis that the variance of the inefficiency error term is zero, i.e. all the disturbance term can be attributed to the white noise data error. A rejection of the null hypothesis indicates that a random effects model is preferable over simple OLS.

A1.62 When applied to our model, the null hypothesis is clearly rejected.

1.1.1.2 Testing fixed effects versus random effects

A1.63 The Hausman test is one way to test the validity of the random effects assumption against fixed effects, and hence to assist in the choice between fixed versus random company specific unmeasured effects. The test compares two estimators:

(k) one consistent both under the null hypothesis of no correlation and also under the alternative hypothesis that the unmeasured inefficiency terms (omitted variables) have an influence on the observed explanatory variables (the fixed effect estimator); and

(l) one that is consistent, and typically efficient, under the null hypothesis only (the random effects estimator, i.e. no correlation between the unmeasured error term and the explanatory variables).

A1.64 The test compares the estimated values from the two models. As they both should be consistent under the null hypothesis, they should not be too different from each other if that hypothesis is true. A significant and systematic difference indicates that the null hypothesis is unlikely to hold.

A1.65 If this is the case, and if fixed effects models are not relevant for the reasons stated above, then a possibility for consistent estimates is a further extension to the random effect model via instrumental variables. A possible alternative to pure random effect estimator is the Hausman-Taylor estimator, where transformations of the explanatory variables serve as their own instruments, providing a way between fixed and random effects estimation.[28]

A1.66 The model in our applications for direct operating cost and non-network operating cost pass the Hausman test at the 5 per cent level, lending further credibility to the choice of random effects over fixed effects.

1.1.1.3 Testing the homoskedasticity assumption

A1.67 To test for heteroskedasticity in ε_{it}, we can use the fixed effects residuals, as the fixed effects estimator is consistent also under the random effects assumption that α_i are i.i.d. and independent of the explanatory variables (basis also for the Hausman test).

A1.68 This makes the available tests computationally less burdensome. One possible test is a variant of the Breusch-Pagan test that regresses the squared residuals on a constant and the set of J explanatory variables thought to affect heteroskedasticity. Under the null hypothesis of homoskedasticity, the test statistic, computed as N(T-1) times the R-squared of the auxiliary regression, is distributed chi-squared with J degrees of freedom.

---

[28] Hausman, J. A. and Taylor, W. E. (1981) “Panel Data and Unobservable Individual Effects”, Econometrica, 49, pp 1377-1398.

A1.69 When applied to our direct operating cost model ,the null hypothesis of homoskedasticity is rejected at the 5 percent significance level, but not at the 2.5 percent level. This indicates that the residuals could well be heteroscedastic, which could be a result of an omitted variable that cannot be considered as a random variable (as the random effect model does).

Glossary of terms

A1.70 The purpose of this section is to give explanations for some of the statistical jargon used above, as it is applied in this paper.

Term	Explanation
dependent variable	the variable of which behaviour we are trying to model
random error term	here; purely random error that is independent of everything and carries no information about anything, part of the total disturbance term
unmeasured inefficiency disturbance term	here; anything we are unable to directly measure that influences the dependent variable, also omitted variables, part of the total disturbance term
disturbance term	here; a combination of the error term and unmeasured disturbance term
homoscedastic	a variable is homoscedastic if its variance is constant over time and over companies
heteroscedastic	opposite of homoscedastic, variance not constant
degrees of freedom	the number of independent observations available for estimation
unbiased estimator	the expected value of the estimator is the true underlying value
consistent estimator	the expected value of the estimator tends to the true underlying value as the sample size tends to infinity.

ANNEX 2: THE DETAILS OF THE EXCHANGE RATE

A1.71 The exchange rate used for this exercise is constructed as a weighted combination of two factors:

A1.71.1 One to be applied to the wage part of the operating expenditure (which we estimate to be around 75 per cent of the total). This would correspond to the ratio between the average Irish and US nominal employment cost in the telecoms sector. This corresponds to the relative compensations of the labour category "Telecommunications equipment installers and repairers" in Ireland and in the U.S., taking into account national insurance contributions and other compulsory employer costs. The compensations for the Ireland part of the computation have been computed using eircom salary data and therefore cannot be quoted in this document.

A1.71.2 One to be applied to the non-wage part of the direct operating expenditure (which we believe relates to tools and other materials and estimate to cover the remaining 25 per cent). This would correspond to a nominal exchange rate adjusted for PPP. According to the latest OECD statistics, the PPP for Ireland relative to U.S. dollars in 2002 has been 1.01 (pushing the conversion from dollars to euros 1 per cent up for that share of the direct operating expenditure).

A1.72 The U.S. wage data was downloaded from the Bureau of Labour Statistics website. The origin of the data is from the Occupational Employment Survey (OES), category 49-2022: "Telecommunications Equipment Installers and Repairers, Except Line Installers".

A1.73 The wages from the OES are straight-time, gross pay, exclusive of premium pay. Base rate, cost-of-living allowances, guaranteed pay, hazardous-duty pay, incentive pay including commissions and production bonuses, tips, and on-call pay are included. Excluded are back pay, jury duty pay, overtime pay, severance pay, shift differentials, non-production bonuses, employer cost for supplementary benefits, and tuition reimbursements.

A1.74 The data is given for each state in 2001. The state shares calculated were used to convert the data into company specific values and then an average annual wage was calculated over the 23 companies in the sample. The annual wage was then divided by the 2080 hours per year (the amount of hours assumed in deriving the annual wage by OES) to give the average hourly wage.

A1.75 In Ireland the compulsory employee compensation includes employers national insurance and pensions contributions. According to the Employer Cost of Employee Compensation series from the U.S. Department of Labour, wages and salaries were 72.9 percent of the total cost for Private Industry in 2001. However, the "total benefits" used to derive that share include insurance contributions, annual leave, overtime etc pay, as well as legally required benefits and pensions contributions. Pensions and savings contributions, together with other legally required benefits have been estimated to be 11.3 percent of total contribution. This was judged to be the closest categories to the Irish national insurance and pensions contributions. Therefore the “additional cost” adjustment factor for the U.S. implied that wages are 88.7 percent of the Irish comparable labour cost.

A1.76 The overall effective exchange rate is estimated as approximately 0.82 euros to a dollar.