COMPUTER AIDED MODELLING OF COSTS IN THE ROAD HAULAGE TRANSPORT

Doc dr.hab. Bernard Kubiak

Department of Organization
of Data Processing
University of Gdansk
Gdańsk, Poland

ABSTRACT

This note provides a concise summary of the author's work entitled "Problems concerning the variability of costs in the road haulage industry". Analysis of a large number of variables about which data might reasonably be expected to be available is used to develop cost models containing relatively few variables. Linear and log-linear models are fitted by regression according to a taxonomic procedure in which a criterion based on a measure of distance between standardised variables is used in deciding which variables to include. Use of the resulting estimated costs in the context of break-even analysis in the management of road haulage undertakings is discussed.

1. Introduction

Cost modelling is a form of quantitative analysis not based on conventional accounting methods.

Dealing with modelling, I tried to understand cost behaviour through the analysis of actual costs in relation to other data. In other words, I tried to answer the question "How do costs react to changes in crucial factors?".

It appeared to be a very difficult problem to solve, because the mechanism of incurring costs is a stochastic one. Thus, sometimes it was even impossible to say how costs change in relation to other measurable and controllable factors such as total run time, number of vehicles, total capacity of vehicles, number of drivers, and total working hours of drivers.

So, the first main task which was faced was to define and explain the mechanisms that determine the functional form of costs and their variation.

For this purpose, I had to define and propose a cost model based on available information, which will make it possible to obtain crucial and relevant cost-related factors and use the best quantitative methods.

It is important to emphasise that I have only assumed available cost information. In fact certain types of data that would be useful for the decision-makers concerned with costs are not available generally.

By applying a taxonomic method for selecting the factors which determine costs, I was able to explain the variation in costs and divide total costs and their components into constant (fixed) and variable parts from standard accounting data available from 13 road transport enterprises.

The investigation enabled me to establish which factors (independent variables) should be measured exactly and should be followed with particular attention in studying the variation in costs.

Such invastigations have not been carried out before in the Polish road haulage industry. This approach has gone beyond the traditional treatment of the subject. It was also a good opportunity to examine whether the routine and traditional way of explaining relationships between costs and relevant factors is adequate.

2. Model construction

Before the methodology of modelling of costs in the road haulage industry is presented, it is necessary to say something about the construction of the model.

I was particularly interested in behavioural equations, which describe the behaviour of individuals or groups in the economy, especially such familiar economic relationships as cost functions.

All the variables contained in a model are related causally, and we shall discuss them later on.

Let us first consider the form of the model I used. I assumed that our dependent variable (costs) is related to more than one independent variable. For that purpose we use multiple regression, because presumably we shall need two or more explanatory variables or predictors. A value of each dependent variable and corresponding values of the independent variables were available for every month over a 3-year period.

So we write

$$Y_{ji} = a_{jo} + a_{j16}X_{16i} + a_{j17}X_{17i} + \dots + a_{j48}X_{48i} + U_j \qquad (1)$$

where:

Y_{ji} for j=1,2,...,15 , is total cost or a component of cost Y_j of output in period i (i=1,2,...,36) (see Table 1 for a list of variables Y_j)

a_{jo} is the intercept term

a_{jk} (k=16,17,...,48) are the unknown population structural parameters (regression coefficients)

X_{ki} (k=16,17,...,48) are the values of the 33 independent (explanatory) variables X_k as observed in period i (i=1,2,...,36) (see Table 2 for a list of the variables X_k)

U_{ji} is a measure of the deviation of the estimated value $\hat{Y}_j$ from the observed value Y_j in period i (the error term)

Y_1	fuel, oil, lubricant costs
Y_2	tyres, inner tubes costs
Y_3	depreciation of vehicles with engines
Y_4	depreciation of trailers
Y_5	repair and maintenance costs
Y_6	hire charges
Y_7	driver wages and insurance depending on vehicle-km
Y_8	driver wages and insurance depending on the working time of vehicles
Y_9	driver wages and insurance not depending on the working time of vehicles
Y_{10}	approved driver expenses (overnight accommodation, meals, etc)
Y_{11}	costs of spares and essential equipment
Y_{12}	vehicle insurance costs
Y_{13}	overheads for the departments
Y_{14}	overheads for the organization (not including Y_{13})
Y_{15}	total costs = $\sum_{j=1}^{14} Y_j$

Table 1: Depedent variables (costs) Y_j

(i)	In physical measures	
X_{16}	goods moved	(tonne-km)
X_{17}	vehicle-distance	(vehicle-km)
X_{18}	loaded vehicle-distance	(vehicle-km loaded)
X_{19}	test running	(vehicle-km)
X_{20}	supply	(vehicle-days)
X_{21}	capacity	(usable vehicle-days)
X_{22}	capacity supplied	(operated vehicle-hours)
X_{23}	capacity supplied	(operated vehicle-hours)
X_{24}	total run time	(hours)
X_{25}	loading and unloading time	(hours)
X_{26}	goods lifted	(tonnes)
X_{27}	number of powered vehicles	
X_{28}	number of trailers	
X_{29}	total load capacity of powered vehicles	(tonnes)
X_{30}	total load capacity of trailers	(tonnes)
X_{31}	total number of employees of the department	(people)

X_{32}	number of drivers	(people)
X_{33}	number of driver assistants	(people)
X_{34}	total working hours of lorry drivers	(hours)
X_{35}	total working hours of driver assistants	(hours)
(ii)	Indices (ratios)	
X_{36}	index of utilisation	(X_{22}/X_{20})
X_{37}	index of utilisation	(X_{22}/X_{27})
X_{38}	index of utilisation	(X_{18}/X_{17})
X_{39}	index of utilisation	$(X_{26}/X_{29}+X_{30})$
X_{40}	index of utilisation	(X_{16} completed/X_{16} potential)
X_{41}	index of utilisation	(X_{24}/X_{23})
(iii)	Averages	
X_{42}	average inventory capacity per vehicle	(tonnes/vehicle)
X_{43}	average load per vehicle	(tonnes/vehicle)
X_{44}	average design speed of fleet	(km/hour)
X_{45}	average operating speed of fleet	(km/hour)
X_{46}	mean haul	(km)
X_{47}	average vehicle working time per day	(hours/day)
X_{48}	average loading and unloading time per tonne	(hours/tonne)

Table 2: Independent veriables X_k

Equation (1) assumes a linear relationship, but in the road haulage industry there usually exists a non-linear relationship between costs and explanatory variables, for example a log-linear relationship

$$Y_{ji} = a_o X_{16i}^{a_{j16}} X_{17i}^{a_{j17}} \;.....\; X_{48i}^{a_{j48}} e^{U_{ji}} \qquad (2)$$

Both equations (1) and (2) were estimated by least squares. For the least squares computation of parameters, we can take the natural logarithms of both sides of equation (2), thus

$$\ln Y_{ji} = \ln a_o + a_{j16} \ln X_{16i} + a_{j17} \ln X_{17i} + \ldots + a_{j48} \ln X_{48i} + U_{ji}$$

We define the following new variables and the constant A.

$$\hat{Y}_{ji} = \ln Y_{ji}$$
$$\hat{a}_o = \ln a_o$$
$$X'_{16i} = \ln X_{16i}$$
$$X'_{17i} = \ln X_{17i}$$
$$X'_{18i} = \ln X_{18i}$$

Then we can rewrite the regression equation as

$$\hat{Y}_{ji} = \hat{a}_0 + \hat{a}_{j16}X'_{16i} + \hat{a}_{j17}X'_{17i} + \ldots + \hat{a}_{j48}X'_{48i} + U_{ji}$$

which is linear in all the variables, so we have reduced the problem to one of multiple linear regression. Other transformations are possible for some other types of non-linear equations.

I have estimated a few hundred different types of relationship and I could not establish which form of relationship between the costs and their explanatory variables gives the best fit in general.

Cost relationships are very complex and not smooth. Usually several functional forms fit the data equally well. Relationships are continuosly undergoing change as a result of changes in economic, technical and organizational conditions. In addition, errors in measurement can affect which form of relation appears to fit the data best.

That is why it is important to observe many factors, particularly those which most determine costs. It was rather difficult to obtain within the time available for the study the necessary number of independent variables to explain the variation in cost. It was even more difficult to choose the important ones and construct good cost models by least squares regression.

There is a further problem. It is very often the case with economic data, especially for costs, that a close relationship exists between the explanatory variables themselves (multicollinearity).

I found that a high correlation exists between different independent variables (X_k and X_l). When such a linear relation exists among two or more of the explanatory variables, it is difficult to measure their separate influence upon the dependent variable (cost).

There was little doubt that multicollinearity existed between the explanatory variables. I have tried to avoid or to reduce collinearity using the available tools of statistical analysis. I have used a taxonomic method to select meaningful explanatory variables, which has a low correlation between themselves.

In that way, I could identify their separate effects upon the costs to be estimated. But it was not always possible, because multicollinearity is an inherent characteristic of much economic, technical and organizational data about which little can be done. relationships here are often multiple and complicated.

Returning to the taxonomic method, it is necessary to emphasise that it was complicated, but we can identify the following steps.

(1) Establish the aim and scope of the investigation

(2) Establish the dependent variables (total costs Y_{15} and component costs $Y_1, Y_2, \ldots, Y_{14}$)

(3) Establish the independent variables (which describe the economic, technical and organizational relations of production)

(4) Standardise the independent and dependent variables (to reduce the range of size and to make different variables comparable)

(5) Examine the matrix of correlation coefficients (between dependent and independent variables and between one independent variable and another)

(6) Examine the matrix of distances between dependent and independent variables (W_{jk}) and between one independent variable and another (C_{kl}) where the distance between two variables is mean of the 36 absolute differences between values taken by the two variables in the same month after standardisation

(7) Establish a certain criterion (C) for the exclusion of independent variables from the cost model

(8) Identify the independent variables highly correlated with the dependent variable (distance between independent variables and dependent variables W_{jk} is lower than the criterion C)

(9) Identify the independent variables which satisfy the condition $C_{kl} \geq C$ and where there is a low correlation between themselves.

3. The estimation of regression model, and measurement of explanatory power

After the choice of explanatory variables I was able to quantify and measure the relationships between the independent and dependent variables.

The relationships between costs and their explanatory variables were not strictly linear in many cases. However, I used a linear regression because linear equations have given a sufficiently close fit to the observations.

In many cases I found that linear regression with one independent variable will give a good fit to the observations, but there were also situations in which I found the fit unsatisfactory. The errors were too large and the regression coefficients were not significant.

There are many reasons for this; one reason might be found in the fact that costs changed markedly during the period studied. Since the hypothesised linear relationships are only approximations, one would not expect them to be stable under large variable changes.

Generally, I can say that the most important determinants of the explained variable (costs) have been included in the regression equations. So, I could employ standard econometric methods to make estimates of the population parameters and to conduct test of hypothesis (using, for example Student's t and the Durbin-Watson statistics).

Let us now examine the results of applying the taxonomic procedure to data for the road haulage industry in Poland,

Cost components	Number of enterprise													Entire road haulage industry	
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	version I	version II
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Y_1	24	22	24,39	22,31,32,42	24	17,39	39	22	24,40	44,26,30,41	22	17	17	24	36,43
Y_2	18	31	16	44,40.	28,26	29	39,42	32,39	44	20	18	28,32	18	17	18
Y_3	16	17,27	17	17,29,30,40, 42	18	16	17,30	16,44	18,34,41	18,26,39,42	22,27,43	16,36	18	17,24	18
Y_4	28	17,27	34,36,39	45	30	29	39,42	30	26	29,19,26,40	32	41,20,29, 46	35,40	30	17
Y_5	24,33	19,21,33,35, 42	40,19,43	27,18,26,41, 45	25	34,44	17,30	31,17,37, 45	28,41	24,30,39,42	32	16	42,37,39,47	29	29
Y_6	33,43,47	48,33,35,39, 47	28	42,24	–	–	26,19	24	–	19,32	31	41,29,38	27,25	33,47	36
Y_7	32,42	43,19,39	25	36,30,38,41, 42	41,26	24	34,40	24	16,34,46	43	32	34,26,28, 30	36,25,33	33,37	30
Y_8	16	25,28,30	32,19	29,45	18	29,36,37	34,35, 40,44	26	23,27,34	16	17,33,35	16,36	22,25	24	24
Y_9	48,17,18, 20,21,22, 25,32,33, 45	30	33	27,45	48,20,31, 46	41	19,46	24	23,27,34	48,19	34,43	39	45,23,24,25, 34,39	34	48,26
Y_{10}	26,33,41, 43	30	22	42,17,18,21, 22,24,32,37	28	16	41,36	37,33,34, 35	47,34	26,20,37	47,33,38	40,29,33, 35,42	36,20,30	29,41	18,45
Y_{11}	28,39,42, 48	42,22	28	29,38	23	41,24,25, 34,37,39, 30	28,35	37,28,30, 31,32,33 42	30	38,30	47	27	19,35,36	33	18,45
Y_{12}	34	32	22	42,17,18,20, 22,36,37,44	46	44,19,36, 48	29,37	45	48,19,34, 35,41	36,30,42, 46,47	45,46	45,25	45,20,23	34	34
Y_{13}	33,20,25, 28,30,31, 34,40,47	29,39,46	22	34,29	21,38,43	24	34,19, 40	46,40,48	26	46	48	41,20,29, 46	28,38,46	34,23	34,23
Y_{14}	28	34	28	30	16,29	40	20,25	31	36	39	32	23,48	28,32	29	29
Y_{15} version I	24	17,27	16,35	40,44,47	17	16	26,44	32,36,37, 39,44	23,40	16	32,27,44,43	16	18	27	47
Y_{15} version II	4,40	17,27,6	16,12,15	40,8,10,33	17	–	–	32,35	23,14	16,3	32,6,13,27,44	16,7,15	1	–	–

Source: Author's results based on data from the above enterprises.

Table 3: The most significant factors for each enterprise in road haulage industry selected by the taxonomic procedure, ranked by revealed significance.

ENTERPRISE	REGRESSION				COSTS				φ^2	R	S_e	EL	Z_L
	Y = f(X)		X = f(Y)		FIXED		VARIABLE						
	a	b	a	b	IN Zł	%	IN Zł	%					
1	2	3	4	5	6	7	8	9	10	11	12	13	14
HAULAGE SECTOR	1,032	-115900	0,641	370700	-115.900		903.700	100,0	0,339	0,813	392.710	1,15	85,5
I	0,677	215720	1,148	- 20330-	215.720	23,8	689.780	76,2	0,223	0,881	65.062	0,76	14,6
II	0,884	- 14100	0,931	177000	- 14.100		818.000	100,0	0,177	0,911	38.200	1,02	10,8
III	0,657	261290	0,531	493330	261.290	29,0	638.480	71,0	0,650	0,591	90.918	0,71	12,0
IV	0,447	479590	0,965	826000	479.590	32	1.000.000	68,0	0,518	0,650	112.000	0,68	9,6
V	0,954	95500	0,941	- 19100	95.500	13,0	657.000	87,0	0,103	0,956	37.700	0,87	15,0
VI	0,824	88200	0,667	190000	88.200	16,0	455.364	84,0	0,447	0,740	60.700	0,84	16,0
VII	0,331	233250	0,124	454340	233.250	58,3	166.680	41,7	0,959	0,203	84.685	0,42	20,7
VIII	0,833	146000	0,874	50700	146.000	21,0	545.000	79,0	0,272	0,850	30.000	0,79	7,9
IX	0,628	260000	1,250	-166000	260.000	36,0	457.000	64,0	0,216	0,890	26.100	0,64	7,5
X	0,431	178000	0,758	328000	178.000	37,0	299.000	63,0	0,659	0,588	53.300	0,63	13,1
XI	0,514	153000	0,640	428000	153.000	28,0	402.000	72,0	0,671	0,575	64.100	0,72	13,5
XII	0,919	- 19694	0,887	590000	- 19.694		523.544	100,0	0,528	0,690	51.530	1,04	9,7
XIII	1,015	74100	0,287	463000	74.100	10,0	694.000	90,0	0,706	0,546	152.000	0,90	22,5

Source: Author's results based on data from the above enterprises.

Table 4a: Estimate of regression parameters of $Y_1 = aX_{17} + b$

Form of model (with standard errors of coefficients in brackets) and associated t-values	Corresponding 5 per cent value of t	Standard deviation of cost (in zloty)	Proportion of variation explained (R^2)	$1-R^2$	R	Autocorrelation coefficient	Durban-Watson d-statistic
$Y_{15} = 1774900 + 3.85X_{17}$ (518250) (0.56) 3.42 6.90	2.0739	243832	0.680	0.320	0.825	0.0683	1.6722
$Y_{15} = 220050000 + 2223000X_{40} + 2423400X_{44} - 195350000X_{47}$ (193060000) (3010700) (16269000) (188310000) 1.14 0.73 0.15 1.07	2.0860	23100000	0.097	0.903	0.311	-0.1307	2.2025
$Y_{15} = 798160 + 6.12X_{17}$ (241110) (0.35) 3.31 17.72	2.0739	189300	0.934	0.066	0.967	0.3704	1.1084
$Y_{15} = 5478330 + -52.1523X_{27}$ (4155747) (48.4471) 1.32 1.08	2.0739	7270900	0.004	0.996	0.061	0.2343	1.5508
$Y_{1} = -147210 + 34.443X_{24}$ (52377) (1.991) 2.81 17.30	2.0739	30760	0.932	0.068	0.962	0.962	1.1174
$Y_{1} = -22824000 - 60513X_{22} + 10818X_{31} - 12319X_{32} + 1563200X_{42}$ (6468400) (69696) (4006) (6268) (436120) 3.528 0.865 2.70 1,98 3.584	2.0930	108848	0.534	0.466	0.730		2.0414
$Y_{1} = 240708 + 19.625X_{24}$ (15063) (0.535) 15.98 36.68	2.0739	123400	0.813	0.187	0.902		0.9675
$Y_{8} = 421100 + 1.0615X_{18}$ (16917) (0.0348) 24.9 30.5	2.0739	54800	0.977	0.023	0.988		1.9419
$Y_{8} = 353990 + 0.4321X_{17}$ (71634) (0.0909) 4.94 4.75	2.0739	37258	0.500	0.500	0.708		1.8794
$Y_{8} = -92696 + 38.40X_{24}$ (29510) (1.03) 3.14 3.73	2.0739	245906	0.817	0.183	0.904		
$Y_{13} = -225270 + 42.22X_{21} - 4422X_{38} - 14681X_{43}$ (106230) (10.23) (234910) (12311) 2.12 4.13 0.02 1.19	2.0860	29900	0.621	0.379	0.788		2.5717
$Y_{13} = -111650 + 204.32X_{29} - 1044.7X_{46}$ (203720) (138.23) (504.6) 0.55 1.48 2.07	2 0796	20182	0.227	0.773	0.477		2.3673
$Y_{13} = -4752 + 1.36X_{35} + 5.21X_{24}$ (13768) (0.55) (0.70) 0.35 2.47 7.44							
$Y_{14} = -688700 + 168.61X_{20} + 33.05X_{25}$ (213160) (36.6) (16.79) 3.23 4.61 1.97	2.0739	64700	0.721	0.279	0.849		1.9026
$Y_{14} = -1253000 + 3955X_{31}$ (575980) (1340) 2.15 2.95	2.0739	97530	0.284	0.716	0.533		2.3513
$Y_{14} = 8.6125X_{29}^{0.6421}$ (0.2515) (0.0343) 8.64 18.7	2.0739	157047	0.531	0.469	0.729		1.3982

Source: Author's results based on data from the above enterprises.

Table 4b: Regression models in road haulage transport in Poland.

and the results of the analysis of cost behaviour.

First, the factors that were identified as affecting costs are shown in Table 3. Examples of the regression coefficients, and the goodness-of-fit are given in Tables 4a and 4b.

4. Economic criteria in management of a road haulage undertaking

In dealing with the assessment of the management of the road haulage industry, ideally we should like to maximise profits or minimise losses. To do this we must find answers to the following questions.

(1) What rates should the public haulage contractor charge for its services?

(2) What level of service should the public carrier offer to maximise profits (or minimise losses)?

(3) What cost and profit decisions should the public haulage contractor make?

In looking for the answers to these questions, it might be helpful to apply ex ante static analysis, known as break-even analysis, which assumes constant or changing selling prices (rates). The first kind of break-even analysis, usually called conventional break-even analysis, represents total revenue and total cost by straight lines, and assumes that profit is to be a linear function of service (or unit sales). This is the main reason why conventional break-even analysis cannot easily be used for the maximization of profit or the minimization of loss (see Figure 1).

Much more useful for these purposes is the second kind of break-even analysis, which does not assume that services and sales can be increased without changing rates and that the common carrier operates at the same efficiency at all levels. But even in this case there will be doubts about whether the public carrier will aim to maximize profits, because he may simply wish not to make a loss (see Figure 2).

Consider the above-mentioned break-even analysis. We can expect it to show us revenues, costs and profits for different levels of unit sales and different rates.

Thus, it is, in fact, perfectly possible that break-even analysis can easily be used to explore changes in the activity levels over time, because it shows the results of using different price-quantity relations.

In both analyses (conventional and extended to include other factors) it was necessary to make some substantive assumptions which, in fact, limit their applications in practice.

In spite of these constraints, break-even analysis has been used quite often to review and check past results in pricing and output decisions of the road haulage and other industries. It is interesting to add that the same analysis could be used to examine many practical problems of

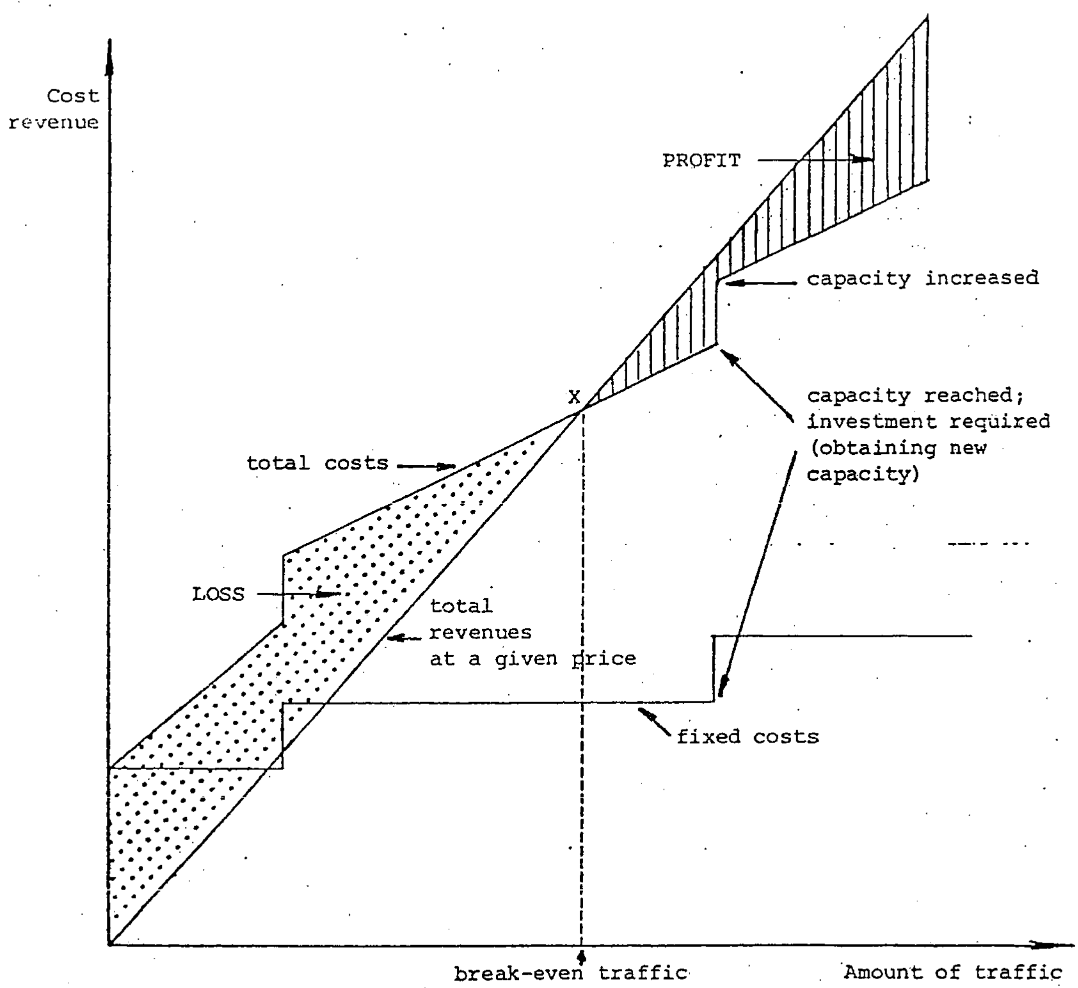

Figure 1: Conventional break-even analysis. Constant rates.

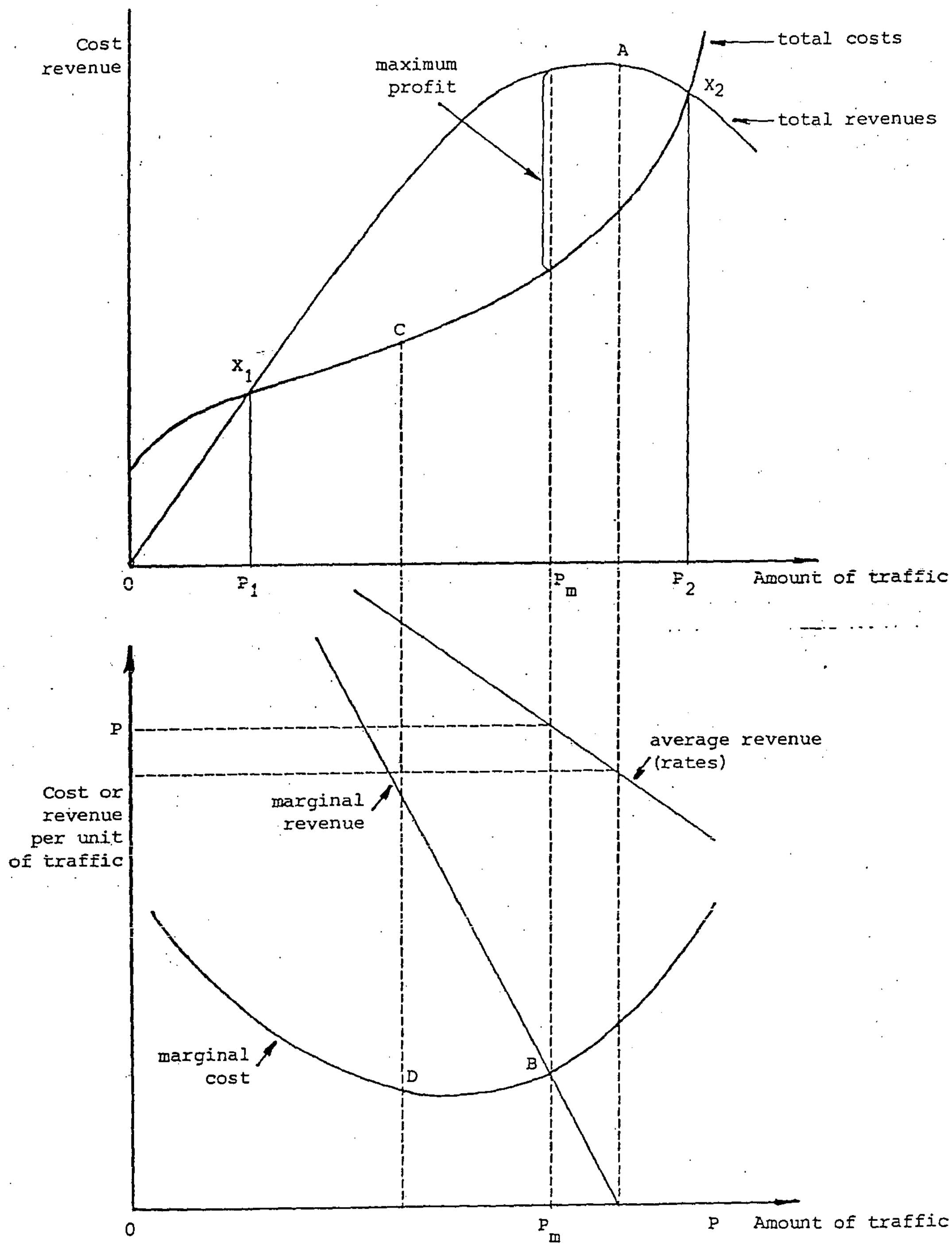

Figure 2: Break-even analysis. Changing rates.

investment decision making.

In practice, however, operational managers of hauliers prefer to draw break-even charts assuming the present rates for the sake of simplicity.

The lowest break-even point is determined by the rate equal to the variable cost per unit (service) and the total fixed costs. Quite frequently fixed costs do not enter at all into the present pricing decision because it could be difficult to find the sales volume which equates the true marginal cost (which is not just variable cost) and marginal revenue (optimal rate).

Other areas of application not dealt with here of these methods and results are:

(i) predicting of costs

(ii) limitation of costs

(iii) assessing the elasticity of costs and reliability of costs and profitability of transport services.

On the basis of broad material I have collected from 13 transport undertakings, I have demonstrated the following.

(1) There exists the possibility of selecting meaningful explanatory variables, which determine a high proportion of the total variation of costs.

(2) The costs of transporting consignments can be modelled by means of regression equations, mostly linear.

(3) Costs (total and components) can be divided into constant and variable parts, in spite of the considerable differences in the degree of their variability, especially in cross-section. There is also a pronounced seasonal pattern (observed over 4 quarters).

(4) Decision making in operation of vehicles can be improved with knowledge of relationships between operating cost and the quality of the management and operating staff.

5. References

[1] Zdzisław Hellwig, (ed.) *Elementy Rachunku Ekonomicznego*, Praca Zbiorowa, PWE, Warszawa, 1972.

[2] Bernard Kubiak, *Analiza czynników determinujących koszty przewozów ładunków w transporcie samochodowym*, Zeszyty Naukowe Wydziału Ekonomiki Produkcji, Uniwersytet Gdański, Gdańsk, 1975.

[3] Bernard Kubiak, *Modele kosztów w transporcie samochodowym*, *Informatyczne Systemy Rachunku Kosztów*, Warszawa, 1976.

[4] Bernard Kubiak, *Problems concerning the variability of costs in the road haulage industry*, Gdansk University, Gdansk, 1978 (Polish with English summary).

[5] Wiesław Pluta, *Metoda wyboru zmiennych objaśniających w modelach symptomatycznych*, "Przegląd Statystyczny", nr2, 1972.