Sixth International Conference on Construction in the 21st Century (CITC-VI)
"Construction Challenges in the New Decade"
July 5-7, 2011, Kuala Lumpur, Malaysia

Road Construction Cost Prediction Models Based on Regression Analysis

Evaggelia A. Antoniou, Georgios N. Aretoulis
Postgraduate Student, Lecturer, Aristotle University of Thessaloniki, Thessaloniki,Greece
eaantoni@civil.auth.gr, garet@civil.auth.gr

Panagiotis Papaioannou,
Professor, Aristotle University of Thessaloniki, Thessaloniki,Greece
papa@vergina.eng.auth.gr

Elsa Adamantidou
Ph.D, Lawyer, Legal Adviser of Public Procurement Monitoring Unit
Rechercher in Centre of International and European Economic Law, Greece
ea-law@hol.gr

Abstract

The ability to predict the final cost of construction projects based on limited initial input data could be a very valuable tool for every project manager and/or construction enterprise. This paper focuses on the models of Trefor P. Williams and their application in Greek road construction projects. An overview and description of each model is provided and also their performance is assessed. These models can predict with satisfactory precision the cost at completion of road construction projects based on initial tender offers. The study applies these models in 28 selected highway construction project cases conducted in central and northern Greece and discusses their performance. The analysis of the models is taking place in various groups of sample projects, based on projects' budgets and geographical locations.

Keywords

Prediction Models, Correlation Analysis, Regression Analysis, Final Cost Prediction

1. Introduction

During the implementation of a construction project, the meticulous "Project Evaluation" at the beginning loses much of its importance during latter stages because of various uncertainties related to the environment resulting in varying degrees of cost overruns. Therefore, particular attention is required to project cost control during project implementation (Nandi and Dutta, 1988). In general, projects that span over a long period of time tend to present time and cost overruns (Eyers, 2001). The goal in any project is to achieve construction quality with minimum time and cost (Papathanasiou, 2003). Studies on the relationship between estimated costs and actual costs of road construction projects has left many policy makers stunned by suggesting that cost overrun is prevalent in the sector and that the magnitudes may be large (Odeck, 2004). Efforts to assess the divergence between estimated costs and actual costs of roads are rare. Only a few studies exist that rigorously compare forecast and actual costs for large groups of highway infrastructure projects (Odeck, 2004).

In competitively bid construction projects, many factors can affect the completed cost of a project. Prices often escalate above the low bid price. The bid price developed by a contractor for a construction project will include the total of the net estimate with appropriate additions for overhead, profits, and risk margin. The low bid is prone to errors in assumptions made by the contractor, and many external events can affect costs during construction. So the ability to predict the final cost of construction projects could be a very valuable tool for project management (Williams, 2003). However, financial offers in competitive bids are strongly affected by the prevailing legislative framework. In the EU the existing framework is the one set by Directive 18/2004 about Public Procurement for public works, services and supplies. The specific directive sets the basic principles and rules which must be followed but at the same time provides adequate margins for its transfer to the respective national legislation about public works. More severe deviations are noticed in other countries worldwide and in the international construction market.

2. Trefor's Predicting models

Trefor P. Williams managed to create correlation models between the initial low bid and final value of a project. Specifically, there was a strong linear relationship between data when they were transformed using logarithms. An extensive statistical analysis resulted in some functional relations using data of roads and highways constructed in New Jersey, New York and Texas. The general form of this linear relationship is:

$$\mathrm{LN}(\mathrm{Y}) = \beta_1(\mathrm{LN}(\mathrm{X})) + (\beta_0) \text{ (Williams, 2003)}$$

Then follow the functional models for each group of projects (depending on their origin), which are summarized in the following table:

Table 1: Trefor's Models

New Jersey	$(\text{Final Cost}) = 0.75 \cdot (\text{Low Bid})^{1.018}$	**(Model 1)**
NJHA	$(\text{Final Cost}) = 1.589 \cdot (\text{Low Bid})^{0.964}$	**(Model 2)**
New York	$(\text{Final Cost}) = 0.984 \cdot (\text{Low Bid})^{1.007}$	**(Model 3)**
Texas	$(\text{Final Cost}) = 1.036 \cdot (\text{Low Bid})^{1.001}$	**(Model 4)**
Corps of Engineers	$(\text{Final Cost}) = 3.142 \cdot (\text{Low Bid})^{0.912}$	**(Model 5)**

That same year Trefor P. Williams published another model. For the compilation of this model he used economic data from previous, constructed projects that he has taken from the Ministry of Transport of New Jersey. With the same procedure of linear regression he produced the following functional relationship which links the low bid of a project and the construction duration (in days) with its final value (Williams, 2002).

$$(\text{Final Cost}) = 1{,}059 \cdot (\text{Low Bid}) + 998{,}27 \cdot (\text{Duration}) - 344118{,}6 \quad \textbf{(Model 6)}$$

This paper examines whether these models could be used to predict the final cost of Greek highway construction projects.

3. Williams' Models Performance for Greek Highway Projects

3.1 Application of the models at the whole sample

A sample of 28 road projects has been selected, that have been constructed in the central and northern Greece. The selected sample presents the following same features: all projects are related to the construction, improvement or enlargement of roads, flexible pavements have been used for the construction, all projects include the same type of work packages (solid works, structures, surfacing, asphalt, signage, lighting), they have been constructed over the past decade (2000-2010), they have a total length ranging from 3,5 km to 14 km, the project funding originated from the same sources and the payment was made in euro. The projects were selected based on their common characteristics. This fact reduces the sample size, but significantly increases the reliability and comparability among the projects' elements.

Firstly, the final price of the Greek projects is estimated, by applying the 6 predicting models of Williams T. P. The estimates are presented in the table below (Table 2). At the same time the actual - final price for each project is provided, in order to make obvious the comparison between the models' forecasts and the actual cost at completion.

Table 2: Models' Prediction and Actual Final Cost

Project	Final Cost	Model 1	Model 2	Model 3
1	6.031.994,70	5.501.481,67	5.039.238,38	6.084.560,59
2	2.114.313,20	1.873.063,29	1.816.598,59	2.095.840,32
3	6.464.234,92	6.143.609,14	5.594.555,89	6.786.644,01
4	3.383.222,76	3.221.588,26	3.035.872,16	3.583.693,73
5	2.103.540,42	2.032.079,36	1.962.320,81	2.271.768,19
6	2.146.332,73	1.808.252,52	1.757.020,54	2.024.091,17
7	297.781,77	359.139,28	380.205,32	409.090,43
8	255.764,42	240.017,28	259.586,31	274.593,41
9	534.659,12	418.824,56	439.790,29	476.285,19
10	3.940.700,93	3.579.184,64	3.354.073,98	3.976.957,89
11	2.500.000,00	2.568.432,74	2.449.635,05	2.864.127,40
12	12.747.331,73	12.834.474,71	11.239.528,44	14.065.407,57
13	2.998.794,64	2.941.663,76	2.785.483,18	3.275.521,48
14	1.371.500,00	1.326.559,09	1.310.329,93	1.489.879,97
15	1.876.681,21	1.670.433,62	1.629.946,39	1.871.424,21
16	1.645.299,00	1.963.947,52	1.899.961,78	2.196.409,27
17	1.230.344,36	1.242.617,09	1.231.678,32	1.396.589,47
18	5.835.800,63	5.276.015,87	4.843.455,87	5.837.837,71
19	4.090.000,00	3.490.757,87	3.275.552,49	3.879.752,39
20	601.200,00	572.931,34	591.695,39	649.332,59
21	448.410,63	482.152,58	502.520,92	547.467,89
22	413.086,26	413.086,26	445.975,36	483.284,49
23	222.430,48	219.537,62	238.562,86	251.405,67
24	1.580.088,69	1.535.480,71	1.504.974,40	1.721.800,01
25	590.975,79	611.131,99	628.989,85	692.144,44
26	504.359,64	543.232,00	562.609,70	616.027,03
27	322.097,65	338.863,06	359.847,34	386.236,54
28	402.152,34	403459,77	424.497,12	458.997,76

Project	Final Cost	Model 4	Model 5	Model 6
1	6.031.994,70	5.836.194,02	4.443.834,11	6.318.351,83
2	2.114.313,20	2.023.096,57	1.692.598,25	2.058.551,55
3	6.464.234,92	6.505.384,84	4.905.796,74	7.140.889,42
4	3.383.222,76	3.448.269,41	2.751.363,85	3.916.850,10
5	2.103.540,42	2.191.865,30	1.820.779,16	2.467.996,75
6	2.146.332,73	1.954.243,29	1.640.034,36	2.109.041,30
7	297.781,77	398.754,30	385.436,33	787.022,06
8	255.764,42	268.292,00	268.631,15	470.793,07
9	534.659,12	463.830,82	442.353,51	608.047,31
10	3.940.700,93	3.824.299,18	3.023.445,08	4.235.209,87
11	2.500.000,00	2.759.577,57	2.245.910,03	3.253.918,44
12	12.747.331,73	13.424.077,86	9.491.799,84	14.123.627,60

Table 2: Models' Prediction and Actual Final Cost (continued)

Project	Final Cost	Model 4	Model 5	Model 6
13	2.998.794,64	3.153.431,69	2.536.188,88	3.984.606,77
14	1.371.500,00	1.441.095,50	1.242.592,60	1.347.884,70
15	1.876.681,21	1.807.689,06	1.527.594,63	2.021.418,55
16	1.645.299,00	2.119.582,90	1.765.991,76	2.884.400,76
17	1.230.344,36	1.351.380,25	1.171.913,45	2.111.072,85
18	5.835.800,63	5.600.923,39	4.280.323,54	5.897.115,48
19	4.090.000,00	3.731.375,12	2.956.439,31	4.141.739,28
20	601.200,00	631.186,57	585.694,95	956.401,75
21	448.410,63	532.709,88	501.827,32	557.679,49
22	413.086,26	470.606,19	448.236,84	644.824,87
23	222.430,48	245.765,54	248.002,45	233.445,14
24	1.580.088,69	1.663.986,58	1.416.552,56	1.996.550,90
25	590.975,79	672.546,10	620.561,78	1.427.333,74
26	504.359,64	598.999,64	558.420,50	923.966,75
27	322.097,65	376.606,82	365.882,77	365.385,59
28	402.152,34	447.093,92	427.787,15	710.968,75

The next table presents in absolute values the percentage variation among the forecasts provided by each model and the actual final cost figure of the project:

Table 3: Models' Absolute Variation

Project	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
1	0,09	0,16	0,01	0,03	0,26	0,05
2	0,11	0,14	0,01	0,04	0,20	0,03
3	0,05	0,13	0,05	0,01	0,24	0,10
4	0,05	0,10	0,06	0,02	0,19	0,16
5	0,03	0,07	0,08	0,04	0,13	0,17
6	0,16	0,18	0,06	0,09	0,24	0,02
7	0,21	0,28	0,37	0,34	0,29	1,64
8	0,06	0,01	0,07	0,05	0,05	0,84
9	0,22	0,18	0,11	0,13	0,17	0,14
10	0,09	0,15	0,01	0,03	0,23	0,07
11	0,03	0,02	0,15	0,10	0,10	0,30
12	0,01	0,12	0,10	0,05	0,26	0,11
13	0,02	0,07	0,09	0,05	0,15	0,33
14	0,03	0,04	0,09	0,05	0,09	0,02
15	0,11	0,13	0,00	0,04	0,19	0,08
16	0,19	0,15	0,33	0,29	0,07	0,75
17	0,01	0,00	0,14	0,10	0,05	0,72
18	0,10	0,17	0,00	0,04	0,27	0,01
19	0,15	0,20	0,05	0,09	0,28	0,01
20	0,05	0,02	0,08	0,05	0,03	0,59
21	0,08	0,12	0,22	0,19	0,12	0,24
22	-	0,08	0,17	0,14	0,09	0,56
23	0,01	0,07	0,13	0,10	0,11	0,05
24	0,03	0,05	0,09	0,05	0,10	0,26
25	0,03	0,06	0,17	0,14	0,05	1,42
26	0,08	0,12	0,22	0,19	0,11	0,83
27	0,05	0,12	0,20	0,17	0,14	0,13
28	0,00	0,06	0,14	0,11	0,06	0,77
Average	**0,07**	**0,10**	**0,11**	**0,09**	**0,15**	**0,38**

The following graph (Figure 1) schematically indicates average values of the percentage difference between the prices of each model prediction and the actual final price of the project:

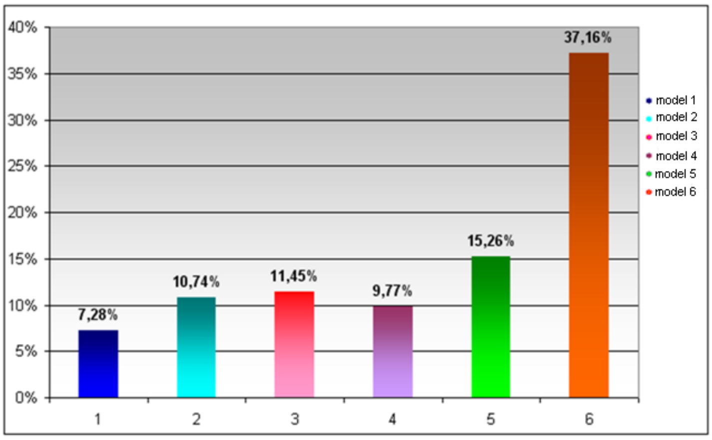

Figure 1: Models' Average Percentage Differences

From the above graph it becomes clear that model 6 is unable to adequately predict the final price of the 28 projects. This model, unlikely to the other 5 models, uses as input data not only the low bid price of each project, but also its duration. The large percentage variations, that the model 6 experiences, therefore, suggest that the duration of construction does not affect in the same way the final cost of the 28 highway projects.

However models 1, 2, 3, 4 and 5 show an average variation around or below 15%. Therefore, they perform quite well. Especially model 1 provides the final price with the average absolute percentage difference of 7.4%.

3.2 Application of the six models according to projects' budget

The project budgets of 27 out of 28 projects of the sample do not exceed 7 million €. Based on that fact the sample could be divided into 3 groups: the first group (including 10 projects with budgets up to 1.000.000 €), the second group (including 9 projects with budgets ranging from 1.000.000 € to 3.000.000 €) and the third group (including 8 projects with budgets ranging from 3.000.000 € to 7.000.000 €). The mean absolute percentage difference between the predicted price of each model and the actual final price of the project for each group has been estimated. The results are presented graphically in the following three figures (2, 3 and 4):

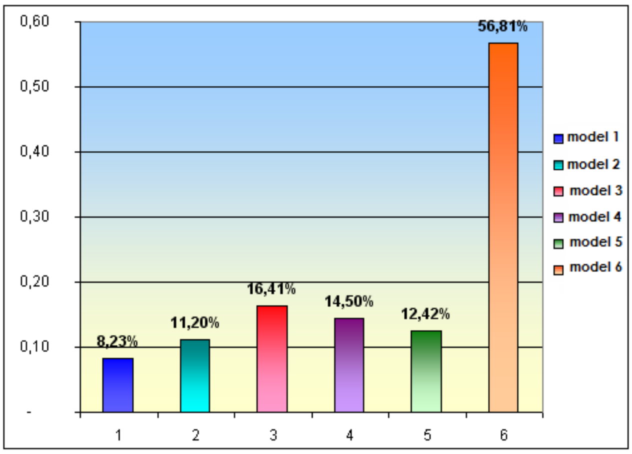

Figure 2: First Group Models' Absolute Percentage Differences

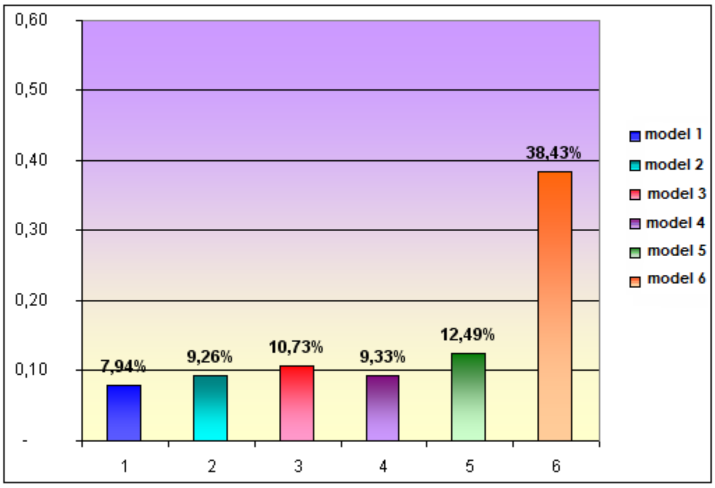

Figure 3: Second Group Models' Absolute Percentage Differences

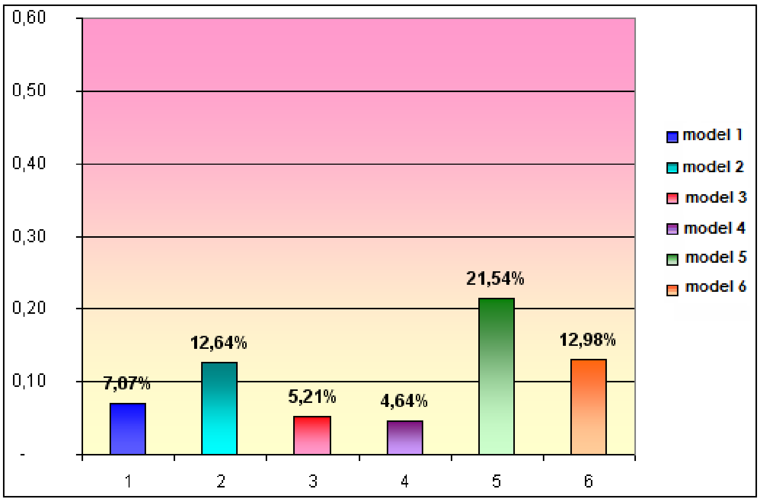

Figure 4: Third Group Models' Absolute Percentage Differences

From the previous diagrams it becomes obvious that the models, which make accurate predictions about the final price of each group are: model 1 for group 1 with a mean absolute difference 8.23%, again model 1 for Group 2 with a mean absolute difference 7.49% and model 4 for Group 3, with a mean absolute difference equal to 4.64%.

Furthermore, the functional models 1, 3 and 4 tend to present more accurate cost estimates as the cost of the projects increases. The same fact comes as a conclusion for the 6th model. This model apart from the original low bid price, also incorporates the duration of the project. This may be an indication that the duration of a project affects the final value when the initial cost is great.

Additionally the model 2 gives similar predictions (ranging from 9.26% to 12.64%) for all three groups, while the model 5 even though it produces almost identical predictions for groups 1 and 2 (12.49% and 12.99%), the 3rd group forecasts present an average absolute difference which is almost double in comparison to groups 1 and 2.

3.2 Application of the six models according to the geographical origin of each project

A further point of interest has been to examine whether the predictions of the 6 models of Williams T. P. are differentiated by the geographical origin of the data. As previously mentioned, the 28 projects used for this research were constructed in central and northern Greece. Therefore, the following two groups have been created in which the 28 projects were divided as follows: First group (including 15 projects of Northern Greece) and the second group (including 13 projects of Central Greece). The mean absolute percentage variation between the predicted and actual value has been estimated in each group. The results are shown graphically in the following two figures (5 and 6):

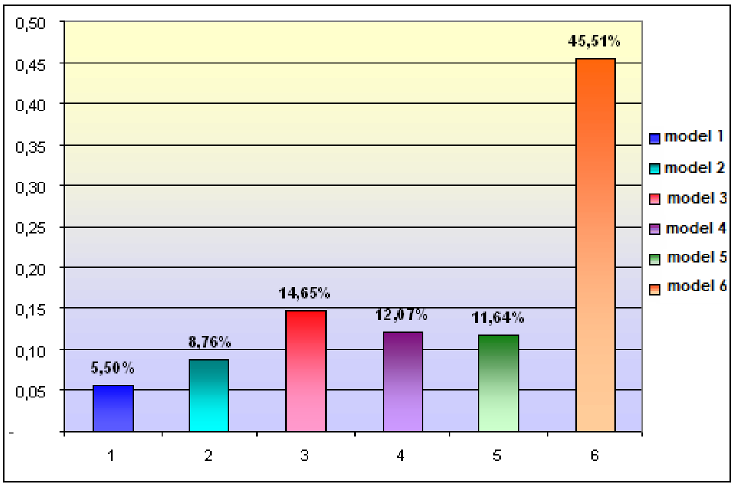

Figure 5: Northern Greece Projects' Absolute Percentage Differences

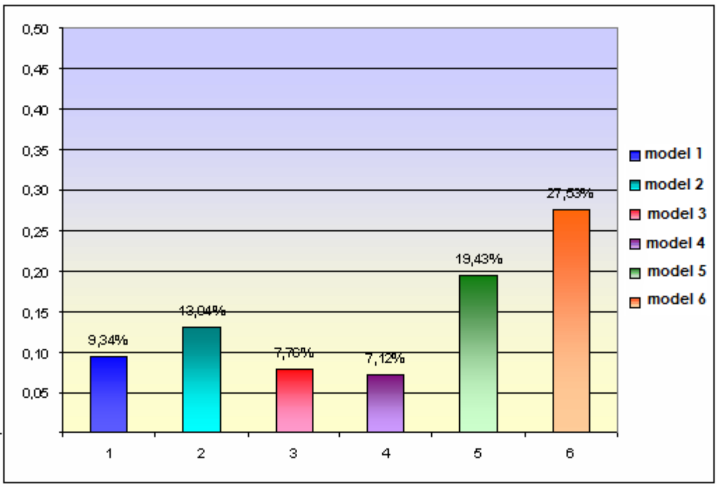

Figure 6: Central Greece Projects' Absolute Percentage Differences

Differences can also be observed in the accuracy of the predictions of the final cost by the six models. Specifically, in the case of Northern Greece projects, model 1 gave the most accurate predictions, while for the projects of Central Greece model 4 was the most accurate one.

4. Conclusions and Future Work

Generally, the selected models provided noticeable results. It seems that there is potential benefit in customizing them, in the Greek construction industry's environment. The three analyses conducted, revealed that from the six models which were examined, the first one exhibits the best overall performance, in predicting the final cost of road construction projects. On the other hand the sixth model could be characterized as the one exhibiting the least accurate predictions. The interesting fact about the sixth model is that it tends to increase its accuracy along with the increase in the project's budget. This is something which can be attributed to the parameter of project duration that was incorporated in the model. It can be roughly speculated that usually greater project budgets are related to greater project durations.

Unfortunately, the degree of adoption or development of such models for the Greek highway construction projects is not extended. The main reason is the lack of available and reliable cost data, which could be used as basis for the creation of the models. Public agencies are reluctant in providing cost data. As a result one of the first steps towards the development of similar models could include the organization of proper databases with historical cost data.

Another point to be considered in the applicability of Williams' models to Greek road projects is the reliability of the project budget in the procurement phase. The estimated project budget relies mainly on the preceding design studies and on the availability of similar historical data. The experience from past Greek projects is that the budget estimates have not been good enough in most cases. The reasons for that are many and definitely beyond the scope of this paper. However, the real difference between the initial cost estimate (budgeted) and the finally realized one would be potentially a good indicator and possibly an important input data in any effort to modify Williams models. Dummy variables reflecting the legal framework governing competitive bidding could be also another important factor that needs to be taken into account.

In conclusion, a number of factors that affect the performance of Williams models should be taken under consideration, when it come to apply them in another environment. These parameters are namely: τhe limited duration and budget of the projects selected, differences in the law and legislation concerning the construction of road projects and especially public projects, the funding approach and methods of payment for the constructed projects, the special characteristics of the domestic economy and last but not least the amount of contractors bidding for the project and their agenda

5. References

Eyers K. (2001). "Belief Network Analysis of Direct Cost Risk in Building Construction". M.S. Thesis, Department of Civil Engineering, University of Toronto, Canada.

Nandi S.N., Dutta B.k. (1988). "Trends in project Cost Control Systems". *OMEGA Int. J. of Mgmt Sci.*, Vol. 16, No. 6, pp. 513-520.

Odeck,J. (2004). "Cost overruns in road construction—what are their sizes and determinants?". *Journal of Transport Policy*, Vol. 11, No. 1, pp 43-53.

Papathanasiou B. (2003). "Concerning the Management of great Construction Projects". *Proceedings of the 16th Conference on Project Management*, Editors: Greek Company of Operational Research, Technical University of Larisa, Larisa, Volume I, pp. 479-482.

Williams, T., P., (2003). "Predicting final cost for competitively bid construction projects using regression models". *International Journal of Project Management*, Vol. 21, pp. 593–599.

Williams, T., P., (2002). "Predicting completed project cost using bidding data". *Journal of Construction Management and Economics*, Vol. 20, No.3, pp. 225-235.

EU Directive 18/2004