S15-3

TREE FORM CLASSIFICATION OF OWNER PAYMENT BEHAVIOUR

Hanh Tran[1], David G. Carmichael[2], Maria C. A. Balatbat[3]

[1] PhD candidate, School of Civil and Environmental Engineering, The University of New South Wales, Australia
[2] Professor, School of Civil and Environmental Engineering, The University of New South Wales, Australia
[3]Senior lecturer, Australian School of Business, The University of New South Wales, Australia
Correspondence to hanh.tran@student.unsw.edu.au

ABSTRACT: Contracting is said to be a high-risk business, and a common cause of business failure is related to cash management. A contractor's financial viability depends heavily on how actual payments from an owner deviate from those defined in the contract. The paper presents a method for contractors to evaluate the punctuality and fullness of owner payments based on historical behaviour. It does this by classifying owners according to their late and incomplete payment practices. A payment profile of an owner, in the form of aging claims submitted by the contractor, is used as a basis for the method's development. Regression trees are constructed based on three predictor variables, namely, the average time to payment following a claim, the total amount ending up being paid within a certain period and the level of variability in claim response times. The Tree package in the publicly available R program is used for building the trees. The analysis is particularly useful for contractors at the pre-tendering stage, when contractors predict the likely payment scenario in an upcoming project. Based on the method, the contractor can decide whether to tender or not tender, or adjust its financial preparations accordingly. The paper is a contribution in risk management applied to claim and dispute resolution practice. It is argued that by contractors having a better understanding of owner payment behaviour, fewer disputes and contractor business failures will occur.

Keywords: Contractual claims, uncertainty, risk, owner classification, regression tree, contractor payment

1. INTRODUCTION

Construction contractors have been shown to have a higher rate of failure compared to other businesses (Carmichael, 2002). Russell (1991) points out that over 60 percent of construction contractor failures involve an economic factor. Lack of liquidity for supporting their daily activities is the more common reason for failure compared to poor management of other resources (Navon, 1996). Effective cash management is therefore one of the more important strategies for the survival of contractors.

Generally, the contract establishes the way owners pay contractors for undertaking construction work. While special payments such as up-front or mobilisation payments may occur, commonly contractors pay for work performed and then invoice the owner based on this work. Claims are usually submitted on a regular basis, and assessed by the owner before being approved and paid. In such cases, the contractor's cash flow and financial status largely depend on how timely the payments are, and in fact, will be severely affected if payments from owners are late and/or incomplete (Carmichael and Balatbat, 2010). When payments are delayed or less than expected, cash for payments to suppliers and subcontractors is short, additional bank interest may be incurred, and if the amount has to go to a dispute resolution forum, extra delay time and cost associated with the dispute will be added. Not all contractors are financially able of withstanding a long dispute resolution time before recovering money owed (El-adaway and Kandil, 2009).

Uncertainties in payments have been addressed and successfully modelled as Markov chains in order to estimate the likelihood of late and incomplete payments by Carmichael and Balatbat (2010). The analysis in this paper develops from this state-of-the-art and presents a classification system of owners based on their payment characteristics. In effect, the analysis groups owners with similarities in their payment behaviour based on their payment history. The classification model addresses the key uncertainties to be considered in claim-payment analysis outlined in Carmichael and Balatbat (2010), namely (1) the delay time in receiving payment, (2) the proportion of a claim likely to be paid by the owner, and (3) the timing and proportion of subsequent instalments (if any) following initial payment.

Based on the contractor's knowledge of past payment practices of the owner, the contractor can determine which group that owner belongs to, and consequently, recognize its typical payment behaviour. Information about the owner-specific type can be combined with the results from the Markov chain modelling in Carmichael and Balatbat (2010) to serve as inputs in the contractor's cash flow estimation and financial analysis. The contractor, therefore, can benefit from more effective cash planning and management during a project. The analysis can also be part of the owner evaluation step when the contractor makes decisions before working with an owner. At the pre-tendering stage, the contractor can perform the analysis to decide whether to tender or not, and to adjust the tender price. Later on, contractors can choose to adopt the most appropriate strategy in collecting payments to suit the owner's payment characteristics. Having predefined typical payment behaviour is useful during tender preparation, which can be a hectic time.

1.1. Background

The issue of payments to contractors has been addressed in many studies. While "payment is important to most firms in the construction process" (Hinze and Tracey, 1994, p.283) and the timing of payments is a key contributor to a contractor's success, contractors face many variabilities and uncertainties in payments (Carmichael, 2002, Carmichael and Balatbat, 2010). The adverse influence of late and incomplete payments on a contractor's business and performance has been highlighted in the literature; for example, contractor non-payment is a cause of disputes escalating (Carmichael 2002, 2010), and "the risk of late payments ... is very common in the industry and has driven many consulting firms to the edge of bankruptcy." (Kometa et al., 1996, p. 275) The situation is similar for subcontractors or specialist contractors. The survey by Hinze and Tracey (1994) shows that the majority of subcontractors are not satisfied with the percent withheld by their general contractors, and (p. 283) "that retainage on ... a subcontractor's work could be substantial and the interest that might be generated would not be trivial." El-adaway and Kandil (2009) propose a risk retention measure for the contractor by buying insurance to relieve the financial and economic burdens of construction claims. Variabilities in payments to contractors have been quantitatively modelled in studies about cash flow forecasting. Chen et al. (2005) use a cost-schedule integration technique in cash flow calculations, which make extensive use of project estimate and schedule data, including payment time. However, these authors admit that an extension of the model to include more detailed payment conditions, various payment time lags and payment frequency is needed. The cash flow model introduced by Park et al. (2005) allows users to take into account uncertainties caused by delayed payments due to owner's circumstances. Other works concerning payment conditions and cash flow modelling include Navon (1994, 1995), Chen and Chen (2005) and Chen et al. (2005). The seminal contribution of Carmichael and Balatbat (2010) adopts information about late and incomplete payments in a completed project to predict payment likelihood in the future. The study takes summaries of outstanding project money against time and models them as a Markov process to estimate the amount uncollectible and the likelihood of payments to be delayed and/or incomplete. These results can be fed into cash flow and accounting risk estimates by contractors when planning new projects.

The idea of classifying, or credit rating, an owner in terms of payment punctuality from the contractor's point of view is new. Owners are often grouped according to their nature, for example, public sector and private sector owners (Flanagan and Norman, 1993), or based on their needs (Chinyio et al., 1998). Examples of studies about owner payment practice and their effects on a contractor's performance and project outcome include Kometa et al. (1996), Kaka (1996), Shash (1993), and Ahmah and Minkarah (1988). It is noted that none of these studies attempt to categorise owners into different types according to their payment histories.

Despite the lack of a systematic evaluation procedure for owner practices, factors relating to an owner's financial capability have always been highly ranked in contractors' decision making. Enshassi et al. (2010) rank an owner's financial capabilities as the most important factor among owner-related factors in a contractor's decision to tender. Similarly, an owner's ability to pay is the top ranked factor by Odusote and Fellows (1992). Shash (1993) ranks owner identity, past profit, and cash flow among the top influencing factors in a contractor's project selection decision. The owner factor is the third important factor in tendering decisions of US contractors (Ahmah and Minkarah, 1988).

1.2. Tree form classification

The present analysis uses a credit rating technique called a classification and regression tree (CART). Introduced by Breiman et al. (1984), a classification and regression tree is a non-parametric technique that produces either classification or regression trees, depending on whether the target variable is categorical or numerical, respectively. CART gives the final classification in a simple form which requires no complicated skill to use but still efficiently classifies new data. Compared to other well-known statistical classification techniques such as discriminant analysis and regression analysis, CART achieves high accuracy in classification and prediction. A study by Lee et al. (2006) shows that CART gives a significantly higher average correct classification rate than that given by linear discriminant analysis and logistic regression. CART is user-friendly and suitable for the contractor's analysis purpose.

The tree form classifier is constructed by repeating binary splits of subsets of an owner population X into two descendant subsets. The owner population X contains a variety of payment profiles ranging from excellent to very poor practice. The construction of a tree revolves around three elements: the selection of the splits; the decisions when to declare a node terminal or to continue splitting it; and the assignment of each terminal node to a class. Each terminal node of the tree can be labelled according to different groups of owners, for example, group 1 to 10 or 1 to 5 ranging from the best to the worst practice respectively.

2. MODEL PRELIMINARIES

2.1. Form of payment data

The method adopts the form of payment profile summaries outlined in Carmichael and Balatbat (2010), that is in terms of total outstanding amounts in weeks after lodgement of claims. Payment profiles can be summarised for individual claim types (for example, progress claims, variation claims, and extension of time claims), and combinations of these.

The progress claim list of a road construction project, costing approximately $114M over a duration of approximately 2.5 years, is used to demonstrate the approach. Table 1 gives selective summaries of outstanding amounts for listed claims at different periods (weekly) following lodgement of claims.

Total claimed amount ($K)	Outstanding amount ($K) at week							
	1	2	3	4	5	6	7	8
114,563	114,563	109,873	98,554	19,344	13,446	13,446	13,446	13,446

Table 1. Selective summaries, case study example.

The outstanding amounts in Table 1 can be expressed as proportions of the total claimed amount and are plotted in Figure 1.

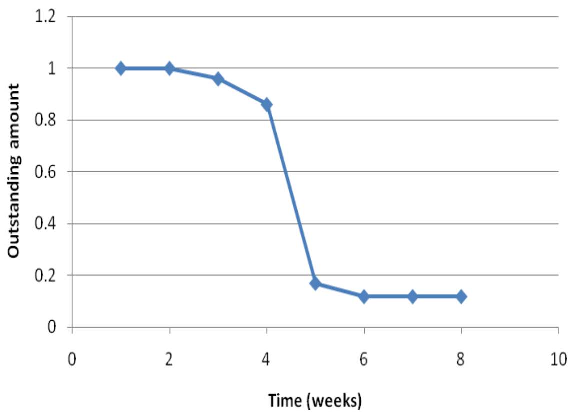

Figure 1. Payment profile, case study example.

The horizontal axis shows the number of weeks that have passed since the time of submitting the claim and the vertical axis shows the outstanding amounts as proportions of the total project value. The limit of 8 weeks duration in the calculation is chosen for illustration purposes only. It is assumed that payments due beyond 8 weeks are to be resolved between the contractor and the owner, may go to formal dispute, and may not be recovered by the contractor. Here 8 weeks represents the time at which the contractor concedes that the payment may not be forthcoming, or the time at which the contractor might instigate dispute proceedings. This time is denoted n in the following. Time intervals other than weeks can be used in the analysis.

A payment profile plot as in Figure 1 contains important information about how the owner responds to contractor's claims on the project. In such a payment profile plot, the payment uncertainties are represented directly or indirectly through the following parameters of the plot:

1. The proportion of the total amount getting paid within the calculation timeframe (here 8 time periods or weeks), calculated by taking the difference between 1 and the last outstanding amount (as a proportion of 1).
2. The average time (response time) taken for the initial payment to be made. This value is the length of the flat part of the curve.
3. The consistency in processing time among claims, represented by the slope of the line.

The paper's tree classification uses these three 'shape' parameters as its predictor variables. These are referred to as y_0, t_0 and α, respectively, below.

2.2. Assumptions

Three assumptions are made. Firstly, since this analysis is for the contractor's internal purposes, all claims made by the contractor are assumed valid; to do otherwise would not help the contractor. Only delays and incompleteness in payment due to the owner initiated causes are included in the payment profile summaries. Secondly, time periods are chosen as weeks, and for definiteness in the calculations, an allowable payment duration is taken to be 8 weeks. Any outstanding amount beyond 8 weeks is considered as needing to be resolved. The choice of time period and duration is for illustration purposes and it does not affect the method. Thirdly, for transparency of calculations, unit claims are used. This means that the outstanding amounts are expressed as proportions of the actual claim. Actual outstanding amounts can be calculated by multiplying the corresponding outstanding values with the actual claim amount.

2.3. Variables

Three shape parameters of the payment profile plot to represent the uncertainties in the payments are used as predictor variables in the tree classification:

y_0 is the proportion of total amount paid by the owner in the 8-week period. It is the total claim amount less the outstanding amount at week 8. For example, y_0 = 0.9 means that if the total amount of the claim submitted is 1, then in 8 weeks the total amount paid by the owner is 0.9.

t_0 is the length of the flat part of the probability curve (of the form of Figure 1) before dropping. t_0 is the time following the submission of the claim to the initial payment made by the owner. t_0 ranges from 0 to 7 weeks.

α represents the downward slope of the payment profile plot. Since the payment profile is a summary of individual claims, the parameter α

represents the consistency in the promptness in responding to each claim. If a payment profile plot has a significantly large downward slope (steep), then that owner tends to take almost the same amount of time to process and pay claims. Conversely, payment profiles with a small downward slope (almost flat) show that the owner responds to claims at very different times.

Among the three variables, y_0 could be thought of as the most important because the total amount recovered over the allotted duration may dominate the thinking of the contractor. However, t_0 and α, the time related variables, are also important because of issues related to interest on funds, cash flow and present worth (net present value).

2.4. Training sample sets

Training or test sets of data are first used to establish a relationship or model from the predictor variables. For this purpose, a data set containing payment profiles for the incomplete payment case (y_0 < 1) and a data set for the complete payment case (y_0 = 1), in the form of outstanding amounts against time, are generated.

Incomplete payment case

The data set chosen contains 273 samples covering most owner payment possibilities in terms of proportion, delay time, and promptness. Values of the predictor variables appearing in the sample are as follows:

y_0 = 0.1, 0.2, 0.3, …, 0.9. y_0 = 0.9 represents 90% of a claim being paid within 8 weeks, whereas y_0 = 0.1 represents only 10% of the claim ending up being paid. By letting y_0 run from 0.1 to 0.9, the sample covers the possibilities that the owner pays almost nothing right up to 90% of the claim within the allotted duration. Because the analysis targets incomplete payments, a y_0 value greater than 0.9 is not included.

t_0 = 0, 1, 2, …, 7. Since the allotted time period is 8 weeks, a delay time of 8 weeks represents no payment. Only integer weeks are used.

α = 1 to 5, where:

1	Response time to claims is almost the same - profile plot has a slope close to 90 degrees to the time axis.
2	Response time to claims is highly consistent - the slope of the profile plot is steep.
3	Response time to claims varies a little - the slope of the profile plot is approximately 45 degrees to the time axis.
4	Response time to claims varies somewhat - slope of the curve is around 60 to 75 degrees to the time axis.
5	Response time to claims varies considerably - almost flat payment profile curve.

To evaluate the quality of a payment profile, a target variable MARK is used. MARK represents the contractor's opinion on each owner payment profile. A higher MARK means a more desirable payment behaviour. MARK is the basis for the construction of the classification model and also is the result of the prediction. In regression analysis and tree modelling, there can be only one target variable for each sample.

In the sample set, MARK is calculated as a function of y_0, t_0, and α as follows:

$$\text{MARK} = 10y_0 + (10 - t_0) + 2\alpha \qquad (1)$$

In Equation (1), the y_0 term is on a scale of 1 to 10 (the factor 10 is applied because actual y_0 values are from 0.1 to 1); the t_0 term is on a scale of 1 to 10 (10 marks are awarded if there is no time delay in payment; for each week late, one mark is deducted); α is also on a scale of 1 to 10 (the factor 2 is applied because α actually is from 1 to 5). The magnitudes of the multipliers are not important; it is the relative ranges of the three terms which are of concern. This choice of function gives the three predictor variables the same range of possible values, implying they are treated as having equal significance.

A linear function is used because it is the simplest type of function and serves the purpose well. Given the calculation of MARK as in Equation (1), the lowest possible MARK in the sample set is 5.1 and the highest possible MARK is 29.

Complete payment case

If the contractor chooses to extend the allotted time period in the analysis, it is possible that 100% of the claimed amount will end up being paid. For example, the contractor may choose to use n = 12 weeks instead. In such cases, even though all the claims are paid in full and the payment profile has y_0 = 1, the timing of the instalments still matters to the contractor's cash flow. An owner who pays 100% quickly should have a higher rank than one that pays late.

Of the three predictor variables used for classification in the incomplete case, y_0 is now equal to 1, leaving t_0 and α as predictor variables for classification. The implications and value ranges of these variables are the same as described above. The data set chosen contains 40 sample payment profiles covering 8 possible values of t_0 (from 0 to 7) and 5 possible values of α (from 1 to 5). The nominal MARK for each payment profile is calculated using Equation (1). Since y_0 is now a constant, 10 y_0 is also a constant, thus the inclusion of y_0 in the calculation of MARK merely shifts the range of MARK by 10 points and does not alter the classification (and as expected, y_0 does not appear in the splitting of the tree). The reason for leaving y_0 in the calculation is so that the MARK values for payment behaviour in this case are of the same order as the MARK values for the incomplete payment case. The 40 samples have MARK values varying from 15 to 30, that is from the worst scenario to best scenario, respectively.

3. MODEL DEVELOPMENT

3.1. Splitting rules

The construction of classification trees is done using the Tree package in the R program (R Development Core Team, 2009; Ripley, 2009).

With MARK estimated as a function of all of the predictor variables as defined in Equation (1), the tree is constructed through binary recursive partitioning, whereby the data set is successively split along coordinate axes of the predictor variables so that, at any node, the split which maximally distinguishes the response variables in the left and the right branches is selected. Splitting continues until nodes are 'pure' (node members have the same properties) or the data are 'too sparse' (in R, the default value of the minimum number of cases in each node is 6). Refer the R manual and Maindonald and Braun (2003).

Each possible split based on each predictor variable (y_0, t_0, α) is assessed in turn, and the split explaining the greatest amount of the deviance in MARK is selected. Deviance is calculated on the basis of a threshold in the predictor variable; this threshold produces two mean values for the response (one mean above the threshold, the other below the threshold). For a given predictor variable (say, y_0), the procedure of splitting given by Maindonald and Braun (2003) and according to the R program manual is as follows:

- Choose a threshold value of the predictor variable.
- Calculate the mean value of the target variable above and below this threshold.
- Use the two means to calculate the deviance.
- Go through all possible values of the threshold (that is, values on the y_0 axis).
- Determine the value of the threshold which gives the lowest deviance.
- Split the data into high and low subsets on the basis of the threshold for this variable.
- Repeat the whole procedure on each subset of the data on either side of the threshold.
- Keep going until no further reduction in the deviance is obtained, or there are too few data points to merit further subdivision. Then a full tree is constructed.
- Prune the tree accordingly to the desired number of terminal nodes or other criteria.

The R program defines the node deviance as

$$D = \sum_j (MARK_j - \mu_{|j|})^2 \qquad (2)$$

where $\mu_{|j|}$ is the mean of all the values of MARK assigned to node j, and the squared differences are summed over all the nodes. At each step, the split is chosen so as to achieve the maximum reduction in D.

3.2. Tree pruning

The procedure in R is that a full tree is constructed and then pruned to the required number of terminal nodes. A full tree, which perfectly fits the data, is the one with as many terminal nodes as there are data samples. Such a tree, therefore, has no explanatory power because it does not group objects with similar properties together. Also, trees too large may be over-elaborated and may respond to random features of the data (Maindonald and Braun, 2003). In the present case, a trial tree of 10 terminal nodes is used because the number 10 is considered a good starting point for the number of owner groups. Trees having more than 10 terminal nodes are not being looked at because they give too many groups of owners, thus deviating from the main purpose of the tool.

The total tree deviances computed for different tree sizes are used to guide the pruning so that the least significant nodes are pruned to reduce the tree size. (The size of the tree is defined as the number of terminal nodes.) For each tree size, the total deviance of the tree is calculated (by summarising all node deviances).

Figure 2 shows the decrease in total tree deviance from almost 7000 to well below 1000 as the size of the tree size increases from 1 to 10.

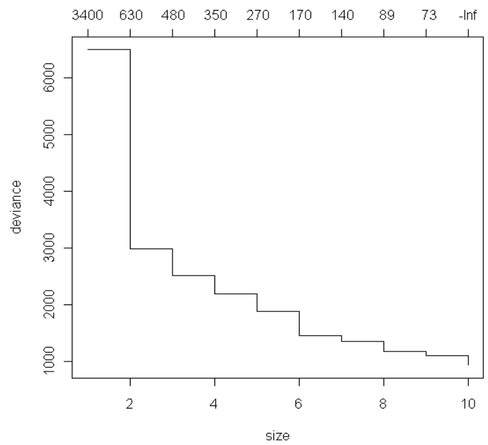

Figure 2. Total tree deviances versus tree sizes.

4. CLASSIFICATION RESULTS

It is noted that when the tree size grows to more than 6 terminal nodes, the tree deviance continues to decrease, however, at a much lower rate than that in the first 5 drops. The total deviance of a 6-node tree is over 1800 and gradually drops to 935.3 when the tree has 10 nodes. Figure 3 gives the average MARK values of the terminal nodes of a 6-node tree for the incomplete payment case.

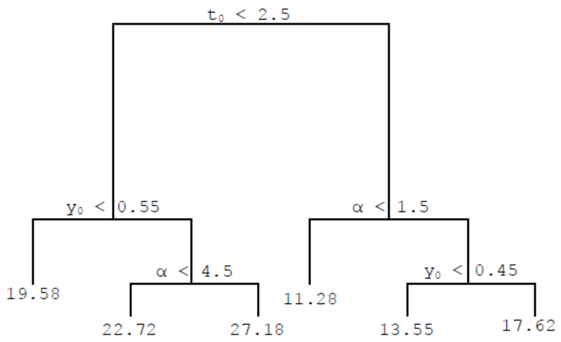

Figure 3. 6-node tree for the incomplete payment case.

A classification tree for the complete payment case with 4 terminal nodes is shown in Figure 4.

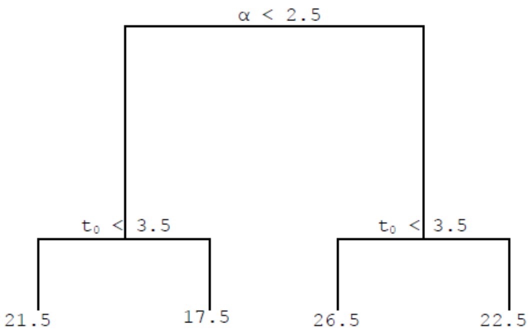

Figure 4. 4-node tree for the complete payment case.

5. OWNER GROUPS

5.1. Incomplete payment case

The incomplete payment tree in Figure 3 divides 273 samples of payment behaviour into 6 groups with relatively distinctive average MARK values, implying that this is a good grouping of owners. Owners belonging to this group did not pay 100% of all the claims on their past projects (there may be some claims paid in full, but other claims were not 100% paid). The six groups are labelled I1 to I6 according to six levels of practice from poor to excellent. The prefix I refers to 'incomplete'. Owners in the incomplete payment case can fall into one of the following types:

Group I1: MARK = 11.28 – Poor. This is the group with the lowest MARK, which is equivalent to having the poorest payment practices among the 6 groups. Owners in this group on average pay about 54% (standard deviation of 24%) of the total amount claimed but their delay in payments is substantial. Claims are usually not paid until the 6th, 7th or 8th week after submission. This type of owner could be regarded as strongly undesirable to work with.

Group I2: MARK = 13.55 – Poor. This group contains payment profiles in the low desirability range. The mean y_0 value of this group is 0.27, which is much lower than that of group I1. That is, on average, only 27% (standard deviation of 10%) of the total claim amount is likely to be paid. The reason that profiles in this group have higher average MARK values than those in Group I1 is the lower t_0 values (implying more prompt payments) and higher α (processing time is almost the same for different claims). In short, owners in this group usually pay 4 weeks after receiving the claims.

Group I3: MARK = 17.63 – Medium. Group I3 contains profiles which are better than those in Group I2 in terms of y_0, which is the total paid proportion. Owners in this group on average pay 70% (standard deviation of 14.5%) of what is claimed. The average processing time is also 4 weeks and does not vary greatly between claims.

Group I4: MARK = 19.58 – Medium. On average, owners in this group have higher MARK values than those in group I3 because they are much more responsive to claims. Some claims can be paid within 1 or 2 weeks after being lodged, but variability in the processing time of claims is large (mean α = 3.45). However, the average amount ending up being paid is low, only 34% (standard deviation of 12.5%).

Group I5: MARK = 22.72 – Good. The average paid proportion by owners in this group is 76% (standard deviation of 11%), which is the highest among owner groups I1 to I6. Owners in this group are usually very responsive to claims; on average, their claims are paid within 1 or 2 weeks after receiving an invoice. However, their promptness in responding to claims is not as consistent as those in group I6, resulting in a lower average MARK. Still, owners in this group could be considered as desirable to work for.

Group I6: MARK = 27.18 – Excellent. This group has the most desirable payment behaviour. Owners in this group on average pay 74% (standard deviation of 11.4%) of the claim amount very quickly. Most of them will pay 90% or more of the claim amount. These owners make prompt payments for all claims, usually in the first weeks after receiving the claim.

5.2. Complete payment case

The complete payment tree in Figure 4 divides 40 samples of payment behaviour of the complete payment case into 4 groups with average MARK values of 17.5, 21.5, 22.5, and 26.5. Owners belonging to these four groups have always paid 100% of their claims to the contractors in the past, therefore, they are preferred over group I1 owners. There are 4 subgroups labelled C1 to C4 according to four levels of practices. The prefix C refers to 'complete'. Owners in the complete payment case can fall into one of the following types:

Groups C2 and C4: Owners respond quickly to claims (t_0 mean = 1.5), that is, on average, they pay claims within 1 to 2 weeks after receiving an invoice. They either have the response time to individual claims almost the same: α = 4 – Group C4; or some claims which are paid very promptly while some are paid very late: α = 1– Group C2.

Group C1 and C3: Owners respond slowly to claims (t_0 mean = 5.50), that is, on average, claims are paid 5 to 6 weeks after invoicing. They either have the response time for individual claims almost the same: α = 4 – Group C3; or some claims which are paid very promptly while some are paid very late: α = 1 – Group C1.

5.3. Classification for the case study profile

The case study data are represented by the payment profile in Figure 1. Using the plot and the definitions of the predictor variables y_0, t_0, and α, their values are determined as follows:

y_0 = (33 - 23.59)/33 = 0.285

t_0 = 2

α = 1

According to the tree in Figure 3, this payment profile falls into Group I4 which has an average MARK of 19.58. The description for this group is:

> Group I4: MARK = 19.58 – Medium. On average, owners in this group have higher MARK values than those in Group I3 because they are much more responsive to claims. Some claims can be paid within 1 or 2 weeks after being lodged, but variability in the processing time of claims is fairly large (mean α = 3.45). However, the average amount ending up being paid is low, only approximately 34% with a standard deviation of 12.5%.

The description given is relatively close to the actual patterns of the claim list in Tables 1 and Figure 1, implying an accurate classification.

5.4. Approach for contractor

The classification for the case study data shows that with pre-established payment behaviour groups, the remaining work for contractors is straightforward. In order to perform a classification, the steps to be taken by the contractor are as follows:

- Use information from projects which are similar to the upcoming project in terms of type, financial conditions, payment terms etc. For such projects, choose a relevant time period (days, weeks, months) and the number of time periods (n) that pass before a claim could be classed as needing to be resolved.
- Summarise total outstanding amounts against the plot of the payment profile. From that calculate:
    - The proportion of the total claimed amount ending up being paid by the owner within n time periods (y_0 value).
    - The time following the claim lodgement that the owner waits before making the first payment (t_0).
    - The consistency in processing time of claims, by selecting the value of α which most closely represents the payment rate of the owner in the scale earlier.
- Follow the splits in the regression tree in Figures 3 or 4 to determine which group the owner belongs to and find the description of the typical payment behaviour of that group. The typical payment behaviour of owners belonging to that group, as well as the recommendations, may be taken as a guide and should be used with other information on the owner. This information then supports the contractor's decision making, for example, whether to contract with such an owner, or by how much the contractor's tender price should be increased to allow for late and incomplete payments from the owner.

6. CONCLUSION

The tree form classification model provides a simple and practical way for a contractor to classify an owner's payment practice based on information about payments on past projects. The calculations can be performed in a few simple steps and the tree is ready to use. It requires simple processing of the data which the contractor already has in hand. The analysis can be done for individual claim types (for example, progress, variation, delay or latent conditions), or combinations of claim types depending on the contractor's wish. The choice of time periods and allowable duration is up to the contractor and the tree is still valid.

Future work. The analysis presented in this paper could be extended by considering different ways of allocating MARK by changing the priorities of the predictor variables to suit a contractor's opinions. There could be more than one classification tree to serve different analysis purposes. For each owner group, a typical payment profile can be nominated to feed into the Markov chain modelling of Carmichael and Balatbat (2010) so that financial estimations can be made. Next, it may serve as inputs to financial analysis tools such as cash flow forecasting and other statistical inference processes to use for prediction purposes. The methodology could also be converted to a spreadsheet tool requiring only user inputs of summaries of payment data and choice of weightings of the predictor variables. The analysis is one part of complete owner evaluation procedures for contractors, subcontractors, and consultants.

REFERENCES

Ahmad, I. and Minkarah, I. (1988), Questionnaire survey on bidding in construction, *Journal of Management in Engineering*, Vol. 4, No. 3, pp. 229-243.

Balatbat, M. C. A., Lin, C.-Y. and Carmichael, D. G. (2010), Comparative performance of publicly listed

construction companies: Australian evidence, *Construction Management and Economics*, Vol. 28, No. 9, pp. 919-932.
Breiman, L. J., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984), *Classification and Regression Trees*, Chapman and Hall, New York.
Carmichael, D. G. (2002), *Disputes and International Projects*, A A Balkema, Rotterdam.
Carmichael, D. G. and Balatbat, M. C. A. (2010), A contractor's analysis of the likelihood of payment of claims, *Journal of Financial Management of Property and Construction*, Vol. 15, No. 2, pp. 102-117.
Chinyio, E. A., Olomolaiye, P. O., Kometa, S. T. and Harris, F. C. (1998), A needs-based methodology for classifying construction clients and selecting contractors, *Construction Management and Economics*, Vol. 16, No. 1, pp. 91-98.
El-Adaway, I. H. and Kandil, A. A. (2009), Contractors' claims insurance: A risk retention approach, *Journal of Construction Engineering and Management*, Vol. 135, No. 9, pp. 819-825.
Enshassi, A., Sheif, M. and El Karriri, A. (2010), Factors affecting the bid, no bid decision in the Palestinian construction industry, *Journal of Financial Management of Property and Construction*, Vol. 15, No. 2, pp. 118-142.
Flanagan, R. and Norman, G. (1993), *Risk Management and Construction*, Blackwell, Oxford.
Hinze, J. and Tracey, A. (1994), The contractor-subcontractor relationship: the subcontractor's view, *Journal of Construction Engineering and Management*, ASCE, Vol. 120, No. 2, pp. 274-287.
Kaka, A. P. (1996), Towards more flexible and accurate cash flow forecasting, *Construction Management and Economics*, Vol. 14, No. 1, pp. 35-44.
Kometa, S. T., Olomolayie, P., O. and Harris, F. C. (1996), A review of client-generated risks to project consultants, *International Journal of Project Management*, Vol. 14, No. 5, pp. 273-279.
Lee, T. S., Chiu, C. C., Chou, Y. C. and Lu, C. J. (2006), Mining the customer credit using classification and regression tree and multivariate adaptive regression splines, *Computational Statistics and Data Analysis*, Vol. 50, No. 4, pp. 1113-1120.
Maindonald, J. and Braun, J. (2003), *Data Analysis and Graphics Using R - An Example-Based Approach*, Cambridge University Press.
Navon, R. (1996), Company-level cash-flow management, *Journal of Construction Engineering and Management*, Vol. 122, No. 1, pp. 22.
Odusote, O. O. and Fellows, R. F. (1992), An examination of the importance of resource considerations when contractors make project selection decisions, *Construction Management and Economics*, Vol. 10, No. 2, pp. 137-151.
R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
Ripley, B. (2009), *Tree: Classification and Regression Trees*, R package version 1.0-27.
Russell, J. S. (1991), Contractor failure: analysis, *Journal of Performance of Constructed Facilities*, Vol. 5, No. 3, pp. 163.
Russell, J. S. and Zhai, H. (1996), Predicting contractor failure using stochastis dynamics of economic and financial variables, *Journal of Construction Engineering and Management*, Vol. 122, No. 2, pp. 183-191.
Shash, A. A. (1993), Factors considered in tendering decisions by top UK contractors, *Construction Management and Economics*, Vol. 11, No. 2, pp. 111-118.