Portfolio Management Method for Deadline Planning

Rachel M. Fewster
Department of Statistics
University of Auckland
Private Bag 92019
Auckland
New Zealand
r.fewster@auckland.ac.nz

Emilia Mendes
Department of Computer Science
University of Auckland
Private Bag 92019
Auckland
New Zealand
e.mendes@auckland.ac.nz

Abstract

We introduce a portfolio management method that uses effort estimates to build sets of feasible deadlines for software projects at the bidding stage. Effort estimates can involve considerable error, and this must be taken into account when selecting deadlines. We show how a simple probability model can allow for possible errors. The model is built using a single effort estimate for each current project, together with historical data on estimated and actual effort for former projects.

The probability model is used in two ways: firstly to find the probability of successfully meeting a set of proposed deadlines; and secondly to select deadlines that deliver a fixed probability of success. Rather than treating projects in isolation, we work with the full company portfolio, enabling setbacks in some projects to be balanced by gain in others. Our method is implemented with demonstrations in the tool PROJMAN.

Key Words: *effort estimation; project management; regression model; risk analysis; risk management.*

1. Introduction

Project management is a key issue for software companies, as contract bids must be competitive yet realistic. Small software companies frequently bid for contracts in application areas where they have little or no experience, and the time available for effort estimation can be minimal. Simple, easy to use methods are essential to assist with managing the company portfolio [11].

This paper outlines a simple portfolio management method to provide decision support for project managers. The portfolio approach means that decisions are based on the full set of company projects, allowing a trade-off between projects requiring more effort than expected and those requiring less [12]. We focus attention on effort, as it is the largest and least predictable component of a project's costs. The method can be used for portfolios of any size or composition.

The aims of our portfolio management method are firstly to determine the feasibility of a set of proposed project deadlines, and secondly to suggest a new set of deadlines to attain a desired probability of successful completion. The central idea is that all effort estimates involve error. Rather than trying to eliminate the errors, we model them explicitly and use them as the basis of our feasibility assessments. To this end, we represent errors, or uncertainties, by probability distributions. An estimated effort of 20 person-days, for example, is converted into a probability distribution centered on the value 20, and including a range of values to either side of 20. Exactly how far to either side depends upon the probability distribution chosen, so the model should first be calibrated to historical data. For assessing or setting deadlines, we take into account the range of effort values that might occur for each of the projects in the portfolio.

The only requirement for our method is a record of previous projects. No models for effort estimation are necessary, although they can be incorporated if desired. Although we do require an effort estimate for every project, it need be nothing more sophisticated than subjective opinion. We do not discourage subjective estimates as they reflect the current practices of most small software companies [11].

Our portfolio management method is implemented in the tool PROJMAN, written in the statistical language R. PROJMAN is downloadable, with demonstrations, from `http://www.stat.auckland.ac.nz/~fewster/projman/`. In the following section, we describe how to build, calibrate, and verify a probability model for effort using historical project data. In Section 3 we explain the PROJMAN tool. Demonstrations with industrial portfolio data are provided in Section 4.

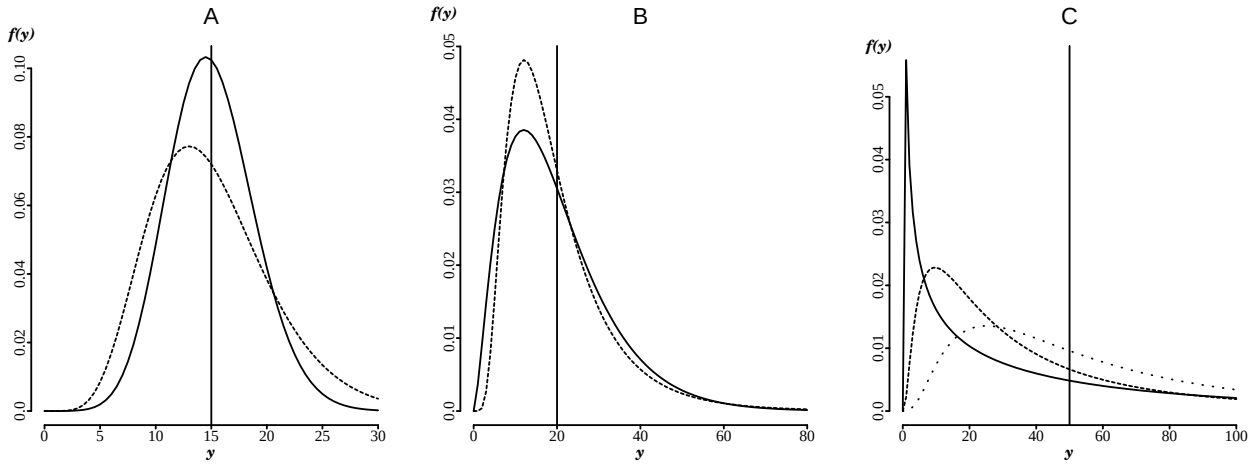

Figure 1. Probability density functions, $f(y)$, for a range of distributions. Part (a) : Poisson distribution (solid line), and Chisquare distribution (dashed line), both with mean = 15. Part (b) : Gamma distribution (solid line), and Lognormal distribution (dashed line), both with mean = 20 and dispersion = 0.4. Part (c) : Gamma distribution (solid line), and Lognormal distribution (dashed line), both with mean = 50 and dispersion = 2; and Lognormal distribution with median specification (dotted line), with median = 50 and dispersion (variance / median^2) = 2.

2. Probability model for project effort

2.1. Probability distributions

The basis of our portfolio management method is a probability model for effort. An effort estimate can be readily formed, but as important as the estimate itself is some idea about the variability or uncertainty associated with the estimate. A probability model aims to quantify this uncertainty. Probability models may be derived in many ways, ranging from statistical regression procedures to subjective opinion. Here we explain what is meant by a probability model and suggest possible ways of constructing one.

The effort required to complete a new project is an unknown quantity Y that may be described by a probability distribution. Probability distributions are characterized largely by three quantities: center, variance, and skewness. The center of a distribution refers to the distribution mean, or sometimes the median. The variance describes the spread of the distribution: diffuse or concentrated. The skewness describes the distribution shape: whether possible values are distributed evenly on either side of the mean (skewness=0), or whether they occur further to one side of the mean than the other. In the context of software projects we expect some positive skew in the distribution of Y, meaning that the probability that the effort will be much higher than estimated is greater than the probability it will be much lower [6].

We focus on six families of probability distributions: the Normal, Poisson, Chisquare, Gamma, Lognormal, and Lognormal-median. The shape of a probability distribution is determined by its family, and is pictured using the probability density function. Examples of distributional shapes are shown in Figure 1. All Normal distributions are symmetric (skewness=0). Poisson and Chisquare distributions have positive skew when the mean is small, but become more symmetric as the mean increases. Chisquare distributions have greater variability and retain greater skew than Poisson (Figure 1(a)). Gamma and Lognormal distributions have pronounced positive skew that can be controlled independently of the mean (Figure 1(b) and (c)). For low mean and variance, the Gamma and Lognormal are quite similar (Figure 1(b)). For high mean and variance, the Gamma distribution has a spike close to zero and a very long upper tail, whereas the Lognormal distribution remains more rounded (Figure 1(c)).

When a distribution has pronounced positive skew, the mean is much greater than the median. The median is the half-way point of a distribution: half of the time a random value will fall below the median, and half of the time it will fall above. The mean, on the other hand, describes the average that would be obtained from a large sample of random values. For highly skewed data it is sometimes more appropriate to describe a distribution in terms of the median than the mean. We consider the Lognormal-median family for this reason. The distributions in this family are ex-

actly the same as those in the Lognormal family, but they are specified in terms of the median rather than the mean. Figure 1(c) shows that the median specification yields a distribution centered more obviously on the stated value (median=50) than the usual specification with mean=50.

To build a probability model for the project effort, Y, we first form an effort estimate on which to center the probability distribution of Y. Throughout this paper, the 'center' of a probability distribution refers to the median for the Lognormal-median family, and to the mean for all other families. The effort estimate could be as simple as a manager's opinion, or it could be the output from a predictive model, based for example on multivariate regression or on machine-learning techniques [2, 10].

Next a family of probability distributions is selected, based on the shape and spread. With some families, such as the Poisson and Chisquare, the single value for distribution center is sufficient to specify the whole distribution. In these cases, the original effort estimate and choice of family make the probability model complete.

With other families, the center specifies only the location of the distribution, and the variability is controlled independently. For these families we need a way of deciding upon the variability for a given center. We do this by devising a rule linking the variance and the center, where the same rule is assumed to hold for all projects. We give this rule the name *dispersion*, a term borrowed from statistical regression modeling. The dispersion is therefore defined to be a function of the distribution variance and center that is assumed constant for all projects.

We have chosen a range of dispersion relationships to associate with the six distribution families detailed here. Our dispersion definitions give a spectrum of modeling options, although many other dispersion choices are also possible. Table 1 shows the chosen dispersion relationships. The Poisson and Chisquare families have fixed dispersion values of $\phi = 1$ and $\phi = 2$ respectively. For the other families, dispersion is constant for all projects but its value must be estimated, usually from data on past projects.

The dispersion for the Normal family is defined to equal the variance, so the amount of variability in Y is constant, regardless of the estimated effort. This is unlikely to be realistic unless all effort estimates are broadly similar, as there is usually less uncertainty associated with an effort estimate of 10 person-days than with one of 100 person-days. It is more realistic to assume that variability increases with estimated effort. The Poisson and Chisquare dispersions fix the ratio of variance to mean, so uncertainty increases for larger projects. Variance is equal to the mean for the Poisson ($\phi = 1$), and equal to twice the mean for the Chisquare ($\phi = 2$). The Gamma and Lognormal dispersions fix the ratio of variance to the square of the mean, so uncertainty can increase dramatically for larger projects.

Table 1. Distributions and dispersions.

Distribution	Dispersion, ϕ	Value
Normal	$\phi = \text{var}$	estimated
Poisson	$\phi = \text{var}/\text{mean}$	$\phi = 1$
Chisquare	$\phi = \text{var}/\text{mean}$	$\phi = 2$
Gamma	$\phi = \text{var}/\text{mean}^2$	estimated
Lognormal	$\phi = \text{var}/\text{mean}^2$	estimated
Lognormal-median	$\phi = \text{var}/\text{median}^2$	estimated

With the Lognormal-median family, the dispersion fixes the ratio of variance to the square of the median. It can be shown that this implies the ratio of variance to the square of the mean is also constant, but we use the median specification for ease.

The choice of distribution family, the effort estimate for distribution center, and the dispersion estimate together specify the probability model entirely. We use the general notation $Y \sim D(e)$, where Y denotes the unknown effort, e denotes the effort estimate or distribution center, and D summarizes the choice of distribution family and dispersion estimate. In the next sections we describe how the estimates of center e and dispersion ϕ are derived, and how the validity of the probability model D is checked.

2.2. Estimating center and dispersion

Effort estimates are used in our probability model only as the central value around which the effort distribution is spread. Consequently, accurate estimates are not crucial, and any method of effort estimation is valid. The subjective effort estimates currently favored by many companies are just as suitable as more formal prediction outputs.

If formal effort estimates are preferred, one possibility is to use a statistical regression model [2, 1]. Estimating effort in this way has the advantage that dispersion, in the definitions used above, is obtained directly from the regression software for certain families and models. Statistical regression is a vast and general framework that allows effort to be formulated as a function of predictor variables. The parameters of the predictor function, which govern the precise contribution of each predictor, are estimated by the statistical method of maximum likelihood. Essentially, this means choosing the parameter values that make our past observations most likely.

A simple example is ordinary least squares regression (OLS), where a typical regression equation for effort, Y,

would be

$$Y = \beta_0 + \beta_1 \times \mathit{size} + \epsilon\,.$$

Here, β_0 and β_1 are parameters to be estimated, *size* is a measure of product size, and ϵ denotes a random error that is assumed to have a Normal distribution with mean 0 and unknown variance σ^2 constant across all projects. The parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are obtained by fitting the model to data from previous projects.

Using OLS regression to estimate effort defines the whole probability model for Y. If a new project has size s and requires effort Y, then Y is assumed to have a Normal distribution with mean $(\hat{\beta}_0 + \hat{\beta}_1 s)$. The dispersion is the estimated variance, $\hat{\sigma}^2$, given directly by the OLS software.

A second possibility is OLS regression with log-transformed response:

$$\log(Y) = \beta_0 + \beta_1 \times \mathit{size} + \epsilon\,,$$

where $\epsilon \sim \text{Normal}(0, \sigma^2)$ as before. It can be shown that this model implies that the probability model for Y belongs to the Lognormal-median family. A project with size s yields a Lognormal distribution for Y, with median $e^{(\hat{\beta}_0 + \hat{\beta}_1 s)}$, and dispersion = variance/median^2 = $e^{\hat{\sigma}^2}(e^{\hat{\sigma}^2} - 1)$.

Beyond OLS regression is a large selection of other regression models. Generalized Linear Models (GLMs) allow non-linear relationships between predictors and effort response, and can model Y with a Normal, Gamma, or Poisson distribution, among others [2]. Once again, dispersion estimates are obtained directly from the fit.

When a formal effort model is not used, and either the Normal, Gamma, Lognormal, or Lognormal-median family has been selected, a dispersion estimate is still required. This can be obtained from historical data by the maximum likelihood method. As with formal effort models, this means that we use the errors in effort estimation that have been observed for completed projects to reflect the likely uncertainty for a new project. In the case of the Normal family, the dispersion equals the variance and is estimated by the mean squared difference between past observations and estimates. For the other three families, estimating the dispersion requires a numerical maximization. R code and demonstrations for dispersion estimation are available from the PROJMAN website.

Effort modeling can be as simple or complicated as required. It is straightforward to write code for maximum likelihood estimation of mean/median and dispersion quantities using a wide range of probability families and predictor functions. All regression modeling involves the selection of a probability distribution, D, which is then used directly in our portfolio management method. The effort estimates obtained from formal models are used by our method in exactly the same way as subjective estimates derived from guesswork.

2.3. Choosing a distribution family

The choice of distribution family for the probability model is central to the success of our method. Different families allow for different levels of variability in Y, and have different amounts of skew. Checking the suitability of the selected family is therefore an important part of the modeling process. Statistical tests for distribution checking are called goodness-of-fit tests.

The probability model for effort Y is formulated as $Y \sim D(e)$. We aim to test the suitability of the choice of distribution family, D. The test is conducted using m completed projects, with estimated efforts $e_1, \ldots, e_m$, actual efforts $y_1, \ldots, y_m$, and estimated dispersion ϕ. It checks whether the observations $y_1, \ldots, y_m$ could reasonably have come from the m distributions characterized by D, with separate centers $e_1, \ldots, e_m$ and common dispersion ϕ. If so, D is assumed also to be suitable for new projects.

Two commonly used statistical goodness-of-fit tests are the Chisquare test and the Kolmogorov-Smirnov test (KS) [4]. One problem is that the tests are designed to check whether a random sample of observations come from a single hypothesized distribution, but in our context we wish to test whether the m observations $y_1, \ldots, y_m$ come from the m different distributions $D(e_1), \ldots, D(e_m)$. We overcome this problem using a transformation of the data as follows.

Let F_i be the cumulative distribution function of distribution $D(e_i)$, defined by $F_i(y) = P$(a random value from distribution $D(e_i)$ is $\leq y$). The function F_i is provided by standard statistical software for all the distribution families we consider. It can be shown that, if $Y \sim D(e_i)$, then $F_i(Y) \sim \text{Uniform}(0,1)$, the Uniform distribution on the interval $(0,1)$. If the choice of distribution family D is correct, this implies that $F_1(y_1), F_2(y_2), \ldots, F_m(y_m)$ are all drawn from the Uniform$(0,1)$ distribution. We can therefore apply the standard Chisquare or KS goodness-of-fit test by testing the sample $F_1(y_1), F_2(y_2), \ldots, F_m(y_m)$ against the Uniform$(0,1)$ distribution.

The Chisquare test operates by binning the data into categories and counting the number of points in each category. If $F_1(y_1), F_2(y_2), \ldots, F_m(y_m)$ are genuinely Uniform$(0,1)$, then roughly 25% of the observations should fall between 0 and 0.25, roughly 25% between 0.25 and 0.5, and so on. If the observed frequencies are very different from those expected, we have evidence that distribution D is not appropriate for these data. The strength of evidence against distribution D is measured by the p-value returned by the test. Small p-values (typically < 0.05) provide strong evidence against D, whereas large p-values provide no evidence against D. The KS test works on a different principle, by comparing the empirical distribution function of the data $F_1(y_1), F_2(y_2), \ldots, F_m(y_m)$ against that of the Uniform$(0,1)$ distribution. Again, evidence against D

is summarized by the p-value.

R code and demonstrations for testing goodness of fit are included on the PROJMAN website. In practice, the tests will perform best when the number m of completed projects is large. Reasonable results may be obtained for $m = 15$ or more.

The goodness-of-fit test on historical data concludes the construction and calibration of a probability model for effort. Next we describe how the model is applied for portfolio management.

3. Portfolio management method

3.1. Portfolio success rate

Once a probability model is in place for effort estimation, it can be used to make decisions about bids and deadlines for future projects. Two applications of the model are suggested. The first is to test the feasibility of a set of proposed deadlines for new projects. The user inputs a proposed schedule, and the model outputs a probability of success. The second application performs the reverse operation: the user inputs a desired probability of success, and the model outputs a possible schedule that achieves that probability.

The company portfolio is defined as a set of n projects under current consideration. These can include projects that are ongoing, new, or still at the bidding stage. For testing deadline feasibility, each project has a specified deadline by which it must be finished. We use d_i to denote the deadline for project i, measured in number of working days from the start day. Projects are labeled from 1 to n according to deadline, so that project 1 has the earliest deadline and project n has the latest. The estimated effort for project i is e_i, measured in person-days. The effort spent on project i to date, in person-days, is written t_i. The company has E employees available for the n projects.

The unknown effort in person-days that will be taken to complete project i is denoted by Y_i. The uncertainty in Y_i is described by the probability model previously constructed. For an unstarted project, Y_i is assumed to be distributed according to the chosen distribution D (one of Normal, Poisson, Chisquare, Gamma, Lognormal, or Lognormal-median), with center e_i and dispersion ϕ, where ϕ is estimated from previous projects. If project i is ongoing, Y_i represents the additional effort needed for project i, given that effort t_i has already been spent.

Given a set of effort estimates $e_1, \ldots, e_n$, and proposed deadlines $d_1, \ldots, d_n$, our aim is to establish whether the deadlines are realistic by finding the probability of portfolio success. For the portfolio to be successful, all projects must be completed by their deadlines. Thus project 1 must be completed by deadline d_1; both projects 1 and 2 must be completed by deadline d_2, and so on. The total available effort in person-days from the start day to deadline d is $E \times d$. We can therefore summarize the deadline criterion as follows:

$$\begin{pmatrix} Y_1 \\ Y_1 + Y_2 \\ \vdots \\ Y_1 + Y_2 + \ldots + Y_n \end{pmatrix} \leq E \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} .$$

The criterion requires all projects $1, 2, \ldots, i$ to be completed by deadline d_i, rather than just project i. This is the key to the portfolio approach, because uncertainty about completion of project i incorporates uncertainty from all previous projects. If $Y_1 + \ldots + Y_i > Ed_i$ for any i, we say that deadline d_i has failed.

It is generally too complicated to find a formula for portfolio success probability, but an accurate value can be obtained in seconds by simulation. A large number of values of $Y_1, \ldots, Y_n$ are generated from their probability distributions. For each set of random Y values, the deadline criterion is checked and the set is recorded as a success or a failure. The estimated probability of success is the proportion of successes observed. Generating 10 000 sets of Y-values will estimate the portfolio success probability with a maximum margin of error of 1%.

If project i is unstarted, standard statistical software can be used to generate a random value for Y_i. If project i is already underway, we need to generate Y_i from the conditional effort distribution for project i, given that the effort must be greater than the value t_i already expended. We can show that such a sample can be obtained from

$$Y_i = F^{-1}\left[\left\{ 1 - F(t_i) \right\} U_i + F(t_i) \right] ,$$

where $F(y)$ is the cumulative distribution function of the unconditional effort distribution for project i (i.e. $F(y) = P(\text{project } i \text{ takes total effort} \leq y)$), and U_i is a uniform random number between 0 and 1. The functions F and F^{-1} are both provided by standard statistical software. Alternatively, managers might prefer to revise their original effort estimates for projects in progress, and sample from a revised unconditional distribution.

The portfolio success rate method is implemented in our portfolio management tool, PROJMAN. Table 2 gives an example application. A portfolio of ten projects is shown, with a range of values for estimated effort, effort to date, and deadline. There are $E = 4$ employees, and the Chisquare distribution has been selected for the probability model. From 10 000 simulations on this portfolio, PROJMAN returns for each deadline the probability of failure and the probability that the deadline is the first to fail. The second probability is the better indicator of 'problem' deadlines, because all previous projects succeeded.

Table 2. Example of the PROJMAN portfolio management tool. A portfolio of 10 projects is shown with estimated effort, e_i (person-days); deadline, d_i (number of working days from start); effort to date, t_i (person-days); and whether or not the deadline is fixed (y/n). The next two columns give the output from 10 000 PROJMAN simulations with E=4 employees using the Chisquare distribution. P(fail) is the probability of failure for deadline d_i, and P(fail 1) is the probability that deadline d_i is the first to fail. The overall probability of portfolio success is 41.07%. The P(fail 1) column sums to 58.93% = 100%–P(overall success). The last three columns show a second trial with a new set of deadlines. The first five projects have fixed deadlines, but the last five open projects are allotted new deadlines to attain an overall probability of success of about 90%. Each open project has been given a first-fail probability of 1.08%.

					Test deadlines		Set and test new deadlines		
Project	Estimate e_i	Deadline d_i	Effort to date t_i	Fixed (y/n)	P(fail) (%)	P(fail 1) (%)	New Deadline d_i	P(fail) (%)	P(fail 1) (%)
P1	20	8	15	y	0.03	0.03	8	0.08	0.08
P2	40	10	35	y	1.37	1.34	10	1.34	1.29
P3	10	12	0	y	2.73	1.54	12	2.56	1.40
P4	20	15	12	y	2.66	1.08	15	2.71	1.19
P5	20	20	5	y	1.96	0.48	20	2.21	0.65
P6	40	25	0	n	25.76	21.61	41.1	3.64	1.08
P7	20	40	0	n	0.43	0.00	25.0	3.11	1.08
P8	20	40	0	n	5.51	0.23	30.2	3.53	1.08
P9	50	45	0	n	49.51	25.59	54.8	3.66	1.08
P10	60	60	0	n	49.63	7.03	70.7	3.80	1.08
					P(overall success) : 41.07%		P(overall success) : 89.99%		

The portfolio in Table 2 has a relatively poor probability of success, with only 41.07% of the 10 000 simulations seeing success for all projects. The riskiest projects are P6 and P9, which would both benefit from amended deadlines. Project P10 has a high probability of failure (49.63%), but from the first-fail probability of 7.03% we see that failure in P10 follows mainly from delays in earlier projects.

Note that in principle it is possible for a deadline to have a high success rate, despite being shorter than the estimated effort for the corresponding project. This is because PROJMAN assumes that projects can be tackled by more than one employee concurrently. If no joint work is possible, the tool should be run separately for each employee's personal portfolio, using $E = 1$.

3.2. Setting deadlines

The second application of our portfolio management method is to suggest a set of deadlines that give a desired probability of overall success. The results are not unique, as there will be many different possible schedules that attain a given success probability, but in many cases a useful guideline can be obtained. The method assumes that joint work is possible on projects.

For this application, a distinction is necessary between the projects for which deadlines have already been agreed, and projects for which deadlines are to be decided. We label the projects as fixed or open: new deadlines will be suggested only for the open projects, although all projects are considered in ensuring that the portfolio meets the specified probability of success.

The desired success probability is written as p. The allowed failure rate is therefore $1 - p$. First we must find the failure rate f for the fixed projects. If this is already greater than $1 - p$, we will not be able to choose the open deadlines for an overall success probability of p. If f is less than $1 - p$, however, this leaves an allowed failure rate of $1 - p - f$ to be shared among the open projects. PROJMAN chooses the open deadlines so that this allowed failure is shared equally between them. If there are m open projects, this means that each open deadline is selected to have probability $(1 - p - f)/m$ of being the first deadline to fail.

To implement this approach we need to construct a priority, or ordering, on the projects. By default, PROJMAN orders fixed projects according to their deadlines, and open projects according to their estimated effort. The user can override the default by specifying an alternative priority. Different orderings produce different results.

Results from the deadline selection routine on the example portfolio are shown in the last columns of Table 2. The overall portfolio success rate is set at 90%. The deadlines obtained are dependent upon the choice of probability model, in this case Chisquare. Different models give different results, so it is important to use the goodness-of-fit tests to verify that the chosen probability distribution gives a realistic impression of effort uncertainty. Further demonstrations of the PROJMAN tool with deadline selection are given on the PROJMAN website.

4. Portfolio management for industrial data

4.1. Tukutuku data.

We demonstrate the PROJMAN tool using 32 Web application projects from the Tukutuku database [9]. Each project is described by 32 variables, from which effort may be estimated. Actual effort upon completion is also known, and ranges from 1 to 223 person-days.

Of the 32 projects, 12 are from Company M, a small company with $E = 5$ employees. We use the Company M data as our company portfolio to demonstrate the PROJMAN tool. The observed efforts for Company M can be used to test with hindsight whether the deadlines suggested by PROJMAN are realistic. To apply PROJMAN to the Company M projects, we also require a separate dataset of completed projects to build and calibrate a probability model. For this we use the remaining set of 20 projects from the Tukutuku database, collected from 13 different companies but similar to those in the Company M portfolio. These historical projects are referred to as Set A.

To show how Company M might use the PROJMAN tool over a period of time, we divide the Company M data at random into two groups of six projects: Group I and Group II. To begin with, we assume that all projects are unstarted, but the Group I projects are under consideration. At a later stage, we assume that the Group I projects have been completed and the Group II projects are under consideration. The observed efforts from Group I are used to update the PROJMAN system and provide new output for Group II.

Figure 2 illustrates the steps of the demonstration. The first step is to derive a model for effort prediction using the projects in Set A. The characteristics of the Tukutuku data suggest two appropriate prediction frameworks [15]: case-based reasoning (CBR), and stepwise regression (SWR). We demonstrate both techniques. Our aim is not to find the best effort model but to show how the PROJMAN tool is used in the two frameworks.

The output from the effort model is used in two ways. Firstly, the model is applied to the Group I projects to estimate efforts $e_1^I, \ldots, e_6^I$ for input to PROJMAN. Secondly, the estimated efforts $e_1^A, \ldots, e_{20}^A$ for Set A, together with the observed efforts $y_1^A, \ldots, y_{20}^A$, are used to construct a probability model $y_i^A \sim D^A(e_i^A)$, as in Section 2. This involves selecting a distribution family for D^A, estimating the dispersion parameter if necessary, and testing goodness of fit. This step differs according to the chosen effort prediction framework. When stepwise regression is used, the distribution D^A is automatically specified by the regression model. For the case-based reasoning approach, D^A must be fitted separately.

Finally, the Group I effort estimates $e_1^I, \ldots, e_6^I$ and the probability model D^A are used as inputs to the PROJMAN system to build or test sets of plausible deadlines for the Group I portfolio. This ends Phase 1 of the demonstration. In Phase 2, the completed Group I projects are used to augment the data from Set A to provide a new set B of historical data. The effort and probability models are refitted to Set B and the new models are used with the Group II projects for input to PROJMAN.

4.2. Case-based reasoning framework (CBR)

Case-based reasoning is a method of effort estimation based on analogy with completed projects [17]. The user identifies a set of variables that influence project effort. For Web applications, suitable variables include number of Web pages, number of images, and so on. A reference set of completed projects is maintained with data on each variable and a record of total effort at completion. Effort is estimated for a new project by finding projects in the reference set that are most similar to the new project on the basis of the chosen variables. The effort values for the similar projects are combined to generate an effort prediction for the new project, typically using a weighted or unweighted mean.

Decisions for the user include the choice of the initial set of variables, how to use these variables to assess similarity between projects, how many reference projects to extract for each effort estimate, and how to combine the reference effort values into a single new effort prediction. With the Tukutuku data, we identified 24 variables for the CBR analysis out of the 32 available. These comprised size variables, such as number of Web pages, text pages, images, animations, and audio/video files, with separate records for new, reused, and outsourced features; diversity variables, indicating the number of different features involved in the application; and variables counting numbers of high effort and low effort features of the application. Similarities were assessed using the unweighted Euclidean distance, with all variables

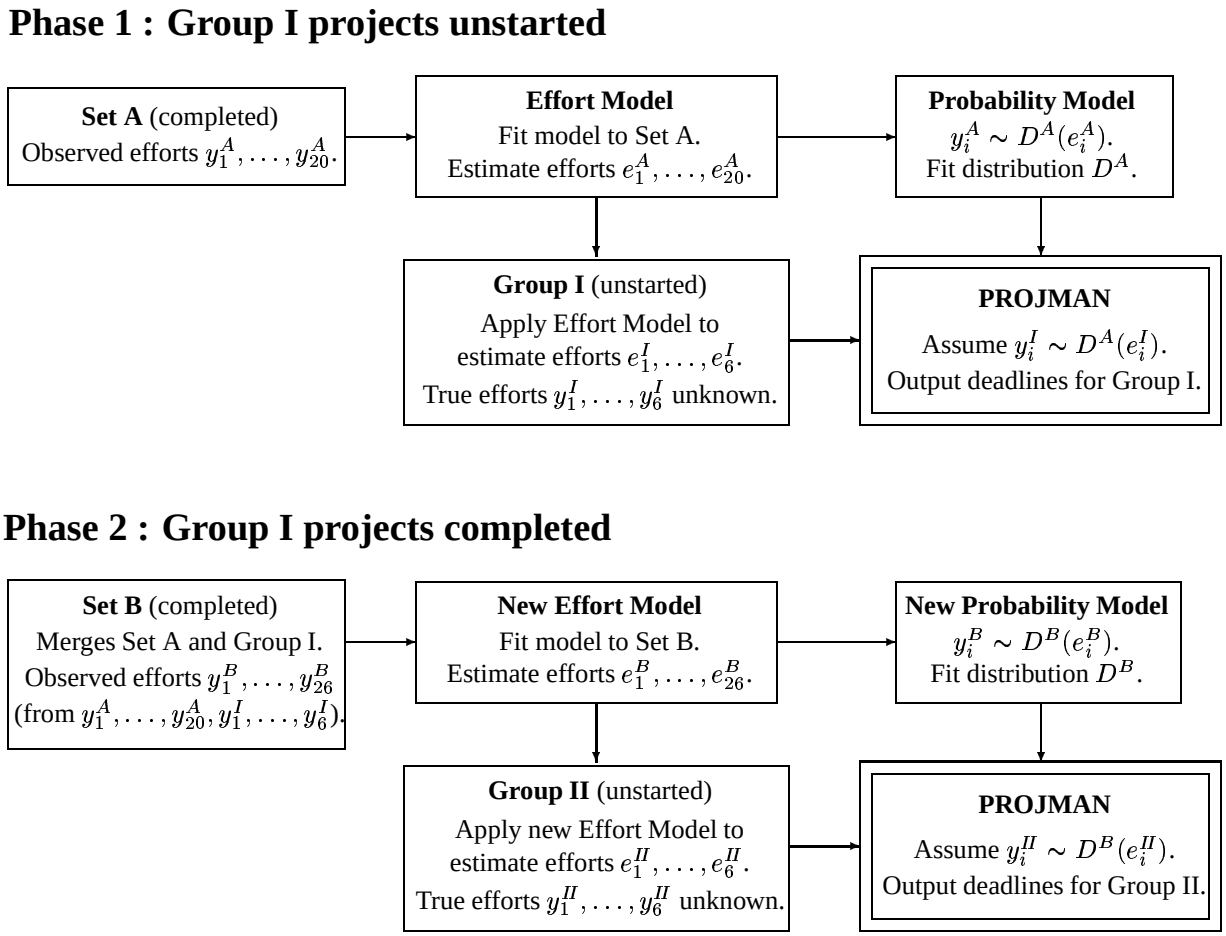

Figure 2. Flow chart showing the steps used to demonstrate the PROJMAN tool with the data from Company M and the historical data Set A. For Phase 1, the true Group I efforts $y_1^I, \ldots, y_6^I$ are assumed unknown. For Phase 2, $y_1^I, \ldots, y_6^I$ are assumed known but the true Group II efforts $y_1^{II}, \ldots, y_6^{II}$ are unknown. The completed Group I projects are used to update the effort and probability models for Phase 2.

scaled between 0 and 1. Because the dataset was relatively small, we extracted only the three closest reference projects for each prediction. The final effort prediction was given by the unweighted mean of the three reference efforts. All CBR analyses were performed using CBR-Works [14], a commercially available CBR tool.

For Phase 1 (Figure 2), we applied the CBR method to the data in Set A. Each project i was given an effort prediction e_i^A based on the three most similar projects in Set A, excluding itself. The actual effort, y_i^A, was already known. We investigated the distributions in Table 1 for fitting the probability model $y_i^A \sim D^A(e_i^A)$. The data were hard to fit because the CBR method tended to underpredict effort, often substantially, so a distribution with extreme positive skew was required. Only the Lognormal-median family gave an acceptable fit, and dispersion was high at 34.9. Chisquare and KS goodness-of-fit tests both gave p-values of about 0.01, indicating a poor fit, but still considerably better than any other family. The distribution captures the upper extreme of the data quite well, but there are fewer observations than expected at the lower extreme. The deadlines provided by PROJMAN may therefore be on the tight side, and a high probability of success should be sought to compensate.

The Set B data were also fitted with the Lognormal-median family, giving a slightly reduced dispersion of 24.3. Although the two dispersion values are numerically quite different, the practical difference in the shape of the fitted distribution is rather small.

4.3. Stepwise regression framework (SWR)

Stepwise regression is a method of selecting a set of predictor variables for a regression model, from a large number of candidate variables [13]. One predictor variable is added to the model at a time, chosen as the variable with the highest partial correlation to the response, taking into account all other predictors currently in the model. The aim is to find the set of predictors that maximize the F-statistic. The F-statistic assesses whether the predictors, taken together, are significantly associated with the response. A new predictor is added only if it increases F by some specified amount. Stepwise regression has frequently been used as a benchmark [17, 3, 16, 8], and is regarded as a good prediction technique in some circumstances [7].

Linear regression models were not suitable for the Set A data due to heteroscedasticity, but ordinary regression with a log-transformed response provided a good fit. The selected model for the Set A data was:

$$\log(y_i^A) = 2.22 + 0.0069(\texttt{NewWP})_i + \epsilon_i^A,$$

where $(\texttt{NewWP})_i$ is the number of new web pages in project i, and $\epsilon_i^A \sim \text{Normal}(0, 1.58)$ for $i = 1, \ldots, 20$. In terms of the probability model $y_i^A \sim D^A(e_i^A)$, this corresponds to the Lognormal-median family with dispersion $e^{1.58}\left(e^{1.58} - 1\right) = 18.5$.

For the Set B data, the selected model was expanded to:

$$\log(y_i^B) = 1.80 + 0.0049(\texttt{NewWP})_i + 0.24(\texttt{TotHigh})_i + \epsilon_i^B,$$

where $(\texttt{TotHigh})_i$ is the total number of high effort features in project i, and $\epsilon_i^B \sim \text{Normal}(0, 1.09)$ for $i = 1, \ldots, 26$. For the probability model $y_i^B \sim D^B(e_i^B)$, this corresponds to the Lognormal-median family with dispersion $e^{1.09}\left(e^{1.09} - 1\right) = 5.8$.

Goodness-of-fit tests indicated that both models gave an excellent fit, with p-values from the Chisquare and KS tests all greater than 0.75. Note that this does not necessarily imply that the models are good predictive models; instead, it means that the extent of the uncertainty in the model has been accurately captured by the Lognormal-median family. There may be better effort models that predict the true effort with less uncertainty. Nonetheless, the dispersion estimates from the stepwise regression method are notably smaller than those from the case-based reasoning approach, so stepwise regression appears to have been a better approach for these data. The excellent fit of the Lognormal-median distribution suggests that the output from the PROJMAN tool can be considered reliable if the same model holds for the Company M data.

All regression analyses were conducted using the statistical software R. We also investigated an alternative stepwise ANOVA procedure [5], but it did not improve upon the stepwise regression results.

4.4. Deadline Planning for Company M

Table 3 shows the outcome of the PROJMAN system applied to the Company M portfolio. The actual effort and efforts estimated by the CBR and SWR approaches are shown for each project. Contrary to appearances, all of the observed efforts are consistent with their estimates. For example, the observation of $y_2 = 38$ for project M2 is at approximately the 10th percentile of the fitted Group I SWR distribution with median $e_2 = 191$ and dispersion 18.5. Even the observation of $y_9 = 233$ for project M9 is only at the 97th percentile for the fitted Group II SWR distribution with median $e_9 = 33$ and dispersion 5.8. This serves to emphasize the vast uncertainty associated with the two effort prediction methods applied to these data, and the need for this uncertainty to be taken into account when deciding upon secure deadlines. We do not necessarily expect the deadlines suggested by PROJMAN to be favorable in the sense of being short, but they are likely to be realistic if the models presented here provide the best available method of effort estimation.

Table 3 also shows suggested deadlines output by PROJMAN to attain both 80% and 50% probability of portfolio success. Occasionally the deadlines suggested by the CBR and SWR approaches are very different, notably for project M2; but generally they are fairly compatible. To test the practical consequences of selecting the wrong model, we used PROJMAN to estimate the probability of portfolio success if the SWR deadlines were used when the truth was represented by the CBR model, and vice versa. The 80% deadlines from SWR yielded 78% success under CBR for Group I, and 64% for Group II. Conversely, the 80% deadlines from CBR yielded 17% success under SWR for Group I, and 78% success for Group II. With the exception of the third value (17%), these results are encouragingly close to the desired 80%. The anomalous 17% is due to the highly discrepant estimates given by CBR and SWR for project M2.

Another question of interest is how conservative the PROJMAN deadlines are for practical purposes. Suggested deadlines that are far longer than necessary will be almost assured of success, but will not enable companies to make their contract bids competitive. PROJMAN allows the user to find the lowest probability of portfolio success at which the observed efforts would meet the corresponding deadlines. Under the SWR approach, this was 21% and 81% for Groups I and II respectively. For the CBR approach, the results were 17% and 62%. All of these probabilities lie well away from the extremes of 0% and 100%, suggesting that the PROJMAN deadlines are within the realistic range. Deadlines provided for 80% success probability will often look unduly conservative, but the SWR Group II data demonstrates that even more conservative deadlines may be

Table 3. Portfolio details. All efforts are measured in person-days. The case-based reasoning and stepwise regression approaches are labeled respectively CBR and SWR. Deadlines are measured in number of working days from the start day, and are provided to attain both 80% and 50% success under the two different probability models.

Group	Project	Estimated efforts, e_i		Actual effort, y_i	Suggested deadlines, d_i			
		CBR	SWR		CBR (80%)	CBR (50%)	SWR (80%)	SWR (50%)
	M1	19	22	42	50	24	56	32
	M2	8	191	38	22	12	386	163
I	M3	3	12	9	8	4	27	16
	M4	3	11	3	9	5	22	12
	M5	42	22	51	118	53	63	36
	M6	27	16	5	72	36	39	23
	M7	6	11	16	15	8	23	14
	M8	3	6	3	7	4	9	5
II	M9	33	33	223	85	42	51	31
	M10	44	77	112	114	56	118	64
	M11	30	11	12	64	34	19	12
	M12	3	7	6	8	5	12	8

Table 4. Experiments with deadlines for Group II projects under the stepwise regression model. Estimated efforts e_i are measured in person-days, while deadlines d_i represent working days from the start. Each trial is conducted with a different set of deadlines. The last two columns for each trial give the output from 10 000 PROJMAN simulations with E=5 employees using the Lognormal-median model. P(fail) is the probability of failure for deadline d_i, and P(fail 1) is the probability that deadline d_i was the first to fail. The overall probability of portfolio success for each trial is given in the final row. For each trial, the numbers in the P(fail 1) column sum to 100%– P(overall success).

		Trial 1			Trial 2			Trial 3		
Proj.	Est. e_i	Deadline d_i	P(fail) (%)	P(fail 1) (%)	Deadline d_i	P(fail) (%)	P(fail 1) (%)	Deadline d_i	P(fail) (%)	P(fail 1) (%)
M8	6	9	3.17	3.17	9	3.01	3.01	16	0.72	0.72
M12	7	12	5.66	3.45	12	5.31	3.24	20	1.28	0.79
M11	11	19	6.32	3.09	19	6.48	3.55	29	1.73	1.06
M7	11	23	8.90	3.56	23	8.74	3.22	35	2.17	0.92
M9	33	51	5.75	3.36	51	6.28	3.78	54	4.65	3.63
M10	77	$\boxed{118}$	5.13	3.23	$\boxed{70}$	18.59	11.55	70	17.48	12.84
		P(overall success) : 80.14%			P(overall success) : 71.65%			P(overall success) : 80.04%		

required in reality (81%). The need for such a large leeway in the SWR II case was due to the high effort required for project M9. There is no obvious feature of the data for project M9 that could have predicted this high effort.

Finally, Table 4 shows a process of deadline experimentation that might be undertaken by a company manager. The projects shown are from Group II under the stepwise regression model. Most of the effort estimates are of small to moderate size, with the exception of M10 which is large. We imagine that the manager is particularly keen to secure the contract for project M10, and aims to reduce the deadline to make as competitive a bid as possible. Applying PROJMAN to suggest deadlines at 80% success probability returns a long deadline of 118 days for project M10 (highlighted in Table 4, Trial 1). In Trial 2, the manager experiments with lowering the deadline for M10 to 70 days: the outcome is an overall portfolio success probability of 71.65%. This might be considered acceptable. However, the manager is also able to force a deadline of 70 days for project M10 and use PROJMAN to calculate renewed deadlines for the other projects in order to raise the overall success probability back up to 80%. This requires compromise on the smaller projects, as shown in Trial 3. Most of the 20% failure allowed for in Trial 3 is caused by the forced project M10.

5. Conclusions

A successful approach to project management and deadline planning depends upon a realistic assessment of the uncertainty attached to effort estimates. Uncertainty models can be built with ease using standard statistical tools and tests. They can be applied to estimates generated either by expert opinion, guesswork, or existing prediction models. In many circumstances, building a model of estimate uncertainty will be considerably easier than finding an accurate model for effort.

A portfolio approach to deadline planning enables a company to provide competitive contract bids by absorbing risks into a company-wide management plan. It is very difficult to balance likely risks by intuition, but with basic statistical software the problem becomes straightforward. The PROJMAN tool takes into account progress on current projects, and allows estimates to be refined or updated at any stage. The probability model can also be updated as required, allowing the management plan to evolve over time.

At present, the methods assume that all employees are able to participate in all projects. In some cases this might not be possible, either due to differential skills of employees, or because projects themselves have a limit on the number of separate contributions that can be worthwhile. For future work, we aim to consider possible adaptations of the method for these situations. Another area for future work concerns the allocation of employee resources to portfolio projects. At present, PROJMAN offers deadlines to reach a required probability of portfolio success, but it does not suggest how employee time should be allocated for these deadlines to be met.

Acknowledgments

We would like to thank all those companies that have volunteered data on their Web projects to the Tukutuku Benchmarking project.

References

[1] R. Fairley. Risk management for software projects. *IEEE Software*, 11(3):57–67, 1994.

[2] R. M. Fewster and E. Mendes. Measurement, prediction and risk analysis for web applications. In *Proceedings of the 7th International Software Metrics Symposium*. IEEE Computer Society, 2001.

[3] G. Kadoda, M. Cartwright, and M. J. Shepperd. Issues on the effective use of CBR technology for software project prediction. In *Proceedings of the 4th International Conference on Case-Based Reasoning, ICCBR*, pages 276–290, Vancouver, Canada, 2001.

[4] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*, volume 2. London, Charles Griffin & Co., 4th edition, 1979.

[5] B. Kitchenham. A procedure for analyzing unbalanced datasets. *IEEE Transactions on Software Engineering*, 24(4):278–301, 1998.

[6] B. Kitchenham and S. Linkman. Estimates, uncertainty and risk. Technical Report TR96-07, Keele University, UK, 1996.

[7] P. Kok, B. A. Kitchenham, and J. Kirakowski. The MERMAID approach to software cost estimation. In *Proceedings of the ESPRIT Annual Conference*, pages 296–314, Brussels, 1990.

[8] E. Mendes, N. Mosley, and S. Counsell. Web metrics: Estimating design and authoring effort. *IEEE Multimedia, Special Issue on Web Engineering*, pages 50–57, 2001.

[9] E. Mendes, N. Mosley, and S. Counsell. Investigating early web size measures for web costimation. In *Proceedings of EASE'03*, 2003.

[10] E. Mendes, I. Watson, C. Triggs, N. Mosley, and S. Counsell. A comparison of development effort estimation techniques for web hypermedia applications. In *Proceedings of the 8th International Software Metrics Symposium*. IEEE Computer Society, 2002.

[11] J. Moses. Learning how to improve effort estimation in small software development companies. In *Proceedings of the 24th International Computer Software and Applications Conference*, 1999.

[12] L. Pickard, B. Kitchenham, S. Linkman, and P. Jones. Can software bidding practices be improved? – Bidding and portfolio management practices in software and other industries. Technical Report TR/SE-003, Keele University, UK, 2000.

[13] L. Schroeder, D. Sjoquist, and P. Stephan. *Understanding Regression Analysis: An Introductory Guide*. Number 57 in Quantitative Applications in the Social Sciences. Sage Publications, Newbury Park, CA, USA, 1986.

[14] S. Schulz. CBR-Works. In *Proceedings of the German Workshop on Case-Based Reasoning, GWCBR'99*. LNAI, Springer-Verlag, 1995.

[15] M. J. Shepperd and G. Kadoda. Comparing software prediction techniques using simulation. *IEEE Transactions on Software Engineering*, 27(11):1014–1022, 2001.

[16] M. J. Shepperd and G. Kadoda. Using simulation to evaluate prediction techniques. In *Proceedings of the 7th International Software Metrics Symposium*, pages 349–358. IEEE Computer Society, 2001.

[17] M. J. Shepperd, C. Schofield, and B. Kitchenham. Effort estimation using analogy. In *Proceedings of ICSE-18*. IEEE Computer Society, 1996.