Modeling Activity Times by the Parkinson Distribution with a Lognormal Core: Theory and Validation

by

Dan Trietsch[1]
Lilit Mazmanyan[1]
Lilit Gevorgyan[1]
Kenneth R. Baker[2]

October, 2010[3]

Abstract

Based on theoretical arguments and empirical evidence we advocate the use of the lognormal distribution as a model for activity times. However, raw data on activity times are often subject to rounding and to the Parkinson effect. We address those factors in our statistical tests by using a generalized version of the Parkinson distribution with random censoring of earliness, ultimately validating our model with field data from several sources. We also confirm that project activities exhibit stochastic dependence that can be modeled by linear association.

[1] College of Engineering, American University of Armenia, Yerevan
[2] Tuck School, Dartmouth College, Hanover NH

Modeling Activity Times by the Parkinson Distribution with a Lognormal Core: Theory and Validation

1. Introduction

Most published work on stochastic scheduling falls into one of two broad categories: (1) machine scheduling models and (2) project scheduling models. A key component of such models is a probability distribution for processing times or activity durations. Machine scheduling models tend to rely on the exponential distribution or the normal distribution. The exponential distribution often yields elegant results in problems that cannot be solved analytically for generic distributions (e.g., Bruno et al. 1981, Ku & Niu 1986). The normal distribution is consistent with assuming that processing times are sums of numerous independent components of uncertainty so that the central limit theorem applies (e.g., Soroush & Fredendall 1994). Project scheduling models, since the seminal work of Malcolm et al. (1959), have mostly relied on the beta distribution because of its flexibility and a claim that it is easy to estimate (Clark, 1962).

In this paper, we advocate the use of the *lognormal* distribution as a model for processing times and activity durations. We enumerate the various theoretical properties that support the use of the lognormal for both machine scheduling and project scheduling models, although our primary concern lies with the latter.

For the most part, the choice of a probability distribution for machine scheduling or project scheduling seems to be driven by convenience rather than empirical evidence. Efforts to validate assumptions about processing time distributions are scarce. For example, we have found no evidence in the literature that the beta distribution has *ever* been validated. Some progress has been made with data on surgery times (May et al. 2000, Strum et al. 2000), showing that the lognormal distribution provides the best fit by far. However, machine scheduling models have rarely considered the lognormal distribution (an exception being Robb and Silver, 1993). In this paper, we validate the use of the lognormal distribution as a model for activity times in several independent datasets obtained from project scheduling applications. By contrast, the beta distribution and the exponential distribution would fail in most of these cases.

Two practical issues arise in attempts to validate a particular probability distribution. One factor is the "Parkinson effect," which is especially relevant in project scheduling: *reported* activity times may violate lognormality because earliness is hidden, not because the lognormal is a poor model. In other words, activities may finish earlier than predicted, or earlier than a given deadline, but there may be no incentive to report any outcome other than finishing on time. In such cases, the reported data contain a bias that obscures the underlying distribution. A second factor is the time scale: empirical data may be collected on a coarse time scale, leading to rounding of the actual times. However, rounding may cause false rejection of lognormality in standard tests, such as Shapiro-Wilk (Royston 1993). These problems may explain why the lognormal has not been widely adopted for machine scheduling applications as well as project applications. In our validations, we recognize the Parkinson effect and the consequences of rounding. We introduce a new version of the Parkinson distribution, and we use statistical tests that account for the presence of ties occurring on a coarse time scale.

The vast majority of papers in both machine scheduling and project scheduling also rely on the assumption of statistical independence, but that is a very strong assumption. One serendipitous feature of the lognormal distribution is that it lends itself to use when statistical dependence is modeled by linear association (Baker & Trietsch 2009). In this paper, we also validate the linear association model for representing dependencies in empirical data, ultimately justifying linearly-associated lognormal processing times with different means but the same coefficient of variation.

Section 2 provides background information from the project scheduling literature and discusses published activity distribution models for PERT. Section 3 presents the theoretical arguments for selecting the lognormal distribution as a model for activity time. Section 4 presents empirical support for that choice. In Section 5 we show how to account for the Parkinson effect and for ties in the data. Section 6 demonstrates that the stochastic dependence we encounter in our datasets can be modeled effectively by linear association. Section 7 contains our conclusion. Rather than starting with a comprehensive literature review, we discuss the main antecedents of our work as we proceed.

2. Activity Time Distributions

The major tools used in project scheduling are two overlapping approaches introduced in the late 1950's: Critical Path Method (CPM) in Kelley (1961) and Program Evaluation and Review Technique (PERT) in Malcolm et al. (1959) and Clark (1962). Of the two, only PERT recognizes the probabilistic nature of activity times within a project. At the heart of the PERT method is a set of assumptions that facilitates a systematic and intuitively-appealing method for modeling stochastic behavior in projects. In this paper we address the most basic element of PERT: fitting a distribution to each individual activity time. According to Clark (1962), the beta distribution has the necessary flexibility, and a good way to estimate its parameters is by eliciting three values, an approach we call the *triplet method*. In particular, estimates are elicited for the minimal possible value (denoted *min*), the mode (*mode*), and the maximal possible value (*max*). These estimates are then used to define the mean μ, and the standard deviation, σ, using the formulas

$$\mu = (min + 4mode + max) / 6 \qquad (1)$$

$$\sigma = (max - min) / 6 \qquad (2)$$

Equation (2) was selected arbitrarily, to resemble a truncated normal between $\pm 2.96\sigma$. Equation (1) was then derived as an approximation for a beta distribution that matches Equation (2) and the estimated *min*, *mode*, and *max* values. (In the interest of simplicity, Malcolm et al. rejected the idea of fitting the parameters of a beta distribution by solving a cubic equation.) Clark (1962) states, "The author has no information concerning distributions of activity times, in particular, it is not suggested that the beta or any other distribution is appropriate." Accordingly, we should not take the selection of the beta distribution and the two estimation formulas too seriously. However, this observation did not prevent considerable research effort on the subject. Theoretically, with three exceptions—noted by Grubbs (1962) as part of a scathing critique of PERT—no beta distribution fits both estimators. To be precise, Equations (1) and (2) provide good approximations for some beta distribution if *mode* is within the range from $min + 0.13(max - min)$ to $min + 0.87(max - min)$; e.g., see Premachandra (2001). This limitation implies a coefficient of variation (*cv*) smaller than about 0.66. When a beta distribution exists that satisfies Equations (1) and (2) approximately, we refer to it as a *PERT-beta*. Typical practitioners are not aware of that technical limit on the location of *mode* relative to *min* and *max*. However, Equations (1) and (2) imply $cv \leq 1$. (Equality would hold only if *mode* = *min* = 0. In such an extreme case, however, we may safely assume a significant error in Equation (1) must ensue.) Considering that *min* = 0 is seldom appropriate and *mode* > *min* is typical, we suspect that the triplet method tends to yield very low *cv* values. For instance, in the numerical examples given by Malcolm et al., the average *cv* is 0.19 and the maximum *cv* is 0.29. Practitioners who implement the formulas cannot avoid underestimating σ when the true range between *min* and *max* is less than 6σ. Furthermore, observations reported by Tversky & Kahneman (1974) suggest that experts asked to provide 98% confidence intervals tend to provide very narrow limits that are exceeded in roughly 30% of instances. By the same token, we should not trust that the true time will fall between the estimates *min* and *max*, as PERT assumes. This possibility reinforces the claim that the triplet method is likely to lead to small *cv* estimates and thus perilously discounts the possibility of large deviations. On top of that, Woolsey (1992) provides anecdotal evidence that practitioners sometimes consider the requirement to generate the *min* and *max* estimates onerous, leading to frivolous results.

In spite of these contrarian arguments, many authors have suggested technical improvements for the fit of the beta to the triplet method. Premachandra (2001) proposes such a refinement and offers a thorough discussion of previous contributions on which his method improves. Yet another approach (surveyed by Premachandra) is to treat *min* and *max* as very high or very low percentiles instead of strict minimum and maximum values. These approaches attempt to stray as little as possible from the beta assumption while correcting particular problems the authors perceive, sometimes at the price of exacerbating problems that the authors do not perceive.

Other authors, however, suggest alternative distributions. In particular, Kotiah & Wallace (1973) propose a doubly-truncated normal distribution (between *min* and *max*) and Mohan et al. (2007) propose the lognormal distribution based either on the elicited parameters *min* and *mode* or on *mode* and *max*. Hahn (2008) suggests a mixed distribution composed of the traditional beta and the uniform distribution. However, his approach relies on elicitation of a fourth parameter, which makes it even harder to implement.

With the notable exception of Grubbs and Woolsey, none of the cited authors questions the basic elicitation approach of PERT. All of them interpret *min* and *max* strictly or as representing very wide confidence limits. Most importantly, none of them makes any attempt to validate their models with practical data; instead, they compare their results to the prescribed beta. Grubbs (1962) remarks that Equations (1) and (2) rely on an implicit assumption that the elicited estimates are not subject to error, but any viable approach must take estimation errors into account. In what follows we establish that *cv* > 0.66 is highly likely in practice, and thus we can conclude that it is highly unlikely that the PERT-beta can be validated.

3. Advantages of the Lognormal Distribution

Baker & Trietsch (2009) present four arguments why the lognormal distribution is attractive for modeling stochastic activity times: (1) it is strictly positive, (2) its *cv* is not restricted, (3) it can approximate sums of positive random variables, and (4) it can represent the relationship between capacity and activity time. In our present context, we add a fifth argument: it can also represent the ratio between actual and estimated activity time. Among the distributions typically employed to model activity times, only the lognormal exhibits all these traits. We next elaborate on the last three.

The lognormal distribution can represent sums of positive continuous random variables (convolutions) because it satisfies the *lognormal central limit theorem* (Mazmanyan et al. 2008; see also Robb 1992): as n grows large, the sum of n independent, positive random variables tends to lognormal. This theorem holds subject to regularity conditions similar to those that apply to the classical central limit theorem. Practitioners often rely on the classical central limit theorem as justification for using the normal distribution to model the sum of a small number of random variables, but it may produce negative realizations. This is especially true when the cv is large, which our data show is an important case. Instead, we use the lognormal central limit theorem as a basis for approximate convolutions of positive random variables. Numerical experience suggests that it is more effective than the normal for small n, too, especially if the components' distributions are skewed to the right, such as Erlang or chi-square. This effectiveness may be explained by another good feature of the lognormal distribution: Mazmanyan et al. show that any convolution of two or more positive random variable must have $f(0) = 0$, where f denotes the density function, and the lognormal distribution satisfies this condition. Among the major distributions that have been suggested for activity times and have unconstrained cv, the lognormal distribution is unique in this respect.

An important special case has components that are lognormal. Let X denote the sum of n independent lognormal random variables with parameters m_j and s_j^2, for $j = 1, \ldots, n$. (The lognormal random variable is the exponent of a core normal random variable with mean m_j and variance s_j^2.) To approximate the distribution of X by a lognormal distribution, we first evaluate the components' means and variances from the following formulas:

$$\mu_j = \exp(m_j + s_j^2/2) \text{ and } \sigma_j^2 = \mu_j^2[\exp(s_j^2) - 1] \tag{3}$$

Next, we add means to obtain μ_X and add variances to obtain σ_X^2. Then we can calculate the squared coefficient of variation as $cv_X^2 = \sigma_X^2/\mu_X^2$. Finally, we obtain m_X and s_X from,

$$s_X^2 = \ln(1 + cv_X^2) \quad \text{and} \quad m_X = \ln\mu_X - s_X^2/2 \tag{4}$$

Such calculations are easy to program and should cause no difficulty in practice.

An interesting feature is that both the sum and the product of independent lognormal random variables are lognormal: the former approximately and the latter analytically. A distribution that has the flexibility to represent both sums and products is likely to fit realistic cases that are influenced by both additive and multiplicative elements.

We say that n positive random variables, Y_j, are linearly associated if $Y_j = BX_j$ where $\{X_j\}$ is a set of n independent positive random variables and B is a positive random variable, independent from $\{X_j\}$. If we weaken the independence assumption and replace it by linear association, the lognormal central limit theorem can still be used. We just have to add up the X_j elements before multiplying by the bias term, B. In practice, B is likely to be influenced by several additive and multiplicative causes, so it is likely to be lognormal by itself. Therefore, if X_j is lognormal, then $Y_j = BX_j$ is also lognormal. Furthermore, by Cramér's theorem, if Y_j is lognormal then both X_j and B must be lognormal. (The theorem states that the sum of two independent random variables cannot be normal unless both are normal.)

The reciprocal of the lognormal distribution is also lognormal (with the same s). In some environments, the activity time is inversely proportional to the effective capacity allocated to an activity. If the total work requirement is a constant but the effective capacity is lognormal, then the activity time will be lognormal. Indeed, the ratio of two lognormal random variables is also lognormal, so even if the total work requirement is lognormal, the ratio of work requirement to capacity—i.e., the activity time—will be lognormal as well. Thus, using the lognormal distribution as our default is convenient for situations in which crashing is permitted.

In our present context, if both activity time and estimate distributions are lognormal, the ratios of activity times to estimates are lognormal. In making that observation, however, we should recognize that estimates and actual times are positively correlated: otherwise, the estimates would be useless. For *any* two positively correlated random variables, assume for the time being that the correlation is due to linear association. Linear association between two random variables X and Y implies the existence of three positive independent random variables R, S, and B such that $X = RB$ and $Y = SB$. If B is a constant, then X and Y are independent and thus associated (Esary et al. 1967). If R and S are constants, then X and Y are proportional to each other and thus associated (with a correlation coefficient of 1). It follows that if X and Y are linearly associated, then $X/Y = R/S$ (where the numerator and denominator are independent). We are mainly interested in instances where the marginal distributions of X and Y are lognormal and are positively correlated. Then, because the convolution of two independent random variables is normal if and only if both of them are normal (by Cramér's theorem), the log-normality of X and Y implies that B, R, S, R/S and X/Y all exist and all are lognormal; that is, positively correlated lognormal variates are always linearly associated. To see this, take logarithms, so that our statements about ratios and products become statements about subtractions and additions. If B is not lognormal, neither X nor Y is lognormal. Nonetheless, $X/Y = R/S$ remains lognormal if R and S are lognormal. That may apply for residuals in a logarithmic regression model.

4. Empirical Validation of Lognormality

In this section we present published and unpublished evidence for the validity of the lognormal distribution for activity times in practice. Robb & Silver (1993) list some early papers that document the validity of the lognormal distribution, mainly for health-related applications such as patient consultation times and sick leave durations. May et al. (2000) report that the lognormal distribution matches surgery time histograms well, and standard tests show that it tends to provide a much better fit than the normal. They fit a lognormal separately to each important combination of determining factors that would be known in advance (such as the exact procedure and the type of anesthesia). Altogether, they identify over 3000 usable groupings, most with fewer than 30 items. Although they show that judiciously setting a shift parameter improves the fit somewhat, they report that the regular (two-parameter) lognormal performs almost as well. (In addition to m and s, a *shifted* lognormal distribution uses a third parameter for the minimum.) We can safely limit ourselves to the regular case because estimating a shift parameter is guaranteed to improve the fit (Royston 1993)—so a slightly better fit is not necessarily indicative of a serious flaw in the two-parameter model—and because the difference between the two versions is reportedly small.

The lognormal can replace the beta even where the beta might have worked (but the converse is less likely to be true). For instance, Premachandra & Gonzales (1996) show a histogram for drilling times based on a sample of almost 650 holes, which we reproduce in Figure 1. Although they do not comment on the distribution, we find that a PERT-beta with *min* = 0, *mode* = 40, and *max* = 250 would match the true mean and variance well. But a lognormal fit, which is depicted in the figure, also works and easily passes a Chi-square goodness-of-fit test.

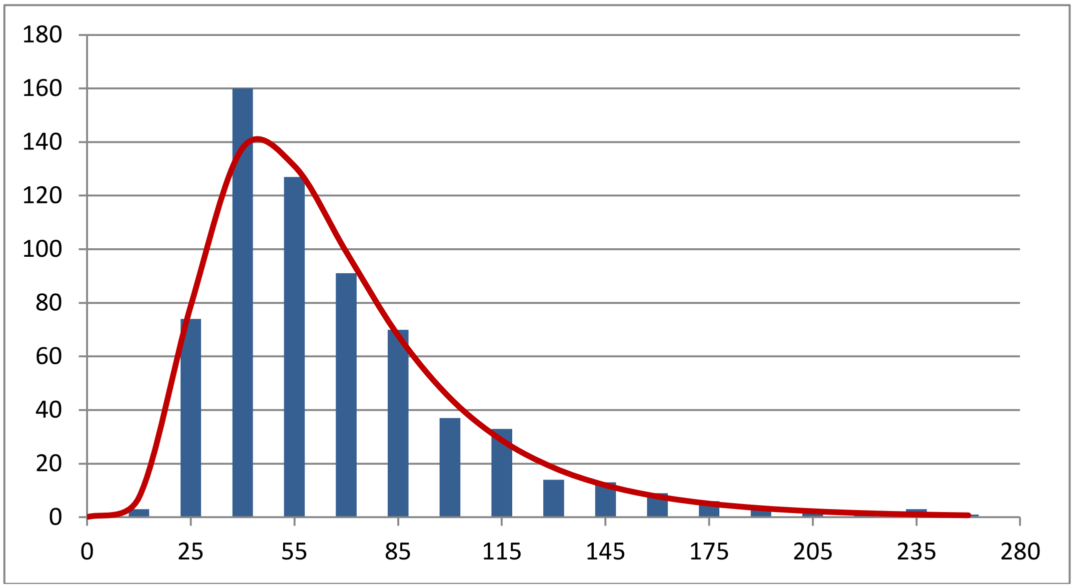


Figure 1. Drilling Times in Minutes (Premachandra & Gonzales, 1996)

Often, we need to compare an activity time to its initial estimate. This type of analysis is important because in practice the initial estimate is sometimes all the information we have in advance for a particular activity, and variation is measured relative *not* to the true mean (which remains unknown) but rather to its estimate. Let p_j denote the duration of activity j, and let e_j be a single point mean estimate that is available for this activity. We model p_j/e_j as lognormal. An equivalent statement is as follows: suppose we log-transform our data, including our estimates, and apply linear regression to estimate $\ln p_j$ by $\ln e_j$, then the residuals would be $\ln p_j - \ln e_j = \ln(p_j/e_j)$. Per our discussion in Section 3, if $\ln p_j$ and $\ln e_j$ are normal (including the case of a constant $\ln e_j$), then $\ln(p_j/e_j)$ is normal even though we assume positive correlation between p_j and e_j. As a rule, when we wish to estimate the parameters of a normal variate, such as $\ln(p_j/e_j)$, the mean and sample variance form a sufficient statistic. Nonetheless, if the distribution is indeed normal, then we can also estimate the mean and variance by a normal Q-Q chart (also known as a Q-Q plot). To obtain such a chart, we sort our sample by increasing order and fit a regression line to the sorted data. A typical equation in such a regression, relating to the activity with the j-smallest $\ln(p/e)$, has the form

$$\hat{m} + z_j \hat{s} = \ln \frac{p_j}{e_j} + u_j \tag{5}$$

where u_j is the error term and z_j is selected in such a way that, under the normality assumption of the transformed random variable, the regression line intercept and slope are good estimates of m and s. For this purpose, the analytically correct values of z_j should match the expected values of the normal order statistic (Shapiro & Francia 1972). However, a much simpler and yet excellent approximation is obtained by Blom's scores (Blom, 1958): set $z_j = \Phi^{-1}[(j - 0.375)/(n + 0.25)]$, the z-value for which the standard normal distribution cumulative distribution function (cdf), $\Phi(z)$, yields a probability of $(j - 0.375)/(n + 0.25)$. We adopt this approach because it can be generalized for the Parkinson effect and because Looney & Gulledge (1985) demonstrated that it can be used as the basis of a statistical normality test similar in structure and power to the Shapiro-Francia test (Shapiro & Francia 1972; see also Shapiro & Wilk 1965 and Filliben 1975). In this test we compare the Pearson linear correlation coefficient between $\ln p_j$ and $\ln e_j$, which can be obtained as the square root of R^2 provided by standard regression analysis, to tabulated values based on simulation. Table 2 in Looney & Gulledge (1985) provides such values for $n \leq 100$ and probabilities of 0.5%, 1%, 2.5%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, 97.5%, 99%, and 99.5%. The need for simulation-based critical values arises because a Q-Q chart employs the order statistic, so its error terms are heteroscedastic and the $\ln(p_j/e_j)$ values are not independent.

To demonstrate this procedure, we use an example which also shows that high coefficients of variation are possible in practice. In a software company in New Zealand, programming task times were estimated by a manager, and tasks were then allocated to programmers, one at a time. Programming times were not recorded, but because each programmer only had one task in progress, the total time between start and finish can be used. In reality, programmers also had other time-consuming tasks, such as participating in meetings and dealing with emergent unscheduled work. Thus, our data exhibits high variance and perhaps we should not expect estimates to be very useful in the first place. However, we face similar situations frequently in practice. Figure 2 shows a Q-Q chart of $\ln(p_j/e_j)$ for $n = 44$, employing Blom's scores. From the regression equation, we obtain estimates of 0.1171 and 2.2787 for m and for s. Noting the high slope (almost 2.28), we might be suspicious before accepting the result. However, by comparing the square root of R^2 (0.987) to the tabulated values of Looney and Gulledge for the $n = 44$ case, we find that it is between 0.985 and 0.989, which mark probabilities of 25% and 50% (for $n = 44$). Evidently, the lognormal can serve as an adequate approximation for this dataset. Furthermore, from the regression output, the t-statistic of the slope is high, at 39.98. For instance, the probability that the slope exceeds 2.0 is 0.999992, which implies a very high coefficient of variation. Thus, the analysis we propose is relevant and plausible for this case. No PERT-beta distribution can fit such a high *cv*.

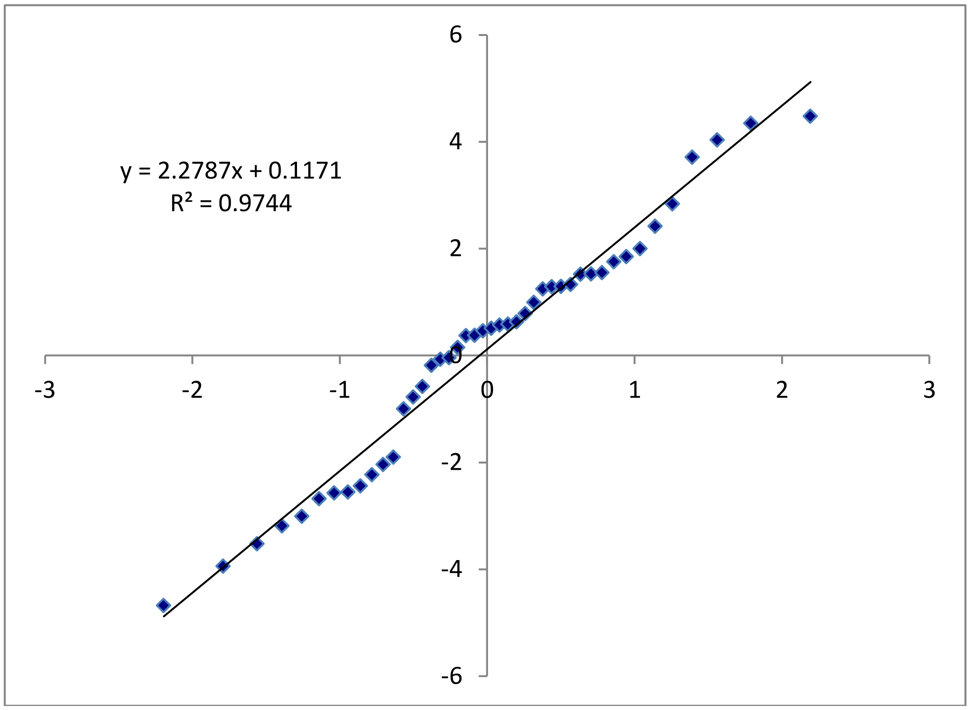


Figure 2. Normal Q-Q Chart of ln(p_j/e_j) for Programming Time

5. The Parkinson Distribution and Correcting for Ties

Parkinson's Law states that "work *expands* so as to fill the time available for its completion" (Parkinson, 1957; emphasis added). The relevance of Parkinson's Law to activity times has been recognized by several authors, including Schonberger (1981) and Gutierrez & Kouvelis (1991). The latter develop theoretical models that include a work expansion element, as per the law. We suggest that Parkinson's observations can also be explained by hidden earliness rather than by explicit expansion; that is, by early work misreported as exactly on time. Because estimates are used for both planning and control, people in charge of performing activities, when asked for estimates, may have an incentive to build some safety time into their estimate. Subsequently, they may hide earliness, to protect their future estimates from being mistrusted and reduced. Accordingly, our model for the Parkinson effect is based on hidden earliness.

Baker & Trietsch (2009) introduce the (pure) *Parkinson distribution*, as follows. Suppose that work is allotted q units of time, but it really requires X, where X is a random variable. In our interpretation of Parkinson's Law, we observe Y, given by

$$Y = \max\{q, X\}$$

and we say that Y has a [pure] Parkinson distribution. One particular family of nine projects investigated in Trietsch et al. (2010) demonstrated a pure Parkinson distribution. But in another family of five projects, some early activities seemed to be reported correctly whereas a large fraction of early activities were reported as precisely on time. In those instances a new version of the Parkinson distribution provided a statistically valid fit. Let p_P denote the probability that an early activity is falsely recorded as precisely on time. We assume that this probability applies to each early activity independently. That is, early activities are recorded correctly with a probability of $(1-p_P)$ and precisely on time otherwise; tardy activities are always recorded correctly. For $p_P = 1$ we obtain the pure Parkinson distribution whereas for $p_P = 0$ we obtain a conventional distribution. Thus the Parkinson distribution generalizes all single variate distributions, with or without the Parkinson effect.

The Parkinson distribution is defined for any internal random variable, X, but as a default we cast the lognormal in that role. Figure 3 depicts a normal Q-Q chart of $\ln(p_j/e_j)$ data from a construction project in Yerevan, Armenia. The left side of the figure depicts the raw data, consisting of 107 points, of which 20 are strictly positive and 38 are zeros. The right side is corrected for the Parkinson effect by removing the 38 "on-time" points and modifying the regression equations (5). Because a proportion p_P is censored, for early activities we replace n in the calculation of z_j by $n(1-p_P)$; that is, $z_j = \Phi^{-1}[(j-0.375)/(n(1-p_P)+0.25)]$. In our example we estimate $p_P = 38/(107-20) = 0.437$, and $n(1-p_P) = 60.26$. An alternative is to ignore early points altogether. This alternative gives more weight to tardy points and thus estimates the right side of the distribution more closely (at the expense of the overall fit). That approach may sometimes be justified because the reverse transformation eventually magnifies errors in tardiness estimates and shrinks earliness errors.

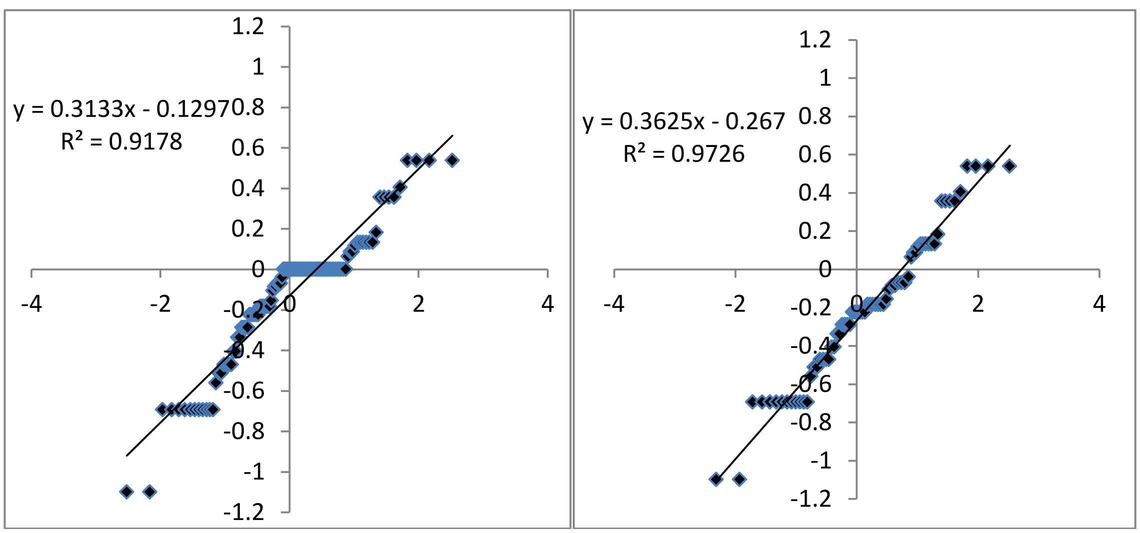


Figure 3. Q-Q Chart of ln(p_j/e_j) in a Construction Project Showing the Parkinson Effect (left) and the Fit of the Parkinson Distribution (right)

We lack formal tests for the hypothesis that the distribution is indeed Parkinson with a lognormal core. Whereas most normality tests require complete samples, our sample is incomplete after removing the "on-time" points. Some tests allow censoring one or even two tails of the distribution (Royston 1993). For instance, if we were to ignore the early points altogether, we could test the hypothesis that the log-transformed tardy points come from a normal distribution with the left tail removed. Similarly, under the assumption of random selection of misreported data, one can treat the 87 – 38 = 49 strictly early points as a censored sample with the right tail removed. The justification is that a random sample from a random sample is a (smaller) random sample. Using the same reasoning, yet another option is to randomly remove a proportion of about p_P of the strictly tardy points: if the Parkinson distribution applies, the result would be a *trimmed* but complete sample of $n(1 - p_P)$ points in which early and tardy points are approximately represented in the correct proportion. In our case, that would require trimming $20 \times 0.436 \approx 9$ strictly tardy points, leaving a trimmed sample of 60 points. Figure 4 depicts two Q-Q charts that can be obtained this way: of the 20 strictly tardy points, two appear in both trimmed samples and the other 18 are divided into two groups of 9 points, each of which appears in only one sample. The differences between the two sides of the figure are visible in the upper right corner, where tardy activities are depicted. Taking the square roots of the R^2 values we obtain 0.9826 and 0.9851. From Table 2 in Looney & Gulledge (1985), p-values of 5%, 10% and 25% are 0.980, 0.984 and 0.988 (for $n = 60$). Thus we see that the trimmed sample on the left of the figure has a p-value between 5% and 10%. The trimmed sample on the right has a p-value over 10%. Accordingly, the normality assumption cannot be rejected at the 5% level but is marginal at the 10% level.

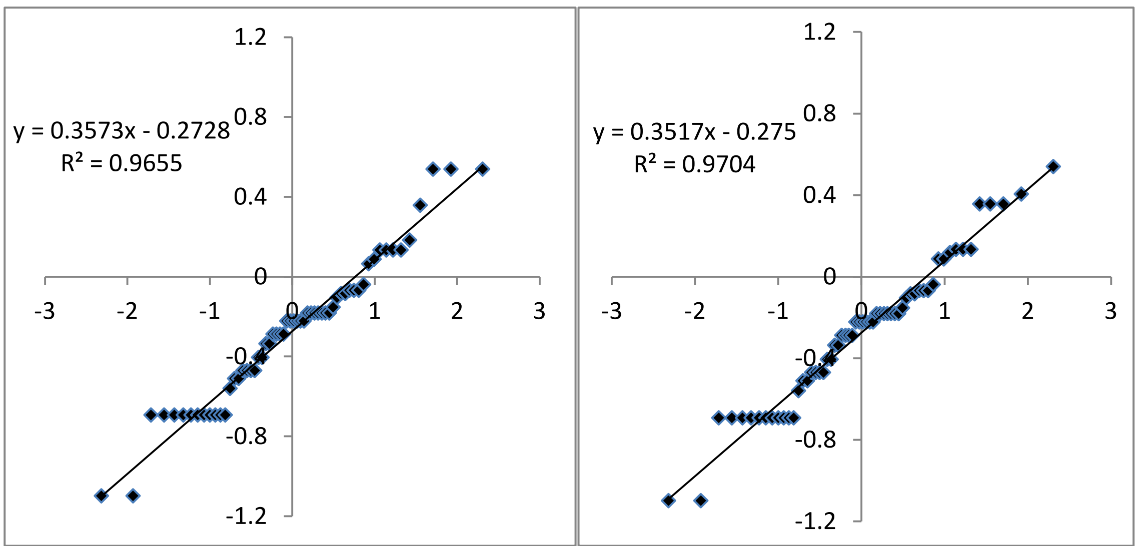
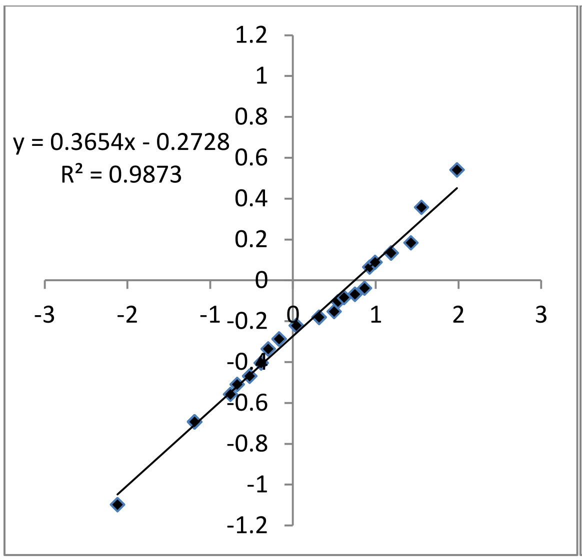


Figure 4. Q-Q Charts of ln(p_j/e_j) for Two Trimmed Samples

We observe strong grouping (ties) in both parts of Figure 4, especially for $\ln(p_j/e_j) = -0.69$. These ties are due to rounding; e.g., when the estimate is 2 time units then all cases where the reported time is rounded to 1 belong to this group. It seems that the weakness of the normality evidence is due to this grouping. Strum et al. (2000) discuss a similar problem that often led to false rejections in their operating time data. Royston (1993) proposes a remedy by averaging out Blom's scores of all tied points in each group (before calculating z_j). Similarly, we can average out the pre-calculated z_j values in each group. Figure 5 shows the results of the latter approach for the trimmed samples of Figure 4. The squared roots of R^2 increase to 0.9936 and 0.9947, and these values are associated with much more respectable p-values of about 75% (the entries for $n = 60$ in Table 2 of Looney & Gulledge for 50%, 75% and 90% are 0.992, 0.994 and 0.9956 whereas our results straddle 0.994). Royston (1993) points out that his remedy may slightly reduce the power of the test to detect non-normality, but the alternative is a gross increase of false rejections. In our case, considering that the samples pass even without the correction, it is highly plausible that the underlying activity times are indeed lognormal. By the same token, our interpretation of the Parkinson effect as hidden earliness cannot be rejected. Qualitatively similar results obtain for the family of five projects studied by Trietsch et al. (2010): three out of five pass the test, at least marginally, even without any remedy; one passes with either one of the remedies; one, with the highest p_P, cannot pass without the Parkinson correction; all pass with high p-values when both remedies are employed. The sorted s estimates for the five cases are 0.565, 0.708, 0.725, 0.797 and 0.801, corresponding to cv values of 0.613, 0.806, 0.831, 0.944 and 0.948. Of these, only one could have been conceivably fitted by a PERT-beta distribution. We note specifically that those projects are eclectic, including community organizing, training people in skills necessary to increase their productivity, agricultural infrastructure development and software programming. The projects relied on subcontractors, so the reported processing times may include queueing outside the project manager's direct control. However, subcontracting is very common in practice. For instance, the Polaris project, on which PERT was initially tested, relied on subcontracting.



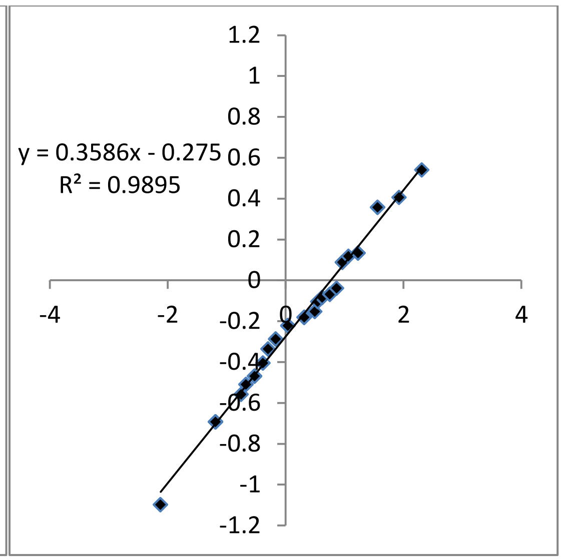


Figure 5. Resolving Ties by Using Average *z*-scores

6. Modeling Dependence by Linear Association

Although we have already demonstrated cases where the PERT-beta could not model cv appropriately, in our datasets we only have single-point mean estimates, rather than PERT-triplets. Hence, we cannot test the validity of the beta assumption directly. But we can test PERT's model for project durations. For that purpose, PERT uses the normal distribution as a model for critical path duration, based on an assumption that the duration can be approximated by a sum of numerous independent random variables. In light of our previous comments, we can invoke the lognormal central limit theorem instead, but if there are many activities in the critical path, the difference is not great. More importantly, it is well known that this approximation ignores the *Jensen gap*; that is, it ignores the fact that the project duration tends to exceed that mean due to Jensen's inequality. Clark (1961) presents an approximate calculation for the true project completion distribution, and thus accounts for the Jensen gap. Malcolm et al. (1959) noted the problem but chose not to incorporate Clark's approximation in the interests of simplicity. Since then, several authors have revisited the issue (e.g., Klingel 1966, Schonberger 1981). However, the Jensen gap may have been a red herring. The lack of calibration and the independence assumption are much more likely to cause problems. To show that, we study the properties of sums of actual activity times; that is, we effectively assume we are dealing with serial projects whose duration is the sum of all activities. Serial projects do not have Jensen gaps, so any problems with PERT for such projects must be directly attributable to its lack of calibration and the independence assumption.

We first analyze a family of five projects introduced by Trietsch et al. (2010). We show that PERT estimates are poor, mostly due to underestimating variance. That underestimation is directly attributable to the ubiquitous statistical independence assumption. We then show that if we use linear association to model statistical dependence, the results are much more plausible. Next we analyze a family of nine projects from the same source and demonstrate that it is imperative to account for consistent bias. Fortuitously, our linear association model already accounts for such bias and the difference is dramatic. Finally, we show that a dataset studied by Hill et al. (2000) can also be explained by linear association.

Ideally, for the type of analysis we pursue, we need data from many projects with many activities, and we need estimates and realizations for all activities. As a rule, activity estimates should be comparable, with the ratio between longest and shortest not exceeding 4 or 5 (Shtub et al. 2005). Otherwise, the variance of the long activities dominates, and it is impossible to analyze the others. In practice, we can use hierarchy to avoid violating this rule, but we cannot do that for historical datasets. In the five-project family, originally we had 185 activities, but we discarded 25 long activities. That still leaves a sample of 160 activities, but we note that five is a too small number of projects. The nine-project family is actually smaller: it comprises only 54 activities and 18 of those are precisely on-time. The Hill dataset is relatively large, but it does not include subtask times. Thus, none of our datasets is ideal. Nonetheless, our main claims are supported with very high statistical significance. Our technical approach to the five and nine-project cases is different from the one taken in Trietsch et al. (2010), but the qualitative results are the same. Specifically, we need to simulate project performance so we can construct empirical distributions. In Trietsch et al. (2010), we first used analysis of the type we demonstrated above to estimate parameters such as m_k, s_k and Parkinson propensities for project k, and then used those parameters to simulate similar projects. Here we introduce a bootstrap approach for that purpose. The parametric bootstrap re-sampling approach is often used to estimate such parameters when datasets are small. However, a nonparametric bootstrap exists that does away with the need to estimate parameters, and that is the one we choose (Johnston & Dinardo 1997).

6.1. Estimating Variance

To implement the nonparametric bootstrap to our data, we rely on the observation that the ratios p_j/e_j are lognormal and thus we should work with ratios or equivalently with their logarithms. We resample the ratio between actual and estimated times. To simulate a new activity time we multiply a re-sampled ratio by the new estimate. For instance, suppose we have historical data for three projects, as given by the three parts of Table 1, where the double index (j, k) indicates activity j of project k. If we were just starting Project 3, our combined history from Projects 1 and 2 would be a list of 9 ratios, 1.25, 1.00, 0.57, 1.00, 0.83, 1.00, 0.50, 2.00, and 0.67. In that history, Project 1 has a higher weight because it has five activities compared to four for Project 2. To simulate one run of Project 3 under the PERT independence assumption we first sample three ratios from the combined history, say 1.25, 0.57 and 2.0. By multiplying these ratios by the estimates of Project 3, namely $e_{1,3} = 4$, $e_{2,3} = 2$, and $e_{3,3} = 1$ and summing the products, we simulate the project duration as 5 + 8/7 + 2 = 8.14. We repeat the same process multiple times to obtain our bootstrapped sample. By sorting the results we obtain an empirical cdf; we denote the empirical cdf of project k by $F_k(t)$. In what follows, for convenience, we treat $F_k(t)$ as if it is continuous. It is a step function, but the assumption is mild if we use a large number of repetitions (we use 10,000 in our experiments, but 1000 is typically enough). In a similar vein, $F_k(t)$ is based on re-sampling from a small set, so repetitions are quite likely. We can avoid repetitions by adding some white noise into the picture. Such white noise could represent rounding effects, for instance. We ran 100 repetitions for Project 3 and obtained a distribution ranging between 3.5 and 13; the run we demonstrated above has a p-value of $F_3(8.14)$ = 0.69 (because 68 of the 100 runs were strictly smaller than 8.14). Next, let C_k denote the actual completion time of serial Project k. For instance, in Table 1, C_1 = 5 + 1 + 4 + 2 + 5 = 17, C_2 = 12 and C_3 = 6. Given the empirical distribution and the actual completion times, we find the value $F_k(C_k)$ for each project; in our example, $F_3(C_3) = F_3(6) = 0.29$.

Index	**1, 1**	**2, 1**	**3, 1**	**4, 1**	**5, 1**
Estimate	4	1	7	2	6
Actual	5	1	4	2	5
Ratio	1.25	1.00	0.57	1.00	0.83

Index	**1, 2**	**2, 2**	**3, 2**	**4, 2**
Estimate	2	4	2	6
Actual	2	2	4	4
Ratio	1.00	0.50	2.00	0.67

Index	**1, 3**	**2, 3**	**3, 3**
Estimate	4	2	1
Actual	1	4	1
Ratio	0.25	2.00	1.00

Table 1. Data for Three Fictional Projects

In this process, the independence assumption is expressed by the fact that we treat all historical ratios as one combined group. For instance, in the run we illustrated the first two ratios were sampled from Project 1 and the third from Project 2. We followed this procedure for the five-project family. The sorted values we obtain for the five project family are depicted on the left side of Figure 6 and listed below:

$$\{0.0009, 0.0071, 0.3883, 0.8320, 0.9968\}.$$

Perhaps surprisingly, this sample passes the Kolmogorov-Smirnov (K-S) uniformity test. That is due to two key facts: (*i*) the sample is very small, so large deviations from the expected value are not significant, and (*ii*) the K-S test is not sensitive to deviation at the tails. Nonetheless, it is easy to show that the results are anything but a sample from a standard uniform. For instance, consider the fact that three out of five points either fall between 0 and 5% or between 95% and 100% (namely, 0.0009, 0.0071 and 0.9968). In fact, these points are also between 0 and 1% or between 99% and 100%. The probability of the first event is given by the probability of three successes or more in a binomial $B(5, 0.1)$, which evaluates to 0.0086 (i.e., it is significant at the 1% level). The probability of the second event is associated with a binomial $B(5, 0.02)$ and is negligible (about 0.00008). We therefore conclude that the true variance of $F_k(t)$ should be higher than the value derived under the independence assumption.

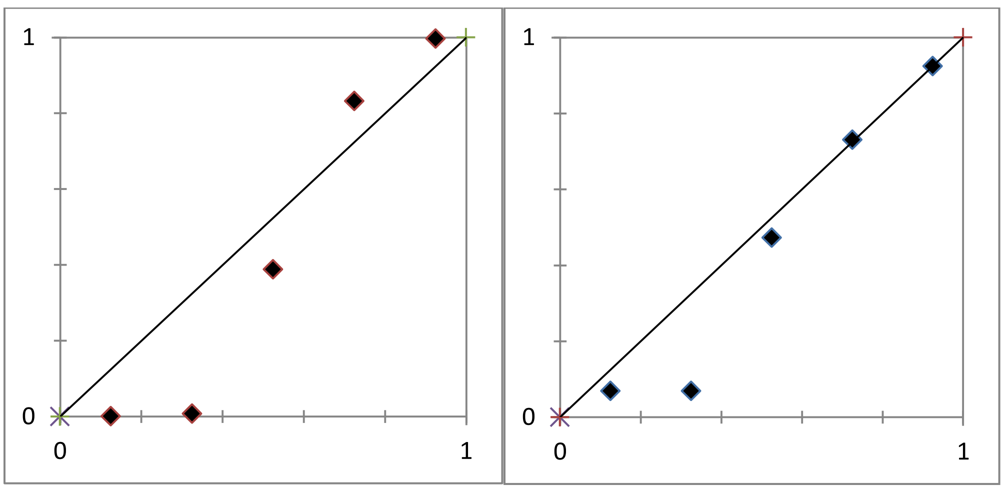


Figure 6. Five Projects: Calibrated Results vs. Linear Association

6.2. Validating Linear Association

We now replace the independence assumption by linear association, which is perhaps the simplest model for positive dependence. Let X_{jk} be an *estimate* of p_{jk}, and we treat it as a random variable. Considering all n_k activities of project k, we assume that $\{X_{jk}\}$ is a set of n_k independent positive random variables. This assumption is important because it is difficult to actually obtain dependent estimates in practice, but we do not impose the same assumption on $\{p_{jk}\}$. Let the *nominal* estimate be $e_{jk} = \mathrm{E}(X_{jk})$, and let $\mathrm{V}(X_{jk})$ denote the variance of the estimate. Random effects that are specific to activity jk are captured by $\mathrm{V}(X_{jk})$. Processing times are linearly associated if $p_{jk} = BX_{jk}$ where B is a positive random variable, independent from $\{X_{jk}\}$. We assume that B is lognormal with parameters $s = s_B$ and $m_B = \ln(\mathrm{E}(B)) - s_B^2/2$. The random variable B models estimation bias for all projects, and b_k is the particular bias realization that applies to project k. Let

μ_B and $V_B = \sigma_B^2$ denote the mean and variance of B. Finally, let L_k denote the duration of a serial project with n_k activities. Our main interest is in the distribution of L_k. Because B and X_{jk} are independent, we have $\mu_{jk} = \mu_B e_{jk}$. (for any k). However, the multiplication of the processing times by the same realization, $B = b_k$, introduces dependence between any pair of processing times in the same project. Specifically, we obtain

$$\sigma_{jk}^2 = (\mu_B^2 + V_B)V(X_{jk}) + V_B e_{jk}^2 = E(B^2)V(X_{jk}) + V_B e_{jk}^2$$

for the variance of activity jk, whereas the covariance of any two activities within the same project is given by

$$COV(p_{ik}, p_{jk}) = \sigma_{ij} = V_B e_{ik} e_{jk} \quad ; \quad i \neq j$$

We require $i \neq j$ in the covariance expression because σ_{jk}^2 is not given by $V_B e_{jk}^2$. However, $V_B e_{jk}^2 = (\sigma_B \Sigma e_{jk})^2$ is the second part of the expression for σ_{jk}^2. Therefore,

$$E(L_k) = \mu_B \sum_{j=1}^{n_k} e_{jk} \quad \text{and} \quad V(L_k) = E(B^2) \sum_{j=1}^{n_k} V(X_{jk}) + (\sigma_B \sum_{j=1}^{n_k} e_{jk})^2$$

The element $(\sigma_B \Sigma e_{jk})^2$ imposes a lower bound of $\sigma_B \Sigma e_{jk}$ on the standard deviation of the duration. If e_{jk} tends to underestimate (overestimate) μ_{jk} then μ_B is larger than 1 (smaller than 1) to compensate, but there is no effect on B's coefficient of variation $cv_B = \sigma_B/\mu_B$. Therefore, in contrast to the case of independent activity durations, the *cv* of project length does not tend to zero as the number of activities on the chain grows large: it always exceeds cv_B. Similarly, the bias in the estimate of mean project length is a multiple of the estimate, namely, $(\mu_B - 1)$.

Our task in this section is to show that this model makes possible estimating activity duration distributions reliably. (Even though we assume a serial structure, we can use linearly associated distributions to simulate projects with any network structure.) Because our dataset is small, we use the bootstrap method again. But in contrast to the previous case, where every activity was re-sampled from the full list of activities in the historical projects, here each simulation repetition is based on re-sampling just one of the historical projects, with a frequency that is proportional to n_k. For instance, if we were to simulate a run for Project 3 in Table 1, we would randomly select either Project 1 (with probability 5/9) or Project 2 (otherwise). Suppose Project 2 is selected, then we sample three ratios out of the set {1.00, 0.50, 2.00, 0.67}, with replacement, and use them to calculate the run's duration as we did before. As a result, if there is indeed a systemic difference in the mean of each historical project, that difference will manifest as variation *between* repetitions. Thus, the variance we observe for L_k will be larger, reflecting $s_B^2 + s_k^2$, where s_k^2 is the value obtained for project k by a normal Q-Q chart *after* the effect of b_k is incorporated. Again we use the empirical distributions obtained by the simulation to test whether they are a sample from a uniform U[0, 1] distribution. The sorted values are listed below and depicted in the left side of Figure 6.

{0.0686, 0.0692, 0.4722, 0.7301, 0.9243}.

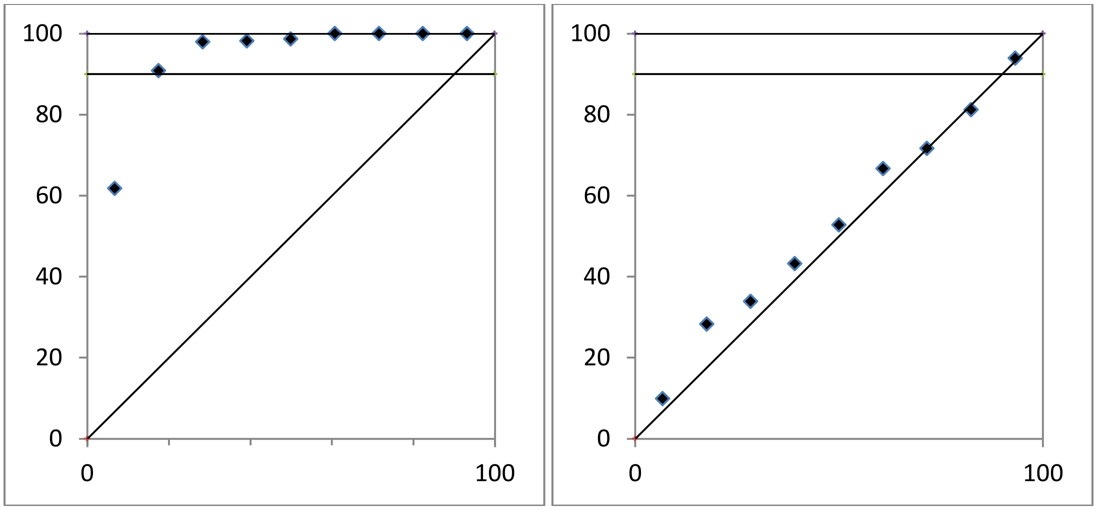


Figure 7. Nine Projects Results by PERT vs. Linear Association

In this instance, there are no points in the two bands below 5% and above 95%. Using the binomial $B(5, 0.1)$ again, the probability of such an event is 59%, and thus it is a plausible result. The sample also passes the K-S test with a higher p-value than the previous one, but that is practically meaningless in this instance. Figure 7 is similar to Figure 6, but it relates to the nine-project family. In that case, by far the best results are obtained under the linear association model, as depicted on the right. On the left, however, we made an adjustment first. To explain that, we point out that our bootstrap application automatically removes any estimation bias. In the five-project case there was some estimation bias but it was not as strong as in the nine-project family. Therefore, for the nine-project case, instead of using $F_k(C_k)$, we first adjusted C_k in such a way that the original bias of the estimate was restored. Equivalently, we could adjust all the ratios in such a manner that their expected effect would be unbiased. As is clear from the left side of Figure 7, bias plays a crucial role in the nine-project family. This time the K-S test rejects the sample with a p-value of 0.000. Most points fall in a band of 10% between 90% and 100%, and a binomial test shows that the event of eight points out of nine falling either in the bottom or top 10%, which follows a binomial $B(9, 0.2)$, is negligible (about 0.00002). By contrast, better results than those on the right might be suspect as too good to be true. In conclusion, although the data is meager, we showed that linear association leads to a plausible empirical project distribution estimate. It is powerful not only for better variance estimation when positive dependence exists in the data but also as a way to remove bias.

We can interpret linear association as a model that accepts statistical independence *within* projects but rejects it *between* projects. This interpretation suggests that we can use standard ANOVA to test the independence assumption, as it is equivalent to the null hypothesis that there is no difference between groups. Indeed, such ANOVA rejects independence for the nine-project family with a p-value of 0.0215, so it is significant at the 0.05 level. The p-value of the five-project family is about 0.0001; however, if we remove one particular project from consideration, the p-value for the remaining four increases to 0.0144 (but remains significant). The pertinent point is that the between-projects variation must not be ignored. For example, Strum et al. (2000)

show that one of the significant explanatory variables for operation time is the surgeon; the other two are the type of operation and the type of anesthesia. In effect, they show that if we consider sets of log-transformed operation times distinguished by surgeon, each has a different mean. But that is precisely what we observe for the five-project family. It follows that to predict the time it would take an unknown surgeon to perform an operation of a particular type with a particular anesthesia, we could use linear association. The between-surgeons variation would then be modeled by *B*.

6.3. Linear Association in the Hill Dataset

We now proceed to study whether the Hill et al. (2000) data exhibits linear association. The data consisted of 506 programming tasks, measured in units of 0.1 days and ranging from 0.1 to 255.4 days, with estimates between 0.5 and 200. Tasks had between 1 and 133 subtasks and required between 1 and 16 programmers. Work was coordinated by six managers who also recorded subjective mean estimates for each task in advance. However, the dataset does not include information about the duration of individual subtasks. We obtained the data, and we use it to show that it is consistent with linear association. To that end we analyze the data from each manager separately, and in one case we trimmed the sample to account for a strong Parkinson effect. That left a sample of 487 tasks, with between 54 and 107 allocated to each manager. Formally speaking, in what follows we may need a triple index of the form *jki* to denote subtask *j* of task *k* under manager *i*. But because we analyze the data for each manager separately, we omit the third index *i*. For each manager, we fit a linear regression model to the log-transformed data using the number of subtasks, n_k, as the explaining variable. Hill et al. noted that this variable was the best choice for explaining the total time required (although they did not use the transformation). Indeed, we found that it provided much tighter estimates than the original subjective estimate, which we thereafter ignore. Among other things, this regression yields residuals and a standard error, SE_Y. We refer to SE_Y^2 as the *overall variance*. Usually, if the residuals are normal, we can estimate *p*-values for each of them. Those *p*-values should be a sample from a standard uniform distribution. In our case, we have good reason to doubt that the residuals are normal, however.

Under the independence assumption, task *k*'s duration is given by the convolution of n_k independent subtasks, denoted X_{jk} (or simply X_j where no confusion may arise). To model linear association, without loss of generality, assume that *B* is lognormal with parameters $s = s_B$ and $m = -s_B^2/2$; that is, E(*B*) = 1. We convolute the X_{jk} elements before multiplying by the bias term, *B*, and we use the lognormal central limit theorem to approximate the convolution of the sum. If the parameters of the convolution are m_X and s_X^2 we obtain parameters $m_k = m_X - s_B^2/2$ and $s_k^2 = s_B^2 + s_X^2$. The value of s_k^2 denotes the variance of project *k* in logarithmic terms, as it might be estimated before the project is run (so there is yet no evidence on the size of b_k), and similarly, m_k is the mean in logarithmic terms. Let s_1^2 denote the variance of a single subtask in log-transformed space (under this manager), and let cv_1^2 denote the squared coefficient of variation of such a subtask without the transformation, then s_k^2 is the following function of n_k,

$$s_k^2(n_k) = s_B^2 + \ln\left(1 + \frac{cv_1^2}{n_k}\right) = s_B^2 + \ln\left(1 + \frac{\exp(s_1^2) - 1}{n_k}\right) \qquad (6)$$

In our case, the overall variance can be approximated by the average variance of all tasks. Suppose now that we guess a value for s_1^2; then we can use Equation (6) and the estimated overall variance to calculate s_B^2. For the guessed s_1^2 we can calculate the desired p-values, because we now know how to correct for the heteroscedasticity that we expect. If the set of p-values is not standard uniform for a particular s_1^2 guess, we might search for a better fit by changing the guess. Indeed, it is straightforward to search for the best guess. We conducted the search with two objectives: (*i*) to pass the K-S test, and (*ii*) to avoid excessive numbers of instances between 0% and 1% and between 99% and 100%. We did so for all six managers, and Table 2 shows the results. The table lists the number of tasks, the regressed m and s for all tasks of that manager, s_1, s_B, the relative contribution of s_B^2 to the overall variance (*ov*), the p-values of the K-S test results and the number of points in the two one-percent bands. The binomial test proved much more sensitive to bad selections than the K-S test. Nonetheless, we found values that satisfied all tests. We see that the relative contribution of s_B^2 is not below 9% but typically it exceeds 40% and may even reach 100%; in this context 100% implies that the effect of s_1^2 is negligible. Thus linear association can explain the results of six managers out of six. In the case of Manager 3, it is also plausible that the results are independent, but that is a special case of linear association. All the binomial counts are plausible.

	Manager 1	*Manager 2*	*Manager 3*	*Manager 4*	*Manager 5*	*Manager 6*
Tasks	107	75	64	81	54	106
m	0.278	0.333	0.223	0.223	-0.102	0.326
s	1.016	1.009	1.094	1.094	0.974	0.996
s_B^2	0.498	0.394	0.045	0.556	0.625	0.375
s_1^2	0.390	0.413	0.971	0.238	0.000	0.885
$s_B^2/\,ov$	0.679	0.453	0.088	0.807	1.000	0.445
K-S p-value	0.317	0.304	0.866	0.522	0.322	0.192
Binomial #	1	2	0	0	1	1

Table 2. Software Development under Six Managers

7. Conclusions and Topics for Future Research

We presented theoretical arguments and field evidence for the suitability of the lognormal distribution as a model for activity durations. The lognormal is positive and continuous, so it exhibits face validity. Nonetheless, practical data is recorded in discrete units—sometimes coarsely. This causes ties in the data, so we must account for such ties in our distribution fitting and hypothesis testing. Furthermore, as predicted by Parkinson's Law, we may encounter many activities that are reported as taking precisely their estimated time. To model the Parkinson effect, we generalized the Parkinson distribution to allow a fraction of early activities to be reported correctly whereas the others are misreported as "on time." We showed examples where both effects were present, yet after accounting for them, the lognormal remains plausible (cannot be rejected statistically). We also showed that projects are unlikely to be composed of truly independent activities.

By invoking linear association to model statistical dependence, we were able to achieve plausible results. Thus we propose linear association as an effective model for dependence in stochastic scheduling. It not only provides a plausible fit but is also tractable. Linear association appears to be the simplest model of dependence that fits the data reasonably well. Furthermore, it is especially tractable for the lognormal distribution in conjunction with the lognormal central limit theorem.

Previous attempts to model activity times for practical applications are associated with PERT, where several parameters are estimated subjectively by experts. Those attempts either rely on the PERT-beta or propose other distributions (including the lognormal) that emulate the PERT elicitation approach. The PERT approach is flawed in several important ways, however. The PERT-beta distribution cannot support practical cases with high *cv*, and we presented cases with high *cv*. It also requires elicited parameters that may be unreliable. When regression-based estimates are possible, we strongly recommend them instead. Otherwise, it may well be better to estimate only one parameter and use some conservative figure for the coefficient of variation.

Although we validated our hypothesis on several datasets, and we encountered no instance where the lognormal random variable was inappropriate (after correction for the Parkinson effect and for ties), we encourage further validation studies as a topic for further research. One issue that may arise is possible leptokurtosis due to subjective estimates: the Hill dataset shows some evidence to that effect. However, because that dataset does not include observations for individual subtasks, we cannot yet provide a confirmation. New research may also lead to better insights for the Parkinson distribution. Although our current version of that distribution seems adequate, it is plausible that only short tasks are subject to the Parkinson effect in some environments. We observed some evidence of that in the Hill database. Thus, there may be room for further development of the Parkinson distribution, but it would require much more data than we now have. Likewise, there is room to explore the value of more sophisticated models than linear association. For example, it may be beneficial to consider several independent common factors, each applying to a subset of activities. Last but not least, some preliminary evidence due to Lipke et al. (2009), in their Table 1, indicates that similar analysis also applies to cost. However, a persuasive validation requires cost-specific research. Such research may also address the relationship between cost and schedule performance.

Based on our results, we believe that stochastic models in scheduling should, as a first cut, represent activity times with lognormal distributions exhibiting the same coefficient of variation, and linear association should be used to represent stochastic dependence. Some lognormal models can be solved by mathematical programming approaches; i.e., they do not require simulation (Baker & Trietsch 2010). Other lognormal models can be solved approximately using similar analytic models. Thus, lognormal scheduling is a challenge not only for stochastic scheduling modelers but also for mathematical programming experts.

Finally, to validate our results in cases where the Parkinson effect is involved, we relied on trimmed samples, thus effectively discarding potentially useful data to render the remaining sample complete. It would be advantageous to develop statistical tests that do not require trimming. Powerful normality tests are akin to testing whether the R^2 value of the appropriate Q-Q chart is sufficiently high (Filliben 1975, Looney & Gulledge

1985). Essentially, such tests rely on simulated results, although in some cases empirical distributions have been fitted to them (Royston 1993). Verrill & Johnson (1988) extended the same approach to censored data, where censoring removes one of the tails of the distribution. Similarly, it should be possible to develop simulation-based tests tailored for the Parkinson distribution with lognormal core.

Acknowledgements: We are grateful to William Graf, Arman Navasardyan (Charity and Development Project Coordination Specialist, Action Against Hunger—ACF-USA), Seda Sahakyan, Kristine Sargsyan (Development Project Coordinator), and Lyn C. Thomas (University of Southampton) for generously providing us with field data. David Robb (Tsingua University) provided useful references and helpful comments on a precursor of this paper.

References

Baker, K.R. and D. Trietsch (2009) *Principles of Sequencing and Scheduling*, John Wiley & Sons, New York.

Baker, K.R. and D. Trietsch (2010) Research Notes for Chapter 6 in *Principles of Sequencing and Scheduling* (Wiley, 2009).
URL: http://mba.tuck.dartmouth.edu/pss/.

Blom G. (1958) *Statistical Estimates and Transformed Beta Variables*, John Wiley & Sons, New York.

Bruno, J., P. Downey and G. Frederickson (1981) Sequencing Tasks with Exponential Service Times to Minimize the Expected Flow Time or Makespan, *Journal of the Association of Computing Machinery* 28, 100-113.

Clark, C.E. (1961) The Greatest of a Finite Set of Random Variables, *Operations Research* 9, 145-162.

Clark, C.E. (1962) The PERT Model for the Distribution of an Activity Time, *Operations Research* 10, 405–406.

Esary, J.D., F. Proschan and D.W. Walkup (1967) Association of Random Variables, with Applications, *Annals of Mathematical Statistics* 38, 1466-1474.

Filliben, J.J. (1975) The Probability Plot Correlation Coefficient Test for Normality, *Technometrics* 17, 111-117.

Grubbs, F.E. (1962) Attempts to Validate Certain PERT Statistics or Picking on PERT, *Operations Research* 10, 912-915.

Gutierrez, G.J. and P. Kouvelis (1991) Parkinson's Law and its Implications for Project Management, *Management Science* 37, 990-1001.

Hahn, E.D. (2008) Mixture Densities for Project Management Activity Times: A Robust Approach to PERT, *European Journal of Operational Research* 188, 450-459.

Hill, J., L.C. Thomas and D.E. Allen (2000) Experts' Estimates of Task Durations in Software Development Projects, *International Journal of Project Management* 18, 13-21.

Johnston, J. and J. Dinardo (1997) *Econometric Methods,* Fourth Edition, McGraw-Hill/Irwin, New York.

Kelley, J.E. (1961) "Critical-Path Planning and Scheduling: Mathematical Basis," *Operations Research* 9(3), 296-320.

Klingel, A.R. (1966) "Bias in PERT Project Completion Time Calculations for a Real Network," *Management Science* (Series B) 13, B-194-201.

Kotiah, T.C.T. and N.D. Wallace (1973) Another Look at the PERT Assumptions, *Management Science* 20, 44-49.

Ku, P.-S. and S.-C. Niu (1986) On Johnson's Two-Machine Flow Shop with Random Processing Times, *Operations Research* 34, 130-136.

Lipke, W., O. Zwikael, K. Henderson and F. Anbari (2009) Prediction of Project Outcome: The Application of Statistical Methods to Earned Value Management and Earned Schedule Performance Indexes, *International Journal of Project Management* 27, 400-407.

Looney, S.W. and T.R. Gulledge (1985) Use of the Correlation Coefficient with Normal Probability Plots, *The American Statistician* 39(1), 75-79.

Malcolm, D.G., J.H. Roseboom, C.E. Clark and W. Fazar (1959) Application of a Technique for a Research and Development Program Evaluation, *Operations Research* 7, 646-69.

May, J. H., D. P. Strum and L. G. Vargas (2000) Fitting the Lognormal Distribution to Surgical Procedure Times, *Decision Sciences* 31, 129-148.

Mazmanyan, L., V. Ohanian and D. Trietsch (2008) The Lognormal Central Limit Theorem for Positive Random Variables (Working Paper).

Mohan, S., M. Gopalakrishnan, H. Balasubramanian and A. Chandrashekar (2007) A Lognormal Approximation of Activity Duration in PERT Using Two Time Estimates, *Journal of the Operational Research Society* 58, 827–831.

Parkinson, C.N. (1957) *Parkinson's Law and Other Studies in Administration*. Houghton-Mifflin.

Premachandra, I.M. and L. Gonzales (1996) A Simulation Model Solved the Problem of Scheduling Drilling Rigs at Clyde Dam, *Interfaces* 26(2), 80-91.

Premachandra, I.M. (2001) An Approximation of the Activity Duration Distribution in PERT, *Computers and Operations Research* 28, 443-452.

Robb, D.J. (1992) Scheduling in a Management Context: Stochastic Tasks and Non-Regular Reward Functions, Unpublished Doctoral Dissertation, Faculty of Management, University of Calgary.

Robb, D. J. and E. A. Silver (1993) Single-Processor Scheduling with Stochastic Tasks and Non-Regular Performance Measures, *Decision Sciences* 24 (6):1085-1108.

Royston, P. (1993) A Toolkit for Testing for Non-Normality in Complete and Censored Samples, *The Statistician*, 42, 37-43. See also Corrigendum (1999) *The Statistician*, 48, Part I, 158.

Schonberger, R.J. (1981) Why Projects are "Always" Late: A Rationale Based on Manual Simulation of a PERT/CPM Network, *Interfaces* 11, 66-70.

Shapiro, S.S. and R.S. Francia (1972) "An Approximate Analysis of Variance Test for Normality," *Journal of the American Statistical Association* 67, 215-216.

Shapiro, S.S. and M.B. Wilk (1965) "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika* 65, 591-611.

Shtub, A., J.F. Bard and S. Globerson (2005) *Project Management: Processes, Methodologies, and Economics*, 2nd edition, Pearson/Prentice Hall, Upper Saddle River, NJ.

Soroush, H.M. and L.D. Fredendall (1994) "The Stochastic Single Machine Scheduling Problem with Earliness and Tardiness Costs," *European Journal of Operational Research* 77, 287-302.

Strum, D.P., A.R. Sampson, J.H. May, L.G. Vargas (2000) Surgeon and Type of Anesthesia Predict Variability in Surgical Procedure Times, *Anesthesiology* 92, 1454–1466.

Trietsch, D., L. Mazmanyan, L. Gevorgyan and K.R. Baker (2010) A New Stochastic Engine for PERT, Working paper. http://mba.tuck.dartmouth.edu/pss/Notes/PERT21.pdf

Tversky A. and D. Kahneman (1974) Judgment under Uncertainty: Heuristics and Biases, *Science* 185, 1124-1131.

Verrill, S. and A. Johnson (1988) Tables and Large-sample Distribution Theory for Censored-data Correlation Statistics for Testing Normality, *Journal of the American Statistical Association* 83, 1192-1197.

Woolsey R.E. (1992) The Fifth Column: The PERT that Never Was or Data Collection as an Optimizer, *Interfaces* 22, 112-114.