The effect of systemic errors on optimal project buffers

Dan Trietsch *

ISOM Department, University of Auckland, Old Choral Hall, 7 Symonds Street, Auckland City 1001, New Zealand

Received 7 September 2004; received in revised form 22 October 2004; accepted 3 December 2004

Abstract

Existing mathematical models for setting buffers for time or cost in project management assume that project activities are statistically independent. This leads to a highly counterintuitive and damaging conclusion that project buffers should become relatively negligible for projects with long chains of activities. We present a model that considers the statistical dependence between activities caused by estimation bias. We show that if relatively high service levels are desired, this imposes a positive lower bound on the buffer as a data-based fraction of the estimated project duration or budget. We also introduce a new approach for collecting data and estimating the parameters necessary to implement the model. This approach places a smaller burden on decision makers than traditional PERT: they provide single point estimates for means, while variance elements and bias correction are computed electronically using historical data.

© 2005 Elsevier Ltd and IPMA. All rights reserved.

Keywords: Time; Cost; Operations

1. Introduction

Recent developments in the practice of project scheduling focus attention on the correct specification of safety buffers. Time buffers are used to protect the schedule, and, similarly, budget contingencies are used for cost control. For example, to prevent wasting earliness, a project buffer is placed at the end of the project schedule. Similarly, where chains of activities merge with the critical path, feeding buffers may be inserted. Note that to specify such buffers is mathematically equivalent to deciding when to start each chain of activities. Yet the size of these buffers is usually specified as an arbitrary fraction of the estimated chain duration. The purpose of this paper is to provide a first step towards optimizing such buffers based on plausible theory and relevant data analysis.

In this paper we study the most basic case: the optimal buffering of projects with \( n \) (not necessarily independent) activities in series, and without intermittent idling. Thus there is only one chain of activities and it is necessarily also the critical path. Assuming that either the due date or the start time is a decision variable, the project buffer is the difference between the due date and the expected completion time. Similar analysis is directly applicable to the determination of a cost buffer as well, by the simple substitution of the due date by a project budget. The main cost of a typical project is the sum of the individual activity costs, although it also includes overhead charges and holding costs that depend on the project duration. Therefore, at least approximately, we can say that cost is additive and not sensitive to the project network structure. Similarly, due to the simplicity of the serial structure, if we assume that earliness is utilized, then the expected duration of the project is equal to the sum of the expected durations of the activities. Our result for the serial structure can serve to develop insights for more realistic project networks, where such
chains are often embedded. In such a case, the (serial) project buffer becomes a feeding buffer for non-critical chains feeding other project activities; e.g., feeding the critical path. Such chains are sub-projects with a serial structure. A useful approach for extending the results we obtain here to more complex project networks is by simulation, but this is beyond our scope.

Henceforth, we will discuss time buffers, keeping in mind that the results also apply to contingency budgeting. For this, we need to know the cumulative distribution function of the project completion time, which, in turn, is a function of the individual activities’ distributions. In Malcolm et al. [1], the PERT developers discussed the need for stochastic estimation of activity durations. Nominally, they proposed fitting a beta distribution. Practically, they required a triplet of estimates for each activity – {min, mode, max} – and then arbitrarily set $standard\ deviation = (max \ - min)/6$ and $mean = (min + 4 \ mode + max)/6$. Assuming independence, they noted that the sum of a serial chain of activities is approximately normal by the central limit theorem. Although they recommended it for tractability, the team recognized that the use of the most likely critical path is optimistic: it ignores the statistical dependence between paths that share activities. They already had a computationally intensive remedy: Clark [2] provided approximate but unbiased calculations for the project duration distribution. Finally, they highlighted a major concern about systemic bias – which is at the heart of this paper and we will discuss it later. Armed with 20/20 hindsight, one may criticize invoking the beta distribution, considering that it was not really utilized, but this is harmless. More seriously, it is difficult, perhaps impossible, to obtain reliable triplets from process owners. Anecdotal evidence to this effect, related to work place politics, was given by Woosley [3]. To this we must add systemic human error as studied by Tversky and Kahneman [4]; e.g., they observed that 98% confidence intervals estimated by experts tend to fail in about one third of the cases (i.e., experts tend to grossly underestimate the true variance).

Britney [5] presented optimal buffers for stochastic project activities on a one-by-one basis. Planned idling occurs between activities unless the preceding buffer is exceeded. He did not, however, optimize the project as a whole. Also, he did not consider the possibility of combining serial buffers. The combined buffer approach was used by Yano [6], who obtained the optimal single buffer for a project (or supply chain) composed of independent serial activities. Yano [7] inserted individual buffers between the serial activities. Ronen and Trietsch [8], Kumar [9], and Chu et al. [10], independently, optimized the ordering times of $n$ parallel project (or supply chain) items with stochastic lead times, where the latest one determines the completion time; i.e., they showed how to coordinate $n$ parallel feeding buffers.

This idea – using planned project- and feeding buffers to achieve improved reliability and avoid [implicit] project delay penalties – had recently been popularized under the title “Critical Chain”, and many practitioners find it useful. Reviews and discussions of this modern development abound; e.g., Herroelen and Leus [11], Leach [12], Raz et al. [13], Trietsch [14]. The basic assumption is that planned idling should not occur within the critical path or within any chain of activities. Nonetheless, it is incorporated where such chains merge (yielding feeding buffers). The size of protective buffers is determined as an arbitrary fraction of the expected chain leading to the buffer (e.g., 50%). Leach [12] suggested the maximum of a buffer based on the traditional independence assumption and the arbitrary fraction. Leach [15] listed 11 reasons why projects are typically longer than expected. He then proposed a larger buffer based on the sum of the two elements mentioned before.

Although Clark’s approximation can handle dependent activities, none of the existing mathematical models for buffer setting accounts for systemic error. We introduce an elementary model for this purpose and show that it yields dependency between activities. We show that if relatively high service levels are desired, this imposes a positive lower bound on the project buffer as a data-based fraction of the expected duration. In contrast, the independence assumption leads to a highly counterintuitive and damaging conclusion that project buffers should become relatively negligible for projects with long chains of activities. We also recommend an approach for collecting data and estimating the parameters necessary to implement our model. This approach places a smaller burden on users than traditional PERT: single point estimates are required instead of triplets. Variance and bias estimates can then be computed by a decision support system that uses historical data, and we show how.

The remainder of the paper is organized as follows. After introducing the main notation, Section 2 presents the [existing] basic buffer optimization model subject to a given project distribution. Section 3 introduces our model with the bias-induced correlation. Section 4 addresses the estimation of the necessary model inputs. Section 5 provides numerical illustrations. Section 6 is the conclusion.

2. Notation and basics

$B$ the bias of time or cost estimates of project activities (a random variable)

$\beta$ $E(B)$, where $E(\cdot)$ is the expected value function

$V_b = \sigma^2_b$ $V(B)$, where $V(\cdot)$ is the variance function

$C$ the cost to postpone the due date (per time unit)
We assume that we have to determine a project due date, $D$, or that one is given with enough slack to make possible a delayed start. The time from the project start to the due date is the planned duration. For convenience, we will treat the start as time zero, so the planned duration is equal to $D$. We also assume that there is an economic cost per time unit, $C$, which provides an incentive to reduce the planned duration; e.g., the customer is more likely to award us the contract if there is an economic cost per time unit, $C$.

3. Modeling positive dependence due to systemic error

It is well known that project activities are sometimes statistically dependent. Nonetheless, with one exception, the literature on setting buffers assumes statistical independence. This includes academic papers and practitioner books. But if we increase the number of activities along the critical path, the independence assumption leads to so-called “optimal” project buffers that, as a fraction of the mean, approach zero asymptotically. Because this is a highly counterintuitive conclusion, the practical approach has always been to set the buffer, or the contingency, as a fraction of the anticipated duration; e.g., see O’Brien [16]. In addition to leading to such a highly counterintuitive conclusion, and to flying in the face of experience, the independence assumption should be challenged on theoretical grounds as well. Leach [15], after citing empirical evidence demonstrating that the independence assumption is not valid in practice, identifies positive bias as the culprit. He then provides 11 causes of positive bias – mostly related to large projects with multiple paths. Our focus here is on a single cause of positive bias, the common tendency to underestimate the durations (and severe punishment for missing due dates in the past will turn optimists to pessimists). The result is a systemic bias across a project. But the magnitude of this bias is a random variable and even its orientation is not known in advance. This introduces a strong...
dependence as far as deviations from plan are concerned. But, operationally, the only variation that concerns us is relative to plan!

Notably, the need to account for estimation bias by some calibrating program has been on the agenda from the earliest PERT days. Malcolm et al. [1] stated that the problem had been raised by many. They then suggested “a program of comparing estimates with actual performance over a period of time to permit ‘calibration’ of the estimators.” Alas there is no evidence that such calibration was ever implemented on a wide scale, if at all. Instead, subsequent papers about bias focused mostly on the optimistic bias due to ignoring near-critical parallel paths that may become critical in reality; e.g., Klingel [17], Schonberger [18]. Thus, the clear recommendation that the PERT team expressed in [1] was ignored, while an issue that they considered more minor – and for which [2] had already provided an approximate remedy – was highlighted.

Furthermore, [1] stated objective with respect to bias was to calibrate the estimates to make them accurate. But there is also a statistical dependence issue that arises. For example, suppose we suspect that there is a random bias error of ±25%, and we find that the first seven activities took 85% of their combined estimated time, then we would probably consider it likely that the next activities will also tend to consume around 85% of their estimate – in other words, we form an informal estimate of 85% for the necessary calibration. This means that we perceive a correlation between the first few activities and the ones that are yet to follow. This, by definition, is a case of statistical dependence (because independent variables are not correlated). In practice, we need to know not only the average bias but also its impact on the covariance of activities. Our purpose here is to address both the need for calibration and the magnitude of the correlation that is involved, so we can draw conclusions for the optimal combined buffer (including elements for bias correction and for safety).

Let the true activity times compose the random vector \( \mathbf{Y} = \{Y_i\} \). If decision makers would have the distribution of \( \mathbf{Y} \), there would be no need for this paper. In reality, however, decision-makers never know \( \mathbf{Y} \), so they must use some estimate that acts as a model of \( \mathbf{Y} \). Accordingly, let \( \mathbf{X} = \{X_i\} \) be a vector of the decision-maker’s models of the actual activity durations, \( \{Y_i\} \). So \( X_i \) is a random variable that represents another random variable. Estimates are based on \( \mathbf{X} \), and not directly on \( \mathbf{Y} \). Systemic error arises because \( \mathbf{X} \) is not a perfect model of \( \mathbf{Y} \). Mathematically, perhaps the simplest possible model for systemic error involves the introduction of an additional independent random variable, \( \mathbf{B} \), that multiplies \( \mathbf{X} \) to obtain \( \mathbf{Y} \). In this paper, we limit ourselves to this basic model. Other potential approaches exist, however, and further research may be justified to identify the best one. We will use \( \beta \) and \( V_b = \sigma_b^2 \) to denote the mean and variance of \( \mathbf{B} \). We assume that \( e_i \), the nominal (single point) estimate of \( Y_i \), is proportional to \( E(X_i) \). For example, if a particular provider of estimates includes a hidden 100% buffer in her estimates, then \( e_i = 2E(X_i) \). Nonetheless, for simplicity (and without loss of generality), we will set \( e_i = E(X_i) \). As long as the assumption that \( e_i \) is proportional to \( E(X_i) \) holds, there always exists a \( \mathbf{B} \) that corrects any deviation introduced by this simplification.

In broad terms, the randomness of \( X_i \) relates to chance events that are specific to \( Y_i \) (as perceived), e.g., problems with raw materials or power supply. \( \mathbf{X} \) also includes some (but not all the) effects of randomness in estimation, since estimates are the result of processes that are not deterministic. Even with the same data, different people at different times typically come up with different estimates. For example, they may forget or neglect different aspects of the job. Random estimation errors are operationally equivalent to random deviation from plan. Because perceived activity-specific chance events and some estimation errors are confounded with each other, they must be represented by the same random variable, \( X_i \). In contrast, \( \mathbf{B} \) models effects that are common to all activities, such as pressure to produce attractive estimates quickly (leading to optimistic bias and omissions), personal safety buffers, weather, economic conditions, etc. Note that \( \mathbf{B} \) is important even if \( \beta = 1 \); i.e., if we take steps to remove bias on average, as suggested by [1], and thus achieve accuracy, \( \mathbf{B} \) would still capture important information about systemic impression. Specifically, those estimation errors that are not confounded with activity-specific chance events.

To present the most basic model, we assume that the elements of \( \mathbf{X} \) are independent of each other. This assumption is often realistic, by which we mean that decision makers typically estimate the elements of \( \mathbf{Y} \) (by the elements of \( \mathbf{X} \)) as if they are independent and therefore the elements of \( \mathbf{X} \) are indeed independent – after all, \( \mathbf{X} \) is a model only. (Modeling dependence into \( \mathbf{X} \) may be useful for some purposes, but requires further research.) Because \( \mathbf{B} \) and \( \mathbf{X} \) are independent of each other, \( \mu_i = \beta \cdot e_i \), but the multiplication by the same realization of \( \mathbf{B} \) introduces [positive] dependence between the elements of \( \mathbf{X} \) even if the elements of \( \mathbf{X} \) are independent. Specifically, \( \sigma_i^2 = \beta^2 \cdot V(X_i) + V(b) + \sigma_b^2 \), and \( \text{COV}(Y_i, Y_k) = V_b \cdot e_i \cdot e_k \forall i \neq k \). We can separate \( \sigma_i^2 \) into two parts, \( \beta^2 \cdot V(X_i) + V(b) \) and \( \sigma_b^2 \). The former equals \( E(B^2) \cdot V(X_i) \) and the latter is a special case of \( V_b \cdot e_i \cdot e_k \). Thus the covariance matrix of \( \mathbf{Y} \) is the sum of a diagonal matrix with elements \( V(X_i) \cdot E(B^2) \) and a full symmetric matrix with elements \( V_b \cdot e_i \cdot e_k \forall i, k \) (i.e., the vector product \( \{e_i \cdot e_j \} \cdot \{e_i \cdot e_j \}^T \)). To correct for the average bias we add a bias correction of \( (\beta - 1) \sum e_i \) to the nominal makespan estimate \( \sum e_i \). Thus we obtain a relative bias correction of \( \beta - 1 \). If \( \beta = 1 \), we
obtain the classic model with independence. Similarly, if $V(B) = 0$ but $\beta \neq 1$, then after the bias correction we again obtain the classic model. Therefore, our model generalizes the classic approach. Finally, monitoring a project over time during its execution always involves estimating the bias that applies to it, either explicitly or implicitly.

Bias correction is the first response to bias, but the standard deviation of the project completion time is vital for rational determination of the safety buffer. Since the elements of $X$ are independent, we have, $\sigma^2 = E(B^2) = \sum_{i=1}^{J} V(X_i) + (\sigma_{0i} \cdot \sum_{i=1}^{J} e_i)^2$. To see this, let $V$ denote the covariance matrix of $Y$, and let $I$ be a column vector in $R^J$ with elements of 1, then the result is obtained directly by the matrix product $I^T V I$. To study the effect of the standard deviation on the optimal buffer further, let $q_1(e) = E(B^2) = \sum_{i=1}^{J} V(X_i)$, and let $q_2(e) = \sigma_{0i} \cdot \sum_{i=1}^{J} e_i$, where $e_1 = \{e_i\}$, then

$$\text{max}\{\sqrt{q_1(e)}, \sqrt{q_2(e)}\} \leq \sqrt{q_1(e)} + \sqrt{q_2(e)}.$$

The central element in the inequality is $\sigma$. $\sqrt{q_2}$ is proportional to the nominal makespan estimate (before the bias correction), $\sum e_i$. So there exists a fraction of this estimate, namely $\sigma_{0i}$, that acts as a lower bound on $\sigma$. Therefore, if we wish to specify a safety buffer against random variation of $k \sigma$ for some $k > 0$, this safety buffer is bounded from below by $k \cdot \sigma_{0i} \cdot \sum_{i=1}^{J} e_i$, which constitutes a fraction of $k \cdot \sigma_{0i}$ of the nominal makespan estimate. Furthermore, when $n \rightarrow \infty$ then $q_1/q_2 \rightarrow 0$ so the same bound serves as the approximate optimal safety buffer. Recall that Leach [12] and Leach [15] suggest relative buffers that are associated with $\text{Max}\{\sqrt{q_1}, \sqrt{q_2}\}$ and $\sqrt{q_1} + \sqrt{q_2}$, respectively, so for a positive $k$ these values provide a lower- and an upper bound for the correct result. (There is no theoretical reason to limit ourselves to $k > 0$, so we do not limit our analysis to this case. Nonetheless, most project managers are uncomfortable with negative buffers and the low service levels they entail.)

4. Estimating model parameters

A tempting approach is to estimate $X_i$ by the classical 3-point method of PERT, thus yielding $e_i$ and $V(X_i)$. It would then only remain to estimate $B$. However, there are major difficulties with this approach (as discussed in Section 1), even without the new requirement to distinguish consciously between systemic bias and individual activity variation causes. Therefore we propose to limit the information elicited from process owners to single point activity estimates, $e_i$, and obtain all the other necessary estimates from historical data with the help of a computerized decision support system (DSS).

An ideal model-based DSS takes simple inputs and converts them to simple-to-use outputs. The conversion itself, however, has to be programmable and efficient but not necessarily simple. With this in mind, we need a programmable model to estimate $\beta$, $\sigma_{bn}$, and $V(X_i)$, and then compute $\sigma_{eij}^2$. This requires a two-stage econometric model, however, because $V(X_i)$ cannot be estimated until we estimate $\beta$.

The historical data we need consists of pairs of activity estimates and realizations. Even though our formal model assumes a serial structure, the network structure of the historical projects is immaterial. We treat activity estimates as explaining variables and realizations as dependent variables. Assume we have $J > 1$ projects in the history, and in this section we use the double index, $i,j$ where $i = 1, 2, \ldots, n_i$ and $j = 1, 2, \ldots, J$. Here $i$ denotes a specific activity and $j$ a specific project with $n_i$ activities. Optionally, we may elect to treat some subprojects as projects in their own right for this purpose. This is reasonable sometimes, e.g., if the subprojects were managed (and thus also estimated) relatively independently or involved distinct sets of resources. For project $j$ our data consists of $n_i$ pairs $(y_{ij}, e_{ij})$, where $y_{ij}$ is the realization and $e_{ij}$ the original activity estimate. One possible estimator of the mean systemic error of project $j$ is given by

$$\hat{\beta}_j = \frac{\sum_{i=1}^{n_j} y_{ij}}{\sum_{i=1}^{n_j} e_{ij}}; \ \forall j$$

leading to

$$\hat{\beta} = \frac{\sum_{j=1}^{J} \frac{n_j \hat{\beta}_j}{n_j}}{\sum_{j=1}^{J} n_j} = \frac{\sum_{j=1}^{J} n_j (\hat{\beta}_j - \hat{\beta})^2}{\sum_{j=1}^{J} n_j - J}.$$

This completes the first stage (and leaves $\sum n_j - J$ degrees of freedom for the second stage).

Our decision to only employ single point activity estimates implies that $V(X_{ij})$ must be modeled as a function of the activity estimate $e_{ij}$ itself. The parameters of this function can only be estimated from historical data. Suppose we start with a quadratic approximation of the form $V(X_{ij}) = x_0 + x_1 \cdot e_{ij} + x_2 \cdot e_{ij}^2$. We first note that, because $V(X_{ij}) \geq 0$, $x_k \geq 0$ is required. Furthermore, in the limit, as $e_{ij} \rightarrow 0$ we expect to obtain $\sigma_{eij}^2 \rightarrow 0$ as well, so $x_0 = 0$ is indicated. Accordingly, we obtain the model $V(X_{ij}) = x_1 e_{ij} + x_2 e_{ij}^2$. Let $CV_{bij}$ denote the coefficient of variation of $X_{ij}$, and recall that $E(X_{ij}) = e_{ij}$, then $CV_{bij} \geq \sqrt{x_2}$ ($\forall i,j$), with equality iff $x_1 = 0$. So if $CV_{bij}$ is approximately constant (for all $i,j$), our model will fit with $x_2$ dominating and $x_1$ may be insignificant. But if long activities often include several independent sub-activities, $x_1$ will be significant.
(To see this consider that if \( x_2 = 0 \) and we combine any number of activities with the same positive \( x_1 \) to one composite activity, its variance will be \( x_1 \) times its combined mean. But this is exactly what we would expect if these activities were independent.) Let \( \rho_{ij} \) denote the residual of \( x_{ij} \), then

\[
\rho_{ij} = \frac{y_{ij}}{\hat{y}_j} - e_{ij}; \quad \hat{\rho}_{ij} = \frac{y_{ij}}{\hat{y}_j} - e_{ij}
\]

and note that \( E(\rho_{ij}) = 0 \). We can now regress \( \rho_{ij}^2 \) by \( e_{ij} \) and \( e_{ij}^2 \) to obtain estimates for \( x_1 \) and \( x_2 \). Since both must be non-negative, if either estimate is negative then to minimize the mean squared error subject to our constraints (\( x_i \geq 0 \)) we must set the offender to zero and regress for the other value.

Over time, as more projects enter the data base, the estimates of \( \beta, \sigma_\beta, x_1, \) and \( x_2 \) should be updated regularly. We need an adaptive, self-calibrating, approach because the very use of a model such as ours is likely to influence the systemic error distribution over time. This can be based on exponential smoothing. One advantage of including such an adaptation mechanism is that it is possible to start the system with relatively inaccurate data and over time it will calibrate itself.

Once in place, the data collection burden this system places on participants is very low. The estimates they are required to provide are limited to the bare minimum – single point activity estimates. As for data to be collected during ongoing projects, since the estimates are already in the data base, it is enough to record the actual performance of each activity. One question remains: how to obtain initial values for such a scheme. This should not be too difficult for most project organizations, since the data required is very elementary. However, if such data is not available, we propose to use generic data from the relevant industry, or even resort to guesstimates. This is permissible because the system will calibrate itself eventually.

Finally, it may be reasonable to assume that \( B \) is lognormal (i.e., that it reflects the multiplicative effects of many small independent biases) while \( \sigma_\epsilon \) tends to be approximately normal. The product of a lognormal random variable and a normal one is neither normal nor lognormal, but – to an extent that depends on the variance elements – its shape is in between the symmetric normal and the skewed lognormal. Thus we may expect a skewed distribution, but not very skewed if \( \sigma_\epsilon / \beta \) is small. Leach [15] cites practical experience that deviations from plan are indeed skewed, as the model predicts. Rough analysis of data from his Fig. 2 [15, p. 37] suggests that time estimates in a particular company had \( \beta \approx 1.8 \) and \( \sigma_\beta \approx 0.55 \). Improvements reduced these to \( \beta \approx 1.06 \) and \( \sigma_\beta \approx 0.3 \). This analysis should be treated with caution, however, because it is based on second-hand data that was originally reported for another purpose.

5. Numerical illustrations

For simplicity, in spite of the recommendation to use a lognormal distribution for \( B \), in the following numerical examples we use an additive normal approximation for the bias effect. Here, the difference is not significant (because \( B \) does not possess a very high coefficient of variation).

5.1. Example 1

Suppose that by analyzing past data we obtain estimates of 1.25 and 0.15 for \( \beta \) and \( \sigma_\beta \) (if we would choose to model \( B \) by a lognormal distribution its parameters would be 0.2160 and 0.1196). This applies to any new project. For a particular project, suppose that our current estimate for a chain of activities is 28 weeks. The variance contribution due to \( \sigma_\beta \) is \( (0.15 \times 28)^2 = 4.2^2 = 17.64 \). Therefore, the standard deviation of the chain is now bounded from below by 4.2 (even when we ignore all other variances). The bias correction required is \((1.25 - 1)28\), i.e., we should add 7 weeks to the estimate.

Now suppose that \( \sum B(X_i) = 19.785 \). After multiplying by \( E(B^2) = 1.585 \), we obtain an updated contribution of 31.36. Taking dependence into account implies adding \( 4.2^2 \) to this value, yielding a total variance of about 49 (7²). If we require a project service level of, say, 95%, then, assuming a normal distribution, we need a safety buffer of \( 1.645 \sigma \), or 11.52, in addition to the bias correction. So our due date should be set to \( 28 + 7 + 11.52 = 46.52 \) (of course, in practice such a result would be rounded). The bias correction plus the safety time add up to 18.52, which is a large increase. But it is justified because our history suggests optimism and sizable dependence, and also because 95% is a high service level.

5.2. Example 2

For the same \( B \) as in Example 1. Suppose \( X \) comprises \( n \) iid activities, each with a normal distribution with \( e_i = 4 \). \( V(X_i) = 0.794^2 \) (so \( E(B^2) \cdot V(X_i) = 1.585 \times 0.794^2 = 1 \) for all \( i \)). Let \( P = 4C \). Determine the optimal project due date for \( n = 1, n = 5, \) and \( n = 25 \). Also, determine the asymptotic behavior of the relative optimal total buffer – defined as \( \beta - 1 + k \cdot \sigma \sum e_i \), i.e., the relative bias correction combined with the relative safety buffer – as a function of \( n \).

5.3. Solution

\( P = 4C \) leads to \( SL = 0.75 \). Using the normal approximation this implies a safety buffer of 0.6745\( \sigma \). In all cases,
we need bias correction elements of \( n \) \((0.25 \times 4 = 1)\); i.e., the bias corrections are 1, 5, and 25, respectively. For \( n = 1 \), our variance is \( 1 + 4^2 \times 0.15^2 = 1.36 = 1.166^2 \) leading to a total buffer of 1.787 (compare to 0.6745 that might be specified without the bias correction). This is 44.7% of the estimated mean. The due date is 5.787. For \( n = 5 \), the variance is \( 5 + 20^2 \times 0.15^2 = 14 \), leading to a due date of 27.524, and the total buffer is 37.6% of the estimated mean. \( n = 25 \) leads to a due date of 135.66, and the total buffer is 35.7% of the estimated mean. Asymptotically, we need a bias correction of 25% and an additional \( 15 \times 0.6745 = 10.1% \) of the estimated mean for safety, thus imposing a lower bound of 35.1% on the total size of the buffer as a fraction of the estimated mean. So most of the protection for \( n = 25 \) is due to expansion and dependence, while a sizable part of the protection for \( n = 1 \) is for processing time variation. Fig. 1 depicts the optimal due date and the total buffer for \( n = 1 \) to \( n = 10 \). We see that when expressed in percents the buffer approaches the limit asymptotically. We also see that the optimal buffer as a function of \( n \) follows an approximately straight line, but it has a positive intercept. Thus, short chains require relatively more protection.

6. Conclusion

The simplest non-trivial project structure is a serial chain of activities buffered by a single project buffer. Without correct analysis of such chains it is impossible to analyze more realistic structures. For project buffers to provide adequate protection, we must take into account positive dependence between activities due to systemic estimation errors and causes that are shared by many activities. We developed a model for this which yields project buffers that tend to grow approximately linearly with the expected estimated duration – i.e., they do not become negligible as the independence assumption implies. We also presented a plausible estimation model that only requires single point estimates and derives bias and variance elements from historical data, which is updated as new projects are completed. When necessary, this process can initialized by using generic industry data.

In the analysis of general project networks, the first step is often to reduce every embedded simple chain by convolution, as if independence applies. Our model suggests that this is inadequate. The next stage of this research should be the study of such networks. One way to
do this is by simulation, but incorporating the insights we provided.

References