Research Notes for Appendix A[*]

Some topics introduced in Appendix A require elaboration. First, regarding the *Parkinson distribution*, our text introduces the probability model and the distribution name for the first time. Gutierrez and Kouvelis (1991) study the stochastic effects of Parkinson's Law based on hypothesized mathematical models of work expansion behavior that might apply. They then compare the expected project completion time under various combinations of due date setting rules and work expansion models. Hasija, Pinker and Shumsky (2010) describe a complementary model that focuses on capacity rather than individual task times but captures a similar phenomenon. The Parkinson distribution is based on a simpler hypothesis than work expansion: earliness is hidden. Our empirical results so far support this hypothesis. Trietsch, Mazmanyan, Gevorgyan and Baker (2010) introduces a more general definition whereby only a fraction of early jobs are subject to the Parkinson effect. We elaborate in the next section, to complement our coverage of the original, *pure Parkinson*, version. More important than the name, the properties of the Parkinson distribution have not been studied extensively yet. One such property is intuitively clear: relative to the core random variable by itself, the Parkinson distribution decreases the variance but at the expense of increasing the mean.[†] For the pure version, we already stated this result—without proof—in the Appendix itself. Here, we prove it for the general version. The proof of the lognormal central limit theorem is another unpublished result, due to Mazmanyan, Ohanyan and Trietsch (2008). Essentially, it boils down to a simple observation that in the limit the lognormal becomes normal, so the regular central limit theorem applies. However, they also showed that the lognormal central limit theorem tends to provide a better approximation than the normal for the sum of a few strictly positive random variables, especially if the random variables are skewed to the right (as per our generic assumption). We repeat these results for completeness. The use of linearly-associated lognormal processing time distributions is just emerging. Trietsch (2005), the original paper on their application to project processing times, did not utilize the lognormal central limit theorem. That refinement—due to Trietsch, Mazmanyan, Gevorgyan and Baker (2010)—was actually motivated by our text. We also discuss how to use the Parkinson distribution in simulation. Preliminary evidence suggests the need to take account of correlations between the parameters involved. For instance, the probability of earliness being unreported seems to be strongly correlated with the common factor of the linear association, so each simulation run should select

---

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

[†] For the single machine piecewise linear safe scheduling template, the Parkinson distribution is wasteful (see the section *Trading Off Due-Date Tightness and Tardiness: The Weighted Case* in our Research Notes for Chapter 7). However, the project stochastic balance model of Chapter 18, by providing explicit release dates for subsequent activities, may call for intentional idling (without using earliness).

correlated values for these parameters (averages are not sufficient). We present one way to do that. Finally, we elaborate on the issue of stochastic dominance and we show that lognormal variates with the same *s* but different *m* parameters are stochastically ordered in the likelihood ratio sense, which implies stochastic ordering in the regular sense.

The Parkinson Distribution

As in our text, but with a slightly different notation, we define the pure Parkinson distribution as

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

where X is the random variable we can observe and Y is a core random variable that we can observe directly only if it exceeds the constant q. Recognizing that in practice *some* early realizations are observable, let p_E denote the probability that $Y \leq q$ and let p_P denote the probability that Parkinson's Law applies to an early processing time. Thus, if $Y \leq q$ then with a probability of p_P we obtain $X = q$ and with the complementary probability, $(1 - p_P)$, we obtain $X = Y \leq q$. By assuming that these probabilities apply independently to different processing times, we obtain the general form of the Parkinson distribution. Another way to write this is

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

where Q is a random variable such that if $Y \leq q$, an event with probability p_E, $Q = q$ with conditional probability p_P; otherwise, $Q = Y$. Finally, the pure Parkinson distribution is a special case with $p_P = 1$, whereas regular random variables can also be modeled as a special Parkinson case with $p_P = 0$.

Bounding the Variance of the Parkinson Distribution

Our purpose is to prove that under the Parkinson distribution $V(X) \leq V(Y)$ but $E(X) \geq E(Y)$ (and if one of the inequalities is strict, the other is also strict). We start with the pure version ($p_P = 1$) and we first prove a more general result for order statistics. Let $\{X_i\}_{i = 1, \ldots, n}$ be a set of random variables and denote their expected values, variances and cdfs by μ_i, σ_i^2 and $F_i(x)$, respectively. Without loss of generality, assume $\mu_1 \leq \mu_2 \leq \ldots \mu_i \leq \mu_{(i + 1)} \leq \ldots \mu_n$. Define the *complementary function* $H_i(x) = 1 - F_i(x)$ for all i and any admissible x. Let $\{X_{[i]}\}_{i = 1, \ldots, n}$ be the corresponding set of order statistics defined for samples with n realizations from the set $\{X_i\}_{i = 1, \ldots, n}$ (one from each X_i). Specifically, $X_{[1]} = \min\{X_i\}$ and $X_{[n]} = \max\{X_i\}$, so we may use the notations *min* and *max* for $X_{[1]}$ and $X_{[n]}$. Our task is to prove the following theorem:

Theorem RNA.1: $\Sigma_{i = 1, \ldots, n}\sigma_{[i]}^2 \leq \Sigma_{i = 1, \ldots, n}\sigma_i^2$.

We stress that we are not requiring stochastic independence and—unlike the case in basic order statistics analysis—we are not assuming that all distributions are identical. Conceptually, we may think of a sample with r independent repetitions of n realizations,

one from each member of the set $\{X_i\}_{i = 1, \ldots, n}$, then as $r \to \infty$ this *conceptual sample* represents the set precisely in the sense that the empirical distribution of the ith column converges to $F_i(x)$. If we add n new columns in which we place the sorted values of each repetition, the empirical distribution of the $(n + i)$th column converges to $F_{[i]}(x)$. The Theorem states that the sum of the variances associated with each of the added columns cannot exceed the respective sum in the unsorted sample. That is, by sorting the sample from smallest to largest and then considering the distributions of the sorted values we reduce the total variance in the system. It is easy to verify that equality is guaranteed if all $F_i(x)$ (and thus also all $H_i(x)$) are non-overlapping (in such a case the sorted columns of the conceptual sample are identical to the original columns). But otherwise the inequality is strict.*

Corollary: $\sigma_{min}^2, \sigma_{max}^2 \leq \sigma_{min}^2 + \sigma_{max}^2 \leq \Sigma_{i = 1, \ldots, n}\sigma_i^2$; for any $n \geq 2$.

We first prove the special case $n = 2$, to which we refer as Theorem RNA.1*a* (we will use the modifier "*a*" to denote this special case in general). This case is actually sufficient to show that the pure Parkinson distribution yields a lower variance at the expense of a higher mean. The following lemma is true by definition, because we are dealing with order statistics:

Lemma RNA.1: $F_{min}(x) \geq \max\{F_i(x)\}_{i = 1, \ldots, n} \geq \min\{F_i(x)\}_{i = 1, \ldots, n} \geq F_{max}(x)$ for any admissible argument x.

That is, X_{min} is stochastically smallest and X_{max} is stochastically largest. For the $n = 2$ case, Lemma RNA.1 suggests that both $F_1(x)$ and $F_2(x)$ are restricted to the envelope between $F_{max}(x)$ and $F_{min}(x)$.

Lemma RNA.2*a*: For $n = 2$, $\mu_{min}^2 + \mu_{max}^2 \geq \mu_1^2 + \mu_2^2$.

Proof.

»» Define $\delta = \mu_1 - \mu_{min}$ and notice that $\delta \geq 0$ by Lemma RNA.1. Because $\mu_1 + \mu_2 = \mu_{min} + \mu_{max}$ we also obtain $\mu_{max} = \mu_2 + \delta$. Substituting $\mu_1 - \delta$ for μ_{min} and $\mu_2 + \delta$ for μ_{max}, $\mu_{min}^2 + \mu_{max}^2 = \mu_1^2 + \mu_2^2 + 2\delta(\mu_2 - \mu_1 + \delta)$. But $\mu_2 \geq \mu_1$ (by assumption) so $2\delta(\mu_2 - \mu_1 + \delta) \geq 0$. (If we were to reverse the assumption that $\mu_2 \geq \mu_1$ then Lemma RNA.1 would imply that $\delta \geq \mu_1 - \mu_2$, and Lemma RNA.2*a* would still hold.) ««

* We are grateful to Victor Ohanyan for pointing out that independence is not required. Victor has also produced a more rigorous proof of the theorem, which does not rely on the conceptual sample.

The following lemma states that the total second moment of the set of distributions does not change by sorting. It is a basic result in the field of order statistics for any moment, and indeed we have already used the same result for the first moment in our proof of Lemma RNA.2*a* (by stating that $\mu_1 + \mu_2 = \mu_{min} + \mu_{max}$). It is easy to see it in terms of the conceptual sample where sorting does not change the actual realizations that we square.

Lemma RNA.3: $\Sigma_{i = 1, \ldots, n} E(X_i^2) = \Sigma_{i = 1, \ldots, n} E(X_{[i]}^2)$.

For $n = 2$,

Lemma RNA.3*a*: Let X_1 and X_2 be independent random variables with finite variances, then $E(X_{min}^2) + E(X_{max}^2) = E(X_1^2) + E(X_2^2)$.

Theorem RNA.1*a*: Let X_1 and X_2 be independent random variables with finite variances, then $\sigma_{min}^2 + \sigma_{max}^2 \leq \sigma_1^2 + \sigma_2^2$.

Proof.

»» By Lemmas 2*a* and 3*a*, $\sigma_{min}^2 + \sigma_{max}^2 = E(X_{min}^2) - \mu_{min}^2 + E(X_{max}^2) - \mu_{max}^2 = E(X_1^2) - \mu_1^2 + E(X_2^2) - \mu_2^2 - 2\delta(\mu_2 - \mu_1 + \delta) = \sigma_1^2 + \sigma_2^2 - 2\delta(\mu_2 - \mu_1 + \delta) \leq \sigma_1^2 + \sigma_2^2$. ««

To extend the proof of Theorem RNA.1*a* to Theorem RNA.1 (for any *n*), using Lemma RNA.3, it is sufficient to show that $\Sigma_{i = 1, \ldots, n} \mu_i^2 \leq \Sigma_{i = 1, \ldots, n} \mu_{[i]}^2$ (which effectively generalizes Lemma RNA.2*a*). To that end we invoke the conceptual sample and we use a particular sorting process that starts with the original sample (as in columns 1 through *n*), ends with the fully sorted sample (as in columns $n + 1$ through $2n$), and such that $\Sigma_{i = 1, \ldots, n} \mu_i^2$ as defined for the columns during the process is monotone increasing. To that end, we define the *adjacent column partial interchange* (ACPI) in the context of a conceptual sample: for columns *i* and $(i + 1)$, compare the realizations one by one (from repetition 1 to repetition *r*), exchange any pair for which the entry in column *i* strictly exceeds that of column $(i + 1)$, and update μ_i and $\mu_{(i + 1)}$. As a result μ_i decreases by some positive value δ and $\mu_{(i + 1)}$ increases by the same value. Recall that the initial sample is sorted so that $\mu_i \leq \mu_{(i + 1)}$ and after the ACPI the same relationship is strengthened. Hence, by the arguments we used to prove Lemma RNA.2*a*, $\mu_i^2 + \mu_{(i + 1)}^2$ must also increase. To complete the proof observe that by performing the ACPI $o(n^2)$ times we can obtain the fully sorted sample.

To prove that the pure Parkinson distribution reduces the variance, it is sufficient to use the version of the corollary for $n = 2$. Of the two elements in the pure Parkinson

distribution, q has a variance of zero so the variance of the maximum—the pure Parkinson random variable itself—is bounded from above by the variance of the core element minus the variance of the minimum. Showing that the mean must increase is trivial (unless $\Pr\{Y \le q\} = 0$, in which case neither the mean nor the variance change).

Finally, to extend the proof to the more general case ($0 \le p_P \le 1$), we slightly adapt the ACPI procedure as applied to the two columns representing Y and q. Instead of exchanging for every instance where the entries are in the wrong order, we now take a side lottery with a probability of p_P to decide whether to perform the exchange.

The Lognormal Central Limit Theorem for Positive Random Variables[*]

In practice, the central limit theorem is often invoked for the convolution of few independent random variables by a normal approximation. The fact that the normal approximation is appropriate for the sum of a very large number of independent random variables is interpreted as permission to use it for a small number of random variables. This may be justified by necessity, but in scheduling applications—where all processing times are nonnegative—we can do better by using the lognormal approximation instead of the normal. Among popular distributions, the lognormal distribution is unique because it satisfies elementary conditions that convolution of a small number of continuous nonnegative random variables must satisfy. Yet it converges to the normal when the number of random variables grows large, in which case we know that the normal approximation is appropriate. Therefore, one can use it as the basis of an alternative central limit theorem for nonnegative random variables.

Let X denote the convolution of $n \ge 2$ independent, nonnegative and continuous random variables with positive means and finite coefficients of variation. We may refer to these random variables as *components*. If we denote the mean of component j by μ_j and its variance by σ_j^2, then it is well known that the mean and variance of X, μ_X and σ_X^2, are given by $\mu_X = \Sigma\mu_j$ and $\sigma_X^2 = \Sigma\sigma_j^2$ (due to statistical independence). We do not require the components to be identically distributed, but they must satisfy the regularity conditions of the central limit theorem (CLT); i.e., when $n \to \infty$, no single component should dominate the convolution. Equivalently, we require that as $n \to \infty$, $\mu_j/\mu_X \to 0$ and $\sigma_j^2/\sigma_X^2 \to 0$ for all j. Denote the density function of X by $f_X(x)$ and that of component j by $f_j(x)$. Whereas $f_j(0) > 0$ is allowed (e.g., if the component is distributed exponentially), we can show by a limiting argument that, because $n \ge 2$, $f_X(0) = 0$. For small n, if the components have very high coefficients of variation, the coefficient of variation of X may also be high (although it must tend to zero when n grows large). Finally, when $n \to \infty$, the CLT should apply. We list these observations as three conditions:

(*i*) $f_X(x) = 0$ for $x \le 0$,

(*ii*) σ_X/μ_X is unbounded,

(*iii*) as $n \to \infty$, $f_X(x) \to \dfrac{1}{\sigma_X\sqrt{2\pi}} \exp\left(-\dfrac{1}{2}\left(\dfrac{x-\mu_X}{\sigma_X}\right)^2\right)$.

[*] This section is an almost verbatim reproduction of Mazmanyan, Ohanyan & Trietsch (2008), which remains unpublished. The main motivation for their work was to support our text.

The vast majority of conventional distributions do not satisfy all three conditions. The normal distribution is disqualified by (*i*), because the normal random variable may be negative. When the probability of a negative realization becomes sufficiently negligible, the normal does not comply with (*ii*) either; for instance, if we require the probability of a negative realization to be below 0.00135, $cv \leq 1/3$ is necessary. Once we limit ourselves to nonnegative random variables, distributions that satisfy condition (*i*) often violate condition (*ii*). Conversely, most distributions that satisfy condition (*ii*), such as Weibull or gamma, violate condition (*i*) because they rely on $f_X(0) > 0$ in cases with high σ_X/μ_X. The most notable exception is the lognormal. We show that it satisfies all three conditions.

Condition (*i*) is satisfied by the lognormal distribution because as $x \to 0^+$, $\ln x \to -\infty$. Condition (*ii*) is satisfied because in the lognormal case *cv* is unconstrained. To show that condition (*iii*) is satisfied, all the following claims are subject to the stipulation that $n \to \infty$. By the law of large numbers, $cv = \sigma_X/\mu_X \to 0$ (because $\mu_j > 0$ and σ_j/μ_j is finite for all *j*). But $s^2 = \ln(1 + cv^2)$ so as $cv \to 0$, $s^2 \to cv^2$ and, equivalently, $s \to cv$. Also, for any *x* in the support of the distribution, $x/\mu_X \to 1$ *almost surely*. Therefore, $xs \to \sigma_X$ and thus $xs\sqrt{2\pi} \to \sigma_X\sqrt{2\pi}$. It remains to show that $(\ln x - m)/s \to (x - \mu_X)/\sigma_X$. We can write $x = \mu_X(x/\mu_X)$, so $\ln x = \ln \mu_X + \ln(x/\mu_X)$. But $x/\mu_X \to 1$ so $\ln(x/\mu_X) \to (x - \mu_X)/\mu_X$. Recall that $m = \ln \mu_X - s^2/2$ and $s^2/2 \to 0$, so $m \to \ln \mu_X$. Substituting these values for $\ln x$ and *m* we obtain $(\ln x - m)/s \to (x - \mu_X)/s\mu_X$. Finally, $s\mu_X \to \sigma_X$, thus completing the proof.

Table RNA.1: The relative MAD of *k*-Erlang and Chi-Square approximations with *k* d.f.

Erlang Case				Chi-Square Case			
k=	Normal	Lognormal	Ratio	d.f.	Normal	Lognormal	Ratio
1	0.314351	0.123926	0.394229	1	0.599782	0.202787	0.338102
2	0.159869	0.070044	0.438132	2	0.314351	0.123926	0.394229
3	0.107003	0.048841	0.456447	3	0.212118	0.089476	0.421823
4	0.08038-	0.037492	0.466434	4	0.159869	0.070044	0.438132
5	0.064357	0.030422	0.472708	5	0.128215	0.057551	0.448865
6	0.053658	0.025595	0.477011	6	0.107003	0.048841	0.456447
7	0.046008	0.022090	0.480143	7	0.091803	0.042421	0.462083
8	0.040266	0.019429	0.482526	8	0.080380	0.037492	0.466434
9	0.035798	0.017341	0.484399	9	0.071483	0.033589	0.469893
10	0.032223	0.015657	0.485909	10	0.064357	0.030422	0.472708
12	0.026857	0.013112	0.488196	12	0.053658	0.025595	0.477011
14	0.023024	0.011278	0.489844	14	0.046008	0.022090	0.480143
16	0.020147	0.009894	0.491089	16	0.040266	0.019429	0.482526
18	0.017910	0.008813	0.492062	18	0.035798	0.017341	0.484399
20	0.016120	0.007945	0.492843	20	0.032223	0.015657	0.485909
25	0.012897	0.006375	0.494257	25	0.025784	0.012599	0.488656
30	0.010748	0.005323	0.495205	30	0.021490	0.010541	0.490507
40	0.008062	0.004002	0.496395	40	0.016120	0.007945	0.492843
50	0.006450	0.003206	0.497112	50	0.012897	0.006375	0.494257
75	0.004300	0.002142	0.498073	75	0.008599	0.004267	0.496156
100	0.003225	0.001608	0.498555	100	0.006450	0.003206	0.497112
150	0.002150	0.001073	0.499040	150	0.004300	0.002142	0.498073
200	0.001613	0.000805	0.499284	200	0.003225	0.001608	0.498555

To illustrate the efficacy of using the lognormal approximation, Mazmanyan, Ohanyan and Trietsch test it for the *k*-Erlang distribution with various *k*. They report the mean absolute deviation (MAD) of the lognormal approximation, as a fraction of the mean, and compare it with that of the normal distribution. The results are given in the first four columns of Table RNA.1. By the table, it is evident that the relative MAD of the lognormal approximation is at most 50% of the normal's in this case (i.e., the approximation is at least two times better). Furthermore, it is even more advantageous where it counts most, for small *k*. Very similar results apply for the chi-square distribution, as shown in the subsequent columns of the table. Due to the larger variance of the chi-square distribution, convergence seems to be exactly twice as fast for the Erlang case. Thus, the lognormal approximation avoids violating conditions (*i*) and (*ii*) and outperforms the normal sizably for low *k* values. It also outperforms the normal approximation for high *k* values for which the normal is highly unlikely to violate condition (*i*) in practice. On the one hand, when using the lognormal approximation for symmetric components, if the normal approximation is sufficiently unlikely to yield a negative result then the advantage goes to the normal (because a symmetric result is advantageous in that case). On the other hand, however, typical nonnegative random variables are usually skewed to the right in practice.

Estimating the Parkinson Distribution Parameters and Using it in Simulation

The results we present here are based on Trietsch et al. (2010). First, we need a way to tell whether the Parkinson distribution provides a reasonable approximation for typical processing times in our past projects. Then, if the fit appears adequate, we need a way to use the information available from historical (complete) projects to generate plausible Parkinson distributions for new projects. We assume that *Y*—the internal random variable—is lognormal, and we have a history of several activity times that are linearly associated with a common factor that is also a random variable. There are two approaches to model the common factor. One, which is convenient when there is sufficient data, is to fit a lognormal distribution to the common factor data (which can then be used in simulation). An alternative is to use the empirical distribution provided by history (without fitting any theoretical distribution to it). Below we focus on the alternative. Recall that if processing times are actually independent we obtain a special case of linear association with a constant common factor. Hence our assumption is essentially a decision to test and account for correlation by basic linear association.

We assume that a history of *K* projects is available and, as our first step, we wish to validate that these projects fit the Parkinson distribution with a common factor model reasonably well. The second step is to estimate distributions for new projects, which can then be used for simulation. For validation, for each project we treat the other $K - 1$ projects as "history." As a result, we obtain for each project a simulated distribution, and we can associate its actual completion time with a probability (cdf value). The set of *K* probabilities thus obtained can be tested statistically and thus provide validation for the model. An early version of Trietsch et al. (2010) reported that this approach was successful for a set of nine projects that exhibited the pure Parkinson distribution (with $p_P = 1$). In the current version, the authors analyze a smaller data set, with five projects, using the general Parkinson distribution. They report that this application passes the validation test as well. (In the first version, written before the generalized definition was

proposed, one of the projects could not be analyzed by data from the others because it had an excessive p_P, whereas the other projects in the family could be analyzed as if p_P were 0.) These data were given in very crude units: all estimates were in months (which the authors interpreted as four weeks) and with one exception, all realizations were given in weeks (the exception took three days, which was interpreted as half a week). Hence, further research and validation—preferably by others—is necessary. With this caveat, the analysis indicates that parameters such as p_P and p_E are highly correlated. Similarly, in the nine-project case, the authors observe a strong correlation between the common factor and the coefficient of variation of Y. Even more complex correlations cannot be ruled out. (Such correlations must be taken into account explicitly in subsequent simulations.) For the pure Parkinson case we can estimate the coefficient of variation of log$Y$, and the logarithm of the common factor by using regression for activities with $Y > q$. We sort ratios in increasing order; that is, the activity with the j-smallest ln(p/e) of project k is denoted by the index jk. Omitting the error term, if this activity is tardy we obtain for it the equation

$$\hat{\ln}(b_k) + z_{jk}\hat{s}_k = \ln\frac{p_{jk}}{e_{jk}} \qquad (1)$$

where—using Blom's scores—$z_{jk} = \Phi^{-1}[(j - 0.375)/(n_k + 0.25)]$ (the z-value for which the standard normal distribution cdf yields a probability of $(j - 0.375)/(n_k + 0.25)$), b_k is the common factor, and s_k is the slope. ln(b_k) and s_k are the parameters of the lognormal distribution of p_{jk}/e_{jk} that we usually denote by m and s. That is, we treat the transformed values ln(p_{jk}/e_{jk}) as our dependent variables, and we obtain estimators for their mean and standard deviation, ln(b_k) and s_k. In this analysis we assume that all non-tardy activities are "on time" and thus they are useless for the purpose of estimating the parameters of Y. In the general Parkinson case we can also use similar equations for strictly early activities. That requires an adaptation, however, because the effective sample size from which the strictly early activities are drawn is smaller than n_k: for strictly early activities, instead of using $z_{jk} = \Phi^{-1}[(j - 0.375)/(n_k + 0.25)]$, we have $z_{jk} = \Phi^{-1}[(j - 0.375)/((1 - p_P)(n_k + 0.25))]$. If the regression line thus obtained fits all the strictly tardy and strictly early activities reasonably well (with a high R^2 value and with residuals that appear random and normal), we can accept the model and proceed to step 2, using it in simulation. Trietsch et al. (2010) report a reasonable fit.

To simulate the Parkinson with linear association for a new project, we can now rely on the distribution obtained by all K historical projects. A simple and effective way to do that for a small K is adopted in that paper. Working in the logarithmic transform space, if the mean bias, slope and p_P parameters of project i are denoted by m_i, s_i, and p_{Pi}, and we wish to simulate project k by a stored sample with r rows, then we use the parameters m_i, s_i, and p_{Pi} of the ith project in $[n_i / \sum_{i \neq k} n_i]r$ of these rows. By using the parameters of each project $i \neq k$ in the history together (as a block), we automatically account for any empirical correlation in the data. Essentially, what we are doing here is using an empirical copula to model any correlation that might exist among the

components of our multivariate distribution.* A modification that requires further research is to discount older projects so that the effective copula we are using favors recent projects. This could be desirable because it is possible that users will learn to reduce their average bias as a result of measuring it routinely, which our system does. If so, old projects' common factors should be given lesser weight.

On Stochastic Orders and the Lognormal Distribution

The subject of stochastic orders is very rich (e.g., see Shaked & Shanthikumar 1994), and in the text we limited ourselves to the two most basic and useful examples, by expectation and in the regular stochastic sense ($\leq_{ex}$ and $\leq_{st}$). We noted that $\leq_{st}$ implies $\leq_{ex}$. In our Research Notes for Chapter 6 we introduced the strongest stochastic dominance form, $\leq_{as}$, where the dominance occurs with probability 1 (that is, *almost surely*). We noted that $\leq_{as}$ implies $\leq_{st}$. Yet another form of stochastic dominance is by likelihood ratio, denoted $\leq_{lr}$, and in terms of strength it lies between $\leq_{st}$ and $\leq_{as}$; that is, $\leq_{as}$ implies $\leq_{lr}$, which in turn implies $\leq_{st}$. We say that $X \leq_{lr} Y$ if $f_X(t)/f_Y(t)$ is monotone nonincreasing in t, where $f_X(t)$ and $f_Y(t)$ are the density functions of X and Y when X and Y are continuous, or mass functions when they are discrete (in which case $t = 0, 1, ...$). In this context, we interpret division of a positive number by zero as $+\infty$. For example, it is easy to show that if X and Y are exponential such that $\mu_X \leq \mu_Y$, then $X \leq_{lr} Y$. We now show a similar result when X and Y are lognormal with the same *cv* (i.e., with the same s). Without loss of generality, assume $\mu_X < \mu_Y$ and $t > 0$. We obtain

$$\frac{f_X(t)}{f_X(t)} = \frac{\frac{1}{t\sqrt{2\pi s^2}}\exp\left[\frac{-(\ln t - m_X)^2}{2s^2}\right]}{\frac{1}{t\sqrt{2\pi s^2}}\exp\left[\frac{-(\ln t - m_Y)^2}{2s^2}\right]} = \frac{\exp\left[\frac{-(\ln t - \ln \mu_X + s^2/2)^2}{2s^2}\right]}{\exp\left[\frac{-(\ln t - \ln \mu_Y + s^2/2)^2}{2s^2}\right]} =$$

$$\exp\left[\frac{(\ln t - \ln \mu_Y + s^2/2)^2 - (\ln t - \ln \mu_X + s^2/2)^2}{2s^2}\right] =$$

$$\exp\left[\frac{(2\ln t - \ln \mu_X - \ln \mu_Y + s^2)(\ln \mu_X - \ln \mu_Y)}{2s^2}\right] =$$

$$\exp\left[\frac{\ln(\mu_X/\mu_Y)}{s^2}\ln t + g(s, \mu_X, \mu_Y)\right] = \exp[g(s, \mu_X, \mu_Y)]t^{\ln(\mu_X/\mu_Y)/s^2}$$

* As summarized by the Wikipedia, a copula is a function linking marginal variables into a single multivariate distribution. If we take a sample from a copula that represents a multivariate distribution with correlations among its components, we obtain a collection of points in the relevant space that form cloud shapes reflecting the correlation structure. To visualize this, consider a sample of heights and weights of men. One might expect some correlation between the two components, so the copula would have higher density in the neighborhood of a diagonal line with a positive slope (similar to a regression line).

where $g(s, \mu_X, \mu_Y)$, which does not depend on t, is given by

$$g(s,\mu_X,\mu_Y) = [s^2 - \ln(\mu_X \mu_Y)]\frac{\ln(\mu_X / \mu_Y)}{2s^2}$$

In the derivation, we utilize the identity $a^2 - b^2 = (a + b)(a - b)$ and the fact that sums (differences) of logarithms are logarithms of products (ratios). Because $\mu_X < \mu_Y$ and $s^2 > 0$, $\ln(\mu_X / \mu_Y)/s^2 < 0$. In addition, $\exp(g)$ is a positive constant, so we obtain a monotone nonincreasing function of t, as required.

References

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

Hasija, S., E. Pinker and R.A. Shumsky (2010) "Work Expands to Fill the Time Available: Capacity Estimation and Staffing Under Parkinson's Law," *Manufacturing and Service Operations Management* 12, 1–18.

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

Shaked, M. and J. G. Shanthikumar (1994) *Stochastic Orders and their Applications*, Academic Press, New York.

Trietsch, D., L. Mazmanyan, L. Gevorgyan and K.R. Baker (2010) "A New Stochastic Engine for PERT," Working Paper.

Trietsch, D. (2005) "The Effect of Systemic Errors on Optimal Project Buffers," *International Journal of Project Management* 23, 267-274.