Aalto University
School of Science
Degree programme in Engineering Physics and Mathematics

Santtu Saijets

Optimal strategies for selecting project portfolios using uncertain cost estimates

Mat-2.4108 Independent Research Projects in Applied Mathematics.

Espoo, February 13, 2014

Supervisor: Prof. Ahti Salo

Instructors: Dr. Eeva Vilkkumaa, Dr. Juuso Liesiö

Contents

1 Introduction

Many organizations seek to achieve their goals by allocating resources to projects (Salo et al. 2011). For instance, municipalities select projects in order to maintain a certain standard of services. They need to ensure the number of e.g. day care centres in a specific region while minimizing the total costs spend on the build and uphold of the facilities. Typically, costs are minimized based on ex ante estimates about projects' expected ex post costs. However, uncertain cost estimates predicted from the historical data or gathered from the experts usually differs a lot from the realized projects' costs. This is why it is hard to choose the true optimal portfolio that minimizes the realized ex post costs. Furthermore, systematic project selection process selects most likely projects whose costs has been underestimated and the decision-maker is likely to experience post-decision disappointment when true costs are realized (Brown 1974, Harrison and March 1984, Smith and Winkler 2006). For example, Flyvbjerg et al. (2002) compared the estimated and realized costs of public transportation infrastructure projects and concluded that the realized costs of the selected projects were on average 28% higher than estimated.

Bayesian modeling of estimation uncertainties can be shown to decrease the realized portfolio cost and to mitigate the post-decision disappointment experienced by the decision-maker (for an overview, see e.g. Gelman et al. 2004). Bayesian modeling of uncertainties also makes it possible to study the value of obtaining additional estimates about the projects' costs by facilitating the computation of the expected decrease in the portfolio cost if such estimates were acquired (La Valle 1968, Marchak and Radner 1972, Gould 1974, Laffont 1980, Delquié 2008). Value of information is studied mainly through simulations (Harrison and March 1984, Keisler 2004). Vilkkumaa et al. (2014) presented analytic results of the value of information in the case where goal is to maximize the selected portfolio value and both projects' values and estimates were normally distributed.

In this work we developed a Bayesian model to support portfolio cost minimization in the presence of uncertain cost estimates. Furthermore, we derive analytic results of the value of obtaining additional estimates to support targeting of re-evaluations of projects when projects' costs and cost estimates are log-normally distributed. In particular, we show that additional information should be acquired about projects whose (i) initial cost estimate is near the selection threshold and (ii) posterior variance is relatively large.

Rest of the study is structured as follows. In section 2 we develop the framework for the project selection under uncertainty. Simulated and analytic results for the value of information and possible applications are presented in section 3. Section 4 concludes.

2 Project portfolio selection under uncertainty

Consider $1, \dots, m$ project candidates out of which the decision-maker (DM) wants to select a subset, i.e. a portfolio. The selected portfolio is represented by binary vector $z = [z_1, \dots, z_m]$ such that $z_i = 1$ if and only if project i is selected. The objective is to minimize the cost of the portfolio subject to some relevant constraints. For instance, the value of the portfolio may need to meet or exceed some predetermined threshold. Other possible constraints may arise from mutually exclusive projects (e.g. project A can only be selected if project B is not selected, and vice versa) or logical interdependencies (e.g. project A can only be selected if project B is selected). Constraints such as these define the set Z of feasible portfolios.

The projects' true costs are $c = [c_1, \dots, c_m]'$. These costs are independent realizations of random variables $C_i \sim f(c_i)$. We assume that distributions $f(c_1), \dots, f(c_m)$ are known. The DM tries to select a portfolio that minimizes the expected portfolio cost and fulfils portfolio requirements mentioned above. If the DM could observe the true costs at the time of the selection decision, the optimal portfolio $z(c)$ would be determined by solving the optimization problem

$$z(c) = \arg\min_{z \in Z} \; zc.$$

The DM cannot, however, observe the true costs, but only the estimates thereof. These estimates $c^E = [c_1^E, \dots, c_m^E]'$ are realizations of conditionally independent random variables $(C_i^E | C_i = c_i) \sim f(c_i^E | c_i)$ where $f(c_i^E | c_i)$ is known for all i and c_i. These estimates are assumed to be unbiased, i.e. $\mathbb{E}[C_i^E \mid C_i = c_i] = c_i$. The optimal portfolio based on estimates c^E is obtained by solving the optimization problem

$$z(c^E) = \arg\min_{z \in Z} \; zc^E.$$

Making the selection decision based on uncertain estimates makes it likely that the selected portfolio will be suboptimal ex post. Moreover, despite the fact that the projects' cost estimates are unbiased a priori, the true cost of the selected portfolio is expected to be higher than estimated, causing the DM to experience post-decision disappointment. This is because those projects whose cost have been underestimated are more likely to be selected.

Figure 1 shows one realization of a portfolio selection problem under uncertainty. Here projects' costs c_i are realizations of a log-normally distributed random variable C_i with a mean of 1 million and a variance of 0.01 million. Cost estimates c_i^E are realizations of random variable $(C_i^E | C_i = c_i) = E_i c_i$, where $E_i \sim \mathrm{LogN}(-\frac{0.1^2}{2}, 0.1^2)$. First parameter chosen such that estimates are unbiased. It can be readily checked that $\mathbb{E}[C_i^E \mid C_i = c_i] = \mathbb{E}[E_i]c_i = c_i$. Projects values v_i are known to the DM and they are realizations of random variable $V_i = c_i + \mathrm{N}(\frac{c_i}{5}, \frac{c_i}{15})$. Marker size is proportional to the project's value. Costs are minimized with portfolio value restricted to be over 4.25 million. True optimal portfolio $z(c)$ is marked with black markers and $z(c^E)$ is marked with circled markers. We can see that there is only one true optimal project in $z(c^E)$. Because costs of projects A and J are underestimated, they are included in $z(c^E)$ and true optimal projects D and I are omitted. Estimated, realized, and optimal portfolio costs, post-decision disappointment, and portfolio value of projects in figure 1 are presented in table 1. We can see that the realized portfolio cost is 3.8% greater than the optimal portfolio cost.

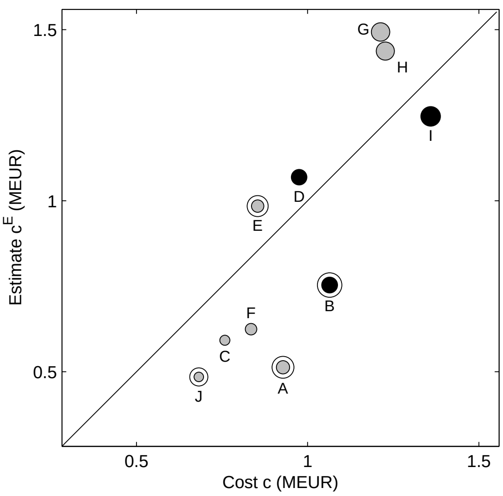


Figure 1: True costs and cost estimates of 10 projects whose costs are log-normally distributed with a mean of 1 million and variance of 0.01 million. Costs are minimized with portfolio value restricted to be over 4.25 million. Marker size is proportional to the project's value. The true optimal portfolio $z(c)$ is marked with black markers. Optimal portfolio $z(c^E)$ selected based on the estimates is marked with circled markers.

Estimated portfolio cost	2.74
Realized portfolio cost	3.53
Post-decision disappointment	0.79
Optimal portfolio cost	3.40
Portfolio value	4.25

Table 1: Estimated, realized, and optimal portfolio costs, post-decision disappointment, and portfolio value of projects in figure 1.

Modeling projects' costs with log-normal distribution is reasonable because (i) the domain of the log-normal distribution is non-negative and (ii) relatively large portion of probability mass is centred around small and moderate values so that probability of considerably large values is small (Keisler 2004). The log-normality also allows the cost estimates to be situated more likely on the left side of the true projects' costs.

3 Bayesian modeling of cost uncertainties in portfolio selection

3.1 Bayesian uncertainty model

Instead of making decisions based on the uncertain estimates c_i^E alone, one can use Bayesian analysis to obtain the distribution for the projects' true costs given the estimates. In particular, the posterior distribution $f(c_i|c_i^E)$ for the projects' true costs given the estimates $(C_i|C_i^E = c_i^E)$ can be obtained from the prior and likelihood distributions $f(c_i)$ and $f(c_i^E|c_i)$ through Bayes' rule $f(c_i|c_i^E) \propto f(c_i)f(c_i^E|c)$.

In general, there is no closed form expression for the posterior distribution. Consider, however, the case in which $f(c_i) = \mathrm{LogN}(\mu_i, \sigma_i^2)$ and $f(c_i^E|c_i) = c_i\mathrm{LogN}(-\frac{\tau_i^2}{2}, \tau_i^2)$ so that indeed the assumption of unbiased estimates $\mathbb{E}[C_i^E \mid C_i = c_i] = c_i$ holds. Proposition 1 below states that in this case a closed-form expression can be obtained. All proofs are in Appendix A.

Proposition 1 *Assume $C_i \sim LogN(\mu_i, \sigma_i^2)$ and $C_i^E = c_iE_i$, where $E_i \sim LogN(-\frac{\tau_i^2}{2}, \tau_i^2)$. Then,*

$$\left(C_i \mid C_i^E = c_i^E\right) \sim \mathrm{LogN}\left(\frac{\sigma_i^2}{\sigma_i^2 + \tau_i^2}(\ln(c_i^E) - \frac{\tau_i^2}{2}) + \frac{\tau_i^2}{\sigma_i^2 + \tau_i^2}\mu_i, \frac{\sigma_i^2\tau_i^2}{\sigma_i^2 + \tau_i^2}\right).$$

Posterior distribution $f(c_i|c_i^E)$ can be used, for instance, to compute the projects' expected true costs given the cost estimates,

$$c_i^B = \mathbb{E}[C_i|C_i^E = c^E] = \int_0^\infty c_i f(c_i|c_i^E)dc_i. \quad (1)$$

Given the assumptions of Proposition 1, for instance, we have

$$c_i^B = (c_i^E)^{\frac{\sigma_i^2}{\sigma_i^2+\tau_i^2}} \cdot \exp\left(\frac{\tau_i^2(\sigma_i^2+\mu_i)}{\sigma_i^2+\tau_i^2}\right). \quad (2)$$

The expected costs $c^B = [c_1^B, \ldots, c_m^B]'$ can then be used as a basis for portfolio selection:

$$z(c^B) = \arg\min_{z\in Z}\ zc^B. \quad (3)$$

To study the average performance of (3), we a define random variable

$$C_i^B = \mathbb{E}[C_i|C_i^E] = \int_0^\infty c_i f(c_i|C_i^E)dc_i. \quad (4)$$

which can be obtained from (1) by replacing the observed estimate c_i^E with the random variable C_i^E.

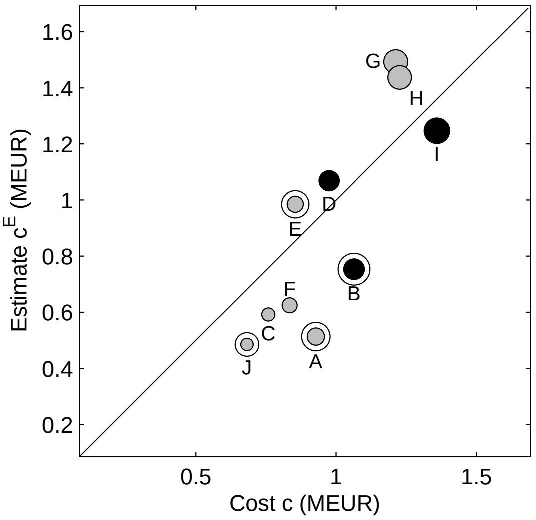
By definition, $z(c^B)$ minimizes the expected portfolio cost. Moreover, Proposition 2 states that $z(c^B)$ eliminates the expected positive gap between the true and estimated portfolio cost, i.e., post-decision disappointment. The proof is analogous to that of Proposition 3 in Vilkkumaa et al. (2014).

Proposition 2 *Let C^E be a conditionally unbiased estimator of C and $z(c^B)$ satisfy* (3)*. Then,*

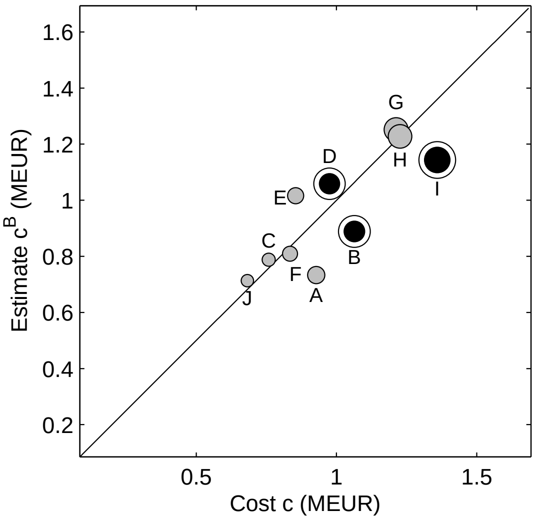
$$\mathbb{E}[z(c^B)c^B - z(c^B)C|C^E = c^E] = 0,$$

for all c^E, and hence $\mathbb{E}[z(C^B)C^B - z(C^B)C] = 0$, where C^B is given by (4)*.*

Figure 2 shows how the selection presented in figure 1 changes when it is done based on estimates c^B obtained from (2). We can see that all of the estimates move closer to the prior mean. Furthermore, portfolios $z(c)$ and $z(c^B)$ are the same. Estimated, realized, and optimal portfolio costs, post-decision disappointments, and portfolio values of projects in figures 2a and 2b are presented in table 2. We can see that post-decision disappointment decreases by 0.48 million, even though, estimated portfolio cost increases only by 0.35 million.



(a) Costs vs. estimates



(b) Costs vs. Bayes estimates

Figure 2: True costs and cost estimates for projects in figure 1 and their Bayes estimates c^B. Costs are minimized with portfolio value restricted to be over 4.25 million. Marker size is proportional to the project's value. The true optimal portfolio is marked with black markers. Optimal portfolio selected based on the estimates is marked with circled markers.

	Estimates	Bayes estimates
Estimated portfolio cost	2.74	3.09
Realized portfolio cost	3.53	3.40
Post-decision disappointment	0.79	0.31
Optimal portfolio cost	3.40	3.40
Portfolio value	4.25	4.26

Table 2: Estimated, true, and optimal portfolio costs, post-decision disappointments, and portfolio values of projects in figures 2a and 2b.

3.2 Value of additional information

Bayesian analysis can also be used to study the expected value of obtaining additional cost estimates for the projects prior to actually acquiring these estimates. In keeping with the standard definition, we define the expected value of information C^E (EVI) as the expected decrease in portfolio cost when selecting the portfolio by taking this information into account (La Valle 1968, Marchak and Radner 1972, Gould 1974, Laffont 1980, Delquié 2008).

Definition 1 *The expected value of information* C^E *is*

$$\mathrm{EVI}[C^E] = \min_{z \in Z} \; z\mathbb{E}[C] - \mathbb{E}[\min_{z \in Z} \; z\mathbb{E}[C|C^E]].$$

This definition can be applied when one or more project evaluation rounds have been completed. In what follows, $\mathbb{E}[C]$ denotes the vector of expected project costs resulting from all the earlier evaluations. Prior to observing c^E, the additional information is a random variable C^E so that the expected projects' costs that have been revised based on the additional information are represented by the random variable $\mathbb{E}[C|C^E]$. C^E may contain additional estimates for a single project or multiple projects at a time.

Because $\mathbb{E}[C|C^E]$ is random, the computation of EVI requires solving a stochastic optimization problem with binary decision variables. In general, this is done through simulation by sampling values of c^E, obtaining $\mathbb{E}[C|C^E = c^E]$ and solving the optimization problem $\min_{z \in Z} \; z\mathbb{E}[C|C^E = c^E]$. Figure 3 shows the EVI for the projects in Figures 1 and 2. Project D has the greatest EVI which means that re-evaluations of D are most likely to change the optimal portfolio and decrease the portfolio cost.

To derive analytic results for EVI, we consider the case in which C_i^E is independent of C_j for all $i \neq j$, there is only one feasibility constraint on the number of projects in the portfolio, and one additional evaluation is acquired for project i only. If project i is not in the current optimal portfolio, the portfolio cost changes only if the evaluation c_i^E is low enough so that $\mathbb{E}[C_i|C_i^E]$ becomes smaller than the highest expected project cost in the current portfolio, denoted by x^+. In this case, the portfolio cost changes by $x^+ - \mathbb{E}[C_i|C_i^E]$. If project i is in the current portfolio, the portfolio cost changes only if the evaluation c_i^E is high enough so that $\mathbb{E}[C_i|C_i^E]$ becomes greater than the lowest expected project cost not in the current portfolio, denoted by x^-, in which case the portfolio cost changes by $\mathbb{E}[C_i|C_i^E] - x^-$. Prior to observing c_i^E, the EVI is computed by taking expectations over random C_i^E.

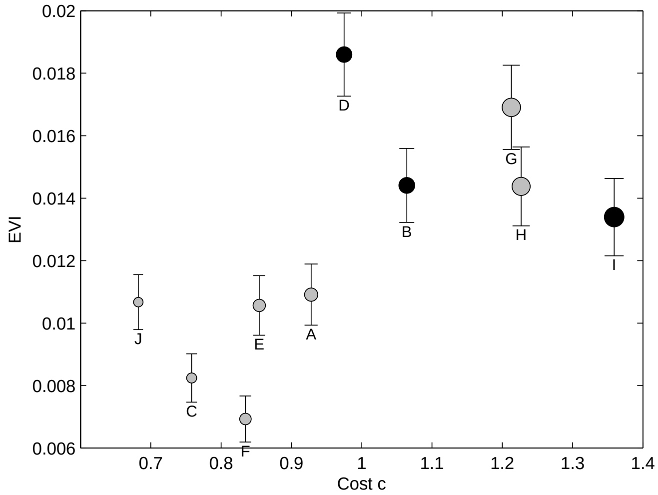


Figure 3: Simulated EVI for the projects presented in figure 1. Average values of 5000 simulation rounds with 95% confidence intervals.

Proposition 3 *Let* $Z = \{z \in \{0,1\}^m \mid \sum_i z_i = b\}$, $b \in \{1,\ldots,m-1\}$ *and let* C_i^E, C_j *be independent for all* $i \neq j$ *and* $z^* \in \text{argmin}\{z\mathbb{E}[C] \mid z \in Z\}$. *The expected value of an additional evaluation* C_i^E *of project* i *is*

$$\text{EVI}[C_i^E] = \begin{cases} \mathbb{E}\left[\max\{0, x^+ - \mathbb{E}\left[C_i|C_i^E\right]\}\right], & \text{if } z_i^* = 0 \\ \mathbb{E}\left[\max\{0, \mathbb{E}\left[C_i|C_i^E\right] - x^-\}\right], & \text{if } z_i^* = 1 \end{cases}, \tag{5}$$

where $x^+ = \max_j\{\mathbb{E}[C_j]|z_j^* = 1\}$ *and* $x^- = \min_j\{\mathbb{E}[C_j]|z_j^* = 0\}$.

Let us consider the case in which $f(c_i) = \text{LogN}(\mu_i, \sigma_i^2)$ and $f(c_i^E|c_i) = c_i\text{LogN}(-\frac{\tau_i^2}{2}, \tau_i^2)$. Proposition 4 below states that under these assumptions, a closed-form expression can be obtained for the distribution of $\mathbb{E}[C_i|C_i^E]$.

Proposition 4 *Let the assumptions of Proposition 1 hold. We can derive*

$$\mathbb{E}[C_i|C_i^E] \sim \text{LogN}(\mu_i + \frac{\sigma_i^2\tau_i^2}{2(\sigma_i^2 + \tau_i^2)}, \frac{\sigma_i^4}{\sigma_i^2 + \tau_i^2}). \tag{6}$$

Using the closed-form expression for $\mathbb{E}[C_i|C_i^E]$, we obtain a closed-form expression for EVI as well.

Proposition 5 *Let the assumptions of Proposition 1, Proposition 3, and Proposition 4 hold. The expected value of an additional evaluation of project* *i is*

$$\mathrm{EVI}[C_i^E] = f(y_i,\rho_i) = \begin{cases} x^- \left[y_i \Phi\left(\frac{\ln(y_i)}{\rho_i} + \frac{1}{2}\rho_i\right) - \Phi\left(\frac{\ln(y_i)}{\rho_i} - \frac{1}{2}\rho_i\right)\right], & \text{if } z_i^* = 1 \\ x^+ \left[\Phi\left(\frac{\ln(y_i)}{\rho_i} + \frac{1}{2}\rho_i\right) - \frac{1}{y_i}\Phi\left(\frac{\ln(y_i)}{\rho_i} - \frac{1}{2}\rho_i\right)\right], & \text{if } z_i^* = 0 \end{cases},$$

where $y_i = \min\{\frac{x^+}{x_i}, \frac{x_i}{x^-}\} \in [0,1]$, $x_i = e^{\mu_i + \frac{1}{2}\tau_i^2} = \mathbb{E}[C_i^B]$, $\rho_i = \frac{\tau_i^2}{\sqrt{\tau_i^2+\sigma_i^2}}$, *and* Φ *denotes the cumulative density function of the standard normal distribution. Function* $f(y_i,\rho_i)$ *is non-negative and increasing in* y_i *and* ρ_i*.*

Proposition 5 implies that it pays off to obtain additional estimates about those projects with (i) initial cost estimates close to the selection threshold ($\frac{x^+}{x_i}$ or $\frac{x_i}{x^-}$ close to 1) and (ii) large posterior variance $(e^{\rho_i^2} - 1)x_i^2$. Results are intuitively reasonable because when expected cost is near the selection threshold, new evaluation might push the expected project cost over the selection threshold and the optimal portfolio would change. Same applies with cost's posterior variance. With large variance, it is possible that new evaluation pushes the expected project cost over the selection threshold and the optimal portfolio changes.

Figure 4 shows how EVI helps to select which projects should be re-evaluated. In this example, 20 projects out of 100 are to be selected. Here, on the left side are the projects that belong to the current optimal portfolio and on the right side are the projects that don't belong to the current optimal portfolio. Contours of EVI are also shown in the figure. The projects' costs are realizations of independent log-normally distributed random variables with a common prior mean 1. The project population consists of two types (50 projects each). The costs of type 1 projects have more variability and can be more accurately evaluated than type 2 projects. This is reflected by parameters $\mathrm{Var}[C_i] = 0.5$ and $\mathrm{Var}[E_i] = 0.05$ for type 1 projects and by parameters $\mathrm{Var}[C_i] = 0.3$ and $\mathrm{Var}[E_i] = 0.1$ for type 2 projects. Assume that there are resources available for re-evaluating 30 projects. One possible strategy would be to choose 30 projects with the lowest expected costs for re-evaluating. Another strategy could be to use EVI for selecting projects to be re-evaluated. The selected projects for these two strategies are illustrated in Figure 4 such that those 30 projects with the lowest expected costs are denoted with black markers and those with the highest EVI with dashed ellipses. Figure shows that EVI suggests to re-evaluate both optimal and non-optimal projects that are near the selection threshold. In the case that two projects have the same expected cost, EVI suggests to re-evaluate the project with the better evaluation accuracy, hence, greater ρ_i.

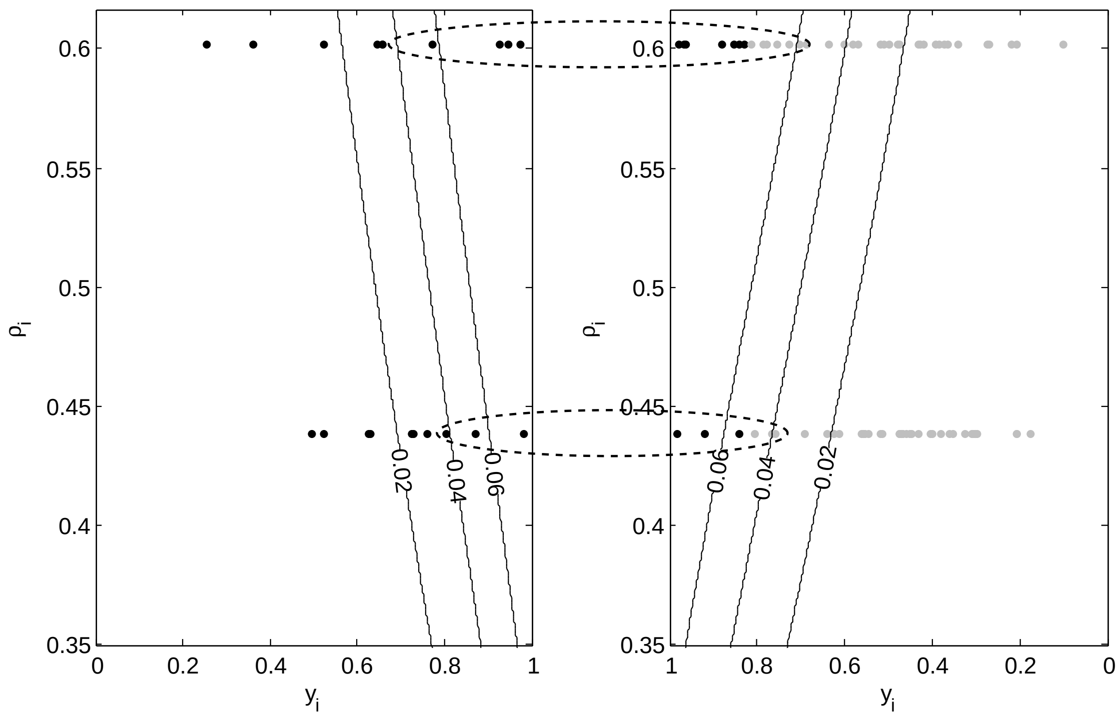


Figure 4: Contours of EVI$[C_i^E]$ for log-normally distributed project costs and evaluation errors (see Proposition 5). Selection of 20 out of 100 projects. 50 projects with $\mathbb{E}[C_i] = 1$, $\mathrm{Var}[C_i] = 0.5$ and $\mathrm{Var}[E_i] = 0.05$ and 50 projects with $\mathbb{E}[C_i] = 1$, $\mathrm{Var}[C_i] = 0.3$ and $\mathrm{Var}[E_i] = 0.1$. 30 projects with the lowest costs marked with black markers. 30 projects with highest EVI marked with dashed ellipses.

3.3 Optimal division of resources between project funding and evaluation

Project evaluation can be expensive and time-consuming. Therefore, it is important to consider how the total resources should be divided between project funding and evaluation. One strategy for doing this could be to compute the EVIs for all projects and, if there was at least one project whose EVI would exceed the cost of obtaining one additional evaluation, to re-evaluate the project with the maximal EVI. This process could be repeated several times. Due to time constraints, however, such an approach may not be feasible, but it may be necessary to submit multiple projects for re-evaluation simultaneously.

Figure 6 compares the performances of such a batch-mode approach in a setting where 20 out of 100 projects are to be selected, and projects are selected for re-evaluation in each round based on four different strategies: (i) complete re-evaluation of all 100 projects, (ii) re-evaluation of 30 projects with the highest EVI, (iii) re-evaluation of 30 projects with the lowest expected costs, and (iv) re-evaluation of 30 randomly selected projects. The average performance of these strategies is computed using 5000 simulation rounds such that $\mathbb{E}[C_i] = 1$, $\mathrm{Var}[C_i] = 0.3$, and $\mathrm{Var}[E_i] = \{0.1, 0.3, 0.5\}$. Distribution of E_i with $\mathrm{Var}[E_i] = 0.5$ is illustrated in figure 5. Applied evaluation costs are 0.5% and 1% of the prior mean cost. From figure 6 we can see that strategy with 30 highest EVI outperforms all other strategies in all cases. The figure also shows that the better the evaluation accuracy is, the less profitable the re-evaluations are. The found optimum, however, gets better along with the evaluation accuracy.

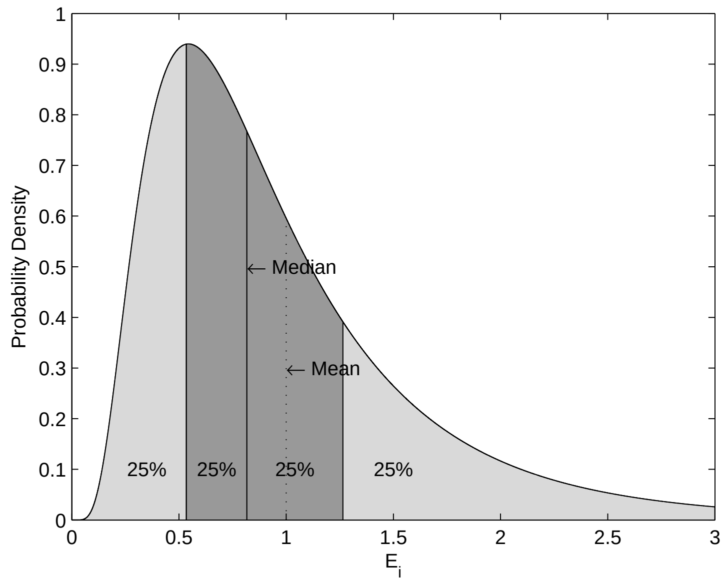


Figure 5: Probability density function of log-normally distributed random variable E_i with a mean of 1 and a variance of 0.5. Dark gray area illustrates how far from the median 50% of the values lie.

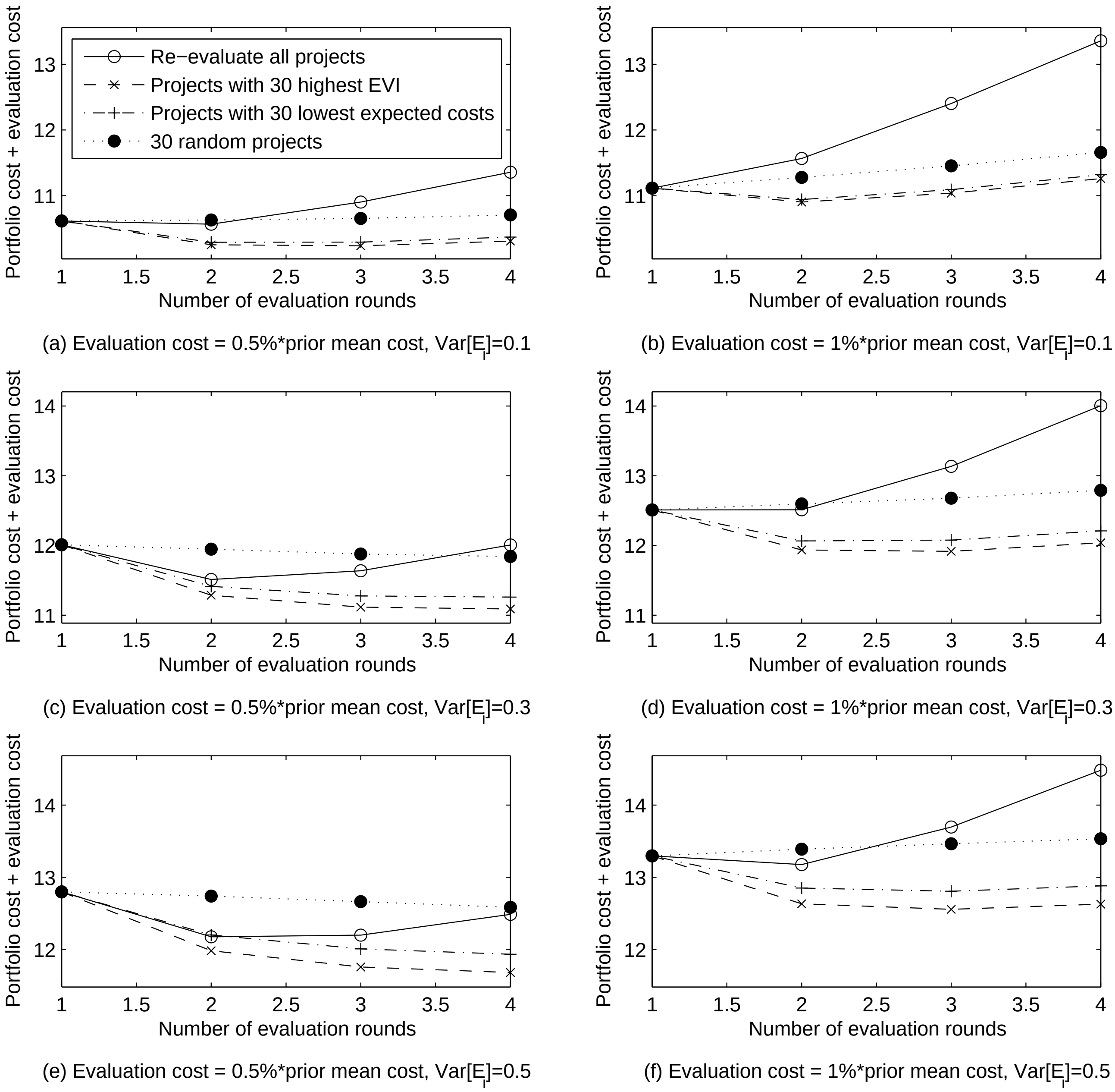


Figure 6: Performance of four different re-evaluation strategies measured by portfolio cost plus evaluation costs. Applied parameters are $\mathbb{E}[C_i] = 1$, $\mathrm{Var}[C_i] = 0.3$, and $\mathrm{Var}[E_i] = \{0.1, 0.3, 0.5\}$. Evaluation costs are 0.5% and 1% times prior mean cost.

Clearly, there are various different kinds of strategies to select projects for re-evaluation. For instance, the number of evaluation rounds and the number of projects to be re-evaluated on each evaluation round can be changed. Assuming that the same number of evaluations is acquired on each round, the optimal number of rounds and evaluations can be determined from the optimization problem

$$\min_{k,e}[\mathbb{E}[\min_{z \in Z} \; z\mathbb{E}[C|C^E(k,e)]] + (n + k(e-1))c_e], \tag{7}$$

where n is the number of project proposals, e is the number of evaluation rounds and c_e is the cost of evaluating one project. $C^E(k,e)$ denotes the random variable which represents the cost estimates based on the initial evaluations for all projects as well as the additional evaluations obtained in rounds $2,\dots,e$ for projects with the k highest EVIs in these rounds. Here, we assume that $e \leq 4$, because execution of more than four evaluation rounds is considered to be impractical in real life portfolio selection problems.

Problem (7) can be solved through simulation by sampling values of the objective function for different combinations of (k,e). Figure 7 shows the optimal division of resources, the optimal number of evaluation rounds e and the optimal number of evaluations k for different evaluation costs and evaluation accuracies. We can see that the total number of evaluations decreases and total money spend on evaluations and project funding increases when the evaluation cost increases. Comparing of figures 7a and 7b shows that with better evaluation accuracy the total number of evaluations is smaller and the found portfolio cost is better. Also, in both cases, the realized portfolio cost increases along with the evaluation costs, even though, share of the money spend on project funding decreases.

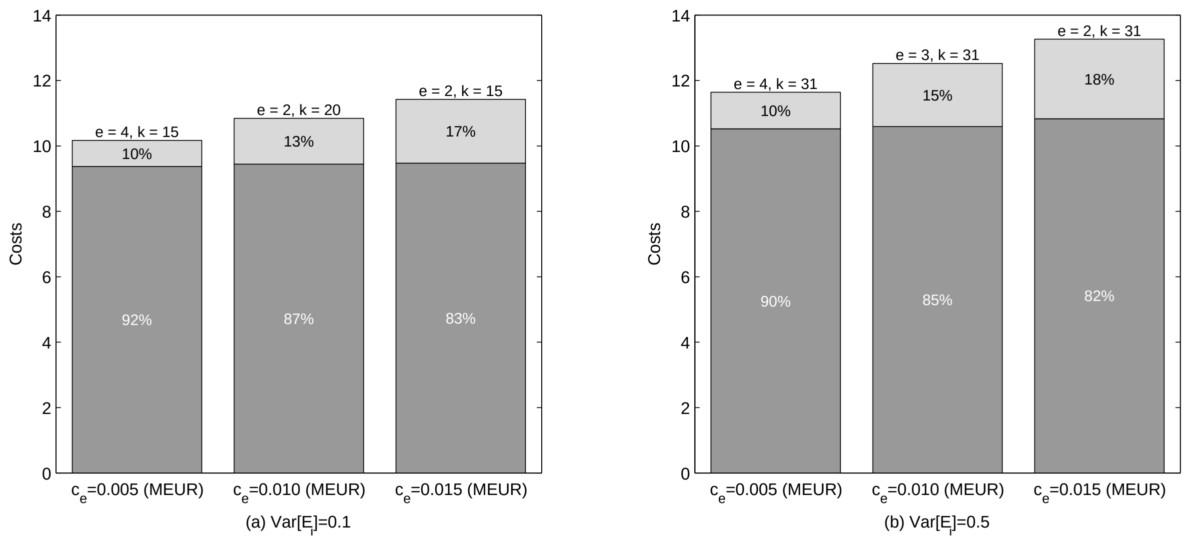


Figure 7: Optimal division of resources for different evaluation costs and evaluation accuracies. Dark and light gray bars indicate the share of resources allocated to project funding and evaluation, respectively.

4 Discussion and conclusions

In this study we developed a Bayesian model to account for cost uncertainties in project portfolio selection. Bayesian modeling of uncertainties was shown to minimize the expected cost of the selected portfolio and to eliminate the expected positive gap between the true and estimated portfolio cost, i.e., post-decision disappointment. Moreover, with Bayesian modeling it is possible to study the value of obtaining additional estimates and determine the projects to be re-evaluated. Analytic results were derived for expected value of additional information in the case where projects' costs and cost estimates were log-normally distributed. A general guideline would be to re-evaluate projects whose (i) current expected cost is near the selection threshold and (ii) posterior variance is relatively large. Furthermore, we studied how resources should be divided between project funding and evaluation and showed how this division depends on the evaluation cost and evaluation accuracy.

Flyvbjerg et al. (2002) did comparative studies of actual costs and estimated costs in the transportation infrastructure projects and concluded that the cost estimates used in the decision making for project selection are highly, systematically, and significantly deceptive. With the proposed Bayesian framework decision makers can gather more precise information about projects' costs and mitigate the experienced post-decision disappointment. Additionally, our study showed that it can be more cost-efficient to re-evaluate only a subset of project proposals and a method for selecting the subset was presented. Performed cost-efficiency analyses showed that presented method for selection of projects to be re-evaluated performed better than conventional methods like short-list method or complete re-evaluation.

Our results can be extended in several ways. Firstly, it would be important to apply the model to real data. Usually companies track cost-efficiency of past executed projects, thus, distribution parameters needed in presented framework could be estimated from the historical data. Secondly, framework could be extended by developing a model that accounts both value and cost uncertainties and correlations between them two. This way number of targets of application could be substantially increased.

Appendix A

Proof of Proposition 1

By assumption, $C_i \sim \text{LogN}(\mu_i, \sigma_i^2)$ and $C_i^E = c_i E_i$, where $E_i \sim \text{LogN}(-\frac{\tau_i^2}{2}, \tau_i^2)$. Then, $\ln(C_i) - \frac{\tau_i^2}{2} \sim \text{N}(\mu_i - \frac{\tau_i^2}{2}, \sigma_i^2)$. With normal prior and log-normal likelihood $(C_i^E \mid C_i = c_i) \sim \text{LogN}(\ln(c_i) - \frac{\tau_i^2}{2}, \tau_i^2)$, the posterior distribution for the unknown first parameter of the likelihood distribution becomes (see e.g. Fink 1997)

$$\left(\ln(C_i) - \frac{\tau_i^2}{2} \mid C_i^E = c_i^E\right) \sim \text{N}\left(\frac{\sigma_i^2}{\sigma_i^2+\tau_i^2}\ln(c_i^E) + \frac{\tau_i^2}{\sigma_i^2+\tau_i^2}(\mu_i - \frac{\tau_i^2}{2}), \frac{\sigma_i^2\tau_i^2}{\sigma_i^2+\tau_i^2}\right) \Rightarrow$$

$$\left(\ln(C_i) \mid C_i^E = c_i^E\right) \sim \text{N}\left(\frac{\sigma_i^2}{\sigma_i^2+\tau_i^2}(\ln(c_i^E) - \frac{\tau_i^2}{2}) + \frac{\tau_i^2}{\sigma_i^2+\tau_i^2}\mu_i, \frac{\sigma_i^2\tau_i^2}{\sigma_i^2+\tau_i^2}\right) \Rightarrow$$

$$\left(C_i \mid C_i^E = c_i^E\right) \sim \text{LogN}\left(\frac{\sigma_i^2}{\sigma_i^2+\tau_i^2}(\ln(c_i^E) - \frac{\tau_i^2}{2}) + \frac{\tau_i^2}{\sigma_i^2+\tau_i^2}\mu_i, \frac{\sigma_i^2\tau_i^2}{\sigma_i^2+\tau_i^2}\right). \quad \square$$

Proof of Proposition 3

Proof is done by observing how the evaluation C_i^E changes the expected portfolio value $\gamma = \max_{z \in Z(b)} \; z\mathbb{E}[C] = z^*\mathbb{E}[C]$, where $Z(b) = \{z \in \{0,1\}^m \mid \sum_{i=1}^{m} z_i \le b\}$ and $z^* \in Z(b)$. For any $i \in \{0, \ldots, m\}$, C_i^E, C_j are independent whenever $i \neq j$ and thus $\mathbb{E}[C_j|C_i^E] = \mathbb{E}[C_j]$. Hence by definition

$$\begin{aligned}
\text{EVI}[C_i^E] &= \gamma - \mathbb{E}\left[\min_{z\in Z(b)} z\mathbb{E}[C|C_i^E]\right] \\
&= \gamma - \mathbb{E}\left[\min_{z\in Z(b)}\left(z_i\mathbb{E}[C_i|C_i^E] + \sum_{\substack{j=1 \\ j\neq i}}^{m} z_j\mathbb{E}[C_i]\right)\right] \\
&= \gamma - \mathbb{E}\left[\min\{\min_{\substack{z\in Z(b) \\ z_i=0}} z\mathbb{E}[C], \mathbb{E}[C_i|C_i^E] + \min_{\substack{z\in Z(b-1) \\ z_i=0}} z\mathbb{E}[C]\}\right]. \qquad (8)
\end{aligned}$$

If $z_i^* = 0$, we have

$$\min_{\substack{z\in Z(b) \\ z_i=0}} z\mathbb{E}[C] = \gamma, \; \min_{\substack{z\in Z(b-1) \\ z_i=0}} z\mathbb{E}[C] = \gamma - x^+,$$

which can be substituted into (8) to obtain

$$\begin{aligned}
\mathrm{EVI}[C_i^E] &= \gamma - \mathbb{E}\left[\min\{\gamma, \mathbb{E}[C_i|C_i^E] + \gamma - x^+\}\right] \\
&= \gamma - \mathbb{E}\left[\gamma + \min\{0, \mathbb{E}[C_i|C_i^E] - x^+\}\right] \\
&= -\mathbb{E}\left[\min\{0, \mathbb{E}[C_i|C_i^E] - x^+\}\right] \\
&= \mathbb{E}\left[\max\{0, x^+ - \mathbb{E}[C_i|C_i^E]\}\right].
\end{aligned}$$

If $z_i^* = 1$, we have

$$\min_{\substack{z\in Z(b)\\ z_i=0}} z\mathbb{E}[C] = \gamma - \mathbb{E}[C_i] + x^-, \quad \min_{\substack{z\in Z(b-1)\\ z_i=0}} z\mathbb{E}[C] = \gamma - \mathbb{E}[C_i],$$

which can be substituted into (8) to obtain

$$\begin{aligned}
\mathrm{EVI}[C_i^E] &= \gamma - \mathbb{E}\left[\min\{\gamma - \mathbb{E}[C_i] + x^-, \mathbb{E}[C_i|C_i^E] + \gamma - \mathbb{E}[C_i]\}\right] \\
&= -\mathbb{E}\left[\min\{-\mathbb{E}[C_i] + x^-, \mathbb{E}[C_i|C_i^E] - \mathbb{E}[C_i]\}\right] \\
&= -\mathbb{E}\left[\min\{-\mathbb{E}[C_i] + x^- - \mathbb{E}[C_i|C_i^E] + \mathbb{E}[C], 0\}\right] - \mathbb{E}\left[\mathbb{E}[C_i|C_i^E]\right] + \mathbb{E}[C_i] \\
&= -\mathbb{E}\left[\min\{x^- - \mathbb{E}[C_i|C_i^E], 0\}\right] \\
&= \mathbb{E}\left[\max\{\mathbb{E}[C_i|C_i^E] - x^-, 0\}\right]. \quad \square
\end{aligned}$$

Proof of Proposition 4

Let C_i^E be random. Then, equation (2) gives

$$\ln(\mathbb{E}\left[C_i \mid C_i^E\right]) = \frac{1}{\sigma_i^2 + \tau_i^2}(\sigma_i^2 \ln(C_i^E) + \sigma_i^2\tau_i^2 + \tau_i^2\mu_i). \tag{9}$$

From the likelihood distribution $(C_i^E \mid C_i = c_i) \sim \mathrm{LogN}(\ln(c_i) - \frac{\tau_i^2}{2}, \tau_i^2)$ we get

$$\ln(C_i^E) \sim \mathrm{N}(\ln(C_i) - \frac{\tau_i^2}{2}, \tau_i^2).$$

Because the prior distribution is $\ln(C_i) \sim \mathrm{N}(\mu_i, \sigma_i^2)$, we can write

$$\ln(C_i^E) = \ln(C_i) - \frac{\tau_i^2}{2} + \delta_i = \mu_i + \epsilon_i - \frac{\tau_i^2}{2} + \delta_i, \tag{10}$$

where $\epsilon_i \sim \mathrm{N}(0, \sigma_i^2)$, $\delta_i \sim \mathrm{N}(0, \tau_i^2)$ and $\epsilon_i \perp \delta_i$. Substituting (10) into (9) gives

$$\begin{aligned}
\ln(\mathbb{E}[C_i|C_i^E]) &= \frac{1}{\sigma_i^2 + \tau_i^2}(\sigma_i^2(\mu_i + \epsilon_i - \frac{\tau_i^2}{2} + \delta_i) + \sigma_i^2\tau_i^2 + \tau_i^2\mu_i) \\
&= \underbrace{(\mu_i + \frac{\sigma_i^2\tau_i^2}{2(\sigma_i^2 + \tau_i^2)})}_{\text{constant, } Var(\cdot)=0} + \underbrace{\frac{\sigma_i^2}{\sigma_i^2 + \tau_i^2}(\epsilon_i + \delta_i)}_{\text{random, } \mathbb{E}[\cdot]=0} \\
&\sim \mathrm{N}(\mu_i + \frac{\sigma_i^2\tau_i^2}{2(\sigma_i^2 + \tau_i^2)}, \frac{\sigma_i^4}{\sigma_i^2 + \tau_i^2}).
\end{aligned}$$

From this it follows that

$$\mathbb{E}[C_i|C_i^E] \sim \text{LogN}(\mu_i + \frac{\sigma_i^2\tau_i^2}{2(\sigma_i^2+\tau_i^2)}, \frac{\sigma_i^4}{\sigma_i^2+\tau_i^2}). \quad \square \tag{11}$$

Proof of Proposition 5

We first need to derive closed-form representations for partial expectations $\mathbb{E}[X|X < k]$ and $\mathbb{E}[X|X > k]$. Let $X \sim \text{LogN}(\mu, \sigma^2)$. Partial expectation $\mathbb{E}[X|X > k]$ is defined

$$\mathbb{E}[X|X > k] = \int_k^\infty xf(x)dx \Big/ \Pr(X > k), \tag{12}$$

where $f(x)$ is the probability density function of random variable X. From (12) we need to solve $g(k) = \int_k^\infty xf(x)dx$. From the properties of log-normal distribution we have

$$g(k) = \int_k^\infty \frac{x}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} dx.$$

With a change of variables $y = \frac{\ln(x)-\mu}{\sigma}$ and $dx = \sigma e^{\sigma y+\mu}$ we get

$$\begin{aligned} g(k) &= \int_{\frac{\ln(k)-\mu}{\sigma}}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2}} \sigma e^{\sigma y+\mu} dy = \int_{\frac{\ln(k)-\mu}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}+\sigma y+\mu} dy \\ &= \int_{\frac{\ln(k)-\mu}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\sigma)^2+(\mu+\frac{\sigma^2}{2})} dy. \end{aligned}$$

Second change of variables $v = y - \sigma$ and $dy = dv$ gives

$$\begin{aligned} g(k) &= \int_{\frac{\ln(k)-\mu-\sigma^2}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v)^2+(\mu+\frac{\sigma^2}{2})} dv = e^{\mu+\frac{\sigma^2}{2}} \int_{\frac{\ln(k)-\mu-\sigma^2}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} dv \\ &= e^{\mu+\frac{\sigma^2}{2}} \left(1 - \Phi\left(\frac{\ln(k)-\mu-\sigma^2}{\sigma}\right)\right) = e^{\mu+\frac{\sigma^2}{2}} \Phi\left(\frac{\mu+\sigma^2-\ln(k)}{\sigma}\right), \end{aligned}$$

where Φ is the cumulative probability function of the standard normal distribution. Then,

$$\begin{aligned} \mathbb{E}[X|X > k] &= e^{\mu+\frac{\sigma^2}{2}} \Phi\left(\frac{\mu+\sigma^2-\ln(k)}{\sigma}\right) \Big/ \Pr(X > k) \\ &= e^{\mu+\frac{\sigma^2}{2}} \Phi\left(\frac{\mu+\sigma^2-\ln(k)}{\sigma}\right) \Big/ \Phi\left(\frac{\mu-\ln(k)}{\sigma}\right) \end{aligned} \tag{13}$$

Partial expectation $\mathbb{E}[X|X<k]$ is defined

$$\mathbb{E}[X|X<k] = \int_0^k xf(x)dx \Big/ \Pr(X<k). \tag{14}$$

From (14) we need to solve $h(k) = \int_0^k xf(x)dx$. This can be formulated

$$h(k) = \int_0^k \frac{x}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} dx. = \int_{-\infty}^{\frac{\ln(k)-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}+\sigma y+\mu} dy$$

$$= \int_{-\infty}^{\frac{\ln(k)-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\sigma)^2+(\mu+\frac{\sigma^2}{2})} dy. = \int_{-\infty}^{\frac{\ln(k)-\mu-\sigma^2}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v)^2+(\mu+\frac{\sigma^2}{2})} dv$$

$$= e^{\mu+\frac{\sigma^2}{2}} \int_{-\infty}^{\frac{\ln(k)-\mu-\sigma^2}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} dv = e^{\mu+\frac{\sigma^2}{2}} \Phi\left(\frac{\ln(k)-\mu-\sigma^2}{\sigma}\right)$$

Then,

$$\mathbb{E}[X|X<k] = e^{\mu+\frac{\sigma^2}{2}} \Phi\left(\frac{\ln(k)-\mu-\sigma^2}{\sigma}\right) \Big/ \Phi\left(\frac{\ln(k)-\mu}{\sigma}\right). \tag{15}$$

Denote $C_i^B = \mathbb{E}[C_i|C_i^E]$. From (11) we get $C_i^B \sim \text{LogN}(\mu_i + \frac{\sigma_i^2\tau_i^2}{2(\sigma_i^2+\tau_i^2)}, \frac{\sigma_i^4}{\sigma_i^2+\tau_i^2})$. Let $\rho_i^2 = \frac{\tau_i^4}{\tau_i^2+\sigma_i^2}$ and $\xi_i = \frac{\tau_i^2\sigma_i^2}{\tau_i^2+\sigma_i^2}$, hence, $\rho_i^2+\xi_i^2 = \tau_i^2$. Using equations (13) and (15) gives

$$\mathbb{E}[C_i^B|C_i^B > x^-]$$

$$= e^{\mu_i+\frac{\sigma_i^2\tau_i^2}{2(\sigma_i^2+\tau_i^2)}+\frac{\frac{\sigma_i^4}{\sigma_i^2+\tau_i^2}}{2}} \Phi\left(\frac{\mu_i+\frac{\sigma_i^2\tau_i^2}{2(\sigma_i^2+\tau_i^2)}+\frac{\sigma_i^4}{\sigma_i^2+\tau_i^2}-\ln(x^-)}{\frac{\sigma_i^2}{\sqrt{\sigma_i^2+\tau_i^2}}}\right) \Big/ \Phi\left(\frac{\mu_i+\frac{\sigma_i^2\tau_i^2}{2(\sigma_i^2+\tau_i^2)}-\ln(x^-)}{\frac{\sigma_i^2}{\sqrt{\sigma_i^2+\tau_i^2}}}\right)$$

$$= e^{\mu_i+\frac{1}{2}\tau_i^2} \Phi\left(\frac{\mu_i+\frac{1}{2}\tau_i^2+\frac{1}{2}\rho_i^2-\ln(x^-)}{\rho_i}\right) \Big/ \Phi\left(\frac{\mu_i+\frac{1}{2}\xi_i^2-\ln(x^-)}{\rho_i}\right)$$

and

$$\mathbb{E}[C_i^B|C_i^B < x^+]$$

$$= e^{\mu_i + \frac{\sigma_i^2\tau_i^2}{2(\sigma_i^2+\tau_i^2)} + \frac{\frac{\sigma_i^4}{\sigma_i^2+\tau_i^2}}{2}} \Phi\left(\frac{\ln(x^+) - \mu_i - \frac{\sigma_i^2\tau_i^2}{2(\sigma_i^2+\tau_i^2)} - \frac{\sigma_i^4}{\sigma_i^2+\tau_i^2}}{\frac{\sigma_i^2}{\sqrt{\sigma_i^2+\tau_i^2}}}\right) \Bigg/ \Phi\left(\frac{\ln(x^+) - \mu_i - \frac{\sigma_i^2\tau_i^2}{2(\sigma_i^2+\tau_i^2)}}{\frac{\sigma_i^2}{\sqrt{\sigma_i^2+\tau_i^2}}}\right).$$

$$= e^{\mu_i + \frac{1}{2}\tau_i^2} \Phi\left(\frac{\ln(x^+) - \mu_i - \frac{1}{2}\tau_i^2 - \frac{1}{2}\rho_i^2}{\rho_i}\right) \Bigg/ \Phi\left(\frac{\ln(x^+) - \mu_i - \frac{1}{2}\xi_i^2}{\rho_i}\right)$$

If $z_i^* = 0$, equation (5) implies

$$\begin{aligned}
&\text{EVI}[C_i^E] \\
&= \mathbb{E}\left[\max\{0, x^+ - C_i^B\}\right] \\
&= \Pr\left(C_i^B < x^+\right)\left(x^+ - \mathbb{E}[C_i^B|C_i^B < x^+]\right) \\
&= \Phi\left(\frac{\ln(x^+) - \mu_i - \frac{1}{2}\xi_i^2}{\rho_i}\right)\left(x^+ - \mathbb{E}[C_i^B|C_i^B < x^+]\right) \\
&= \Phi\left(\frac{\ln(x^+) - \mu_i - \frac{1}{2}\tau_i^2 - \frac{1}{2}\rho_i^2}{\rho_i}\right) x^+ - e^{\mu_i + \frac{1}{2}\tau_i^2} \Phi\left(\frac{\ln(x^+) - \frac{1}{2}\tau_i^2 - \frac{1}{2}\rho_i^2}{\rho_i}\right)
\end{aligned}$$

If $z_i^* = 1$, equation (5) implies

$$\begin{aligned}
&\text{EVI}[C_i^E] \\
&= \mathbb{E}\left[\max\{0, C_i^B - x^-\}\right] \\
&= \Pr\left(C_i^B > x^-\right)\left(\mathbb{E}[C_i^B|C_i^B > x^-] - x^-\right) \\
&= \Phi\left(\frac{\mu_i + \frac{1}{2}\xi_i^2 - \ln(x^-)}{\rho_i}\right)\left(\mathbb{E}[C_i^B|C_i^B > x^-] - x^-\right) \\
&= e^{\mu_i + \frac{1}{2}\tau_i^2} \Phi\left(\frac{\mu_i + \frac{1}{2}\tau_i^2 + \frac{1}{2}\rho_i^2 - \ln(x^-)}{\rho_i}\right) - \Phi\left(\frac{\mu_i + \frac{1}{2}\tau_i^2 + \frac{1}{2}\rho_i^2 - \ln(x^-)}{\rho_i}\right) x^-
\end{aligned}$$

Denote $x_i = e^{\mu_i + \frac{1}{2}\tau_i^2} = \mathbb{E}[C_i^B]$ and $y_i = \min\{\frac{x^+}{x_i}, \frac{x_i}{x^-}\}$. We know that $y_i \in [0, 1]$ because if $z_i^* = 1$, $x^- - \mathbb{E}[C_i^B] \geq 0$, and if $z_i^* = 0$, $\mathbb{E}[C_i^B] - x^+ \geq 0$. Thus, EVI can be formulated

$$\text{EVI}[C_i^E] = f(y_i, \rho_i) = \begin{cases} x^-\left[y_i\Phi\left(\frac{\ln(y_i)}{\rho_i} + \frac{1}{2}\rho_i\right) - \Phi\left(\frac{\ln(y_i)}{\rho_i} - \frac{1}{2}\rho_i\right)\right], & \text{if } z_i^* = 1 \\ x^+\left[\Phi\left(\frac{\ln(y_i)}{\rho_i} + \frac{1}{2}\rho_i\right) - \frac{1}{y_i}\Phi\left(\frac{\ln(y_i)}{\rho_i} - \frac{1}{2}\rho_i\right)\right], & \text{if } z_i^* = 0 \end{cases}.$$

We need to study how $f(y_i, \rho_i)$ changes with respect to y_i and ρ_i. Consider the case with $z_i^* = 1$. Then, $y_i = \frac{x_i}{x^-}$, and

$$\begin{aligned}
\frac{\partial f(y_i,\rho_i)}{\partial y_i} &= x^- \left(\Phi\left(\frac{\ln(y_i)}{\rho_i} + \frac{1}{2}\rho_i\right) + \frac{1}{\rho_i}\varphi\left(\frac{\ln(y_i)}{\rho_i} + \frac{1}{2}\rho_i\right) - \frac{1}{y_i\rho_i}\varphi\left(\frac{\ln(y_i)}{\rho_i} - \frac{1}{2}\rho_i\right)\right) \\
&> x^- \left(\frac{1}{\rho_i}\varphi\left(\frac{\ln(y_i)}{\rho_i} + \frac{1}{2}\rho_i\right) - \frac{1}{y_i\rho_i}\varphi\left(\frac{\ln(y_i)}{\rho_i} - \frac{1}{2}\rho_i\right)\right) \\
&= \frac{x^-}{\rho_i\sqrt{2\pi}} \left(e^{-\frac{1}{2}\left(\frac{\ln^2(y_i)}{\rho_i^2} + \ln(y_i) + \frac{\rho_i^2}{4}\right)} - \frac{1}{y_i} e^{-\frac{1}{2}\left(\frac{\ln^2(y_i)}{\rho_i^2} - \ln(y_i) + \frac{\rho_i^2}{4}\right)}\right) \\
&= \frac{x^-}{\rho_i\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln^2(y_i)}{\rho_i^2} + \frac{\rho_i^2}{4}\right)} \left(\frac{1}{\sqrt{y_i}} - \frac{1}{\sqrt{y_i}}\right) = 0,
\end{aligned}$$

so that $f(y_i,\rho_i)$ is increasing in y_i. Also,

$$\begin{aligned}
\frac{\partial f(y_i,\rho_i)}{\partial \rho_i} &= x^- \left(\varphi\left(\frac{\ln(y_i)}{\rho_i} - \frac{1}{2}\rho_i\right)\left(\frac{\ln(y_i)}{\rho_i^2} + \frac{1}{2}\right) - y_i\varphi\left(\frac{\ln(y_i)}{\rho_i} + \frac{1}{2}\rho_i\right)\left(\frac{\ln(y_i)}{\rho_i^2} - \frac{1}{2}\right)\right) \\
&= \frac{x^-}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln^2(y_i)}{\rho_i^2} + \frac{\rho_i^2}{4}\right)} \left(\sqrt{y_i}\left(\frac{\ln(y_i)}{\rho_i^2} + \frac{1}{2}\right) - \sqrt{y_i}\left(\frac{\ln(y_i)}{\rho_i^2} - \frac{1}{2}\right)\right) \\
&= \frac{x^-\sqrt{y_i}}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln^2(y_i)}{\rho_i^2} + \frac{\rho_i^2}{4}\right)} > 0,
\end{aligned}$$

so that $f(y_i,\rho_i)$ is increasing in ρ_i. The proof for $z_i^* = 0$ is similar. $\square$

References

K. C. Brown. A note on the apparent bias of net revenue estimates for capital investment projects. *Journal of Finance*, 29:1215–1216, 1974.

P. Delquié. The value of information and intensity of preference. *Decision Analysis*, 5(3):129–139, 2008.

D. Fink. A compedium of conjugate priors. Technical report, 1997. <http://www.johndcook.com/CompendiumOfConjugatePriors.pdf> Retrieved 15.07.2014.

B. Flyvbjerg, H. Skamris, and S. Buhl. Underestimating costs in public works projects: Error or lie? *Journal of the American Planning Association*, 68 (3):279–295, 2002.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis (2nd ed.).* London UK: Chapman & Hall, 2004.

J. P. Gould. Risk, stochastic preference, and the value of information. *Journal of Economic Theory*, 8:64–84, 1974.

J. R. Harrison and J. G. March. Decision making and post-decision surprises. *Adminstrative Science Quarterly*, 29:26–42, 1984.

J. Keisler. Value of information in portfolio decision analysis. *Decision Analysis*, 1(4):177–189, 2004.

I. H. La Valle. On cash equivalents and information evaluations in decisions under uncertainty - part i: Basic theory. *Journal of the American Statistical Association*, 36:252–276, 1968.

J.-J. Laffont. *Essays in the economics of uncertainty.* Harvard Business Press, Boston, 1980.

J. Marchak and R. Radner. *Economic theory of teams.* Yale University Press, 1972.

A. Salo, J. Keisler, and A. Morton. *Portfolio Decision Analysis: Improved Methods for Resource Allocation.* Springer, New York, 2011.

J. Smith and R. Winkler. The optimizer's curse skepticism and postdecision suprise in decision analysis. *Management Science*, 52(3):311–322, 2006.

E. Vilkkumaa, J. Liesiö, and A. Salo. Optimal strategies for selecting project portfolios using uncertain value estimates. *European Journal of Operational Research*, 233:772–783, 2014.