

Designated Drivers

Written on October 17, 2012 by [Tony Welsh](#)

There has been a lot of talk this year about the use of drivers in schedule risk analysis. The idea is that the correlations between task durations are the result of one or more outside influences (the drivers) and that their influence on task duration can and should be modeled directly. As a software vendor, I have considered supporting drivers and probably will do so in the future, but I do have some concerns.

Before going into these concerns, I should explain the idea as I understand it. The sampled duration for a task is modified by one or more randomly generated multiplicative factors representing the drivers. The value of each driver is drawn from a distribution roughly centered on 1. For example, if a task is affected by two drivers these might be represented by two triangular distributions $T(0.7, 1.0, 1.2)$ and $T(0.9, 1.0, 1.4)$. On a particular iteration, we might sample 1.04 and .98, so the value sampled for the task duration is multiplied by 1.04 times .98, or about 1.02.

While I can see some circumstances where this idea might be useful, my concerns relate to its inappropriate use as a substitute for correlations. It is not easy to know what the final distribution of the task duration is or how it is correlated with other tasks tied to one or more of the same drivers. Protagonists say that real people do not think in terms of correlation coefficients and so do not need to know what these are, but that leads to new problems.

Firstly, one has to estimate the distribution of some pretty abstract things; instead of estimating the distribution of how long a task will take, one has to estimate the distribution not of the factor itself (say the weather) but of the effect of this factor on task durations. This is such an abstract concept that its value is not knowable even after the event.

And once you have determined that the driver effect ranges say from 0.8 to 1.3 this applies to all tasks tied to that driver. So it is impossible to model the situation where two tasks are affected by the same outside influence but to different degrees. (You can make the correlation coefficients different, assuming you care about these, but only by changing one or both of the standard deviations in ways you may not want.) And it is impossible to model negative correlations.

I have also come across a fallacy resulting from a misunderstanding about what exactly is meant by two tasks being influenced by the same factor. It is important to understand that this means the two tasks are both influenced by a particular realization of that factor. That is, if on a particular Monte Carlo iteration the sampled value is 1.15, then this factor applies to all tasks tied to that driver. So, you cannot model a factor like estimation uncertainty using a driver, as I have actually heard suggested, because the uncertainty could go either way on each individual task; the fact that all estimates are subject to this source of uncertainty does *not* make them correlated. I fear that such misunderstandings will result in bad models, and hence bad results.

Drivers also require more data entry than directly specifying correlations, and very much more if one wants to do what-if comparisons of different degrees of correlation. (Changing the degree of correlation between two tasks, while leaving their distributions approximately the same, requires changing at least 6 parameters, whereas the same exercise using correlations requires changing only 1 or 2 parameters and leaves the distributions *exactly* the same.)

One advantage of drivers is that they can be applied to events which may not happen at all. You can apply a probability of a driver being active, allowing them to be used to model improbable but possibly catastrophic risks of the sort generally found in a risk register. This is not usually the primary reason for doing quantitative risk analysis, but I think this is the only sensible context in which to use drivers.

My conclusion is therefore that drivers have a place in quantitative risk analysis, but not as a way of modeling correlations.