Historical Data Analysis

John Owen, *President*
Barbecana, Inc.

Introduction

The purpose of the Schedule Risk Analysis is to model the impact of uncertainty on future forecast dates.

Ideally it will be possible to capture best-case, most likely, and worst-case duration estimates, also known as a 3-point estimate, for all the tasks in the schedule. Unfortunately, this is not always possible. The most common issue is that the estimating process is already complete and only the most likely duration estimate was provided to the scheduler.

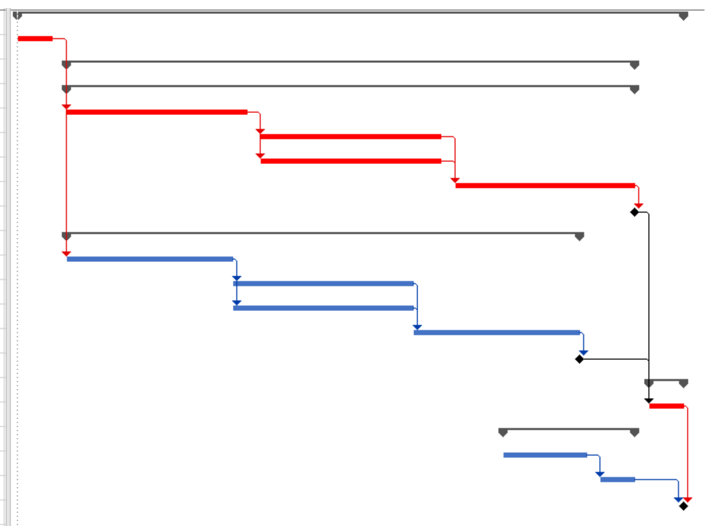
So now we have a schedule with no defined estimates for uncertainty, forecasting a completion date at some point in the future. …and then you are asked to perform a schedule risk analysis!

Generic Uncertainty

There is actually value in performing an analysis with some very generic assumptions about uncertainty.

Consider this simple schedule: -

Sample Project	**70 days**	**6/1/20 8:00 AM**	**9/4/20 5:00 PM**
Initiate	5 days	6/1/20 8:00 AM	6/5/20 5:00 PM
Development	**60 days**	**6/8/20 8:00 AM**	**8/28/20 5:00 PM**
Hardware	**60 days**	**6/8/20 8:00 AM**	**8/28/20 5:00 PM**
HW Task 1	20 days	6/8/20 8:00 AM	7/3/20 5:00 PM
HW Task 2	20 days	7/6/20 8:00 AM	7/31/20 5:00 PM
HW Task 3	20 days	7/6/20 8:00 AM	7/31/20 5:00 PM
HW Task 4	20 days	8/3/20 8:00 AM	8/28/20 5:00 PM
HW Complete	0 days	8/28/20 5:00 PM	8/28/20 5:00 PM
Software	**54 days**	**6/8/20 8:00 AM**	**8/20/20 5:00 PM**
SW Task 1	18 days	6/8/20 8:00 AM	7/1/20 5:00 PM
SW Task 2	18 days	7/2/20 8:00 AM	7/27/20 5:00 PM
SW Task 3	18 days	7/2/20 8:00 AM	7/27/20 5:00 PM
SW Task 4	18 days	7/28/20 8:00 AM	8/20/20 5:00 PM
SW Complete	0 days	8/20/20 5:00 PM	8/20/20 5:00 PM
Integration	**5 days**	**8/31/20 8:00 AM**	**9/4/20 5:00 PM**
Integration	5 days	8/31/20 8:00 AM	9/4/20 5:00 PM
Marketing	**15 days**	**8/10/20 8:00 AM**	**8/28/20 5:00 PM**
Brochure Development	10 days	8/10/20 8:00 AM	8/21/20 5:00 PM
Brochure Printing	5 days	8/24/20 8:00 AM	8/28/20 5:00 PM
Delivery	0 days	9/4/20 5:00 PM	9/4/20 5:00 PM



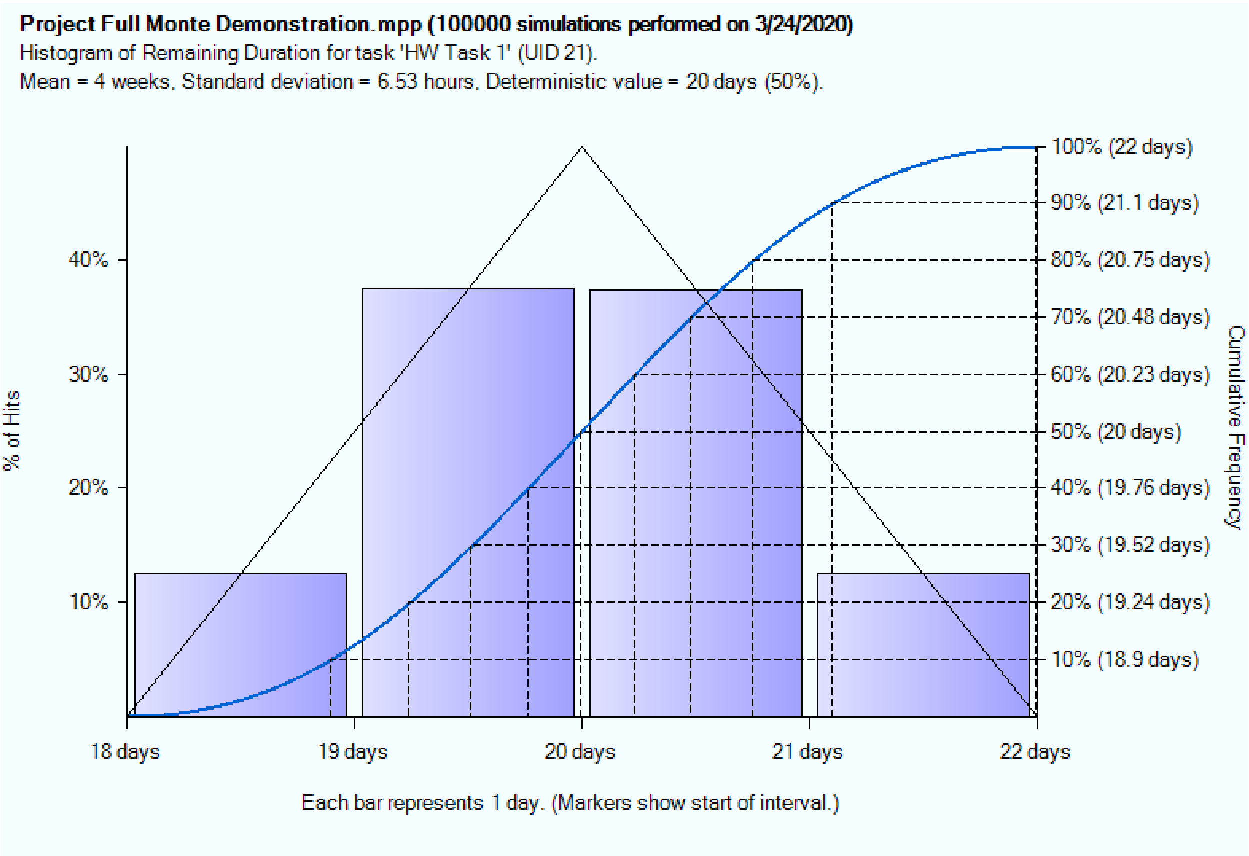
The schedule is predicting a final delivery on September 4, 2020. The question is: How realistic is that date?

Unfortunately, we only have the single point duration estimates used in the schedule as the task duration. So, what can we do to model uncertainty without ruffling the feathers of either the estimators or management?

Most people would agree that there is some element of uncertainty with any task duration estimate so one approach is to simply say we will say every estimate is subject to +/- some value. For example, we could say that every estimate is likely to vary by +/- 10%. On average the work is expected to complete in the estimated duration but may finish up to 10% early or up to 10% late. That's a very fair assumption and unlikely to ruffle feathers. As for the duration distribution type, we will use Triangular as this is the most common distribution type used.

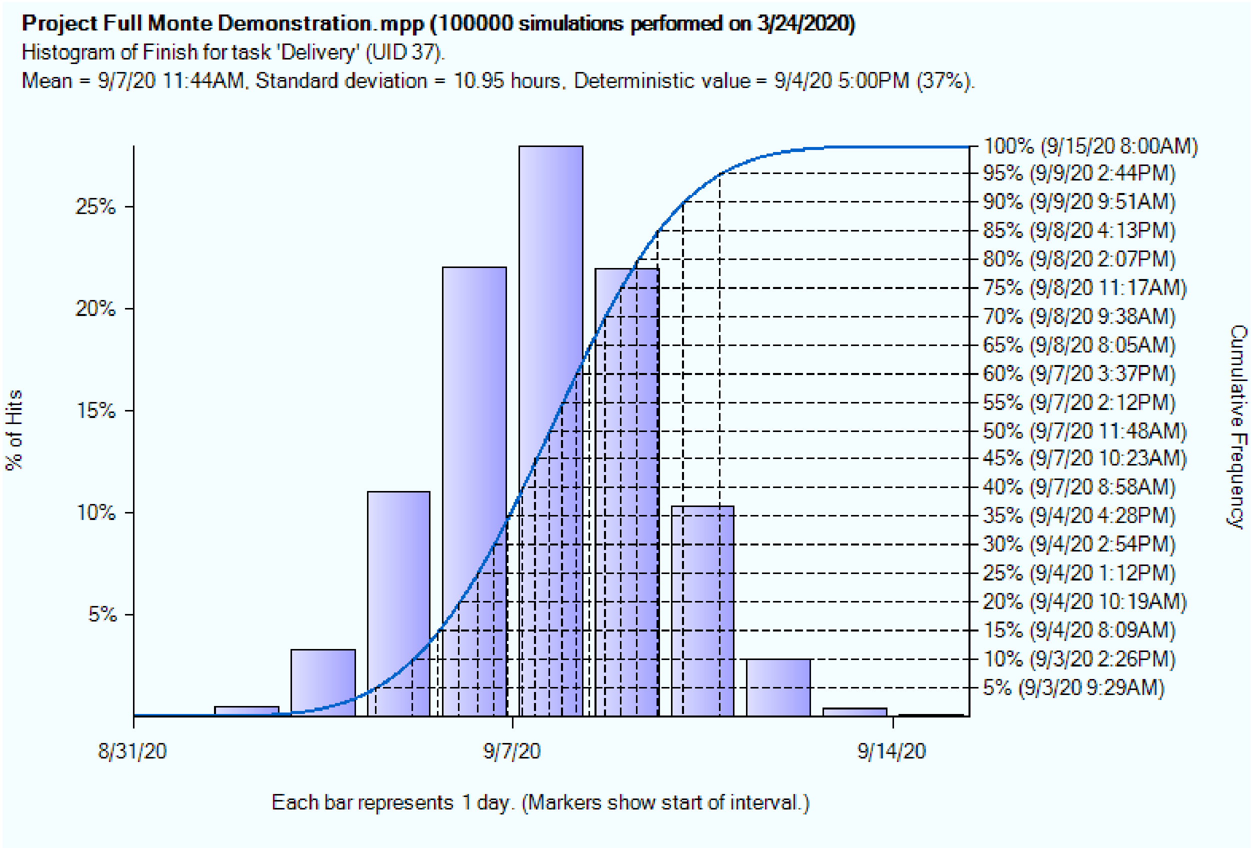
For any individual task, the probability of completing the task in the estimated duration will be 50%.

Here is the Duration Histogram for task HW Task 1, estimated to take 20 days.



The software is showing a 50% chance of completing the work in 20 days – the estimate duration. It may complete is as little as 18 days and as much as 22 days (the +/- 10%).

So, what do we expect the chance of completing the whole project by September 4 to be? Let's view the completion date histogram following the analysis.



It may be a surprise that the software is only predicting a 37% chance of delivering on September 4 despite every individual task, on average, finishing in its estimated duration. Why?

This simple exercise demonstrates the impact of an effect called **Merge Bias**. Fundamentally the probability of any task starting on time is likely to decrease if it has more than one predecessor. This is because, for the successor to start, all the logical predecessors must be complete and if they are subject to uncertainty, even symmetrical uncertainty, then the longest delay will prevail and push out the successor.

Even this simple schedule has four merge points where multiple predecessors merge into a single successor, and each can cause a merge delay.

Note the 80% confidence value (right hand Y-Axis) is September 8.

However, uncertainty is not generally as kind as +/- 10%. How can we improve the realism of our assessment of uncertainty without going back to the estimators for further information?

Historical Analysis

Full Monte can perform an analysis of past projects to compare actual with estimated duration.

Note: If you are using Microsoft Project, then it is required that a baseline was captured before status was updated. This is not a requirement with Primavera P6.

To perform the analysis, open a past project in Full Monte and click View, Open Named View, History.

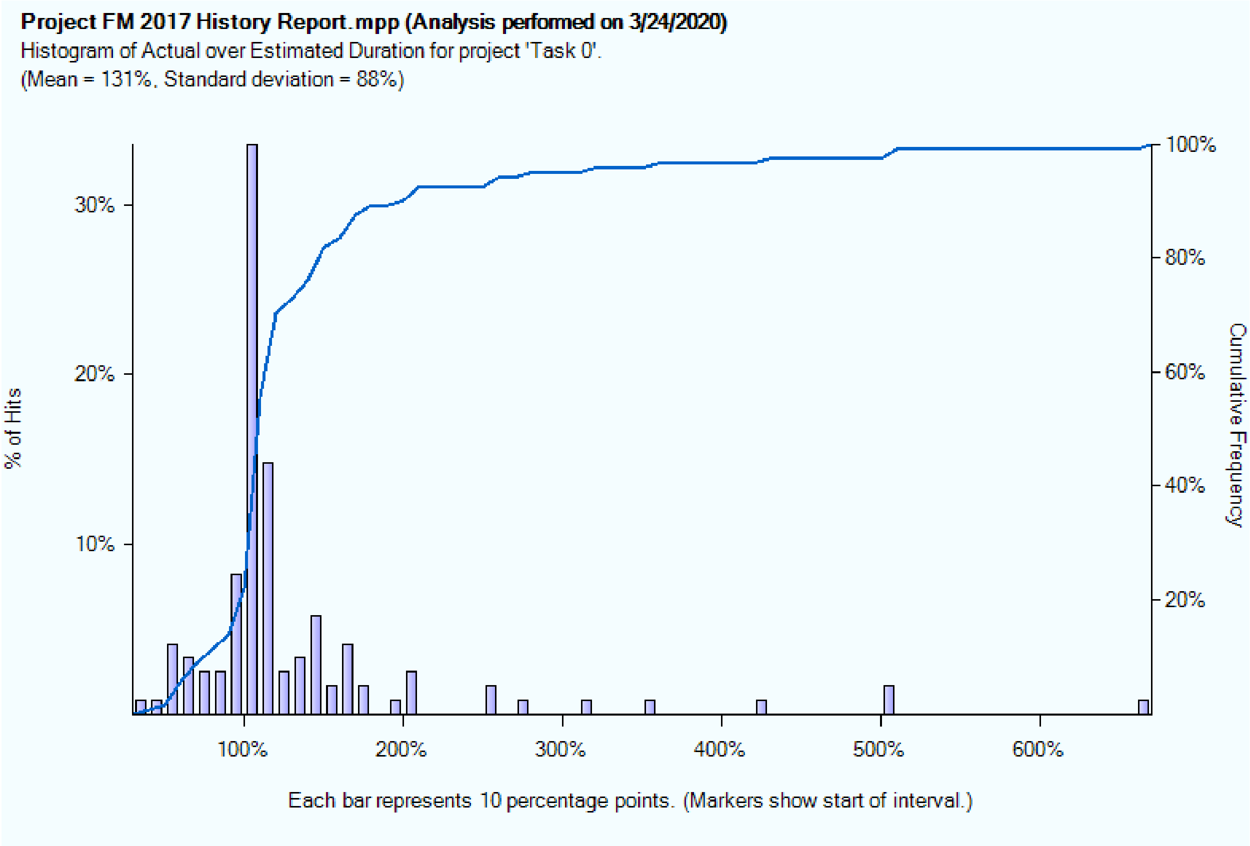
The view will look something like the following:

ID	Task Name	Percent Complete (MSP)	Baseline Duration	Actual Duration	Actual over Estimated Duration
0	⊟ Task 0	33%	0	88 days	Graph
1	⊟ Task 1	33%	540 days	88 days	Graph
176	⊟ Task 176	22%	448 days	42 days	Graph
177	⊟ Task 177	40%	441 days	89 days	Graph
178	⊟ Task 178	46%	270 days	29 days	Graph
179	Task 179	100%	25 days	29 days	116%
180	Task 180	100%	20 days	22 days	110%
181	Task 181	100%	10 days	25 days	250%
182	Task 182	100%	30 days	44 days	146%
183	Task 183	100%	20 days	21 days	105%
184	Task 184	100%	30 days	34 days	113%
185	Task 185	100%	20 days	20 days	100%
189	⊟ Task 189	100%	32 days	60 days	Graph
190	Task 190	100%	10 days	11 days	110%
191	Task 191	100%	12 days	18 days	150%
192	Task 192	100%	5 days	10 days	200%
193	Task 193	100%	5 days	21 days	420%
196	⊟ Task 196	100%	320 days	221 days	Graph
197	Task 197	100%	59 days	94 days	159%

The tabular view is showing the original estimated duration, in this case from the Microsoft project baseline and the actual duration calculated from the Actual Start/Finish Dates.

The Actual over Estimated column will allow us to view the information graphically. Graphs can be viewed at any level of the project structure so we can compare data from different elements of the project (e.g. Design vs Manufacturing etc.).

Here is the chart for the overall project.



The height of the bars indicates the percentage of tasks in each percentage ratio.

In the example, 33.6% of the tasks completed in their estimated duration (Actual/Estimated=1) which shows as the peak at the 100% marker.

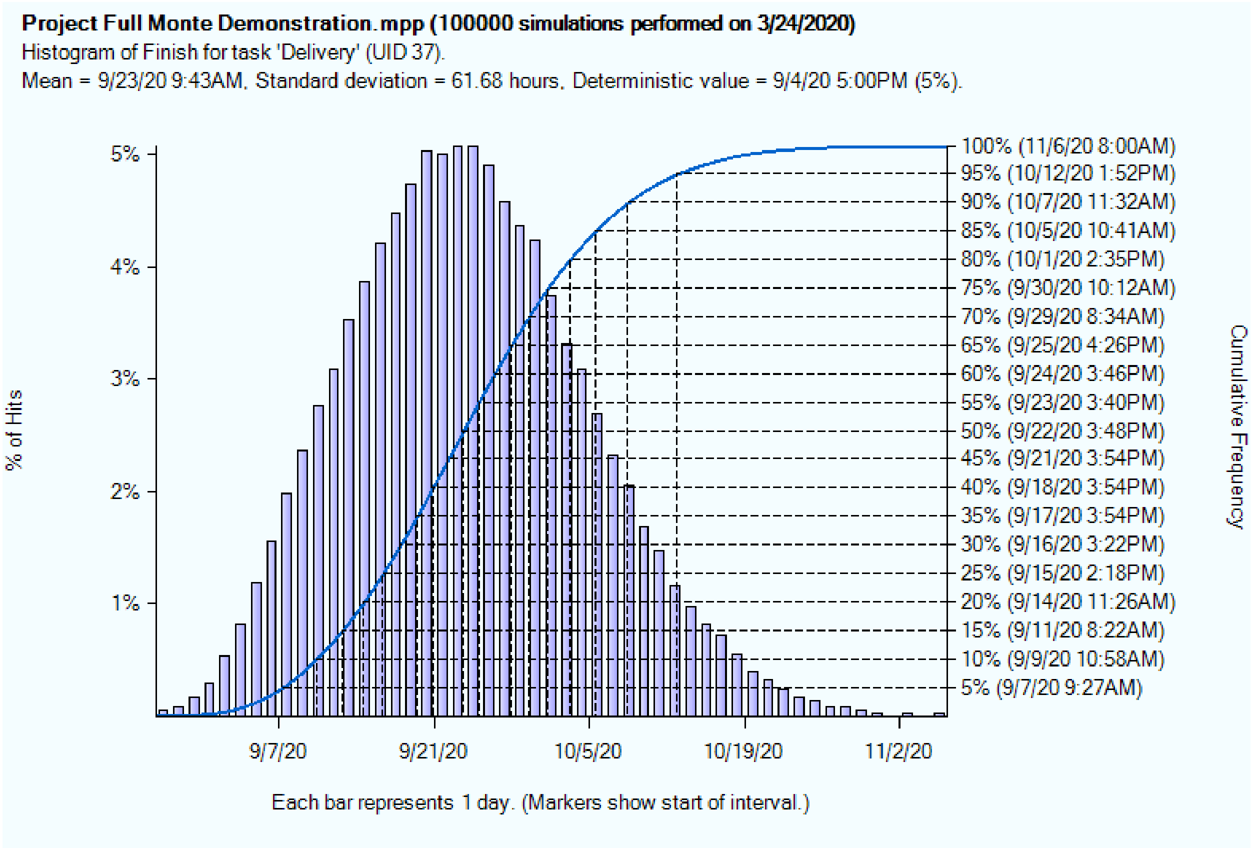
However, we can see significant variation, with some tasks completing in as little as 30% of their estimated duration and some as much as 660% of their estimated duration!

There can be several reasons for this variation:

- Data Entry Error (Actual dates were entered incorrectly)
- Very short duration estimates (a task estimated to take 1 minute took 6 minutes)
- Tasks were impacted by risks that had not been properly identified and mitigated
- It was just a bad estimate

In the example above, the task shown at 660% was a 3-week task that ultimately took nearly 20 weeks. A risk occurred that had not been mitigated.

We can choose to ignore unusual outliers in the data but even so there is significant variation. Based on this data it could be justified to suggest future uncertainty parameters, for similar work, of best-case = 50% of estimate, most likely = 100% of estimate and worst-case of around 180% of estimate. The shape of the distribution tends towards a Beta distribution. Let's apply that to our schedule.



Using uncertainty data based on performance from the completed project has produced some challenging results. The probability of delivery on September 4 is now just 5%. The 80% confidence date is pushed out to October 1.

The key message here is that the uncertainty is based on how we have performed before. It wasn't plucked from thin air. If we haven't learned from our mistakes, be they estimating or engineering issues, and improved our processes going forward then history will repeat itself.

Knowledge is Power

How we use the above information depends on our circumstances. If we have the luxury of telling the customer when we will deliver, then we can commit to October 1 with an 80% chance of delivery by that date. If we want less risk, we could commit to October 7 with a 90% chance of delivery by that date.

But, if we already committed to delivery on September 4, we have some work to do. Can we reduce uncertainty by re-estimating high-risk work? Can we find less risky methods? Can we use better resources to reduce durations? Can we alter the schedule, so the riskier work is not on the critical path?

Ultimately the schedule will need to show a delivery date earlier than September 4 so that when uncertainty is applied the, for example, 80% date becomes September 4 and not October 1.

See Barbecana document '**Sensitivity and Risk Path Analysis**' for more information on achieving required dates.