A Comparison of Correlation in Technical and Programmatic Risk and Cost Risk

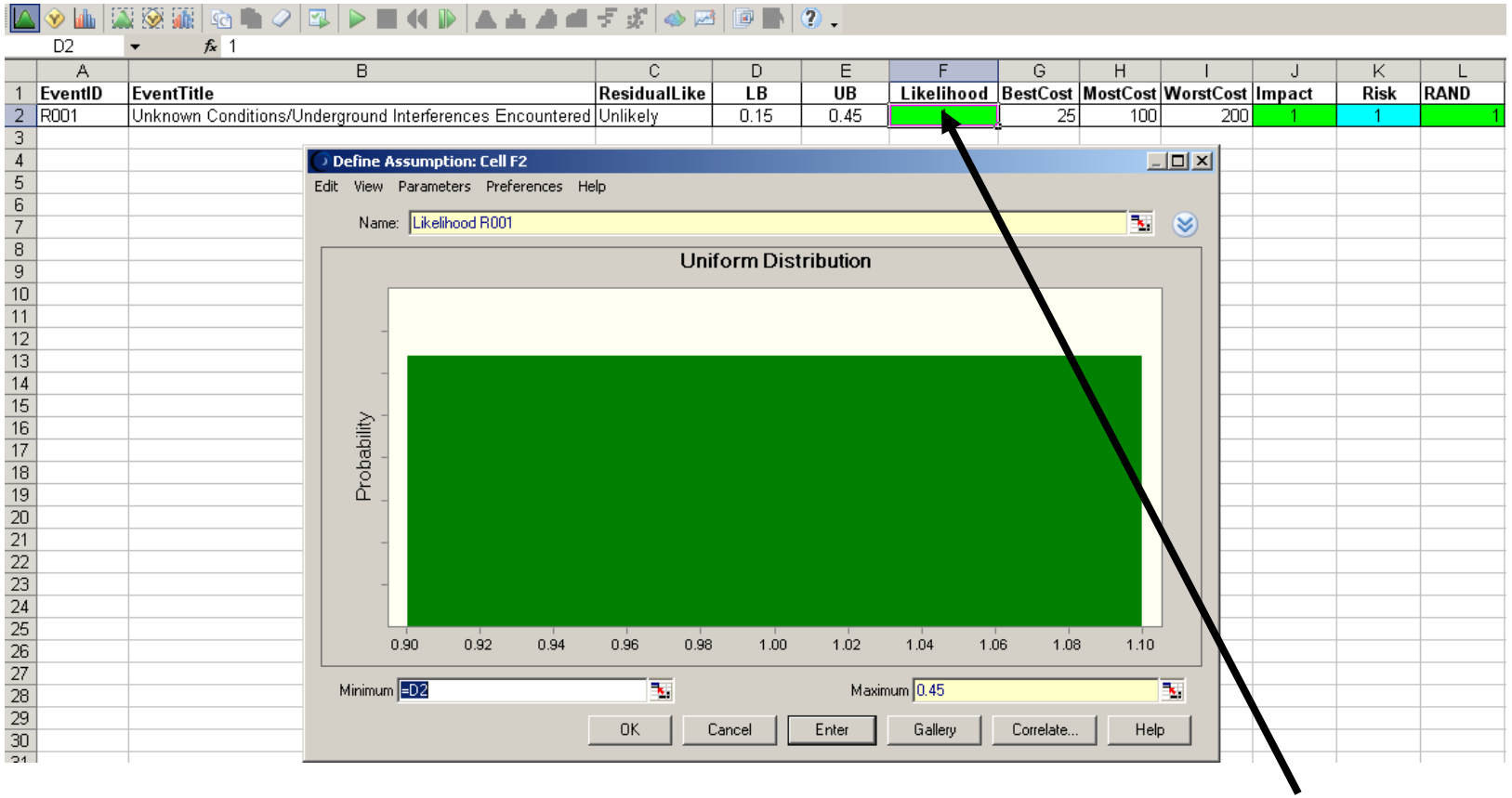
Joseph F. Krupa Ph.D.

Principal Technical Advisor
Savannah River Nuclear Solutions, LLC
June 2011

What was the question to be answered?

- Correlation has long been known to impact Cost Risk
- Cost Risk is the variability in the estimate due to uncertainties in estimate inputs (weather impacts, wage rates, labor productivity, task definition in hours, escalation, etc.)
- There were assertions that correlation had a significant impact on Technical and Programmatic risk
- There were no suggestions on how this was to be modeled, or what to expect.

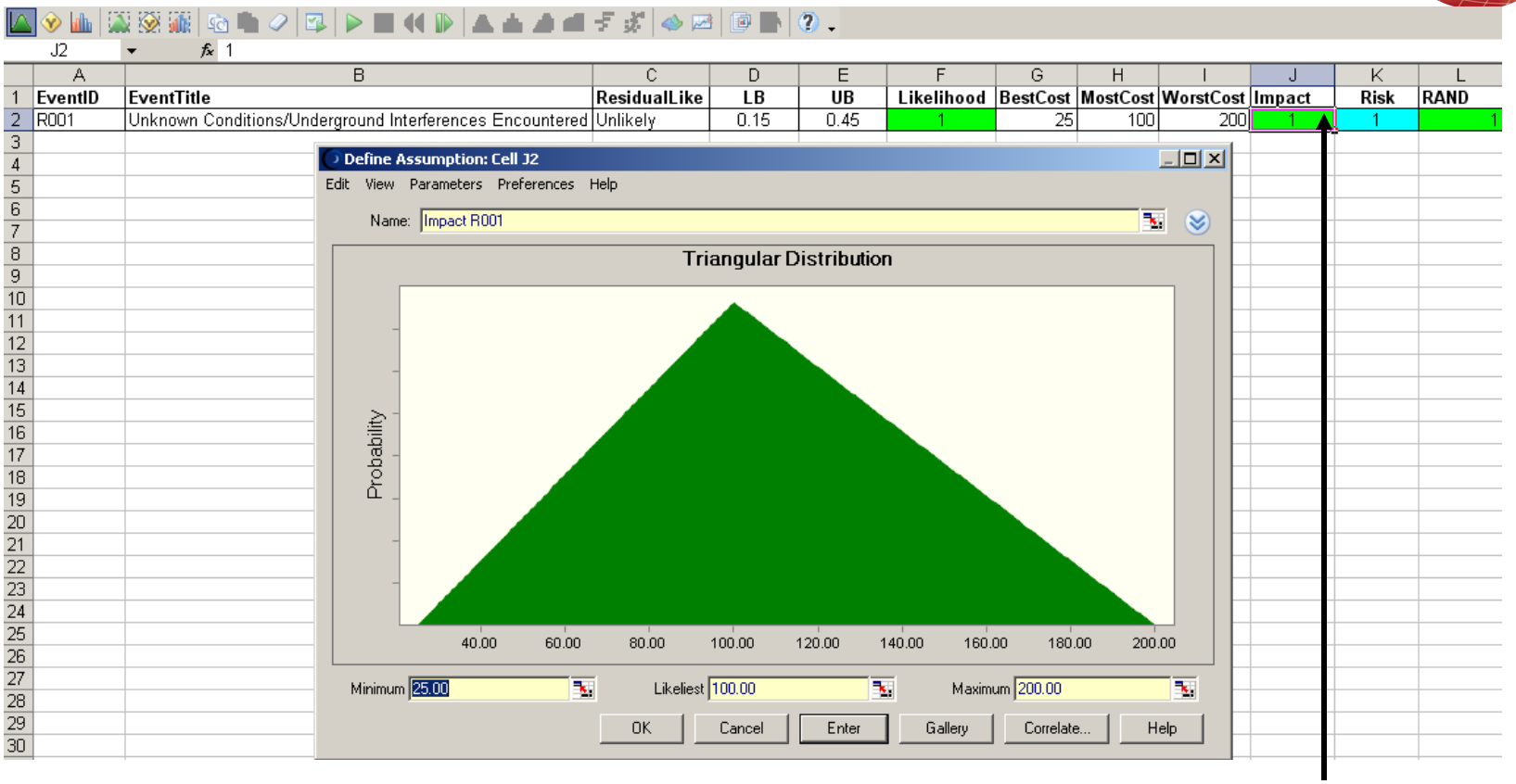
Modeling T&P Risk Using Crystal Ball™



	A	B	C	D	E	F	G	H	I	J	K	L
1	EventID	EventTitle	ResidualLike	LB	UB	Likelihood	BestCost	MostCost	WorstCost	Impact	Risk	RAND
2	R001	Unknown Conditions/Underground Interferences Encountered	Unlikely	0.15	0.45		25	100	200	1	1	1

Risk is determined by probability times impact. Likelihood is a probability distribution for risk occurrence.

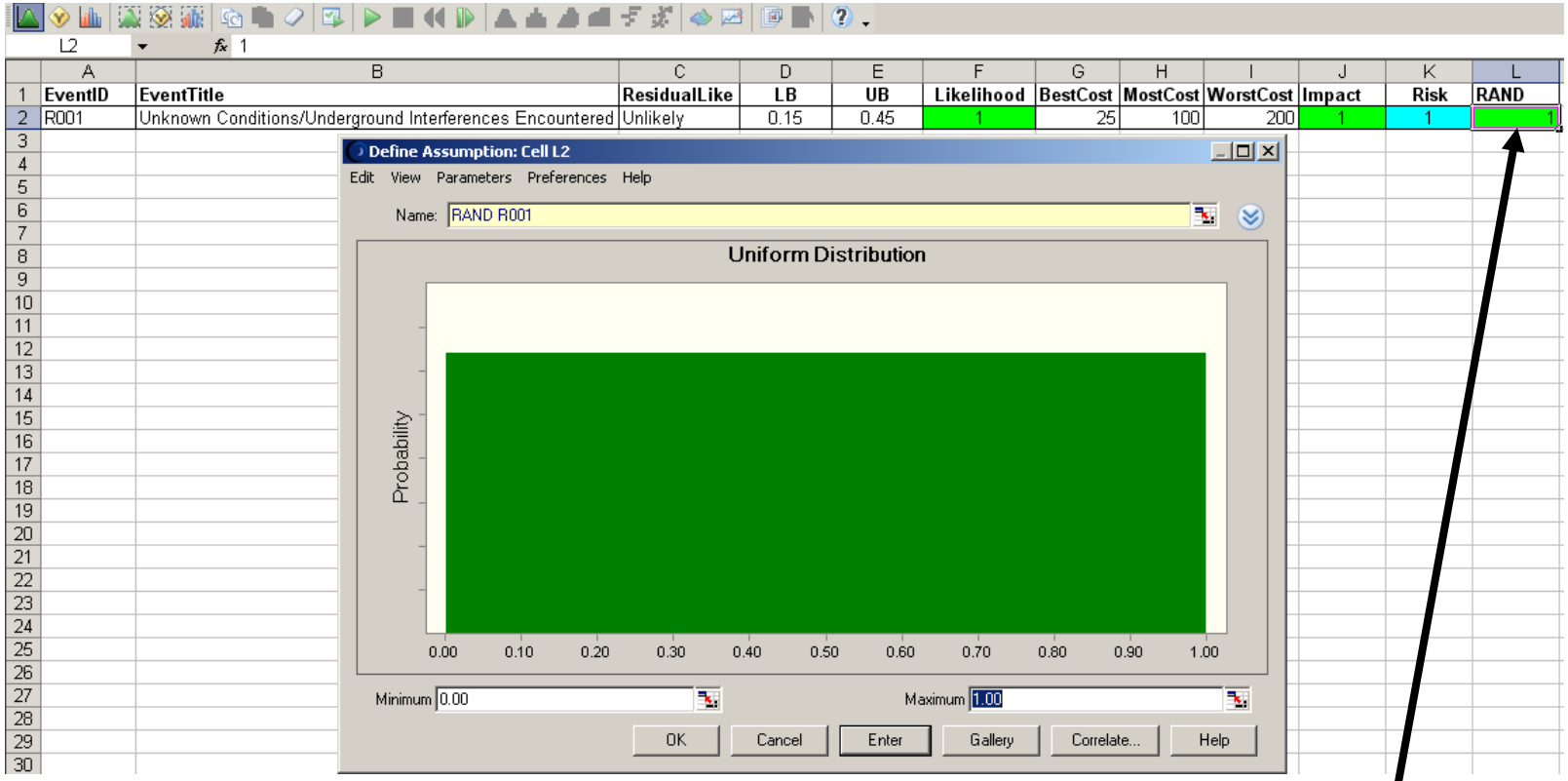
Modeling T&P Risk Using Crystal Ball™ -2



	A	B	C	D	E	F	G	H	I	J	K	L
1	EventID	EventTitle	ResidualLike	LB	UB	Likelihood	BestCost	MostCost	WorstCost	Impact	Risk	RAND
2	R001	Unknown Conditions/Underground Interferences Encountered	Unlikely	0.15	0.45	1	25	100	200	1	1	1

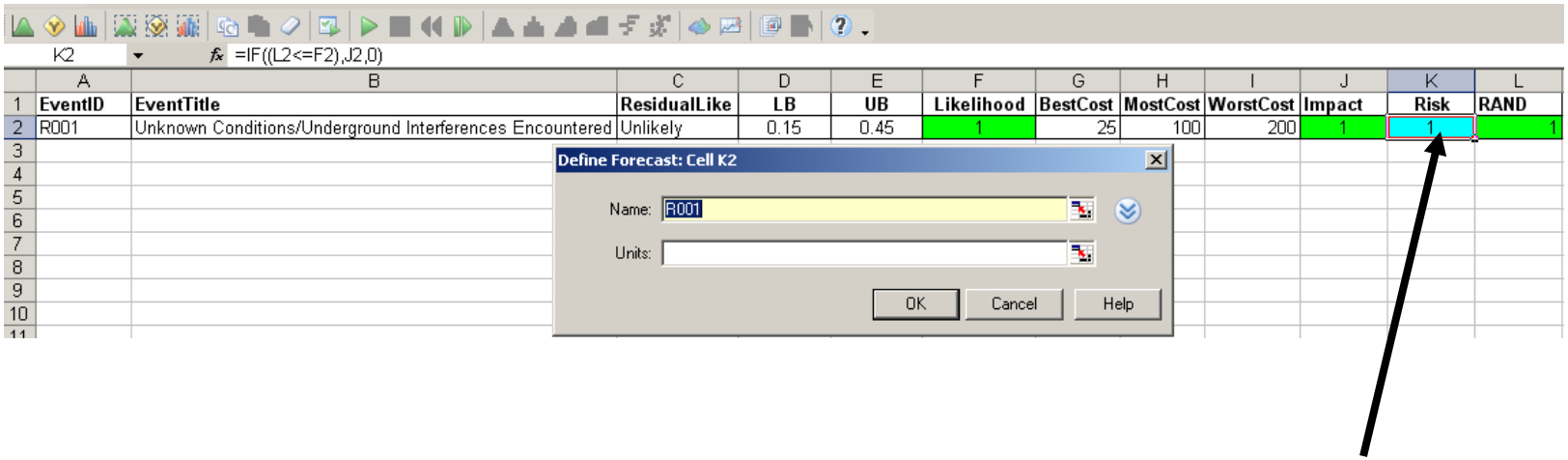
Here the impact is defined as a triangular distribution based on the best, most likely and worst impact if the risk occurs.

Modeling T&P Risk Using Crystal Ball™ -3



This cell sets up a U(0,1) random variable for use in risk occurrence determination. It also enables correlation

Modeling T&P Risk Using Crystal Ball™ -4



	A	B	C	D	E	F	G	H	I	J	K	L
1	EventID	EventTitle	ResidualLike	LB	UB	Likelihood	BestCost	MostCost	WorstCost	Impact	Risk	RAND
2	R001	Unknown Conditions/Underground Interferences Encountered	Unlikely	0.15	0.45	1	25	100	200	1	1	1

Cell K2 shows the defined forecast. If the value of the random variable (L2) is less than or equal to the likelihood variable (F2), then K2 is set to the impact variable J2 for that iteration. If it is not J2 is set to zero (the risk did not occur).

This is the correlated risk model studied

	A	B	C	D	E	F	G	H	I	J	K	L
1	EventID	EventTitle	ResidualLike	LB	UB	Likelihood	BestCost	MostCost	WorstCost	Impact	Risk	Rand
2	R001	Unknown Conditions/Underground Interferences Encountered	Likely	0.45	0.75	1	25	100	200	1	1	1
3	R002	Procurement Delays	Very Unlikely	0	0.15	1	10	10	25	1	1	1
4	R003	Access Road Design/Build Contract	Unlikely	0.15	0.45	1	50	100	250	1	1	1
5	R004	Delayed Design/Bid Specifications	Unlikely	0.15	0.45	1	50	100	200	1	1	1
6	R005	Delayed Design From Communications Infrastructure	Likely	0.45	0.75	1	20	120	240	1	1	1
7	R006	Permit Delays	Unlikely	0.15	0.45	1	25	75	125	1	1	1
8	R007	Delayed Approvals & Support	Unlikely	0.15	0.45	1	75	125	250	1	1	1
9	R008	Construction Activity During Training Exercises	Likely	0.45	0.75	1	20	30	40	1	1	1
10	R009	Access Road Interference Wlth Set Aside Area/Threatened & Endangered Species	Unlikely	0.15	0.45	1	25	50	100	1	1	1
11	R010	Violation of General Storm Water Permit Conditions and Controls	Unlikely	0.15	0.45	1	25	75	150	1	1	1
12	R013	Fire Water Supply	Very Unlikely	0	0.15	1	150	500	1000	1	1	1
13	R014	Rerouted Traffic During Construction	Very Likely	0.75	1	1	0.5	5	10	1	1	1
14	R015	Subcontractors Fail To Perform	Unlikely	0.15	0.45	1	10	25	50	1	1	1
15	Total										13	

Correlated risk groups are shown with the same color (green, yellow or pink). Uncorrelated risks are shown without a color (risks 2, 7, 13 and 15).

How do you set up risks to be correlated?

- Set up random variates for a group to be correlated using the correlation matrix
- Set up the likelihood random variates for a group to be correlated.
- Note that since we are only looking at occurrences being correlated (and not magnitudes), we don't need to set up impact random variates to be correlated

This is the Crystal Ball™ correlation matrix

	R001 Like (Sheet1)	R001 RAND (Sheet1)	R003 Like (Sheet1)	R003 RAND (Sheet1)	R004 Like (Sheet1)	R004 RAND (Sheet1)	R005 Like (Sheet1)	R005 RAND (Sheet1)	R006 Like (Sheet1)	R006 RAND (Sheet1)	R008 Like (Sheet1)	R008 RAND (Sheet1)	R009 Like (Sheet1)	R009 RAND (Sheet1)	R010 Like (Sheet1)	R010 RAND (Sheet1)	R014 Like (Sheet1)	R014 RAND (Sheet1)
R001 Like (Sheet1)	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.600	0.000	0.000	0.000	0.000	0.000	0.600	0.000
R001 RAND (Sheet1)		1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.600	0.000	0.000	0.000	0.000	0.000	0.600
R003 Like (Sheet1)			1.000	0.000	0.600	0.000	0.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R003 RAND (Sheet1)				1.000	0.000	0.600	0.000	0.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R004 Like (Sheet1)					1.000	0.000	0.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R004 RAND (Sheet1)						1.000	0.000	0.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R005 Like (Sheet1)							1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R005 RAND (Sheet1)								1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R006 Like (Sheet1)									1.000	0.000	0.000	0.000	0.600	0.000	0.600	0.000	0.000	0.000
R006 RAND (Sheet1)										1.000	0.000	0.000	0.000	0.600	0.000	0.600	0.000	0.000
R008 Like (Sheet1)											1.000	0.000	0.000	0.000	0.000	0.000	0.600	0.000
R008 RAND (Sheet1)												1.000	0.000	0.000	0.000	0.000	0.000	0.600
R009 Like (Sheet1)													1.000	0.000	0.600	0.000	0.000	0.000
R009 RAND (Sheet1)														1.000	0.000	0.600	0.000	0.000
R010 Like (Sheet1)															1.000	0.000	0.000	0.000
R010 RAND (Sheet1)																1.000	0.000	0.000
R014 Like (Sheet1)																	1.000	0.000
R014 RAND (Sheet1)																		1.000

The correlation used was 0.6.
Uncorrelated variables show 0.0

What correlation looks like for risks 1, 8 &14

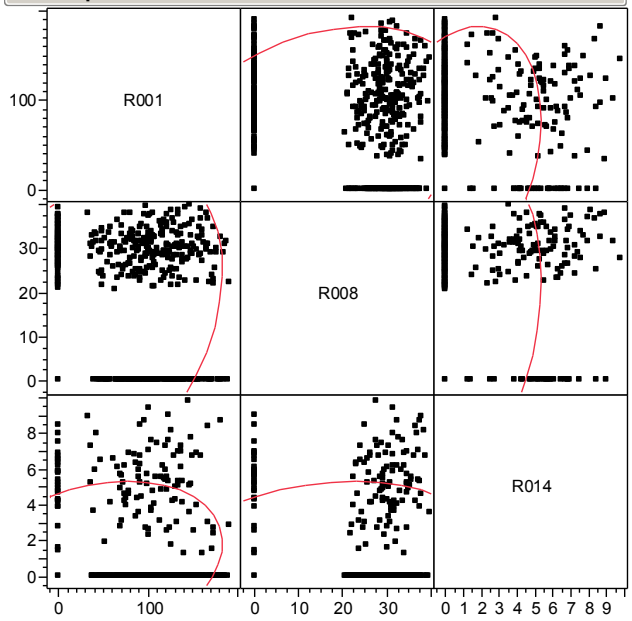
Multivariate

Correlations

	R001	R008	R014
R001	1.0000	0.3496	0.2450
R008	0.3496	1.0000	0.2928
R014	0.2450	0.2928	1.0000

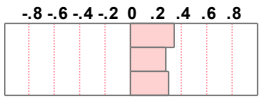
The correlations are estimated by REML method.

Scatterplot Matrix



Nonparametric: Spearman's ?

Variable	by Variable	Spearman ?	Prob>|?|
R008	R001	0.3485	<.0001*
R014	R001	0.2821	<.0001*
R014	R008	0.3024	<.0001*



Note that correlation is <0.6 but all statistically significant. Lower value is due to variability in risk occurrence

What correlation looks like for risks 3,4 & 5

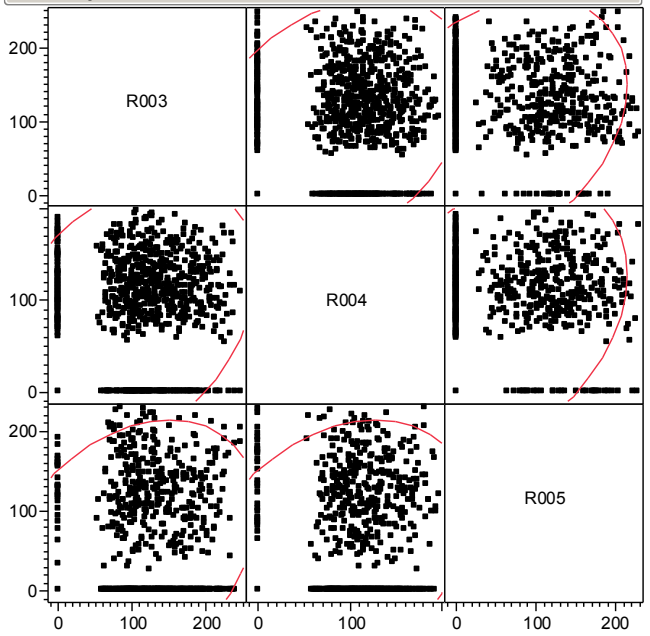
Multivariate

Correlations

	R003	R004	R005
R003	1.0000	0.3443	0.3288
R004	0.3443	1.0000	0.3189
R005	0.3288	0.3189	1.0000

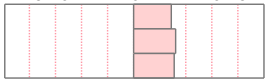
The correlations are estimated by REML method.

Scatterplot Matrix



Nonparametric: Spearman's ?

Variable	by Variable	Spearman ?	Prob>|?|
R004	R003	0.2922	<.0001*
R005	R003	0.3202	<.0001*
R005	R004	0.3087	<.0001*



Again, correlation is <0.6 but significant.

What does correlation look like for risks 6,9 & 10?

Correlation for risks 6, 9 & 10 were calculated by JMP8™ The correlation is smaller than the input correlation because of the non-occurrence of risks in some iterations.

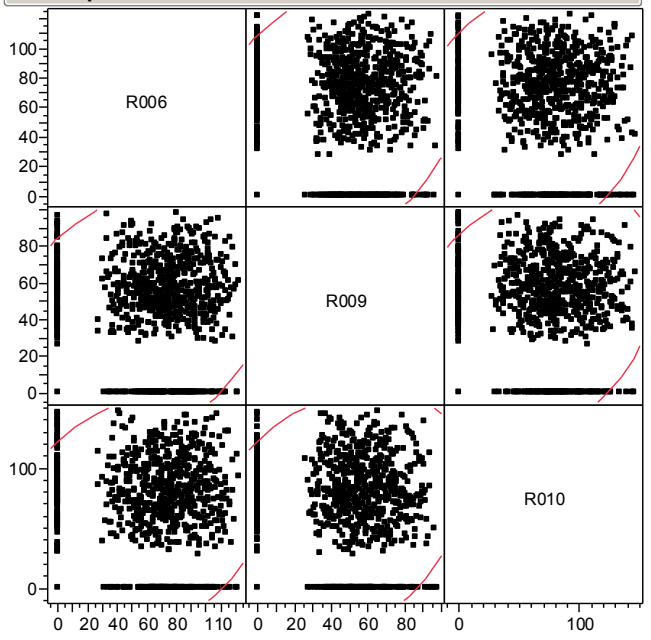
Multivariate

Correlations

	R006	R009	R010
R006	1.0000	0.3332	0.3428
R009	0.3332	1.0000	0.3241
R010	0.3428	0.3241	1.0000

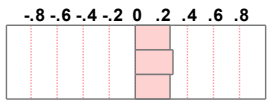
The correlations are estimated by REML method.

Scatterplot Matrix



Nonparametric: Spearman's ?

Variable	by Variable	Spearman ?	Prob>|?|
R009	R006	0.2699	<.0001*
R010	R006	0.2948	<.0001*
R010	R009	0.2681	<.0001*



What does correlation for uncorrelated risks look like?

Risks two and seven have a small negative correlation. All other pairs of uncorrelated risks show no correlation.

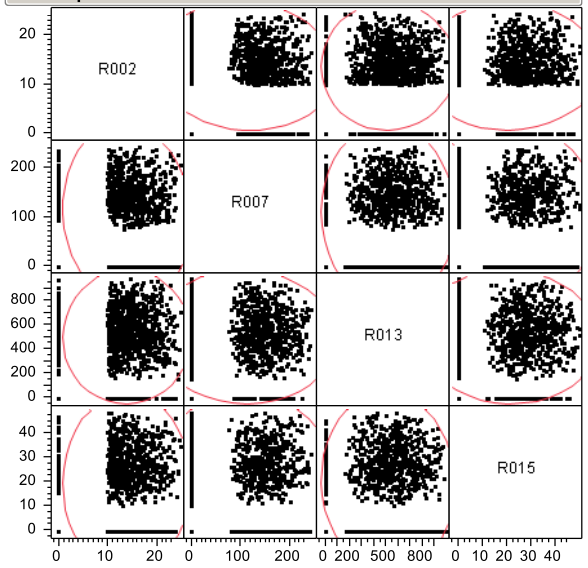
Multivariate

Correlations

	R002	R007	R013	R015
R002	1.0000	-0.0679	0.0047	0.0099
R007	-0.0679	1.0000	-0.0275	-0.0028
R013	0.0047	-0.0275	1.0000	-0.0365
R015	0.0099	-0.0028	-0.0365	1.0000

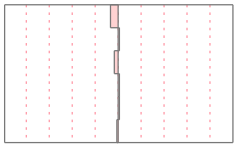
The correlations are estimated by REML method.

Scatterplot Matrix

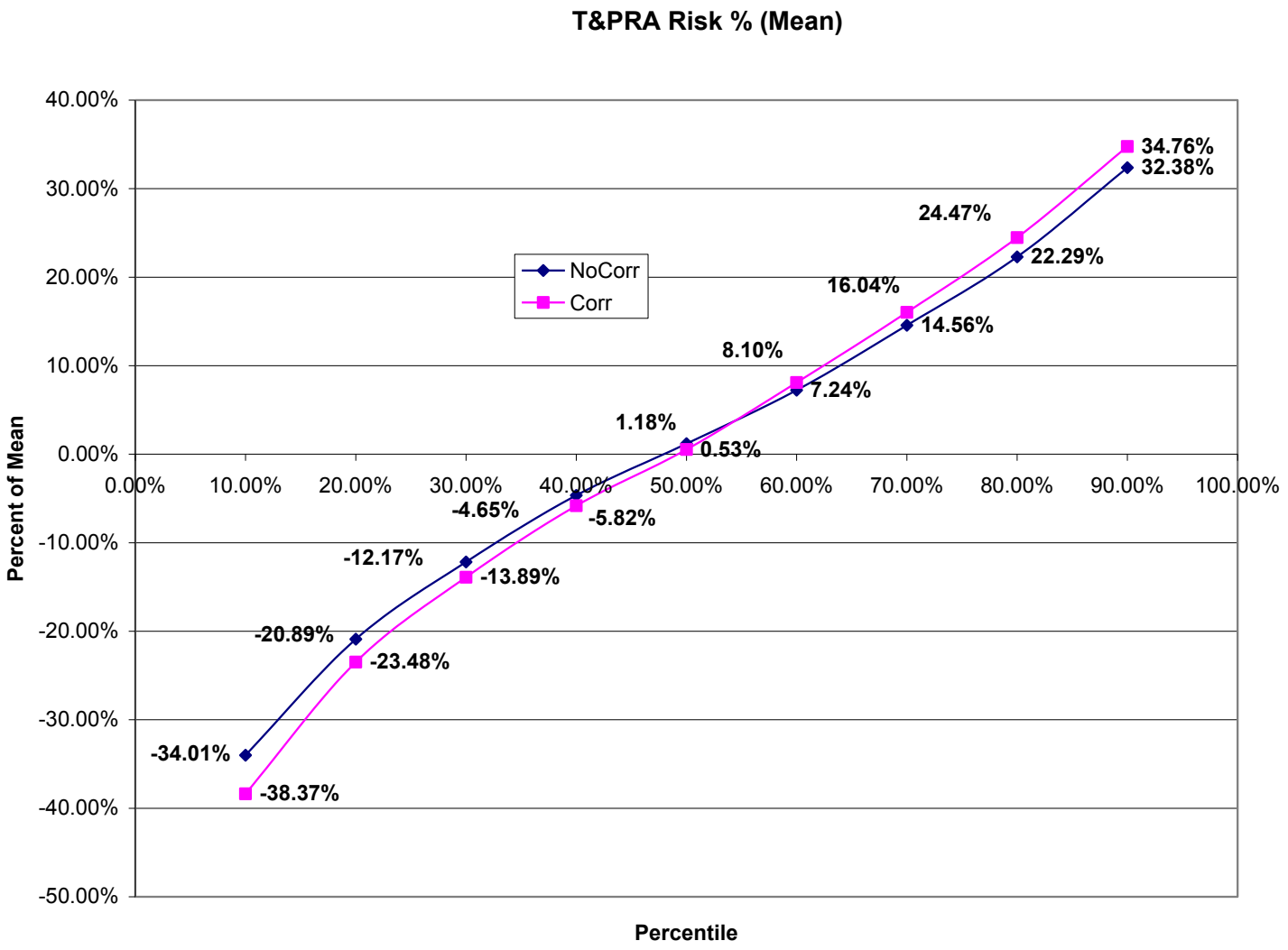


Nonparametric: Spearman's ρ

Variable	by Variable	Spearman ρ	Prob>|ρ|
R007	R002	-0.0668	0.0346*
R013	R002	0.0075	0.8130
R013	R007	-0.0321	0.3103
R015	R002	0.0084	0.7915
R015	R007	0.0080	0.7997
R015	R013	-0.0148	0.6391



What is the impact of correlation on T&P Risk?



What is the difference between correlation in T&P Risk and Cost Risk?

- T&P Risk correlation increases the likelihood that, if an impact value shows up in an iteration for a Risk A that is correlated with Risk B, Risk B is more likely to show up for that iteration.
- In cost risk, correlation factors into every iteration

What is the Cost Risk Model?

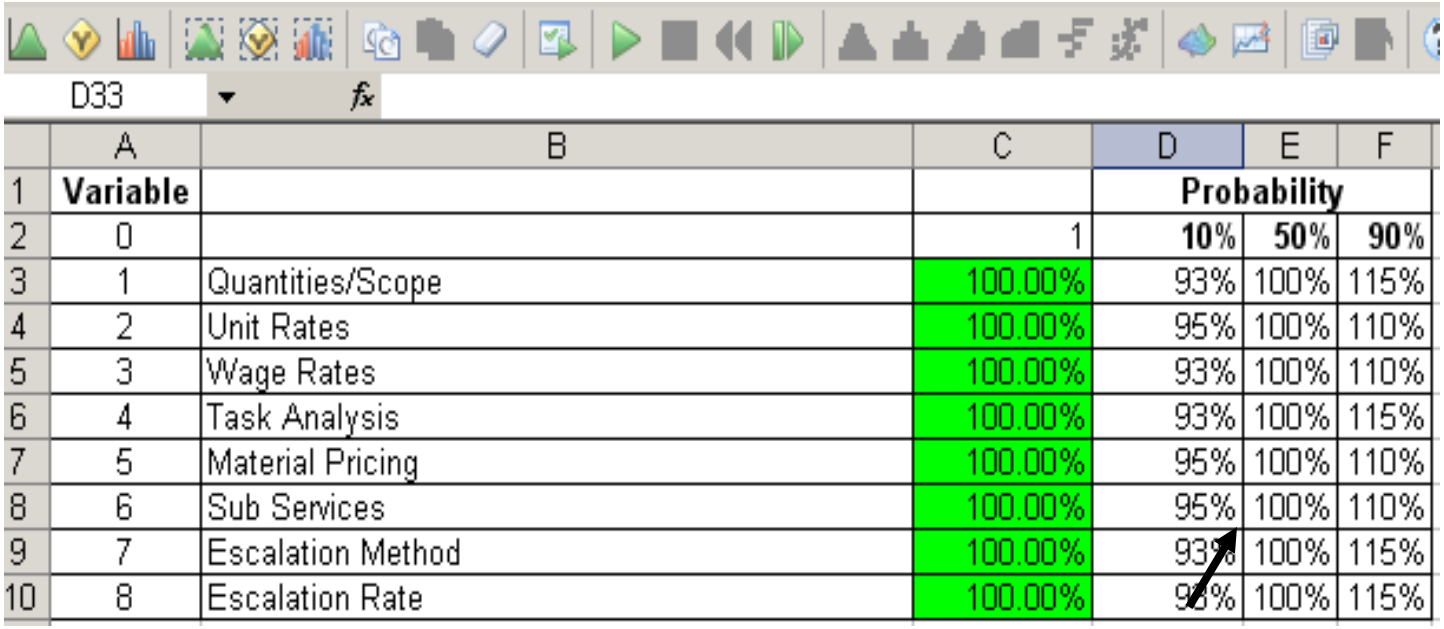
$$TotalCost = \sum_{i=1}^{n} T_i \prod_{j=1}^{m} v_j$$

T = Term (cost of an estimate element)

V = variable (distribution for a factor that influences Term cost)

Direct Cost Estimate is considered to be the median value i.e. v = 1 at the 50th percentile

Sample variable template

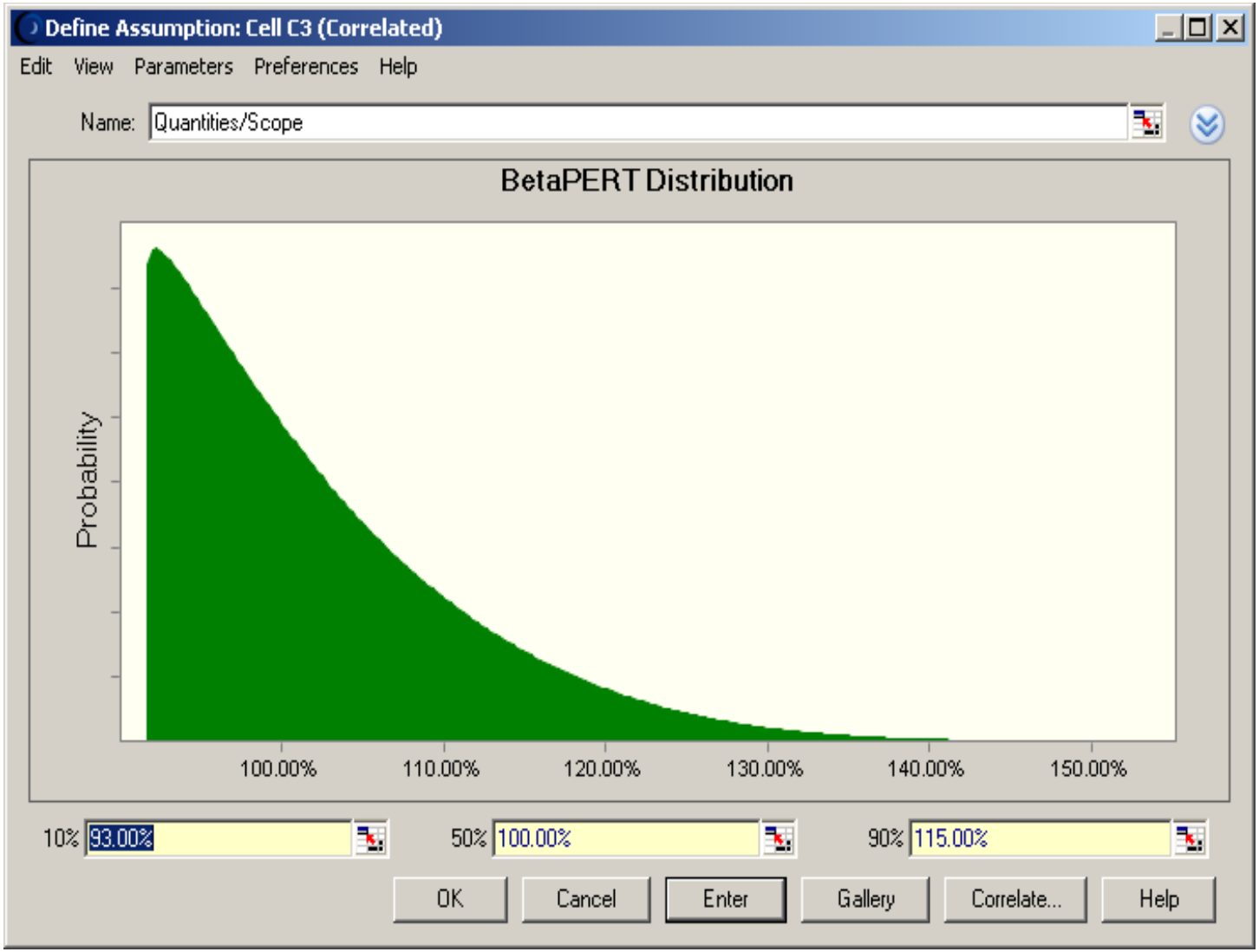


D33

	A	B	C	D	E	F
1	**Variable**			**Probability**		
2	0		1	**10%**	**50%**	**90%**
3	1	Quantities/Scope	100.00%	93%	100%	115%
4	2	Unit Rates	100.00%	95%	100%	110%
5	3	Wage Rates	100.00%	93%	100%	110%
6	4	Task Analysis	100.00%	93%	100%	115%
7	5	Material Pricing	100.00%	95%	100%	110%
8	6	Sub Services	100.00%	95%	100%	110%
9	7	Escalation Method	100.00%	93%	100%	115%
10	8	Escalation Rate	100.00%	93%	100%	115%

Values at 10%, 50% and 90% establish a distribution used as a multiplication factor. Note that it is more likely for costs to increase than decrease.

For this analysis Beta PERT distribution was used



This is the terms template

Estmated Direct Cost Excluding Contingency ($K)	11770
Sum of terms	**11770**
Difference	**0**

			VARIABLES					
TERM	**DESCRIPTION**	**WEIGHT**	**1**	**2**	**3**	**4**	**5**	**Factor**
1	Vegetation Removal	900	1	2	3	4	6	900.00
2	Soil Cover Design -M&O	200	1	2	3	4	5	200.00
3	Soil Cover Construction-Sub	6400	1	2	3	4	6	6400.00
4	Project Support-M&O	70	1	2	3	4	5	70.00
5	Stock Pile Soils -- Labor	2200	1	2	3	4	0	2200.00
6	Stock Pile Soils -- "PECMC"	1400	1	2	5	0	0	1400.00
7	Soil Cover Installaton Support -- Sub	300	1	2	3	4	5	300.00
8	Escalation	300	7	8	6	4	1	300.00
	Total							11770.00

Estimate direct cost is compared to sum of terms to ensure correct values have been entered.

Factor is the product of the weight and the iteration values for the variables selected.

Correlation Matrix for Cost Risk

	Quantities/Scope (Input)	Unit Rates (Input)	Wage Rates (Input)	Task Analysis (Input)	Material Pricing (Input)	Sub Services (Input)	Escalation Rate (Input)
Quantities/Scope (Input)	1.000	0.000	0.000	0.600	0.000	0.000	0.000
Unit Rates (Input)		1.000	0.600	0.000	0.600	0.600	0.600
Wage Rates (Input)			1.000	0.000	0.600	0.600	0.600
Task Analysis (Input)				1.000	0.000	0.000	0.000
Material Pricing (Input)					1.000	0.600	0.600
Sub Services (Input)						1.000	0.600
Escalation Rate (Input)							1.000

What is the final correlation of wage rates and unit rates?

Note that since correlation effects every iteration, final value is close to initial

0.57 ≈ 0.6

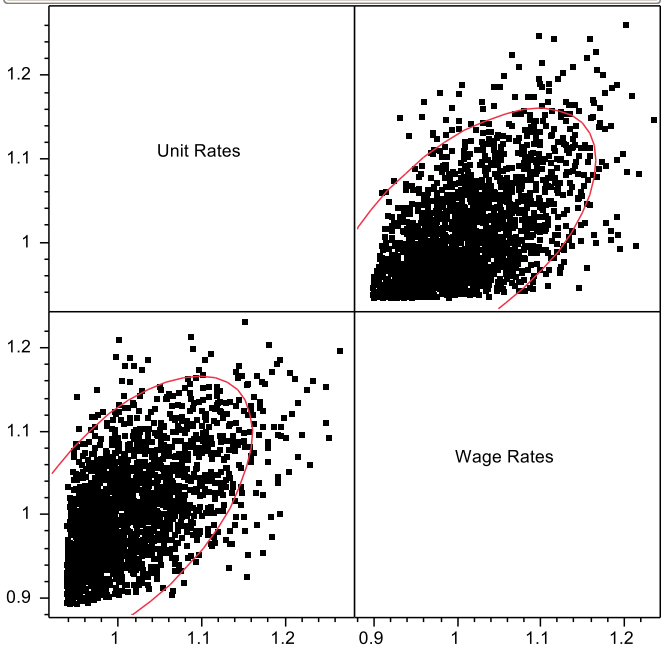
Multivariate

Correlations

	Unit Rates	Wage Rates
Unit Rates	1.0000	0.5741
Wage Rates	0.5741	1.0000

The correlations are estimated by REML method.

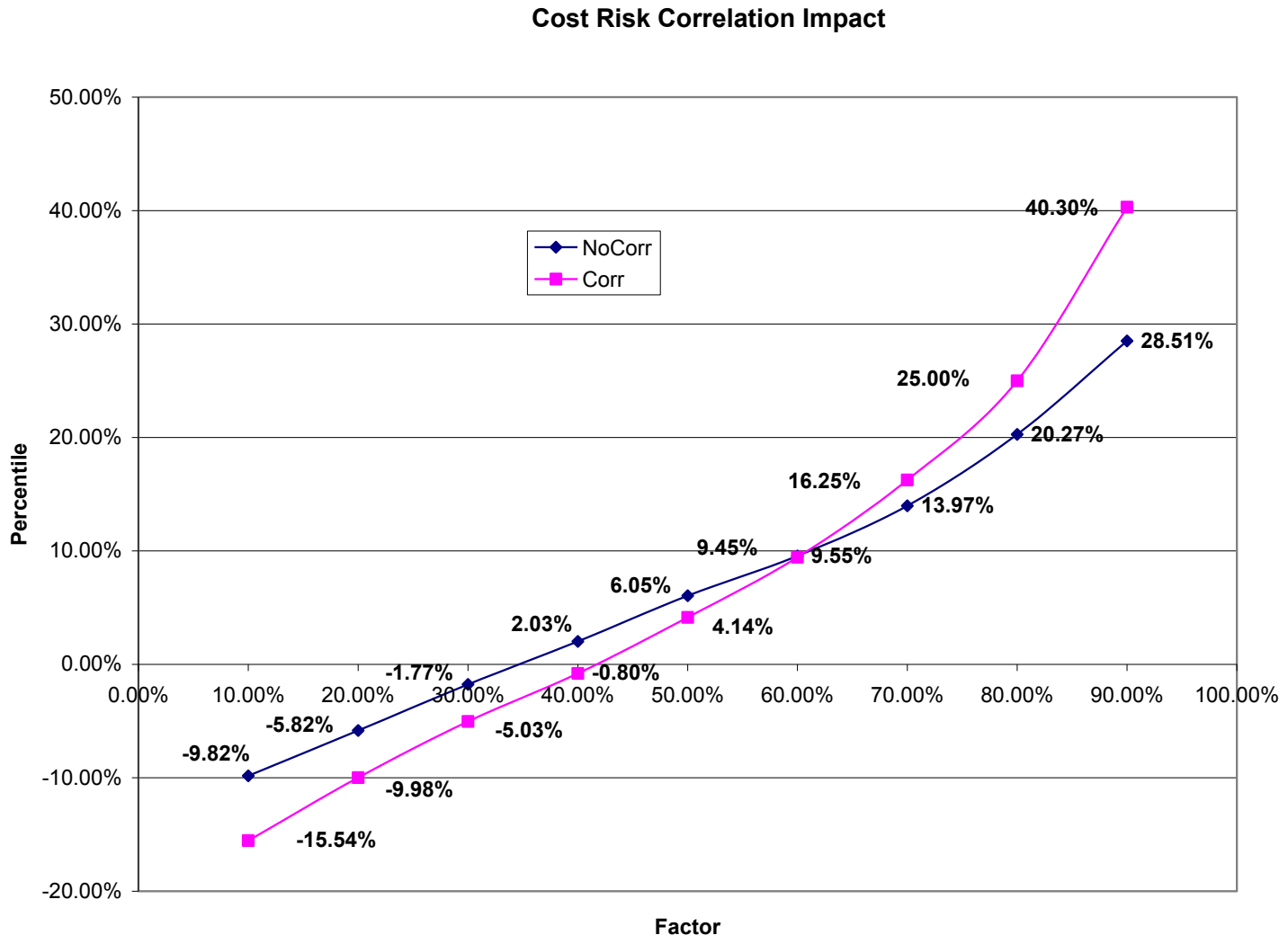
Scatterplot Matrix



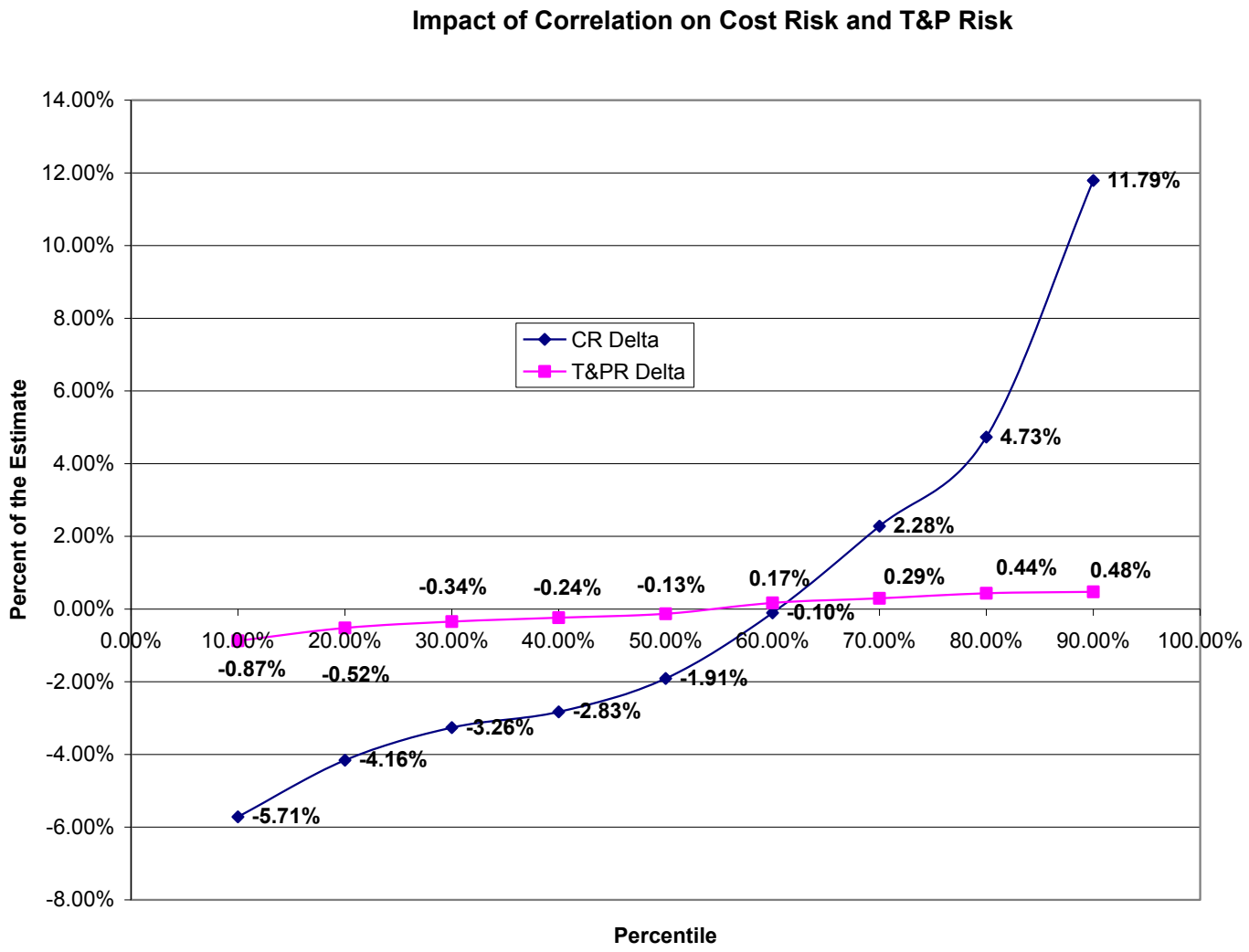
Nonparametric: Spearman's ρ

Variable	by Variable	Spearman ρ	Prob>|ρ|	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Wage Rates	Unit Rates	0.5712	<.0001*	

What is the impact of correlation on cost risk?



What is the impact of correlation?



Conclusions

- Correlation in Cost Risk has a significantly greater impact than correlation in T&P Risk.
- Since only a moderate impact was seen for three sets of correlated T&P Risk, it is not likely that T&P correlation will have a significant impact on T&P risk

Questions ?