Making Statistical Analysis Accessible:

The RAMS Tool for Regression & Risk Analysis

Matt Lytton, Eric Druker, Greg Hogan

2009 ISPA/SCEA Professional Development and Training Workshop
St. Louis

Contents

Why RAMS?: Putting Tools in the Hands of the Analyst

- Although the scarcity and quality of data remains the largest barrier in parametric estimating, it is also true that many analysts do not pursue every available option when developing CERs
  - Many of the statistical techniques needed for CER development require advanced degrees in mathematics, which are not always present among cost estimators
- Additionally, despite their being statistical methods for quantifying risk and uncertainty around Cost Estimating Relationship (CER)-based estimates, many analysts still use the standard error of the estimate, or even worse, subjective measures to evaluate uncertainty
  - Almost all non-statistical methods for quantifying risk and uncertainty underestimate both
- Although there are many tools in the marketplace that perform this type of analysis, they are either not always available, or not always useful to analysts in the field
  - Not every contract wants to pay for commercial tools that may or may not help in finding CERs – especially when there's always the "engineering judgment" as backup
  - The commercial tools contain so many features that beginner analysts may have problems sorting through them all to find the correct method

Why RAMS?: Putting Tools in the Hands of the Analyst

- To help our consultants apply analytical techniques in their cost estimates and risk analysis, Booz Allen Hamilton developed the Regression & Risk Analysis Methodology Streamliner (RAMS) tool

- The Goals of RAMS
  - Provide analysts a tool they can use to perform *the most common* types of regression & risk analysis and allow the tool to interact with the most commonly used COTS risk tools
  - Develop the tool internally, so that all of Booz Allen's consultants have access to it
  - Maximize the application of statistical methodologies in cost estimating & risk analysis
  - Include error-checking capabilities to minimize mistakes in the analysis

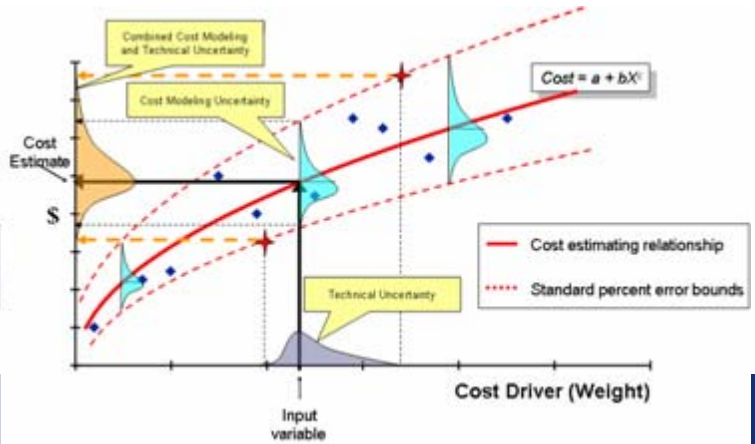
RAMS is About Putting Tools in the Hands of the Analyst

The RAMS Tool: Features

- Once the user enters their data, RAMS automatically performs the following regression techniques
- Ordinary Least Squares – statistics are updated in real time as the user enters data
  - Linear
  - Log-Linear
  - Multilinear
  - Log-Linear with intercept ($ax^b + c$, $ae^{bx} + c$)
- Generalized Least Squares
  - Minimum Percent Error
  - Minimum Unbiased Percent Error (in development)
- RAMS also includes error-checking capabilities by checking to ensure that regressions are significant, have homoscedastic residuals and that there is no multicollinearity between the independent variables

The RAMS Tool: Features

- Once a CER has been developed, the RAMS tool automatically converts the associated estimate into a risk distribution
  - For OLS regressions, either a t or log-t distribution based on the prediction intervals around the estimate is used to characterize cost risk[1]
  - For GLS regressions the tool will use the bootstrap method to develop the prediction intervals and will then convert them into risk distributions[2]
- This distribution can then be used in either an inputs-based or an outputs-based simulation
- The RAMS tool allows cost drivers to be defined as distributions and includes interfaces for COTS risk tools (such as Crystal Ball) to allow for easy integration

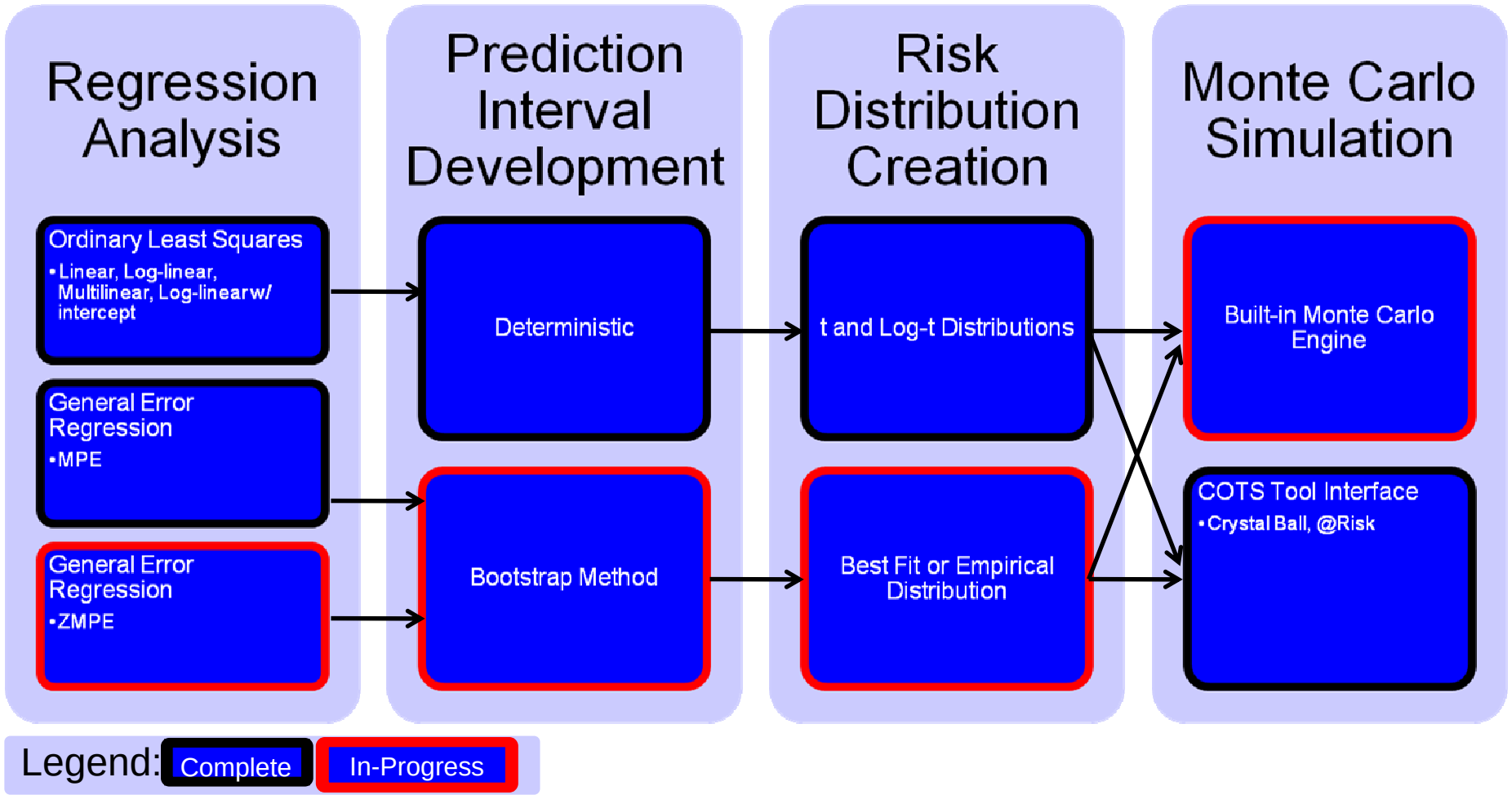


RAMS Automates CER Development From Regression to the Development of the S-Curve

[1] *Taking the Next Step: Turning OLS CER-Based Estimates into Risk Distributions*. Kanick, Christina; Druker, Eric. 2008 SCEA Conference. June 2008. Los Angelos

[2] *Prediction Bounds for General-Error-Regression CERs*. Book, Steve. DoDCAS. February 2006. Williamsburg

RAMS Tool: Features



Using RAMS

1. The analyst enters their bivariate or multivariate data on the front sheet
   - Statistics for all 4 OLS models (linear, power, exponential, logarithmic) are updated as the user enters their data
   - A scatter plot of the data is automatically generated

2. The analyst selects a regression technique and clicks a button to develop the CER
   - The RAMS tool performs the regression analysis, finds the best fit-CER and checks to ensure there are no errors in the regression

3. The analyst clicks a button to develop a risk adjusted estimate
   - The RAMS tool finds the distribution around the point estimate given either a value or distribution for the cost driver(s)
   - Selected confidence levels are also displayed on the S-curve
   - A Crystal Ball interface is included so the analyst can run a Monte Carlo simulation

- All of the worksheets developed by RAMS can be copied into external cost models with full functionality

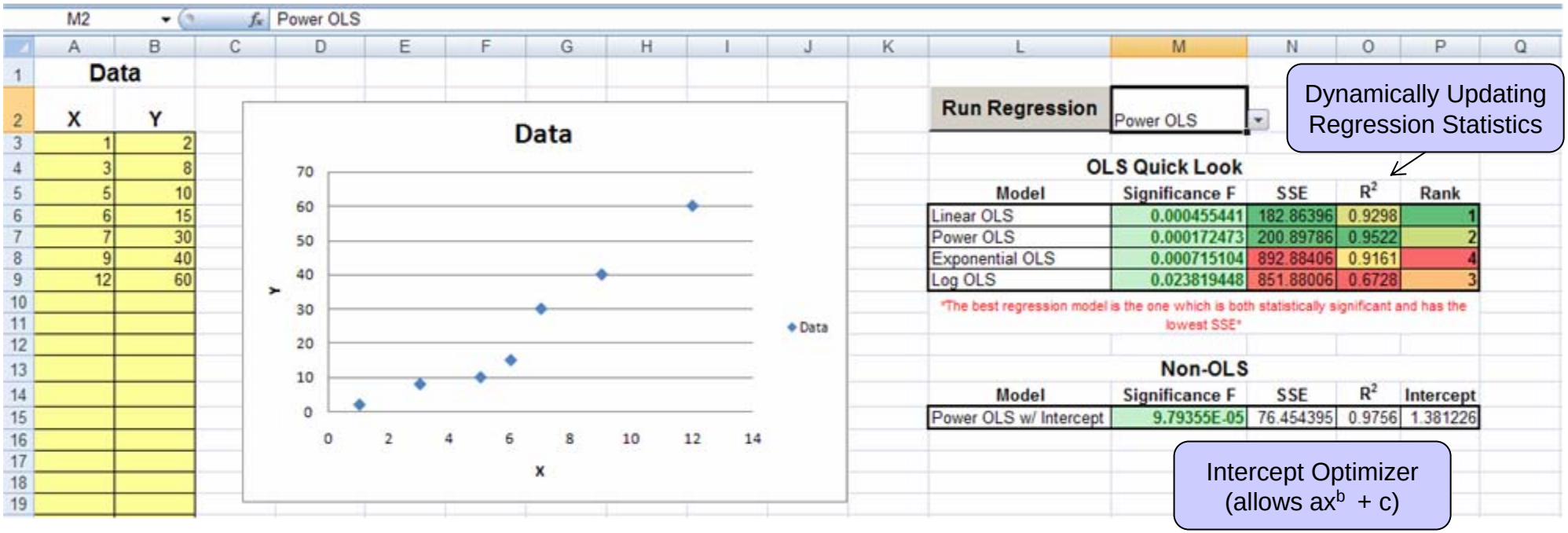
Conclusions

- RAMS has been used on several estimating/risk analysis efforts and has proven a valuable assett on each
  - It minimizes the time needed to develop statistically based CERs and associated risk analysis
- Because it only contains the features needed to develop CERs and perform risk analysis, it has an extremely steep learning curve
- RAMS is currently included in the firm's regression and risk analysis training
  - As analysts learn the techniques they are provided with a tool to help them perform the analysis
- Now, a short demonstration of the RAMS tool

RAMS Demo

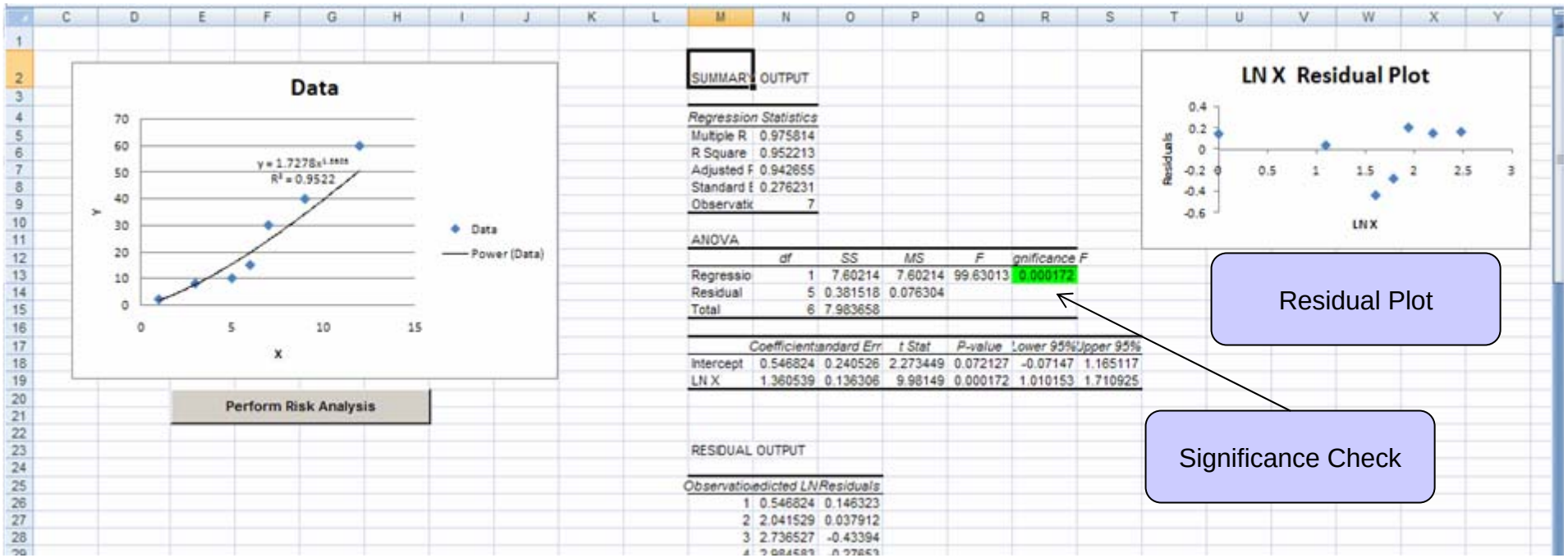
Screenshots

Screenshot – Bivariate Data Entry



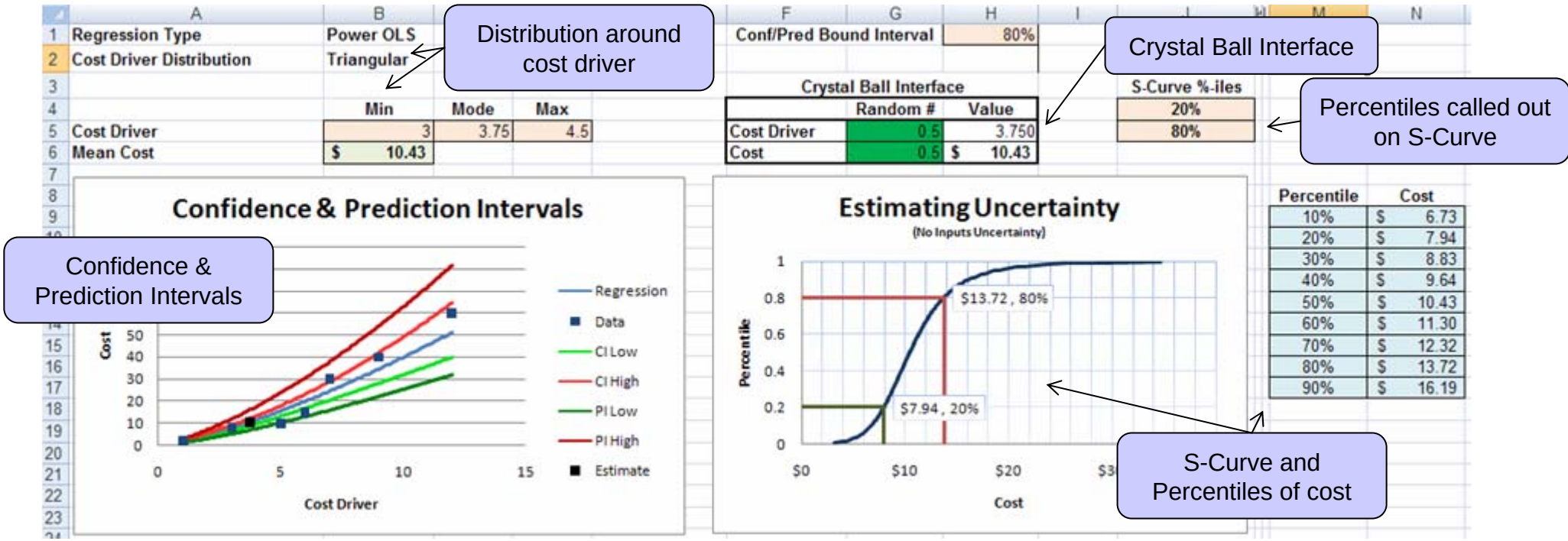
- Once data is entered, tool automatically performs “Quick Look” analysis, determining the best fit among the standard OLS models (Linear, Power, Exponential, Log)
- The desired model is selected from a drop down menu and the regression is automatically performed with the click of a button

Screenshot – Bivariate Regression



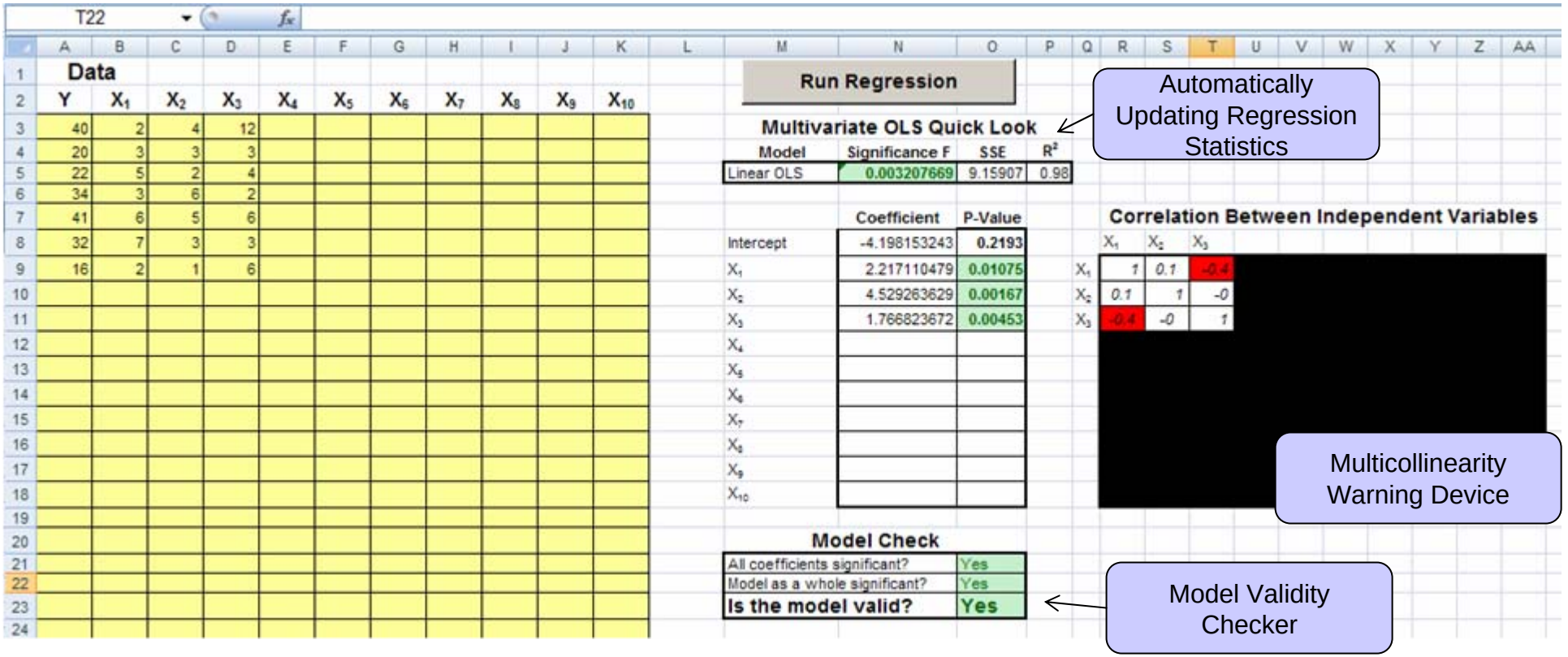
- RAMS automatically charts the best-fit line, plots the residuals, and determines if the regression is significant
- With the click of a button, RAMS then uses the ANOVA statistics to perform risk analysis

Screenshot – Bivariate Risk Analysis



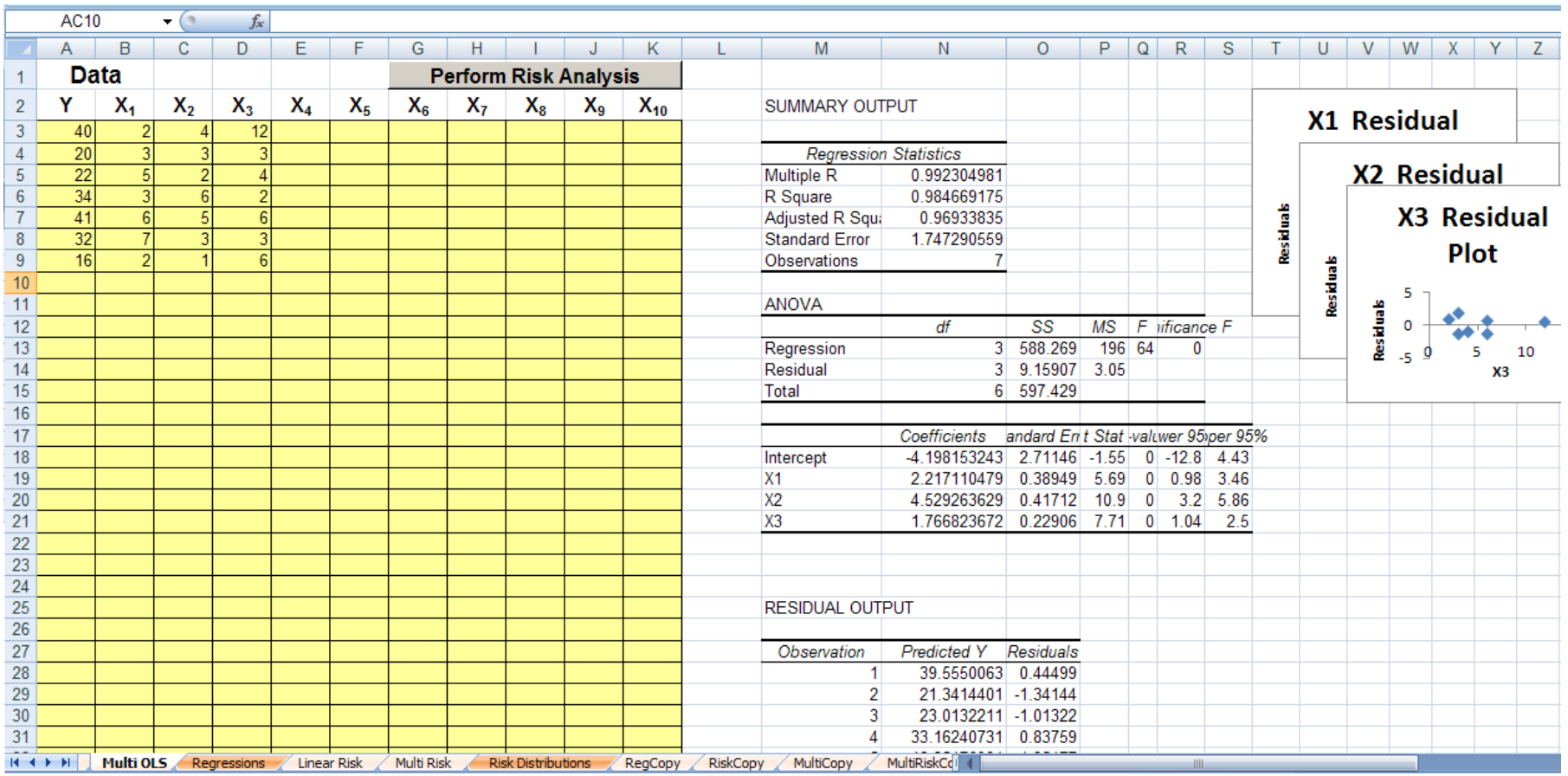
- Above is a sample risk analysis performed on a CER-based estimate

Screenshot – Multivariate Data Entry



- RAMS also accepts multivariate data
- As with bivariate OLS regression, statistics are updated in real time with a correlation matrix for the X variables added to warn when multicollinearity could arise

Screenshot: Multivariate Regression



AC10

Data						Perform Risk Analysis				
Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
40	2	4	12							
20	3	3	3							
22	5	2	4							
34	3	6	2							
41	6	5	6							
32	7	3	3							
16	2	1	6							

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.992304981
R Square	0.984669175
Adjusted R Squ	0.96933835
Standard Error	1.747290559
Observations	7

ANOVA

	df	SS	MS	F	ificance F
Regression	3	588.269	196	64	0
Residual	3	9.15907	3.05		
Total	6	597.429			

	Coefficients	andard Er	t Stat	-valuwer 95	per 95%
Intercept	-4.198153243	2.71146	-1.55	0	-12.8	4.43
X1	2.217110479	0.38949	5.69	0	0.98	3.46
X2	4.529263629	0.41712	10.9	0	3.2	5.86
X3	1.766823672	0.22906	7.71	0	1.04	2.5

RESIDUAL OUTPUT

Observation	Predicted Y	Residuals
1	39.5550063	0.44499
2	21.3414401	-1.34144
3	23.0132211	-1.01322
4	33.16240731	0.83759

Multi OLS | Regressions | Linear Risk | Multi Risk | Risk Distributions | RegCopy | RiskCopy | MultiCopy | MultiRiskCo

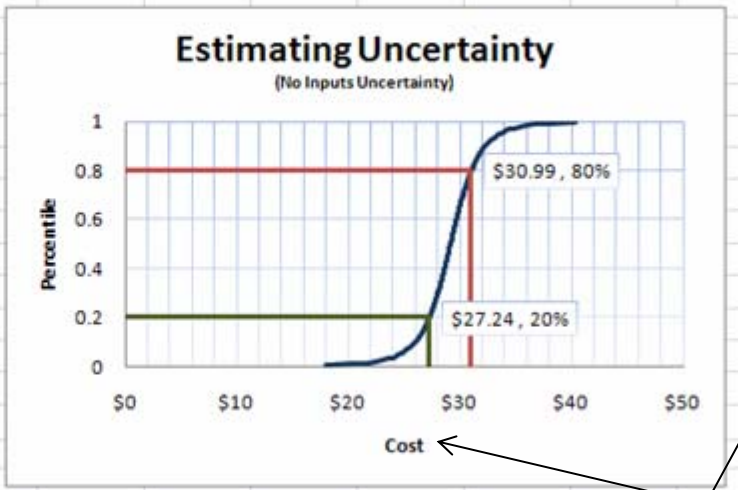
Screenshot – Multivariate Risk Analysis

S3

	A	B	C	D	E	F	G
1	**Cost Drivers**					**Crystal Ball Interface**	
2		**Distribution**	**A**	**B**	**C**	**Random #**	**Value**
3	**Intercept**						
4	X_1	Triangular	2	4	5	0.49969605	3.731524278
5	X_2	Normal	4	2		0.93072358	6.96240341
6	X_3	Point	4				4
7	X_4						
8	X_5						
9	X_6						
10	X_7						
11	X_8						
12	X_9						
13	X_{10}						
14	**Cost**					0.5	$ 42.68

Definition of Parameters:
Point - A: Point Estimate
Normal: A - Mean, B - Standard Deviation
Triangular: A - Min, B - Mode, C - Max

Crystal Ball Interface



Percentile	Cost
10%	$ 25.98
20%	$ 27.24
30%	$ 28.00
40%	$ 28.59
50%	$ 29.12
60%	$ 29.65
70%	$ 30.23
80%	$ 30.99
90%	$ 32.25

S-Curve %-iles
20%
80%

S-Curve and Percentiles of cost

Contacts

- For more information on the RAMS tool, or to become a beta-tester, please contact:
  - Eric Druker
    - Druker_eric@bah.com
    - (314) 368-5850
  - Greg Hogan
    - Hogan_gregory@bah.com
    - (703) 609-9134
  - Matt Lytton
    - Lytton_matthew@bah.com
    - (540) 798-6590