Early Warning Model for Acquisition Program Cost and Schedule Growth

15 April 2008
Dan Davis
Agenda

• Background
• Approach
• Results
• Way ahead
Background

• Prior related studies
  – The Rayleigh Analyzer, Theory and Application (Vol. 1)/ AT902C1 (Ducovich, Houser, and Lee)
    • LMI study for OSD
  – The Rayleigh Analyzer, User’s Manual (Vol. 2)/AT902C2 (Ducovich, Houser, and Lee)
    • User’s guide for Excel Add-on by LMI
Background, cont’d

- Prior related studies (cont.)
Early Warning Model

\[ c(t) = d \left( 1 - e^{-\alpha t^2} \right) \]

- Cumulative cost as a function of time (in millions of $)
- Scale factor. How much will the program cost?
- Proportion of work completed at time \( t \).
- Shape parameter. When will peak spending rate occur?
- Duration time (in years). What is the current duration of the program?
Early Warning Model, cont’d

The parameter $d$ tells us the height of the curve. What is the upper bound on cumulative cost?

The parameter $\alpha$ tells us the shape of the curve. When does the peak spending rate occur?

This curve inflects when the rate curve below reaches a maximum.
Early Warning Model

Reason for Project

- Previous study (CRM D0015902.A2/final):
  - Developed prototype analytical technique
  - Validated early warning utility of technique
  - Developed prototype executive management tool
    - Contract assessment module
    - Plan assessment module
    - Encorporated preliminary business insights
  - Demonstrated practicality of prototype
  - Validated for R&D contracts only
  - Used information sets that were available only early in the contract life (the 3rd, 4th, and 5th submissions)
  - Did not calculate standard errors or risk regions
Early Warning Model

Reason for Project, cont’d

- Sponsor liked prototype and wanted to add:
  - Criteria for when to use Rayleigh model
  - Validation using full range of information sets over entire life of contracts
  - Improved data loading instructions
  - Version 2 algorithms
  - Exploration of expanding model to:
    - Procurement contracts
    - Program level assessments
  - Expanded and improved business insight feature
  - Improved user interface
Study tasks:
- Collect data
- Update optimization algorithm
- Account for the "missing 30 percent" of cost growth
  - Calculate confidence region
  - Calculate contract cost and schedule risk
  - Develop Rayleigh spline option
    - Incorporate Over-target Baseline
- Revalidate using information sets over entire life of contract
- Update plan assessment module
- Improve data upload
- Explore use with procurement contracts
- Explore use with program level analysis
- Improve user interface
- Improve and expand business insights
Early Warning Model

Approach

- Update model algorithm ✓
- Confirm that Non-linear Least Squares with restricted parameters (NRP) is best implementation of model ✓
- Revalidate with information sets over full life of contract ✓
  - Confirm Rayleigh is best fit ✓
  - Confirm Rayleigh predictions are most accurate ✓
  - Confirm use of unadjusted data is better or as good as analysis with inflation-adjusted data ✓
  - Confirm model is still best even as “business practices” evolve ✓
  - Confirm Rayleigh provides reliable early warning ✓
    - Confirm trend of Rayleigh predictions converges ✓
    - Confirm trend of Rayleigh predictions converges the quickest ✓
    - Confirm trend of Rayleigh predictions is stable (non-volatile) ✓
  - Confirm Rayleigh applicability does not depend on service type ✓
Early Warning Model

Approach, cont’d

• Incorporate contract-level cost and schedule risk ✓
  – Calculate approximate standard errors and confidence regions ✓
  – Calculate cost risk ✓
  – Calculate schedule risk ✓
• Investigate use of the model with program level budget data ✓
• Test applicability of model to procurement contracts ✓
• Upgrade user interface
• Develop Rayleigh spline package for tool to account for “missing 30%” ✓
• Upgrade “business insight” prompts in tool ✓
• Explicitly link model to AOP-like management process
Summary of results

• Database consisted of 107 completed contracts
• In 100% of the contracts the NRP method outperformed unrestricted NLLS
  – NRP less volatile (more stable)
  – NRP converged faster
  – NRP gave better predictions
Summary of results, cont’d

• Rayleigh (NRP) yields an R2:
  – Greater or equal to .9 in 93 contracts (87%)
  – Between .8 and .9 in 9 contracts (8%)
  – Between .7 and .8 in 4 contracts (4%)
  – Less than .6 in 1 contract (1%)

This means that Rayleigh NRP achieves a good fit to the data in all but 1 case out of 107.
Summary of results, cont’d

- Database included 36 Navy, 36 AF, and 35 Army contracts
- Navy average R2 = .947
- AF average R2 = .951
- Army average R2 = .941
- Standard t-test shows that no service average differs significantly from overall average or R2 = .947
Summary of results, cont’d

Model: \( R^2 = a + b \times \text{WSDATEIndex} \)

<table>
<thead>
<tr>
<th>a_hat</th>
<th>b_hat</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.952</td>
<td>-0.00057</td>
</tr>
<tr>
<td>(.0277)</td>
<td>(.001443)</td>
</tr>
</tbody>
</table>

Estimate of \( a \) is not significantly different for the average \( R^2 \) and the estimate of the slope term is not significantly different from 0.
Summary of results, cont’d

• 1970s average R2 is .96
• 1980s average R2 is .94
• 1990s average R2 is .93
• 2000s average R2 is .95
• Overall average R2 is .95
Summary of results, cont’d

• Rayleigh (NRP) average R2 is .947
• Rayleigh with nominal data is better or almost as good as Rayleigh with real inflation-adjusted data 102 out of 107 cases (95%)
  – And Rayleigh with nominal data has higher R2 (.947 > .933)
Summary of results, cont’d

• Rayleigh (NRP) is best EAC generator over the life of a contract 70% of the time
  – Rayleigh NRP is best or second best 82% of the time (88 out of 107)
• EAC1 is the best 2% of the time
• EAC2 is the best 4% of the time
• EAC3 is the best 13% of the time
• The contractor is the best 4.5% of the time
• The PM is the best 6.5% of the time

The contractor or the PM estimate profile is the worst 51% of the time
Summary of results, cont’d

• All estimates converge to final realized cost and duration over time
• Rayleigh (NRP) estimate of cost converges fastest 93% of the time
• Rayleigh (NRP) is the unique leading indicator 17% of the time

The other EAC calculation methods and the contractor and the PMs estimate trends **never** are the unique leading indicator.
Predictions converge fastest, are stable, and are relatively most accurate.

Prediction of duration is Almost dead on

Fit of Rayleigh (NRP) is almost perfect
Review Rayleigh fit to data

- Criteria to evaluate
  - Fit score
  - Relative accuracy score
  - Absolute convergence score
  - Relative convergence score
  - Leading indicator score
- Give examples of good cumulative scores
- Review bad scores
  - Explain why Rayleigh doesn’t appear to work
  - Develop rules of thumb for application of Rayleigh
  - Develop cautionary comments for use of Rayleigh
Cost and schedule risk (148-8)

Eventual cost will be $960.5 M with a duration of 8.1 years

Time is now 1.56 years after work started

<table>
<thead>
<tr>
<th></th>
<th>Cost Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk that cost will exceed PM's estimate</td>
<td>91%</td>
</tr>
<tr>
<td>Risk that cost will exceed PM's estimate by 10%</td>
<td>88%</td>
</tr>
<tr>
<td>Risk that cost will exceed PM's estimate by 25%</td>
<td>84%</td>
</tr>
<tr>
<td>50/50 Cost = $910M</td>
<td>50%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Schedule risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk that duration will exceed PM's estimate</td>
<td>97%</td>
</tr>
<tr>
<td>Risk that duration will exceed PM's estimate by 10%</td>
<td>94%</td>
</tr>
<tr>
<td>Risk that duration will exceed PM's estimate by 25%</td>
<td>88%</td>
</tr>
<tr>
<td>50/50 duration = 8.5 years</td>
<td>50%</td>
</tr>
</tbody>
</table>
Rayleigh fit to program RDT&E funding profiles: Summary

- 39 programs
  - 16 Navy programs
  - 9 Army programs
  - 13 Air Force programs
  - 1 DoD program
- Overall average fit: $R^2 = .979$
  - Navy average fit: $R^2 = .978$
  - Army average fit: $R^2 = .987$
  - Air Force average fit: $R^2 = .976$
  - DoD average fit: $R^2 = .98$
Way ahead

- Develop risk region and risk analysis
- Expand business insights
- Explore application to production contracts
- Explore application to program/budget data?
Conclusion

• Questions/comments?
Back-up

- Back-up slides
Early warning model
NRP Rayleigh with nominal data gives best fit

Five cases when Rayleigh with real inflation data did noticeably better started work in 1974, 1982, 1983, 1984, and 1986. Only one of these occurred during a high inflation period.

Inflation rates for RDT&E)
2 Rayleigh Spline (148-9)

Result of vertical summation of 2 Rayleigh stages

Original plan

Added work content 5.1 years after work start
2 Rayleigh Spline (148-9)

Stage 1: Original Plan

Stage 2: Original plan plus added contract content at 5.1 years after work start

A bimodal spend rate pattern
2 Rayleigh Spline and Single Rayleigh comparison (148-9)

Even the best linear fit only had $R^2=.991$
3 Rayleigh Spline (148-8)

Result of vertical summation of 3 Rayleigh stages

Original plan

- cost
- total_spline_fit
- stage 1
- stage 2
- stage 3

Added work content 1.8 years after work start

Over-target baseline and replan 4.6 years after work start

- time in years from work start
- cost in millions of $
3 Rayleigh Spline (148-8)

A trimodal spend rate pattern

Stage 1: Original Plan

Stage 2: Original plan plus Added contract content 1.8 Years after work start

Stage 3: Original plan plus added contract content plus over-target baseline at 4.6 years after work start
3 Rayleigh Spline and Single Rayleigh comparison (148-8)

3 Rayleigh Spline

Single Rayleigh

R2=.994

R2=.988
RDT&E funding profiles

Program 549

Program 581, First 10 years

Program 581, last 9 years

Program 582

R2 = .99
R2 = .979
R2 = .964
R2 = .914
RDT&E funding profiles, cont’d

Program 390

Program 223

Program 197

Program 364

R2 = .982

R2 = .951

R2 = .984

R2 = .989
RDT&E funding profiles, cont’d

Program 101

Program 542, first 8 years

Program 542, last 7 years

Program 191

R2 = .991

R2 = .925

R2 = .978

R2 = .997
RDT&E funding profiles, cont’d

Program 282
R2 = .988

Program 334
R2 = .995

Program 517
R2 = .967

Program 289
R2 = .999
RDT&E funding profiles, cont’d

Program 178

Program 212

R2 = .972

R2 = .996
RDT&E funding profiles

Program 219

Program 341

Program 601

Program 278

R2 = .993

R2 = .984

R2 = .999

R2 = .986
RDT&E funding profiles, cont’d

Program 260

\[
\begin{align*}
\text{millions of $} & \quad 0 & \quad 200 & \quad 400 & \quad 600 & \quad 800 \\
0 & \quad 5 & \quad 10 & \quad 15 & \quad 20 & \quad 30
\end{align*}
\]

\[R^2 = .976\]

Program 280

\[
\begin{align*}
\text{millions of $} & \quad 0 & \quad 500 & \quad 1000 \\
0 & \quad 5 & \quad 10 & \quad 15
\end{align*}
\]

\[R^2 = .992\]

Program 831

\[
\begin{align*}
\text{millions of $} & \quad 0 & \quad 200 & \quad 400 & \quad 600 & \quad 800 \\
0 & \quad 5 & \quad 10 & \quad 15 & \quad 20
\end{align*}
\]

\[R^2 = .973\]

Program 148

\[
\begin{align*}
\text{millions of $} & \quad 0 & \quad 1000 & \quad 2000 & \quad 3000 \\
0 & \quad 5 & \quad 10 & \quad 15 & \quad 20
\end{align*}
\]

\[R^2 = .993\]
RDT&E funding profiles, cont’d
RDT&E funding profiles

Program 376

Program 200, first 14 years

Program 200, last 16 years

Program 273

R2 = .989

R2 = .986

R2 = .92

R2 = .993
RDT&E funding profiles, cont’d

Program 176

Program 265

Program 252

Program 555

R2 = .979

R2 = .994

R2 = .989

R2 = .936

5/9/2008 3:51 PM
RDT&E funding profiles, cont’d

Program 503

Program 560

Program 302

Program 248

R2 = .915

R2 = .997

R2 = .995

R2 = .994
RDT&E funding profiles, cont’d

Program 210

Program 166

R2 = .954

R2 = .995
RDT&E funding profiles

Program 198

millions of $

0 5 10 15 20

time in years from program start

R2 = .98
Research Task: Questions to Be Addressed

- How can an executive effectively use questionable EVM data for management decisions?
- Can new tools be developed or “old” tools modified to give earlier warning of impending contract execution problems?
Briefing Agenda

- Summarize Rayleigh model (Version 1)
- Summarize results of validation (Version 1)
- Describe tool
  - One module for “traditional” analysis
  - One module for assessment of an original plan before actual cost data have been collected
- Potential impact of study
- Rayleigh, Version 2
- Recommendations
The Rayleigh Model

Cumulative cost as a function of time (in millions of $)

\[ c(t) = d \left(1 - e^{-\alpha t^2}\right) \]

Scale factor. How much will the program cost?

Proportion of work completed at time \( t \).

Shape parameter. When will peak spending rate occur?

Duration time (in years). What is the current duration of the program?
An Example of a Rayleigh Schedule

The parameter $d$ tells us the height of the curve. What is the upper bound on cumulative cost?

The parameter $\alpha$ tells us the shape of the curve. When does the peak spending rate occur?

This curve inflects when the rate curve below reaches a maximum.
Rayleigh Model Advantages

- The model is applied only to R&D contracts
- Rayleigh is a plausible model of cumulative cost accrual over the life of a contract
- The model is based on dollars that have not been adjusted for inflation
- The model depends only on standard currently available EVM data (no new reports)
- The model only requires 3 actual cost submissions and a budget
Rayleigh Model Advantages

• The model does not depend on the reliability of Earned Value (BCWP) submissions
• The model predicts both EAC and completion date
• The model predicts the path of actuals to completion date
• The model is Excel-based using standard Solver add-in
Validation of the Rayleigh model

• Compared accuracy of predictions considering cost at completion and completion time

• Methods compared
  – Rayleigh estimate
  – Contractor estimate
  – PM Estimate
Validation (cont)

- Methods compared (cont)
  - EAC1 (BAC/CPI plus max of contractor and PM time estimate)
  - EAC 2 (Actuals+(BAC-EV)/(.8CPI+.2SPI) plus max of contractor and PM time estimate)
  - EAC 3 (Actuals +(BAC-EV)/(CPI X SPI) plus max of contractor and PM time estimate)
  - Note: EAC1, EAC2, and EAC3 methods do not produce an independent estimate of duration time
Validation (cont)

- Selected programs for regression analyses
- Selected only R&D programs
- Selected complete programs
  - Eliminated programs less than 90% complete to get valid baselines
  - Eliminated programs with over 2 years between work start and first submission to evaluate early warning utility
Validation (cont)

• Began with entire CAS database
• Selected 74 programs
• Consisting of 115 contracts
• Earliest start date 1/1/1970
• Latest start date 8/1/2002
• All services included
Rayleigh validation results

Rayleigh Composite Accuracy Ranking

Percent

Rayleigh best  Rayleigh 2d  Rayleigh 3d  Rayleigh 4th  Rayleigh 5th  Rayleigh 6th

81.1  1.4  0.3  2.3  4.9  10
How much better were Rayleigh predictions? (EAC)

- All estimates underestimate final cost over 78% of the time. When they underestimate cost:
  - Rayleigh underestimates final cost on average by 30%
  - The contractor underestimates on average by 35%
  - The PM underestimates on average by 37%
  - The EAC1 method underestimates on average by 34%
  - The EAC2 method underestimates on average by 34%
  - The EAC3 method underestimates on average by 32%
How much better were Rayleigh predictions? (time)

- All estimates underestimate final contract duration over 73% of the time. When they underestimate duration:
  - Rayleigh underestimates duration on average by 24%
  - The contractor underestimates on average by 35%
  - The PM underestimates on average by 55%
Conclusions from database validation

- Rayleigh yields best estimate of final cost
- Rayleigh yields best estimate of time duration
Conclusions from database validation (cont)

- Rayleigh is still short of final cost on average by 30%
- A basic assumption of all EAC techniques is that we know full scope at the time of prediction and we fit the sparse data with a single Rayleigh curve
- Earlier CNA study ("Program Cost Growth: The Navy’s Experience 1983-2004")
  - Total cost growth is level from 1978-2004
  - Within the total, the “overrun" component is declining and the “changes” component is increasing
- We think the bulk of the 30% shortfall is attributable to contract changes
Executive Cost and Schedule Assessment (XCaSA) tool

- Executive Plan Assessment Module (XPAM)
  - Allows executive to assess plan realism before any actuals are submitted
- Executive Contract Assessment Module (XCAM)
  - Allows executive to assess contract performance after at least 3 submissions of actuals
XCaSA (cont)

- XCAM (cont)
  - Incorporates Rayleigh estimates
  - Displays “traditional” analysis for comparison
  - Incorporates relevant DCMA tripwires
New metrics (XCAM)

- Cost Overrun Vulnerability Index:
  \[ COVI = \frac{EAC_{Rayleigh}}{BAC_N} \]

- Schedule Slip Vulnerability Index:
  \[ SSVI = \frac{\hat{t}_{Rayleigh}}{\hat{t}_{N,PM}} \]
New features (XPAM)

- Plan Validity Index
- “What if” drills
XCaSA advantages

• User friendly
• Interactive
• Provides useful information early in life of contract
• Uses built-in Solver add-in with widely used Excel spreadsheet software
• Provides business insights
Dashboard view of XCAM

**Program Name**: EFV
**Contract Name**: SDD
**Contract Number**: M67854-01-C-0001
**As of Date of data**: 24-Oct-06
**Start date of contract**: Feb-01
**PM estimated completion date**: Sep-08
**PM contract duration (years)**: 7.59
**Rayleigh estimated completion date**: Apr-10
**Rayleigh contract duration (years)**: 9.17

**Cost Status**
- **Comment**: On Cost Vulnerability Index (Line 20).
- **Questions**:
  - Are you adequately budgeted?
  - If not, what are your plans to resolve the budget issue?
  - Do you have a cost reduction program?
  - What are Nunn-McCurdy implications?

**Schedule Status**
- **Comment**: On Schedule Slip Vulnerability Index (Line 21).
- **Questions**:
  - What is the basis for your schedule assessment?
  - Do you routinely do schedule risk assessments?
  - What are the implications of slipping the PM’s schedule?
  - What are the most important critical path tasks over the next 12 months?

**Future Cost Performance**
- **Comment**: On TCPI (Line 22).
- **Cost Variance Graph**:
  - If SV has recent flat or positive slope, explain improvement in CV.
  - If SV has flat or recent positive slope, explain improvement in SV.
  - Why is PM’s cost projection more pessimistic than contractor’s?
  - Is remaining MR sufficient to cover remaining risk?

**Cost Performance**
- **Cost Variance Graph**:
  - If the TAB is increasing, explain the increasing contract scope.
  - Does your EAC include projected increases in TAB?
  - When will changes stop?
Dashboard view of XPAM

This plan is initially frontloaded. Explain why the plan has most of the effort in the first half of the program. Are labor resources in place to support a rapid build-up? There is low effort in later phases. Is this consistent with the test plan and other work content (for example, production start)? Is this consistent with risk assessment?
XPAM Advantages

• Only XPAM can assess the initial plan
  – Current EVM diagnostics cannot assess the plan until after submission of some number of full EVM data
  • Often more than a year after contract start
XCaSA Tool Status

- Tested XCAM with multiple current programs
- Tested XPAM with notional initial program management baselines
Potential Impact

• Improve oversight of programs
• Obtain early assessments of plan and contract execution
• Make better informed tradeoff decisions
• Make EVM tool of choice across the government
Rayleigh, Version 2

- Update model algorithm and revalidate
- Incorporate contract-level cost and schedule risk
- Investigate use of the model with program level budget data
- Test applicability of model to procurement contracts
- Upgrade user interface
- Determine feasibility of developing Monte Carlo policy simulation package with Rayleigh spline for tool to account for “missing 30%”
- Upgrade “insight” prompts in tool
- Link model to AOP-like management process
Update model algorithm and revalidate

• Re-estimated using non-linear least squares with restricted monotonically transformed parameters (NRMP)
• Improves efficiency of computation
• Enables approximation of risk region
• Revalidated with over 2500 information sets
Update model algorithm and revalidate: results

![Graph showing predicted cumulative cost in millions of $ over time in years from work start when prediction is made.

![Graph showing predicted contract duration in years over time in years from work start when prediction is made.](https://www.iceaaonline.com)
Results of more comprehensive analysis

• Rayleigh estimates final cost with greater than or equal accuracy 67.4% of the time
• Rayleigh estimates final contract duration with greater than or equal accuracy 58.5% of the time
• Rayleigh estimate of final cost converges to within 10% of the final actual cost with greater than or equal speed 92% of the time
• The average speed of convergence for the Rayleigh final cost estimate is .48
Results of more comprehensive analysis

Average $R^2 = .93$

Model: $r_2 = a + b \times WSDATEINDEX$

$\hat{a} = .95$  
$\hat{b} = -.00126$

(.038108)  
(.001978)
Results of more comprehensive analysis

- Compared Rayleigh with “level of effort” model
- Rayleigh a better “fit” 63% of the time
- Rayleigh better or virtually the same “fit” 85% of the time
- Linear model is unambiguously better “fit” 15% of the time
Results of more comprehensive analysis

- Estimated covariance matrix
- Calculated confidence ellipse in parameter space
- Mapped ellipse into “final-cost/final-duration” space
- Graphed confidence region
Results of more comprehensive analysis

The graph shows the cumulative cost in millions of dollars over time, with the time in years from work start ranging from 0 to 11. The graph includes lines for actual cost, rayl_fit, conf_region_up, conf_region_down, prediction, and region_limits.
Results of more comprehensive analysis