Data Collection and Analysis Supporting Defendable Cost Estimates

Arlene Minkiewicz, Chief Scientist
PRICE Systems, LLC
arlene.minkiewicz@pricesystems.com

TruePlanning®
by PRICE® Systems

Agenda

- Introduction
- Historical Prospective on Cost Estimation
- Data Driven Estimation and its Importance in the current economic client
- Pilot Case Studies
  - Software Data Collection
  - Automotive Data Collection and Analysis
- Lessons Learned

Introduction

- Cost estimation in the Aerospace and Defense world emerged as a necessary around the time of World War II
- All sorts of mathematical and experiential models have been applied to the problem with varying degrees of success
- More and more organizations are pushing for models tuned to their specific data and experience
- Many of these organizations lack the infrastructure, processes and tools to collect and analyze project data efficiently
- This paper presents two pilot projects (both on-going) to inject data driven estimation into two very different organizations.

Historical Perspective

- General H.H. “Hap” Arnold at the end of WWII

“During the war, the Army, Army Air Forces, and the Navy have made unprecedented use of scientific and industrial resources. The conclusion is inescapable that we have not yet established the balance necessary to insure the continuation of teamwork among the military, other government agencies, industry, and the universities. Scientific planning must be years in adva[nce] of the actual research and development work.”

- Leaders took this to heart – state of the art weapons systems are necessary to ensure military superiority

- Beginning of period of time during which the Department of Defense (DOD) is plagued with issues surrounding cost!

Historical Perspective

Operations Research applied to Military Affairs

Operations Research and Statistics applied to Industrial Affairs

PRICE H as an RCA Only Tool

Proliferation of Commercially Available Cost Estimation Models


General H.H. "Hap" Arnold


Frank Collbohm

RAND Created


Frank Frieman


J. Gordon Davis

PMI Created

OSD CAIG Formed

ISPA Founded

PC Project Mgmt EVM Tools

Parametric Estimating Initiative

Korean ISPA Chapter

Great Depression

World War II

Severe DoD Project Cost Overruns

Berlin Wall Breached

War on Terrorism

Homeland Defense

Defense Transformation

Gov't & Industry Business Modernization Cost Overruns

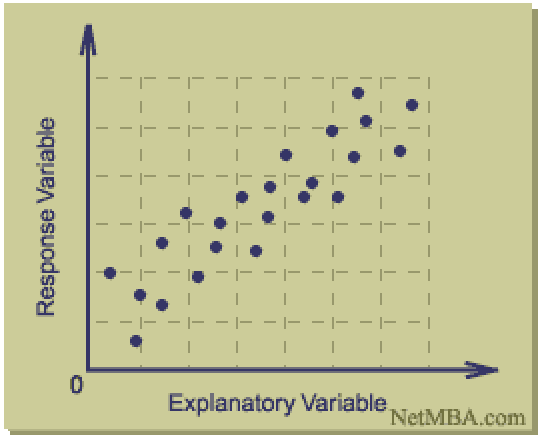
1930 1940 1950 1960 1970 1980 1990 2000 2010 2020

Optimize tomorrow today.™

TruePlanning® by PRICE® Systems

Data Driven Estimation

- Not a new idea – just a new focus on what's behind an estimate
- Regardless of estimating methodology – a data driven approach lends a sense of credibility and defendability to any estimate
- A 'data-driven' estimate may be…..
  - Based on data from analogous systems
  - Based on a model using analogous systems to determine model inputs
  - Based on a model calibrated to historical data
- A 'data-driven' estimate must be
  - Transparent as to what data was used
  - Clear about how that data was used



Two Case Study Pilots

- Software data collection and analysis to support more defendable estimate
  - Army Program Executive Office for Simulation, Training, and Instrumentation (PEO STRI)
  - No data collection processes in place
  - Pilot involves one large, complex program – two contractors – one does primarily development, one primarily integration
- Data collection and analysis to support more informed decisions for an automotive manufacturer
  - Low volume auto maker
  - Collecting tons of cost and technical data at the part level
  - Need processes and tools to make sense of that data and mine it into usable form
  - Pilot involves about 100 cars of 9 different models

Software Pilot - Motivation

- PEO STRI personnel take their mission of delivering training to the warfighter seriously and intend to deliver best of breed solutions.
- Estimates presented to the Office of the Deputy Assistant Secretary of Defense for Cost and Economics (ODASA-CE) are met with the question "Where's the data that supports this estimate?"
- Unsatisfactory responses inspire large risk margins resulting in less budget available to the program
- Furthermore, without insight into projected costs for specific capabilities, it is difficult to make trade-offs to accomplish reduced costs
- PEO STRI wants to establish a consistent, repeatable process to collect cost and technical data for each capability they develop in their programs

Software Pilot – Implementation and Process

- Lots of collaboration between PEO STRI personnel, contractors and the PRICE team resulting in…..
  - Integrated Product Team (IPT) developed
  - Data Item Dictionary created to define….
    - What data to collect
    - How often to collect it
    - How to align project specific data to more general categories (across multiple contractors)
  - Requirement added to Contract Deliverable Requirements List (CDRL) making data collection a contractor requirement
  - Creation of automated code counting tools by wrapping USC's Code Counting tool with Excel
  - Software Resource Report (SRR) developed to align with output of code counting tool and automate mapping processes

Software Pilot – Software Resource Report

Section 2.5 PRODUCT SIZE REPORTING								(Column Will not print)	
SECTION 2.5.1 Capability Requirement Name	SECTION 2.5.2 Standardized Capability Name	SECTION 2.5.3 Language	SECTION 2.5.4 % Auto Generated	SECTION 2.5.5 New Code	SECTION 2.5.6 Deleted Code	SECTION 2.5.7 Modified Code	SECTION 2.5.8 Unmodified Code	SECTION 2.5.9 Application Type	SECTION 2.5.10 Estimated % of Total Effort (Build)
CR 65774 Reduced System Footprint - SCAMT Multiple Applications on Single Node	Complex Capability	Java		1,891	3,740	628	2,492,491	Assessment/Analysis Functions - Simulation (Constructive)	12.00%
CR 65774 Reduced System Footprint - SCAMT Multiple Applications on Single Node	Complex Capability	XML		63	36	36	8,242,067	Assessment/Analysis Functions - Simulation (Constructive)	
CR 67136 Crowd Formation	Moderate Capability	Java		978	-	-	2,499,444	Assessment/Analysis Functions - Simulation (Constructive)	3.00%
CR 67136 Crowd Formation	Moderate Capability	XML		779	-	-	8,208,573	Assessment/Analysis Functions - Simulation (Constructive)	
CR 67136 Crowd Formation	Moderate Capability	XLS File		121	-	-	815,757	Assessment/Analysis Functions - Simulation (Constructive)	
CR 67136 Crowd Formation	Moderate Capability	CSV file		7	-	-	276,976	Assessment/Analysis Functions - Simulation (Constructive)	
CR 67154 Dynamic UHRB Aperture Consistent Ordering	Complex Capability	C++		12	41	34	673,807	Assessment/Analysis Functions - Simulation (Constructive)	2.00%
CR 67154 Dynamic UHRB Aperture Consistent Ordering	Complex Capability	Java		238	87	98	2,496,600	Assessment/Analysis Functions - Simulation (Constructive)	
CR 67156 Effects Guidance Matrices	Moderate Capability	Java		1,791	10	15	2,497,419	Assessment/Analysis Functions - Simulation (Constructive)	35.00%
CR 67156 Effects Guidance Matrices	Moderate Capability	XML		7	-	3	8,209,143	Assessment/Analysis Functions - Simulation (Constructive)	
CR 67156 Effects Guidance Matrices	Moderate Capability	XLS File		14,066	-	-	813,585	Assessment/Analysis Functions - Simulation (Constructive)	
CR 67156 Effects Guidance Matrices	Moderate Capability	CSV file		24	-	-	276,972	Assessment/Analysis Functions - Simulation (Constructive)	
CR 67164 Raytrace Feature Composition	Complex Capability	C++		982	4	35	676,299	Assessment/Analysis Functions - Simulation (Constructive)	13.00%

Software Pilot – Progress to Date

- Data collection occurs every 10 weeks.
- The first iteration was a thrown away
  - Automation failures in the contractor configuration
  - Uncertainty of some data definitions
  - Additional discussions required by IPT to iron out questions and uncertainties
- Three successful iterations have occurred since then – each resulting in slight refinement to the process, DID and/or SRR
- Some practice calibration exercises have led to additional changes to data collection process and the DID
- Current focus is on determination of best long term strategy for storage of this data so it can be accessed and analyzed on a per capability basis

Software Pilot – Lessons Learned

- Entire team needs to understand motivation for data collection
  - Alleviate contractor concern about being ‘measured’
  - Facilitate determination of data collection targets
  - Need buy in from the data collectors
- Automation is key to success
  - Creates consistency
  - Creates efficiency
  - Reduces impact on contractors
  - Promotes harmony
- Communication and teamwork are paramount
- Flexibility necessary from both the contractors and PEO STRI
  - Not everything is possible or practical
  - Compromises are sometimes necessary

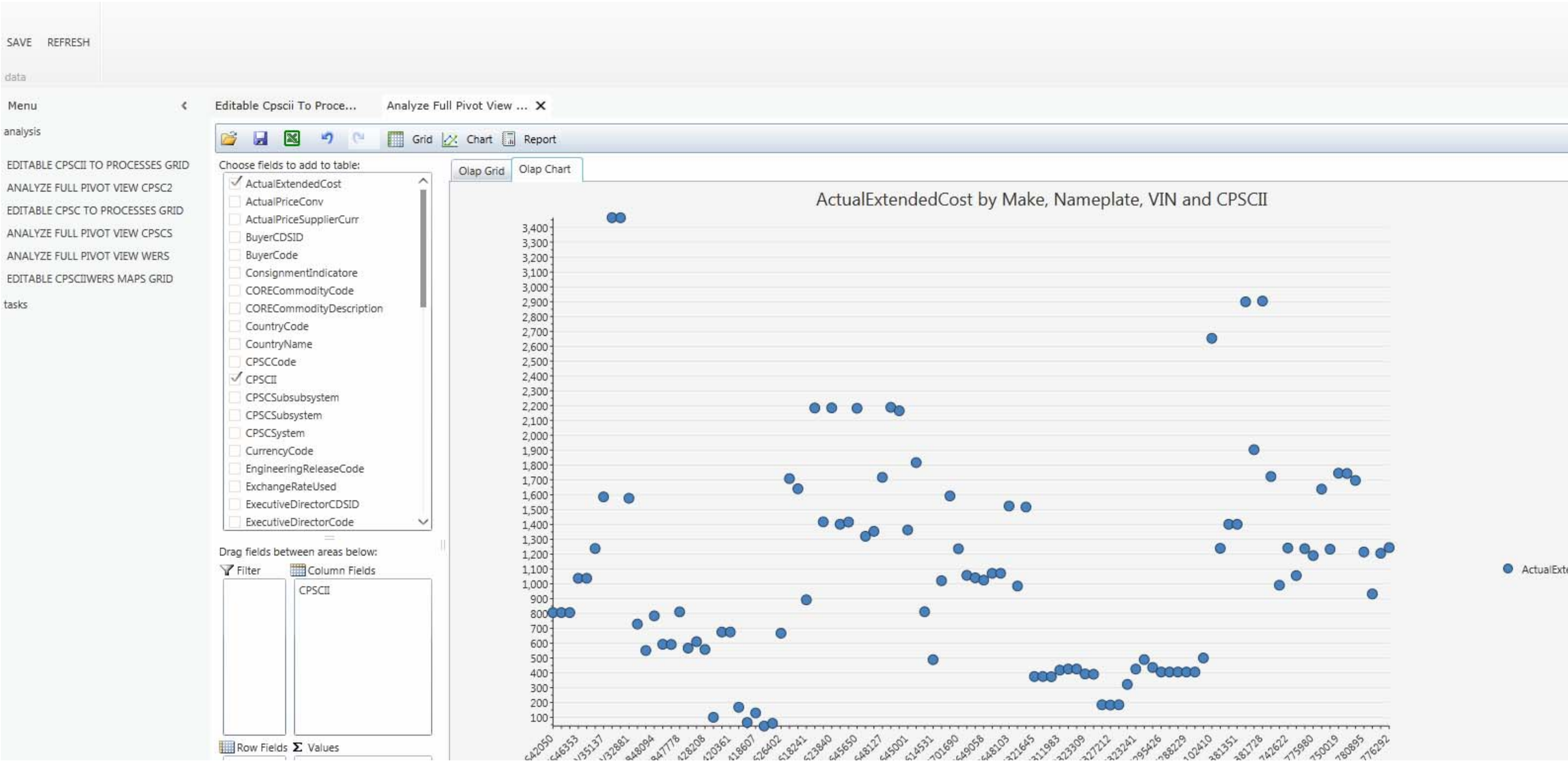
Automotive Pilot - Motivation

- Auto manufacturer makes low volume luxury cars and has a vast quantity of cost and technical data for each automobile at the part level.
- This data is collected by various systems and is stored in various files and formats
- The auto manufacturer is currently unable to predict the cost of an a new or modified vehicle until about two months before the first one rolls out of the plant
- They would like to be able to predict at concept what the cost of the new or modified automobile will be using a question based estimating approach

Automotive Pilot – Implementation and Process

- Initial meetings with customer were necessary to understand their mission and identify sources of potential data
- Best potential cost drivers were identified to be the feature codes associated with each individual vehicle
- The sub assembly level was identified as the level for analysis (e.g the seat, dashboard, center console)
- Automation was created to
  - Align cost and part information at the sub-assembly level
  - Align costs with feature codes
  - Add visualization capabilities

Data collection and visualization

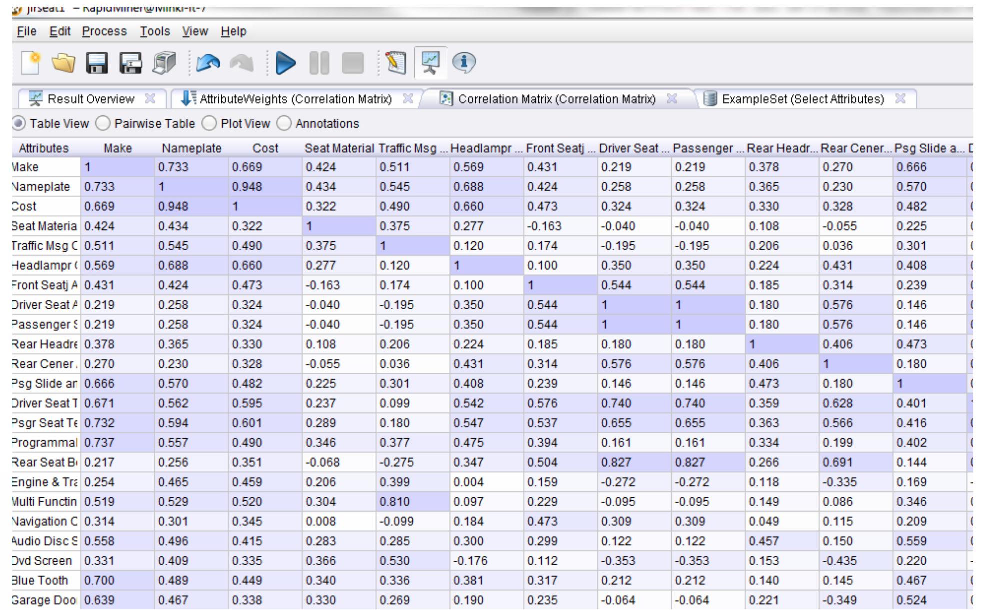


Optimize tomorrow today.™

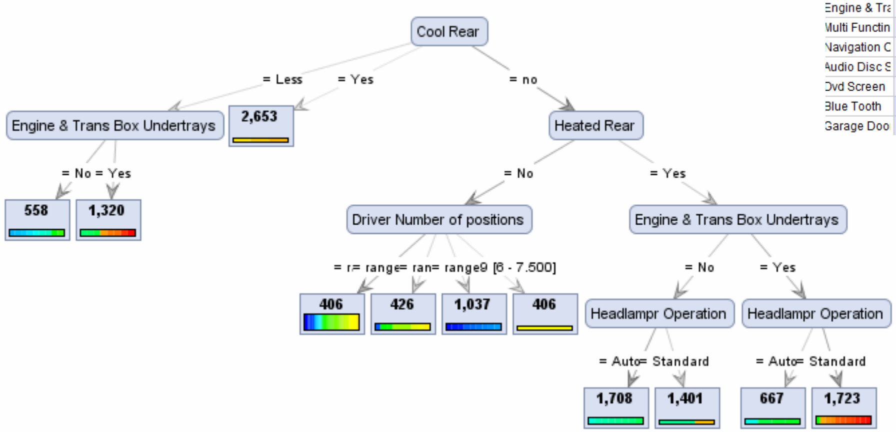
Automotive Pilot – Progress to Date

- One sub-assembly was identified for the pilot analysis and data for 100 vehicles was collected
  - Automation was applied to extract cost and part information for each vehicle into a single file.
  - Each sub-assembly has a finite number of possible feature codes associated with it
  - Each of these needed to be identified and searched for in the individual files associated with the vehicle Identification Number(VIN). Automation was applied to create mappings between feature codes and costs.
- Data mining techniques were used to identify likely trends
- Regression techniques are currently being applied to create the CERs underlying the question based estimating process
- Further automation and process change being considered to align the disparate data collection systems in the plant.

Data Analysis

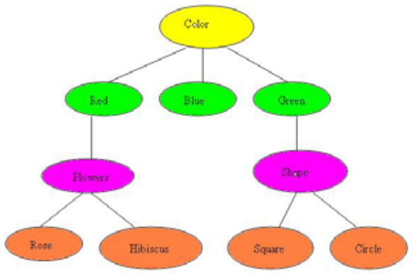


Attributes	Make	Nameplate	Cost	Seat Material	Traffic Msg ...	Headlampr ...	Front Seatj ...	Driver Seat ...	Passenger ...	Rear Headr...	Rear Cener...	Psg Slide a...
Make	1	0.733	0.669	0.424	0.511	0.569	0.431	0.219	0.219	0.378	0.270	0.666
Nameplate	0.733	1	0.948	0.434	0.545	0.688	0.424	0.258	0.258	0.365	0.230	0.570
Cost	0.669	0.948	1	0.322	0.490	0.660	0.473	0.324	0.324	0.330	0.328	0.482
Seat Materia	0.424	0.434	0.322	1	0.375	0.277	-0.163	-0.040	-0.040	0.108	-0.055	0.225
Traffic Msg C	0.511	0.545	0.490	0.375	1	0.120	0.174	-0.195	-0.195	0.206	0.036	0.301
Headlampr (	0.569	0.688	0.660	0.277	0.120	1	0.100	0.350	0.350	0.224	0.431	0.408
Front Seatj A	0.431	0.424	0.473	-0.163	0.174	0.100	1	0.544	0.544	0.185	0.314	0.239
Driver Seat A	0.219	0.258	0.324	-0.040	-0.195	0.350	0.544	1	1	0.180	0.576	0.146
Passenger S	0.219	0.258	0.324	-0.040	-0.195	0.350	0.544	1	1	0.180	0.576	0.146
Rear Headre	0.378	0.365	0.330	0.108	0.206	0.224	0.185	0.180	0.180	1	0.406	0.473
Rear Cener.	0.270	0.230	0.328	-0.055	0.036	0.431	0.314	0.576	0.576	0.406	1	0.180
Psg Slide ar	0.666	0.570	0.482	0.225	0.301	0.408	0.239	0.146	0.146	0.473	0.180	1
Driver Seat T	0.671	0.562	0.595	0.237	0.099	0.542	0.576	0.740	0.740	0.359	0.628	0.401
Psgr Seat Te	0.732	0.594	0.601	0.289	0.180	0.547	0.537	0.655	0.655	0.363	0.566	0.416
Programmal	0.737	0.557	0.490	0.346	0.377	0.475	0.394	0.161	0.161	0.334	0.199	0.402
Rear Seat Be	0.217	0.256	0.351	-0.068	-0.275	0.347	0.504	0.827	0.827	0.266	0.691	0.144
Engine & Tra	0.254	0.465	0.459	0.206	0.399	0.004	0.159	-0.272	-0.272	0.118	-0.335	0.169
Multi Functin	0.519	0.529	0.520	0.304	0.810	0.097	0.229	-0.095	-0.095	0.149	0.086	0.346
Navigation C	0.314	0.301	0.345	0.008	-0.099	0.184	0.473	0.309	0.309	0.049	0.115	0.209
Audio Disc S	0.558	0.496	0.415	0.283	0.285	0.300	0.299	0.122	0.122	0.457	0.150	0.559
Dvd Screen	0.331	0.409	0.335	0.366	0.530	-0.176	0.112	-0.353	-0.353	0.153	-0.435	0.220
Blue Tooth	0.700	0.489	0.449	0.340	0.336	0.381	0.317	0.212	0.212	0.140	0.145	0.467
Garage Doo	0.639	0.467	0.338	0.330	0.269	0.190	0.235	-0.064	-0.064	0.221	-0.349	0.524



Auto Pilot – Lessons Learned

- Automation certainly key to success
  - First pass of data mining was done virtually manually. Addition of automation pares a weeks worth of work down to seconds.
- Communication
  - Understand the mission of the data collection project
  - Understand the lexicon of the customer
- Analysis of large quantities of data requires visualization and data mining capabilities. Tools such as Rapid-Miner, R and Excel are essential
- For data driven estimating – quality of data is more important than quantity – too much data can create noise and misleading results
- Flexibility



Conclusions

- The need for data driven estimation is driving many organizations in Aerospace and Defense and industry in general to look for new ways to collect and use data
- Data driven estimation is really nothing new – we have been using data (real or experiential) to justify our estimates forever. The new requirement is transparency and openness about the data and how it is being used.
- To support this effort organizations may need to establish formal processes for collecting their project data.
- When doing this it is important to understand the importance of
  - Communication
  - Automation
  - Flexibility
  - Buy in across all the stakeholders

Questions

Arlene.minkiewicz@pricesystems.com

TruePlanning®
by PRICE® Systems

Backup Slides