Data-Driven Estimating - Quantifying Electronics Complexity Using Public Data

Abstract

The parametric cost estimation industry is being challenged to increase the defendability of their estimates, mainly through comparison to analogous data. However, companies are more protective than ever of their high quality cost data. While this data is often obtainable, it usually comes with strings attached (non-disclosure agreements, etc.) that preclude its usage in defending your estimate. At the same time, government and military databases with cost and technical data are being exposed to the public, and while the noise introduced by granularity issues and the mapping process of this data presents its own set of challenges, this trend also contains the key to a solution.

This paper discusses our research into quantifying electronics complexity for inclusion in our hardware cost estimation model, while meeting the challenges of this new reality. I will discuss our research approach to this problem, our final solution structure, and interesting insights obtained along the way. The process is based on data obtained with the IHS Haystack tool, which contains parts and logistics data on over 100 million items in the U.S. Federal Supply Catalog across 70+ Army, Navy, Air Force, and other government, military and commercial databases. I will discuss issues we encountered in molding available data to a format that fits our cost models, validating our assumptions, and dealing with incomplete data. The end result is a calculator that guides users through quantifying complexity of a vast selection of electronic items in a way that is defendable by comparison to the public data on which it is based.

Introduction

In recent years, the parametric cost estimation industry is experiencing a push for greater defendability of their cost estimates. One of the main issues in defending a parametric cost estimate is defending the selection of subjectiveparameter values, because subjective parameters allow for bias to creep into a cost estimate.In this paper, I'll be discussing a project we've completed in providing guidance for selecting a value for one of the most important parameters in the TruePlanning for Hardware model: Manufacturing Complexity for Electronics. In doing so, I hope to illustrate some of the challenges and solutions for using public data to build and defend an estimate, when that data only partially matches the ideal granularity, comprehensiveness and quality.

Manufacturing Complexity for Electronics is a technology index in the form of a numerical scale for the electronic portion of the item being modeled. It is intended to measure the electronicitem's technology, its producibility (component make-up, packaging density, test and reliability requirements, etc.), and yield. Manufacturing Complexity for Electronics can be thought of as representing the production cost per unit of weight for electronics, as illustrated in Figure 1.

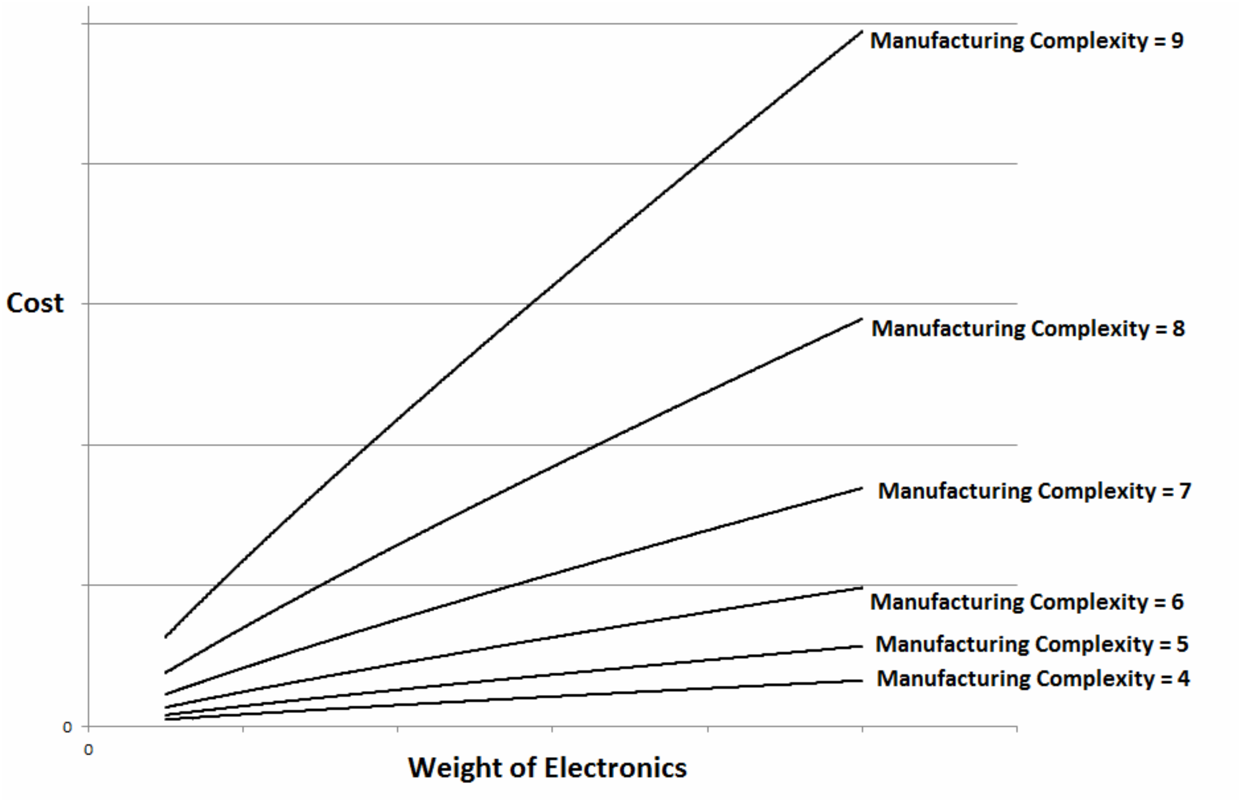

Figure 1: Cost vs. Weight Curves, as defined by Manufacturing Complexity for Electronics

For this project, our goal was to update our guidance to the cost estimator for selecting a value for the Manufacturing Complexity parameter, while arming them with the ability to defend their selections by justifying it with analogous data. The requirements for this guidance are:

1. It must cover a vast majority of electronic equipment types and technologies estimated by the Aerospace and Defense market.

2. It must be derived from actual electronic cost data that can be redistributed publicly.

3. It must enable the user to trace recommendations back to the cost data from which it came.

Research Method

Scope

The research began with scoping the problem. The first requirement (covering a vast majority of electronic equipment types/technologies) means that we needed to gain a better grasp on how we can ensure we cover all kinds of electronics used by the aerospace and defense market. We began by meeting with electronic engineers in this market, trying to scope out this issue. These meetings led us to classify electronics two different ways: Electronic Function and Equipment Type.

Electronic function is a way of describing the general functions that electronics can perform. Each electronic function can be found in many different types of electronic equipment. Examples include amplification, sensing/detecting, transduction, signal processing, switching, power conditioning/converting, voltage/current regulation, oscillation, etc. A full list can be found in Appendix A. Equipment types, on the other hand, are electronics that share a similar purpose. They often use many different electronic functions to meet their purpose. Examples include communications, navigation, control systems, countermeasures, GPS, etc. The list is potentially endless if including rare equipment types, but a manageable list could be compiled that covers a vast majority of projects within the aerospace and defense community. A preliminary list of the equipment types that we have currently analyzed canbe found in Appendix B, though we will continue adding to the list before the release of this new guidance, and in the future as well, based upon requests by our users.

Data Requirements

The main drivers of the hardware model include the following:

- Quantity
- Weight (of Electronics and Structure)
- Manufacturing Complexity (for Electronics and Structure)
- Operating Specification[1]

These represents the base set of inputs required for our model to estimate unit production cost[2]. However, if one of these drivers is unknown, but the unit production cost is known, the model can be run “in reverse” to calculate the unknown input. This method, called calibration, can be used to objectively measure a value for Manufacturing Complexity for Electronics, and is the basic method used in this project.

The next challenge was finding data sources that contain all the necessary information to perform this calibration. Of course, it turns out there were no data sources that perfectly matched this data format. However, we did find a set of data sources that, when combined, can be used to derive all the needed inputs. In this section, I will discuss the data that we’ve found to be available, the problems encountered in adapting this data to fit our calibration process, and the techniques used to mitigate these problems.

Knowing the base set of information needed for the calibration process, we set out to find data sources. Many sources were considered. Data was available from various customers and from projects we’ve

[1] Operating Specification describes the intended operating environment and defines the degree of reliability, portability, security, documentation, and testing requirements. For example, hardware produced for manned spacecraft has very high reliability, quality, and documentation requirements.

[2] This represents the base set of inputs for estimating Production-phase costs. Other inputs are available to drive Development, Operation and Support phases. Also, more inputs are available to describe different Production-phase scenarios, such as learning curves, integration complexities, etc, but are not main drivers in this case study.

worked on in the past, but much of it was not publicly shareable, as it came with non-disclosure agreements or other strings attached. Electronic part databases and search engines such as GlobalSpec.com, Octopart.com, etc. were considered, as they contain the components used to build practically any type of electronic equipment. However, building a cost estimate of assembled, completed, equipment based solely upon component purchase prices was insufficiently accurate.

The dataset that best met our needs was the U.S. Federal Supply Catalog, which contains parts and logistics data on over 100 million items, including practically every electronic equipment type used by the aerospace and defense community. All items that could be ordered as replenishment spares by the U.S. government are entered in the U.S. Federal Supply Catalog, which is a useful level of detail for data collection. The data sources include over 70 US Army, Navy, Air Force and other government, military and commercial-related databases, and the IHS Haystack tool was used to organize this data for us [1].The next section (Data Collection) discusses how the data were grouped and collected by the categories in Appendix A. Then, the following section (Data Transformation) discusses the specific data fields available for each data point, and how they are used to work with our calibration process.

Data Collection

The data we've used from the U.S. Federal Supply Catalog is at the level of replenishment spares. Often, U.S. military spares for electronics are at the circuit card assembly (CCA) level, though there are plenty of exceptions. Some items are spared by the "box of electronics," such as flight recorders for aircraft (i.e. the black box), or as entire pieces of equipment, such as display electronics (which come as a complete display).

For each category of electronics from Appendix A, we first checked if these items were spared at the CCA level. If so, we tried to collect data on CCAs of that type, as this removed uncertainty from our judgments of the ratio of structure to electronics. In other cases, we collected data at whatever level the items were generally spared, trying to keep the data granularity as close to pure electronics as possible, or at the level where it was easiest to determine the ratio of structure to electronics.

Many items in the U.S. Federal Supply Catalog were categorized according to Item Name Code (INC) from the NATO Codification System[2], which was used frequently to identify items within the same category. Others were identified based on text searches through a Technical Characteristics database, which often contains detailed descriptions of the equipment. Once an appropriate method was developed to query items that fit into the same category, all of these data points were downloaded in a separate spreadsheet that contains every data field we've used for the analysis, along with the full set of technical characteristics for our users to reference later. All of this data is made available to users of our tool so they can justify and defend their estimates by pointing to these analogous items.

Data Transformation (Mapping Available Data to our Calibration Format)

As discussed earlier, the calibration process requires the following to measure Manufacturing Complexity for Electronics:

- Unit Production Cost
- Weight of Electronics and Structure
- Operating Specification
- Manufacturing Complexity for Structure (when item contains structure)

In this section, I'll talk about the data that's available for items in the U.S. Federal Supply Catalog, and how it fits into our calibration process.

Unit Production Cost

This was perhaps the most challenging piece of information to find from the available data. The cost data most commonly available is called ML-C Unit Price. This price comes from a Defense Logistics Agency (DLA) database, and is basically the average unit price chargedby the DLA to the Service Branch for replenishment spares of the item. The ML-C Unit Price has lots of costs wrapped up in it, which can be broken down as follows:

$$ML-C\ Unit\ Price = Acquisition\ Cost + Logistics\ Overhead\ Costs$$

Acquisition Costcan be broken down into the Unit Production Cost, as well as an amount for the manufacturer's G&A, Fee/Profit, and Cost of Money. The amount for G&A and Fee/Profit are unknown; However, the DCAA audits manufacturers to ensure that prices for replenishment spares are "fair and reasonable." Multiple subject matter experts we've spoken with have stated that it's reasonable to assume a value 15% of the purchase price for G&A and Fee/Profit, though the DCAA doesn't explicitly state this number. For now, we'll operate under this assumption, but we will revisit this assumption to check its validity or find a more appropriate number in the "Final Calibration Step" section. Cost of Money is simply the interest that could be earned if the amount invested in spare parts was invested elsewhere. A reasonable number is 2% of the ML-C Unit Price. Finally, Logistics Overhead cost is a charge for the cost to the Service Branches' logistics organization in managing the orders for spare parts, assuring they reach the final destination, stocking, etc. For the US Air Force, this is handled by the Material Support Division (MSD), and data is frequently available on how much specifically is charged by the MSD for logistics overhead costs. So, for items that don't spell out the Logistics Overhead amount, we can use the average amount charged by the MSD. Putting all this information together, we get the following equation for Unit Production Cost:

$$Unit\ Production\ Cost = (MLC\ Unit\ Price - Avg.MSD\ Overhead\ Cost) * 0.85\ (G\&A) * 0.85 \left(\frac{Fee}{Profit}\right) * 0.98\ (Co$$

I should also note that all dollar amounts in the database are tied to a specific date, so costs are normalized to remove the effects of inflation before executing this algorithm. Finally, Unit Production Cost must be tied to an order quantity to account for the effects of learning curves. A very slight learning curve was assumed, due to the fact that these items are replenishment spares which are already far down the learning curve, items are produced in small batches, and electronics manufacturing is a highly automated process. We also must associate UPC with an order quantity. In cases where procurement history data is available, we've used the average order quantity. In other cases, we've taken the average of other items in the same category. Order quantities tend to be small, an average of ~24 for the data we've looked at thus far. This order quantity assumption that sometimes gets applied also has a relatively minor effect on the final numbers due to the very slight learning curve.

Weight of Electronics and Weight of Structure

In determining the item's weight, the data most commonly available for items inthe U.S. Federal Supply Catalog was the Unit Pack Weight, along with a Packaging Type code that contains a description of the packaging (material , type (such as bag or box),durability, etc). Unit Pack Weight includes the Weight of Structure, Weight of Electronics, and the Weight of Packaging. In order to find the weights, we simply need to subtract the weight of packaging, and determine the ratio of structure to electronics.

To remove the weight of packaging, we read through each Packaging Type description, and came up with percentages of the total weight that were likely to be packaging, and applied this factor to find the Unpackaged Item Weight.

A majority of the items analyzed were circuit card assemblies, for which it was simple to estimate the ratio of structure to electronics, as well as to find percentages to use for Packaging Type. For items that were not circuit card assemblies, we used a ratio based upon data for similar types of equipment from projects we've been involved with, or we applied a set of predefined rules of thumb that we have for certain types of equipment based on previous projects. In other rare cases,we simply used an expert's estimate that we believe to be reasonably accurate. In the "Final Calibration Step" section, I will discuss methods we've applied to validate our packaging ratio assumptions from this section.

Determining Other Key Input Values

There are 2 remaining cost drivers that we need to determine values for (Operating Specification and Manufacturing Complexity for Structure) which are relatively straightforward. There was a large database of Technical Characteristics for all the items we used, and many of these characteristics clearly spelled out the Operating Specification. For example, the "End Item Identification" field was frequently filled out, and if it stated the item was intended for an F-16 aircraft, weused an Operating Specification value for "Military Airborne" items. Other characteristics include FSC Application Data, which would describe the item as "Airborne Radar System" or "Ground Station," or it might name a specific equipment type from the Joint Electronics Type Designation System[3], which describes the operating specification in its codification system.Finally, determining Manufacturing Complexity for Structure was often very straightforward, as the structure was usually only the structural components of a circuit card, or the box for various boxes of electronics. In the cases of special items for which the item contained additional structure along with the electronics, we used our Manufacturing Complexity for Structure generator tool that comes with the TruePlanning® hardware estimation model.

Additional Drivers

While many categories of electronics were within a narrow range of complexities, others had a much wider range. In these cases, it didn't make sense to deliver guidance of just a single value. Rather, these categories could have additional characteristics that might explain some of the variation. Some of the characteristics make sense to examine for every category of electronics (such as density, operating specification, and date of technology), while others might be specific to the category they are in (such as frequency for radio frequency electronics). Appendix C shows some example charts demonstrating these relationships.

Operating Specification and Density were found to be consistently good at explaining variance in the data, and have been included as additional drivers of complexity in our guidance. Technology Date's explanatory power seems to change on a case-by-case basis for different electronics categories. When there is a consistent trend, we performed additional research, attempting to find a technical characteristic that might explain the trend. In some cases, older data was removed in favor of less noisy, more recent data. There are also cases with clusters of higher or lower complexity that are associated with specific programs.Clusters indicate the need for further research to understand what is different about the electronics for those programs, which may give clues as to areas for further research on category-specific drivers.

Some items may still have a significant amount of variability, even after adjusting for the characteristics described above. For these items, the variability may be due to different intended uses that require more advanced technology. For example, RF amplifiers can be used in simple walkie talkies (low complexity), in state-of-the-art electronic warfare equipment (high complexity), or anywhere in between. That's why RF Amplifiers have such a large standard deviation, as seen in Appendix A.One of the final adjustments, the "Technology Adjustment," will be included in our guidance, and is designed to account for this effect. The degree of change by adjusting this input will be driven by the standard deviation for that specific electronics category found in our dataset. This allows users to pick how complex the technology is compared to other kinds of equipment within the same category, while ensuring the amount of the adjustment is data-driven and justifiable.

Final Calibration Step

In the previous sections, I described the calibration process used to measure Manufacturing Complexity for Electronics, as well as the data collection and analysis methods that plug into the calibration process.

In executing the process, we came up with a list of complexity values for every category of electronics listed in Appendix A. However, during the process, we made a few assumptions that we were uncertain about, specifically the assumption that manufacturers charge 15% G&A and 15% Fee/Profit, as well as estimations of the weight of item packaging. In this section, I'll discuss methods we used to validate these assumptions.

For each category of electronics, we've operated under the same assumptions for G&A, Fee/Profit, and estimations of the weight of packaging. By using a smaller set of data with known values, we can check these assumptions, which I will demonstrate below. Even if the values we've assumed are incorrect, the **relative** Manufacturing Complexity values of each item should be correct, while the **absolute** values would be all biased in one direction (either too high or too low).

We have a dataset for electronic items that contains high quality, detailed data, but most of it cannot be shared due to non-disclosure agreements(NDA) that we've signed with the owner of the data. While we cannot share this data with our customers, we can use it to check the validity of the results of the Manufacturing Complexity calibration process discussed in this paper. For electronic categories that have entries in both our private and public databases, we can compare the two and gain some interesting insights.

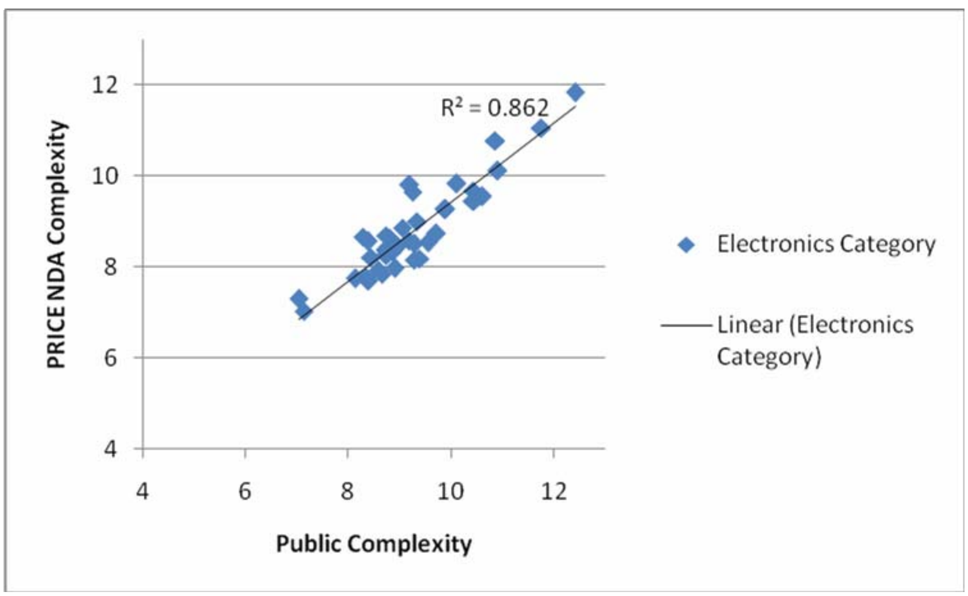

Figure 2: Scatter Plot Comparing Manufacturing Complexity Calibrated from Public vs. PRICE NDA Data

"Public Complexity" is the Manufacturing Complexity for Electronics derived from data in the U.S. Federal Supply Catalog using the calibration process described in this paper.

"PRICE NDA Complexity" refers to the Manufacturing Complexity for Electronics derived from high quality electronics data in PRICE's internal database that is protected by NDA.

Figure 2 demonstrates that the **relative** complexities from the calibration process discussed in this paper are correct, due to the tight correlation when comparing results from the public and PRICE NDA databases. Figure 3, however, shows that the **absolute** values appear to be biased negatively. With no bias, we would expect Figure 3 to show points randomly scattered around the line y=0. However, it appears they are randomly scattered around the line y= -0.293, which indicates our assumptions for G&A, Fee/Profit, and Packaging Type are resulting in complexities that are, on average, 0.293 points too low on the Manufacturing Complexity scale.

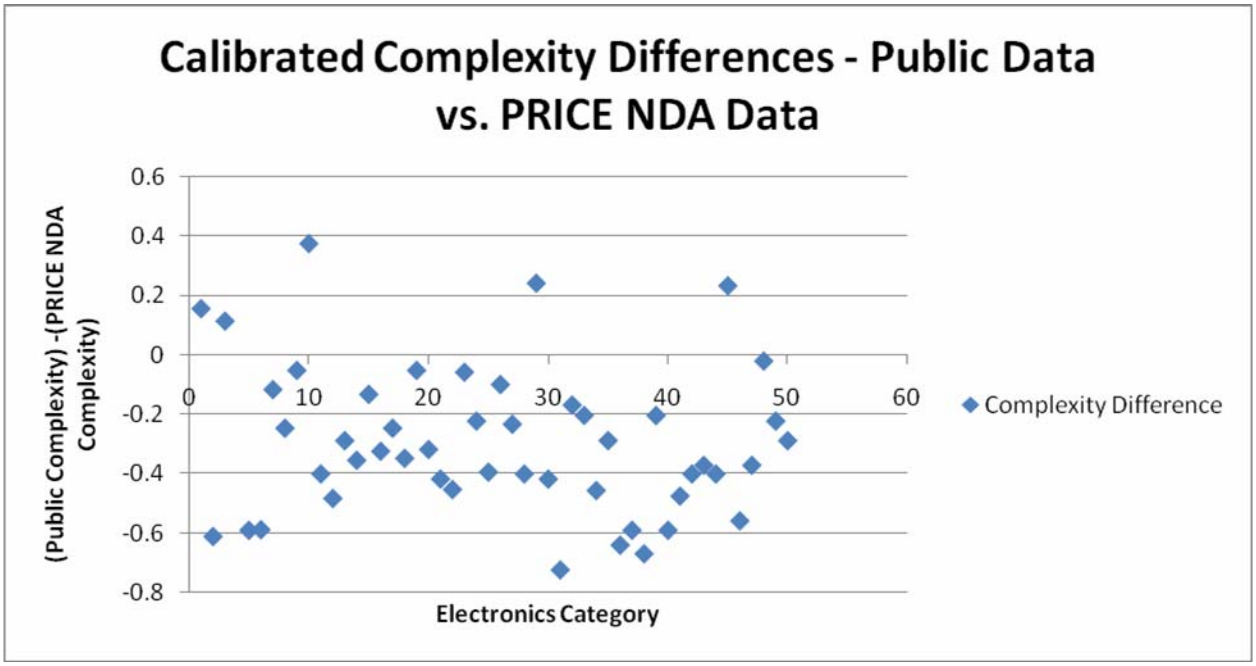

Figure 3: Chart Showing Complexity Differences Between the Public and PRICE NDA Databases for Matching Electronic Categories

Moving forward with the analysis, we can further examine the data to see if any packaging types are consistently biased when compared to other packaging types, and adjust the Packaging Weight Ratio assumption discussed earlier. In other cases, the values for G&A and Fee/Profit may be to blame for biased results, and these can be adjusted as well. The amount of these adjustments will be such that, when reanalyzed, this chart will show points randomly spread around the line y = 0. This will effectively remove any bias based on uncertainties in the assumptions, yielding Manufacturing Complexities that are correct both relatively and absolutely.

DefendingYour Estimate

The results of this research project will be made available to users of the TruePlanning® cost estimation software, implemented in our Hardware models. As the aerospace and defense community places more emphasis on making defendable estimates, we wanted structure this solution so that our complexity guidance can hold up to a high level of scrutiny.

Let’s take an example of a cost estimate for an aircraft that contains some avionics. The Product Breakdown Structure (PBS) might look something like this:

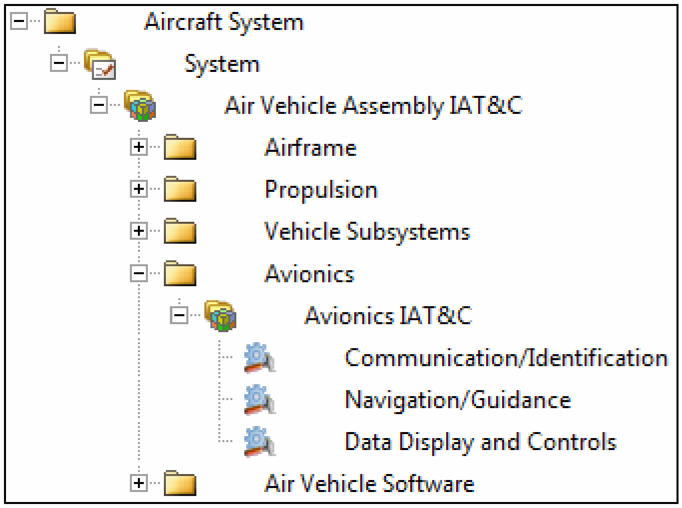

Figure 4: Aircraft System – Product Breakdown Structure (PBS) Example

For each electronic item, the user describes the equipment by entering cost drivers such as Quantities, Weights, Operating Specification, and Schedule based on information from the Project Office, Weight Statements, and other official sources. Other drivers are entered to further describe the scenario, such as Percent of New Design and others. Manufacturing Complexity for Electronics is chosen using the guidance in the model, based on the research presented in this paper. Once cost drivers are entered for each item in the PBS, we run the model and produce a cost estimate.Then, the scrutiny begins, and managersor auditors start asking "How did you estimate that item?"

Defending the inputs such as quantities, weights, and schedule is easy enough, as they came from the project office or weight statements. For Percent of New Design, you can show how the equipment being modeled is similar to existing items, proving you can reuse portions of the design. The difficulty often comes in defending the value for Manufacturing Complexity for Electronics. With this new guidance, the answer is generated for you by TruePlanning®:

"The complexity value is based on data for analogous items. The data comes from publically available sources, including the DLA, government procurement history databases, etc. This data was mined, cleansed, categorized, and normalized by the cost researchers at PRICE Systems. The principal cost estimating relationship is a cost per pound curve from similar equipment. For the "Data Display and Controls" equipment, there are 10 analogous items.Here is a graph showing the estimate relative to the analogous data, generated by TruePlanning.®"

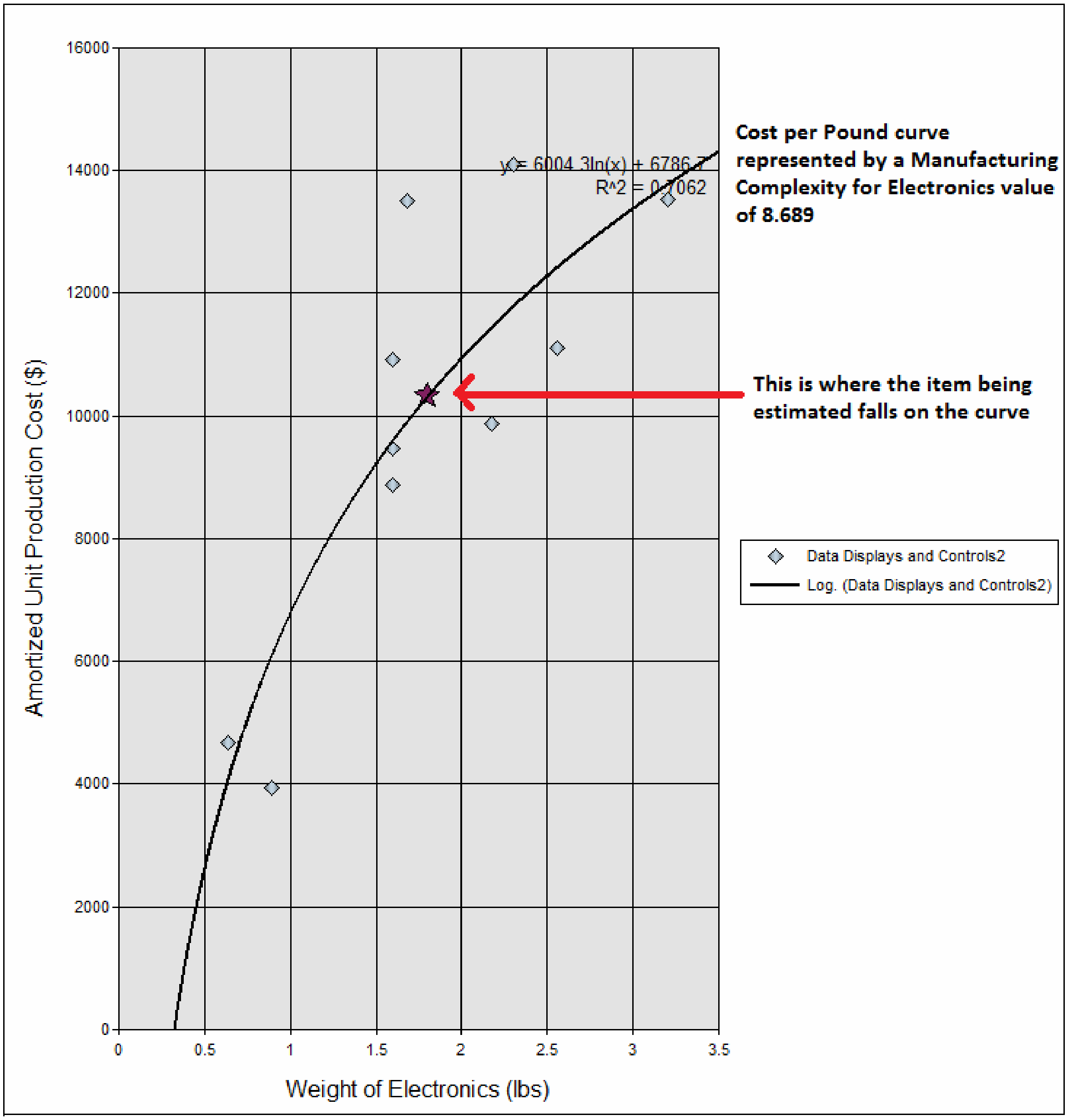

Figure 5: Avionics Data Displays and Control - Estimated Cost vs Analogous Item Costs

Conclusion

This effort is still underway as we continue adding to the Equipment Type list, and recalibrating the uncertain assumptions as necessary. Final calibration values will be adjusted as we collect more data for these additional categories, and as new data becomes available for existing categories. However, the proof of concept is complete, the preliminary results (in Appendices A and B) look very promising, and we are well on our way to finalizing and publishing the new guidance in the next release of our TruePlanning® cost estimation software.

Working with public data that doesn't match your ideal granularity, comprehensiveness and format can be challenging. However, there is often much information that can be gleaned from this kind of data. In this case, the relative complexities could be extracted from the publicly available data, while a separate dataset was used to calibrate to the correct absolute complexities. By striving to understand the context of the data, and extracting the useful information contained within while also addressing any missing links with other methods, you can often arrive at a solution.

Bibliography

[1] IHS Haystack Gold – Main Website.http://www.ihs.com/products/product-design-sourcing/component-supplier-data/haystack-military-government.aspx. 3/25/2013.

[2] NATO Codification System. http://www.nato.int/structur/AC/135/ncs_guide/english/e_index.htm. 3/25/2013

[3] Joint Electronics Type Designation System description. http://en.wikipedia.org/wiki/Joint_Electronics_Type_Designation_System. 3/25/2013.

Appendix A – Electronic "Function" Categories – Preliminary Results*

Category	Num Data Points	Manufacturing Complexity for Electronics**	Standard Deviation
RF Receiver	71	5.672997996	0.723855617
RF Transmitter	96	5.779296455	0.881168088
Voltage Regulator	75	5.849781327	0.816859457
Audio Amplifier	27	5.98050853	1.079334506
Oscillator	114	6.159627701	0.862932123
Ultrasonic Transducer	10	6.169872899	0.453461128
Power Converter	5	6.174941324	0.264359678
Power Supply	153	6.268657878	1.187093085
Video Display	7	6.318298511	0.632168127
EO/IR Sensor/Detector	11	6.58102463	0.479291602
Power Conditioner	10	6.701217884	1.267998067
Mixed D/A Video Processor	30	6.772393781	0.74327255
Digital Signal Processor	44	6.877437875	0.827833222
Video Driver	1	6.951321863	
MPU/MCU	13	7.200749417	1.086186619
RF Exciter	1	7.365183105	
EO/IR Processor	1	7.434485199	
Audio Switch	3	7.596551933	0.253987363
EO/IR Emitter	8	7.733841358	0.589875086
Video Detector	4	7.995106963	0.562747374
RF Amplifier	86	7.995895265	1.192718174
Digital Memory Board	5	8.006249797	0.194194092
Audio Processor	3	8.181145382	0.38090207
RF Signal Processor	10	8.565291915	0.778118449
Frequency Standard	26	8.680046037	0.727628014
Analog Video Processor	3	8.955729797	0.387523107

*****Preliminary Results (unadjusted by Final Calibration Step) for Electronic Function Categories.

** Manufacturing Complexity for Electronics normalized to operating specification of 1.4 (Military Ground)

Appendix B - "Equipment Type" Categories – Preliminary Results*

Category	Num Data Points	Manufacturing Complexity for Electronics**	Standard Deviation
Avionics - Flight Control	58	5.739140365	0.541916428
Avionics - Navigation	395	6.360869547	1.046959153
Avionics - Electronic Support Measures	3	6.499198852	0.393015743
VHF Radio	2	6.556011963	0.055579353
Avionics - Communication	424	6.567045431	0.924500274
Avionics - Dipping Sonar	29	6.583571921	0.858323284
Avionics - HUD - Heads Up Display	12	6.765909209	0.632469119
Data Link	7	6.846399658	0.213473396
Avionics - Weather Radar	7	6.934387821	1.033023435
GPS User Equipment	17	6.990838713	0.574004548
Telemetry	5	7.085089329	0.844089424
Avionics - Flight Recorder	27	7.246000139	0.54533827
Avionics - FLIR - Forward Looking Infrared	3	7.462099254	0.459293222
Ships - Sonar	166	7.479608536	1.189639096
Avionics - Radar	50	7.541975958	0.878339668
Ships - Radar	16	7.571304234	0.81085177
GPS Control Systems	3	7.683280627	0.183400612
Avionics - Displays	10	7.743040631	0.599256931
Avionics - Defensive Aid System	6	7.844803839	1.071828441
Compass	7	7.846375951	0.71239854

***Preliminary Results (unadjusted by Final Calibration Step) for Electronic Equipment Type Categories.

** Manufacturing Complexity for Electronics normalized to operating specification of 1.4 (Military Ground)

Appendix C

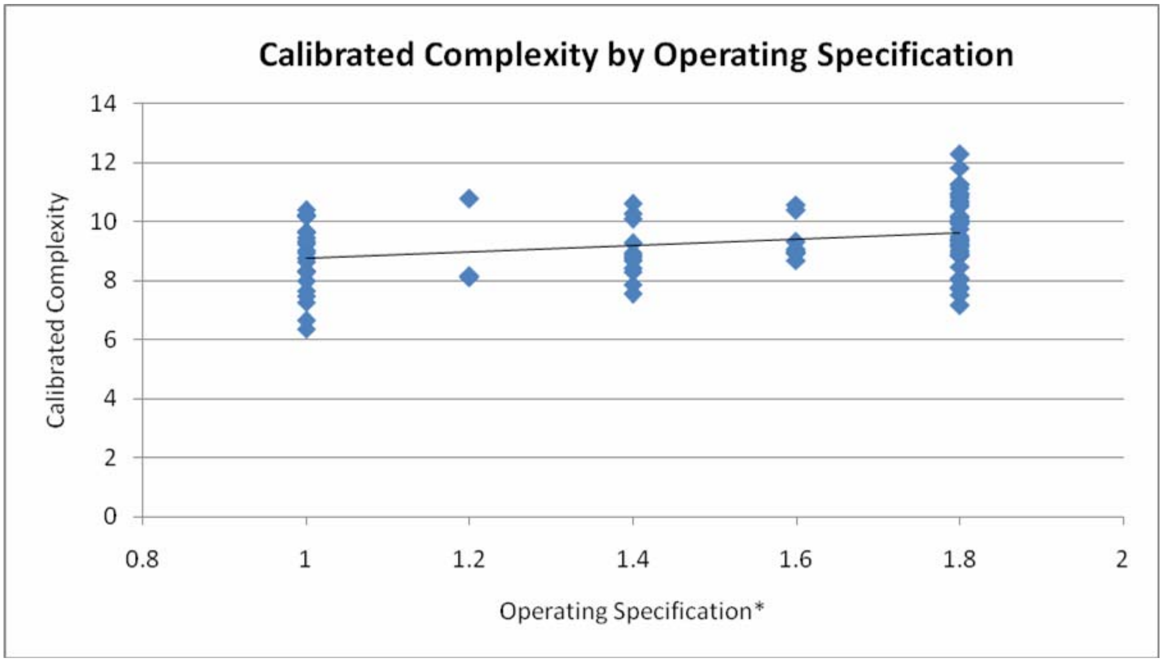

Figure C1: Example - Calibrated Manufacturing Complexity for Electronics by Operating Specification.

*Operating Specification value of 1.8 equates to Military Airborne, 1.6 Military Ships, 1.4 Military Mobile, 1.0 Military Fixed Ground

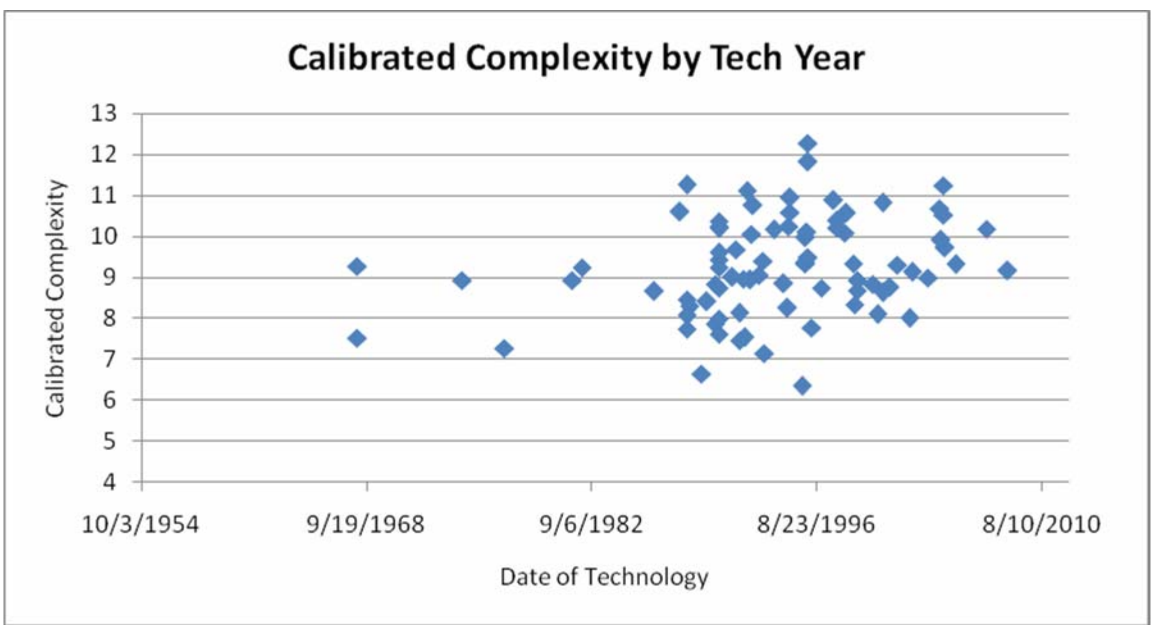

Figure C2: Example - Calibrated Manufacturing Complexity for Electronics by Date of Technology.

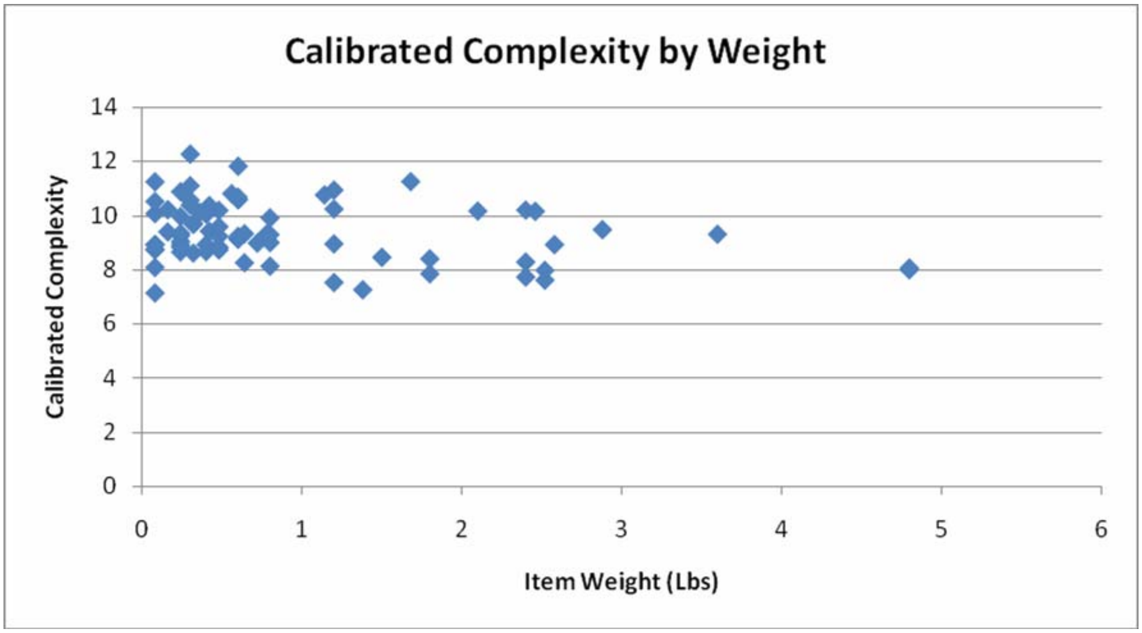

Figure C3: Example - Calibrated Manufacturing Complexity for Electronics by Item Weight.