2

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

THE USE OF STATISTICAL MEASURES TO VALIDATE SYSTEM DYNAMICS MODELS

by

Todd David Stephan

March 1992

Thesis Advisor: Tarek K. Abdel-Hamid

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b. OFFICE SYMBOL (If applicable) 37	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943	7b. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5006
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Naval Postgraduate School 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
The Use of Statistical Measures to Validate System Dynamics Models

12. PERSONAL AUTHOR(S)
Todd David Stephan

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, month day)	15. PAGE COUNT
Master's thesis	FROM______TO______	March, 1992	75

16. SUPPLEMENTARY NOTATION
The views expressed in this paper are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Software development; model validation; simulation

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

System dynamics modelers have often been criticized for their informal methods of model validation and for not using more formal, quantifiable measures to lend confidence to the validation process. Numerous proponents of the system dynamics approach have highlighted this shortcoming, however, and have suggested a variety of appropriate statistical measures which could be used in the model validation process.

The objective of this thesis is to complement earlier validation efforts of the Abdel-Hamid and Madnick System Dynamics Model of Software Development by submitting the model to a battery of appropriate statistical measures. The model is evaluated with statistics which have been used by others in the system dynamics field. The evaluation makes two different comparisons. First, an evaluative comparison is made between data generated by the model and actual data of two real software projects. Then, an evaluative comparison is made between model generated data and data obtained by direct experimentation for two different experiments, using the model's gaming interface. The two evaluations serve to promote confidence in the model.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSICIATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Abdel-Hamid	22b. TELEPHONE (Include Area Code) (408) 646-2686	2c. OFFICE SYMBOL AS/Ah

The Use of Statistical Measures to Validate System Dynamics Models

by

Todd David Stephan
Captain, United States Marine Corps
BA, Washington State University, 1982

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS

From the

NAVAL POSTGRADUATE SCHOOL
March 1992

Author: Todd David Stephan
Todd David Stephan

Approved by: T. Hamid
Tarek K. Abdel-Hamid, Thesis Advisor

Linda Gorman
Linda Gorman, Second Reader

David R. Whipple, Chairman
Department of Administrative Sciences

ABSTRACT

System dynamics modelers have often been criticized for their informal methods of model validation and for not using more formal, quantifiable measures to lend confidence to the validation process. Numerous proponents of the system dynamics approach have highlighted this shortcoming, however, and have suggested a variety of appropriate statistical measures which could be used in the model validation process.

The objective of this thesis is to complement earlier validation efforts of the Abdel-Hamid and Madnick System Dynamics Model of Software Development by submitting the model to a battery of appropriate statistical measures. The model is evaluated with statistics which have been used by others in the system dynamics field. The evaluation makes two different comparisons. First, an evaluative comparison is made between data generated by the model and actual data of two real software projects. Then, an evaluative comparison is made between model generated data and data obtained by direct experimentation for two different experiments, using the model's gaming interface. The two evaluations serve to promote confidence in the model.

TABLE OF CONTENTS

I. INTRODUCTION

A. BACKGROUND

System Dynamics modelers have been criticized for their qualitative, informal methods of model evaluation and for not utilizing quantitative, objective measures of model validation. As stated by Sterman "...the validity of system dynamics models is often questioned even when their correspondence to historical behavior is quite good...the failure to present formal analysis of historical behavior creates an impression of sloppiness and unprofessionalism." (Sterman, 1984, p. 51) Numerous proponents of the system dynamics approach have highlighted this shortcoming, however, and many have suggested various means to tackle the problem (Barlas 1989, Forrester and Senge 1980, Naylor 1971, Rowland 1978, Sterman 1984).

There is, however, an even more basic issue that warrants discussion before specifically addressing the problem of validation. The issue, as discussed by Barlas and Carpenter (1990), is a result of two differing philosophies of science, the traditional logical empiricist philosophy and the more recent relativist philosophy. Where the logical empiricist "...assumes that knowledge is an objective representation of reality and that theory justification can be an objective formal process." (Barlas an Carpenter, 1990, p. 148) And the relativist advocates "...knowledge is relative to a given society, epoch, and scientific world view. Theory justification is therefore a semiformal, relative social process." (Barlas and Carpenter, 1990, p. 148) The authors argue that the relativist philosophy is the applicable philosophy to hold for the system dynamics methodology in the context of model validation. The relativist philosophy has a certain appeal, in that the empiricist would espouse a given model to be an objective, absolute representation of reality and as such, the model could be empirically evaluated as being true (or false) (Barlas and Carpenter 1990). The relativist would view a given model as only one of many ways to portray reality, with no model being able to claim absolute objectivity, although one model may be more effective than another (Barlas and Carpenter 1990). Those who are familiar with and use the system dynamics methodology could equate easily to the relativist viewpoint.

The validation of a system dynamics model is, thus, not a simple matter of subjecting a model to some standard set of classic statistical tests. As pointed out by Barlas "System Dynamics models have certain characteristics that render standard statistical tests inappropriate." (Barlas, 1989, p. 59) This does not mean that the validation process for a system dynamics model should be solely qualitative. It means that a system dynamics modeler needs to employ tests, both quantitative and qualitative, that can serve to evaluate a given model.

As stated by Forrester and Senge, "There is no single test which serves to 'validate' a system dynamics model. Rather, confidence in a system dynamics model accumulates gradually as the model passes more tests and as new points of correspondence between the model and empirical reality are identified." (Forrester and Senge, 1980, p. 209) This point is emphasized by many in discussions of model validation (see for example Barlas and Carpenter 1990, Richardson and Pugh 1981, Sterman 1984). The consensus is that validating system dynamics models, should imply a continuous cycle of confidence building tests throughout the iterative development of a model. In essence, the utility of a simulation model depends upon the confidence that the model users have in the model. Each test should not serve as an end in itself, but merely as one of many steps which serve to build that confidence.

Richardson and Pugh, address the issue of model validity in several different perspectives. The first of those issues involves validity and model purpose "...it is meaningless to try to judge validity in the absence of a clear view of model purpose." (Richardson and Pugh, 1981, p. 310) Richardson and Pugh also discuss model validity in terms of a model's suitability and consistency. In doing so they pose two questions: "Is the model suitable for its purposes and the problem it addresses?" and "Is the model consistent with the slice of reality it tries to capture?" (Richardson and Pugh, 1981, p. 312) Since no model can claim absolute truth, the best that can be hoped for is that the model be suitable for its purpose and consistent with reality.

B. PURPOSE OF THESIS RESEARCH

The focus of this thesis is the evaluation of the ability of the software development system dynamics model developed by Abdel-Hamid and Madnick (1991) to satisfactorily match the historical data of the system it was designed to model. Sterman (1984) described the evaluation of a model's historical fit as a weak test by itself, while noting that "Failure to satisfy a client or reviewer that a model's historical fit is satisfactory is often sufficient grounds to dismiss the model and its conclusions." (Sterman, 1984, p. 52) systems (Rowland and Holmes 1978, Senge 1973). These tests will be applied to the Abdel-Hamid and Madnick model (Abdel-Hamid and Madnick 1991) in four different cases which compare actual output to the model's output. A description of each test follows:

1. Mean-Square-Error (MSE) Test

The mean-square-error (MSE), a measure of forecast error, is defined as:

$$\frac{1}{n} \cdot \sum_{t=1}^{n} (S_t - A_t)^2$$

where

n = Number of observations ($t = 1, \ldots, n$)

S_t = Simulated value at time t

A_t = Actual value at time t

The MSE measures the deviation of the simulated variable from the actual value over a given time period. The advantages of this measure are that large errors are weighted more heavily than small ones and that errors of opposite sign do not cancel each other out (Sterman 1984). By taking the square root of the MSE, the forecast error can be put into the same units as the variable in question. This measure is referred to as the root-mean-square (RMS) simulation error (Pindyck and Rubenfield 1991).

2. Root-Mean-Square Percent Error (RMSPE) Test

A more convenient measure of forecast error is the root-mean-square percent error (RMSPE), which provides a normalized version of the error and is defined as:

$$\sqrt{\frac{1}{n}\cdot\sum_{t=1}^{n}\left[\frac{(S_t - A_t)}{A_t}\right]^2}.$$

This also measures the deviation of the simulated variable from the actual value over a given time period, but puts it into percentage terms (Pindyck and Rubenfield 1991).

3. Theil Statistics Test

The MSE and the RMSPE measure the size of the total error between the actual and the simulated data. The MSE can also be decomposed into the Theil statistics (Sterman 1984, Pindyck and Rubenfield 1991) to assist in revealing the sources of the error. The sources of error are given in terms of bias, variance, and covariance. The decomposition of the MSE into the Theil statistics is as follows:

$$\frac{1}{n}\sum_{t=1}^{n}(S_t - A_t)^2 = \left(\overline{S} - \overline{A}\right)^2 + (S_S - S_A)^2 + s(1-r)S_S \cdot S_A$$

where:

$$\overline{S} = \frac{1}{n}\sum S_t$$

and

$$\overline{A} = \frac{1}{n}\sum A_t.$$

S_S and S_A are the standard deviations of S and A and

$$S_S = \sqrt{\frac{1}{n}\sum\left(S_t - \overline{S}\right)^2}$$

and

$$S_A = \sqrt{\frac{1}{n}\sum\left(A_t - \overline{A}\right)^2}.$$

r is the correlation coefficient between simulated and actual data:

$$\frac{\frac{1}{n}\sum\left(S_t - \overline{S}\right)\left(A_t - \overline{A}\right)}{S_S \cdot S_A}.$$

By dividing each of these proportions by the MSE, the inequality proportions U^M, U^S, and U^C are derived as

$$U^M = \frac{\left(\overline{S} - \overline{A}\right)^2}{\frac{1}{n}\sum\left(S_t - A_t\right)^2},$$

$$U^S = \frac{\left(S_S - S_A\right)^2}{\frac{1}{n}\sum\left(S_t - A_t\right)^2},$$

$$U^C = \frac{2(1-r)S_S S_A}{\frac{1}{n}\sum\left(S_t - A_t\right)^2}.$$

The proportions U^M, U^S, and U^C represent the amount of error in the MSE due to bias, variance, and covariance respectively. Note also, that $U^M + U^S + U^C = 1$ as the sum of the three represents the total MSE.

Bias (U^M) measures the degree to which the average values of the simulated and actual values differ. In conventional statistical terms, an estimate is biased if estimates are made repeatedly and the mean for those estimates does not approach the actual value of the parameter, as the number of estimates grows (Bush and Mosteller 1955, p. 199). Therefore it is more appealing that a model's estimates be unbiased, that is, the expected value of the estimator approaches the population value, as the number of sample estimates increases. Large bias (indicated by large U^M and a large MSE) is an indicator of systematic error between the model and reality and could be potentially troubling. Systematic error may indicate that there is some variable or parameter in the real system which is not reflected correctly in the model. It is unlikely that a model which adequately reflects reality would produce these results. Bias errors could indicate specification of parameter errors within the model. On the other hand, not all bias errors are detrimental to a model. This could be the case if U^M is large but the size of the error itself is small (small MSE/RMSPE) or there are acceptable simplifying assumptions present. As stated previously, if an error is systematic, even if it is large, it may still be acceptable provided that it does not compromise the purpose of the model. "In terms of testing the validity of a model...a model should have predictive power, it should be able to forecast...the degree of precision being sufficient if increased accuracy did not lead to different conclusions." (Bloomfield 1986, p. 94) If a closer goodness-of-fit does not serve to provide the user of the model with a clearer understanding of the software development process, then confidence should not be adversely affected. It may still be prudent, however, for the modeler to re-examine the parameters impacting that variable.

The variance proportion (U^S) measures how well the model's estimate matches the degree of variability in the actual value. For instance, a large U^S suggests that the simulated series has fluctuated considerably while the actual series has fluctuated very little, or vice versa. A large variance proportion may also be an indicator of a systematic error.

The covariance proportion (U^C) measures the unsystematic error (the error remaining after deviations from average and average variabilities have been accounted for). This portion of error is the least troublesome of the three. Unsystematic error suggests that an exogenous event influenced the system behavior. The presence of unsystematic error does not compromise a model's ability to suit its purpose, as it is not within a model's scope to forecast based on random external noise. To do so could defeat the purpose for which a model is intended.

4. Theil Inequality Coefficient Test

The final test which will be employed is the Theil Inequality Coefficient (Rowland and Holmes 1978, Theil 1961, 1966). The inequality coefficient is defined as:

$$U = \frac{\sqrt{\frac{1}{n}\sum (S_t - A_t)^2}}{\sqrt{\frac{1}{n}\sum S_t^2} + \sqrt{\frac{1}{n}\sum A_t^2}}.$$

The inequality coefficient (U) will always be between 0 (perfect predictions) and 1 (worst predictions).

Of course, with these tools in hand, one must then ask what defines an acceptable level of goodness-of-fit in order to instill confidence in the model. Research into this area of study has shown that within the software development field, there is no standard of acceptable tolerance that a model of this nature should adhere to, for it to be deemed "valid" or acceptable. In general, however, these tests can effectively build confidence in a model (Barlas 1989, Rowland and Holmes 1978, Sterman 1984) if:

(1) Errors are small (RMSPE less than 10%) and unsystematic (concentrated in U^S and U^C). (Sterman 1984) An RMSPE of 10% is used as the guideline for an acceptable tolerance level in this study and is derived from two sources. The first is Sterman (1984, p. 56) "The RMS percent errors are below ten percent...While the small total errors in most variables show the model tracks the major variables, the several large errors might raise questions about the internal consistency of the model or the structure controlling those variables." While not explicitly stated, an acceptable error tolerance level of 10% is implied within his analysis. The other basis is from Veit (1976 p. 540) "Generally speaking, if the model can reproduce the historical values of key variables within 10% then the structure of the model is probably sound. In other words, all of the variables and sectors are linked together in such a way that the model is a fair representation of the real world...If the structure of the model is correct, it will vary the values of the variables at variable rates over time in such a way that they reproduce historical data fairly closely."

(2) Large errors, but due to excluded modes, simplifying assumptions, or noise in historical data, such that the nature of the error does not adversely impact the model's purpose. (Sterman 1984)

(3) The Theil Inequality coefficient is less than 0.4, "...one may arbitrarily identify TIC values above 0.7 as corresponding to rather poor models, TIC values between 0.4 and 0.7 for average-to-good models, and TIC values below 0.4 as very good or excellent models." (Rowland and Holmes 1978, p. 40)

The Quattro Pro 3 spreadsheet application, by Borland International, was used to compute the statistics. A representative spreadsheet layout for each formula presented and analyzed is given in Appendix A. This analysis will use these statistics to form the basis of the model evaluation.

II. THE ABDEL-HAMID MADNICK SYSTEM DYNAMICS MODEL OF SOFTWARE DEVELOPMENT

A. MODEL PURPOSE

The software systems development model by Abdel-Hamid and Madnick, is based on the feedback principles of system dynamics (Abdel-Hamid and Madnick 1991). The purpose of the model is to serve as a vehicle which "...enhances our understanding of, provides insight into, and makes predictions about the process by which software development is managed...intended to provide a general understanding of the nature of the dynamic behavior of a project (e.g., how work force level and productivity change over time and why) rather than to provide point-predictions (e.g., exactly how many errors will be generated.)" (Abdel-Hamid and Madnick, 1989, pp. 1426-1437). Through this model, the developers have endeavored to provide a means by which managers and researchers, can gain a better understanding of the managerial side of the software development process. This has proven to be a complicated process, which is yet to be fully understood or comprehended, by both academia and management professionals.

For this model to accomplish its purpose, it must reasonably portray a given software development project as it would actually unfold under given management policy decisions and situations. Users of the model must also have an acceptable degree of confidence in the model's forecasting ability. However, the model's purpose is not to make point predictions or to derive an optimal solution to a given situation. Rather, it is to gain understanding and insight into the complex process of managing software projects.

The engineering functions of software development have experienced significant advances in recent years. Improvements in areas such as structured programming, structured design, formal verification, language design for more reliable coding, and diagnostic compilers continue to be introduced to the field (Abdel-Hamid and Madnick 1989). In contrast, the managerial side of software development has received relatively little attention from researchers (Abdel-Hamid and Madnick 1989). This dearth of research may certainly be a contributing factor to the managerial problems which characterize the software industry today. As stated by Brenton R. Schlender "...software remains the most complex and abstract activity man has yet contrived." (Schlender 1989, p. 112) This model also serves to broaden the range and scope of research which has been conducted in the somewhat brief history of software development.

B. MODEL DEVELOPMENT AND STRUCTURE

The model was developed from an extensive field study of software project managers in five organizations. The study consisted of three information gathering steps (Abdel-Hamid and Madnick 1989). The first step involved a series of interviews with software project managers at three organizations. From the information gathered in this phase and from the modelers' own experience in software development, a skeleton of a system dynamics software development model was established. The next step was an extensive literature review, which served to fill many knowledge gaps and resulted in a more detailed model. The final step was another round of intensive interviews with software project managers at three organizations. In this round of interviews, only one of the three project managers was from the initial interview group.

From these three steps, a highly detailed, quantitative simulation model was developed which integrates managerial activities (e.g., planning, controlling, and staffing) with software production type activities (e.g., design, coding, reviewing, and testing). The model contains over one hundred causal links and four major subsystems (human resource management, software production, control, and planning). It has been designed for use on medium sized, organic type software projects (i.e., projects that are 10,000 to 250,000 lines of code and conducted in familiar in-house environments). For a detailed discussion of the model's actual structure and formulation, see Abdel-Hamid and Madnick (1989 and 1991).

III. ANALYSIS OF DE-A AND DE-B PROJECTS

A. DESCRIPTION OF PROJECTS

One of the initial model validation efforts for the Abdel-Hamid and Madnick model (Abdel-Hamid Nov. 1990, Abdel-Hamid and Madnick 1989) involved a case study at NASA's Goddard Space Flight Center (NASA was not among the five organizations studied during model development) (Abdel-Hamid Nov. 1990). The case study involved the simulation of two separate software projects at NASA, the DE-A and DE-B projects. The validation procedure used a graphical comparison of actual data against the model's data. Both projects were designed for the purpose of designing, implementing, and testing software systems for processing telemetry data and providing attitude determination and control for NASA's DE-A and DE-B satellites. The development and target operations machines for both projects were the IBM S/360-95 and-75, and the programming language was FORTRAN. Initial project estimates and actual results are given in Table 3-1.

TABLE 3-1

DEA-A		**Initial Estimates**	**Actuals**
Size	(DSI)	16,000	24,400
Cost	(man-days)	1,100	2,239
Schedule	(working days)	320	380
DEA-B			
Size	(DSI)	19,600	25,700
Cost	(man-days)	1,345	2,200
Schedule	(working days)	320	335

(Note: DSI = Delivered Source Instructions)

B. DE-A AND DE-B PROJECT VARIABLES

The analysis of the DE-A and DE-B projects involves a comparison of three variables (SCHEDULE estimate, WORKFORCE size, and cost in MAN-DAYS) in terms of actual project results versus the model's results. The variable comparisons are made at different time intervals throughout the projects' lifecycles. The reason for comparison at different time intervals vice comparing just the final outcome, is that the model's purpose is to gain an understanding of the entire software development process, not just the final result.

The SCHEDULE variable is an estimate of how long it will take to complete the project from start to finish. For example, on day 40 after the project had commenced, the project managers estimated that the project would be complete on the 320th day of elapsed time, whereas on day 280, they had revised the completion day to the 330th day. Thus, the analysis of the SCHEDULE variable is a comparison of the project managers' actual estimated schedule completion time versus the model's estimated schedule completion time. The WORKFORCE variable represents the desired staffing level at a given time (comparison of the actual number staff desired vs. model generated). The MAN-DAYS variable is a measure of the project's accumulated cost (in man-days) at a given time (comparison of the actual cost vs. model generated).

1. DE-A Project Results and Analysis

The input data tables used to calculate the statistics for the actual results and the model results for each of the variables is given in Appendix B.

Table 3-2 provides the RMSPE, MSE, Theil's Inequality Statistics, and Theil's Inequality Coefficient for each of the DE-A project variables.

TABLE 3-2. ERROR ANALYSIS OF DE-A PROJECT

			INEQUALITY STATISTICS			
Variable	**RMSPE(%)**	**MSE**	U^M	U^S	U^C	**TIC**
SCHEDULE	.98	10.6	.01	.28	.71	.00
MAN-DAYS	9.3	22178	.04	.12	.84	.05
WORKFORCE	17.6	.9	.29	.3	.41	.06

As can be seen from Table 3-2, SCHEDULE and MAN-DAYS have an RMSPE below 10%, while WORKFORCE is above the 10% level. All three variables have a TIC value well below the .40 level.

The SCHEDULE variable shows an extremely low RMSPE (.98%), indicating that the difference between the actual results and the model results is very small. This indicates that the model matched very well with the actual schedule estimates made by the project managers. On average, the model differed from the actual estimates by only three days (square root of the MSE). The decomposition of the MSE into the inequality statistics reveals that the source of the small error was unequal covariance (unsystematic error). As such, the nature of the error is not a major concern since the model's purpose is not point prediction. The two series are plotted in Figure 3-1.

The MAN-DAYS variable shows a 9.3% difference, on average, between the actual cost and the model's forecasted cost over the project's

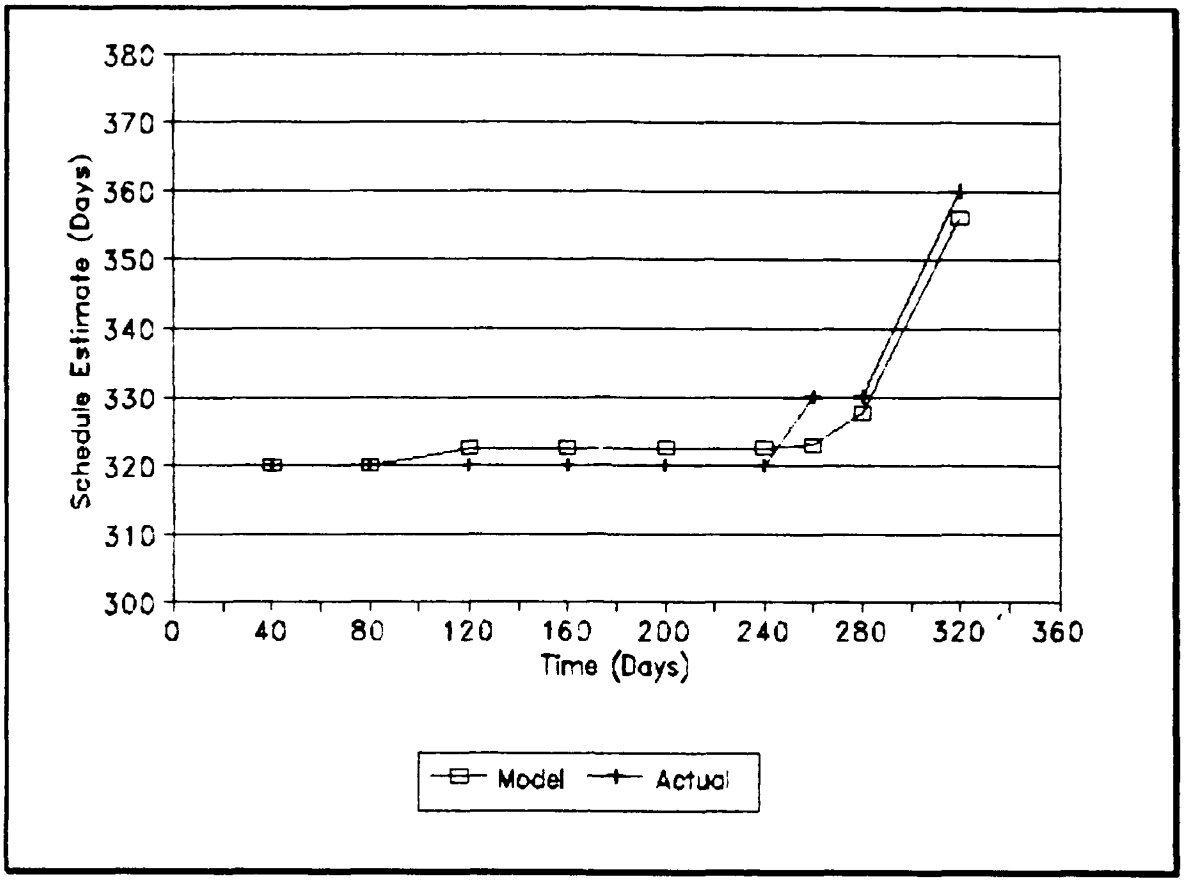

Figure 3-1. DE-A SCHEDULE Actual vs. Model

duration. In absolute terms, this equates to an average difference between the model cost and actual cost of 149 man-days (square-root of the MSE). This of course, is well below the 10% error tolerance level and suggests that structure of the model is sound. The inequality statistics suggest that the majority of the error is unsystematic (e.g., 84% of the error due to covariance), which is quite acceptable. Additionally, the simulated cost trend matches the actual cost trend quite well. This can be seen graphically as well in Figure 3-2, where the point by point differences are obvious, but the general slopes appear to be very close.

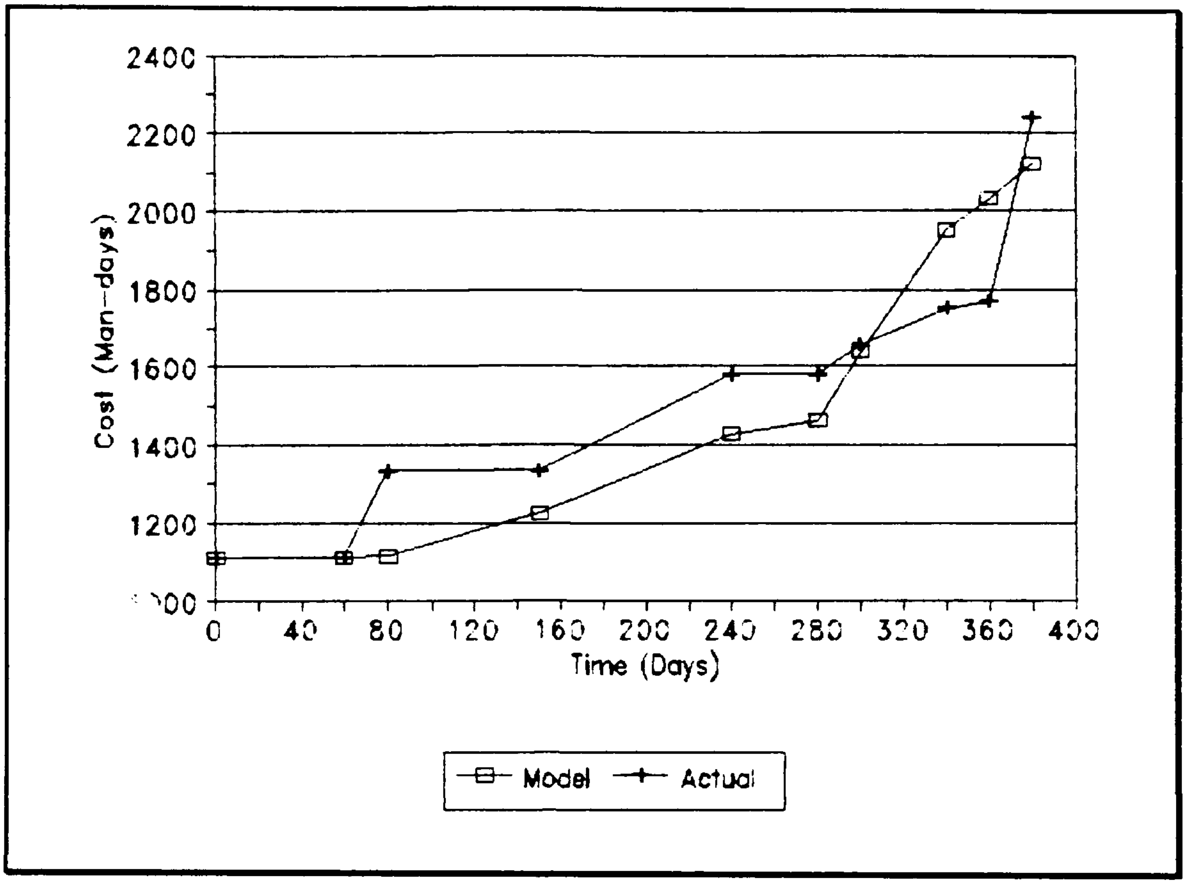

Figure 3-2. DE-A MAN-DAYS Actual vs. Model

The WORKFORCE variable displays the largest RMSPE at 17.6%. In actual terms, the model shows an average difference from the actual workforce size of .95 people over the course of the project's life. The inequality statistics do not indicate that the error is concentrated in any one source. Rather, the error is evenly distributed between the three sources. While the majority of error is in unequal covariance (41%) or unsystematic in nature, it does not dominate. That 29% and 30% of the error is due to bias and variance, respectively, could be of concern because both are potential indicators of systematic error. This could potentially compromise the model's usefulness. In this case however, the trend of the model matches that of the actual data very closely and the difference between the average values of the two series (the error due to bias), is small enough as to not adversely impact the purpose of the model. The reasoning behind this is that the purpose of the model is to provide insight into the dynamic behavior of a project, not point prediction. Any adjustment of the model's parameters to make a closer fit, would not necessarily increase ones ability to glean further insight or understanding (Bloomfield 1986). Therefore, the error in this case is unsystematic with respect to the purpose of the model. Plotting the model results vs. the actual results (Figure 3-3), shows the small differences in the point by point match and highlights the very similar trend pattern of each series.

2. DE-B Project Results and Analysis

The input data tables for the actual results and the model results for each of the variables is given in Appendix C. Table 3-3 provides the calculated RMSPE, MSE, Theil's Inequality Statistics, and Theil's Inequality Coefficient for each data set of the DE-B project variables being analyzed.

TABLE 3-3. ERROR ANALYSIS OF DE-B PROJECT

			INEQUALITY STATISTICS			
Variable	**RMSPE(%)**	**MSE**	U^M	U^S	U^C	**TIC**
SCHEDULE	2.5	64.3	.68	.02	.30	.01
MAN-DAYS	2.8	3405	.07	.28	.64	.02
WORKFORCE	11.0	1.0	.17	.10	.72	.07

As can be seen from Table 3-3, SCHEDULE and MAN-DAYS have an RMSPE well below 10%, while WORKFORCE is above the 10% level, as was the case with the DE-A project. Additionally, all three variables have a TIC value well below the .40 level. Each variable will be discussed separately and analyzed using the inequality statistics.

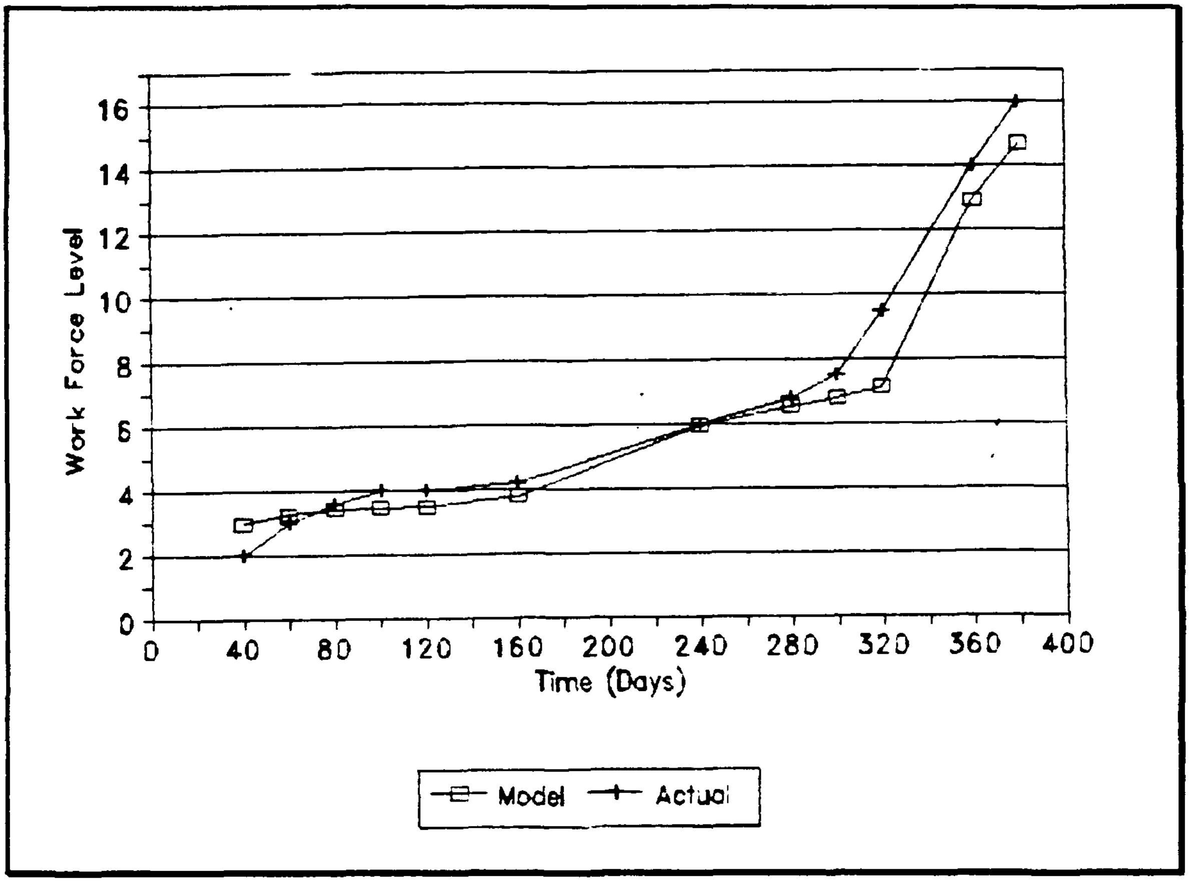

Figure 3-3. DE-A WORKFORCE Actual vs. Model

SCHEDULE shows a very low RMSPE indicating that the magnitude of the error is very small and that the model matched the real system quite well. The inequality statistics reveal that the major source of the error can be attributed to bias, or possibly a systematic difference between the model and reality, which is a potential problem. The graph of the two series (Figure 3-4) shows that the project managers did not adjust their schedule estimates until day 260. According to DeMarco (1982) this is typical "Once an original estimate is made, it's all too tempting to pass up subsequent opportunities to estimate by simply sticking with your previous numbers. This often happens even when you know your old estimates are substantially off. There are a few possible explanations for this effect: It's too early to show slip...If I re-estimate now, I risk having to do it again later (and looking bad twice)...As you can see, all such reasons are political in nature." The model does in fact take this

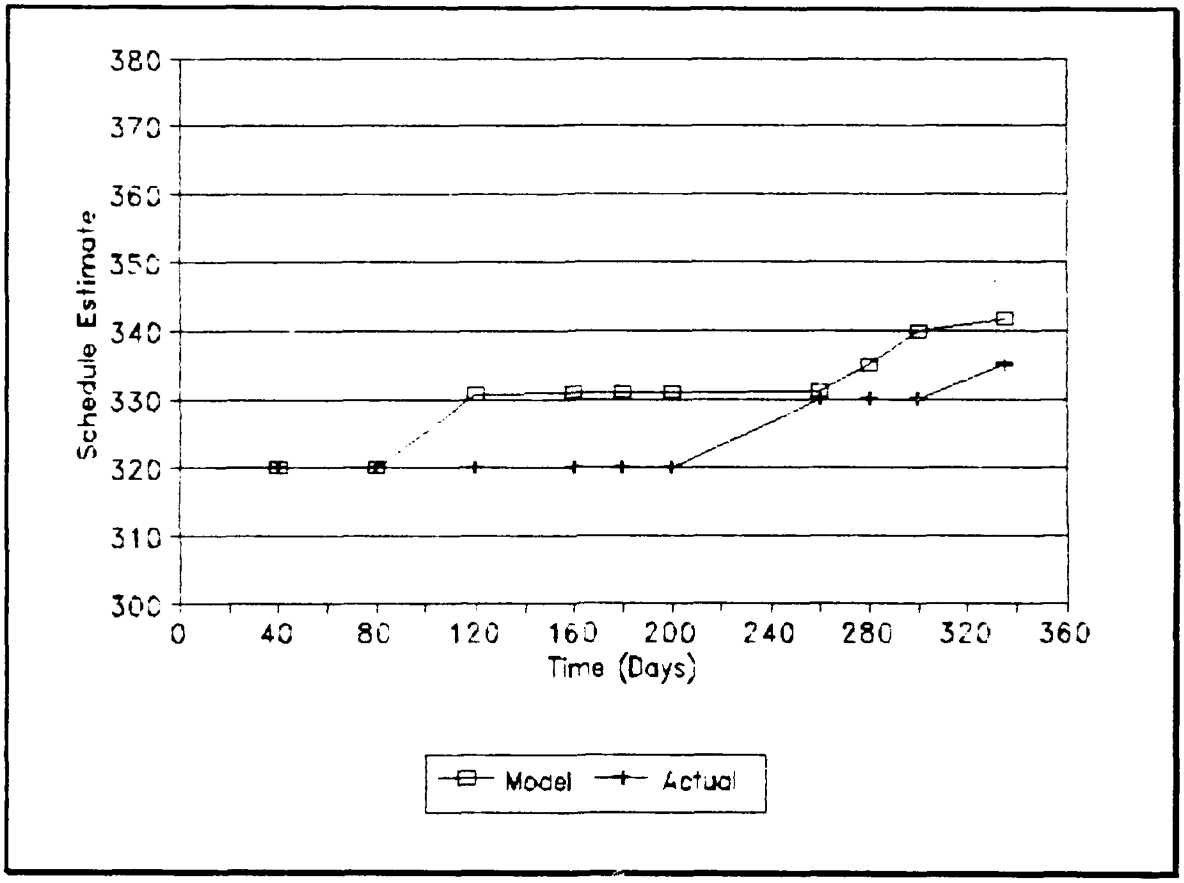

Figure 3-4. DE-B SCHEDULE Actual vs. Model

system component into account and the small error size is evidence of its presence. The model does capture the major portion of this component. The simplifying assumptions in this regard do not jeopardize the model's purpose, as it does not degrade any general understanding of the nature of the SCHEDULE estimates within the system.

MAN-DAYS, like SCHEDULE has a very low RMSPE (2.8%), indicating that the magnitude of this error is also very small and that it also approximates reality quite well. Additionally, the inequality statistics show that the preponderance of the error is concentrated in unequal covariance (64%) and variance (28%). The small and unsystematic error does not in any way detract from the model's ability to serve its purpose. Specifically, the small impact of outside noise does not affect a user's ability to gain insight into the cost structure, reflected by the model's MAN-DAYS variable. A plot of the model vs. the actual cost (Figure 3-5) helps to illuminate the model's ability to match reality for this variable.

As with the DE-A project, WORKFORCE displays the highest RMSPE (11%) of the three variables for the DE-B project. Although in this project the error is not as great as in the DE-A project. The source of the error is concentrated mainly in the unequal covariance proportion (72%) and is an unsystematic type of error. Once again, the model captures the general trend of the real system, even though it varies on a point by point basis (see Figure 3-6). This does not detract in any way from the model's ability to demonstrate the dynamic nature of the work force structure during the project's lifecycle.

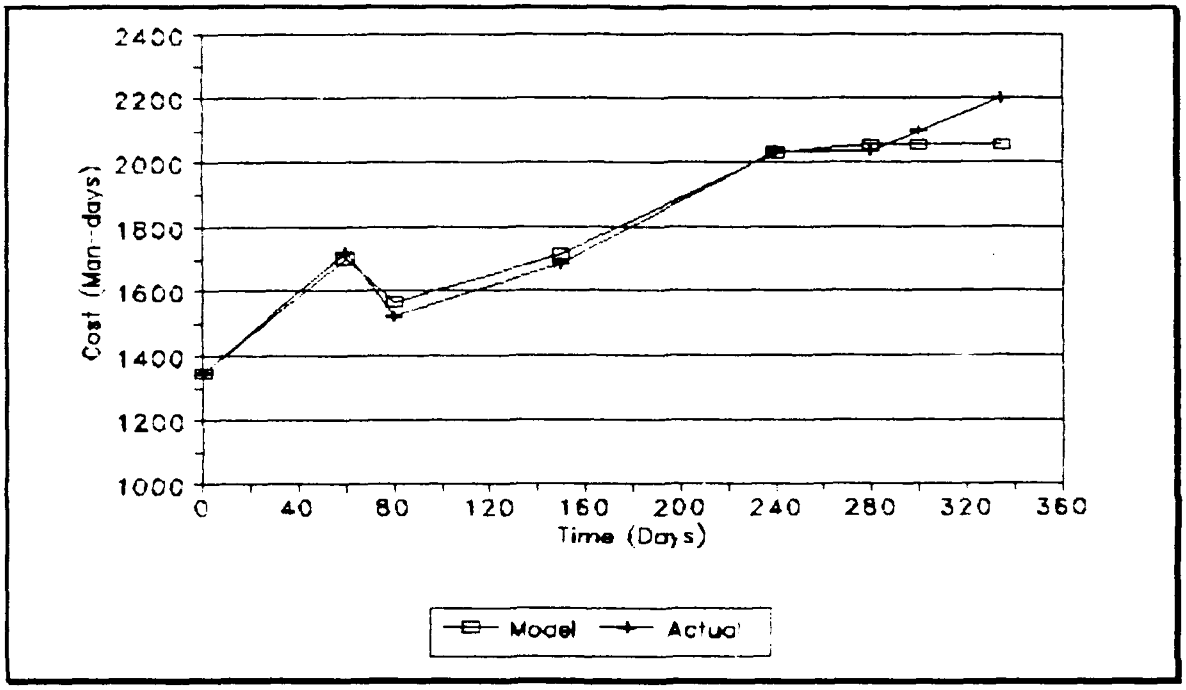

Figure 3-5. DE-B MAN-DAYS Actual vs. Model

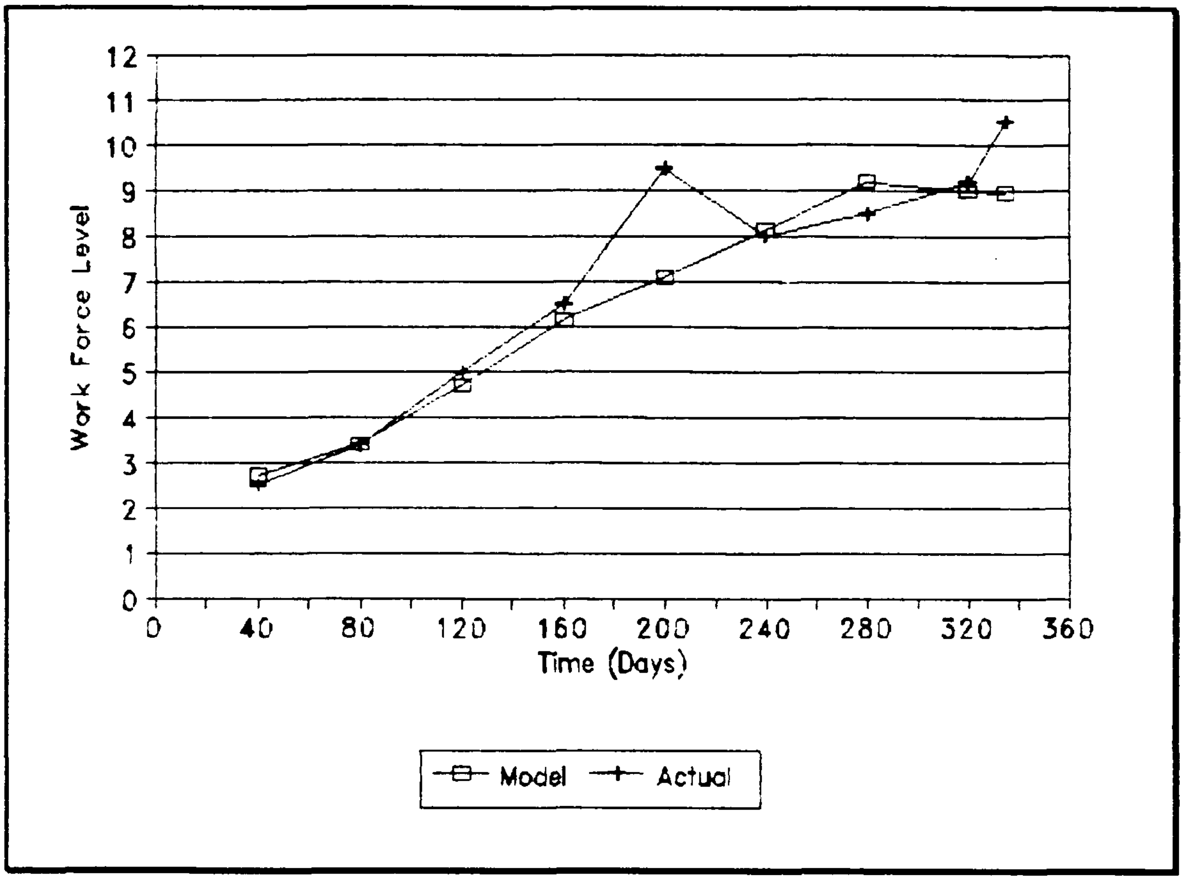

Figure 3-6. DE-B WORKFORCE Actual vs. Model

C. CONCLUSIONS

The final statistical measure which remains to be addressed for both projects is the TIC measure. As can be seen from Tables 3-2 and 3-3, the TIC values for both projects are below .10. All of these values are well below the general guidelines given by Rowland and Holmes, where TIC values below .4 would equate to a very good or excellent model "...one may arbitrarily identify TIC values above 0.7 as corresponding to rather poor models, TIC values between 0.4 and 0.7 for average-to-good models, and TIC values below 0.4 as very good or excellent models." (Rowland and Holmes 1978, p. 40) Thus, the analysis of TIC values for all variables, indicates that the model performs extremely well at portraying the reality of the given software development projects. None of the tests which were conducted were able to detract from the model's ability to suit its purpose, as all errors were either 1) small (RMSPE under 10%) and unsystematic or 2) large errors, but unsystematic and 3) all TIC values were well below .40. Consequently, the foregoing tests should serve only to build confidence in the model's utility towards understanding the software development process.

IV. ANALYSIS OF RONAN'S AND BAKER'S EXPERIMENTS

A. THE DIRECT EXPERIMENTATION METHOD

Another method of comparison which can be used to test a model's goodness-of-fit, besides that of direct comparison with actual historical data, is direct experimentation. Direct experimentation uses an interactive game based on the model being tested. Subjects in the game assume a given role in the system which has been modeled and are required to make a specific decision(s). The subjects are placed in the same decision making setting assumed in the model, they receive the same information set as the model, and try to meet the same goals as the model. The subjects are then free to make their decision in any way that they want. Of course, the decision of the model is based on the explicit rule set contained in the model's structure. It is then possible to compare the decision(s) made by the subjects in the experiment to the decision(s) made by the model. This comparison can be used to confirm or disconfirm the decision rule contained in the model and thus, promote confidence in the model. (Sterman 1987)

The goodness-of-fit tests used in the previous chapter measured the ability of the model to capture three elements of the reality of the system that the model was designed to capture. Direct experimentation provides another measurement of goodness-of-fit, from a somewhat different perspective. The direct experimentation method is a comparison of the subjects' behavior with that of the model's, for given variables or decision rules, within the same environment. The assumptions underlying the model environment must also exist in the subjects experimentation environment. Direct experimentation will not reveal if those assumptions are incorrect. It can, however, be used to promote confidence in the model by showing that "... given the institutional structure people behave the same way the model presumes them to behave." (Sterman 1987, p. 1577) Therefore, direct experimentation can serve as a useful tool for examining the accuracy of a given decision rule, for a given variable's output.

The same statistics used to evaluate goodness-of-fit for the DE-A and DE-B projects (RMSPE, MSE, Theil Inequality Statistics, and TIC), will be used in the following direct experimentation comparisons. Additionally, an alternative method of analyzing the model's ability to match that of the subjects is proposed. This method is based on the work done by Sterman (1987 and 1989) and will be introduced and applied to a subset of Baker's Experiment (Baker 1992). The main focus of the analysis, however, will remain on the previously defined statistics. The proposed computation is as follows:

$$\frac{\sqrt{\sum (A_t - S_t)^2}}{S_t}$$

where

S_t = Simulated (model) value at time t

A_t = Actual (experimental subjects) value at time t

The purpose of the proposed measure is to examine the computed value at each time t and to analyze the nature of the changes in the computed value over the entire lifecycle of the project.

B. RONAN'S EXPERIMENT

1. Experiment Description

The basic design of Ronan's experiment ("Experiment two") was to use graduate students as surrogate software project managers for decision making purposes (Ronan 1990). The subjects utilized Abdel-Hamid's and Madnick's System Dynamics Model of Software Project Management (SDM) gaming interface, to input decisions and to provide feedback at each of the decision making intervals (once every 20 days). The experiment was designed to create identical SDM projects which differed only by the initial man-day cost estimate. The initial constraints of the software project the subjects worked with were based on the DE-A project. The project variables within the SDM were identical, with the exception of the initial man-day cost. The subjects' decided upon the desired staffing level for the remainder of the project at each interval (based on information generated by the SDM gaming interface). Their goal was to decide on the staffing level which they felt would allow the project to finish on an acceptable schedule and while avoiding excessive cost overrun.

Ronan's objective was to compare the desired staffing level decisions of software project managers managing identical projects throughout the development phase. The only difference was that their man-day cost was initially under-estimated, over-estimated, or perfectly estimated. The subjects were divided into four groups, with the 8 or 9 students in each group designated by a "G-number". The group with the perfectly estimated initial cost was designated "G-1900" for an initial estimate of 1900 man-days. Two groups received over-estimated initial costs, "G-2185" and "G-2470". The under-estimated group was "G-1460".

Each subject, within each group made his own staffing decision based on the initial conditions and the subsequent information provided by the SDM gaming interface. A group desired staffing level, for each of the four groups at a given interval, was computed based on the combined results of each subject within a group.

The same initial conditions which were provided to the subjects for decision making, were than input into the SDM to compare the model's work force level decisions, with each of the four student groups.

2. Ronan Experiment Results and Analysis

The input data tables for the actual results and the model results for each of the variables is given in Appendix D. Table 4-1 provides the calculated summary statistics (RMSPE, MSE, Theil's Inequality Statistics, and TIC) for each data set of Ronan's Experiment.

TABLE 4-1. ERROR ANALYSIS OF RONAN'S EXPERIMENT (WORKFORCE LEVEL)

			INEQUALITY STATISTICS			
Variable	RMSPE(%)	MSE	U^M	U^S	U^C	TIC
G-1460	21.5	1.8	.16	.12	.72	.11
G-1900	18.1	1.6	.50	.28	.22	.11
G-2185	11.0	.6	.42	.15	.43	.06
G-2470	17.0	1.5	.56	.00	.44	.09

As can be seen from Table 4-1, the TIC values for each of the groups are well below the .40 level, suggesting that the model does an excellent job of matching the subjects decisions. The RMSPE ranges from a low of 11% to a high of 21.5%. This error range of the WORKFORCE variable in Ronan's experiment, is not unlike the error range exhibited in the DE-A and DE-B projects for this same variable (17.6% and 11% respectively). The subjects' actual values versus the model values are plotted in Figures 4-1, 4-2, 4-3, and 4-4 for G-1460, G-1900, G-2185, and G-2470 respectively. Obviously, the RMSPE values all exceed the 10% level and merit further analysis.

In general, the inequality statistics do not demonstrate the presence of clearly unsystematic error. Although, for the group with the initially underestimated cost (G-1460), the majority of the error is concentrated in covariance, which does indicate unsystematic error. For the remaining three groups however, much of the error is concentrated in the bias proportion. This could be an indicator of systematic error between the model and the experimental groups. A large, systematic error could be potentially troublesome, as it would limit the model's usefulness as a research and education tool. Or in the least, lead to questions of its usefulness.

One possible explanation for the existence of bias between the model and the student subjects, lies in the difference between the subjects' environment and the model environment. The experiment strived to place the subjects in the same environmental context which the model is based on. In contrast, the model is not designed to mimic the environment which the students are in. Therefore, it is possible for there to be various environmental factors which affect students' decisions, but are not reflected in the model's parameters. Whether or not it would be important to adjust

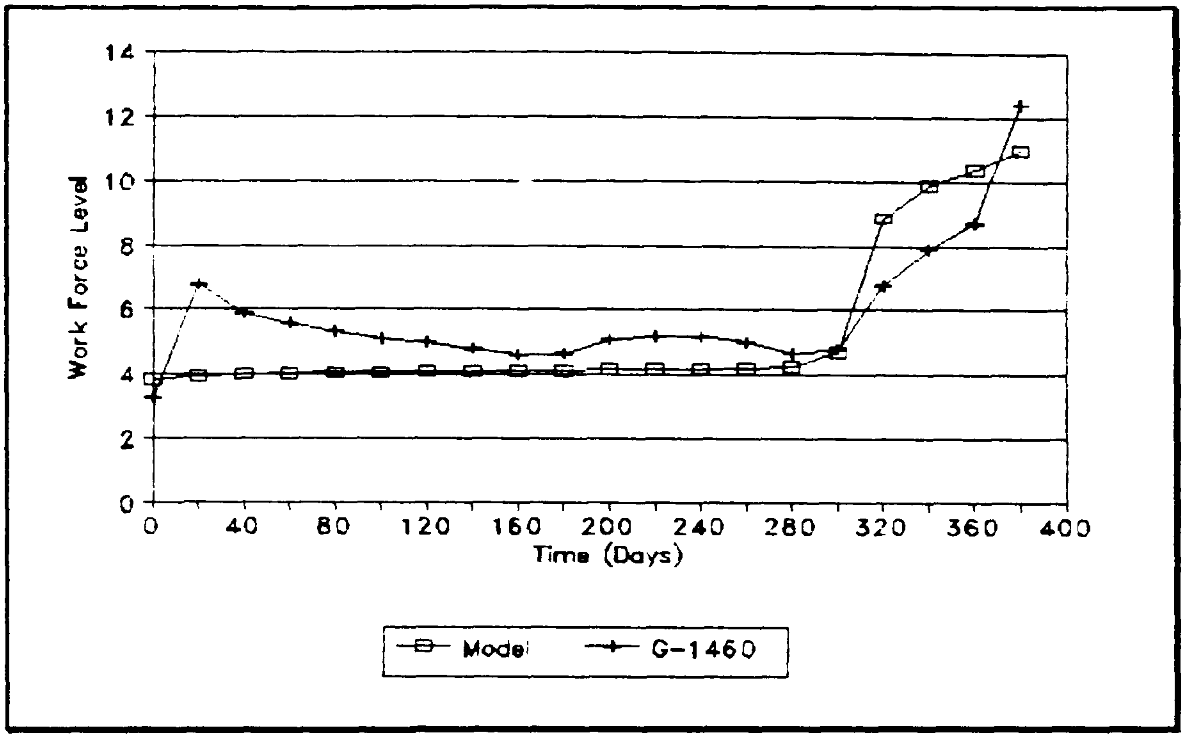

Figure 4-1. Ronan Experiment G-1460 Actual vs. Model

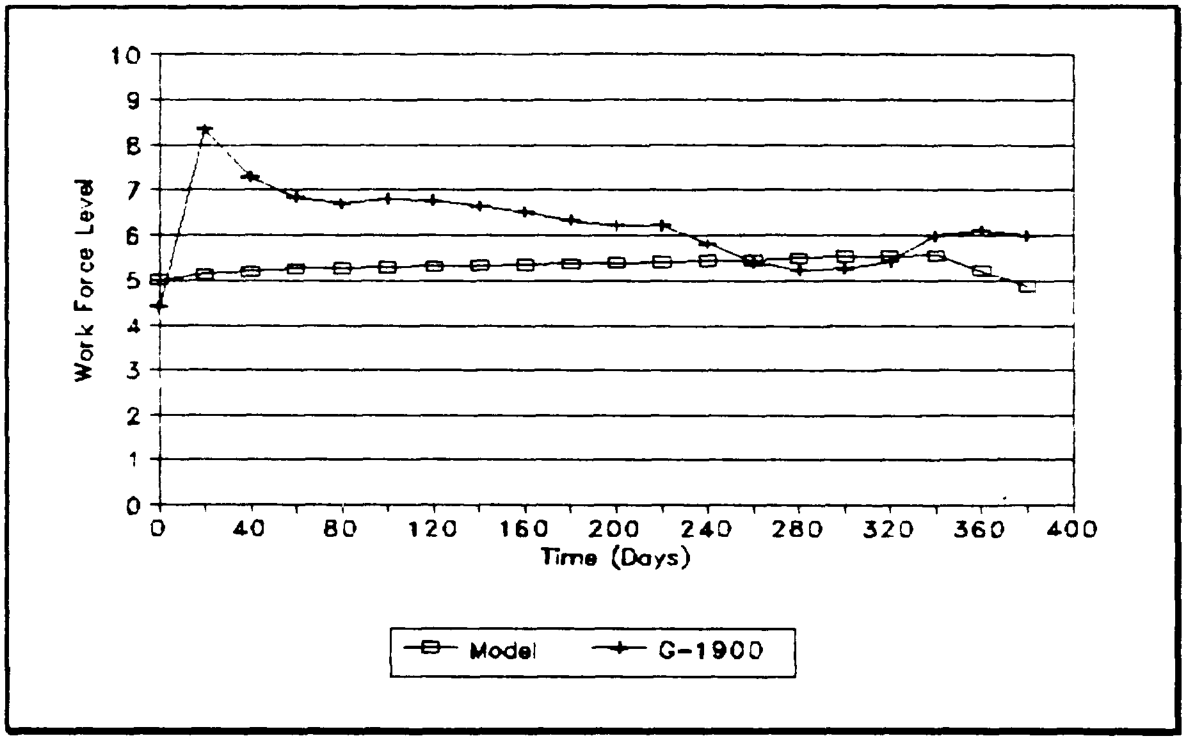

Figure 4-2. Ronan Experiment G-1900 Actual vs. Model

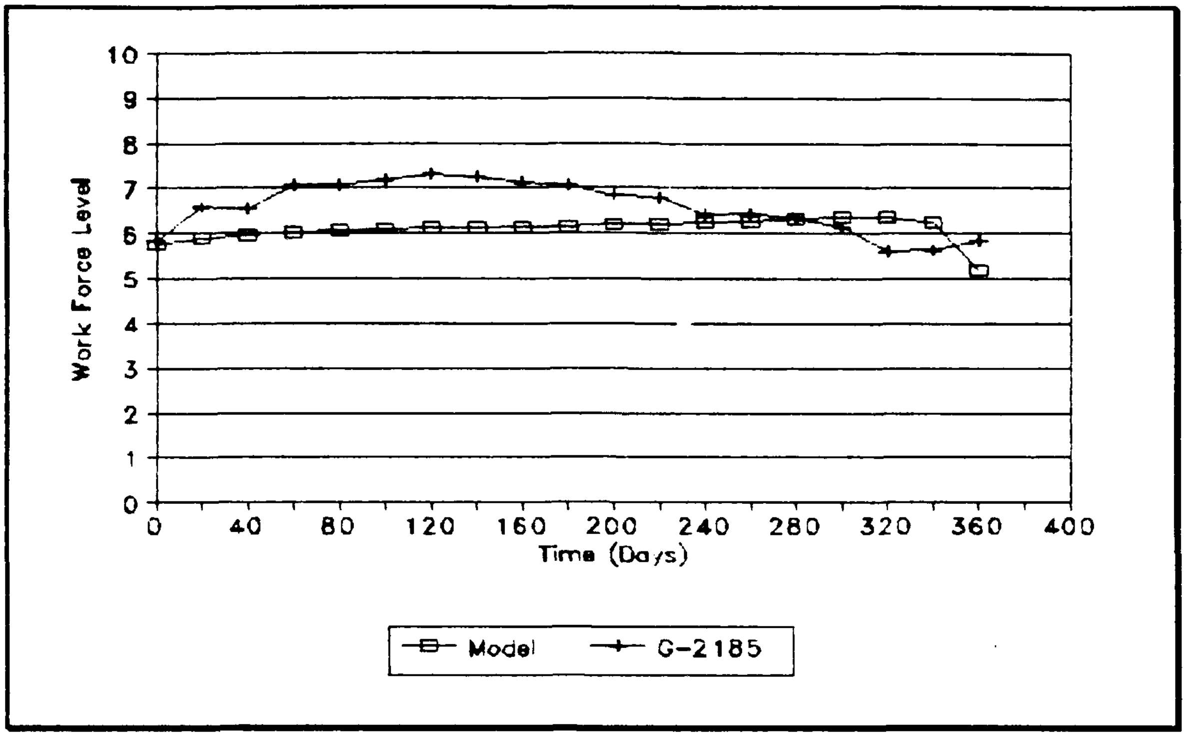

Figure 4-3. Ronan Experiment G-2185 Actual vs. Model

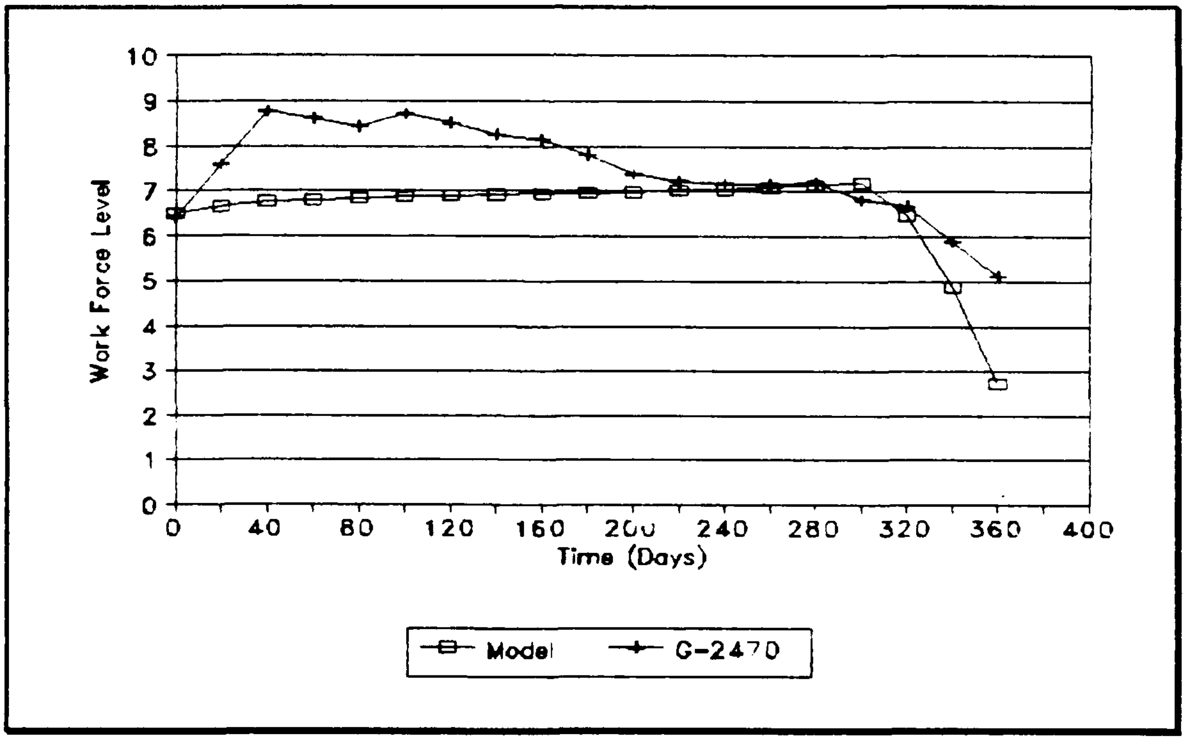

Figure 4-4. Ronan Experiment G-2470 Actual vs. Model

the model to account for these factors, if indeed they could be identified, would be up to the discretion of the model user. One would have to consider the impact of making any calibrations, as it could corrupt the structure of the software project environment being modelled.

As an example, within the software engineering curriculum, at the institution from which the subject students were used, the lesson of Brooks' Law was introduced, with some emphasis. Brooks' Law proclaims that adding more people to a late project only makes it later (Brooks 1975). The subjects knowledge of Brooks' Law could provide an explanation for the WORKFORCE level decisions made in the early stages of each group's projects. As seen in Figures 4-1, 4-2, 4-3, and 4-4, each of the groups tended to add more people early in each of their projects. This could be deduced as a means of avoiding the crux of Brooks' Law. That is, by adding people early in the project, I won't have to worry about the dilemma of Brooks' Law later on, because I can avoid the problem of a late project altogether. The figures also show that the model WORKFORCE decisions were well below those of the subjects during those initial stages. The impact of the students' knowledge of Brooks' Law alone, could explain a large portion of the error between the model and the subjects. Of course there is no evidence to support this and it is intended only as an example.

The existence of this error should not, however, degrade confidence in the model, provided that the modeler, or model user, recognizes the importance of the simulation environment. To a certain extent, it may be possible to calibrate the SDM gaming interface to more closely reflect the subjects environment, without disturbing the integrity of the software development environment, which the model is designed to emulate. Of course, the extent of any calibration would depend greatly on the intended purpose of the experiment.

C. BAKER'S EXPERIMENT

1. Experiment Description

The design of Baker's experiment is essentially the same as Ronan's (Baker 1992). Graduate students were used as surrogate software project managers, they used the SDM gaming interface to input their decisions and to provide updated status reports on the software project, and the initial constraints of their software project were based on the DE-A project.

In Baker's experiment there were two groups of subjects (Group A and Group B). Each group started off with the same initial conditions and the same objective, to complete the project as close as possible to the original estimates of schedule and cost. The difference between the two groups, was that Group A's project grew gradually in size from 320 tasks (one task equals approximately 50 lines of code) to 610 tasks, by day 100 of the project simulation. Group B's project size remained at 320 tasks through the 100th day of the project simulation, after the Day 80 status report (40 day decision making intervals), the subjects received a message on their screen that the project size had just been increased to 610 tasks, due to increased requirements. The project size then remained constant for the remainder of the project simulation for both groups. The subjects were required to input two decisions (staffing level and project cost estimate) at each simulated 40 day interval.

2. Baker Experiment Results and Analysis

The input data tables for the actual results and the model results for each of the variables is given in Appendix E. Table 4-2 and 4-3 provide the RMSPE, MSE, Theil's Inequality Statistics, and TIC for each data set of Baker's Experiment.

TABLE 4-2. ERROR ANALYSIS OF BAKER'S EXPERIMENT (GROUP A)

			INEQUALITY STATISTICS			
Variable	RMSPE(%)	MSE	U^M	U^S	U^C	TIC
WORKFORCE	23.9	1.7	.26	.32	.41	.12
MAN-DAYS	8.6	11936	.41	.01	.58	.04

TABLE 4-3. ERROR ANALYSIS OF BAKER'S EXPERIMENT (GROUP B)

			INEQUALITY STATISTICS			
Variable	RMSPE(%)	MSE	U^M	U^S	U^C	TIC
WORKFORCE	25.6	2.3	.53	.11	.36	.13
MAN-DAYS	11.0	23110	.45	.17	.39	.05

The statistics presented in Tables 4-2 and 4-3 do not differ dramatically from those presented in Ronan's experiment. The TIC values for all of the variables, are all well below the .40 level, indicating a very good or excellent model. The RMSPE values for WORKFORCE are somewhat high, although not significantly higher than in Ronan's experiment. The RMSPE values for MAN-DAYS straddle the .10 level, indicating that the model structure as it relates to MAN-DAYS is probably sound. Additionally, the breakdown of the inequality statistics does not clearly reveal errors which are unsystematic, as was the case in the Ronan experiment. A plot of the model generated decisions versus the student subject generated decisions for Group A WORKFORCE, Group A MAN-DAYS, Group B WORKFORCE, and Group B MAN-DAYS is displayed in Figures 4-5, 4-6, 4-7, and 4-8 respectively. Essentially the same discussion which was presented in the analysis of Ronan's experiment, in regards to the simulation environment and Brooks' Law, is also applicable to Baker's experiment and will not be reiterated. Therefore, even though the size of the errors could be construed as being large (RMSPE's above 10%) and possibly systematic, the nature of these errors can be acceptable to a user of the model. As such, the scope of these errors does not necessarily degrade ones confidence in the model.

The computations for the proposed alternative analysis measure, for Group A, WORKFORCE variable are presented in Table 4-4. The intent of this measure is to analyze the nature of the changes in the difference between the model and the experiment subjects, over a project's lifecycle for a given variable. This is only a proposed measure, however, and as such requires further analysis as to its suitability. It is presented here to serve as a basis for further research. The computation for each value at time t is as follows:

$$\frac{\sqrt{\sum (A_t - S_t)^2}}{S_t}.$$

As with the previous statistics, this measure was computed using the Quattro Pro 3 spreadsheet application. The actual input values and the spreadsheet documentation used to derive the values for the WORKFORCE variable of Group A, are given in Appendix F.

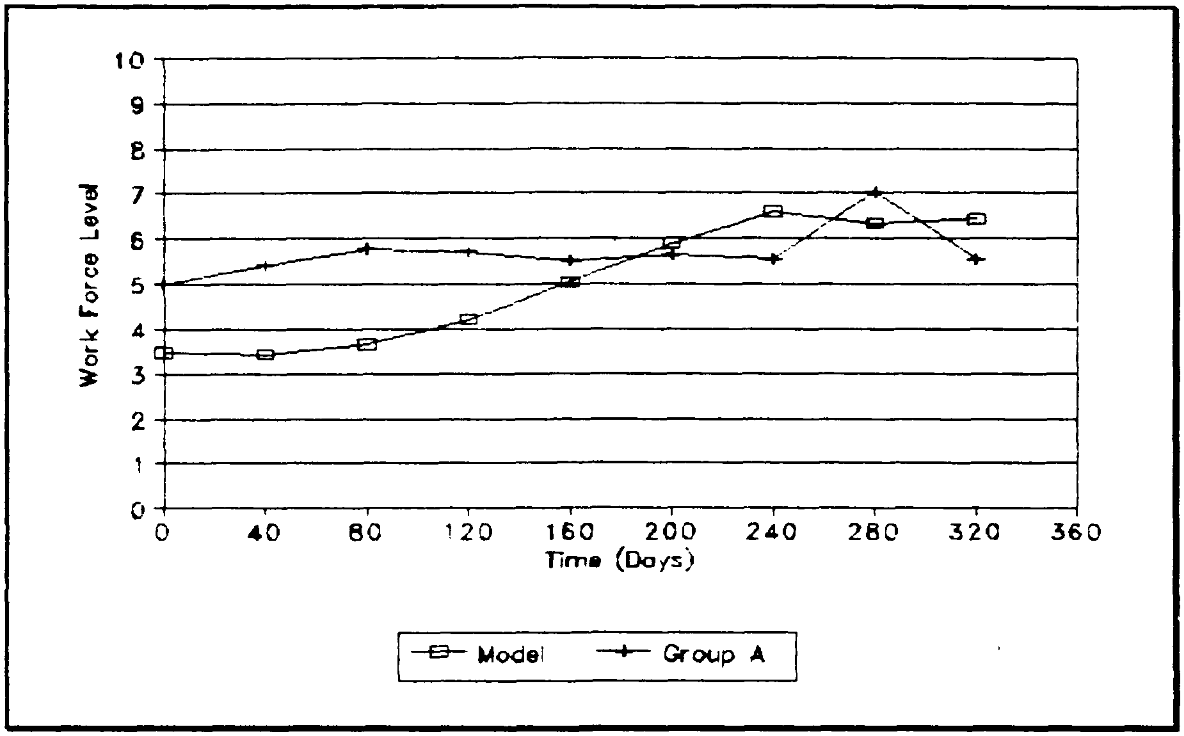

Figure 4-5. Baker Experiment Group A WORKFORCE Actual vs. Model

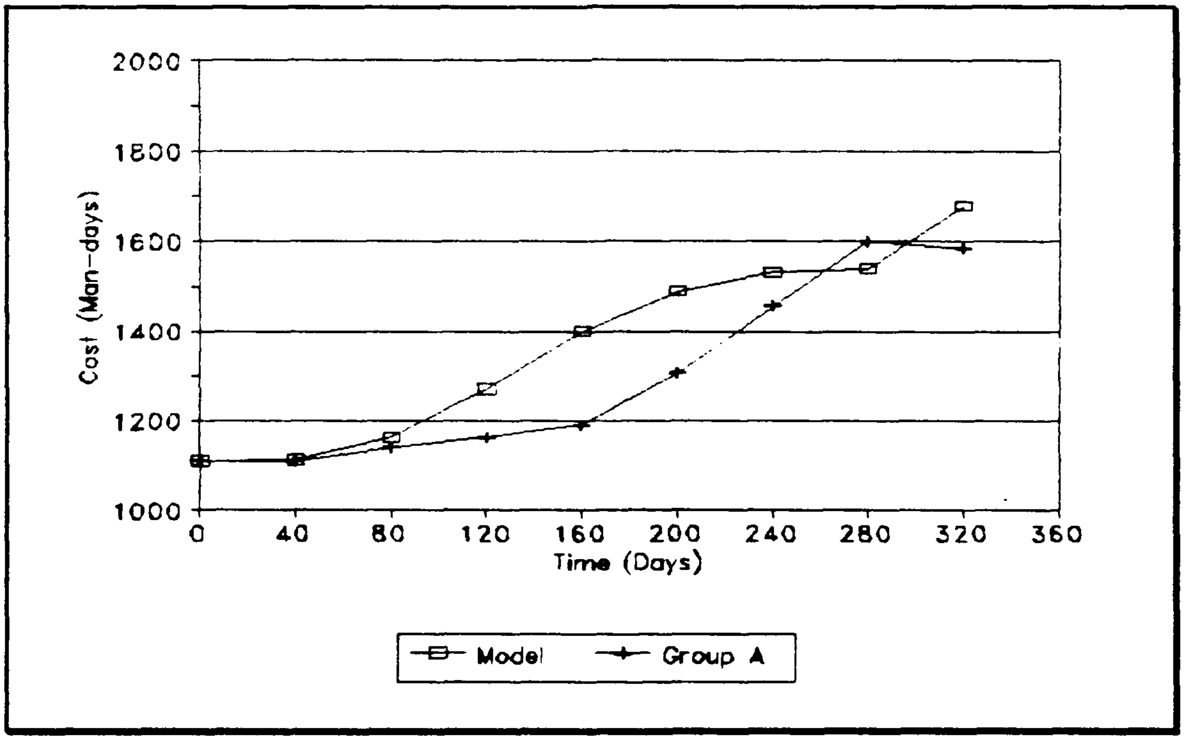

Figure 4-6. Baker Experiment Group A MAN-DAYS

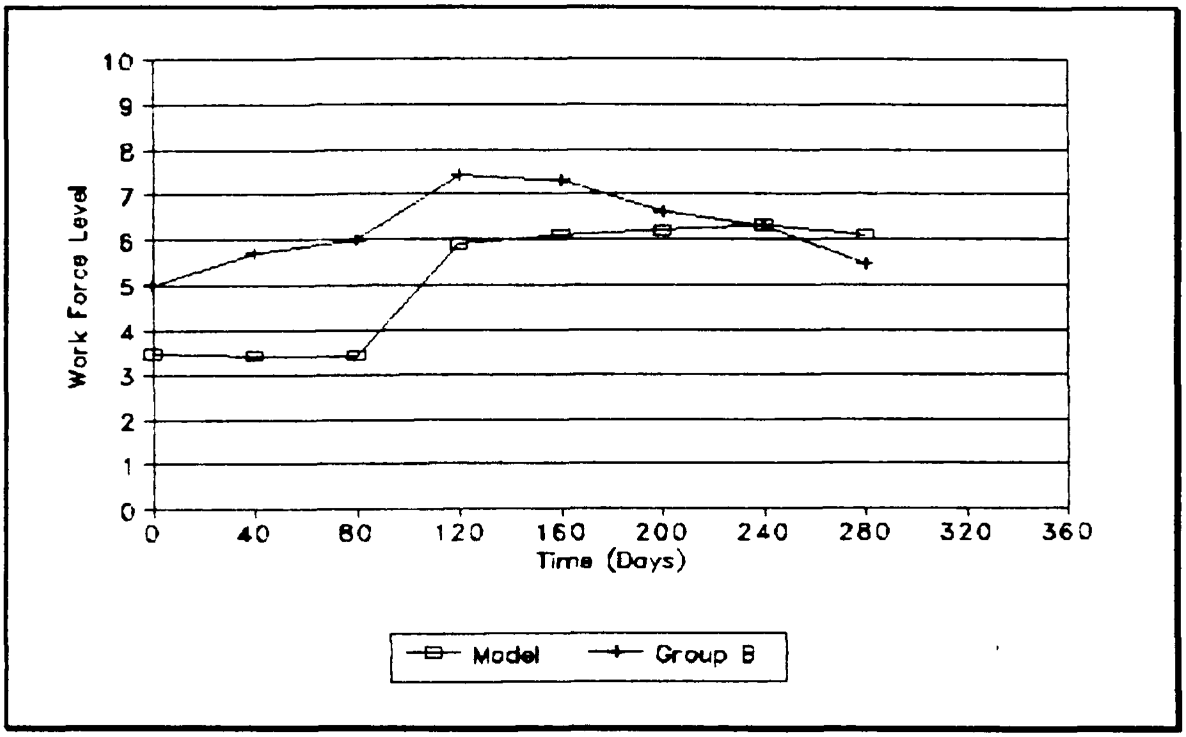

Figure 4-7. Baker Experiment Group B WORKFORCE Actual vs. Model

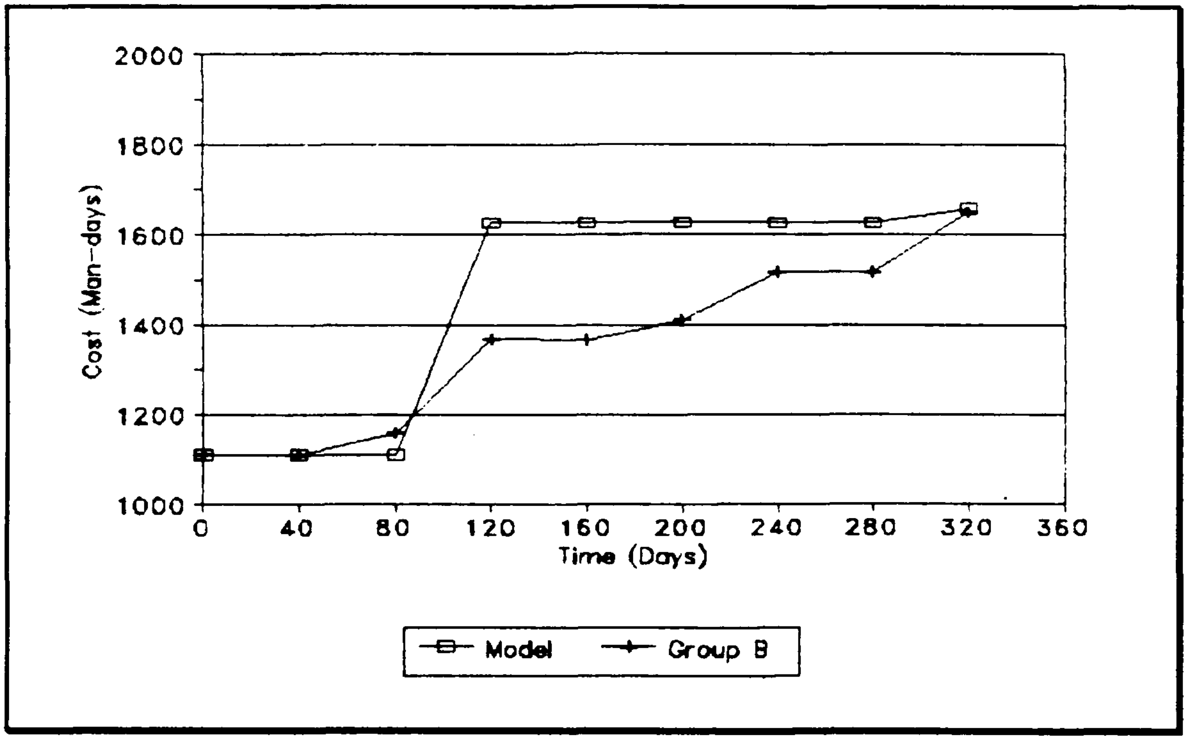

Figure 4-8. Baker Experiment Group B MAN-DAYS Actual vs. Model

TABLE 4-4. COMPUTED VALUE AT TIME INTERVAL *t*

0	**40**	**80**	**120**	**160**	**200**	**240**	**280**	**320**
2.07	2.85	3.18	2.62	1.90	1.67	1.51	1.32	.97

D. CONCLUSIONS

The analysis of the test statistics for the direct experimentation method, is not as clear as it was in the case of comparison with actual project results. The results of the analysis reveal that further research would be prudent, as the results were somewhat mixed. The TIC values, on all counts, suggest that the model is an excellent one in terms of forecasting ability. The RMSPE and the breakdown of the inequality statistics, however, indicate that caution and a thorough understanding of the purpose and use of the model is essential. While this may seem intuitively obvious to some, irregardless of the validation results, it is still worth noting. Additionally, an alternative measure for analyzing the nature of the error differences, between the model and the experimental subjects, over a project's lifecycle, was presented. Further research is required to further explore the potential of this proposed measure.

V. CONCLUSIONS

A. ACCOMPLISHMENTS

Using several statistical measures to evaluate the goodness-of-fit of the Abdel-Hamid and Madnick model (Abdel-Hamid and Madnick 1991), this increases confidence in the validity and utility of the system dynamics model of software development by Abdel-Hamid and Madnick. Of course, how much confidence one has in the model depends greatly upon one's view of the model and its relationship reality. There are basically two potential users of the model. The first are software project managers. They would use the model to study the effects of varying management policies/decisions relevant to a given project's lifecycle. The second is academics. Their primary purposes for using the model would be to gain an understanding of the complexities of the software development process and as a teaching tool.

Those users then, must have confidence in the model's ability to be a reasonably true reflection of reality if they are to make use of the model. If the model's goodness-of-fit is not adequate to suit their needs or expectations, then that alone may be reason enough to discard the model. The goal of this research effort was to complement the earlier validation steps with a battery of statistical measures. The approach taken to achieve this goal, was to incorporate general statistical measures given by others within the system dynamics field of study (TIC and RMSPE). Where a TIC below .40 equated to a very good or excellent model (Rowland and Holmes 1978) and an RMSPE less than .10 indicated that the model's structure was probably sound (Veit 1978).

The error between the real system and the model was then broken down into its sources in terms of bias, variance, and covariance using the Theil inequality statistics. From there, the type of error was determined. Was it an error which the model should reasonably be able to capture (systematic error)? Or, was it from some influence outside of the system being modeled (unsystematic error) and therefore, reasonable for the model to not capture the error? If determined that the error was systematic in nature, then the modeler may need to reexamine the estimation parameters. If the error was unsystematic, than the user must accept the assumption that it is not reasonable to expect the model to capture the exogenous influences. In the case of unsystematic errors, it is possible for the modeler to insert dummy variables into the model to create a closer goodness-of fit. However, doing so may upset the feedback structure inherent in a system dynamics model and the model would no longer be a reflection of the system it is attempting to emulate. This would only defeat the purpose of creating the model.

This research also conducted a comparison of the model using direct experimentation. While this type of comparison is somewhat different than that of a comparison with actual project results, it provides a useful analysis tool. The experimentation analysis, while not as clear cut as the comparison with actual results, did not serve to undermine any confidence that one would vest in the model. The experimentation analysis also introduced another measure of analysis based on work done by Sterman (1987 and 1989). This measure has potential for use in future analysis of the model.

The results of this research have demonstrated that the system dynamics model of software development by Abdel-Hamid and Madnick, displays a reasonable and acceptable degree of goodness-of-fit. This study therefore, should serve to build confidence in the model's usefulness as a tool for providing a better understanding of the managerial aspects of the software development process.

B. FUTURE DIRECTION

The process of model validation should be evolutionary. Since the nature of a system dynamics model is dynamic, the confidence building process, or validation process, should also be dynamic in order to keep pace with the model. There is no single test which can and should be used for validation efforts. Rather, a multitude of varying tests should be incorporated to test the suitability and utility of a model for its given environment. As the model itself grows and adapts to its environment, so to should the testing process.

There exists a multitude of directions that one could take for future testing of this model. Several are given here:

(1) Collect results from other software projects that are suited to this particular model and conduct the same type of analysis as presented in this thesis.

(2) Conduct further analysis of the direct experimentation method using the tests of significance presented by Sterman (1989).

(3) Utilize the six-step behavior validation procedure presented by Barlas (1989) on the DE-A and DE-B projects.

(4) Utilize spectral analysis techniques to compare the DE-A and DE-B project results to the model results.

(5) Conduct further research into the proposed analysis measure presented in Chapter IV to determine its suitability and applicability to system dynamics models.

APPENDIX A

The following are the cell formulas, as computed in the Quattro Pro V. 3.0 spreadsheet application. The formulas for the DE-A project only, are presented. The only difference between the DE-A project and the other computations is the data set. Otherwise, all other values and formulas remain the same for the other projects and experiments.

Three spreadsheets were built (MSE and RMSPE computation spreadsheet, Theil's Inequality Statistics spreadsheet, and TIC spreadsheet). The MSE and RMSPE spreadsheet was the base spreadsheet (FORMMSE), where all of the initial data entry was made. The other two spreadsheets linked to the first spreadsheet with the link [FORMMSE] and the specific cell address.

1. Computations for Mean Square Error (MSE) and Root Mean Square Percentage Error (RMSPE)

```
A1: (T) [W17] 'PROJECT:
B1: (T) 'DE-A
A2: (T) [W17] 'Mean Square Error (MSE) of
D2: (T) [W9] 'MAN-DAYS
E2: (T) [W15] '  Data
A5: (T) [W17] ^t
B5: (T) ^St
C5: (T) [W7] ^At
D5: (T) [W9] ^St-At
E5: (T) [W15] ^(St-At)**2
F5: (T) [W9] ^St-At/At
G5: (T) [W15] ^(St-At/At)**2
A6: (T) [W17] 0
B6: (T) 1111
C6: (T) [W7] 1111
D6: (T) [W9] +B6-C6
E6: (T) [W15] +D6^2
F6: (T) [W9] +D6/C6
G6: (T) [W15] +F6^2
```

```
A7: (T) [W17] 60
B7: (T) 1111
C7: (T) [W7] 1111
D7: (T) [W9] +B7-C7
E7: (T) [W15] +D7^2
F7: (T) [W9] +D7/C7
G7: (T) [W15] +F7^2
A8: (T) [W17] 80
B8: (T) 1115.5
C8: (T) [W7] 1336
D8: (T) [W9] +B8-C8
E8: (T) [W15] +D8^2
F8: (T) [W9] +D8/C8
G8: (T) [W15] +F8^2
A9: (T) [W17] 150
B9: (T) 1225.7
C9: (T) [W7] 1336
D9: (T) [W9] +B9-C9
E9: (T) [W15] +D9^2
F9: (T) [W9] +D9/C9
G9: (T) [W15] +F9^2
A10: (T) [W17] 240
B10: (T) 1428.3
C10: (T) [W7] 1582
D10: (T) [W9] +B10-C10
E10: (T) [W15] +D10^2
F10: (T) [W9] +D10/C10
G10: (T) [W15] +F10^2
A11: (T) [W17] 280
B11: (T) 1461.5
C11: (T) [W7] 1582
D11: (T) [W9] +B11-C11
E11: (T) [W15] +D11^2
F11: (T) [W9] +D11/C11
G11: (T) [W15] +F11^2
A12: (T) [W17] 300
B12: (T) 1637.5
C12: (T) [W7] 1657
D12: (T) [W9] +B12-C12
E12: (T) [W15] +D12^2
F12: (T) [W9] +D12/C12
G12: (T) [W15] +F12^2
A13: (T) [W17] 340
B13: (T) 1951.5
C13: (T) [W7] 1750
D13: (T) [W9] +B13-C13
E13: (T) [W15] +D13^2
F13: (T) [W9] +D13/C13
G13: (T) [W15] +F13^2
A14: (T) [W17] 360
B14: (T) 2029.9
```

```
C14: (T) [W7] 1769
D14: (T) [W9] +B14-C14
E14: (T) [W15] +D14^2
F14: (T) [W9] +D14/C14
G14: (T) [W15] +F14^2
A15: (T) [W17] 380
B15: (T) 2121.5
C15: (T) [W7] 2239
D15: (T) [W9] +B15-C15
E15: (T) [W15] +D15^2
F15: (T) [W9] +D15/C15
G15: (T) [W15] +F15^2
A16: (T) [W17] @COUNT(A6..A15)
E16: (T) [W15] @SUM(E6..E15)
G16: (T) [W15] @SUM(G6..G15)
A18: (T) [W17] 'MSE=
B18: (T) +E16/A16
F18: (T) [W9] 'RMSPE=
G18: (T) [W15] (G16/A16)^0.5
```

2. Computations for Theil's Inequality Statistics (U^M, U^S, and U^C)

```
A1: (T) [W15] 'PROJECT:
B1: (T) [W33] 'DE-A
A2: (T) [W15] 'THEIL Inequality Statistics for
D2: (T) [W12] 'MAN-DAYS
E2: (T) '  Data
A5: (T) [W15] ^(1)
B5: (T) [W33] ^(2)
C5: (T) [W15] ^(3)
D5: (T) [W12] ^(4)
E5: (T) ^(5)
F5: (T) [W30] ^(6)
A6: (T) [W15] +[FORMMSE]A5
B6: (T) [W33] +[FORMMSE]B5
C6: (T) [W15] +[FORMMSE]C5
D6: (T) [W12] ^St-S(mean)
E6: (T) ^At-A(mean)
F6: (T) [W30] ^(4)*(5)
A7: (T) [W15] +[FORMMSE]A6
B7: (T) [W33] +[FORMMSE]B6
C7: (T) [W15] +[FORMMSE]C6
D7: (T) [W12] +B7-$B$19
E7: (T) +C7-$C$19
F7: (T) [W30] +D7*E7
A8: (T) [W15] +[FORMMSE]A7
B8: (T) [W33] +[FORMMSE]B7
```

```
C8: (T) [W15] +[FORMMSE]C7
D8: (T) [W12] +B8-$B$19
E8: (T) +C8-$C$19
F8: (T) [W30] +D8*E8
A9: (T) [W15] +[FORMMSE]A8
B9: (T) [W33] +[FORMMSE]B8
C9: (T) [W15] +[FORMMSE]C8
D9: (T) [W12] +B9-$B$19
E9: (T) +C9-$C$19
F9: (T) [W30] +D9*E9
A10: (T) [W15] +[FORMMSE]A9
B10: (T) [W33] +[FORMMSE]B9
C10: (T) [W15] +[FORMMSE]C9
D10: (T) [W12] +B10-$B$19
E10: (T) +C10-$C$19
F10: (T) [W30] +D10*E10
A11: (T) [W15] +[FORMMSE]A10
B11: (T) [W33] +[FORMMSE]B10
C11: (T) [W15] +[FORMMSE]C10
D11: (T) [W12] +B11-$B$19
E11: (T) +C11-$C$19
F11: (T) [W30] +D11*E11
A12: (T) [W15] +[FORMMSE]A11
B12: (T) [W33] +[FORMMSE]B11
C12: (T) [W15] +[FORMMSE]C11
D12: (T) [W12] +B12-$B$19
E12: (T) +C12-$C$19
F12: (T) [W30] +D12*E12
A13: (T) [W15] +[FORMMSE]A12
B13: (T) [W33] +[FORMMSE]B12
C13: (T) [W15] +[FORMMSE]C12
D13: (T) [W12] +B13-$B$19
E13: (T) +C13-$C$19
F13: (T) [W30] +D13*E13
A14: (T) [W15] +[FORMMSE]A13
B14: (T) [W33] +[FORMMSE]B13
C14: (T) [W15] +[FORMMSE]C13
D14: (T) [W12] +B14-$B$19
E14: (T) +C14-$C$19
F14: (T) [W30] +D14*E14
A15: (T) [W15] +[FORMMSE]A14
B15: (T) [W33] +[FORMMSE]B14
C15: (T) [W15] +[FORMMSE]C14
D15: (T) [W12] +B15-$B$19
E15: (T) +C15-$C$19
F15: (T) [W30] +D15*E15
A16: (T) [W15] +[FORMMSE]A15
B16: (T) [W33] +[FORMMSE]B15
C16: (T) [W15] +[FORMMSE]C15
D16: (T) [W12] +B16-$B$19
E16: (T) +C16-$C$19
```

```
F16: (T) [W30] +D16*E16
A17: (T) [W15] +[FORMMSE]A16
F17: (T) [W30] @SUM(F7..F16)
A19: (T) [W15] "Mean=
B19: (T) [W33] @AVG(B7..B16)
C19: (T) [W15] @AVG(C7..C16)
E19: (T) "r=
F19: (T) [W30] ((1/A17)*(F17))/((B20)*(C20))
A20: (T) [W15] "Std Dev=
B20: (T) [W33] @STD(B7..B16)
C20: (T) [W15] @STD(C7..C16)
A23: (T) [W15] "UM=
B23: (T) [W33] ((B19-C19)^2)/([FORMMSE]B18)
A24: (T) [W15] "US=
B24: (T) [W33] ((B20-C20)^2)/[FORMMSE]B18
A25: (T) [W15] "UC=
B25: (T) [W33] (2*(1-F19)*B20*C20)/[FORMMSE]B18
A26: (T) [W15] "Total=
B26: (T) [W33] @SUM(B23..B25)
```

3. Computations for Theil's Inequality Coefficient (TIC)

```
A1: (T) [W41] 'PROJECT:
B1: (T) [W53] 'DE-A
A2: (T) [W41] 'THEIL Inequality Coefficient for
D2: (T) [W41] 'MAN-DAYS
E2: (T) [W15] '  Data
A5: (T) [W41] ^(1)
B5: (T) [W53] ^(2)
C5: (T) [W41] ^(3)
D5: (T) [W41] ^(4)
E5: (T) [W15] ^(5)
F5: (T) [W15] ^(6)
A6: (T) [W41] +[FORMMSE]A5
B6: (T) [W53] +[FORMMSE]B5
C6: (T) [W41] +[FORMMSE]C5
D6: (T) [W41] ^(St-At)**2
E6: (T) [W15] ^St**2
F6: (T) [W15] ^At**2
A7: (T) [W41] +[FORMMSE]A6
B7: (T) [W53] +[FORMMSE]B6
C7: (T) [W41] +[FORMMSE]C6
D7: (T) [W41] +[FORMMSE]E6
E7: (T) [W15] +B7^2
F7: (T) [W15] +C7^2
A8: (T) [W41] +[FORMMSE]A7
B8: (T) [W53] +[FORMMSE]B7
C8: (T) [W41] +[FORMMSE]C7
```

```
D8: (T) [W41] +[FORMMSE]E7
E8: (T) [W15] +B8^2
F8: (T) [W15] +C8^2
A9: (T) [W41] +[FORMMSE]A8
B9: (T) [W53] +[FORMMSE]B8
C9: (T) [W41] +[FORMMSE]C8
D9: (T) [W41] +[FORMMSE]E8
E9: (T) [W15] +B9^2
F9: (T) [W15] +C9^2
A10: (T) [W41] +[FORMMSE]A9
B10: (T) [W53] +[FORMMSE]B9
C10: (T) [W41] +[FORMMSE]C9
D10: (T) [W41] +[FORMMSE]E9
E10: (T) [W15] +B10^2
F10: (T) [W15] +C10^2
A11: (T) [W41] +[FORMMSE]A10
B11: (T) [W53] +[FORMMSE]B10
C11: (T) [W41] +[FORMMSE]C10
D11: (T) [W41] +[FORMMSE]E10
E11: (T) [W15] +B11^2
F11: (T) [W15] +C11^2
A12: (T) [W41] +[FORMMSE]A11
B12: (T) [W53] +[FORMMSE]B11
C12: (T) [W41] +[FORMMSE]C11
D12: (T) [W41] +[FORMMSE]E11
E12: (T) [W15] +B12^2
F12: (T) [W15] +C12^2
A13: (T) [W41] +[FORMMSE]A12
B13: (T) [W53] +[FORMMSE]B12
C13: (T) [W41] +[FORMMSE]C12
D13: (T) [W41] +[FORMMSE]E12
E13: (T) [W15] +B13^2
F13: (T) [W15] +C13^2
A14: (T) [W41] +[FORMMSE]A13
B14: (T) [W53] +[FORMMSE]B13
C14: (T) [W41] +[FORMMSE]C13
D14: (T) [W41] +[FORMMSE]E13
E14: (T) [W15] +B14^2
F14: (T) [W15] +C14^2
A15: (T) [W41] +[FORMMSE]A14
B15: (T) [W53] +[FORMMSE]B14
C15: (T) [W41] +[FORMMSE]C14
D15: (T) [W41] +[FORMMSE]E14
E15: (T) [W15] +B15^2
F15: (T) [W15] +C15^2
A16: (T) [W41] +[FORMMSE]A15
B16: (T) [W53] +[FORMMSE]B15
C16: (T) [W41] +[FORMMSE]C15
D16: (T) [W41] +[FORMMSE]E15
E16: (T) [W15] +B16^2
F16: (T) [W15] +C16^2
```

```
A17: (T) [W41] +[FORMMSE]A16
D17: (T) [W41] +[FORMMSE]E16
E17: (T) [W15] @SUM(E7..E16)
F17: (T) [W15] @SUM(F7..F16)
A19: (T) [W41] "U=
B19: (T) [W53]
((+D17/A17)^0.5)/(((+E17/A17)^0.5)+((+F17/A17)^0.5))
```

APPENDIX B

Input data files for DE-A project, SCHEDULE variable, MAN-DAYS variable, and WORKFORCE variable at time t. Where:

S_t = Simulated (model) value at time t

A_t = Actual value at time t

SCHEDULE DATA TABLE

t	S_t	A_t
40	320.01	320
80	320.01	320
120	322.55	320
160	322.55	320
200	322.56	320
240	322.57	320
260	323.04	330
280	327.48	330
320	356.2	360

MAN-DAYS DATA TABLE

t	S_t	A_t
0	1111	1111
60	1111	1111
80	1115.5	1336
150	1225.7	1336
240	1428.3	1582
280	1461.5	1582
300	1637.5	1657
340	1951.5	1750
360	2029.5	1769
380	2121.5	2239

WORKFORCE DATA TABLE

t	S_t	A_t
40	2.96	2
60	3.27	3
80	3.43	3.6
100	3.47	4
120	3.50	4
160	3.80	4.2
240	5.95	6
280	6.55	6.8
300	6.80	7.5
320	7.13	9.5
360	12.92	14
380	14.67	16

APPENDIX C

Input data files for DE-B project, SCHEDULE variable, MAN-DAYS variable, and WORKFORCE variable at time *t*. Where:

S_t = Simulated (model) value at time *t*

A_t = Actual value at time *t*

SCHEDULE DATA TABLE

t	S_t	A_t
40	320.01	320
80	320.01	320
120	330.86	320
160	330.87	320
180	330.87	320
200	330.87	320
260	331.04	330
280	335.00	330
300	339.90	330
335	341.85	335

MAN-DAYS DATA TABLE

t	S_t	A_t
0	1111	1111
60	1111	1111
80	1115.5	1336
150	1225.7	1336
240	1428.3	1582
280	1461.5	1582
300	1637.5	1657
335	1951.5	1750

WORKFORCE DATA TABLE

t	S_t	A_t
40	2.73	2.5
80	3.42	3.4
120	4.71	5
160	6.15	6.5
200	7.09	9.5
240	8.14	8
280	9.20	8.5
320	8.98	9.2
335	8.96	10.5

APPENDIX D

Input data files for Ronan's Experiment, WORKFORCE variable at time *t* for G-1460, G-1900, G-2185, and G-2470. Where:

S_t = Simulated (model) value at time *t*

A_t = Actual value at time *t*

G-1460 DATA TABLE

t	S_t	A_t
0	3.8	3.3
20	3.9	6.8
40	4.0	5.9
60	4.0	5.6
80	4.0	5.3
100	4.1	5.1
120	4.1	5.0
140	4.1	4.8
160	4.1	4.6
180	4.1	4.6
200	4.1	5.1
220	4.1	5.2
240	4.2	5.2
260	4.2	5.0
280	4.2	4.7
300	4.7	4.8
320	8.9	6.8
340	9.9	7.9
360	10.4	8.7
380	11.0	12.4

G-1900 DATA TABLE

t	S_t	A_t
0	5.0	4.4
20	5.1	8.4
40	5.2	7.3
60	5.2	6.8
80	5.3	6.7
100	5.3	6.8
120	5.3	6.8
140	5.3	6.6
160	5.3	6.5
180	5.4	6.3
200	5.4	6.2
220	5.4	6.2
240	5.4	5.8
260	5.5	5.4
280	5.5	5.2
300	5.5	5.3
320	5.5	5.4
340	5.5	6.0
360	5.2	6.1
380	4.9	6.0

G-2185 DATA TABLE

t	S_t	A_t
0	5.8	5.8
20	5.9	6.6
40	6.0	6.5
60	6.0	7.1
80	6.1	7.1
100	6.1	7.2
120	6.1	7.3
140	6.1	7.3
160	6.1	7.1
180	6.2	7.1
200	6.2	6.8
220	6.2	6.8
240	6.2	6.4
260	6.3	6.4
280	6.3	6.3
300	6.4	6.1
320	6.3	5.6
340	6.2	5.6
360	5.2	5.8

G-2470 DATA TABLE

t	S_t	A_t
0	6.5	6.4
20	6.7	7.6
40	6.7	8.8
60	6.8	8.6
80	6.8	8.4
100	6.9	8.7
120	6.9	8.5
140	6.9	8.2
160	6.9	8.2
180	7.0	7.8
200	7.0	7.4
220	7.0	7.2
240	7.0	7.2
260	7.1	7.2
280	7.1	7.2
300	7.2	6.8
320	6.5	6.7
340	4.9	5.9
360	2.7	5.1

APPENDIX E

Input data files for Baker's Experiment, WORKFORCE variable and MAN-DAYS variable at time t, for Group A and Group B. Where:

S_t = Simulated (model) value at time t

A_t = Actual value at time t

GROUP A WORKFORCE DATA TABLE

t	S_t	A_t
0	3.5	5
40	3.4	5.4
80	3.6	5.8
120	4.2	5.7
160	5.1	5.5
200	5.9	5.6
240	6.6	5.6
280	6.3	7.0
320	6.4	5.6

GROUP A MAN-DAYS DATA TABLE

t	S_t	A_t
0	1111.0	1111
40	1111.5	1111
80	1161.9	1140
120	1268.8	1162
160	1398.9	1188
200	1487.5	1306
240	1529.6	1456
280	1539.1	1600
320	1677.8	1583

GROUP B WORKFORCE DATA TABLE

t	S_t	A_t
0	3.5	5.0
40	3.4	5.7
80	3.4	6.0
120	5.9	7.4
160	6.1	7.3
200	6.2	6.6
240	6.3	6.3
280	6.1	5.5

GROUP B MAN-DAYS DATA TABLE

t	S_t	A_t
0	1111.0	1111
40	1111.0	1108
80	1111.0	1158
120	1624.4	1366
160	1624.4	1364
200	1624.4	1407
240	1624.4	1516
280	1626.2	1516
320	1654.6	1649

APPENDIX F

The following listings are the input data tables and the cell formula documentation, for the WORKFORCE variable of Baker's Experiment, for the proposed measure. The input data table used to compute the measure is presented first, followed by the cell formula computations.

1. Input data table for the student data (ACTUAL) and the Model data for time *t* of the project's lifecycle.

GROUP A WORKFORCE
ACTUAL AND MODEL VALUES FOR TIME T

NAME	0	40	80	120	160	200	240	280	320
BELL	5	5	5.3	5.5	5.5	5.6	6.2	7.5	
BITTNER	5	5	6	6	6	7	7	6	
BRANLEY	5	5.96	6.5	7.7	7.7	10.2	10.2	9	
CHELOUCHE	5	4	4	3	3	3	4	8	7
CULPEPPER	5	5	5	7	7	7	4	5	
FEY	5	5	5	4.8	5.6	8	5	10	
FOSTER, T	5	5	5	5	5	5	3	3	3
HODGKINS	5	5.5	5.5	6	6.5	6.5	6.8	6.8	
IVEY	5	6	6	6	6	5	4	4	
LACO	5	5.5	5.5	4.9	4.9	4.8	4.8	4.8	
LOCKHART	5	4.5	4	3.5	3.5	5.5	6	5.5	6.5
MAIN	5	7	7	7.5	8.5	8.5	8.5		
METCALF	5	6	6	5.5	2.6	2	2	4	4
PASADILLA	5	5	6	6	7	8	8		
PENCE	5	6	6	6	5	5	5	7	7
POSEY	5	6	6	6	5	4	4	3	3
SABENE	5	5	5	1	1	1	5	4	6
SALTERS	5	5.5	10.5	10.5	10.5	6	5		
STEELE	5	6	6	6	4.5	4.5	4.5	6	7
TOY	5	5	5.5	5	5	5	6	7	4
WRIGHT	5	5	5	6	5	5	6	6	8
YOUNG	5	6	6	6.5	6.5	7	7		
MODEL	3.47	3.43	3.64	4.19	5.05	5.89	6.59	6.33	6.44

2. Each line of the documentation represents a separate cell block address, with the respective contents of that given cell. For example, in the first listing the cell block address A1: contains the text "GROUP A WORKFORCE". Linkage between this spreadsheet and the data input spreadsheet is made with file links. For example, the first link is listed at cell address B11, where ([AWFMOD]B5-[AWFACT]B5) represents the formula $(A_t-S_t)^2$ for the student named BELL at $t = 0$.

```
A1: [W14] 'GROUP A WORKFORCE
A5: [W14] 'Computation for the square root of the sum of squared errors
A6: [W14] 'for time t
B8: [W6] '(At-St)^2 Table
B10: [W6] 0
C10: [W6] 40
D10: [W7] 80
E10: [W7] 120
F10: [W6] 160
G10: [W6] 200
H10: [W6] 240
I10: [W6] 280
J10: [W6] 320
A11: [W14] 'BELL
B11: (F2) [W6] ([AWFMOD]B5-[AWFACT]B5)^2
C11: (F2) [W6] ([AWFMOD]C5-[AWFACT]C5)^2
D11: (F2) [W7] ([AWFMOD]D5-[AWFACT]D5)^2
E11: (F2) [W7] ([AWFMOD]E5-[AWFACT]E5)^2
F11: (F2) [W6] ([AWFMOD]F5-[AWFACT]F5)^2
G11: (F2) [W6] ([AWFMOD]G5-[AWFACT]G5)^2
H11: (F2) [W6] ([AWFMOD]H5-[AWFACT]H5)^2
I11: (F2) [W6] ([AWFMOD]I5-[AWFACT]I5)^2
J11: (F2) [W6] ([AWFMOD]J5-[AWFACT]J5)^2
A12: [W14] 'BITTNER
B12: (F2) [W6] ([AWFMOD]B6-[AWFACT]B6)^2
C12: (F2) [W6] ([AWFMOD]C6-[AWFACT]C6)^2
D12: (F2) [W7] ([AWFMOD]D6-[AWFACT]D6)^2
E12: (F2) [W7] ([AWFMOD]E6-[AWFACT]E6)^2
F12: (F2) [W6] ([AWFMOD]F6-[AWFACT]F6)^2
G12: (F2) [W6] ([AWFMOD]G6-[AWFACT]G6)^2
H12: (F2) [W6] ([AWFMOD]H6-[AWFACT]H6)^2
I12: (F2) [W6] ([AWFMOD]I6-[AWFACT]I6)^2
J12: (F2) [W6] ([AWFMOD]J6-[AWFACT]J6)^2
A13: [W14] 'BRANLEY
B13: (F2) [W6] ([AWFMOD]B7-[AWFACT]B7)^2
C13: (F2) [W6] ([AWFMOD]C7-[AWFACT]C7)^2
D13: (F2) [W7] ([AWFMOD]D7-[AWFACT]D7)^2
E13: (F2) [W7] ([AWFMOD]E7-[AWFACT]E7)^2
F13: (F2) [W6] ([AWFMOD]F7-[AWFACT]F7)^2
G13: (F2) [W6] ([AWFMOD]G7-[AWFACT]G7)^2
H13: (F2) [W6] ([AWFMOD]H7-[AWFACT]H7)^2
I13: (F2) [W6] ([AWFMOD]I7-[AWFACT]I7)^2
J13: (F2) [W6] ([AWFMOD]J7-[AWFACT]J7)^2
```

```
A14: [W14] 'CHELOUCHE
B14: (F2) [W6] ([AWFMOD]B8-[AWFACT]B8)^2
C14: (F2) [W6] ([AWFMOD]C8-[AWFACT]C8)^2
D14: (F2) [W7] ([AWFMOD]D8-[AWFACT]D8)^2
E14: (F2) [W7] ([AWFMOD]E8-[AWFACT]E8)^2
F14: (F2) [W6] ([AWFMOD]F8-[AWFACT]F8)^2
G14: (F2) [W6] ([AWFMOD]G8-[AWFACT]G8)^2
H14: (F2) [W6] ([AWFMOD]H8-[AWFACT]H8)^2
I14: (F2) [W6] ([AWFMOD]I8-[AWFACT]I8)^2
J14: (F2) [W6] ([AWFMOD]J8-[AWFACT]J8)^2
A15: [W14] 'CULPEPPER
B15: (F2) [W6] ([AWFMOD]B9-[AWFACT]B9)^2
C15: (F2) [W6] ([AWFMOD]C9-[AWFACT]C9)^2
D15: (F2) [W7] ([AWFMOD]D9-[AWFACT]D9)^2
E15: (F2) [W7] ([AWFMOD]E9-[AWFACT]E9)^2
F15: (F2) [W6] ([AWFMOD]F9-[AWFACT]F9)^2
G15: (F2) [W6] ([AWFMOD]G9-[AWFACT]G9)^2
H15: (F2) [W6] ([AWFMOD]H9-[AWFACT]H9)^2
I15: (F2) [W6] ([AWFMOD]I9-[AWFACT]I9)^2
J15: (F2) [W6] ([AWFMOD]J9-[AWFACT]J9)^2
A16: [W14] 'FEY
B16: (F2) [W6] ([AWFMOD]B10-[AWFACT]B10)^2
C16: (F2) [W6] ([AWFMOD]C10-[AWFACT]C10)^2
D16: (F2) [W7] ([AWFMOD]D10-[AWFACT]D10)^2
E16: (F2) [W7] ([AWFMOD]E10-[AWFACT]E10)^2
F16: (F2) [W6] ([AWFMOD]F10-[AWFACT]F10)^2
G16: (F2) [W6] ([AWFMOD]G10-[AWFACT]G10)^2
H16: (F2) [W6] ([AWFMOD]H10-[AWFACT]H10)^2
I16: (F2) [W6] ([AWFMOD]I10-[AWFACT]I10)^2
J16: (F2) [W6] ([AWFMOD]J10-[AWFACT]J10)^2
A17: [W14] 'FOSTER, T
B17: (F2) [W6] ([AWFMOD]B11-[AWFACT]B11)^2
C17: (F2) [W6] ([AWFMOD]C11-[AWFACT]C11)^2
D17: (F2) [W7] ([AWFMOD]D11-[AWFACT]D11)^2
E17: (F2) [W7] ([AWFMOD]E11-[AWFACT]E11)^2
F17: (F2) [W6] ([AWFMOD]F11-[AWFACT]F11)^2
G17: (F2) [W6] ([AWFMOD]G11-[AWFACT]G11)^2
H17: (F2) [W6] ([AWFMOD]H11-[AWFACT]H11)^2
I17: (F2) [W6] ([AWFMOD]I11-[AWFACT]I11)^2
J17: (F2) [W6] ([AWFMOD]J11-[AWFACT]J11)^2
A18: [W14] 'HODGKINS
B18: (F2) [W6] ([AWFMOD]B12-[AWFACT]E12)^2
C18: (F2) [W6] ([AWFMOD]C12-[AWFACT]C12)^2
D18: (F2) [W7] ([AWFMOD]D12-[AWFACT]D12)^2
E18: (F2) [W7] ([AWFMOD]E12-[AWFACT]E12)^2
F18: (F2) [W6] ([AWFMOD]F12-[AWFACT]F12)^2
G18: (F2) [W6] ([AWFMOD]G12-[AWFACT]G12)^2
H18: (F2) [W6] ([AWFMOD]H12-[AWFACT]H12)^2
I18: (F2) [W6] ([AWFMOD]I12-[AWFACT]I12)^2
J18: (F2) [W6] ([AWFMOD]J12-[AWFACT]J12)^2
A19: [W14] 'IVEY
B19: (F2) [W6] ([AWFMOD]B13-[AWFACT]B13)^2
C19: (F2) [W6] ([AWFMOD]C13-[AWFACT]C13)^2
D19: (F2) [W7] ([AWFMOD]D13-[AWFACT]D13)^2
E19: (F2) [W7] ([AWFMOD]E13-[AWFACT]E13)^2
```

```
F19: (F2) [W6] ([AWFMOD]F13-[AWFACT]F13)^2
G19: (F2) [W6] ([AWFMOD]G13-[AWFACT]G13)^2
H19: (F2) [W6] ([AWFMOD]H13-[AWFACT]H13)^2
I19: (F2) [W6] ([AWFMOD]I13-[AWFACT]I13)^2
J19: (F2) [W6] ([AWFMOD]J13-[AWFACT]J13)^2
A20: [W14] 'LACO
B20: (F2) [W6] ([AWFMOD]B14-[AWFACT]B14)^2
C20: (F2) [W6] ([AWFMOD]C14-[AWFACT]C14)^2
D20: (F2) [W7] ([AWFMOD]D14-[AWFACT]D14)^2
E20: (F2) [W7] ([AWFMOD]E14-[AWFACT]E14)^2
F20: (F2) [W6] ([AWFMOD]F14-[AWFACT]F14)^2
G20: (F2) [W6] ([AWFMOD]G14-[AWFACT]G14)^2
H20: (F2) [W6] ([AWFMOD]H14-[AWFACT]H14)^2
I20: (F2) [W6] ([AWFMOD]I14-[AWFACT]I14)^2
J20: (F2) [W6] ([AWFMOD]J14-[AWFACT]J14)^2
A21: [W14] 'LOCKHART
B21: (F2) [W6] ([AWFMOD]B15-[AWFACT]B15)^2
C21: (F2) [W6] ([AWFMOD]C15-[AWFACT]C15)^2
D21: (F2) [W7] ([AWFMOD]D15-[AWFACT]D15)^2
E21: (F2) [W7] ([AWFMOD]E15-[AWFACT]E15)^2
F21: (F2) [W6] ([AWFMOD]F15-[AWFACT]F15)^2
G21: (F2) [W6] ([AWFMOD]G15-[AWFACT]G15)^2
H21: (F2) [W6] ([AWFMOD]H15-[AWFACT]H15)^2
I21: (F2) [W6] ([AWFMOD]I15-[AWFACT]I15)^2
J21: (F2) [W6] ([AWFMOD]J15-[AWFACT]J15)^2
A22: [W14] 'MAIN
B22: (F2) [W6] ([AWFMOD]B16-[AWFACT]B16)^2
C22: (F2) [W6] ([AWFMOD]C16-[AWFACT]C16)^2
D22: (F2) [W7] ([AWFMOD]D16-[AWFACT]D16)^2
E22: (F2) [W7] ([AWFMOD]E16-[AWFACT]E16)^2
F22: (F2) [W6] ([AWFMOD]F16-[AWFACT]F16)^2
G22: (F2) [W6] ([AWFMOD]G16-[AWFACT]G16)^2
H22: (F2) [W6] ([AWFMOD]H16-[AWFACT]H16)^2
I22: (F2) [W6] ([AWFMOD]I16-[AWFACT]I16)^2
J22: (F2) [W6] ([AWFMOD]J16-[AWFACT]J16)^2
A23: [W14] 'METCALF
B23: (F2) [W6] ([AWFMOD]B17-[AWFACT]B17)^2
C23: (F2) [W6] ([AWFMOD]C17-[AWFACT]C17)^2
D23: (F2) [W7] ([AWFMOD]D17-[AWFACT]D17)^2
E23: (F2) [W7] ([AWFMOD]E17-[AWFACT]E17)^2
F23: (F2) [W6] ([AWFMOD]F17-[AWFACT]F17)^2
G23: (F2) [W6] ([AWFMOD]G17-[AWFACT]G17)^2
H23: (F2) [W6] ([AWFMOD]H17-[AWFACT]H17)^2
I23: (F2) [W6] ([AWFMOD]I17-[AWFACT]I17)^2
J23: (F2) [W6] ([AWFMOD]J17-[AWFACT]J17)^2
A24: [W14] 'PASADILLA
B24: (F2) [W6] ([AWFMOD]B18-[AWFACT]B18)^2
C24: (F2) [W6] ([AWFMOD]C18-[AWFACT]C18)^2
D24: (F2) [W7] ([AWFMOD]D18-[AWFACT]D18)^2
E24: (F2) [W7] ([AWFMOD]E18-[AWFACT]E18)^2
F24: (F2) [W6] ([AWFMOD]F18-[AWFACT]F18)^2
G24: (F2) [W6] ([AWFMOD]G18-[AWFACT]G18)^2
H24: (F2) [W6] ([AWFMOD]H18-[AWFACT]H18)^2
I24: (F2) [W6] ([AWFMOD]I18-[AWFACT]I18)^2
J24: (F2) [W6] ([AWFMOD]J18-[AWFACT]J18)^2
```

```
A25: [W14] 'PENCE
B25: (F2) [W6] ([AWFMOD]B19-[AWFACT]B19)^2
C25: (F2) [W6] ([AWFMOD]C19-[AWFACT]C19)^2
D25: (F2) [W7] ([AWFMOD]D19-[AWFACT]D19)^2
E25: (F2) [W7] ([AWFMOD]E19-[AWFACT]E19)^2
F25: (F2) [W6] ([AWFMOD]F19-[AWFACT]F19)^2
G25: (F2) [W6] ([AWFMOD]G19-[AWFACT]G19)^2
H25: (F2) [W6] ([AWFMOD]H19-[AWFACT]H19)^2
I25: (F2) [W6] ([AWFMOD]I19-[AWFACT]I19)^2
J25: (F2) [W6] ([AWFMOD]J19-[AWFACT]J19)^2
A26: [W14] 'POSEY
B26: (F2) [W6] ([AWFMOD]B20-[AWFACT]B20)^2
C26: (F2) [W6] ([AWFMOD]C20-[AWFACT]C20)^2
D26: (F2) [W7] ([AWFMOD]D20-[AWFACT]D20)^2
E26: (F2) [W7] ([AWFMOD]E20-[AWFACT]E20)^2
F26: (F2) [W6] ([AWFMOD]F20-[AWFACT]F20)^2
G26: (F2) [W6] ([AWFMOD]G20-[AWFACT]G20)^2
H26: (F2) [W6] ([AWFMOD]H20-[AWFACT]H20)^2
I26: (F2) [W6] ([AWFMOD]I20-[AWFACT]I20)^2
J26: (F2) [W6] ([AWFMOD]J20-[AWFACT]J20)^2
A27: [W14] 'SABENE
B27: (F2) [W6] ([AWFMOD]B21-[AWFACT]B21)^2
C27: (F2) [W6] ([AWFMOD]C21-[AWFACT]C21)^2
D27: (F2) [W7] ([AWFMOD]D21-[AWFACT]D21)^2
E27: (F2) [W7] ([AWFMOD]E21-[AWFACT]E21)^2
F27: (F2) [W6] ([AWFMOD]F21-[AWFACT]F21)^2
G27: (F2) [W6] ([AWFMOD]G21-[AWFACT]G21)^2
H27: (F2) [W6] ([AWFMOD]H21-[AWFACT]H21)^2
I27: (F2) [W6] ([AWFMOD]I21-[AWFACT]I21)^2
J27: (F2) [W6] ([AWFMOD]J21-[AWFACT]J21)^2
A28: [W14] 'SALTERS
B28: (F2) [W6] ([AWFMOD]B22-[AWFACT]B22)^2
C28: (F2) [W6] ([AWFMOD]C22-[AWFACT]C22)^2
D28: (F2) [W7] ([AWFMOD]D22-[AWFACT]D22)^2
E28: (F2) [W7] ([AWFMOD]E22-[AWFACT]E22)^2
F28: (F2) [W6] ([AWFMOD]F22-[AWFACT]F22)^2
G28: (F2) [W6] ([AWFMOD]G22-[AWFACT]G22)^2
H28: (F2) [W6] ([AWFMOD]H22-[AWFACT]H22)^2
I28: (F2) [W6] ([AWFMOD]I22-[AWFACT]I22)^2
J28: (F2) [W6] ([AWFMOD]J22-[AWFACT]J22)^2
A29: [W14] 'STEELE
B29: (F2) [W6] ([AWFMOD]B23-[AWFACT]B23)^2
C29: (F2) [W6] ([AWFMOD]C23-[AWFACT]C23)^2
D29: (F2) [W7] ([AWFMOD]D23-[AWFACT]D23)^2
E29: (F2) [W7] ([AWFMOD]E23-[AWFACT]E23)^2
F29: (F2) [W6] ([AWFMOD]F23-[AWFACT]F23)^2
G29: (F2) [W6] ([AWFMOD]G23-[AWFACT]G23)^2
H29: (F2) [W6] ([AWFMOD]H23-[AWFACT]H23)^2
I29: (F2) [W6] ([AWFMOD]I23-[AWFACT]I23)^2
J29: (F2) [W6] ([AWFMOD]J23-[AWFACT]J23)^2
A30: [W14] 'TOY
B30: (F2) [W6] ([AWFMOD]B24-[AWFACT]B24)^2
C30: (F2) [W6] ([AWFMOD]C24-[AWFACT]C24)^2
D30: (F2) [W7] ([AWFMOD]D24-[AWFACT]D24)^2
E30: (F2) [W7] ([AWFMOD]E24-[AWFACT]E24)^2
```

```
F30:  (F2)  [W6]  ([AWFMOD]F24-[AWFACT]F24)^2
G30:  (F2)  [W6]  ([AWFMOD]G24-[AWFACT]G24)^2
H30:  (F2)  [W6]  ([AWFMOD]H24-[AWFACT]H24)^2
I30:  (F2)  [W6]  ([AWFMOD]I24-[AWFACT]I24)^2
J30:  (F2)  [W6]  ([AWFMOD]J24-[AWFACT]J24)^2
A31:  [W14]  'WRIGHT
B31:  (F2)  [W6]  ([AWFMOD]B25-[AWFACT]B25)^2
C31:  (F2)  [W6]  ([AWFMOD]C25-[AWFACT]C25)^2
D31:  (F2)  [W7]  ([AWFMOD]D25-[AWFACT]D25)^2
E31:  (F2)  [W7]  ([AWFMOD]E25-[AWFACT]E25)^2
F31:  (F2)  [W6]  ([AWFMOD]F25-[AWFACT]F25)^2
G31:  (F2)  [W6]  ([AWFMOD]G25-[AWFACT]G25)^2
H31:  (F2)  [W6]  ([AWFMOD]H25-[AWFACT]H25)^2
I31:  (F2)  [W6]  ([AWFMOD]I25-[AWFACT]I25)^2
J31:  (F2)  [W6]  ([AWFMOD]J25-[AWFACT]J25)^2
A32:  [W14]  'YOUNG
B32:  (F2)  [W6]  ([AWFMOD]B26-[AWFACT]B26)^2
C32:  (F2)  [W6]  ([AWFMOD]C26-[AWFACT]C26)^2
D32:  (F2)  [W7]  ([AWFMOD]D26-[AWFACT]D26)^2
E32:  (F2)  [W7]  ([AWFMOD]E26-[AWFACT]E26)^2
F32:  (F2)  [W6]  ([AWFMOD]F26-[AWFACT]F26)^2
G32:  (F2)  [W6]  ([AWFMOD]G26-[AWFACT]G26)^2
H32:  (F2)  [W6]  ([AWFMOD]H26-[AWFACT]H26)^2
I32:  (F2)  [W6]  ([AWFMOD]I26-[AWFACT]I26)^2
J32:  (F2)  [W6]  ([AWFMOD]J26-[AWFACT]J26)^2
A34:  [W14]  'TOTAL
B34:  (F2)  [W6]  @SUM(B11..B32)
C34:  (F2)  [W6]  @SUM(C11..C32)
D34:  (F2)  [W7]  @SUM(D11..D32)
E34:  (F2)  [W7]  @SUM(E11..E32)
F34:  (F2)  [W6]  @SUM(F11..F32)
G34:  (F2)  [W6]  @SUM(G11..G32)
H34:  (F2)  [W6]  @SUM(H11..H32)
I34:  (F2)  [W6]  @SUM(I11..I32)
J34:  (F2)  [W6]  @SUM(J11..J32)
A38:  [W14]  '
B38:  (F2)  [W6]  +B34^0.5/[AWFACT]B28
C38:  (F2)  [W6]  +C34^0.5/[AWFACT]C28
D38:  (F2)  [W7]  +D34^0.5/[AWFACT]D28
D38:  (F2)  [W7]  +E34^0.5/[AWFACT]E28
F38:  (F2)  [W6]  +F34^0.5/[AWFACT]F28
G38:  (F2)  [W6]  +G34^0.5/[AWFACT]G28
H38:  (F2)  [W6]  +H34^0.5/[AWFACT]H28
I38:  (F2)  [W6]  +I34^0.5/[AWFACT]I28
J38:  (F2)  [W6]  +J34^0.5/[AWFACT]J28
```

LIST OF REFERENCES

Abdel-Hamid, T. K. and Madnick, S. E., *Software Project Dynamics: An Integrated Approach*, Prentice Hall Inc., 1991.

Abdel-Hamid, T. K. and Madnick, S. E., "Lessons Learned from Modeling the Dynamics of Software Development," *Communications of the ACM*, v. 31, n. 12, December 1989, 1426-1455.

Abdel-Hamid, T. K., "A Multiproject Perspective of Single-Project Dynamics," Paper submitted to *IEEE Transactions on Software Engineering*, November 1990.

Baker, D. L., *An Experimental Investigation of the Effects of Software Size Increase on Software Project Management Behavior*, Masters Thesis, Naval Postgraduate School, Monterey, California, March 1992.

Barlas, Y. and Carpenter, S., "Philosophical Roots of Model Validation: Two Paradigms," *System Dynamics Review*, v. 6, n. 2, 1990, 148-166.

Barlas, Y., "Multiple Tests for Validation of System Dynamics Type of Simulation Models," *European Journal of Operational Research*, v. 42, 1989, 59-87.

Bloomfield, B. P., *Modelling the World*, Basil Blackwell, 1986.

Brooks, F. P., *The Mythical Man Month*, Addison-Wesley, 1975.

Bush, R. R., and Mosteller, F., *Stochastic Models for Learning*, Wiley, 1955.

DeMarco, T., *Controlling Software Projects*, Yourdon Press, New York, 1982, 19.

Forrester, J. W. and Senge, P. M., "Tests for Building Confidence in System Dynamics Models," *TIMS Studies in the Management Sciences*, v., 14, 1980, 209-228.

Naylor, T. H., *Computer Simulation Experiments with Models of Economic Systems*, Wiley, 1971.

Pindyck, R. S. and Rubenfield, D. L., *Econometric Models and Economic Forecasts*, McGraw-Hill Inc., 1991, 336-342.

Richardson, G. P. and Pugh, A. L. III, *Introduction to System Dynamics Modeling with Dynamo,* The MIT Press, 1981, 310-314.

Ronan, D., *Decision Making Heuristics and Biases in Software Project Management: An Experimental Investigation,* Masters Thesis, Naval Postgraduate School, Monterey, California, March 1990.

Rowland, J. R. and Holmes, W. M., "Simulation Validation with Sparse Random Data," *Computers and Electrical Engineering,* v. 5, 1978, 37-49.

Schlender, B. R., "How to Break the Software Logjam," *Fortune,* September 1989.

Senge, P. M., "Some Issues in Evaluating the Validity of Social Models," *Proceedings 1973 Summer Computer Simulation Conference, 1973,* 1176-1181.

Sterman, J. D., "Appropriate Summary Statistics for Evaluating Historical Fit of System Dynamics Models," *Dynamica,* v. 10, 1984, 51-66.

Sterman, J. D., "Misperceptions of Feedback in Dynamic Decision Making," *Organizational Behavior and Human Decision Processes,* v. 43, 1989, 301-335.

Sterman, J. D., "Testing Behavioral Simulation Models by Direct Experiment," *Management Science,* v. 33, No. 12, December 1987, 1572-1592.

Theil, H., *Applied Economic Forecasting,* North-Holland, Amsterdam, 1966, 26-35.

Theil, H., *Economic Forecasts and Policy,* North-Holland, Amsterdam, 1961, 30-37.

Veit, K. P., "System Dynamics in Corporate Long-Range Strategic Planning," *Managerial Applications of System Dynamics,* The MIT Press, 1978.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
Cameron Station
Alexandria, Virginia 22304-6145

2. Library, Code 0142 2
Naval Postgraduate School
Monterey, California 93943-5002

3. Dr. Tarek K. Abdel-Hamid, Code AS/Ha 7
Department of Administrative Sciences
Naval Postgraduate School
Monterey, California 93943-5000

4. Professor Linda Gorman, Code AS/Gr 1
Department of Administrative Sciences
Naval Postgraduate School
Monterey, California 93943-5000

5. Todd Stephan 1
856 Lomita St.
Monterey, California 93940

6. Computer Technology Curricular 1
Office, Code 37
Naval Postgraduate School
Monterey, California 93943-5000

7. Commandant of the Marine Corps 1
Code TE 06
Headquarters, U.S. Marine Corps
Washington, D.C. 20380-0001