A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY

Military Operations Research

Volume 3 Number 1
1997

A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY

MORS
MILITARY OPERATIONS RESEARCH SOCIETY

Military Operations Research

A publication of the Military Operations Research Society

The Military Operations Research Society is a professional society incorporated under the laws of Virginia. The Society conducts a classified symposium and several other meetings annually. It publishes proceedings, monographs, a quarterly bulletin, *PHALANX*, and a quarterly journal, *Military Operations Research*, for professional exchange and peer criticism among students, theoreticians, practitioners and users of military operations research. The Society does not make or advocate official policy nor does it attempt to influence the formation of policy. Matters discussed or statements made in the course of MORS symposia or printed in its publications represent the opinions of the authors and not the Society.

Military Operations Research, the journal of the Military Operations Research Society (ISSN 0275-5823) is published quarterly by the Military Operations Research Society, 101 South Whiting Street, Suite 202, Alexandria, VA 22304-3418. The domestic subscription price is $40 for one year and $75 for two years; international rates are $80 for one year and $150 for two years. Periodicals Postage Paid at Alexandria, VA. and additional mailing offices.

POSTMASTER: Send address changes to *Military Operations Research*, the journal of the Military Operations Research Society, 101 South Whiting Street, Suite 202, Alexandria, VA 22304.

Table of Contents

Volume 3, Number 1

1997

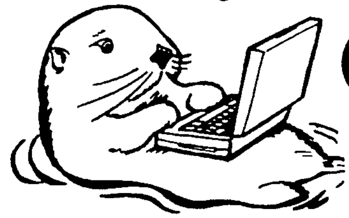
THE MILITARY OPERATIONS RESEARCH SOCIETY SYMPOSIUM (MORSS)

Currently, an average of over 950 military operations research analysts attend the annual Military Operations Research Society Symposium (MORSS), a national professional society under the sponsorship of the Joint Staff, US Army, US Navy, US Air Force, US Marine Corps and the Office of the Secretary of Defense. The attendees represent the services, the private sector, academe, and other government agencies. The MORS Symposium provides them with a unique opportunity to hear the thoughts of influential military leaders. It also provides them with the opportunity to exchange information and examine completed or ongoing research in a classified setting.

The symposium consists of a series of meetings centered around 32 working groups. The working groups also meet in composite sessions to address a wider spectrum of topics, which are of interest to their associated composite group. In addition, MORS offers general sessions and tutorials. The general/special sessions include thematically related invited papers of broad interest, selected best working group papers, workshop reports and sessions, and an education session. The tutorials are generally educational sessions on tools and techniques of operations research.

The 66th MORS Symposium will be held at the Naval Postgraduate School (NPS) in Monterey, California on 23-25 June 1998. This year's theme, *Preparing for Military Operations Research in the 21st Century*, was echoed in remarks recently made by the Chief of Naval Operations concerning graduate education at NPS: "Students will expand their breadth of knowledge in a particular discipline and will reinvigorate their ability to successfully analyze and solve the complex challenges we face. These important skills will help guide our Navy into the 21st Century through fresh thinking and innovation." Because of its mission, the Naval Postgraduate School is the perfect setting for our society to prepare a path to the next century.

23-24-25 June 1998



66th MORSS

Naval Postgraduate School

This year's symposium will consist of a keynote session on Tuesday, three general sessions, eight composite group/working group sessions, three tutorial sessions during lunch, two poster sessions, and the social event Wednesday evening.

So, mark your calendars now for the 66th Military Operations Research Society Symposium at the historic Naval Postgraduate School from 23-25 June 1998. Get ready to enjoy one of the most beautiful spots on the face of the earth — the Monterey Peninsula. Help us provide a path to the next century by contributing to a successful symposium with an early response to the Call for Papers and submitting your "approved for public release" abstract in January. And while you're convincing your boss to let you go to the Symposium, convince him or her to go with you! If you have any doubts at all on presenting a paper, have any other questions, have not received the ACP, or would like to assist the program staff, please contact the MORS office at (703) 751-7290 or CDR Kirk Michealson at (703) 697-0064.

Executive Summaries

MULTIPOLAR NUCLEAR STABILITY: INCENTIVES TO STRIKE AND INCENTIVES TO PREEMPT

by Jerome Bracken

First-strike stability, or incentive to preempt in a strategic nuclear environment, underlies many discussions of strategic force structure, posture and arms control. In the bipolar context it is generally agreed that defenses are destabilizing, particularly at medium-to-high levels. The impact of additional armed sides on stability, however, is not well understood. The objective of this paper is to shed light on this multipolar stability question, particularly on the stability implications of small-to-medium-size defenses.

GAIN-SHARING, SUCCESS-SHARING AND COST-BASED TRANSFER PRICING: A NEW BUDGETING APPROACH FOR THE DEPARTMENT OF DEFENSE (DOD)

by Francois Melese

Once the dust settles and so-called "core" DoD support activities are defined, the issue of incentives must be addressed. Do current budgeting systems inadvertently punish cost-savings? Does cost-based "transfer pricing" offer a viable alternative? This paper by Francois Melese offers some answers. It also pioneers the application of an analytical OR approach to budget and incentive problems. A new budgeting proposal is introduced to help govern peacetime relationships among operating forces and internal support activities. This new budgeting approach integrates cost-based "transfer pricing" with the popular business practice of "gain-sharing" and a novel incentive program called "success-sharing." The combined budgeting and incentive package is designed to improve the management of financial resources and business operations throughout the DoD.

MODELING LOSS EXCHANGE RATIOS AS INVERSE GAUSSIAN VARIATES: IMPLICATIONS

by D. H. Olwell

Loss Exchange Ratios (LER) are widely used as a summary of the results of a simulated battle. We have found from repeated simulations of the same battles that the LER for a given battle varies widely. This has policy and statistical implications. Attacking some of the statistical issues, such as what is a good model for LER and how to estimate its parameters, sheds light on the policy issues, such as how many runs are required and how accurate is the output.

We examine in detail LER from replications of battles simulated on JANUS and CASTFOREM, and propose a new statistical model for this output, based on the Inverse Gaussian distribution. Since this member of the exponential family is not widely known, we include a primer. We make the point that the LER can not be adequately summarized by just its mean; a second parameter is necessary to fully describe the model.

Executives should routinely ask their modelers how many runs of a simulation were conducted in a simulation study, how that number was selected, how the LER (or other summary output) is best described, and how much improvement has been demonstrated over the controls. This article provides the analyst with powerful tools to answer those questions.

CLUSTERIZATION OF ALTERNATIVES IN THE ANALYTIC HIERARCHY PROCESS

by R. Islam, M.P. Biswal and S.S. Alam

Despite the impressive success of the Analytic Hierarchy Process (AHP) in solving a wide range of discrete multi-criteria decision making (MCDM) problems, a major drawback of its use to solve a large scale problem is the huge amount of work to make all the necessary pairwise comparisons. R. Islam, M.P. Biswal and S.S. Alam have developed a clustering procedure to solve a large scale MCDM problem by AHP. In order to show the viability of the proposed procedure they have considered the problem of choosing the best transport aircraft from the available twenty alternatives.

EVALUATING THE EFFECTIVENESS OF SHOOT-LOOK-SHOOT TACTICS IN THE PRESENCE OF INCOMPLETE DAMAGE INFORMATION

by Yossi Aviv and Moshe Kress

The advent of new technologies, and in particular the development of efficient detection sensors, have produced advanced long-range and accurate weapon systems such as HELLFIRE, SLAM, GBU-15 and FOG-M. A common problem that is associated with the operation of these expensive weapon systems is that of Bomb Damage Assessment (BDA). Stating it simply, the question is: how to utilize these weapon systems in the most efficient way when damage information is not complete?

In this paper, Yossi Aviv and Moshe Kress analyze several shooting tactics and show that a certain simple tactic may be best in terms of fire efficiency and operational convenience. This result sheds some light on doctrinal issues that may be related to the operation of thse weapon systems.

Multipolar Nuclear Stability: Incentives to Strike and Incentives to Preempt*

Jerome Bracken

Yale University
New Haven, CT 06520
Phone: (301)654-9320
Fax: (301)652-0817
Email: jerbracken@aol.com

APPLICATION AREA:
Strategic Arms Control

OR METHODOLOGY:
Game Theory

A. OVERVIEW

This paper presents a concise, parsimonious theory of multipolar nuclear stability. The framework is intended to enable understanding of force structure and other relationships which affect incentives to strike and incentives to preempt, unilaterally or in coalitions.

The paper addresses warfare involving three armed sides and one unarmed side. There are presently five major nuclear powers: the United States, Russia, Britain, France and China. The aggregation into three armed sides and one unarmed side analyzed in the example presented here is (1) United States, Britain and France, (2) Russia, (3) China and (4) "Rest of the World".

B. BACKGROUND

First-strike stability in a bipolar world is defined and analyzed in Kent and Thaler (1989) and, with particular attention to the impact of defense, in Kent and Thaler (1990). The concept of first-strike stability, addressing the incentive to preempt for both sides due to the "reciprocal fear of surprise attack", was first suggested by Schelling (1960).

Incentives to strike and first-strike stability in a multipolar world are defined and analyzed in Bracken and Shubik (1993). Five nuclear powers are organized into all possible coalitions of two sides.

First-strike stability and strategic defense is the focus of Best and Bracken (1993) and Best and Bracken (1995), where five powers are organized into all possible coalitions of two sides. The cost function from Kent and Thaler (1989, 1990) is adopted as the basic measure of utility, and a multipolar extension of their bipolar first-strike stability measure is presented. The insight is derived that though defenses are uniformly destabilizing in a bipolar world, defenses which are large relative to the smaller powers and small relative to the larger powers are stabilizing in the multipolar world.

First-strike stability with three players acting separately is introduced in Best and Bracken (1994).

The present paper allows three players to act separately or to form two coalitions. There are twelve possible wars. A measure of utility is suggested. Behavior of the three armed sides is modeled. Both incentives to strike and first-strike stability are addressed. An example is given with several variations. Alternative utility functions are discussed. Of particular interest is the role of the unarmed, fourth side in influencing the incentives to strike and the first-strike stability of the three armed sides.

C. RESOURCES AND MEASURES

1. Resources

There are three armed sides and one unarmed ("rest of the world") side. The data describing the four sides are:

V_1, V_2, V_3, V_4

= value targets of sides 1, 2, 3, 4

O_1, O_2, O_3

= offensive weapons of sides 1, 2, 3

D_1, D_2, D_3, D_4

= defensive weapons of sides 1, 2, 3, 4

Some of the offensive weapons of sides 1, 2 and 3 may be vulnerable.

During the war, all weapons are expended. The value targets before and after the war are:

$V_1^1, V_2^1, V_3^1, V_4^1$

= value targets before the war

$V_1^2, V_2^2, V_3^2, V_4^2$

= value targets after the war

**This is an expansion with modifications of a previous paper. Copyright information form the original publication follows:*

Jerome Bracken, "Multipolar Nuclear Stability: Incentives to Strike and Incentives to Preempt", In Melvin L. Best, jr., John Hughes-Wilson and Andrei Piontkiowsky (editors), *Strategic Stability in the Post-Cold War World and the Future of Nuclear Disarmament*, pages 203–220. Printed in the Netherlands.

2. Utility

The utility of a side is dependent on the state of the world with respect to value targets. The utilities of Side i before and after the war are:

$$U_i^1 = \frac{V_i^1}{V_1^1 + V_2^1 + V_3^1 + V_4^1}$$

$$U_i^2 = \frac{V_i^2}{V_1^2 + V_2^2 + V_3^2 + V_4^2}$$

That is, the utilities of Side i are the fractions of the world's value targets possessed by Side i before the war and after the war, respectively.

3. Incentive to Strike

Incentive to strike for Side i is the ratio of the utility of Side i after the war to the utility of Side i before the war, as follows:

$$I_i = \frac{U_i^2}{U_i^1}$$

If incentive to strike is greater than one, Side 1 has a larger fraction of the world's total value targets after the war than before the war and is better off from striking. If it is equal to one, Side 1 has the same fraction and is the same before and after striking. If it is less than one, Side i has a smaller fraction and is worse off from striking.

Later we will introduce a measure which is not simply a ratio of the fraction of the world's value targets possessed by a side, since one or more sides may abhor a war, or may at least abhor the loss of their own value. However, the above function is the most simple measure of the warfare-related change in relative value of a side as the result of a war.

4. First-Strike Stability (Incentive Not to Preempt)

First-strike stability differs from incentive to strike in that it compares the utility of a side if it strikes first to the utility of a side if it strikes other than first, and then examines all sides to determine which side fares relatively worst from not striking first.

First-strike stability of Side i is as follows:

$$S_i = \frac{U_i^2 \text{ of worst outcome for Side i for all first strikes by other sides}}{U_i^2 \text{ of worst outcome for Side i for all first strikes by Side i}}$$

The first-strike stability of Side i is the ratio of the minimum U_i^2, or worst outcome, from striking other than first, to the minimum U_i^2, or worst outcome, from striking first. The first-strike stability lies between 0 and 1.

The computation of first-strike stability of Side i further assumes that possible first strikes by others will include two sides in coalition against Side i. However, first strikes by Side i will not include another side in coalition with Side i. In other words, others can act in a coalition against Side i, but Side i cannot plan on acting in a coalition.

The first-strike stability of all of the sides taken together is:

$$S = \text{minimum } (S_1, S_2, S_3)$$

This can be interpreted as follows. The first strike will be undertaken by the side with the lowest first-strike stability, so all sides have the same incentive to strike first as has the side with the lowest first-strike stability.

D. POSSIBLE WARS AND ASSUMED STRATEGIES

There are twelve possible wars among three armed sides and one unarmed side. The armed sides may act separately in the following six orders: 1 2 3, 1 3 2, 2 1 3, 2 3 1, 3 1 2 and 3 2 1. The armed sides may form two coalitions and act in the following orders: 1 23, 2 13, 3 12, 12 3, 13 2 and 23 1. One or more of the three armed sides may attack the unarmed side, depending on the calculation of the optimal allocations.

For each sequence of actions, the strikers know the sequence and the behavioral motivations of the other strikers.

1. Three Sides Acting Separately

When three sides act separately they may attack in any of the six orders. Denote the first striker by a, the second striker by b and the third striker by c. Denote the fourth side by d.

The first striker a attacks b forces, c forces, b value, c value and d value, maximizing the utility of the first striker. The second striker b attacks c forces, a value, c value and d value, maximizing the utility of the second striker. The third striker c attacks a value, b value and d value, maximizing the utility of the third striker.

The optimal strategies of the three sides are found by working backward from the end of the war. Each first and second strike leads to an optimal third strike. Each first strike leads to an optimal second strike. Knowing the optimal third strike, the optimal second strike for each first strike can be chosen. Knowing the optimal second strike, the optimal first strike can be chosen. The optimal first, second and third strikes yield utilities for a,b,c and d.

2. Three Sides Acting in Two Coalitions

When three sides form two coalitions, there are two types of wars – one side attacks two sides and two sides attack one side. In all cases, the fourth side is attacked. Denote the sides by a,b,c and d.

In the first three wars, where the first coalition of one side attacks the second coalition of two sides, the first coalition chooses its counterforce and countervalue allocation to maximize its utility, and the second coalition chooses its countervalue allocation to maximize its utility. The first coalition of side a attacks b forces, c forces, b value, c value and d value. Then the second coalition of sides b and c attacks a value and d value – it maximizes its utility by destroying the maximum possible number of value targets of sides a and d. The optimal choice of the first coalition anticipates the optimal choice of the second coalition.

In the second three wars, where the first coalition of two sides attacks the second coalition of one side, the first coalition chooses its counterforce and countervalue allocation to maximize the minimum utility of its two individual members. The second coalition of one side chooses its countervalue allocation to maximize its utility. The first coalition of sides a and b attacks c forces, c value and d value. Then the second coalition of side c attacks a value, b value and d value. The optimal choice of the first coalition anticipates the optimal choice of the second coalition.

E. EXAMPLE

1. Data

The resources in the example are as follows:

Value targets V_1, V_2, V_3, V_4

$$= 3000, 2000, 1000, 2000$$

Offensive weapons O_1, O_2, O_3

$$= 4000, 3000, 1000$$

Defensive weapons D_1, D_2, D_3, D_4

$$= 0, 0, 0, 0$$

Side 1 is the United States, Britain and France, Side 2 is Russia and Side 3 is China. Side 4 is the rest of the world.

Value target data are consistent with the upper range of the Committee on International Security and Arms Control, National Academy of Sciences (1988) and with the medium range of the Congressional Budget Office (1991). Offensive weapon data are consistent with START II levels of the United States and Russia, present weapon levels of Britain and France, and a substantially expanded arsenal for China.

The fractions of offensive weapons of Sides 1, 2 and 3 assumed to be vulnerable are .5, .67 and .67, respectively. In other words, one-half of the side 1 weapons are invulnerable and one-third of the side 2 and side 3 weapons are invulnerable.

The attrition to vulnerable offensive weapons (weapon targets) and to value targets are calculated from the following equations:

Weapon Targets Killed = Weapon Targets X (1 − exp [− (3.0 Attacking Weapons/Weapon Targets)])

Value Targets Killed = Value Targets X (1 − exp [− (1.5 Attacking Weapons/Value Targets)])

Here exp [− Z] denotes the negative exponential operation on Z and

(1 − exp [− (Parameter X Attacking Weapons/Targets)])

is the proportion of targets killed.

For ratios of attackers to targets of .25 to 2.0, with the parameters 3.0 for vulnerable weapon targets killed and 1.5 for value targets killed, the attrition equations yield the following proportional damages:

Attackers/Targets	Vulnerable Weapon Targets Killed	Value Targets Killed
.25	.53	.31
.5	.78	.53
1.0	.95	.78
2.0	.998	.95

There is a decreasing marginal utility of assigning weapons to targets. Weapon targets are relatively more lucrative because weapon targets are assumed to involve systems with multiple warheads (ICBMs) and systems with many weapons vulnerable per target (submarines in port and bombers on bases); weapon targets killed per weapon expended is high for small attacks. This is reflected in the larger parameter used in the weapon target attrition equation.

2. Allocation Possibilities for Twelve Possible Wars

There are twelve possible wars to be examined. For each war certain allocation possibilities are examined, which are discussed in this section.

There are three different types of sequential optimization problems among the three sides. There are six cases in which the three sides act separately, three cases where one side attacks two sides and three cases where two sides attack one side. For each stage of the war the optimization process allows the attacker to choose one allocation from among a number of available possible allocations. These possible allocations are chosen to span the space of available choices. The optimization is thus on a finite grid of choices and is limited by the density of the grid.

For the six cases of three sides acting separately, the first striker has 93 choices of allocation to five target classes, the second striker has 45 choices of allocation to four target classes and the third striker has 10 choices of allocation to three target classes.

The 93 choices of the first striker are generated as follows. There are five combinations of 1.0 to one of the five target classes, ten combinations of .5 each to two of the five target classes, ten combinations of .33 each to three of the five target classes, five combinations of .25 each to four of the five target classes, and one choice of .2 each to all of the five target classes, – this totals 31 choices. There can also be additional numbers of weapons drawn off to attack Side 4 value targets. The amounts drawn off to attack Side 4 value targets can be 0.0, 0.1 or 0.2 of the total weapons. Thus there are 3 × 31 = 93 possible choices.

Following are the proportions of weapons allocated to the five target classes for the particular case of .2 each to all of the five target classes, with proportions .0, .1 and .2 of all allocations drawn off to attack Side 4 value targets (equivalently, d value):

extra to Side 4 value targets	b forces	b value	c forces	c value	d value
.0	.2	.2	.2	.2	.2
.1	.18	.18	.18	.18	.28
.2	.16	.16	.16	.16	.36

The 45 choices of the second striker are generated as follows. There are four combinations of 1.0 to one of the four target classes, six combinations of .5 each to two of the four target classes, four combinations of .33 each to three of the four target classes and one choice of .25 each to four target classes – this totals 15 choices. There can also be additional numbers of weapons drawn off to attack Side 4 value targets. The amounts drawn off to attack Side 4 value targets can be 0.0, 0.1 or 0.2 of the total weapons. Thus there are 3 × 15 = 45 possible choices.

The 10 choices of the third striker allow all combinations of 1.0 to one of the three target classes, .5 each to two of the three target classes, .33 each to the three target classes and a mix of .5 to one target class and .25 each to two target classes.

Thus from the choices of the three strikers acting separately there are 93 × 45 × 10 = 41,850 possible paths to be followed.

For the three cases of one side attacking two sides, the first striker has 180 choices of attacking five types of target classes and the second striker has 25 choices of attacking two types of target classes. Thus there are 180 × 25 = 9,500 possible paths.

For the three cases of two sides attacking one side, the first striker has 10 choices of attacking three target classes and the second striker has 10 choices of attacking three target classes, for a total of 100 possible paths.

3. Computational Implementation

The computational implementation is a modification of the model documented in Bracken (1993), which treats three armed sides and one unarmed side. That model, however, is based on the costs of Kent and Thaler (1989, 1990) and considers incentive to preempt only; incentives to strike to improve after-war position as compared to before-war position are not addressed therein. The present analysis modifies that model to define utilities and to consider both incentive to strike and first-strike stability.

4. Results

Table 1 gives results of the calculations for the twelve possible wars. Presented are value targets before the war, value targets after the war, utilities before the war, utilities after the war, incentives to strike and first-strike stability.

These results are for the optimal strategies as described above. That is, for example, for the order 1 2 3, Side 1 maximizes his utility knowing that Side 2 will maximize his utility knowing that Side 3 will maximize his utility, and Sides 1, 2, 3 and 4 have value targets after the war which equal 1044, 134, 165 and 319, respectively. All of the 41,850 paths are explored to generate the optimal behavior of Sides 1, 2 and 3. Side 4 has no choices, but simply obtains the result of the optimal behavior of the other sides.

Asterisks on incentives to strike denote the first striker in a sequence. In the first six wars, where the three sides act separately, the utilities of the first strikers are always improved – all of the incentives to strike are equal to or greater than one. In the next three wars, where one side atacks two sides, incentive for Side 1 is greater than one and incentives for Sides 2 and 3 are less than one. In the last three wars, where coalitions must form, incentives are all greater than one; the incentives for Sides 2 and 3 to combine are least.

Table 1 gives first-strike stability of each side in the three columns. To the right is the first-strike stability of all sides taken together.

Also, first-strike stability is evaluated three ways. Measure 1 is that defined previously; namely, first-strike stability S is the minimum of the first-strike stabilities of the three sides. It may involve ratios of small numbers – the utility for going other than first (in the numerator) may be very small and the utility for going first (in the denominator) may also be very small. This may result in a low first-strike stability, though preemption may not be reasonable since the side with these properties may well not be motivated to strike first.

To alleviate this problem, the next two first-strike stability measures are modifications of the first measure, defined to override ratios of small utilities for one or more sides by considering explicitly both after-war utility and before-war utility. Measure 2(.5) sets first-strike stability equal to 1.0 when the incentive to strike is equal to or less than .5; that is, if the ratio of after-war utility to before-war utility is less than one-half, a side will not preempt even if it may be relatively better off striking than waiting should war occur. Similarly, Measure 3(.9) sets stability equal to 1.0 when the incentive to strike is equal to or less than .9. Measure 2(.5) and Measure 3(.9) essentially require that, even if it is relatively better off in after-war terms if it strikes first than if it waits and some one else strikes first, no side will strike first if its utility after the war is less than .5 or .9, respectively, of its utility before the war.

First-strike stabilities for the three sides for Measure 1 are .46, .46 and .51. Sides 1 and 2 are equally-motivated and Side 3 is least-motivated to preempt. First-strike stabilities for Measure 2(.5) are .46, .46 and 1.00. Sides 1 and 2 are equally- motivated to preempt and Side 3 is stable – Side 3's lowest measure of incentive to strike of .24 places it below the cutoff for Measure 2(.5). First-strike stabilities for Measure 3(.9) are .46, 1.00 and 1.00; Side 1 is most-motivated to preempt. This illustrates how first-strike stability of three sides individually and of three sides taken together changes as the measure changes.

Overall, for all three measures, first-strike stability is .46, although the incentives of the three sides differ within the measure.

Table 1. Results for Example

Value Targets Before the War					
	Side 1	Side 2	Side 3	Side 4	
	3000.	2000.	1000.	2000.	
Value Targets After the War					
Order					
1 2 3	1044.	134.	165.	319.	
1 3 2	892.	173.	91.	141.	
2 1 3	785.	721.	223.	284.	
2 3 1	296.	446.	223.	222.	
3 1 2	892.	181.	325.	81.	
3 2 1	336.	446.	223.	235.	
1 23	1104.	446.	50.	446.	
2 13	2141.	648.	363.	506.	
3 12	3000.	2000.	160.	128.	
12 3	2760.	1765.	72.	145.	
13 2	2164.	307.	613.	419.	
23 1	1104.	799.	399.	358.	
Utilities Before the War					
	.38	.25	.13	.25	
Utilities After the War					
1 2 3	.63	.08	.10	.19	
1 3 2	.69	.13	.07	.11	
2 1 3	.39	.36	.11	.14	
2 3 1	.25	.38	.19	.19	
3 1 2	.60	.12	.22	.05	
3 2 1	.27	.36	.18	.19	
1 23	.54	.22	.02	.22	
2 13	.59	.18	.10	.14	
3 12	.57	.38	.03	.02	
12 3	.58	.37	.02	.03	
13 2	.62	.09	.17	.12	
23 1	.41	.30	.15	.13	
Incentives to Strike					
1 2 3	*1.67	.32	.80	.77	
1 3 2	*1.83	.53	.56	.44	
2 1 3	1.04	*1.43	.89	.56	
2 3 1	.67	*1.50	1.50	.75	
3 1 2	1.61	.49	*1.76	.22	
3 2 1	.72	1.44	*1.44	.76	
1 23	*1.44	.87	.19	.87	
2 13	1.56	* .71	.79	.55	
3 12	1.51	1.51	*.24	.10	
12 3	*1.55	*1.49	.12	.12	
13 2	*1.65	.35	*1.40	.48	
23 1	1.11	*1.20	*1.20	.54	
First-Strike Stability					All Sides
Measure 1	.46	.46	.51		.46
Measure 2 (.5)	.46	.46	1.00		.46
Measure 3 (.9)	.46	1.00	1.00		.46

F. VARIATIONS OF EXAMPLE

1. Defenses

Three defenses are investigated, as follows:

$D_1, D_2, D_3, D_4 = 1000, 1000, 0, 0$

$D_1, D_2, D_3, D_4 = 1000, 1000, 0, 1000$

$D_1, D_2, D_3, D_4 = 2000, 2000, 0, 0$

The first defense is 1000 perfect interceptors each for Side 1 and Side 2 and none for Side 3 and Side 4. This defense is large enough to protect Side 1 and Side 2 from strikes by Side 3 alone. Side 3 can still act in coalition with Side 1 or Side 2, and Side 3 can still attack Side 4.

The second defense is 1000 perfect defenders each for Sides 1, 2 and 4 and none for Side 3. Now, Side 4 is protected from attack by Side 3.

The third defense is 2000 perfect interceptors each for Side 1 and Side 2 and none for Side 3 and Side 4. Now, Side 1 and Side 2 can partially defend themselves against each other. This is the type of defense which induces first-strike instability. One side can attack with more weapons than can be confronted by the defense of the other, with the penetrating attackers allocated to counterforce and countervalue missions such that they kill vulnerable weapons and value targets; the relatively few surviving weapons of the second striker cannot effectively penetrate the defense of the first striker and thus few value targets of the first striker are destroyed. Whichever side strikes first does much better and thus both sides are highly motivated to preempt.

Table 2 presents results for defense 1000, 1000, 0, 0. Comparing Table 2 with Table 1 (the undefended case), incentives to strike for Side 1 and Side 2 are higher in Table 2 while incentives to strike for Side 3 are far lower. Side 3 is removed from being a unilateral problem; for the orders 3 1 2 and 3 2 1, incentives to strike drop from 1.76 and 1.44 in Table 1 to .83 and .31 in Table 2. As in Table 1 the results are for the optimal allocations for each of the twelve sequences of moves.

For Measure 1, first-strike stability of Side 1 increases from .46 to .60, first-strike stability of Side 2 decreases from .46 to .31 and first-strike stability of Side 3 decreases from .51 to .07; thus overall first-strike stability decreases from .46 to .07. But Measure 1 involves for Side 3 a ratio of very small utilities since Side 3 is severely damaged both as a first striker or as other than a first striker. For Measure 2(.5), overall first strike stability decreases from .46 to .31; Side 2 is the least stable. For Measure 3(.9), however, overall first-strike stability increases from .46 to .60, since only Side 1 has incentive to strike exceeding .9.

Table 3 gives results for defense 1000, 1000, 0, 1000. Comparing Table 3 with Table 2, incentives to strike for Sides 1, 2 and 3 are usually lower. Of particular interest is that, for Measure 2(.5), first-strike stability increases from .31 to .59, and for Measure 3(.9), first-strike stability increases from .60 to 1.00 (due to Side 1 having incentive to strike for sequence 1 23 of .85). The qualitative result is that, overall, first-strike stability is increased by defending Side 4, the unarmed side.

Table 4 gives results for defenses 2000,2000,0,0. Comparing Table 4 with Table 1 (the undefended case), incentives to strike for Side 1 and Side 2 are significantly higher. Incentive to strike for Side 3 is significantly lower. First-strike stability for Side 1 is higher and first-strike stability for Side 2 is lower. First-strike stability, overall, is lower for all measures.

Comparing Table 4 with Table 2 (defenses for Side 1 and Side 2 have been increased from 1000 to 2000), incentives to strike for Side 1 and Side 2 increase. For Side 2, incentive to strike for order 2 13 goes from .85 (worse off) to 1.30 (better off). This results in overall first-strike stability for Measure 3(.9) being significantly lower – decreasing from .60 in Table 2 to .38 in Table 4. This is the qualitative effect discussed earlier. Going from 0 to 1000 Side 1 and Side 2 defenders increases first-strike stability, but going from 1000 to 2000 Side 1 and Side 2 defenders decreases first-strike stability to below the undefended case of Table 1; this is particularly true if one believes that Measure 3 (.9) is the most plausible of the three measures.

Of note in Table 4 is that for almost all attack orders Side 1 and Side 2 have greater survival of value targets than in Table 1 and Table 2. Defenses both decrease their losses from striking each other and encourage them to strike Side 3 and Side 4.

2. Vulnerable Weapons

Total weapons of Sides 1, 2 and 3 are 4000, 3000 and 1000, respectively. The variation is to

Table 2. Results for Defense 1000,1000,0,0

Value Targets Before the War					
	Side 1	Side 2	Side 3	Side 4	
	3000.	2000.	1000.	2000.	
Value Targets After the War					
Order					
1 2 3	3000.	876.	190.	102.	
1 3 2	2336.	211.	325.	287.	
2 1 3	1820.	2000.	160.	122.	
2 3 1	1820.	2000.	160.	135.	
3 1 2	2336.	285.	325.	194.	
3 2 1	1820.	2000.	160.	135.	
1 23	3000.	1375.	223.	360.	
2 13	1104.	649.	1000.	307.	
3 12	3000.	2000.	72.	27.	
12 3	3000.	2000.	5.	0.	
13 2	1104.	649.	1000.	307.	
23 1	3000.	676.	1000.	446.	
Utilities Before the War					
	.38	.25	.13	.25	
Utilities After the War					
1 2 3	.72	.21	.05	.02	
1 3 2	.74	.07	.10	.09	
2 1 3	.44	.49	.04	.03	
2 3 1	.44	.49	.04	.03	
3 1 2	.74	.09	.10	.06	
3 2 1	.44	.49	.04	.03	
1 23	.61	.28	.05	.07	
2 13	.36	.21	.33	.10	
3 12	.59	.39	.01	.01	
12 3	.60	.40	.00	.00	
13 2	.36	.21	.33	.10	
23 1	.59	.13	.20	.09	
Incentives to Strike					
1 2 3	*1.92	.84	.36	.10	
1 3 2	*1.97	.27	.82	.36	
2 1 3	1.18	*1.95	.31	.12	
2 3 1	1.18	*1.94	.31	.13	
3 1 2	1.98	.36	*.83	.25	
3 2 1	1.18	1.94	*.31	.13	
1 23	*1.61	1.11	.36	.29	
2 13	.96	*.85	2.61	.40	
3 12	1.57	1.57	*.11	.02	
12 3	*1.60	*1.60	.01	.00	
13 2	*.96	.85	*2.61	.40	
23 1	1.56	*.53	*1.56	.35	
First-Strike Stability					All Sides
Measure 1	.60	.31	.07		.07
Measure 2 (.5)	.60	.31	1.00		.31
Measure 3 (.9)	.60	1.00	1.00		.60

Table 3. Results for Defense 1000,1000,0,1000

	Side 1	Side 2	Side 3	Side 4	All Sides
Value Targets Before the War					
	3000.	2000.	1000.	2000.	
Value Targets After the War					
Order					
1 2 3	3000.	1194.	76.	1287.	
1 3 2	3000.	1194.	76.	1287.	
2 1 3	1104.	2000.	91.	700.	
2 3 1	1104.	945.	1000.	945.	
3 1 2	3000.	1194.	76.	1287.	
3 2 1	1104.	945.	1000.	945.	
1 23	1894.	1081.	1000.	2000.	
2 13	1936.	1155.	1000.	1155.	
3 12	3000.	2000.	72.	83.	
12 3	3000.	2000.	5.	307.	
13 2	1104.	649.	1000.	649.	
23 1	669.	945.	1000.	945.	
Utilities Before the War					
	.38	.25	.13	.25	
Utilities After the War					
1 2 3	.54	.21	.01	.23	
1 3 2	.54	.21	.01	.23	
2 1 3	.28	.51	.02	.18	
2 3 1	.28	.24	.25	.24	
3 1 2	.54	.21	.01	.23	
3 2 1	.28	.24	.25	.24	
1 23	.32	.18	.17	.33	
2 13	.37	.22	.19	.22	
3 12	.58	.39	.01	.02	
12 3	.56	.38	.00	.06	
13 2	.32	.19	.29	.19	
23 1	.19	.27	.28	.27	
Incentives to Strike					
1 2 3	*1.44	.86	.11	.93	
1 3 2	*1.44	.86	.11	.93	
2 1 3	.76	*2.05	.19	.72	
2 3 1	.74	*.95	2.00	.95	
3 1 2	1.44	.86	*.11	.93	
3 2 1	.74	.95	*2.00	.95	
1 23	*.85	.72	1.34	1.34	
2 13	.98	*.88	1.53	.88	
3 12	1.55	1.55	*.11	.06	
12 3	*1.51	*1.51	.01	.23	
13 2	*.87	.76	*2.35	.76	
23 1	.50	*1.06	*2.25	1.06	
First-Strike Stability					All Sides
Measure 1	.59	.82	.07		.07
Measure 2 (.5)	.59	.82	1.00		.59
Measure 3 (.9)	1.00	1.00	1.00		1.00

Table 4. Results for Defense 2000,2000,0,0

	Side 1	Side 2	Side 3	Side 4	All Sides
Value Targets Before the War					
	3000.	2000.	1000.	2000.	
Value Targets After the War					
Order					
1 2 3	3000.	446.	105.	64.	
1 3 2	3000.	446.	105.	87.	
2 1 3	1820.	2000.	50.	15.	
2 3 1	1820.	2000.	50.	39.	
3 1 2	3000.	446.	105.	87.	
3 2 1	1820.	2000.	50.	39.	
1 23	3000.	1573.	138.	176.	
2 13	3000.	2000.	1000.	151.	
3 12	3000.	2000.	72.	27.	
12 3	3000.	2000.	5.	0.	
13 2	1820.	2000.	1000.	47.	
23 1	3000.	446.	1000.	100.	
Utilities Before the War					
	.38	.25	.13	.25	
Utilities After the War					
1 2 3	.83	.12	.03	.02	
1 3 2	.82	.12	.03	.02	
2 1 3	.47	.51	.01	.00	
2 3 1	.47	.51	.01	.01	
3 1 2	.82	.12	.03	.02	
3 2 1	.47	.51	.01	.01	
1 23	.61	.32	.03	.04	
2 13	.49	.33	.16	.02	
3 12	.59	.39	.01	.01	
12 3	.60	.40	.00	.00	
13 2	.37	.41	.21	.01	
23 1	.66	.10	.22	.02	
Incentives to Strike					
1 2 3	*2.21	.49	.23	.07	
1 3 2	*2.20	.49	.23	.10	
2 1 3	1.25	*2.06	.10	.02	
2 3 1	1.24	*2.05	.10	.04	
3 1 2	2.20	.49	*.23	.10	
3 2 1	1.24	2.05	*.10	.04	
1 23	*1.64	1.29	.23	.14	
2 13	1.30	*1.30	1.30	.10	
3 12	1.57	1.57	*.11	.02	
12 3	*1.60	*1.60	.01	.00	
13 2	*1.00	1.64	*1.64	.04	
23 1	1.76	*.39	*1.76	.09	
First-Strike Stability					All Sides
Measure 1	.76	.38	.08		.08
Measure 2 (.5)	.76	.38	1.00		.38
Measure 3 (.9)	.76	.38	1.00		.38

raise the number of weapons which are vulnerable from 2000, 2000 and 667 to 3000, 2700 and 900.

This would correspond to, for Side 1, 1000 of the 4000 weapons on submarines at sea and all land-based missiles and bombers vulnerable, for Side 2, 300 of the 3000 weapons at sea or on untargetable missiles and, for Side 3, 100 of the 1000 weapons at sea or on untargetable missiles.

The earlier case to which this comparison is made is that of Table 2, for defense 1000, 1000, 0, 0.

Comparing Table 5, where there are more vulnerable weapons, with Table 2, now Side 2 can attack Sides 1 and 3, in the order 2 13, and be better off – incentive to strike for order 2 13 increases from .85 to 1.26. This is because, in Table 2, Side 2 is not motivated to attack Side 1 weapons, leaving them to destroy much of Side 2 in response, but in Table 5 Side 2 is motivated to attack Side 1 weapons since more Side 1 weapons are vulnerable. Side 1 weapons instead attack the undefended Side 4.

First-strike stability decreases. For Measure 2(.5) it decreases from .31 to .21. For Measure 3 (.9) it decreases from .60 to .21.

G. ALTERNATE UTILITY FUNCTION

The cost function of Kent and Thaler (1989;1990) is

$$\text{Cost} = \text{Damage to Self} - \text{Weight} \times \text{Damage to Others} + \text{Weight},$$

where Weight is a parameter designed to discount damage to others and the equation defines cost between 0 and 1 + Weight. Weight is usually set to .3, resulting in valuing damage to self significantly more than damage to others and resulting in cost being defined from 0 to 1.3. If Weight is equal to 1.0, all sides are treated the same and cost is defined from 0 to 2.

The utility function presented in Section C above and used in the example does not allow a side to discount the effects on the other sides. All sides are treated the same.

The following utility function reflects all three of these criteria: (1) one's percent of all value surviving, (2) the undesirability of a war, and (3) a preference for one's own value. Here V denotes the total value held Others by the other sides.

$$U_i^1 = \frac{V_i^1}{V_1^1 + V_2^1 + V_3^1 + V_4^1} \times \frac{V_i^1 + \text{Weight} \times V_{\text{Others}}^1}{V_i^1 + \text{Weight} \times V_{\text{Others}}^1}$$

$$U_i^2 = \frac{V_i^2}{V_1^2 + V_2^2 + V_3^2 + V_4^2} \times \frac{V_i^2 + \text{Weight} \times V_{\text{Others}}^2}{V_i^1 + \text{Weight} \times V_{\text{Others}}^1}$$

That is, the utility before the war is the percentage of the value held by Side i, while the utility after the war is the percent of value held by Side i times the percent of the original value left, with Side i setting a weight on how much he cares about the others with respect to original value.

Tables 6 and 7 give results for Weight = 1 (Side i values everyone the same) and for Weight = .2 (Side i values others .2 times as much as himself). Comparing both Table 6 and Table 7 with Table 1, for the original utility function, there is now less incentive to strike. This is because all sides have an incentive to preserve the world. With respect to the effect of Weight alone, the qualitative effect of going from Table 6 to Table 7 is to decrease most incentives to engage in warfare and to increase first-strike stability. These sensitivities are because when Weight = .2 all sides are more insensitive to the destruction of others and thus perform more countervalue attacks. Any initiator thus faces more damage to his own value.

H. SUMMARY OF INSIGHTS

This theory of multipolar nuclear stability attempts to encompass the minimum number of entities which affect the problem.

The resources are: (1) value targets of Sides 1,2,3,4, (2) offensive weapons of Sides 1,2,3 and (3) defenses of Sides 1,2,3,4. Some of the offensive weapon resources may be vulnerable. There are two attrition functions, counterforce and countervalue.

The behavioral motivations are captured in the utility function – incentive to strike and first-strike stability.

Table 5. Results for Defense 1000,1000,0,0 and Increased Vulnerability of Offensive Weapons

Value Targets Before the War					
	Side 1	Side 2	Side 3	Side 4	
	3000.	2000.	1000.	2000.	
Value Targets After the War					
Order					
1 2 3	3000.	876.	268.	129.	
1 3 2	2336.	211.	325.	287.	
2 1 3	1961.	2000.	180.	35.	
2 3 1	1961.	2000.	180.	40.	
3 1 2	2336.	285.	325.	194.	
3 2 1	1961.	2000.	180.	32.	
1 23	3000.	1453.	259.	294.	
2 13	3000.	2000.	1000.	329.	
3 12	3000.	2000.	72.	27.	
12 3	3000.	2000.	5.	0.	
13 2	1104.	649.	1000.	307.	
23 1	2336.	1470.	347.	42.	
Utilities Before the War					
	.38	.25	.13	.25	
Utilities After the War					
1 2 3	.70	.21	.06	.03	
1 3 2	.74	.07	.10	.09	
2 1 3	.47	.48	.04	.01	
2 3 1	.47	.48	.04	.01	
3 1 2	.74	.09	.10	.06	
3 2 1	.47	.48	.04	.01	
1 23	.60	.29	.05	.06	
2 13	.47	.32	.16	.05	
3 12	.59	.39	.01	.01	
12 3	.60	.40	.00	.00	
13 2	.36	.21	.33	.10	
23 1	.56	.35	.08	.01	
Incentives to Strike					
1 2 3	*1.87	.82	.50	.12	
1 3 2	*1.97	.27	.82	.36	
2 1 3	1.25	*1.92	.35	.03	
2 3 1	1.25	*1.91	.35	.04	
3 1 2	1.98	.36	*.83	.25	
3 2 1	1.25	1.92	*.35	.03	
1 23	*1.60	1.16	.41	.23	
2 13	1.26	*1.26	1.26	.21	
3 12	1.57	1.57	*.11	.02	
12 3	*1.60	*1.60	.01	.00	
13 2	*.96	.85	*2.61	.40	
23 1	1.49	*1.40	*.66	.04	
First-Strike Stability					All Sides
Measure 1	.78	.21	.07		.07
Measure 2 (.5)	.78	.21	1.00		.21
Measure 3 (.9)	.78	.21	1.00		.21

Table 6. Results for Alternate Utility Function with Weight = 1.0

Value Targets Before the War					
	Side 1	Side 2	Side 3	Side 4	
	3000.	2000.	1000.	2000.	
Value Targets After the War					
Order					
1 2 3	3000.	2000.	50.	146.	
1 3 2	3000.	446.	1000.	3.	
2 1 3	3000.	2000.	67.	0.	
2 3 1	1820.	2000.	105.	0.	
3 1 2	3000.	446.	1000.	3.	
3 2 1	958.	1601.	1000.	26.	
1 23	2399.	945.	223.	732.	
2 13	1527.	2000.	132.	15.	
3 12	3000.	1601.	130.	85.	
12 3	2540.	2000.	1000.	2000.	
13 2	1777.	307.	1000.	2000.	
23 1	1104.	2000.	1000.	19.	
Utilities Before the War					
	.38	.25	.13	.25	
Utilities After the War					
1 2 3	.38	.25	.01	.02	
1 3 2	.38	.06	.13	.00	
2 1 3	.38	.25	.01	.00	
2 3 1	.23	.25	.01	.00	
3 1 2	.38	.06	.13	.00	
3 2 1	.12	.20	.13	.00	
1 23	.30	.12	.03	.08	
2 13	.19	.25	.02	.00	
3 12	.38	.20	.02	.01	
12 3	.32	.25	.13	.22	
13 2	.22	.04	.13	.22	
23 1	.14	.25	.13	.00	
Incentives to Strike					
1 2 3	*1.00	1.00	.05	.06	
1 3 2	*1.00	.22	1.00	.00	
2 1 3	1.00	*1.00	.07	.00	
2 3 1	.61	*1.00	.11	.00	
3 1 2	1.00	.22	*1.00	.00	
3 2 1	.32	.80	*1.00	.01	
1 23	*.80	.47	.22	.33	
2 13	.51	*1.00	.13	.01	
3 12	1.00	.80	*.13	.04	
12 3	*.85	*1.00	1.00	.89	
13 2	*.59	.15	*1.00	.89	
23 1	.37	*1.00	*1.00	.01	
First-Strike Stability					All Sides
Measure 1	.40	.15	.38		.15
Measure 2 (.5)	.40	.15	1.00		.15
Measure 3 (.9)	1.00	.15	1.00		.15

Table 7. Results for Alternate Utility Function with Weight = .2

	Side 1	Side 2	Side 3	Side 4	All Sides
Value Targets Before the War					
	3000.	2000.	1000.	2000.	
Value Targets After the War					
Order					
1 2 3	2538.	743.	105.	36.	
1 3 2	2089.	446.	581.	132.	
2 1 3	1392.	1735.	408.	469.	
2 3 1	1352.	930.	465.	222.	
3 1 2	1294.	602.	642.	511.	
3 2 1	1237.	930.	465.	426.	
1 23	1919.	945.	223.	1023.	
2 13	3000.	770.	1000.	770.	
3 12	3000.	2000.	160.	128.	
12 3	2760.	1765.	72.	145.	
13 2	2309.	783.	675.	474.	
23 1	1820.	910.	455.	411.	
Utilities Before the War					
	.38	.25	.13	.25	
Utilities After the War					
1 2 3	.50	.09	.01	.00	
1 3 2	.37	.04	.08	.01	
2 1 3	.17	.30	.05	.04	
2 3 1	.19	.13	.06	.02	
3 1 2	.17	.07	.10	.05	
3 2 1	.16	.13	.06	.04	
1 23	.28	.11	.02	.12	
2 13	.47	.07	.14	.07	
3 12	.49	.31	.01	.01	
12 3	.46	.27	.01	.01	
13 2	.37	.09	.09	.04	
23 1	.28	.11	.06	.04	
Incentives to Strike					
1 2 3	*1.34	.35	.08	.01	
1 3 2	*1.00	.17	.66	.04	
2 1 3	.44	*1.19	.38	.16	
2 3 1	.51	*.52	.50	.08	
3 1 2	.47	.27	*.79	.21	
3 2 1	.43	.52	*.50	.18	
1 23	*.73	.45	.18	.48	
2 13	1.27	*.30	1.15	.28	
3 12	1.31	1.26	*.12	.03	
12 3	*1.22	*1.10	.05	.04	
13 2	*.98	.34	*.74	.16	
23 1	.73	*.46	*.46	.14	
First-Strike Stability					All Sides
Measure 1	.59	.58	.43		.43
Measure 2 (.5)	.59	1.00	1.00		.59
Measure 3 (.9)	1.00	1.00	1.00		1.00

The three sides act in any of twelve ways – six orders in which they may strike separately and six orders in which they may act in two coalitions.

There are thus sixteen dimensions for resources and attriton: the eleven resource levels, the three weapon vulnerability parameters and the two attrition equations. They are integrated by the utility measure and the incentive measures in examining all of the twelve wars. The motivations to strike and to preempt are the focus of the analysis.

Several strong conclusions can be drawn:

1. Results are sensitive to all of the sixteen resource and attrition dimensions. Some of the sensitivities are presented in the paper.

2. Stability increases and decreases can be explained as the sixteen dimensions are changed.

3. An interesting finding with respect to defenses is that if each of the large nuclear powers has enough defense to block the offensive weapons of the third, smaller, nuclear power, but not enough defenses to block a large fraction of the offensive weapons of the other large nuclear power, first-strike stability increases. Adding defenses to the unarmed side also increases stability.

Specific results of the analysis of the example are obtained for four levels of defense. First-strike stability Measure 3(.9), where only those sides with a ratio of utility after the war to utility before the war larger than .9 will initiate, yields the following:

Defense of Sides 1, 2, 3, 4				First-Strike Stability
0,	0,	0,	0	.46
1000,	1000,	0,	0	.60
1000,	1000,	0,	1000	1.00
2000,	2000,	0,	0	.38

These numerical results correspond to the conclusions discussed above. Stability increases with smaller defenses but decreases with larger defenses.

I. FURTHER COMMENTS

Two topics are addressed in this section. The first topic is a discussion of the logic of why small defenses are stabilizing in a world of three nuclear-armed powers when one of the powers is small. The second topic is a discussion of why the proposed model is needed to explore sensitivities to key parameters.

1. Discussion of Insight That Small Defenses Are Stabilizing

The following argument can serve to clarify the basic logic of the more detailed deductive model and computational results presented above.

The basic bipolar argument of Schelling (1960), built on by others, is that deterrence stability requires that both sides have a secure second-strike capability to conduct a punishing attack upon the other.

Assume that there are two superpowers and a smaller third power without defenses. If the defenses of both superpowers are small relative to each other's nuclear arsenals, even after a preemptive attack by one superpower the other can still deliver a punishing second strike upon his attacker. Therefore with these small defenses the superpower balance would remain.

The stability of the nuclear balance is more problemmatic, however, with the existence of a third, but smaller, nuclear power. This power may be tempted, in a crisis, to launch its weapons rather than be denied a second strike capability. Since the two superpowers are large relative to the third, the third power could not hope to deliver a preemptive first strike with the hope of avoiding a punishing attack, but, by striking first, the third power could be expected to deliver some punishment to a threatening superpower and might be tempted to do so if convinced the superpower was likely to attack first.

However, if both superpowers have sufficient defenses to deny the third power's preemptive attack, the third power has no incentive to strike first. This leads to the conclusion that defenses of two large sides which are large relative to a third side but small relative to each other are stabilizing.

The above argument seems to be quite straightforward, and it can be argued that the complexity of the present paper is not necessary to reach the conclusion. This is true, but the insight did not emerge prior to the model development. Rather, it resulted from the research and is now quite clear. It is therefore reported as a conclusion of the research.

2. Discussion of Model Scope

Results of three-stage optimized wars with three armed sides and one unarmed side can be quite sensitive to all of the parameters comprising the model – value targets, offensive weapons, percent of offensive weapons which are vulnerable, defensive weapons, attrition equations for counterforce effectiveness against offensive weapons which are vulnerable and attrition equations for countervalue effectiveness against value targets.

Sensitivity analyses can identify the directions of effects of changes in the parameters, but it is difficult to estimate the amounts of the effects. For instance, increasing the parameter for counterforce effectiveness of one side can lead to large changes in weapon allocations and in the final status of surviving value targets on all sides.

Another very interesting example is the behavior of the entire model as a function of the defensive weapon levels. Since the attacker must exhaust all of the defenders of a side before killing any vulnerable offensive weapons or value targets, there is a fixed price for attacking any defended side. If there are one, two or three sides with defending weapons, the attacker must choose which prices to pay – how many weapons to allocate – and how to split the allocations between counterforce and countervalue.

The multistage and discontinuous nature of the decision process seems to be virtually impossible to sort out in the absence of actual computation. Even though the model is as parsimonious as possible, its behavior as a function of the input parameters cannot be predicted in advance. The qualitative argument in Section I.1 above is important, and gives insight into the multipolar world. However, the model is necessary for exploring the effects of specific force levels of the resources of the three armed sides and one unarmed side.

J. CAVEAT

The quantitative results given in the tables might be changed somewhat if a more detailed grid of permissible allocations were considered. It would be desirable to verify with more detailed calculations these approximate computational results.

K. ACKNOWLEDGMENTS

I am indebted to the following colleagues for various contributions to my understanding of this general class of problems and the application of game-theoretic concepts thereto: Marvin Alme, Melvin Best, Glenn Kent, Thomas Quint and Martin Shubik.

I am further indebted to the referees for helpful comments. In particular, one referee suggested the logical argument clarifying the structure of the analysis presented in Section I.1 and another referee suggested a number of detailed changes which have improved the paper.

None of the above are responsible for any errors in the present paper.

L. REFERENCES

Best, Melvin and Jerome Bracken (1993), "First-Strike Stability in a Multipolar World", in Reiner K. Huber and Rudlof Avenhaus (editors), International Stability in a Multipolar World: Issues and Models for Analysis, Baden-Baden: Nomos Verlagsgesellschaft.

Best, Melvin and Jerome Bracken (1994), "First-Strike Stability in a Multipolar World: Three Independent Sides", in A. Zicihchi (editor), International Seminar on Nuclear War and Planetary Emergencies: 18th Session, Erice, Italy, Singapore: World Publishing Company.

Best, Melvin and Jerome Bracken (1995), "First-Strike Stability in a Multipolar World", Management Science, Volume 41, pages 298-321.

Bracken, Jerome (1993), "First Strike Stability in a Multipolar World: Three Independent Sides – Computer Program Description", report to U.S. Arms Control and Disarmament Agency.

Bracken, Jerome and Martin Shubik (1993), "Worldwide Nuclear Coalition Games: A Valuation of Strategic Offensive and Defensive Forces", Operations Research, Volume 41, pages 655-668.

Committee on International Security and

Arms Control, National Academy of Sciences (1988), Reyjavik and Beyond: Deep Reductions in Strategic Nuclear Arsenals and the Future Direction of Arms Control, Washington, DC: National Academy Press.

Congressional Budget Office (1991), The START Treaty and Beyond, Washington, DC: U.S. Government Printing Office.

Kent, Glenn A. and David E. Thaler (1989), "First-Strike Stability: A Methodology for Evaluating Strategic Forces", R-3765-AF, The RAND Corporation.

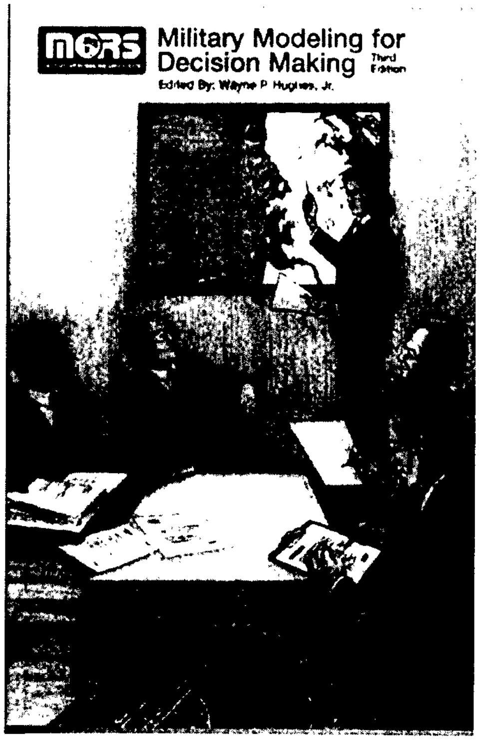
Kent, Glenn A. and David E. Thaler (1990), "First-Strike Stability and Strategic Defense: Part II of A Methodology for Evaluating Strategic Forces," R-3918-AF, The RAND Corporation.

Schelling, Thomas C. (1960), The Strategy of Conflict, Cambridge, Massachusetts: Harvard University Press.

Military Modeling for Decision Making, Third Edition

Edited By Wayne P. Hughes, Jr., Captain USN, Retired

There comes a time in every profession when it is appropriate to record the core wisdom generated by and shared among the practitioners and theorists of their craft. After 40 years of military operations research being practiced, MORS felt that this time had come and, in 1989, published the first edition of *Military Modeling*. After selling out of the first and second print runs, MORS decided that, to stay abreast with the computer generation and of the modern modeling tools, a completely revised third edition was in order!



This monograph is the latest step in MORS' dedication to furthering the technical base and professionalism of our membership and the MOR community at large. The purpose of MORS is "to enhance the quality and effectiveness of classified and unclassified military operations research." A key medium for accomplishing that purpose is professional publications such as *Military Modeling for Decision Making*. Through these publications we are enhancing the knowledge of our current members, and more importantly, providing a collated reference for the young analysts yet to join our ranks. It is frankly the latter which motivated the MORS volunteers and staff who labored to produce this monograph. Professor Wayne Hughes, father of this monograph in all its editions, and his group of authors willingly devoted a significant quantity of their scarce time and profound professional knowledge and experience to author and/or revise chapters to reflect the latest technologies, techniques, and applications. We are all quite proud of the result, and trust that it will be used for many years to come by analysts as they employ *Military Modeling for Decision Making*.

Please Send Me _____ copies of Military Modeling for Decision Making at $40 per copy plus $5 for freight.

Name ______________________________

Address ______________________________

____

____

City ______________ State ________ Zip Code ________

VA Residents add $1.80 sales tax per copy

❐ CASH ❐ CHECK ❐ Visa ❐ MC ❐ AMEX **Credit card #** ______________ **Expiration date:** ________ **Billing zip code:** ________ **Name on Credit Card (print):** ______________ Phone Number: ______________ **Signature:** ______________	TOTAL $

101 South Whiting Street • Suite 200 • Alexandria, VA 22304 • (703) 751-7290 • Fax (703) 751-8171
E-mail morsoffice@aol.com

Gain-Sharing, Success-Sharing and Cost-Based Transfer Pricing: A New Budgeting Approach For The Department of Defense (DoD)*

Francois Melese
Associate Professor of Economics

Defense Resources Management Institute
Naval Postgraduate School
Monterey, CA 93943
Phone: (408)656-2009
Fax: (408)656-2139
Email: fmelese@nps.navy.mil

APPLICATION AREA:
Resource Analysis, Social Science Methods, Logistics

OR METHODOLOGY:
Dynamic Programming, Cost Analysis

GAIN-SHARING, SUCCESS-SHARING AND COST-BASED PRICING

"In principle... a system ought to encourage individuals to do what is right by rewarding them for carrying out... desirable policies"
(Weitzman [1976] p.251)

INTRODUCTION

The objective of this paper is to model a new incentive structure for government activities that rewards cost-savings and efficiencies. The model combines "cost-based pricing" (see Pavia [1995]) with the popular business practice of "gain-sharing" (see Welbourne & Mejia [1995]), and a new incentive program we call "success-sharing".

The first section of the paper reviews these three concepts. The next section identifies incentive problems that result from traditional public budgeting practices and offers cost-based "transfer pricing" as an alternative. The third section explores ongoing efforts to implement a transfer pricing system in the Department of Defense (DoD). The paper concludes by offering a new budgeting approach that integrates cost-based transfer pricing with gain-sharing and success-sharing.

Under "fully-distributed" cost-based pricing, firms allocate all costs to their various outputs and then use those costs to set prices. (Pavia [1995] p.1060) In a study of over 500 Fortune 1000 firms, 83 percent reported using fully-distributed costs to establish prices. (Govindarajan & Anthony [1983]) Government mandates also require many regulated firms, such as public utilities, to use cost-based pricing. (see Vogelsang& Finsinger [1979], Laffont & Tirole [1986], and Sappington & Sibley [1988])

Another important application of cost-based pricing is in the construction of "transfer prices." (e.g. see Benke & Edwards [1980] or Magee [1986]) Many large, complex, vertically-integrated organizations include separate activities that conduct internal exchanges of goods and services. The challenge faced by these organizations is to govern the relationships among internal activities to promote the goals of the organization as a whole. Many private firms solve this problem through the use of internal transfer prices.

Two important lessons come out of the transfer pricing literature. (e.g. see Kovac & Troy [1989], Eccles [1985], Bruns & Kaplan [1987], and Rogerson [1995]) First, to promote the goals of the organization, transfer prices must correctly reflect costs. Second, internal activities must somehow be rewarded for using transfer price signals to pursue organizational goals. These two lessons hint at a model that combines a cost-based transfer pricing system with a program of organizational incentives.

This paper applies lessons from the transfer pricing literature to a unique subset of government activities—those that "earn" their budgets. Internal DoD support activities financed through the Defense Business Operations Fund (DBOF) offer an illustration. Charging their "customers" (operating forces) regulated cost-based transfer[1] prices, DBOF activities "... sell goods or services to customers with the intent of recovering the total cost of providing those goods and services." (DBOF Handbook [1995] p. 2-1)

Unfortunately, it is well documented that, by itself, "... cost-based pricing may introduce economic inefficiencies such as the failure... to control costs of production and a lack of incentives... to invest in cost-reducing innovations." (Pavia [1995] p.1061)[2] This observation is troublesome in the case of DBOF. Designed to "foster a business-like customer/provider approach" the aim of DBOF was to "improve the delivery of support services to the Department's operating forces while reducing the cost of operations." (DBOF Handbook[1995] p.1-1)

Along with DBOF, four approaches have been discussed to help reduce the cost of support in DoD: i.) "outsourcing" and

*I would like to thank Jim Blandin, Schroeder Dodds, Paul Hough, Jim Howard, C.J. LaCivita, Philippe Michel, Andy Rumbaugh, Neil Seiden, Mike Stroup and two reviewers of this journal for helpful comments and suggestions. The views expressed are my own. They do not necessarily reflect the policy of the U.S. Navy, the Office of the Under Secretary of Defense (Comptroller), or the Department of Defense.

(*)The Defense Business Operating Fund (DBOF) that I refer to in this paper has recently been replaced by what is called the Defense Working Capital Funds. The name has been changed to protect the innocent. For the purposes of my paper, all of the operating concepts of revolving funds are still in effect. The break-up of DBOF into five separate working capital funds re-emphasizes the management oversight responsibility each of the Services has for their activities. The Office of the Under Secretary of Defense Comptroller (OUSDC) still develops policy and has oversight responsibilities for review of the Services Working Capital Fund budgets. The day-to-day management of the Defense Working Capital Funds facilities remains firmly in the hands of the Services.

"privatization",[3] ii.) competition, iii.) deregulation, and iv.) organizational incentives (see the SECDEF's Annual Report [1996], and Directions for Defense (DD) [1995]). The first two approaches are outward looking. They focus on the opportunity for current DoD support to be provided by private firms. The last two approaches are inward looking. They focus on improving conditions under which work is currently done within DoD.[4]

This paper is inward looking. It focuses on internal DoD budgeting and management issues, and the role of organizational incentives. The dual objective of the paper is: first, to offer a framework to help upper-level management evaluate organizational incentives in cost-based pricing systems; and second, to persuade the Operations Research Community to join in this effort to model critical budgeting issues.

Citing the need to "reduce the cost of support to help fund higher priority needs," the Congressionally mandated *Report of the Commission on Roles and Missions of the Armed Forces* (DD [1995]) recommends that the DoD "reduce the cost of... support... through increased outsourcing and better management." (p.ES-3) The Department is encouraged to identify its "core competencies" or "inherently governmental functions" and outsource or privatize other responsibilities. (p.ES-6) The Report concludes that remaining functions "can benefit... by **pursuing business practices used in the private sector**" (DDp.3-1), and urges the adoption of "**[b]etter organizational incentives**... " (DD[1995] p.4-16).

The Secretary of Defense generally concurs,[5] emphasizing that (together with the Military Departments) the DoD Comptroller and various Under Secretaries have been "... examining new approaches to **create incentives for achieving greater savings and efficiencies**... " (SECDEF WJ Perry, DoD News Release No. 470-95, 8/25/95 p.10) This paper combines a model of cost-based pricing with two incentive programs designed to reward cost-savings: gain-sharing and success-sharing.

The goal of gain-sharing is to encourage activities to make an effort to reduce current costs in return for some of the immediate gains from the cost savings. In a cost-based pricing system, cost-savings occur whenever an activity succeeds in driving its actual total costs below its "earned" budget. Under gain-sharing a portion of the cost-savings is rebated to the activity as a reward (in the form of employee bonuses, extra vacation time, etc.). However, a recent survey article by Welbourne & Mejia [1995] expresses concern that, while numerous case studies suggest gain-sharing can help bring costs down over time,[6] "[t]he bulk of the extensive gainsharing literature is atheoretical and exploratory in nature." (p.584)

The dynamic, discrete-time optimization model introduced in this paper helps to address this concern. Moreover, this simple deterministic model leads to an important discovery. Gain-sharing alone may not be enough. The dynamics of the model suggest that while gain-sharing is necessary to encourage cost-savings, it may not be sufficient. Under certain conditions, what we call "success-sharing" is required to complement gain-sharing in order for an activity to have a sufficient incentive to invest in cost-reducing innovations.

Success-sharing offers an additional opportunity to reward activities. Suppose a "permanent" stream of cost-savings is obtained from a cost-reducing innovation. Success-sharing takes place when the success of the cost savings generated over time is shared with the activity responsible for those savings. However, it is more common for the success of cost savings not to be shared with the activity responsible for the savings. Instead, future cost savings are typically passed along to customers in the form of reduced prices in future periods.

For example, the stated policy for setting transfer prices charged by DBOF providers is that: "If costs are reduced, then prices will be reduced the following year to pass along the savings to the customer." (Business Management Directorate [1993]) Thus "[r]educed... costs translate to reduced prices... " (Isosaari [1996]p.17) However, providers "have to be confident that prices will stay high enough to recoup [investments in cost-reducing innovations]... [since] if once investments are made, regulators slash prices, [those providers] may be leery of investing again." ("The Regulatory Experiment," The Economist, Jan. 28, 1995 p.64)

In the case of DBOF, investment costs for both capital assets and management improvements are recovered through depreciation expenses (or "capital surcharges") factored into future year rates. In general, under gain-sharing, activities face a difficult trade-off. While cost-reduction efforts increase current-year pay-outs, resulting price cuts tend to squeeze potential gains from future cost savings. Unless future revenues are discounted at an unreasonably high rate, gain-sharing alone may not be sufficient to encourage investments in cost-reducing innovations. In this case success-sharing has an important role to play.

Success-sharing refers to the degree to which any "permanent" savings are shared between customers and the activity responsible for the cost savings. By definition, there is no success-sharing if the activity responsible for the cost savings does not share in the permanent stream of savings. In this case, any long-run benefits are enjoyed entirely by customers who pay new lower prices that fully reflect the cost savings.

Multiyear defense contracts offer an illustration. These contracts are often written in such a way that they effectively guarantee defense firms 100% gain-sharing, but no success-sharing. According to Rogerson [1994], the "DoD essentially makes the following bargain with the firm. In return for revealing its ability to lower costs, DoD will let the firm keep the benefits [the difference between the negotiated price and actual costs] for the duration of the multiyear in which costs are lowered."(p.78) These "benefits" correspond to 100% gain-sharing over the period of the contract. However, "on subsequent contracts, DoD will take the benefits itself."(p.78) Thus, "when a defense firm discovers a way to lower its production costs, previously negotiated prices are not changed. Firms are able to keep profits created by such cost reductions until negotiations take this new efficiency into account and lower prices on future contracts."(p.72) Any profits earned through cost reductions can be thought of as pay-outs from a 100% gain-sharing program. However, since future contracts translate the firm's cost reductions directly into lower prices, there is no success-sharing.

Success-sharing only takes place if, in subsequent periods, the regulated (or "negotiated") price is "stabilized" or only partially ratcheted down. In this case, while customers (e.g. the DoD) might still enjoy the benefit of new lower prices, those prices do not fully reflect the cost-savings achieved. Thus, success-sharing offers an extra incentive to suppliers to lower costs, since they are allowed to capture part of the future stream of benefits from their cost-reducing investments.[7]

The next section briefly introduces management implications of traditional public budgeting practices, and offers cost-based transfer pricing as an alternative. Although internal DoD support activities are used as an illustration, the results of the paper are not unique to Defense. The model applies to any regulated activity subject to unit-cost-based (transfer) pricing. The third section explores ongoing efforts to implement transfer pricing in DoD. The fourth and fifth sections model a new budgeting approach that combines cost-based transfer pricing with gain-sharing and success-sharing. The model's notation, assumptions, and stationary equilibrium are presented and discussed. An interpretation of the results, some policy guidance, and directions for future research appear in the concluding section.

TRADITIONAL BUDGETING VS COST-BASED TRANSFER PRICING

In traditional public budgeting systems, government activities submit budget requests each fiscal year to cover the costs of their operations. For example, internal DoD "providers" (support activities) that operate under direct appropriations first submit a budget to Congress, and then receive appropriated funds to generate support for their "customers" (operating forces).

Unfortunately, the traditional budgeting approach has several drawbacks. First, from the point of view of providers, an old adage concerning the appropriate size of budgets is revealing: "some is good, more is better, and... too much is just right." In principle, budget requests are formulated by providers to cover the expected costs of satisfying their customer's requirements. In practice, there is little incentive for support activities to request smaller budgets—even in periods of "down-sizing." Thus, according to Niskanen [1971], Fox [1988], Rogerson [1994], and others, under traditional budgeting there is a built-in bias to "maximize budgets" rather than to seek cost-savings. A parallel issue is that customers (operating forces)—and not providers—are usually in a better position to determine the level of support they require. Thus, under traditional budgeting a real concern in DoD is that "[m]erely reducing our military—in and of itself—will not address the problem of controlling support costs which, if left unchecked will steal scarce resources from the operating forces." (Maroni [1993]p.2)

Another drawback to traditional budgeting is that, while operating forces are in the best position to determine their own support requirements, they remain largely insulated from cost concerns. In contrast, support activities that receive direct appropriations are forced to tackle cost issues every time they submit a budget. Thus providers tend to be more aware of the cost implications of different levels (and composition) of support than are the customers that receive that support. In fact, under this system, customers sometimes view support as "free."

Unless budgets are unlimited, resource management decisions at all levels, including the operations level, need to take resource costs into account in order to maximize effectiveness (i.e. readiness, deterrence, or combat capability). However, under traditional budgeting "... operating forces [have] neither the responsibility nor the flexibility to make trade-off decisions in determining the optimal amount of support required to sustain readiness." (Maroni[1993]p.3)

Finally, traditional budgeting tends to punish cost savings, manufacture inefficiencies, and often contributes to an explosion of rules and regulations. A noted authority on public budgeting observes that "if departments save money, they run the risk the government will recapture the savings,... [and thus] efficient departments may be penalized while inefficient ones are rewarded." (Schick [1988] p.531) Worse yet, cost savings achieved one year make it harder to secure requested budgets in subsequent years: "A bureaucrat who failed to spend [or obligate] his entire budget would be in danger of having his budget cut the next year." (Stiglitz [1986]p.173) Thus traditional budgeting tends to manufacture inefficiencies by leading to familiar "use-it-or-lose-it" year-end spending sprees.[8]

The typical bureaucratic response to combat year-end spending sprees is to impose a detailed set of guidelines that constrains activities' spending, and the timing of that spending, through the fiscal year. This leads to new, constraining regulations, and to costly monitoring and auditing of activities.[9] One particularly insidious consequence is that government managers increasingly view their role as insuring strict compliance with regulations and "protecting" programs, not in cutting costs or increasing efficiencies. Thus adding layers of regulations further handcuffs management without addressing the underlying perverse incentives to avoid cost-savings.[10] Acknowledging these problems, the principal deputy Comptroller of DoD, Alice Maroni [1993], emphasizes that managers: "... need to move from a mindset focused on how fast can appropriated funds be obligated and spent, to how much can the cost of certain goods and services be reduced." (p.2)

On the basis of these observations, a good alternative to traditional budgeting would: a) satisfy customer demands, b) increase cost visibility, c) reduce the burden of excessive rules and regulations, and d) reward cost-savings and efficiencies. In DoD, this would have the dual impact of lowering support costs while improving the timeliness and quality of support provided to operating forces.

A cost-based transfer pricing system offers one alternative to traditional budgeting. A permanent concern of top management of large firms and organizations is to insure that users of internally supplied intermediate (or "support") products make efficient use of those products in producing final outputs. Another concern is to insure that internal providers supply those products as efficiently as possible. A standard solution adopted by commercial firms is to use internal transfer prices. (Magee [1986])

In a transfer pricing system, "customers" of internally supplied intermediate products purchase those products from internal "providers." The cost-based transfer prices charged for these intermediate products are designed to encourage customers to make efficient decisions by making them aware of the cost of producing those products. Meanwhile, combining cost-based transfer pricing with organizational incentives (gain-sharing and success-sharing), and/or with the threat of competition (from internal sources, or from outsourcing), can help drive internal providers to make more efficient production and investment decisions.[11]

The general consensus is that for transfer prices to provide the most efficient resource allocation signals, they should be based on marginal costs—the additional costs of producing the last unit of an intermediate product. (e.g. see Rogerson [1995]) In practice, however, transfer pricing systems are often based on unit costs—calculated by dividing the total costs of producing an intermediate product over some period, by the number of units produced that period.

Three accounting characteristics help explain the popularity of unit-cost-based transfer pricing systems: First, traditional accounting systems were not designed to collect marginal cost data, and in any case, unit cost calculations are less data-intensive. Second, unlike economist's who assume a "U-shaped" (or quadratic) average total cost curve (where unit costs are high at low output levels, fall to a minimum as output expands, and then increase again as output expands further), an (often implicit) accounting convention is to assume that unit costs are independent of output. If unit costs are constant over a wide range of production, then this implies marginal costs equal unit costs over that range. Finally, in the case of multi-product firms, overhead (G&A) costs are difficult to assign to specific units of a single product. Instead, overhead is typically spread over the various products according to some rule, and then averaged into the price/cost charged for specific units of a product. (e.g. see Rogerson [1995])

A first simplifying assumption of the model is that transfer prices are set on the basis of unit costs. A second simplifying assumption is that all relevant costs can be directly assigned to individual outputs.[12] The next section reviews the transfer pricing system currently implemented in DoD, and develops some additional assumptions for the new budgeting model that follows.

UNIT-COST-BASED TRANSFER PRICING

The most widespread use of cost-based transfer pricing in a government setting is found in DoD's Defense Business Operations Fund (DBOF). The DBOF was established in October 1991 by the Secretary of Defense.[13] It consists of all supply and logistics organizations ("providers") within DoD that sell their outputs to other organizations ("customers") within DoD. Over $70 billion per year of support (almost one third of the defense budget) is funded under DBOF, and over 300,000 civilian and military personnel are employed in DBOF activities. This section focuses on some key concepts used in budgeting support activities under DBOF and contrasts current DBOF policies with new policy proposals offered in the model.

In order to be included in the DBOF financial structure as a "business area," support functions must meet four criteria: i.) outputs can be identified; ii.) an approved accounting system is available; iii.) customers can be identified; and iv.) benefit/costs of establishing a buyer-seller relationship can be evaluated. (DBOF Handbook [1995]) Business areas include logistics activities that distribute, maintain and replace materiel to give combat units the equipment and support services they need, when they need them. Examples of DBOF business areas include: supply management, including the purchase, maintenance, storage, repair, and transportation of supplies and equipment; financial and accounting services; publications services; commissaries; information services; and some research and development.[14]

The purpose of DBOF was to "more closely relate the support infrastructure with the force structure" and to "improve the delivery of support services... while reducing the cost." (DBOF Handbook [1995]p.1-1) Four specific objectives are cited in the DBOF Handbook: (1) to "identify the full cost of support;" (2) to "measure performance on the basis of cost/output (i.e. unit cost) goals;" (3) to "reduce DoD support costs through better business practices;" and (4) to "foster efficiency and productivity improvements." (p.1-3) Each objective is briefly discussed below.

(1) "Identify the full cost of support": The ultimate goal is to reveal all labor, materials and capital costs that contribute to each output of a given support activity at as disaggregated a level as possible. This increased cost visibility is designed to facilitate the cost accounting required to derive cost-based transfer prices "charged" to customers of DBOF activities.

The full cost of support includes civilian labor, military labor, material, and other direct costs, depreciation expenses, property maintenance, and "acceleration of labor" (i.e. the cost of fringe benefits). In a multi-product organization, the development of activity-based costing (ABC) can help to identify the full cost of support.[15] (Brimson [1991]) For ease of exposition, the model focuses on support functions that produce at least one measurable output to which all relevant costs can be assigned.

(2) "Measure performance on the basis of cost/output (i.e. unit cost) goals": Dividing a support activity's total costs in one period by the output produced that period yields an average total cost measure. Analogous to the accountant's constant unit cost assumption, under limited

(or asymmetric) information, this average total cost measure can be used as an estimate of the unit cost of support in a subsequent period.

"Organizations financed through DBOF sell goods or services to 'customers' with the intent of recovering the total cost of providing those goods and services." (DBOF Handbook [1995]p.2-1) This can be accomplished in two ways: First, through direct budget authority based on unit cost targets, or second, by allowing the activity to "earn" its budget through sales to customers at (transfer) prices which cover its costs.

In the first case, budget authority is provided equal to a support activity's total yearly output multiplied by its unit cost target. In this case the unit cost target set by OSD as the regulator of an activity could be thought of as the price paid by OSD as the final customer of that activity's product. This is particularly useful in the case where support activities offer significant "positive externalities." For example, when benefits are not completely captured by individual customers and "spill-over" to the DoD as a whole (e.g. like joint US-International defense management education), such activities can be "centrally funded" (using unit-cost-based price targets) by a single "customer" (such as OSD) who acts as a representative for the larger interests of DoD as a whole.[16]

The second case reflects current DoD policy as represented by DBOF. In this case, rather than support activities receiving unit-cost-based budget appropriations, customers (i.e. operational commands) request budgets to accommodate their support purchases. In turn, support activities are authorized to sell their outputs to customers at regulated cost-based transfer prices designed to cover their total costs. Unit cost goals can then be used as a measure of efficiency by comparing "actual unit cost experience against planned corporate expectations." (DBOF Handbook [1995] p.3-16)

(3) "Reduce DoD support costs through better business practices": Within business areas, support organizations (providers) operate like commercial businesses, selling goods and services to customers. Customers establish their requirements and are charged for the cost of the products or services provided.[17] Thus DBOF providers "earn" their budgets based on the quantity of goods and services they sell. Under DBOF, customers—typically combat or operating units—fund their requests with appropriations from Congress (i.e. with Procurement, Operations & Maintenance (O&M), and/or Research, Development, Test & Evaluation (RDT&E) money). The expectation is that "when these costs are... visible to the operating forces they will... make better resource allocation decisions in determining the levels of support they require for day-to-day operations." (Maroni [1993]p.3) Thus DBOF creates a "business" (or "customer-provider") relationship between military operating forces and support organizations. More importantly, it helps to link mission operations with the cost to support those operations.

Military Departments and Defense Agencies that have business areas financed under DBOF are responsible for the day-to-day management and operation of their respective business areas. However, when it comes to setting transfer prices, "[t]here are few restrictions by actual statute on DBOF activities in the establishment of rates." (Isosaari [1996]p.19)

The Military Departments establish prices with oversight provided by the Office of the Under Secretary of Defense (Comptroller) (OUSD(C)).[18] Since OUSD(C) has oversight responsibility for generating unit-cost-based transfer prices, it effectively acts as the price regulator for support activities. Another assumption of the model is that the Comptroller (OUSD(C)) regulates transfer prices with the objective of encouraging cost reductions over time.

(4) "Foster efficiency and productivity improvements": A primary difference between DBOF business areas and private firms is that, by Congressional statute, DBOF activities must operate on a cumulative, non-profit (or "break-even") basis. Thus, activities financed under DBOF sell their goods and services to customers with the sole intent of recovering the total cost of providing those goods and services: "DBOF businesses strive to break even in prices charged to customers." (DBOF Handbook [1993])

However, as a DBOF business area sells goods or services, it earns revenues. The difference between revenues from sales and the actual costs incurred at any point in time is called the "Net Operating Result" (NOR). In general, during budget execution, a business area's NOR will either be positive (indicating profits) or negative (indicating losses). The "Accumulated Operating Result" (AOR) is the ultimate profit or loss realized from the operations of the business activity. Ideally, DBOF prices are set to achieve an AOR in the budget year of zero (see DBOF Handbook [1995].

According to the DBOF Handbook [1995]: "[R]ates must be adjusted by the activity's manager to offset prior year gains or losses, thereby achieving zero net profit and loss." (p.20) Thus activities are penalized with tougher targets (i.e. lower prices) for cost reductions that lead to a positive NOR. The struggle to break-even can lead to inefficient behavior. In a recent study, Pavia [1995] emphasizes that unit-cost-based transfer pricing alone may not be sufficient to induce providers to lower costs. (p.1061) In the absence of further cost controls (e.g. from detailed regulations, organizational incentives, or the threat of competition), DBOF activities' lack of a profit motive could "lead to large losses, taking years to recoup, and may lead to unusually high rates that may cause the alienation of valued customers." (Friend [1995]p.4) Since losses lead to price increases in subsequent years, as rates climb, further reductions in customer demand could force providers to spread fixed costs over fewer units, thus driving prices up even higher, eventually leading to what Friend [1995] has called the "death spiral." (p.5) One "... complaint about DBOF has been the rapid boosting of... rates." (Friend [1995]p.11) However, the DBOF Handbook [1995] emphasizes that: "the primary responsibility of DBOF activities is to provide services and products to its customers at the lowest cost." (p.21)[19]

This paper offers a new framework that helps address these issues. The model combines unit-cost-based transfer pricing with two organizational incentive programs designed to foster efficiency and productivity improvements: "gain-sharing" and "success-sharing".

Under "gain-sharing," a portion of any surplus (or "profits") due to cost-reductions is rebated to the activity in the form of employee bonuses, extra vacation time, etc. Under "success-sharing," rather than cutting future transfer prices to eliminate "accumulated profits," prices are ratcheted down to reflect a part, but not all of the cost reductions achieved. Thus, while success-sharing allows customers to benefit from cost reductions, it also rewards support activities with part of the future stream of benefits from their cost-reducing innovations. A useful, if imperfect analogy (and a possible way to distribute benefits derived from gain-sharing and success-sharing) is offered by current compensation practices familiar to some public sector wage earners. Gain-sharing would be similar to a (one-time) lump-sum bonus or a "merit increase," while success-sharing would be comparable to a performance-based promotion to a higher wage category or a "step increase" (a permanent increase in wage).

The next section introduces a model which illustrates analytically the incentives created by gain-sharing and the more potent incentives offered through success-sharing. The results of the model indicate that when support activities can invest in one period to lower their unit costs in a subsequent period, gain-sharing is a necessary, but not sufficient condition to create the incentives for them to do so. Success-sharing may be required to augment gain-sharing in order to encourage cost reductions over time.

THE MODEL

Assuming that "the primary responsibility of DBOF activities is to provide services and products to its customers at the lowest cost" (DBOF Handbook [1995]p.21), this section models a new budgeting approach for the Department of Defense. The objective of the model is to develop an incentive structure for internal support activities that rewards cost-savings and efficiencies. The model combines cost-based transfer pricing with the popular business practice of gain-sharing, and a new incentive program we call "success-sharing." Gain-sharing and success-sharing programs encourage support activities to lower costs by offering the opportunity to reward employees for cost-reducing innovations.[20]

The DoD Comptroller (OUSD(C)) currently has oversight responsibility for cost-based transfer prices set under DBOF. The Comptroller "finalizes and approves the stabilized rates that business activities may charge customers in a Program Budget Decision (PBD)." (Friend [1995]p.7) As a consequence, the Comptroller is modeled as a regulator (or "planner"), who sets maximum allowable transfer prices charged by sole providers of "core" support outputs.

The Comptroller's underlying objective in the model is to motivate activities to uncover and exploit cost-reducing innovations. The Comptroller offers gain-sharing and success-sharing programs, and regulates unit-cost-based transfer prices, to reward unit-cost reductions over time. Meanwhile, support activities seek investment strategies in unit-cost-reducing innovations that maximize the discounted present value of their rewards from gain-sharing and success-sharing programs, while simultaneously satisfying customer demands. The analysis begins with a discussion of the "principal's" (i.e. Comptroller's) problem, and then focuses on decisions taken by the "agents" (i.e. support activities). The model reveals conditions under which agents are likely to carry out the principal's objectives.

The model assumes each business area consists of a single support activity that produces one primary output, Q, to which all relevant costs can be assigned (where Q refers to the quantity of output or "workload" i.e. the number of units produced and sold per period). If the total cost function is given by TC(Q), then TC(Q)/Q = C(Q) are average total (or "unit") costs, while marginal costs are given by TC'(Q) = MC(Q).

Each activity is assumed to operate somewhere on the economist's standard "U-shaped" (quadratic) unit cost function, C(Q).[21] The unit cost function has a unique minimum at Q^*, where: $C'(Q^*) = 0$. Moreover, since $C''(Q) > 0$, for all Q: at output levels $Q < Q^*$, unit costs decrease in Q (i.e. $C'(Q) < 0$); while at output levels $Q > Q^*$, unit costs increase in Q (i.e. $C'(Q) > 0$). Finally, since $C'(Q) = (1/Q)[MC(Q)-C(Q)]$: *marginal costs* are below unit costs when $Q < Q^*$; above unit costs when $Q > Q^*$; and the same as unit costs when $Q = Q^*$. The model focuses on three possibilities. An activity can operate: i) on the decreasing portion of its unit cost function (where $Q < Q^*$ and $MC(Q) < C(Q)$); ii) at the minimum point (where $Q = Q^*$ and $MC(Q) = C(Q)$); or iii) on the increasing portion of its unit cost function (where $Q > Q^*$ and $MC(Q) > C(Q)$).

The model assumes that cost saving measures (i.e. unit-cost reducing investments) undertaken by an activity in the previous period, t-1, result in some quantifiable, permanent reduction in the *entire unit-cost function* in the current period, t, as well as in all subsequent periods. In this model, changes in output (or sales), Q, result in movements along the unit-cost function, C(Q), while (past) investments in unit-cost reductions, say I_{t-1}, lower the entire unit-cost function i.e. cost-reducing investments translate into the same unit-cost savings for any output level, Q. Thus, the cumulative stock of unit-cost savings achieved by t, say K_t, acts like a shift parameter on the initial unit-cost function, $C(Q)$, lowering the entire function by the amount K_t, but preserving the minimum point at Q^*. (also see Sweeney [1981])

Suppose the Comptroller allows activity managers to invest, I_t, each period in whatever alterations to the production process they choose in order to generate bonuses through gain-sharing and success-sharing.[22] Unit-cost savings are assumed to occur only after some initial investment, $I_0 > 0$, takes place (i.e. the initial stock of unit-cost savings is zero, $K_0 = 0$). In order to launch cost saving efforts, the Comptroller could offer seed money, I_0, at time t = 0, and allow the activity to invest what it chooses, I_t, in each subsequent period.

According to the DBOF Handbook [1995], it is the responsibility of the management of each DBOF business area to "Identify and justify... those improvements which will produce future gains in effectiveness and efficiency."(p.2-9) These improvements (or "cost-reducing innovations") could be as simple as minor workplace modifications that boost morale, or as complex as labor-saving (management education, worker training, etc.) and/or capital-saving (adopting new software or internet applications or EOQ inventory policies, etc.) technical changes in the existing production process.

It is also stated that capital investments to finance these changes must "increase the utility of existing assets for more than one accounting period, or... substantially increase operating efficiency over more than one accounting period." (DBOF Handbook [1995]p.3-11) The model assumes that investments undertaken by the activity in the past period, I_{t-1}, result in some permanent, quantifiable operating efficiency, $f(I_{t-1})$, captured as a (lagged) increase in the stock of cost savings, $K_t - K_{t-1}$, that lowers the entire unit-cost function for all subsequent periods. (also see Sweeney [1981]) Thus the change in the stock of unit-cost savings from any period t-1 to t is given by:

$$K_t - K_{t-1} = f(I_{t-1}); \text{ where } f'(I) > 0. \quad (1)$$

Equation (1) represents the evolution of unit-cost savings as a function of cost-saving investment in the previous period. The model assumes that past investment, I_{t-1}, translates into cost reductions, $f(I_{t-1})$, that add to the (permanent, cumulative) stock of unit-cost savings achieved over time i.e. $K_t = K_{t-1} + f(I_{t-1})$. Moreover, with $K_0 = 0$, writing, $K_t = \Sigma_{j=0}^{t-1} f(I_j)$, reveals that K_t is simply the cumulative stock of all unit-cost savings achieved up to time t-1.

Recent DBOF procedures established for capital budgeting (DBOF Handbook [1995]) suggest the Comptroller may have some historical basis for understanding the process described by (1). Note that K_t consists of a historical stream of unit cost savings (i.e. $K_t = f(I_0) + \ldots + f(I_{t-1})$) made up of individual unit-cost "success stories," $f(I_j)$, that are (in principle) observable.

The cumulative impact of cost reductions on the initial unit cost function, C(Q), is similar to that which might result from (and, in fact, could originate from) combining a "learning curve" with the original unit cost function. The cumulative stock of unit-cost savings, K_t, acts like a shift parameter on the initial unit-cost function, C(Q). For example, the unit cost in period $t - 1$, for any output level Q, is given by $C(Q) - K_{t-1}$. However, the investment, I_{t-1}, lowers unit costs in the next period, t, by $f(I_{t-1})$, to $C(Q) - K_{t-1} - f(I_{t-1}) = C(Q) - K_t$. In the model, cost-reducing investments lower the entire unit-cost function, but preserve the minimum point, Q^*. A useful avenue for future research is to examine the implications of cost-reducing investments that shift the minimum point, Q^* (i.e. that lower the cost function and simultaneously: increase the point of minimum efficient scale (or "full capacity")—a shift to the right; or reduce the point of minimum efficient scale (or "full capacity")—a shift to the left).

It is reasonable to assume that the Comptroller does not know the precise shape of an activity's unit cost function. For example, the Deputy Comptroller of DoD, Alice Maroni [1993], reveals that "In developing the FY1993 defense budget, DBOF rates were established... (and customer accounts were sized)... based on the best judgement that could be made at that time... [However] [w]e have much to learn about the workload, cost, and revenue trends being experienced... "(p.4)

Under such limited (or "asymmetric") information it is common for "[p]lanners... [to] use recent performance as a... basis for setting future indicators." (Weitzman [1976]p.253) Thus, an important simplifying assumption is that future transfer prices are developed from prior year unit cost experience. Given this scenario, and following Vogelsang & Finsinger [1979] and Sappington & Sibley [1988], support activities' cost and output data in the model are assumed to be revealed to the Comptroller (price regulator) with a one-period lag. As a consequence, the Comptroller (price regulator) sets a stabilized transfer price for each period, P_t, partly based on last period's unit costs, $C(Q_{t-1}) - K_{t-1}$.

Under current DBOF policy, "[r]ates remain in effect for a fiscal year to be used to bill the customer for work or service." (Isosaari [1996]p.20) Thus, the Comptroller sets a stabilized price, P_t, for a support activity's output for a given fiscal year, t, and then holds that price constant during the year of execution. This "stabilized rate" policy was originally designed "to protect appropriated fund customers (operating forces) from unforseen cost changes and thereby enable customers to more accurately plan and budget for DBOF support requirements."(DBOF Handbook [1995]p.3-8) However, the fact that transfer prices remain unchanged over the fiscal year delivers another advantage. This so-called "regulatory lag" opens the door for a gain-sharing initiative to reward cost savings. According to Rogerson [1994]: when the "... regulatory adjustment of prices in response to cost reductions... lag[s] behind the actual achievement of cost reductions [this] creates an incentive for cost-efficiency." (p.65)

Since unit cost information in the model is revealed to the regulator with a one-period lag, the potential exists for a support activity to generate a surplus (or "profit") by driving its actual unit costs below the stabilized transfer price, P_t, during the period of regulatory lag. According to Laffont & Tirole [1993], Rogerson [1994] and others, in situations where the agent (the support activity) has better information than the principal (the regulator) about costs, it will generally be optimal for the principal to offer the agent a contract (e.g. gain-sharing) that leaves the agent with some economic profit (or "surplus"), in order to give the agent an incentive to reduce costs. Thus, combining gain-sharing with a "stabilized" unit-cost-based pricing policy can encourage the constructive exploitation of regulatory lags.

According to Rogerson [1994]: "[a] general theme of the principal-agent literature is that in situations with asymmetric information, incentive schemes which cause the agent to reduce costs often necessarily also leave the agent with economic profit... This suggests... [the] creation of efficiency incentives. [For example], using incentive schemes such as regulatory lag that leave profit to the agent may be a particularly desirable policy for DoD to consider."p.76 (also see Demsetz [1968], Laffont & Tirole [1993], and Riordan [1993]).

Thus, combining cost-based transfer pricing and gain-sharing incentives with regulatory lags offers one approach to help overcome the principal-agent problem. This "incentive scheme" rewards extra revenues to support activities who are successful in achieving cost savings, i.e. in cutting actual unit costs below the previous period's unit costs—or below the regulated price, P_t. For the remainder of the paper, the term "surplus" (as opposed to "profits") will be used to describe these extra revenues. The objective is to distinguish earnings derived from cost-reducing innovations (i.e. "surplus"), from presumably less desirable earnings (i.e. "profits") that might be extracted from an internal organization's monopoly (or market) power, or from some other scheme.

Gain-sharing initiatives allow an activity to retain a fraction ($g \epsilon [0,1]$) of any surplus, say S_t, earned through unit-cost reductions in a given year, t. The activity's gain-sharing bonus (gS_t) is the fraction of any surplus the support activity is entitled to retain, and to distribute internally, to reward cost savings. The remainder, $(1 - g)S_t$, are actual savings (to DoD) for the Fiscal Year.

The smaller the Comptroller sets the gain-sharing parameter, g, the greater the share of savings (surplus) that accrues to DoD, but the less incentive the support activity has to reduce costs and generate those savings. Conversely, the larger the Comptroller sets the gain-sharing parameter, g, the lower the share of savings that accrue to DoD, but the greater the incentive the support activity has to reduce costs.

Consider two extremes. With 100 percent gain-sharing ($g = 1$), the support activity retains all its cost savings (for employee bonuses, etc.), and there is no immediate gain to DoD. Alternatively, with no gain-sharing ($g = 0$), it will be demonstrated that there is still no gain to DoD in the model, since there is no incentive for cost savings to occur over time. In the latter case (although monitoring "performance" is not formally modeled), if the Comptroller sets $g = 0$ and then attempts to "impose" lower unit cost targets (or cut transfer prices unilaterally), the Comptroller is likely to incur more burdensome monitoring costs to insure forced savings are not achieved at the expense of quality, or through "cost-shifting" or through some other "creative scheme."

Brief experiments in DoD with so-called "productivity gain-sharing" returned up to 50 percent (i.e. $g = 0.5$) of cost savings to the activity responsible.[23] (see Alderman [1993], Orvis, et.al. [1992], and Shycoff [1992]) However, according to the then acting Comptroller: "[t]he remainder of the savings will remain in the DBOF or operating budget and will be reflected in the next fiscal year's unit cost goals and price reduction to the customers." (Don Shycoff [1992]p.3) Thus, under productivity gain-sharing, beating the unit-cost-based price target earned support activities a share of current year cost savings, but made the target harder to beat in subsequent years.

In contrast, "success-sharing" shares the success of any permanent cost savings between customers and the activity responsible for those savings. Under success-sharing, while customers might still enjoy the benefit of new lower prices, the regulator does not lower prices to the full extent of the cost-savings achieved. This leaves support activities with future unit-cost (price) targets that are easier to beat than they might otherwise have been. As a consequence success-sharing tends to encourage further cost savings. Success-sharing initiatives (represented by the parameter, $s \epsilon [0,1]$) are designed to share the permanent (cumulative) unit-cost savings, K_t, between customers and the activity responsible for those savings.

The Comptroller is assumed to set (or regulate) transfer prices each period, P_t, based on past (observed) unit-costs, $C(Q_{t-1}) - K_{t-1}$, modified to account for (expected) unit-cost savings, $f(I_{t-1})$, generated from the most recent (observed) investment, I_{t-1}, with an allowance for success-sharing. Thus the stabilized price an activity can charge for each unit of output over the period t, is given by,

$$P_t = C(Q_{t-1}) - (1 - s)K_t; \qquad (2)$$

where: P_t consists of the original unit cost function, $C(Q_{t-1})$, adjusted for the (cumulative) downward shift in unit-cost savings, $K_t = K_{t-1} + f(I_{t-1})$, plus whatever allowance is made for success-sharing, sK_t.

The smaller the Comptroller sets the success-sharing parameter, s, the lower the price to customers, but the less incentive the support activity has to reduce costs. Conversely, the larger the Comptroller sets the success-sharing parameter, s, the higher the price to customers, but the greater the incentive the support activity has to reduce costs in subsequent periods.

Consider two extremes. With 100 percent success-sharing ($s = 1$), regulated transfer prices are permanently "stabilized" and never reflect any investment in cost savings. In this case, the Comptroller shares all subsequent (permanent) unit-cost savings with the support activity responsible, leaving customers without the benefit of lower prices. Alternatively, in the usual case of no success-sharing ($s = 0$), after a one-period lag, regulated transfer prices always fully reflect any unit-cost savings achieved. However, while a small s grants immediate benefits (price relief) to customers, the support activity does not have as large an incentive to invest in cost-reducing innovations, and thus long-term savings may be disappointing (i.e. to customers, DoD, Congress, or taxpayers).

The Comptroller's underlying objective in the model is to develop an incentive structure that motivates activities to invest in cost-reducing innovations that lead to permanent cost-savings. The Comptroller offers gain-sharing (g), success-sharing (s), and regulates unit-cost-based transfer prices, P_t, to encourage unit-cost reductions over time. In turn, support activities seek an investment strategy to reduce their unit costs over time, such as to maximize their returns (or share of the "surplus") from gain-sharing and success-sharing programs. The remainder of this section will focus on a support activity's response to the combined incentives of the unit-cost-based transfer pricing rule (equation (2)), gain-sharing (g), and success-sharing (s).

The last condition imposed on the support activity in the model is a requirement to satisfy demand at the regulated price. Most regulated public utilities face a similar mandate. A support activity must produce the quantity of output demanded by its customers at the unit-cost-based regulated transfer price set by the Comptroller. Assuming that a customer's demand for the support activity's output is sensitive to price, the relationship between the price charged and the quantity demanded from the activity becomes an important component of the analysis.

The means by which customers justify and obtain resources from DBOF activities is through DoD's Planning, Programming, and Budgeting System (PPBS). Once transfer prices are established, then customers determine how much support they will purchase at those prices. Resources required by customers to purchase business area products are subsequently identified in budget request document. "[C]ustomers determine the amount of goods and services they expect to purchase. . . and prepare their budget documents based on the projected rates and prices for those goods and services." (DBOF Handbook [1995])p.3-6)

Here we are concerned with activities whose ("natural" or internal) monopoly position would, in the absence of regulatory oversight (or unit-cost-based transfer pricing), allow these activities to independently determine the (monopoly) price they could charge customers for their product. The study of industrial organization suggests that as the product price increases, less is demanded by customers of firms with market power. For example, in the case of repairs "[a]s. . . prices climb, operating unit commanders [i.e. customers] who have limited funds available may economize and reduce the number of units submitted for repair." (Friend [1996]p.5)

Thus, while a support activity must satisfy demand at the regulated price, the quantity demanded of a support activity's output is generally sensitive to that price. The demand function that captures this relationship can be written in two different ways: either as,

$$Q_t = F(P_t);\ \text{or as, } P_t = F^{-1}(Q_t) \equiv D(Q_t); \quad (3)$$

where: $D'(Q_t) < 0$. In the model, it is convenient to follow the economist's convention and use the latter, "inverse demand function," to represent customer demand. In the case where demand is set exogenously by policy-makers, and is thus insensitive to price, the demand equation in (3) is said to be perfectly "inelastic."

Faced with the Comptroller's binding regulatory pricing constraint (equation (2)), and a quadratic unit cost function, say AC(Q) = C(Q)-K, a support activity seeks investments in cost-reducing innovations (governed by (1)), that satisfy customer demands (according to (3)) and maximize returns from gain-sharing (g) and success-sharing (s). Rewards from gain-sharing in any period, t, are given by gS_t; where S_t is the surplus earned during the period of regulatory lag.

The activity's surplus in the model is the difference between its total revenues (price times quantity sold) and its total costs (which include production and investment related expenses). Alternatively, the total surplus, S_t, consists of the "profit" generated on each unit sold, multiplied by actual sales, minus the total organizational costs of current investment. The amount of profit generated on each unit sold is the difference between the stabilized unit-cost-based transfer price, P_t, that the activity is allowed to charge its customers over the period of regulatory lag, and the actual (or realized) unit cost of producing each unit, $C(Q_t) - K_t$. However, multiplying the per-unit-profit, $P_t - [C(Q_t) - K_t]$, by actual sales, Q_t, yields only part of the total surplus, S_t. Investments in cost-reducing innovations and organizational costs associated with those investments need to be subtracted out.[24] These costs are captured by the investment cost function, $h(I_t)$, where: $h'(.) > 0$, and $h''(.) > 0$.

The support activity can make a surplus or deficit or break even over the period of regulatory lag, depending on whether or not it covers all of its costs when it charges the regulated unit-cost-based transfer price, P_t. These costs include production costs as well as investment costs. Thus, the surplus function in any period t can be written as:

$$S_t = P_tQ_t - [C(Q_t) - K_t]Q_t - h(I_t); \quad (4)$$

The first term on the RHS of (4) represents earned revenues, or the allowed price times the actual quantity sold. The second term on the RHS of (4) represents actual total production costs, or the actual unit costs incurred (in brackets) times the quantity sold. The last term, $h(I_t)$, is an investment cost function that reflects diminishing returns to cost-reducing investments. It is designed to capture the organizational (or "total system") cost of adopting new technologies (e.g. actual investment costs, together with any training costs, morale problems, etc.) to reduce its future unit production costs.

In the model, the Comptroller regulates transfer prices according to equation (2), and offers gain-sharing (g) and success-sharing (s) incentive programs that reward cost-reductions over time. An activity's rewards from gain-sharing is given by gS_t; where $g\epsilon[0,1]$ reflects the share of the surplus, S_t (given by equation (4)), that the Comptroller decides to reward during the period of regulatory lag. Meanwhile, an activity's rewards from success-sharing depend on the new price, P_t (given by equation (2)), set at the end of each period (or fiscal year), where $s\epsilon[0,1]$ reflects the share of the unit-cost savings the Comptroller decides to allow the activity to keep in future periods.[25] The next section models a support activity's response to this incentive structure.

THE SOLUTION

The support activity's objective is to choose an investment strategy to generate cost-reductions over time, that maximizes the discounted present value of its gain-sharing and success-sharing returns from its future stream of surpluses, while simultaneously satisfying customer demands. More formally, the activity's problem is to maximize,

$$W = \sum_{0}^{\infty}\{g/(1+r)^t\}S_t; \quad (5)$$

where: the surplus at any time, t, is given by,

$$S_t = P_tQ_t - [C(Q_t) - K_t]Q_t - h(I_t); \quad (4)$$

the (inverse) demand the activity must satisfy is,

$$P_t = D(Q_t); \quad (3)$$

the regulated transfer price charged by the activity and set by the Comptroller is,

$$P_t = C(Q_{t-1}) - (1-s)K_t; \quad (2)$$

the change in the stock of unit cost savings from past investment decisions is,

$$K_t - K_{t-1} = f(I_{t-1}); \quad (1)$$

gain-sharing and success-sharing parameters set by the Comptroller are, respectively, $g\epsilon[0,1]$ and $s\epsilon[0,1]$; and the rate at which a support activity discounts the future is given by, $r\epsilon[0,1]$.

While this problem is too complex to obtain a complete analytical solution, useful insights can be derived from studying the (long-run) stationary equilibrium. To solve for the stationary equilibrium, it is useful to focus on the output consequences of an activity's investment decisions.

The connection between an activity's investment choices and the resulting output consequences is intuitive and fairly immediate. From equation (1), investment decisions translate into cost-reductions that are eventually captured, in equation (2), as regulated prices. In turn, the price charged by an activity impacts customers' demands for the output through equation (3). As a consequence, any investment strategy essentially has an output counterpart. While, in principle, either output or investment can be used for purposes of obtaining stationary equilibrium results, in practice, solving the model in terms of output requires considerably less assumptions. In any case, the stationary equilibrium is achieved at some stationary equilibrium output level, say Q_E, when no further cost-reducing investment is initiated.

Given the properties of the objective function together with the constraints, the first order condition requires that, at the optimum, for any two periods t and t+1, the increased profit from a small increase in output one period should be just offset by the discounted loss of profit in the subsequent period, or

$$dS_t / dQ_t + (1/(1+r))(dS_{t+1}/dQ_t) = 0.^{26} \quad (6)$$

In order to investigate the implications of the first order condition given by (6), the support activity's surplus function, S_t, given by (4), must be written exclusively as a function of output, i.e. Q_{t-1}, Q_t, $and Q_{t+1}$. (see equations (4a&b) in Appendix 1) This requires two further simplifying assumptions. First, changes in the stock of cost savings are assumed to be a fraction ($a\epsilon[0,1]$) of past investment, I_{t-1}, or

$$K_t - K_{t-1} = f(I_{t-1}) = aI_{t-1}. \quad (1')$$

Second, the investment cost function in the surplus equation, (4), is given a functional form that reflects diminishing returns to cost-saving investments, or

$$h(I_t) = (1/2)bI_t^2; \text{ where } b > 0, \text{ and } h', h'' > 0. \quad (4')$$

Together, these two assumptions yield an expression for (6) exclusively in terms of output (see Appendix 1),

$$B(Q_{t-1}, Q_t, Q_{t+1}, Q_{t+2}) = 0. \quad (7)$$

Meanwhile, substituting (3) into (2), and using (1'), investment can also be written exclusively in terms of output (see Appendix 1),

$$I_t = (1/a)(1/(1-s))\{[C(Q_t) - C(Q_{t-1})] - [D(Q_{t+1}) - D(Q_t)]\}. \quad (8)$$

The stationary equilibrium output level is given by $Q_E = Q_{t-1} = Q_t = Q_{t+1} = Q_{t+2}$. It is immediately clear from (8) that, once the stationary equilibrium output level, Q_E, is attained no further cost-reducing investment will occur. At that point, the marginal (organizational) costs of any (further) investment outweighs the marginal benefits from gain-sharing and success-sharing.

At the stationary equilibrium, condition (7) can be written more explicitly as,

$$B(Q_E) = [g/(1-s)(1+r)]\{s(1+r)[MC(Q_E) - MR(Q_E)] - rC'(Q_E)Q_E\} = 0, \quad (9)$$

where: $MC(Q_E) = [C(Q_E)] + C'(Q_E)Q_E]$ are marginal production costs (based on the original cost function, C(Q)), and $MR(Q_E) = [D(Q_E) + D'(Q_E)Q_E]$ is the marginal revenue from sales, Q_E.

From (9), as long as an activity is offered some amount of gain-sharing and success-sharing (i.e. with $g\epsilon(0,1)$ and $s\epsilon(0,1)$), if future returns are treated the same as immediate returns (i.e. with a discount rate, $r = 0$), the model solution is analogous to the usual static optimization result for any profit-maximizing firm with market power. Notably, the optimal stationary equilibrium output level, Q_E, is attained where marginal cost equals marginal revenue (i.e. where $C(Q_E) > MC(Q_E) = MR(Q_E)$). The difference from the static result is that, with initial output, $Q^0 < Q^*$, and with $D'(Q) < C'(Q) < 0$, from (2a&b) in Appendix 1, at the stationary equilibrium, the cumulative stock of unit-cost-savings is $K > 0$. Thus, with r = 0, an activity's cost-reducing investments eventually drive its actual unit-costs down to $AC(Q_E) = C(Q_E)-K$; *where* Q_E satisfies $MC(Q_E) = MR(Q_E)$.

However, under the reasonable behavioral assumption that managers of support activities pay closer attention to near-term results, a positive time rate of discount ($r > 0$) is more appropriate. The remaining analysis focuses on this scenario.

Given a positive discount rate, it is useful to consider two reference points to help analyze the stationary equilibrium. The first reference point reflects the possibility the stationary equilibrium is at the initial point (Q^0, P^0), where the regulator first sets a price at which demand intersects the average cost function, or where: $P^0 = C(Q^0) = D(Q^0)$. The initial price, P^0, is assumed to be set at a level such that it just covers the activity's initial cost per unit, $C(Q^0)$, of satisfying the resulting demand (Q^0), or $P^0 = C(Q^0) = D(Q^0)$. A second useful reference point is the minimum of the unit cost function, Q^* (i.e. where $C'(Q^*) = 0$). From (9), the first reference point yields,

$$B(Q^0) = [g/(1-s)(1+r)]\{s(1+r)[MC(Q^0) - MR(Q^0)] - rC'(Q^0)Q^0\}, \quad (9a)$$

while the second reference point yields,

$$B(Q^*) = [gs/(1-s)]\{C(Q^*) - MR(Q^*)\}. \quad (9b)$$

Note from (9a&b), that gain-sharing is necessary for activities to invest in cost-reducing innovations. With g = 0, the result is always the same, i.e. $B(Q^0) = 0$ (and $B(Q^*) = 0$). As a consequence, without gain-sharing, the stationary equilibrium is simply the initial output level, Q^0 (i.e. $Q_E = Q^0$), and the corresponding stationary equilibrium price is simply the initial regulated price, $P^0 = C(Q^0) = D(Q^0)$. Moreover, from (2a&b) in Appendix 1, $K = [1/(1-s)][C(Q^0)-D(Q^0)] = 0$, or in other words, the stock of cost savings at the stationary equilibrium is zero. The implication is that, without gain-sharing, it does not pay to invest in cost-reductions. Thus, gain-sharing is required for cost-reducing investment to occur in the model. As a consequence, the analysis that follows assumes some degree of gain-sharing, i.e. $g > 0$.

From (9a&b) it is possible to narrow the analysis of stationary equilibria down to three cases. Each case is defined according to where an activity first operates on its initial unit cost function, C(Q). An activity can operate: (1) on the decreasing section of its unit cost function (where $Q < Q^*$ and $C'(Q) < 0$); (2) at the minimum point (where $Q = Q^*$ and $C'(Q) = 0$); or (3) on the increasing section of its unit cost function (where $Q > Q^*$ and $C'(Q) > 0$).

It is an empirical question where an activity finds itself on its initial unit cost function. If an activity's unit costs decrease with output (or "scale"), then it enjoys so-called "economies of scale" (see Case (1)). In contrast, if an activity's unit costs increase with output (or "scale"), then it suffers from "diseconomies of scale" (see Case (3)). Strikingly different results are obtained from gain-sharing in these two cases. A summary of the model results (when $g > 0$ and $r > 0$) is presented below, and in an accompanying series of graphs. Details of the calculations can be found in Appendix 2.

Case (1): (See Figure 1)

If an activity initially operates at some point, $Q^0 < Q^*$, on the declining section of its unit cost function (where $C'(Q) < 0$), gain-sharing alone can motivate the activity to drive its costs down over time. The activity benefits by investing in cost-reductions, taking advantage of the regulatory lag in price adjustments.

Thus, in the absence of success-sharing (i.e. with s = 0), gain-sharing alone can motivate activities to invest in cost reductions. The cumulative stock of cost-savings at the stationary equilibrium is, $K_{s=0} > 0$. These investments eventually drive the activity to operate at the minimum point, $Q_E = Q^* > Q^0$, of a lowered unit-cost curve, $AC(Q) = C(Q)-K$. At the stationary equilibrium, Q^*, the price is given by, $P_{s=0} = D(Q^*) = AC(Q^*) = MAC(Q^*)$; where: $MAC(Q) = AC'(Q)Q+AC(Q) = MC(Q)-K$, is the marginal cost function associated with the lowered unit-cost function, AC(Q).

In this case of decreasing unit costs (or "economies of scale"), combining the unit-cost-based pricing rule with gain-sharing alone, not only motivates activities to increase efficiencies and drive costs down over time, but also encourages production to con-

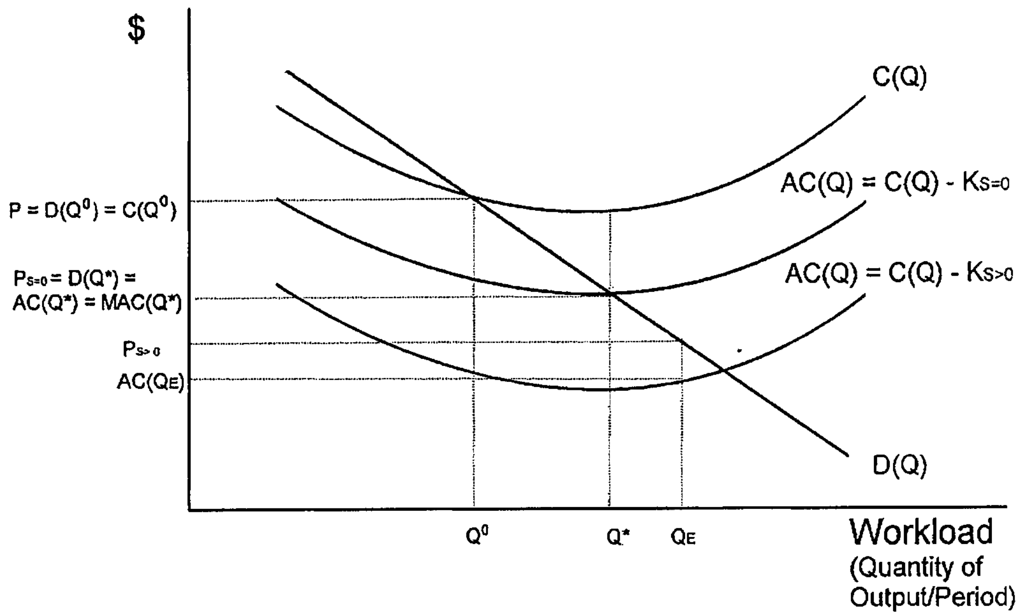


Figure 1.

verge to the economist's "full capacity" (or "minimum efficient scale"), where unit cost equals marginal cost. Thus, an additional advantage of combining the regulator's pricing rule (equation (2)) with gain-sharing ($g > 0$) is that unit-cost-based prices tend to automatically converge to preferred marginal-cost-based prices.

Nevertheless, any degree of success-sharing can promote even further cost-savings, i.e. at the new stationary equilibrium with $s = 0$, $Q_E > Q^* > Q^0$, and $K_{s>0} > K_{s=0} > 0$. It is useful to examine the impact of adding a success-sharing program from both sides of the organization—from the perspective of (internal) "customers," and from the perspective of (internal) "support activities."

From the point of view of customers, although unit costs are lower than they would be under gain-sharing alone, the price they pay is greater than actual unit costs, i.e. $P_{s>0} = D(Q_E) > AC(Q_E)$. However, from the activity's viewpoint, this price-cost difference can be thought of as the necessary (discounted stream of) rewards that motivates the search for further cost-savings. Regardless, customers still benefit from this new budgeting approach. They have more product available at the stationary equilibrium (i.e. $Q_E > Q^*$), and are charged a lower price (i.e. $AC(Q_E) < P_{s>0} < P_{s=0} = AC(Q^*)$), than under gain-sharing alone.

Case (2): (See Figure 2)

If an activity operates at the minimum point, i.e. $Q^0 = Q^*$, where $C'(Q) = 0$, gain-sharing alone will not have an impact. At the stationary equilibrium, $Q_E = Q^0 = Q^*$, and $P_{s=0} = C(Q^0) = D(Q^0)$, so that the stock of cost-savings is, $K = 0$. In this case success-sharing is required to motivate an activity to invest in cost-reductions.

Combining gain-sharing with success-sharing can promote cost-savings. Similar to Case (1), from the customer's viewpoint, although unit costs are lower at the new stationary equilibrium ($Q_E > Q_0 = Q^*$) than under gain-sharing alone, the customer pays more than actual unit costs, i.e. $P_{s>0} = D(Q_E) > AC(Q_E)$. However, from the support activity's viewpoint, this price-cost difference can be thought of as the necessary (discounted stream of) rewards that motivates the search for cost-savings. Regardless, customers still benefit from the new budgeting approach. At the stationary equilibrium, customers have more product available, $Q_E > Q^*$, and pay less for it (i.e. $AC(Q_E) < P_{s>0} < P_{s=0} = C(Q^*)$).

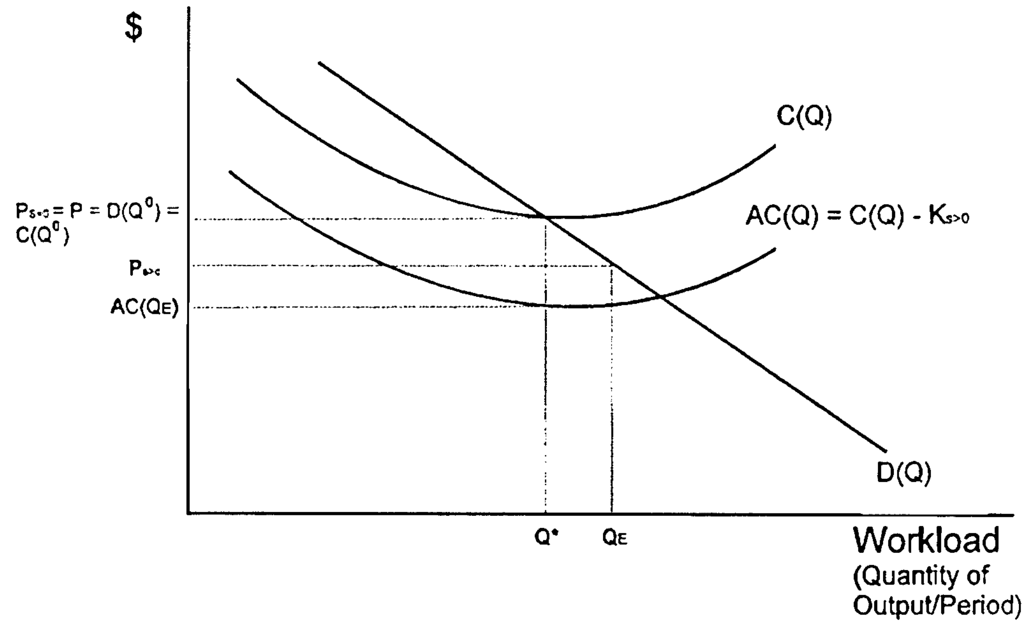


Figure 2.

Case (3): (See Figure 3)

Finally, if an activity initially operates at some point, $Q^0 > Q^*$, on the increasing section of its unit cost function (where $C'(Q) > 0$), even if gain-sharing is substantial (i.e. $g \rightarrow 1$), the threat of a subsequent collapse in the price (to the new, lower unit costs) wipes out the activity's incentive to invest in cost-savings. (see Figure 4)

In fact, with gain-sharing alone, the only (notional) stationary equilibrium is one where $K < 0$. However, a negative stock of cost savings corresponds to an increase in unit-costs!

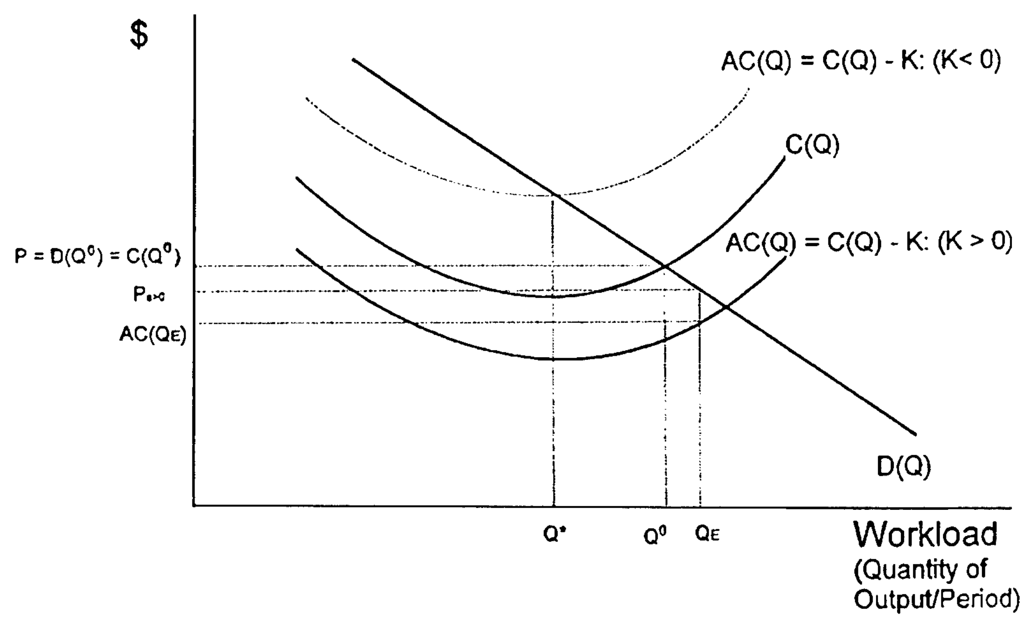


Figure 3.

Since the cumulative stock of cost savings is constrained to be non-negative in the model (i.e. cost-saving investment can only reduce unit-costs), gain-sharing alone does not yield a stationary equilibrium. In order to invest in cost-reductions, activities require a minimum amount of success-sharing.

The intuition for the results obtained in this case (when $g > 0$ and $s = 0$) can be seen using a simple example. Figure 4 illustrates a case where gain-sharing is not sufficient to induce an activity to invest in cost-reductions. If the activity produced 200 units last period (point A) at a unit cost of $5/unit, suppose this is the price the regulator allows it to charge its customers during the current period.

Suppose cost-saving efforts drive the unit cost function in the current period from C(Q) to AC(Q). Then the unit cost of production is driven down to $3/unit at point B. This would generate a maximum surplus during the period of regulatory lag equal to $2/unit, or a total of $400. If there is 50% gain-sharing (i.e. $g = .5$), then the maximum corresponding gain-sharing bonus is $200, to be distributed to the employees at the end of the current period.

However, now suppose the regulated price the activity can charge in the next period drops to the new unit cost level, or $3/unit. Since the quantity demanded by customers is sensitive to the price, this new, lower price generates more demand (250 units) by customers (point C). This increase in quantity demanded implies that production costs per unit for the activity will rise to $6/unit (point D), creating a $3 loss on each unit sold (or a $750 loss).

This means the activity manager faces a decision as to whether a $200 bonus today is worth bearing a $750 loss next period. To make the investment in cost-savings worthwhile to the activity, the discount rate would have to be an unrealistically high 275% (i.e. $200 \geq [1/(1+r)]750 \Rightarrow r \geq 2.75$). Even with 100% gain-sharing (i.e. with $g = 1$, such that the gain-sharing bonus equals the total surplus of $400), the discount rate applied would have to be over 85% (i.e. $r \geq .875$). This simple illustration suggests that, while gain-sharing can be demonstrated to be effective where there are economies of scale, it may not be effective when there are diseconomies of scale.

Worse yet, with diseconomies of scale it is possible activities might have a perverse incentive to increase their unit cost function. If a zero-profit policy is pursued (where surpluses are used as bonuses, and losses are recouped by

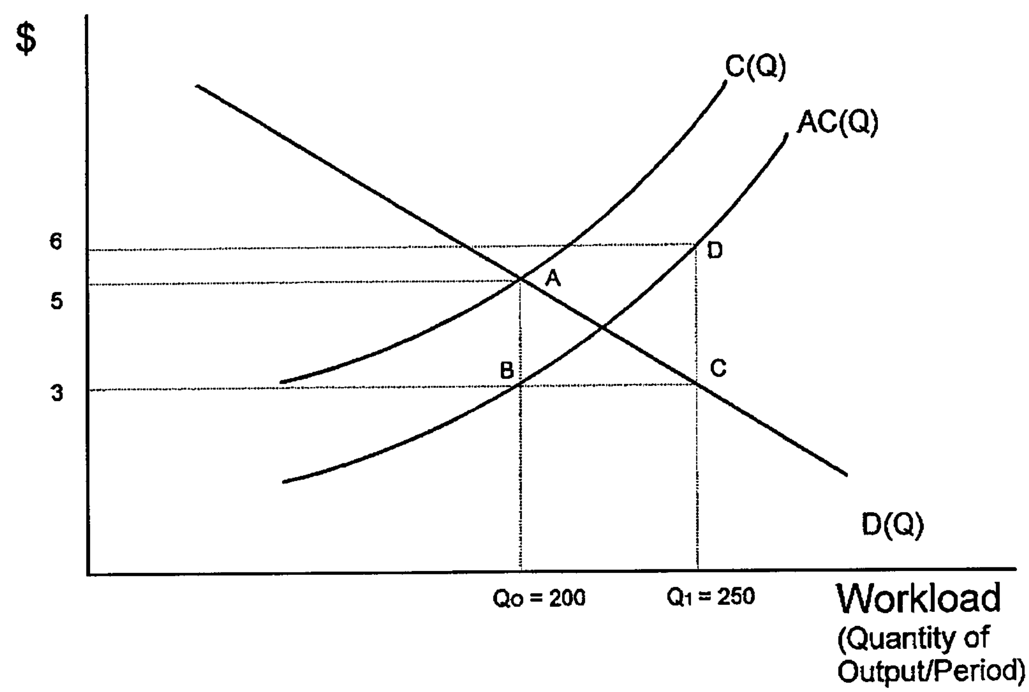


Figure 4.

adjusting future target prices), future target prices would reflect past losses. This means that the subsequent period target price would be the summation of the $6 unit cost and an adjustment factor to recoup the $750 total loss. This higher price would decrease the quantity demanded by the customer to a level perhaps even below the initial level of 200 units (i.e. to some quantity to the left of point A). A pricing system of this sort might either lead to instability, or perhaps as described in Appendix 2, to a stationary equilibrium at the minimum point, Q^*, on a higher unit cost curve, $AC(Q) = C(Q) - K$, where $K < 0$. This suggests that if gain-sharing is offered by itself, current DBOF pricing policy may need to be re-evaluated for those activities that operate under "diseconomies of scale" or "decreasing returns to scale."

One way to overcome this problem, and to induce managers to undertake cost-saving measures, is to prevent next period's target price from falling all the way to the new lower unit cost of production. This is precisely the concept behind "success-sharing." In the case of diseconomies of scale, although gain-sharing is necessary to motivate cost-savings over time, if offered by itself it can be counterproductive. A minimum level, $s[r/(1+r)]$, of success-sharing is required to attain a stationary equilibrium (i.e. $Q_F < Q^0 < Q^*$), where the stock of cost savings is, $K_{s>0} > 0$.

From the viewpoint of (internal) customers, although stationary equilibrium unit-costs are lower than starting unit-costs, $AC(Q_F) < C(Q^0)$, customers pay more than the new unit-costs, i.e. $P_{s>0} = D(Q_F) > AC(Q_F)$. However, from the activity's viewpoint, this price-cost difference can be thought of as the necessary (discounted stream of) rewards that motivates the search for cost-savings. Regardless, the final outcome under a combination of gain-sharing and success-sharing is favorable to the customer. More of the product or service is available, $Q_F > Q^0$, at a lower price than the starting price, i.e. $AC(Q_F) < P_{s>0} < P = C(Q^0)$.

CONCLUSION

This paper offers a new budgeting approach for the Department of Defense. The paper models an incentive structure for government activities that rewards cost-savings and efficiencies. The model combines cost-based-pricing with the popular business practice of gain-sharing, and a new incentive program called success-sharing. The dual objective of the paper was: first, to offer a framework to help upper-level management evaluate organizational incentives in cost-based pricing systems; and second, to persuade the Operations Research community to join in this effort to model critical budgeting issues.

The goal of gain-sharing is to encourage activities to make an effort to reduce their current costs in return for some of the immediate gains from the cost savings. Success-sharing refers to the degree to which any "permanent" success in obtaining cost-savings is shared between customers and the activity responsible for the savings. While success-sharing allows customers to benefit from cost reductions, it also rewards support activities with part of the future stream of benefits from their cost-reducing innovations.

Many large, complex, vertically-integrated organizations include separate activities that conduct internal exchanges of goods and services. The challenge faced by these organizations is to govern the relationships among internal activities to promote the goals of the organization as a whole. Many private firms solve this problem through the use of internal cost-based transfer prices.

This paper applies lessons from the transfer pricing literature to a unique subset of government activities—those that "earn" their budgets. Internal DoD support activities financed through DBOF offer an illustration. DBOF activities sell their goods and services to "customers" (operating forces) at regulated cost-based transfer prices. The primary difference between DBOF business areas and private firms is that, by Congressional statute, DBOF activities must operate on a cumulative, non-profit basis. A critical challenge is to align agent's (or support activities') incentives, with the objective (e.g. cost-reduction) of the "principal" (or Comptroller).

This paper offers a new budgeting approach that encourages support activities to lower costs by offering the opportunity to reward employees for cost-reducing innovations. Economist's have discovered that, both with defense firms and with electricity generating companies, regulatory adjustments of prices in response to cost reductions tend to lag behind the actual achievement of cost reductions, and that this can create an incentive for cost-efficiency. This concept of so-called "regulatory lag" is applied here to the regulation of internal "support activities."

In the model, the "principal" (i.e. the Comptroller) offers gain-sharing and success-sharing programs, and regulates unit-cost-based transfer prices with a lag, to reward unit-cost reductions over time. Meanwhile, the "agents" (i.e. support activities) seek investment strategies in unit-cost reducing innovations that maximize the discounted present value of their rewards from gain-sharing and success-sharing, while simultaneously satisfying customer demands.

One extension of the model would be to develop an explicit game between the principal and the agent(s). In this game, one player, the principal (or Comptroller), chooses a pricing rule, the optimal period of regulatory lag, and gain-sharing and success-sharing programs, to maximize (the discounted present value of) total cost savings over time. Meanwhile, the other player, the agent (or support activity), would operate much as modeled here. The agent would take the pricing rule and regulatory lag as given, and would invest in cost-reductions to maximize the discounted present value of returns from gain-sharing and success-sharing programs, while satisfying customer demands.

This new budgeting approach, with its "built-in" incentive structure, rewards a share of revenues to activities who are successful in achieving costs savings over time. A recent Congressionally-mandated study emphasizes that: "[a] powerful incentive in the DoD would be to give Service Secretaries and heads of defense agencies the authority to retain in their future "top line" planning a substantial portion of any savings [or "surplus"] that can be generated in their department or agency" (DD [1995]p.4-16) This policy could ultimately encourage the implementation of gain-sharing and success-sharing programs at lower levels in the organization.

The results of the model indicate that when support activities can invest in one period to lower their unit costs in a subsequent period, gain-sharing is a necessary, but not sufficient condition to create the incentives for them to do so. Success-sharing may be required to augment gain-sharing in order to encourage cost reductions over time.

Three important results are obtained in the model. Each result depends on where an activity first operates on its initial unit cost function. An activity can operate: (1) on the decreasing section of its unit cost function; (2) at the minimum point; or (3) on the increasing section of its unit cost function.

It is largely an empirical question where an activity finds itself on its initial unit cost function. If an activity's unit costs decrease with output (or "scale"), then it enjoys so-called "economies of scale" (Case (1)). In contrast, if an activity's unit costs increase with output (or "scale"), then it suffers from "diseconomies of scale" (Case (3)). Strikingly different results are obtained from gain-sharing in these two cases.

Given the current defense environment, a support activity may face a number of scenarios. Among these is: a) a cut in the demand for its product; b) increased competition from internal (or external) suppliers of a similar product; or c) an increase in demand for its product (say due to consolidation).

If an activity suffers significant cuts in demand for its output, it is more likely to fall under Case (1). The greater the actual (or threat of) competition, the more likely it is an activity operates at minimum unit cost, Case (2). Finally, if an activity experiences an increase in demand for its output, say due to consolidation, it is more likely to fall under Case 3. Regardless, in the absence of gain-sharing (i.e. with $g = 0$), it makes no difference where an activity operates on its initial unit cost function. There is no incentive for cost-savings in the model without some degree of gain-sharing.[27] In each case reviewed below, results are reported for gain-sharing alone, and then for a combination of gain-sharing and success-sharing.

Case (1): If an activity initially operates with decreasing unit costs, gain-sharing alone, combined with the regulated unit-cost-based transfer pricing rule, not only encourages activities to reduce costs over time, but also eventually results in marginal cost pricing. The activity benefits by investing in cost-reductions, because it can take advantage of the regulatory lag in price adjustments. In this case of decreasing unit costs (or "economies of scale"), combining the unit-cost-based pricing rule with gain-sharing alone, not only motivates activities to increase efficiencies and drive costs down over time, but also encourages production to converge to the economist's "full capacity" (or "minimum efficient scale"), where unit cost equals marginal cost. Thus, an additional advantage of combining the regulator's pricing rule with gain-sharing is that unit-cost-based prices tend to automatically converge to preferred marginal-cost-based prices.

Nevertheless, any degree of success-sharing can promote even further cost-savings. From the point of view of customers, although unit costs are lower than they would be under gain-sharing alone, the price they pay is greater than actual unit costs. However, from the activity's viewpoint, this price-cost difference can be thought of as the necessary (discounted stream of) rewards that motivates the search for further cost-savings. Regardless, customers still benefit from this new budgeting approach. They have more product available and pay a lower price than they would under gain-sharing alone. An extension of the analysis might investigate whether offering gain-sharing alone leads to a bias in the composition of investment—favoring immediate (but more "transient") cost-savings, at the expense of future (but more "permanent") savings that are rewarded under a success-sharing program.

Case (2): If an activity operates at minimum unit costs, gain-sharing alone will not have an impact. However, combining gain-sharing with success-sharing can promote cost-savings. From the customer's viewpoint, although unit costs are lower at the new stationary equilibrium than under gain-sharing alone, the customer pays more than the actual unit costs. However, from the support activity's viewpoint, this price-cost difference can be thought of as the necessary (discounted stream of) rewards that motivates a search for cost-savings. Regardless, customers still benefit from the new budgeting approach. At the stationary equilibrium, customers have more product available and pay less for it.

Case (3): With increasing unit costs, even if gain-sharing is substantial, the threat of a subsequent collapse in the price (to the new, lower unit costs) wipes out the activity's incentive to invest in cost-savings. Gain-sharing alone does not yield a stationary equilibrium. In fact, in the case of "diseconomies of scale," offering gain-sharing by itself can be counterproductive.

In order to invest in cost-reductions, activities require a minimum amount of success-sharing. From the viewpoint of (internal) customers, although unit-costs with success-sharing are lower than starting unit-costs, customers pay more than the actual unit-costs. However, from the activity's viewpoint, this price-cost difference can be thought of as the necessary (discounted stream of) rewards that motivates a search for cost-savings. Regardless, customers still benefit from the new budgeting approach. A combination of gain-sharing and success-sharing results in more of the product or service being made available, at a lower price to the customer.

Although combining gain-sharing and success-sharing with unit-cost-based pricing appears to offer an attractive alternative to conventional public budgeting, the manner in which such a system is implemented is critical to its success. A number of concerns remain to be addressed.

First, the model is silent about the depreciation of capital (either due to "wear & tear" or due to obsolescence). Adding depreciation would be a useful extension that would result in a steady state (investment plan), as opposed to the stationary state (in which there is no further investment) in the model. Another interesting opportunity to extend the model is to recognize that inaccurate forecasts of customer demands can impact other activities. The role of substitutes and complements and uncertainty in demand might be captured in a stochastic demand function. Other extensions could examine the impact of variable returns from cost-reducing investments, and explore the consequences of varying gain-sharing and success-sharing parameters over time.

A second concern is that the proposed budgeting system can only operate where organizational outputs, inputs, and customers are well defined. Moreover, a financial management accounting system is required that reveals all labor, materials and capital costs that contribute to each output at as disaggregated a level as possible. In the case of multiple outputs, any "common" or "joint" costs must be carefully allocated for the system to succeed in lowering overall costs. Moreover, in implementing this system, unit costs must include the (allocated) costs of the financial management accounting system itself, together with any required monitoring costs.

Third, to retain proper organizational incentives, the financial accounting system must be capable of separating exogenous cost changes from an activity's endogenous cost-savings. If exogenous (input) cost increases camouflage an activity's endogenous cost-savings, then incentives may not be awarded when they are in fact deserved. Conversely, if exogenous (input) cost decreases camouflage an activity's endogenous cost increases, then incentives will be awarded when they are not deserved. Further complicating the problem is the fact external (or exogenous) costs are not constant over time. Introducing a stochastic unit cost function in the model could offer further insights.

Fourth, cost savings can clearly be generated by reducing quality. Thus, there may be burdensome monitoring costs (as well as some additional accounting costs) in insuring that claimed cost savings are not achieved at the expense of quality, effectiveness, or through "cost shifting," or through some other creative "rent-seeking" schemes. To avoid these problems, the threat of outsourcing, competition, or the explicit regulation and monitoring of performance, continue to be essential to safeguard quality and effectiveness.

Fifth, the incentive problem must be broken into two parts: i) the "external" incentives provided to the activity (through gain-sharing and success-sharing) to motivate cost savings, and ii) the internal distribution of those incentives to motivate management and workers. The important question of the distribution of internal incentives remains an issue for further study (see footnote 27).

Finally, the success of gain-sharing and success-sharing programs depends on public organizations benefiting from their cost-savings. Earning profits in the public sector is a sensitive and controversial issue. However, it may be useful to educate the public of the important role profits can play in motivating cost-savings. A few precedents could make this job easier. These include beneficial suggestion programs that offer both lump-sum (gain-sharing type) returns, and more permanent (success-sharing type) returns based on the savings enjoyed by an organization over time. Comparisons can also be drawn between employee compensation plans that offer productivity rewards as one-time ("lump-sum" or "annual") bonuses (similar to gain-sharing), as opposed to "step" or "merit" increases—which are permanent increases in salary (similar to success-sharing). Moreover, it may prove useful to use the term "surplus" instead of "profits." The term surplus can be used to distinguish earnings derived from cost-reducing innovations, from the presumably less desirable earnings (i.e. profits) that might be extracted from an activity's internal market power.

Increasing cost awareness and instilling business practices in public activities is an important first step. The next step is to grant activity managers the financial authority and flexibility to invest in manpower, equipment, and other resources to improve quality and lower costs, and to reward them for doing so.

In the private sector, savings result in increased profits or improved effectiveness—metrics for which managers are rewarded. In contrast, "[i]n the Federal sector... most rewards are for strict compliance with rules... [Thus], better organizational incentives are needed." (DD [1995]p.4-17) This paper offers a new budgeting approach with a "built-in" set of organizational incentives. The main conclusion is that, although customers still have to monitor quality, combining "gain-sharing" and "success-sharing" with a cost-based pricing system can motivate substantial cost-savings over time. Moreover, these results are not unique to Defense. The model applies to any regulated activities subject to unit-cost-based pricing.

APPENDIX 1

If a support activity is required to satisfy demand at the regulated price, then (3) can be substituted into (2) yielding expressions for this and next periods stock of cost savings in terms of output. These are respectively,

$$K_t = (1/(1-s))[C(Q_{t-1}) - D(Q_t)], \qquad (2a)$$

and

$$K_t = (1/(1-s))[C(Q_t) - D(Q_{t+1})]. \qquad (2b)$$

Substituting (2a&b) into (1′) yields an expression for investment, I_t, in terms of output that can be given by (4), where $h(I_t) = (1/2)bI_t^2$. As a result of this and (1′), (2), (2a&b), and (3), surplus functions at t and t+1 can be written in terms of Q as follows:

$$S_t = D(Q_t)Q_t - Q_t\{C(Q_t) - (1/(1-s)) \cdot [C(Q_{t-1}) - D(Q_t)]\} - Z_t \qquad (4a)$$

and

$$S_{t+1} = D(Q_{t+1})Q_{t+1} - Q_{t+1}\{C(Q_{t+1}) - (1/(1-s))[C(Q_t) - D(Q_{t+1}]\} - Z_{t+1}, \qquad (4b)$$

where:

$$Z_t = (b/2a^2)(K_{t-1}^2 - 2K_tK_{t+1} + K_t^2),$$

and

$$Z_{t-1} = (b/2a^2)(K_{t+2}^2 - 2K_{t+1}K_{t+2} + K_{t+1}^2),$$

are functions of Q_{t-1}, Q_t, *and* Q_{t+1}, as a result of (2a&b). Using (4a&b) the condition given by (6) can now be expressed exclusively as a function of output, or as,

$$B(Q_{t-1}, Q_t, Q_{t+1}, Q_{t+2}) = 0. \quad (7)$$

APPENDIX 2

Given the initial price, $P^0 = C(Q^0) = D(Q^0)$, from (9a): with a success-sharing program (i.e. $s > 0$),

$$B(Q^0) \geq \text{or} \leq 0, \text{ as } s(1 + r)[C'(Q^0) - D'(Q^0)] - rC'(Q^0)\} \geq \text{or} \leq 0; \quad (9a')$$

In the absence of success-sharing (i.e. $s = 0$), this condition reduces to,

$$B(Q^0) \geq \text{or} \leq 0, \text{ as } -rC'(Q^0)\} \geq \text{or} \leq 0. \quad (9a'')$$

Meanwhile, from (9b): with $s > 0$,

$$B(Q^*) \geq \text{or} \leq 0, \text{ as}\{C(Q^*) - MR(Q^*)\} \geq \text{or} \leq 0, \quad (9b')$$

where, $MR(Q^*) = D'(Q^*)Q^* + D(Q^*)$; while with $s = 0$, the condition reduces to,

$$B(Q^*) = 0. \quad (9b'')$$

The three cases analyzed below are distinguished by the location of the initial output level, Q^0, relative to the minimum of the initial unit-cost function, Q^*: Case (1) examines $Q^0 < Q^*$; Case (2) examines $Q^0 = Q^*$; and Case (3) examines $Q^0 > Q^*$. In each case, the initial price is set where $P^0 = C(Q^0) = D(Q^0)$.

Case 1 ($Q^0 < Q^*$):

Suppose an activity operates on the declining portion of its unit cost function at some initial production level, $Q^0 < Q^*$, where $P^0 = C(Q^0) = D(Q^0)$, and that for any output level, $Q^0 \leq Q \leq Q^*$, $D'(Q) < C'(Q) < 0$. Two possible success-sharing scenarios need to be examined.

Case (1a): If $s = 0$, then it is immediate from (9a″) and (9b″) respectively that: $B(Q^0) > 0$ and $B(Q^*) = 0$. Thus, with no success-sharing, the stationary equilibrium is given by $Q_E = Q^* > Q^0$. Moreover, the fact that at Q^0, $C(Q^0) = D(Q^0)$, combined with a monotonic decreasing (inverse) demand function and $D'(Q)<C'(Q)<0$, $Q^0 \leq Q \leq Q^*$, implies that, $C(Q)>D(Q)$, $Q^0<Q\leq Q^*$. As a consequence, associated with the stationary equilibrium, $Q_E = Q^*$, there is a positive stock of cost-savings, $K_{s=0}$ (i.e. from (2a&b) in Appendix 1, $K_{s=0} = [C(Q^*) - D(Q^*)]>0$).

This indicates that gain-sharing alone (i.e. $g>0$ and $s = 0$) is sufficient for cost-reducing investments to take place. The stationary equilibrium at $Q_E = Q^*>Q^0$ on the new lower average total cost curve, $AC(Q) = C(Q)-K_{s=0}$, is illustrated in Figure 1. Expressing the regulated price as $P = AC(Q)+sK$, reveals a stationary equilibrium price, $P_{s=0} = D(Q^*) = AC(Q^*)$.

It is useful to define the marginal cost function associated with the new lower unit cost curve, $AC(Q)$, as, $MAC(Q) = d[AC(Q)Q]/dQ = AC(Q)+AC'(Q)Q$. Since, at $Q_E = Q^*$, $AC'(Q^*) = 0$, the regulated price at the stationary equilibrium is set equal to marginal costs, (i.e. since $MAC(Q^*) = AC(Q^*)$). In this case, gain-sharing combined with a unit-cost-based transfer pricing rule not only encourages activities to reduce costs over time (i.e. $K > 0$), but, in the limit, also eventually results in marginal cost pricing of the output (i.e. $P_{s=0} = MAC = AC$).

Case (1b): If $s > 0$, since $D'(Q)<C'(Q)<0$, $Q^0 \leq Q \leq Q^*$, from (9a′), $B(Q^0)>0$. However, since $C(Q) > D(Q)$, $Q^0>Q\leq Q^*$, from (9b′), it is also the case that $B(Q^*) > 0$. This implies that neither Q^0 nor Q^* are a stationary equilibrium. Thus, if it exists, the stationary equilibrium is at some output level, Q_E, beyond Q^*, such that $Q_E > Q^* > Q^0$. While gain-sharing alone is sufficient to encourage cost-savings, combining gain-sharing with success-sharing (i.e. $g > 0$ and $s > 0$) motivates further cost-reducing investment i.e. the stock of cost-savings with success-sharing is $K_{s>0} = [1/(1-s)][C(Q_E) - D(Q_E)] > K_{s=0}$, (i.e. the spread between $C(Q)$ and $D(Q)$ at Q_E is larger than at Q^* and $0<s<1$ implies $[1/(1-s)]>1$). Moreover, the stationary equilibrium price is $P_{s>0} = D(Q_E) = AC(Q_E) + sK > AC(Q_E)$. Thus, in the model, further cost savings can be achieved if customers are willing to pay prices above unit costs at the stationary equilibrium. Finally, although marginal production costs at Q_E are greater than unit costs,

$MAC(Q_E) > AC(Q_E)$, the price, $P_{s>0,}$ may or may not cover marginal costs, partly depending on the magnitude of success-sharing, sK (i.e. if success-sharing is sufficiently large, $P_{s>0} \geq MAC(Q_E) > AC(Q_E)$).

Case 2 ($Q^0 = Q^*$):

Suppose an activity operates at the minimum of its unit cost function at some initial production level, $Q^0 = Q^*$, where $P^0 = C(Q^0) = D(Q^0)$, and $D'(Q^0) < C'(Q^0) = 0$. Two possible success-sharing scenarios need to be examined.

Case (2a): If s = 0 and $Q^0 = Q^*$, then it is immediate from (9a″) and (9b″) respectively that: $B(Q^0) = 0$ and $B(Q^*) = 0$. Thus, with no success-sharing, the stationary equilibrium is given by $Q_E = Q^0 = Q^*$. As a consequence, while there is marginal cost pricing at the stationary equilibrium (i.e. $P^0 = MC(Q^0) = C(Q^0)$), since $K_{s=0} = [C(Q^0)-D(Q^0)] = 0$, even with a gain-sharing program (i.e. g > 0), if an activity currently operates at minimum unit costs, in the absence of a success-sharing program (i.e. with s = 0) there is no incentive in the model for investments in cost-reductions over time.

Case (2b): If s > 0 and $Q^0 = Q^*$, since $D'(Q^0) < C'(Q^0) = 0$, and $P^0 = C(Q^0) = D(Q^0) > MR(Q^0)$, then from (9a′), $B(Q^0) > 0$, and from (9b′), it is also the case that $B(Q^*) > 0$. This implies that neither $Q^0 = Q^*$ nor $Q < Q^*$ is a stationary equilibrium. Thus, if it exists, the stationary equilibrium is at some output level, Q_E, beyond Q^*, such that $Q_E > Q^0 = Q^*$. Since the stock of cost savings at this point is, $K_{s>0} = [1/(1-s)][C(Q_E)-D(Q_E)] > K_{s=0} = 0$, combining gain-sharing with success-sharing (i.e. g > 0 and s > 0) is required to motivate cost-reducing investment. Moreover, the stationary equilibrium price is $P_{s>0} = D(Q_E) = AC(Q_E) + sK > AC(Q_E)$. Thus, in the model, cost savings are achieved only if customers are willing to pay prices above unit costs at the stationary equilibrium. Finally, although marginal production costs at Q_E are greater than unit costs, $MAC(Q_E) > AC(Q_E$, the price, $P_{s>0}$, may or may not cover marginal costs partly depending on the magnitude of success-sharing, sK (i.e. if success-sharing is sufficiently large, $P_{s>0} \geq MAC(Q_E) > AC(Q_E)$).

Case 3 ($Q^0 > Q^*$):

Suppose an activity operates on the increasing portion of its unit cost function at some initial production level, $Q^0 > Q^*$, where $P^0 = C(Q^0) = D(Q^0)$, and that for any output level, $Q^* \leq Q \leq Q^0$, $C'(Q) > 0 > D'(Q)$. Two possible success-sharing scenarios need to be examined.

Case (3a): If s = 0, then it is immediate from (9a″) and (9b″) respectively that: $B(Q^0)<0$ and $B(Q^*) = 0$. Thus, with no success-sharing, if it exists, the stationary equilibrium is given by $Q_E = Q^* < Q^0$. However, the fact that at Q^0, $C(Q^0) = D(Q^0)$, combined with a monotonic decreasing (inverse) demand function and $C'(Q) \geq 0 > D'(Q)$. $Q^* \leq Q \leq Q^0$, implies that, $C(Q)<D(Q)$, $Q^* \leq Q < Q^0$. As a consequence, associated with the stationary equilibrium, $Q_E = Q^*$, there is a **negative** stock of cost-savings, i.e. $K_{s=0} = [C(Q^*) - D(Q^*)]<0$. However, this implies the activity will invest to pad (or increase) its costs, and only cost-reducing investments are allowed in the model. So $Q_E = Q^*$ cannot be a stationary equilibrium. It is demonstrated in (3b) below that a minimum degree of success-sharing is required when an activity operates with unit production costs that increase with workload. Moreover, if a regulator has a pricing rule that compensates activities for cost increases, surplus-maximizing activities may have an incentive to make spurious "investments" when they operate under increasing unit cost conditions.

Case (3b): If s > 0, since $C'(Q) \geq 0 > D'(Q)$, $Q^* \leq Q \leq Q^0$, from (9a′), if $0 > s > [r/(1+r)]$, then $B(Q^0) > 0$. Meanwhile, since $C(Q)<D(Q)$, $Q^* \leq Q < Q^0$, if the demand curve is sufficiently inelastic (i.e. customers are relatively insensitive to price), then $C(Q^*) > MR(Q^*)$, and from (9b′), it is also the case that $B(Q^*) > 0$. Marginal revenue at Q^*, $MR(Q^*)$, can be written in terms of the price elasticity of demand ($E_{Q,P} = (\% \Delta Q/ \% \Delta P) = (D'Q/D)<0$): $MR(Q^*) = D'(Q^*)Q^* + D(Q^*) = D(Q^*)[(1/E_{Q,P})+1]$, where inelastic demand is given by $|E_{Q,P}| < 1$. If these conditions hold, then neither Q^0 nor Q^* (nor any other arbitrary reference point, $Q^* < Q < Q^0$) are a stationary equilibrium. Thus, if it exists, the stationary equilibrium is at some output level, Q_E, beyond Q^*, such that $Q_E > Q^0 > Q^*$. Since the stock of cost savings at this point is, $K_{s>0} = [1/(1-s)][C(Q_E)-D(Q_E)] > 0$, combining gain-sharing with a minimum level of success-sharing (i.e. g > 0 and s > [r/(1+r)]) is required to motivate cost-reducing investment. Moreover, the stationary equilibrium price is $P_{s>0} = D(Q_E) = AC(Q_E) + sK > AC(Q_E)$. Thus, in the model, cost savings are achieved only if customers are willing to pay prices above unit costs at the stationary equilibrium. Finally, although marginal production costs at Q_E are greater than unit costs, $MAC(Q_E) > AC(Q_E)$, the price, $P_{s>0}$, may or may not cover marginal costs partly depending on the magnitude of success-sharing, sK (i.e. if success-sharing is sufficiently large, $P_{s>0} \geq MAC(Q_E) > AC(Q_E)$).

REFERENCES

Alderman, K.C., "Defense Business Operations Fund: A Framework for Organizational Gain-Sharing," Armed Forces Comptroller, Fall 1992.

Benke, R.L., and D.J. Edwards, Transfer Pricing: Techniques and Uses, National Association of Accountants, New York, 1980.

Brimson, J., "Activity Accounting: An Activity-Based Costing Approach, New York: John Wiley & sons, 1991.

Bruns, W.J., and R.S. Kaplan (eds), Accounting and Management: Field Study Perspectives, Harvard Business School Press, Boston, MA, 1987.

Business Management Directorate, "The Defense Business Operations Fund," Office of the DoD Comptroller, Pentagon, Wash D.C. [1993]

DBOF Handbook, The Defense Business Operations Fund, Office of the Under Secretary of Defense (Comptroller), Wash. D.C., 1995.

Deming, W.E., Out of Crisis, MIT Press, Cambridge, Mass., 1986.

Demsetz, H., "Why Regulate Utilities?" Journal of Law and Economics, 11, April 1968.

Directions for Defense (DD), "Report of the Commission on Roles and Missions of the Armed Forces," U.S. Government Printing Office, 1995.

Eccles, R.G., The Transfer Pricing Problem, Lexington Books, Lexington, MA, 1985.

Fox, J.R., The Defense Management Challenge, Boston: Harvard Business School Press, 1988.

Friend, G.C., "DBOF and the Stabilized Rate-Setting Process," Navy Comptroller, 1996.

Govindarajan, V. and R.N. Anthony, "How Firms Use Costs Data in Price Decisions," Management Accounting, 65/1, 1983.

Hirschey, M. and J.L. Pappas, "Fundamentals of Managerial Economics" (5th ed.), New York: Dryden Press, 1995.

Isosaari, J.M., "Cost Accounting at a DBOF Activity," Navy Comptroller, 1995.

Kovac, E. and H. Troy, "Getting Transfer Prices Right: What Bellcore Did," Harvard Business Review, 1989.

Laffont, J-J, and J. Tirole, A Theory of Incentives in Procurement and Regulation, Cambridge: MIT Press, 1993.

Magee, R., Advanced Management Accounting, New York: Harper & Row, 1986.

Maroni, A.C., "Improving Financial Management in the Department of Defense," Defense Issues, 8/39, 1993.

Michel, P., Cours de Mathematiques pour Economistes, Paris: Economica, 1984.

Niskanen, W.A., Bureaucracy and Public Economics, Fairfax: Edward Elger, 1994.

Orvis, B.R., Hosek, J.R., and M.G. Mattock, "PACER SHARE Productivity and Personnel Management Demonstration," RAND (R-4127-FMP), Santa Monica, CA, 1992.

Pavia, T.M., "Profit Maximizing Cost Allocation for Firms Using Cost-Based Pricing," Management Science, 41/6, 1995.

Riordan, M., "Profit Regulation of Defense Contractors and Prizes for Innovation," Journal of Political Economy, 97, 1989.

Rogerson, W.P., "Economic Incentives and the Defense Procurement Process," Journal of Economic Perspectives, 8/4, 1994.

Rogerson, W.P., "On the Use of Transfer Prices Within DoD," Logistics Management Institute (PA303RD1), Mclean, VA., 1995.

Schick, A., "Micro-Budgetary Adaptations to Fiscal Stress in Industrialized Democracies," Public Administration Review, Jan/Feb, 1988.

Secretary of Defense (SECDEF), Annual Report to the President and the Congress, U.S. Government Printing Office, March 1996.

Stiglitz, J., Economics of the Public Sector, New York: Norton & Co., 1986.

Sweeney, G., "Adoption of Cost-Saving Innovations by a Regulated Firm," American Economic Review, 17, 1981.

Vogelsang, I., and J. Finsinger, "A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms," Bell Journal of Economics, 10, 1979.

Welbourne, T.M., and L.R. Gomez-Mejia, "Gainsharing: A Critical Review and a Future Research Agenda," Journal of Management, 21/3, 1995.

Weltzman, M. L., "The New Soviet Incentive Model," Bell Journal, 1976.

ENDNOTES

[1] See, for example, W. Rogerson's paper entitled "On the Use of Transfer Prices Within DoD: The Case of Repair and Maintenance of Depot-Level Reparables by the Air Force," Logistics Management Institute Paper PA303RD1, March 1995.

[2] Stiglitz [1986] offers the following example. Suppose there is cost-based pricing where the allowed price for the service provided by a government activity is the cost per unit of the service. Suppose further that performance is difficult to measure; and that the government activity attempts to maximize its budget. Assuming demand for the service is inelastic over some range (i.e. that demand is relatively insensitive to price increases), it would pay a monopoly activity "to **increase the degree of inefficiency** until the price (costs per unit of delivered service) were increased to [the budget maximizing price]." (p.173)

[3] Outsourcing involves "using federal funds to pay a private company to do defense work" while privatization "completely transfers [a DoD activity] to the private sector." (SECDEF Annual Report [1996] p.125)

[4] The defining choice of any firm or organization is to "make-or-buy" required intermediate products. The "buy" options include open market purchases or supplier contracts (outsourcing). The "make" option entails "ownership integration" of support services. The make-or-buy decision defines the boundaries of the organization. These management choices rely heavily on the "transactions costs" literature. (Shugart, et. al. [1994]) Transactions costs consist of (1) search and information costs, (2) bargaining and decision costs, and (3) policing and enforcement costs. The choice of whether to purchase inputs in markets, to contract with suppliers, or to internalize the transaction depends on the marginal benefits and costs of each alternative. Relying on outside suppliers or markets involves price searches, quality concerns and security issues. Contracting with specific suppliers involves negotiations and the writing and enforcement of contracts. Meanwhile, a primary challenge of ownership integration is the "principal-agent" problem. Principals (upper-level management) need to align the interests of their agents (lower level managers and workers) with their own. One approach is to create an incentive structure that rewards efforts that help accomplish the principal's goals, but this necessarily includes the (costly) monitoring of performance. This paper models a specific incentive structure designed to motivate agents (support activities) to accomplish the goal of achieving cost savings and efficiencies over time.

[5] While Secretary of Defense Perry agrees that "outsourcing [certain] commercial activities [such as depot maintenance and material supply management] holds promise to streamline DoD support activities and to achieve cost savings," he concludes "the Department must carefully evaluate the extent to which we can achieve efficiencies." (DoD News Release No. 470-95, 8/25/95) Meanwhile, the Commission acknowledges that, where there are "... difficulties in structuring appropriate contracts and establishing meaningful competition" (DD [1995] p.3-4), or where government activities depend on specialized, defense-unique resources (DoD Handbook [1995]), outsourcing may even be counterproductive. Moreover, expanding the DoD's use of contract support also "requires improving its abilities to create and administer those contracts, and to monitor contractor performance."(DD [1995] p.3-3) Another issue that remains is the current legislation that mandates a 60-40 split of support work between DoD and private contractors.

[6] A recent example is provided by the British Government's attempt to regulate its newly privatized electric utilities ("How to Privatize,"

The Economist, 3/11/95). When a support activity cannot be competitively outsourced, and cannot be competitively supplied in-house, it may still be possible to control the activity via price regulations similar to Britain's so-called RPI-x method. (see "Incredible," The Economist, 3/11/95) There, a regulated price is established and allowed to rise by no more than the retail price index (RPI) less some percentage, x. When the price cap's time is up, the regulator sets a new one for the next period. Under this system a regulated activity profits directly from any cost savings, at least until the next review. Britain's regulations essentially allow activities 100% gain-sharing for any cost savings achieved during the period the regulated price is in force. However, when a new cap is set, cost savings are passed on to consumers who receive the full benefits from the price dropping to the new lower cost level. So far the evidence suggests that cost-based price regulation combined with gain-sharing has "delivered lower real prices to consumers." ("Disgusted," The Economist, 3/11/95)

[7] Although if success-sharing is to be implemented in future contracts for the contractor's benefit, then the DoD might consider sharing cost savings in the current contract (i.e. allowing contractors something less than 100% gain-sharing).

[8] Another concern is that, with time-constrained purchases, the "ease" of a purchase begins to take precedence over its cost-effectiveness. Moreover, as the fiscal year comes to a close, instead of pooling savings remaining towards the end of the year at a more aggregate level and evaluating remaining funding priorities globally, each activity has an incentive to use its budget on its own list of priorities, leading to sub-optimal behavior. Increased oversight of activities (e.g. monitoring, auditing, rules and regulations) often occurs in response to problems such as these.

[9] Schick [1988] cites the extreme case of Ireland where the government "maintain[s] year-round financial control,... departments submit monthly spending plans at the start of the year and month-by-month comparisons of actual and projected expenditure during the year... underexpenditure in any month is treated as a saving for the year and normally is not available for spending in a later month." (p.529)

[10] The standard regulatory response also violates the National Performance Review's and Commission on Roles and Mission's goal to empower managers and workers at lower levels by decentralizing authority and encouraging them to become more entrepreneurial: "DoD managers at all levels must be empowered to make sound business decisions based on broad policy guidance, rather than on detailed rules." (DD [1995] p.3-6)

[11] If customers are given the freedom to choose between internal providers, or to "outsource," and/or if customers have the flexibility to spend their funds on a variety of inputs, this freedom and flexibility can encourage further efficiencies in customer purchasing decisions and generate competitive pressure on internal providers to lower costs and improve the product.

[12] This assumption is not as innocent as it sounds. As noted by Pavia [1995]: "[m]ultiproduct firms usually incur fixed costs which are difficult to attribute to the production of a particular output. This is a troublesome issue for firms that use production costs to establish prices. [Whereas] firms may avoid the problem by only looking at directly attributable [variable] costs," (p.1060) this can pose serious difficulties. For example, if cost-based pricing is combined with gain-sharing, costs that are "counted" might be cut at the expense of those not counted, resulting in higher costs overall. A recent study by Rogerson [1995] offers sobering insights into problems that can arise under DBOF when costs are not assigned correctly in multiproduct organizations. Although not specifically addressed here, these problems merit further attention.

[13] Historically, certain DoD support activities operated in so-called "revolving funds" that charged customers for products and services. The U.S. Military has used two primary types of revolving funds, stock funds and industrial funds. Stock funds were used to procure material and to hold inventory for resale to the operating forces, recovering only the cost of the material itself. Industrial funds provided services such as depot maintenance and transportation, recovering overhead costs in addi-

tion to material costs. The DBOF merged into one revolving fund, nine existing stock and industrial funds, along with five additional defense commercial operations or business activities previously funded with direct appropriated funds.

[14] For detailed information on the composition and scope of each business area, see Appendix C of the DBOF Handbook [1995].

[15] Activity-based costing (ABC) breaks down an organization into activities. The principal function of an activity is to convert resources (labor, materials, etc.) into outputs. While activities consume resources, customers consume activities. ABC systems: 1) identify the activities performed to produce outputs; 2) map the usage of organizational resources to these activities; 3) identify the outputs produced; and 4) link the activity costs to the outputs.

[16] Alternatively, OSD could offer price subsidies that lower the price to in-house (DoD) customers. Also, to the extent internal costs exceed commercial rates, but in-house production is required to protect wartime mission capability, OSD could either explicitly subsidize the difference, or act as the final customer. A useful avenue for future research is to examine the role of subsidies in DoD.

[17] Customers determine and justify their anticipated requirements for goods and services they acquire from the DBOF business areas. Resources required by customers to purchase business area products are subsequently identified in budget request documents. Budget documents are developed using projected rates and prices published by the DBOF business areas.

[18] Final approved rate changes are established by the OUSD(C)and recorded in Program Budget Decision (PBD) documents: "For the DBOF business areas, the OUSD(C) reviews and approves all rates and prices developed... " (DBOF Handbook {1995]p.3-12)

[19] The "death spiral" problem results in so-called "pass-through" funding requests to Congress. However, pass-through funding was originally designed only as a onetime correction of the price structure to bring rates back down to a reasonable level. A conjecture is that this problem will become progressively more acute under current DBOF regulations.

[20] The working hypothesis is that the provision of a clear link between a support activity's success in reducing costs (while preserving or increasing quality), and wages and job security, will motivate managers and employees to generate cost savings. The incentive problem can be broken down into two paprts: i.) external incentives provided to the activity to motivate cost savings, and ii.) the internal distribution of those incentives to motivate management and workers. This paper focuses on external incentives and leaves the important question of the distribution of internal incentives as an issue for further study. However, an expert on the latter offers that "[o]nce 'statistical control' is established, serious work [by management] to improve... [the] economy of production can commence." (Deming [1986] p.354) This paper studies two specific external incentives designed to reward the "economy of production": namely gain-sharing and success-sharing.

[21] This unit-cost function, C(Q), is also assumed to be stable over time. This assumption is important because in reality, cost functions are stochastic and subject to exogenous (random) shocks in input prices. (Deming [1986]) If exogenous shocks (unforeseen resource price increases, natural disasters, wars, etc.) create cost increases which camouflage an activity's true cost saving efforts, then appropriate incentives may not be awarded to the activities when they were, in fact, deserved. Thus, if gain-sharing and success-sharing under unit-cost-based transfer pricing is to be successful, it must be possible for financial accounting systems to separate exogenous cost changes (e.g. higher fuel prices) from an activity's endogenous cost savings (e.g. a new inventory policy). Although not specifically addressed here, this more complex problem deserves further attention.

[22] These cost-reducing innovations could be as simple as introducing modifications to the procurement process that allow credit card purchases and encourage the use of commercial specifications, to more involved changes that require rewriting rules & regulations governing travel, personnel actions, and part-time employment. In the model, it makes no difference whether cost savings result from an application

of new technology, improved employee cooperation and information sharing that reduces errors, rework and wasted materials, or from employee education and training.

[23] Orvis, et.al. [1992] offer an interesting review of one DoD gain-sharing experiment conducted at Sacramento's Air Logistics Center called PACER SHARE. Although difficult to evaluate, apparently "Sacramento display[ed] a tendency toward cost savings under PACER SHARE relative to its baseline,. . . " (p..121)

[24] In contrast to the way investment costs are treated in the model, under current DBOF policies, "[c]apital expenditures. . . for new capital assets. . . are financed through depreciation or capital surcharge rates included in prices." (DBOF Handbook [1995]p.3-11)

[25] Although g and s are fixed over time for purposes of the model, an interesting extension would be to examine the case where gain-sharing and success-sharing is not constant over time.

[27] In the case of DBOF, business areas are organizations within the department, and savings in DBOF business areas are savings in the department, and the savings result in increased operating revenue for the war fighters. Therefore, one might assume that public servants already have the motivation to seek cost-savings. However, due to the current price-setting rules in DBOF and from the history of non-profits, our model assumes further cost-savings could be motivated through organizational incentives that offer something similar to a "profit motive." For example, one of the most striking changes taking place in business today is the dramatic reshaping of compensation plans. Incentive pay plans are rapidly spreading from the executive suite to the shop floor. Incentive pay plans set the compensation of workers and top management according to how well the company achieves a number of preset objectives. The most widely employed is profit- (or gain-) sharing, whereby employees receive annual bonuses based on corporate profit performance. More than 30% of US companies employ some form of profit- or gain-sharing, However, the design of an effective and fair incentive pay plan is a daunting challenge. Pay must be closely linked to performance measures that managers and employees can directly influence. Moreover, the marginal impact of each employee's effort must be separated from the influence of others and more general company-wide or economy-wide influences. Each of these considerations is relevant to the current model, and offers opportunities for future research. (also see Hirschey &Pappas [1995]p.339)

Modeling Loss Exchange Ratios as Inverse Gaussian Variates: Implications*

D. H. Olwell

Department Of Mathematical Sciences
U. S.Military Academy
West Point, New York 10996
Phone: (914)938-5987
Fax: (914)938-2409
Email: olwell@euler.usma.edu

ABSTRACT

This paper outlines a method for estimating and comparing the Loss Exchange Ratio (LER) output of computer combat simulations, and develops methods to establish *a priori* the number of simulation runs required to detect a change in the parameters of the simulation of a given size.

The Loss Exchange Ratio (LER) is a widely used and widely accepted summary statistic for a simulation run involving force-on-force combat models. The LER is surprisingly variable - multiple runs of the same scenario produce a large range of LER.

We assert here that these loss exchange ratios are skewed stochastic random variables, and that they are well modeled by the inverse gaussian (IG) distribution. We discuss technical reasons for preferring the inverse gaussian model over other distributions, particularly the log-normal distribution.

Adopting this IG stochastic model allows us to develop explicit statistical methods for estimating the parameters of this distribution, using its known sampling distributions. We also inherit precise statistical tests for hypothesis testing. Finally, we are able to determine *a priori* the number of simulation runs necessary to detect a change in the distribution of a given size. This is a particularly valuable ability, given the increased reliance of the Army on these simulation models to make procurement and doctrinal decisions. We discuss how these simulation tests fit into the larger scheme of procurement and doctrine decisions

We illustrate with data sets from both the JANUS and CASTFOREM simulations. In particular, we find that the use of the IG model allows us to make more powerful conclusions about the data.

In our discussion of the inverse gaussian distribution, we illustrate the versatility of this lesser known member of the exponential family for modeling positive skewed data and provide a primer on its properties.

We conclude that the IG is a good model for describing the variability of LER with useful estimation and testing properties, and recommend its consideration when modeling LER. We sketch two other promising areas for research which follow from the use of this model.

Key Words: *Loss exchange ratios, sample size, inverse gaussian random variables, JANUS, CASTFOREM, simulation, design of experiments*

INTRODUCTION

Consider two systems which are being considered for acquisition. How does one tell if they are worth the cost of acquiring them, or what their benefits are? How does one choose between the two, if resources are constrained to permit the adoption of only one? The question is particularly difficult if the systems are from different battlefield operating systems, say an air defense weapon system and a communications system.

One strategy for comparing these systems is to model their characteristics, and add them to an existing "base case" force model. For example, we may have a force model which represents a battalion task force. We adjust the model to reflect the addition of new, competing systems. These new force models are used in a suite of scenarios which are executed in a combat simulation, say JANUS or CASTFOREM. The results of the simulation with the new force packages are compared to each other and to the base case. Inferences are drawn about their relative merits. These merits, together with the costs of the systems, can form the basis for rational choices using a cost-benefit analysis.

Such comparisons are not limited to system acquisition: doctrinal changes and force structures can also be modeled and compared using this simulation approach.

A related problem asks, what are the specifications which should be required for a new system? One approach is to construct a model which allows varying capability in the new system, and to simulate at various levels of this capability. One then chooses a response from the simulation, and constructs a model of the response as a function of the level of the capability. It is possible to construct response surface models which examine the effects of making mul-

*This research was supported in part by the Army Research Laboratories and the Mathematical Sciences Center of Excellence, USMA. The comments of an anonymous referee about the use of LER as a metric resulted in improvements of the revised paper, and are gratefully acknowledged. Approved for public release; distribution unlimited.

METHODOLOGY:
Simulation

APPLICATION AREA:
Statistics

tiple changes to the base force model simultaneously. These models help decide how much and which capability to buy. They also allow exploration of the interactions between capabilities and the identification of any resulting synergies.

This methodology requires us to select from the outputs of these simulations for comparison. It then requires a statistically valid means of modeling, estimating, and comparing these responses.

One of the conventional summary statistics for a combat simulation run is the loss exchange ratio (LER), which is the ratio of enemy losses to friendly losses. While this statistic suffers from all the difficulties associated with summarizing a very complex battle with one number, it has found wide acceptance in the operations research community, as well as among decision-makers.

We will use the LER as the response variable for the purposes of this discussion. We note that the methodology developed here is general, and can be applied to other skewed, non-negative measures of effectiveness.

We will not examine the issue of whether the simulations validly represent reality – in other words, if loss exchange ratios for repeated "actual" battles follow the same distribution as the simulations. The question is likely unanswerable, in the first place – exact replications being notoriously difficult to obtain from the records of combat. In the second place, understanding the variability of the simulations themselves is a sufficiently ambitious project for the scope of this paper – especially since it is the simulations which are increasingly being used to shape decisions about acquisitions, doctrine, and force structure.

We also note that LER depends on your point of view; one could easily consider the reciprocal of LER as the measure of performance with no logical difficulty. We shall examine that issue, and see that it does not pose any fatal difficulties.

This paper has the following structure. In section 2, we discuss loss exchange variables and possible models, adopting the inverse gaussian model. In section 3, we discuss estimation of LER parameters using the inverse gaussian model. In section 4, we discuss hypothesis testing. In section 5, we discuss sequential testing methods of LER. Next, in section 6, we examine the power of these tests, and propose a simulation method for determining the appropriate number of runs for a simulation. We then look at two different sets of simulation results in section 7. We close with section 8, conclusions and recommendations. A primer on the inverse gaussian distribution is appended as section 9.

Loss Exchange Ratios

The loss exchange ratio is a widely used summary statistic for combat models. It has theoretical underpinnings in the work of Frederick Lanchester, and his deterministic differential equation models of combat.

It is well known that the output of a computer simulation package such as JANUS or CASTFOREM is variable. The exact same scenario can be simulated repeatedly on these models, and different - sometimes strikingly different - outcomes may result. For example, the boxplot in Figure 1 shows the LERs of 80 runs of the same scenario on the same computer using the same simulation package, CASTFOREM. The maximum LER was 4.5, while the minimum was 0.69. The median was 1.5, while the mean was 1.69788.

The variability and skewness in these data argue strongly against using the single summary statistic, average LER. The LER data need to be described not only with a measure of location, but also with measures of its dispersion and shape. For appropriate statistical description and analysis, we require a statistical model. Lacking such a model, we can not compare the outputs of the competing simulations:

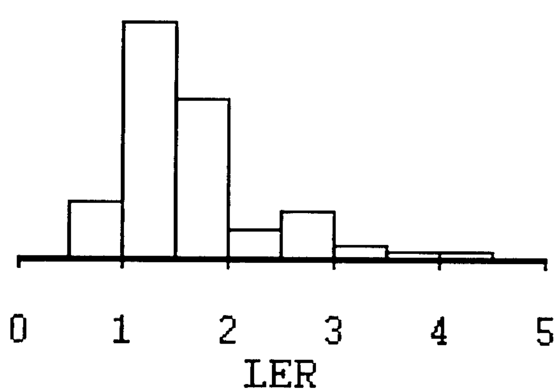


Figure 1. Histogram for loss exchange ratios for 80 simulations of a scenario simulated using CASTFOREM. Data provided by TRAC-WSMR.

we can not determine if differences in response are due merely to chance.

Models

There are several possible models for modeling non-negative skew data. The log-normal, gamma, Weibull, and inverse gaussian distributions immediately suggest themselves.

We desire our model to have several properties. First, the model must fit the data well. Second, the distribution of the maximum likelihood estimators (MLEs) for the model parameters should be known, and tractable. As a minimum, we should be able to find the MLEs without resorting to numerical methods. Third, the theory of estimation and testing for the model should be well developed. Fourth, the parameters of the model should be easily interpretable.

We exclude the gamma and the Weibull distributions for failing to have the second property. The MLEs for these distributions can not be found explicitly, and require numerical approximation. The distribution for the MLEs is not tractable.

The log-normal is a possible model. The distribution of the MLE's is known, and parallels the standard normal distributions. However, there is a real practical difficulty which arises from the logarithmic transformation of the data necessary to conduct statistical testing. For example, statements about the mean of the transformed variable are not statements about the mean of the original variable, but rather the median of the original variable. The mean of the original variable is a function of both the mean and variance of the transformed variable. A direct test for equality of means of the original variable is awkward at best. Similarly, statements about the variance of the original variable are complicated by the fact that it is a function of both the mean and variance of the transformed variables.

In other words, the estimation and hypothesis testing is occurring in the space of the transformed model, not the original data.

If the performance measure of for the model is the ln(LER), then this poses no difficulties. If, however, the decision-maker is interested in the *LER* as the performance measure, then the correct interpretation of the results of the log-normal model is complicated by the transformation back to the original scale.

As a second objection, we will show shortly that the log-normal does not fit our data well.

We prefer a model which fits well and does not require transformation, so that the parameters are immediately useful. As we discuss in the next section, we choose the inverse gaussian distribution.

Why Inverse Gaussian?

The inverse gaussian distribution is a positive skewed distribution with two parameters, μ and λ: μ is the mean of the distribution, and λ is a shape parameter. The MLEs are known and the distributions of the MLEs involve only the inverse gaussian distribution and the Chi-squared distribution. Statistical tests for equality of μ and λ involve only the t-distributions and the F-distributions.

The inverse gaussian distribution fits the data sets we display in this report at least as well as the log-normal. The difference between the two is in the behavior of the right tail, where the log-normal tends to underestimate the quantiles.

For example, Figure 2 is a "QQ" plot of the data set from Figure 1 against the log-normal distribution. Notice that the data set is more heavy tailed to the right than the normal quantiles would suggest. Similarly, Figure 3 shows that the histogram for the transformed data is still skewed to the right. Figure 4 shows the density for the inverse gaussian distribution with MLEs, and this model fits the tails better.

The graphical evidence in Figures 2, 3, and 4 is supported by more formal goodness of fit testing using the Wilks-Shapiro statistic.

Of course, these few data sets hardly constitute conclusive proof that the inverse-gaussian is to be preferred for LER modeling. Such a statement would have to be supported by much more extensive analysis. We merely argue here that the inverse-gaussian distribution is a *plausible* model for LER. We wish to explore the implications of that model.

Accordingly, we make the assumption for the balance of this paper that the LER data set is well modeled by the inverse gaussian distribution, with parameters given by the maximum likelihood estimates.

We mentioned earlier that the LER is dependent on the point of view of the modeler: the inverse LER could as easily be the subject of the model. The distribution of the reciprocal of

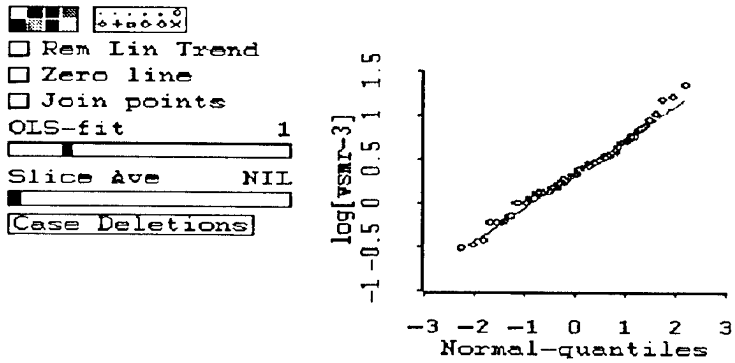


Figure 2. A "QQ" plot of the logarithm of the WSMR data against normal quantiles.

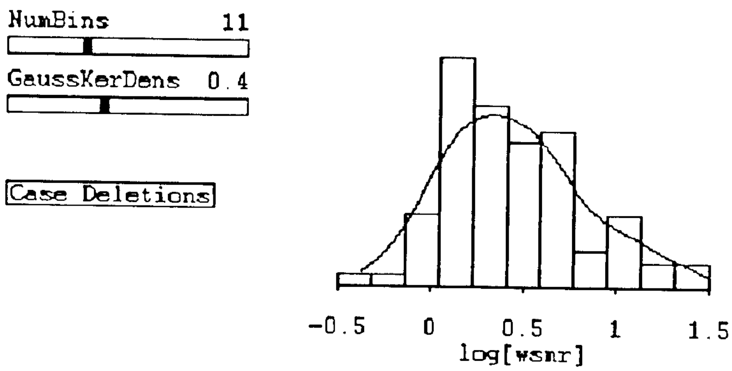


Figure 3. A histogram of the logarithm of the WSMR base data set. A non-parametric smoothing has been applied to the data. Notice that the data set is still skewed to the right, suggesting that the log-normal model may be inappropriate.

an inverse gaussian random variable is known [Chhikara and Folks, 1989] to be a convolution of another inverse gaussian random variable and a Chi-squared random variable with one degree of freedom. In other words, the distribution of the reciprocal of an IG random variable is the sum of a second IG random variable and a Chi-squared random variable. This suggests that the IG model can also be relatively well fit to the reciprocal of LER. As a practical matter, we choose to work with the point of view that results in the heavier right tail. In most models, where Blue outperforms Red, this results in the usual LER. We follow this practice for the balance of the paper.

We also concede that the model sensitivity to point of view is a possible argument against the use of the inverse-gaussian distribution for LER (and against the Weibull and gamma distributions, as well.) We appeal to George Box's dictum:

"All models are wrong, but some are useful." and proceed to demonstrate utility.

Estimation

The MLE estimate of the mean of the IG distribution is the sample average and its distribution is $\hat{\mu} = \bar{X} \sim IG(\mu, n\,\lambda)$. This allows us to construct confidence intervals for the mean of the LER. These confidence intervals are more accurate than ones based on the asymptotic application of the law of large numbers, because the data set is more heavy tailed than the normal distribution. Application of the standard $\bar{X} \pm k\,\hat{\sigma}$ results in an unnecessarily large confidence interval for the mean.

The shape parameter λ has MLE given by

$$\frac{1}{\hat{\lambda}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right) = \frac{1}{n}V \qquad (1)$$

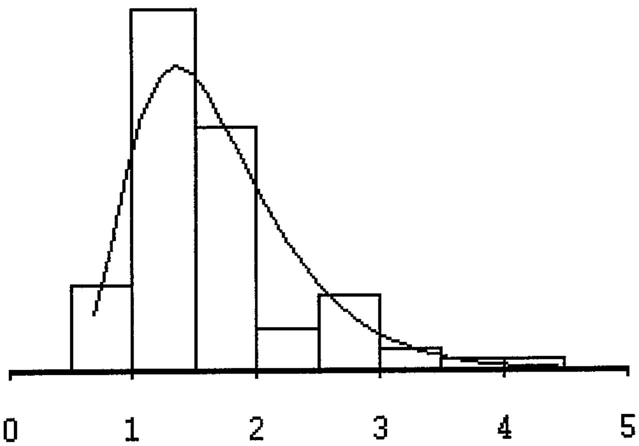


Figure 4. Histogram of the WSMR base data with best fitting IG density.

This estimator is a function of the sufficient statistic $V = \Sigma_{i=1}^{n}(1/X_i - 1/\bar{X})$.

The distribution of $V \sim 1/\lambda\ \chi^2_{n-1}$. This allows confidence intervals to be constructed for λ based on the χ^2 distribution.

For further details, the reader is referred to the primer in the appendix.

The key point is that the distributions of these MLEs involve only the *IG* itself and the χ^2 distributions: they are very tractable. The actual estimates are easily computed.

Estimating the shape parameter seems to be particularly noteworthy, as the skewness and variance of the LER are not routinely reported. The variance of the $IG(\mu, \lambda)$ distribution is given by μ^3/λ, so as the shape parameter increases, the variability decreases.

Closed form expressions for confidence intervals for μ and λ are available in Chhikara and Folks [1989], and again are based on quantiles of standard distributions.

Impact and example

These confidence intervals are narrower than ones based on the asymptotic normal distribution. For example, consider the WSMR base data. A 95% confidence interval, based on a standard normal distribution approximation for the mean which follows from the strong law of large numbers, we obtain

$$\mu \in (1.53895, 1.8568) = (\mu \pm 1.96\sigma/\sqrt{n}) \quad (2)$$

Using the *IG* model and the formulas given in Section 9, we obtain a tighter confidence interval for μ, which also recognizes the skew nature of the data:

$$\mu \in (1.56257, 1.85883) \quad (3)$$

As a result, we have a more precise estimate of the mean, given the available data and the assumed model.

We obtain similar confidence intervals for the λ parameter.

We mention in passing that we are averse to confidence intervals for the mean, preferring instead to assume a Bayesian model with a non-informative prior distribution, which results in probability intervals for the mean. Such Bayesian methods are also outlined in Section 9.

HYPOTHESIS TESTING

The uniform most powerful unbiased tests for the equality of two inverse gaussian population means are known. We consider here the case where neither the mean nor shape parameter is known. References for the other cases are in Section 9.

The rejection region is a function of the sufficient statistics for each sample, $\bar{X}$ and V, and the critical points are given by the t distribution. Details are given in the primer in the appendix.

The uniform most powerful test for the equality of the shape parameters is a function of the sufficient statistics V for each sample, and follows the F distribution. Again, details are in the primer.

These tests allow us to test if the means and shapes of two samples are statistically equivalent. In the context of our problem of comparing the output of two combat simulations, they allow us to test the hypothesis that the outputs came from identical processes.

Moreover, since these tests are based on well fit distributions, they are more powerful than using asymptotically based tests. We see in the examples where these tests allow us to show statistically different results, where the asymptotic methods do not.

The result is that we can make more powerful inference based on the simulations we do run, which saves us computational expense and results in more efficient use of the simulations we do run. For large simulations, this can result in significant economies.

Significance tests also exist for one sided and two sided tests for the mean with λ both known and unknown. Significance tests also are known for the equality of λ with the mean both known and unknown. Additionally, there are two sample versions of the above tests. These cover the usual possibilities, involve only the IG, χ^2, t, and F distributions, and allow simple implementation of exact tests. These tests are outlined in Chhikara and Folks [1989].

Example

Consider the WSMR base data. We wish to test the hypothesis that $\mu=1.5$ against the alternate hypothesis that $\mu \neq 1.5$ The test statistic, from Section 9, follows the t-distribution with $n\text{-}1$ degrees of freedom. We have 79 data points, so our critical value is $t_{crit}=1.99045$ at the 0.05 significance level.

We compute the value of the statistic and obtain:

$$t = \left| \frac{\sqrt{n-1}(\bar{X} - \mu_0)}{\mu_0 \sqrt{\bar{X} V}} \right| = 3.032 > 1.99 \quad (4)$$

We reject the hypothesis that $\mu=1.5$. This accords with the results of our previous section, where 1.5 was not included in our 95% confidence interval for μ, given in Equation 3.

SEQUENTIAL TESTING

It is possible to test if the means of two combat models are equivalent using sequential methods. In these methods, one does not predetermine the number of simulation runs, but rather samples until one can make a decision. The classic method is the sequential probability ratio test.

Wald conjectured [Wald, 1947] and later proved [Wald and Wolfowitz, 1948] that the sequential probability ratio test (SPRT) is optimal for deciding between two point hypotheses in the sense that the expected number of points sampled before a decision could be reached was minimized with the SPRT. A precise statement of these optimality properties of the SPRT in a decision framework can be found in [Ferguson, 1967].

The SPRT considers

$$\Lambda_n = \frac{f(X_1, X_2, \ldots, X_n|\theta_1)}{f(X_1, X_2, \ldots, X_n|\theta_0)} = \prod_{i=1}^{n} \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} \quad (5)$$

where $f(x|\theta)$ is the joint or marginal density as appropriate. The SPRT accepts H_0: $\theta=\theta_0$ if $\Lambda_n \leq A$, accepts H_a:$\theta=\theta_1$ if $\Lambda_n \geq B$ and otherwise continues sampling. This is illustrated in Figure 5, with $A=-3$ and $B=3$, where the null hypothesis would have been rejected at observation number 4.

In practice, we work with the log-likelihood, or $ln(\Lambda_n)$, which results in a cumulative sum. We accept, reject, or continue sampling based on the value of this cumulative sum. As we have written it, the log-likelihood ratio will have a negative expected value when the process is in-control. When the process is well modeled by the alternate hypothesis, the log-likelihood ratio will have a positive expected value. As a result, when the process is in-control, the sum tends downward. When the process is out-of-control at the alternative distribution, the sum tends upward. When the sum is above a certain limit, we have evidence in favor of the alternative hypothesis. When the sum is below a certain limit, we decide in favor of the null hypothesis. When the sum is in-between the limits, we continue to sample.

In the present context, we would apply the SPRT as follows. We would first have our esti-

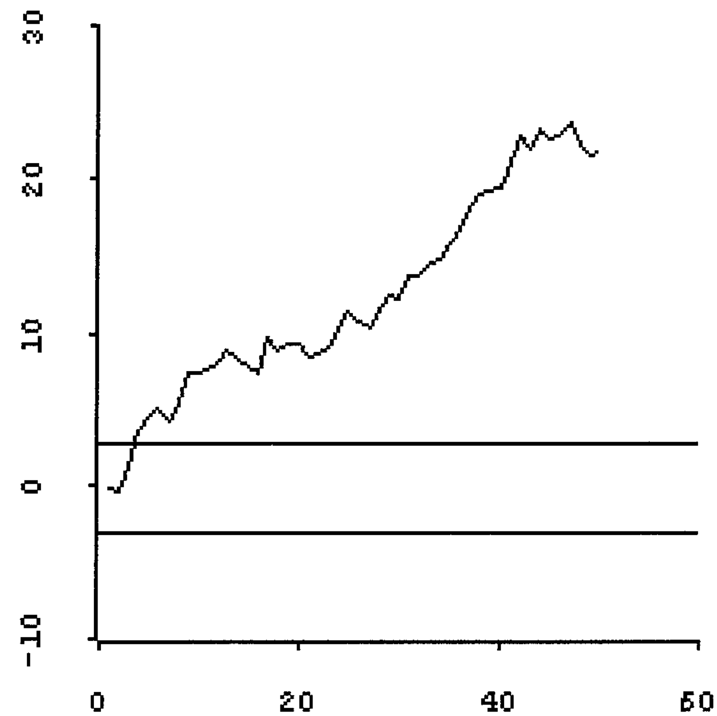


Figure 5. A graphical description of the SPRT.

mate of the base case parameters, which would determine θ_0. We would then select the shift in the parameter for which we desire maximum sensitivity. For example, say our estimate of the mean for the base case was $\mu_0=1.69$. Say further that we wished maximum power to detect if the mean had shifted to $\mu_1=2.00$. We would construct the SPRT with those two point hypotheses, and sample until we reached a conclusion.

The values of the upper and lower limits for the SPRT are set after considering the desired performance of the test in terms of type I (α) and II (β) errors. Exact methods are available, but the usual approximation is to set $A=\alpha/(1-\beta)$ and $B=(1-\alpha)/\beta$.

By using sequential methods, one is guaranteed to reach a decision, and to do so in the fewest average number of simulations. This avoids the situation where one runs, say, the usual thirty trials, fails to reject the null hypothesis, yet doesn't know if 5 more trials would have resulted in the rejection of the null. This approach also avoids the need to do the power calculations discussed next.

NUMBER OF RUNS

To determine the number of simulation runs necessary to detect a difference of parameter of a given size with a given probability, the usual course is to use the power function for the test. The power functions for the inverse gaussian distribution test statistics are not known, however, because the non-central distributions of the test statistics are intractable. In this section, we sketch an approximate method for determining the number of simulation runs necessary.

We assume that we have historical data on the current model, with summary statistics given by $\bar{X}, V_X$, and n_X. This corresponds to the knowledge we would have about the current model after n_X runs.

First, we need to specify two models and error probabilities: the current model, the smallest model change that we wish to detect, the probability of type 1 error (reject the null when it is true) and the probability of type 2 error (accept the null when it is false).

For example, we could identify our current model as represented by the WSMR base case data. We want the probability that we incorrectly say that the model has changed, when it remains constant, to be less than 5%. We desire to be 95% sure that we detect a model shift to $\mu=2.00$, with λ remaining constant. In other words, we want $\alpha=0.05$, $\beta=0.05$. How many trials should be run?

Our setup consists of two samples, one known and one to be drawn. Here the known sample is the WSMR data. We want to know how large the sample should be for the one remaining to be drawn.

Under the null hypothesis that the means are equivalent, the distribution of the test statistic T given by Equation 13 is known to have the t distribution. As a result, we can compute our critical value for the test statistic. For the WSMR data, with its large sample size, we can approximate the critical value by 2.00, regardless of the size of the second sample.

Under the alternate hypothesis, $\mu = 2.00$. We can draw samples of size n repeatedly, compute T, and find the approximate probability that $T<t_{crit}$. This gives us an empirical estimate for β, the probability that we don't detect the model shift to $\mu=2.00$ when it has occurred.

Routines for these simulations are easily implemented. One such LISP implementation is available from the author.

Example

We return to the WSMR base-case data. How many runs do we need to make to be 95% sure to detect a change this large?

We set $n=200$. Of a thousand trials, 977 have a value of T greater than 2.00. We set $n=180$. Then 965 of a thousand trials have a value of $T>2$. We set $n=150$, and find 937 of a thousand trials have a value of $T>2.00$. We could apply a bisection method or a simple interpolation to find that we need to set $n \approx 165$ to achieve our desired design.

We note that these simulations take a few minutes to run on a personal computer, but are much quicker than the corresponding JANUS or CASTFOREM simulations.

We believe that the simulation community is generally unaware of the large number of simulation runs required to have high power for hypothesis tests when the underlying distribution is as variable and skew as the distribution of LER.

EXAMPLES

We present two short examples to support the ideas in this report. The first data set was provided by Mr. Dave Durda of TRAC-WSMR, and is called the WSMR data throughout this paper. The second was provided by Mr. Tom Herbert of the RAND corporation, and is called the RAND data.

WSMR

TRAC-WSMR is responsible for stochastic combat simulation models. One of their models is CASTFOREM. There has been discussion recently about adjusting the way that CASTFOREM assesses damage to systems represented in the model. One proposal was to model degraded states, where instead of a system having a binary state space ("killed" or "not killed"), the system could take on one of several states representing reduced capability.

Three new types of rules were proposed, along with one base case. We call them the base case, and cases one through three. There was interest in whether or not these different rules affected the performance of CASTFOREM, and if so, by how much.

We were provided with the results of 260 simulations of the different rules in CASTFOREM using a standard scenario. The base case and cases two and three were run 80 times each through the standard scenario. The first base case was only run 20 times. We call the base case data set "WSMR", the old rule data set "WSMR1", and the data sets from the two other methods "WSMR2" and "WSMR3".

Boxplots for the LERs from the simulation for each of the models are in Figure 6. We see immediately from the boxplots that the old rules clearly produce different results from the three new rules; the graphical evidence is compelling and sufficient. We move on to the question of whether or not the three new methods produce similar results.

The data sets were each found to be well modeled by the IG distribution.

We compute the confidence intervals for the means of WSMR, WSMR2 and WSMR3, and obtain:

$$\mu_{WSMR} \in (1.5635, 1.858) \qquad (6)$$

$$\mu_{WSMR2} \in (1.410, 1.667) \qquad (7)$$

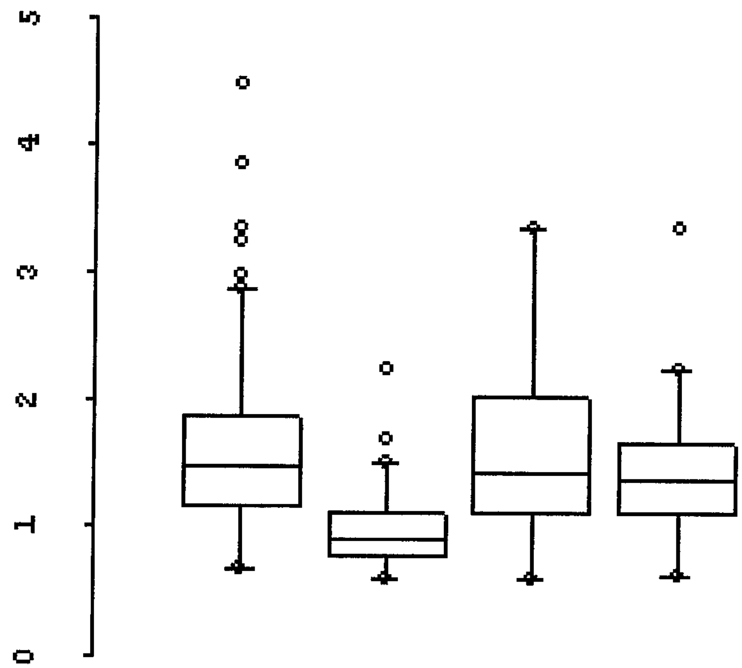


Figure 6. Boxplot of the four WSMR data sets. From left to right, they are the base case, the old binary rules, and two modifications to the base case. Source: White Sands Missile Range, 1996.

$$\mu_{WSMR3} \in (1.3833, 1.6408) \quad (8)$$

It appears from the confidence intervals that the WSMR2 and WSMR3 means are indistinguishable. Can they be distinguished from the base case?

Applying the two sample test developed earlier, we find that WSMR and WSMR2 are not statistically significantly different, as the value of the resulting t statistic is only t=1.7468. The test for equality of means between WSMR and WSMR3, however, has a t statistic value of t=2.071, which is significant at the 0.05 level. We conclude that the WSMR3 set of rules for degraded states has a statistically significantly different impact on LER than the WSMR rules.

We note in conclusion here that if we naively apply the two sample t test which would follow from the inappropriate assumption that WSMR and WSMR3 were normally distributed, or from an asymptotic approximation based on an application of the law of large numbers, we would obtain t=1.62, and **we would not detect the model differences**. Our methods are more powerful than the asymptotic normal approximation.

RAND

We have a second group of data sets, provided by RAND. This data came from trials of the effects of a new weapon system. Three scenarios were run. In the base case, a blue battalion task force attacked a defending red battalion task force. In the second case, the attackers were augmented with a new weapons system. In the third case, the attackers were augmented with two new weapons systems. The simulators sought to demonstrate that the LER was significantly better (from the blue point of view) with the new weapons systems.

RAND conducted thirty runs of each case.

Boxplots for the three cases are presented in Figure 7. From the boxplots, we see again that no formal statistics are necessary to see that the new weapons systems help the blue force. We can obtain confidence intervals for μ and λ to emphasize the point.

We prefer to dwell on a different point: despite the difference in combat simulations between JANUS and CASTFOREM, both produce distributions of LER which are well modeled by the inverse gaussian distribution. We present some graphical evidence in Figures 8, 9

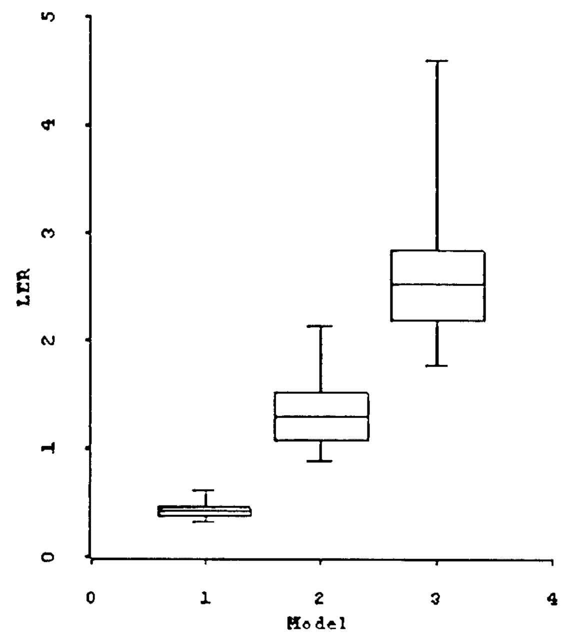


Figure 7. Boxplots for the RAND data. Source: RAND Corporation, 1996.

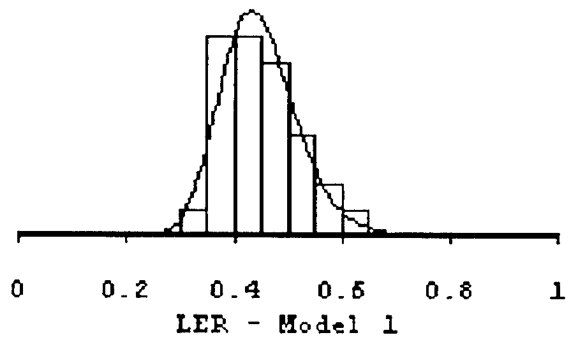


Figure 8. Histogram of the first RAND data set with fitted IG density.

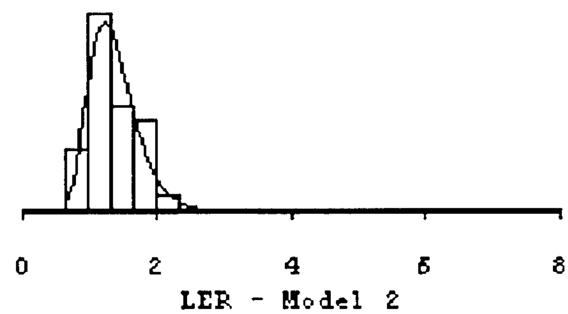


Figure 9. Histogram of the second RAND data set with fitted IG density.

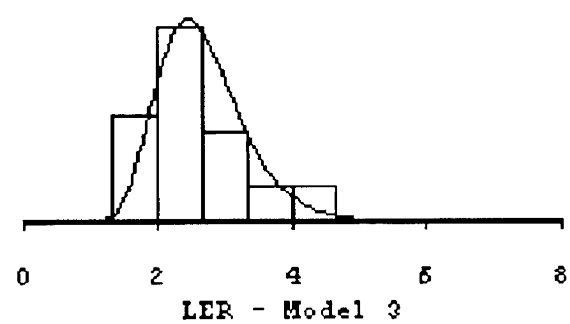


Figure 10. Histogram of the third RAND data set with fitted IG density.

and 10. Formal testing using Wilks-Shapiro and Kolmogorov-Smirnov tests supports this graphical evidence.

CONCLUSIONS AND RECOMMENDATIONS

This is a quick summary of the main points of this paper.

Conclusions

- The inverse gaussian distribution fits well the LER data from JANUS and CASTFOREM presented in this paper.
- The inverse gaussian distribution provides a complete theory of estimation, hypothesis testing, and design of simulation studies for the use of the analyst. This theory is largely based on standard distributions, such as the t, normal, and χ^2 distributions, which are accessible to analysts.
- Methods based on the inverse gaussian distribution are more powerful for analysis of LER problems than methods based on asymptotic normality.
- Using the methods of this paper, it is possible to easily and accurately approximate the number of simulation runs necessary to detect a change in the mean or shape of the distribution of LER results.

Recommendations

The inverse gaussian distribution presents a powerful, tractable model for the distribution of LER from combat simulations. This model allows increased power and precision, and should be considered for use when those properties are desirable.

Further applications of this model should be studied. One promising area is the development of regression models which predict the LER for a given level of JANUS or CASTFOREM parameter associated with some system capability. This could allow the acquisition community to decide on a desired level of capability before setting specifications for systems procurement and design. In particular, regression models based on the inverse gaussian distribution with several predictors seem fruitful for future study.

A second question for study is an expanded examination of the distribution of simulated LER data sets. This paper is a first attempt to find answers to that question.

A SHORT PRIMER ON THE INVERSE GAUSSIAN DISTRIBUTION

This section contains the technical results referenced in the main body, and background on the inverse gaussian distribution for those who may be unfamiliar with it.

The inverse gaussian distribution has its genesis in the analysis of Brownian motion. Following Chhikara and Folks [1989], we characterize the Wiener process $X(t)$ with drift ν and variance σ^2 as follows:

1. $X(t)$ has independent increments; for $t_1 < t_2 < t_3 < t_4$, we have $X(t_2) - X(t_1)$ independent of $X(t_4) - X(t_3)$
2. $X(t_2) - X(t_1)$ is normally distributed with mean $\nu(t_2 - t_1)$ and variance $\sigma^2(t_2 - t_1)$, with $t_2 > t_1$.

Schroedinger [1915] first showed that the distribution of the first time until the process $X(t) > a$ for $\nu > 0$, $a > 0$ was inverse gaussian. See Figure 11 for an illustration.

This characterization of the inverse gaussian is useful for many military applications. Many processes, such as decision times, times to complete a mission, and times to failure, can be well modeled by passage times of Brownian motion with drift.

We have seen how the attrition of forces in a military model can be approximately modeled by the inverse gaussian distribution.

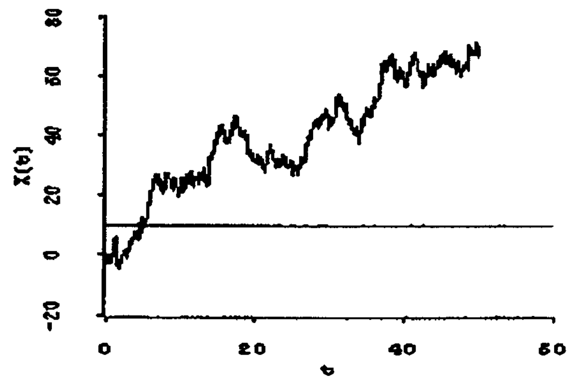


Figure 11. First passage time illustration for Brownian motion with drift.

The inverse gaussian distribution is also useful for modeling of positive, skewed processes in general, even when the underlying mechanics do not immediately suggest a theoretical basis for Brownian motion passage times.

The reader may notice a striking similarity between Figures 5 and 11. Wald [1944, 1945, 1947] showed that the distribution of the stopping times in the SPRT with lower limit $b = -\infty$ and upper limit a was approximately distributed as inverse gaussian. Since the stopping times for the SPRT are discrete random variables and the inverse gaussian is a continuous random variable, the two distributions can only be approximately equal.

Well Known Properties

In this section, we list several well known results about the inverse gaussian distribution, which later prove useful.

PDF

The probability density function (pdf) for $X \sim IG(\mu, \lambda)$ is

$$f(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi}}\, x^{-3/2} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad x > 0 \quad (9)$$

Several density curves for various values of μ and λ can be seen in Figures 12 and 13.

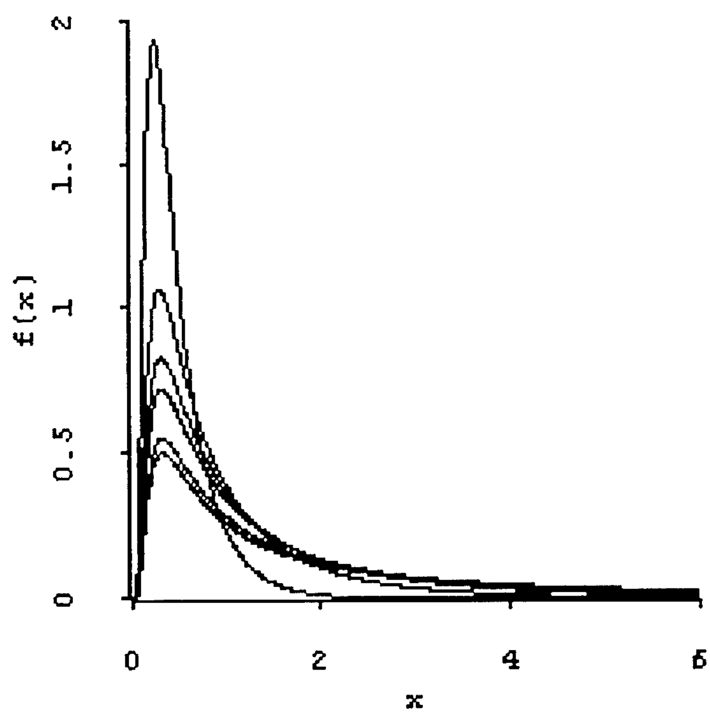


Figure 12. A sheaf of $IG(\mu, 1)$ densities for μ = .5, 1, 1.5, 2, 5, and 10.

CDF

The cumulative distribution function for $X \sim IG(\mu, \lambda)$ is

$$F(x;\mu,\lambda) = \Phi\left[\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right] + \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left[-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right] \quad (10)$$

where $\Phi(x)$ is the CDF of the standard normal distribution. [Chhikara and Folks, 1989]

First Passage Time Interpretation. It is well known [Chhikara and Folks, 1989] that for a Wiener process with positive drift ($\nu > 0$) the first passage time to the barrier is inverse gaussian $IG(\mu, \lambda)$, with $\mu = (a - X(\phi))/\nu$ and $\lambda = (a - X(\phi))^2/\sigma^2$.

Characteristic Function and Moments. The characteristic function for $X \sim IG(\mu, \lambda)$, $\Psi(t) = E(\exp(iXt))$, is

$$\Psi(t) = \exp\left(\frac{\lambda}{\mu}\left[1 - \sqrt{\left(1 - \frac{2i\mu^2 t}{\lambda}\right)}\right]\right).$$

From this, it follows that the mean of X is μ and the variance of X is μ^3/λ.

Member of the Exponential Family. The inverse gaussian distribution is known to belong to the exponential family of order two. Let $\theta = \lambda/\mu^3$. Then the pdf can be expressed as

$$f(x;\lambda,\theta) = \left(\frac{\lambda}{2\pi}\right)^{1/2}\exp(\lambda\theta/2)x^{-3/2} \cdot \exp\left(-\frac{1}{2}\left(\frac{\lambda}{x} + \theta x\right)\right) \quad (11)$$

which is of the natural exponential family

$$c(x)d(\Theta)\exp(a(x)\cdot b(\Theta))$$

with $b(\Theta) = \Theta = (\lambda, \theta)$ and $a(x) = -1/2(x^{-1}, x)$.

Accordingly, for a random sample X from the inverse gaussian, the two dimensional statistic $(\Sigma X, \Sigma(X^{-1}))$ is complete and minimal sufficient.

MLEs of Parameters and Their Distribution. For a random sample $X_1, X_2, \ldots, X_n$ where X_i

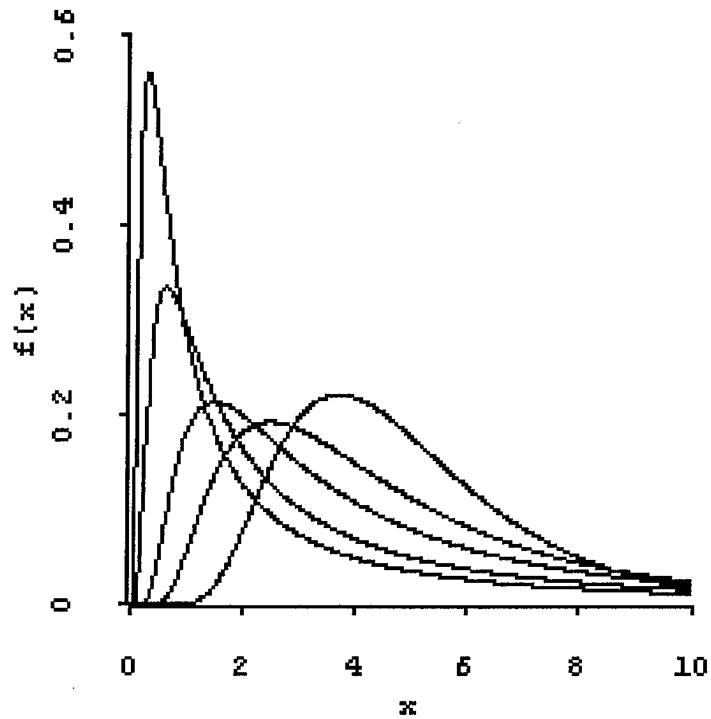


Figure 13. A sheaf of $IG(5, \lambda)$ densities for λ = 1, 2, 5, 10, and 25.

$\sim IG(\mu, \lambda)$, the likelihood function is

$$L(\mu,\lambda)$$

$$i = \left(\frac{\lambda}{2\pi}\right)^{n/2}\left(\prod_1^n x_i^{-3/2}\right)\exp\left(-\lambda \sum_1^n \frac{(x_i-\mu)^2}{2\mu^2 x_i}\right)$$

and the maximum likelihood estimators of μ and λ are easily seen to be

$$\hat{\mu} = \bar{X}$$

and

$$\frac{1}{\hat{\lambda}} = n^{-1}\sum_1^n\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right).$$

These estimators were obtained by Schroedinger [1915].

The distribution of these estimators is also known. Tweedie [Tweedie, 1957a and 1957b] proved the following results.

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed as inverse gaussian with finite first and second moments. Then ΣX_i and $\Sigma X_1^{-1} - n^2(\Sigma X_i)^{-1}$ are independently distributed.

Tweedie showed in his proof that $\bar{X} \sim IG(\mu, n\lambda)$. He defined

$$V = \sum_1^n\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right) \qquad (12)$$

and also proved

$$\lambda V \sim \chi^2_{n-1}$$

Related Distributions. Assume we have two inverse gaussian processes, X and Y, with common but unknown scale parameter λ. Chhikara [1975] derived the uniformly most powerful unbiased (UMP-unbiased) test for the equality of the two process means. We test the hypothesis

$$H_o{:}\mu = \nu \text{ versus } H_1{:}\mu \neq \nu$$

We draw a sample from each population. The rejection region is given by

$$|T| = \left| \frac{\sqrt{n_1 n_2 (n_1 + n_2 - 2)}(\bar{X} - \bar{Y})}{\sqrt{\bar{X}\bar{Y}(n_1\bar{X} + n_2\bar{Y})(V_1 + V_2)}} \right| > t_{1-\alpha/2, n_1+n_2-2} \quad (13)$$

$|T|$ has the folded t distribution with $n_1 + n_2 - 2$ degrees of freedom, a result reported in Chhikara and Folks [1989]. Note that the expression in Equation 6.31 of Chhikara and Folks [1989] is incorrect; the expression in Equation 13 is correct.

Let

$$Y^2 = \frac{(X - \mu)^2}{X\mu^2}$$

Then

$$\lambda Y^2 \sim \chi_1^2$$

that is, λY^2 has the chi-square distribution with one degree of freedom. [Chhikara and Folks, 1989].

Let $X_1, X_2, \ldots, X_{n_X}$ and $Y_1, Y_2, \ldots, Y_{n_Y}$ be independent random samples from $IG(\mu, \lambda)$. Let

$$V_X = \sum_1^{n_X} \left(\frac{1}{X_i} - \frac{1}{\bar{X}} \right) \quad (14)$$

and let V_Y be similarly defined. Then it follows from its form as the ratio of two independent χ^2 random variables and is well known [Chhikara and Folks, 1989] that

$$R = \frac{(n_Y - 1)V_X}{(n_X - 1)V_Y} \sim F_{n_X-1, n_Y-1} \quad (15)$$

where F_{n_X-1, n_Y-1} is the standard F distribution with $n_X - 1$ and $n_Y - 1$ degrees of freedom.

Predictive Distribution for Non-Informative Prior. To obtain natural conjugate priors, it is advantageous to reparameterize the distribution. This poses no logical difficulty in the predictive framework, where the parameters will be integrated out in the process. The parameterization we use sets $\psi = 1/\mu$, and is due to Tweedie.

Lacking data, one would often prefer a non-informative prior distribution. Jeffrey's prior sets $p(\psi, \lambda) \propto \sqrt{|I(\psi, \lambda)|}$. Unfortunately, this prior produces a posterior which is not a proper distribution [Chhikara and Folks, 1989].

If one considers instead a diffuse prior for the parameters, the predictive distribution for the next observation given a series of observations from the inverse gaussian is known and due to Chhikara and Guttman [1982]. The prior used is

$$p(\psi, \lambda) \propto \lambda^{-1},$$

Then let $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ be n independent observations from $IG(\mu, \lambda)$, and Y be an additional observation taken independently of x. For $y > 0$,

$$h(y|\mathbf{x}) = k\left[\frac{\bar{x}\hat{\lambda}}{(n\bar{x} + y)y^3}\right]^{1/2}\left[1 + \frac{(\bar{x} - y)^2\hat{\lambda}}{\bar{x}y(n\bar{x} + y)}\right]^{-n/2}, \quad (16)$$

where

$$k = \frac{S_{t,n}\left((n+1)\sqrt{\dfrac{n}{z(n\bar{x} + y)}}\right)}{\beta\left(\dfrac{1}{2}, \dfrac{n-1}{2}\right) S_{t,n-1}\left(\sqrt{\dfrac{(n-1)}{\nu\bar{x}}}\right)}, \quad (17)$$

$S_{t,n}$ denotes the Student's t cumulative distribution function with n degrees of freedom, and

$$z = n\nu + \frac{n(\bar{x} - y)^2}{\bar{x}y(n\bar{x} + y)}$$

Note that the expression given in Equation 17 for k is correct, and rectifies an existing unnoticed error in both the *Technometrics* article by Chhikara and Guttman [1982] and the Chhikara and Folks monograph [1989].

As an example, consider the following 5 observations from an $IG(\mu, \lambda)$ distribution, with μ and λ unknown: $\{x_1 = 3, x_2 = 4, x_3 = 6, x_4 = 3.5, x_5 = 2.5\}$. Then the graph of the predictive density for the next observation can be seen in Figure 14.

Predictive limits for Y can be obtained by solving the appropriate integral equation using $h(y|\mathbf{x})$. To find the lower predictive limit, $l(\mathbf{x})$, given the data $\mathbf{x}$, for the next observation with probability $\alpha/2$, we solve numerically the integral equation:

$$\int_0^{l(\mathbf{x})} h(y|\mathbf{x})dy = \frac{\alpha}{2} \quad (18)$$

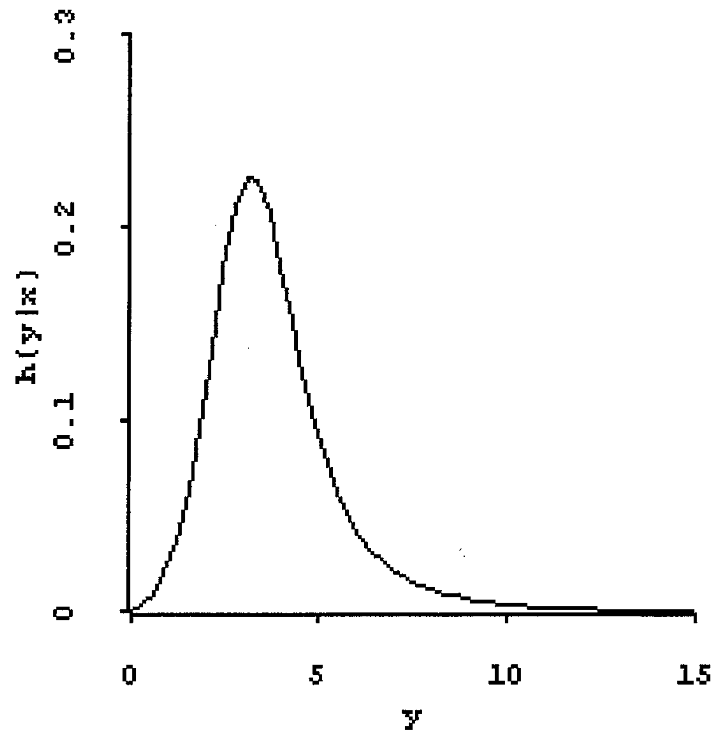


Figure 14. An example of the predictive density for the next observation from an $IG(\mu, \lambda)$, for five previous points; $\{x_1 = 3, x_2 = 4, x_3 = 6, x_4 = 3.5, x_5 = 2.5\}$.

and similarly for the upper predictive limit, $u(x)$.

Olwell [1996] derived the more general joint predictive distribution for the next m observations, given the first n, below. Olwell [1996] also derived tighter predictive limits.

Modeling Advantages Over Other Skewed Distributions. There are three major advantages to using the inverse gaussian distribution to model skewed data. The first is an appeal to the underlying physical properties of the process being modeled. The second discusses the notion of failure rates, and their asymptotic behavior. The third is based on the tractability of the sampling distribution of the inverse gaussian. We discuss each of the three in turn.

If the underlying process can be thought of as a Wiener process, then the use of the inverse gaussian seems especially appropriate. The time to failure of a complex system may depend on the accumulated additive effects of many small perturbations. Compare this with the log-normal distribution, whose application to modeling depends on multiplicative effects, which are often difficult to defend from first principles.

Secondly, consider the failure rate, $r(t)$ of a system as a function of time. We define

$$r(t) = \frac{f(t)}{1 - F(t)},$$

where $r(t)$ denotes the instantaneous rate of failure for a process conditional on its having lasted a certain time. For a Poisson process with time to failure modeled by the exponential distribution with parameter λ, $r(t) = \lambda$; a constant failure rate.

The assumption of constant failure rate is rather strong. Some applications call for a monotonic failure rate: some with an increasing failure rate (IFR); others for a decreasing failure rate (DFR). For these, it is possible to use the Weibull distribution, with density $f(t, \alpha, \beta) = \alpha\beta^{-\alpha} \exp(-(x/\beta)^{\alpha})$, $(x, \alpha, \beta > 0)$. Then $r(t) = \alpha\beta^{-\alpha}x^{\alpha-1}$, and is decreasing for $\alpha < 1$ and increasing for $\alpha > 1$.

In many situations which are characterized by a "burn in" process, it seems appropriate to have an initially increasing then decreasing failure rate [Chhikara and Folks, 1977].

An initially IFR then DFR process is often modeled by the log-normal distribution. For such a process,

$$r(t) = f_{\mu,\sigma}(t)/(1 - \Phi((t - \mu)/\sigma)).$$

This failure rate is non-monotonic; increasing then decreasing asymptotically to zero. For many reliability situations, this asymptotic failure rate of zero seems illogical.

An alternate model uses the inverse gaussian process. Its failure rate is given by:

$$r(t) = \frac{f(t)}{1 - F(t)} \tag{19}$$

$$= \frac{\left(\frac{\lambda}{2\pi t^3}\right)^{1/2} \exp\left(\frac{-\lambda(t-\mu)^2}{2\mu^2 t}\right)}{\Phi\left(\sqrt{\frac{\lambda}{t}}\left(1 - \frac{t}{\mu}\right)\right) - e^{2\lambda/\mu}\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(1 + \frac{t}{\mu}\right)\right)} \tag{20}$$

This failure rate is also non-monotonic, initially increasing then decreasing. However, its asymptotic failure rate is given by

$$r(t) \to \frac{\lambda}{2\mu^2}$$

It is easily shown that the maximum value of $r(t)$ occurs at the value t^* which is the solution to the equation

$$r(t) = \frac{\lambda}{2\mu^2} + \frac{3}{2t} - \frac{\lambda}{2t^2}$$

and the maximum failure rate is $r(t^*)$.

This provides a strong argument for using the inverse gaussian over the log-normal distribution to model lifetimes. It is hard to conceive of physical processes where the failure rate would decrease to zero.

The third argument for using the inverse gaussian is that the sampling distribution of the MLEs of the parameters are known and easy to work with, as above. Using the inverse gaussian avoids the need to transform the data prior to finding MLEs, as is the case with the lognormal distribution.

REFERENCES

Banerjee, Asit K. and G. K. Bhattacharyya (1979) Bayesian Results for the Inverse Gaussian Distribution with an Application. *Technometrics* Vol. 21(2). pp. 247-251.

Bickel, Peter J. and Kjell A. Doksum. (1977) *Mathematical Statistics*. Englewood Cliffs, NJ; Prentice Hall.

Chhikara, R. S. (1975) Optimum tests for the comparison of two inverse Gaussian distribution means. *Australian Journal of Statistics*. Vol. 17. pp. 77-83.

Chhikara, R. S. and J. L. Folks. (1976) Optimum test procedures for the mean of first passage time in Brownian motion with positive drift (inverse Gaussian distribution). *Technometrics*. Vol. 19. pp. 189-193.

Chhikara, R. S. and J. L. Folks. (1977) The Inverse Gaussian Distribution as a Lifetime Model. *Technometrics* Vol. 19. No. 4. pp. 461-468.

Chhikara, R. S. and J. L. Folks. (1989) *The Inverse Gaussian Distribution*. New York: Marcel Dekker.

Chhikara, Raj S. and Irwin Guttman. (1982) Prediction Limits for the Inverse Gaussian Distribution. *Technometrics* Vol. 24. No. 4. pp. 319-324.

Desmond, A. F. and G. R. Chapman. (1993) Modeling Task Completion Data with Inverse Gaussian Mixtures. *Applied Statistics*. Vol. 42. No. 4. pp. 603-613.

Dupuy, Trevor N. (1987) Can We Rely Upon Computer Combat Simulations? *Armed Forces Journal International*. August. pp. 58-63.

Edgeman, Rick L., Robert C. Scott, and Robert J. Pavur. (1988) A modified Kolmogorov-Smirnov Test for the Inverse Gaussian Density with Unknown Parameters. *Communications in Statistics—Simulations*. Vol. 17. No. 4. pp. 1203-1212.

Ferguson, Thomas S. (1967) *Mathematical Statistics: A Decision Theoretic Approach*. New York, Academic Press.

Folks, J. L. and R. S. Chhikara (1978) The Inverse Gaussian Distribution and its

Statistical Application—A Review. *Journal of the Royal Statistical Society B*. Vol. 40. No. 3. pp. 263-289.

Fries, Arthur. (1986) Optimal Design for an Inverse Gaussian Regression Model. *Statistics and Probability Letters*. Vol. 4. pp. 291-294.

Geisser, Seymour (1993) *Predictive Inference: An Introduction* New York: Chapman & Hall.

Helmbold, Robert L. (1990) *Rates of Advance in Historical Land Combat Operations*. Bethesda, Maryland: US Army Concepts Analysis Agency.

IMSL, Inc. (1989) *MATH/LIBRARY: FORTRAN Subroutines for Mathematical Applications*, Edition 1.1. Houston: IMSL, Inc.

Hughes, Wayne P., editor. (1984) *Military Modeling*. Alexandria, VA: Military Operations Research Society.

Lanchester, F. W. (1956) Mathematics in Warfare. In *The World of Mathematics*. Vol. 4. Edited J. R. Newman. New York: Simon and Schuster. pp. 2138-2157.

Olwell, David H. (1996) *Topics in Statistical Process Control* Ann Arbor: University Microfilms.

Savage, I. R. (1962) Surveillance Problems. *Naval Research Logistics Quarterly*. Vol. 9.

Schroedinger, E. (1915) Zur Theorie der fall- und steigversuche an teilchen mit Brownscher bewegung. *Phys. Ze*. Vol. 16. pp. 289-295.

Taylor, H. M. (1965) Statistical Control of a Gaussian Process. *Technometrics*. Vol. 9.

Taylor, James G. (1981) *Force-on-Force Attrition Modeling*. Arlington, VA: Operations Research Society of America.

Taylor, James G. (1983) *Lanchester Models of Warfare*. Volumes I and II. Arlington, VA: Operations Research Society of America.

Tierney, Luke. (1990) *LISP-STAT: An Object Oriented Environment for Statistical Computing and Dynamic Graphics*. New York: Wiley.

Tweedie, M. C. K. (1957a) Statistical properties of inverse Gaussian distributions I. *Annals of Mathematical Statistics*. Vol. 28. pp. 362-377.

Tweedie, M. C. K. (1957b) Statistical properties of inverse Gaussian distributions II. *Annals of Mathematical Statistics*. Vol. 28. pp. 696-705.

Varian, Hal R. (1975) A Bayesian Approach to Real Estate Assessment. In *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage* Eds. Stephen Feinberg and Arnold Zellner. Amsterdam: North-Holland, pp. 195-208.

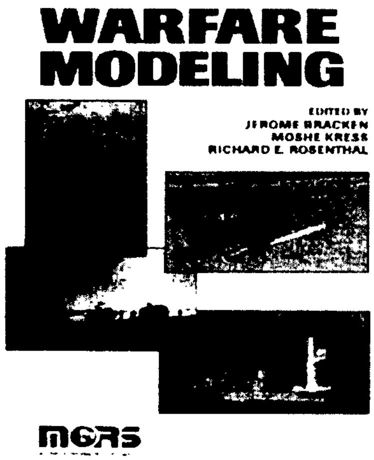
Ventisel, Ve. S. (1964) *Introduction to Operations Research* Moscow: Soviet Radio Publishing House.

Wald, Abraham (1945) Sequential Tests of Statistical Hypotheses. *Annals of Mathematical Statistics*. Vol. 16.

Wald, Abraham. (1947) *Sequential Analysis* New York: Wiley.

Wald, Abraham, and Jacob Wolfowitz. (1948) Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*. Vol. 19. pp. 326-339.

Zollner, Arnold. (1986) Bayesian Estimation and Prediction Using Asymmetric Loss Functions. *Journal of the American Statistical Association*. Vol. 81. No. 394. pp. 446-451.



ORDER FORM

Warfare Modeling

Published by the
MILITARY OPERATIONS RESEARCH SOCIETY
In cooperation with John Wiley & Sons
600 pages, $35.00.

The Military Operations Research Society (MORS) serves the students, developers, practitioners, and clients of military operations research through the ongoing production of high-quality, low-cost meetings, books, and periodicals.

Warfare Modeling was selected for publication by MORS because it contains 25 state-of-the-art chapters, addressing a variety of operations research models that are of central importance in military planning, analysis, and operations. Research on this subject matter seldom appears in the open literature. Most of it appears in reports and documents of government agencies and advisory corporations.

This book is an invaluable reference for military OR professionals as well as for a more general audience of researchers and practitioners. The scope of models included (analytic and simulative, stochastic and deterministic, simple and complex, applied and theoretical, domestic and international) is extensive.

The Book Is Comprised Of Seven Sections:

1. Theater-Level Modeling
2. Mathematical Models of Combat
3. Historical Analysis
4. Weapon System Analysis
5. Command, Control, Communications, and Intelligence
6. Cost Effective Analysis
7. Modeling in Support of Operations Desert Shield and Desert Storm

Warfare Modeling was edited by Jerome Bracken of Yale University, Moshe Kress of Israel's Center for Military Analyses, and Richard E. Rosenthal of the US Naval Postgraduate School. The foreword is by Wayne P. Hughes, Jr., a retired Navy captain and MORS Fellow.

Please Send Me _____ copies of *Warfare Modeling* at $35 per copy

plus $5 for US freight, $8 for Canada, and $10 for all other international shipments per book.

Name ______________________________

Address ______________________________

____

____

City ____________________ State __________ Zip Code ________

VA Residents add $1.58 sales tax per copy

☐ **CASH** ☐ **CHECK** ☐ **Visa** ☐ **MC** ☐ **AMEX** Credit card # ____________________ Expiration date: __________ Billing zip code: __________ Name on Credit Card (print): ____________________ Phone Number: ____________________ Signature: ____________________	TOTAL $

101 South Whiting Street • Suite 200 • Alexandria, VA 22304 • (703) 751-7290 • Fax (703) 751-8171
E-mail morsoffice@aol.com

Clusterization of Alternatives in the Analytic Hierarchy Process

R. Islam, M.P. Biswal and S.S. Alam*

Department of Mathematics, Indian Institute of Technology Kharagpur - 721 302, India

ABSTRACT

The Analytic Hierarchy Process (AHP) attempts to find out the 'best' alternative from several ones by considering a number of conflicting criteria. Despite its potential for solving a wide range of multiple criteria decision making problems, there has been criticisms regarding the labor in filling out pairwise comparison matrices, particularly for large size problems. Even for a medium size problem, AHP requires a large number of pairwise comparisons to form a pairwise comparison matrix. To overcome this difficulty, the present paper experimentally investigates Saaty's suggestion of clustering alternatives. For the experimental purposes, the problem of choosing the best transport aircraft from several ones has been considered.

Key Words: Multicriteria Analysis, Analytic Hierarchy Process, Large-scale Problems.

1. INTRODUCTION

The Analytic Hierarchy Process structures any complex, discrete multi-criteria decision making (MCDM) problem hierarchically. The overall objective of the decision is placed at the top level of the hierarchy and the criteria, subcriteria and the decision alternatives on each descending level. After structuring the hierarchy, pairwise comparisons among the elements belonging to a level with respect to an element belonging to immediately higher level are performed in order to find out their local priority weights. The global weights of the alternatives are obtained by synthesizing the local weights from each level of the hierarchy. The detailed algebraic treatment of AHP is available in Saaty (1990a)

Since its introduction, AHP has been extensively used in private, public, industry, and various governmental sectors. In a review paper, Vargas (1990) has mentioned more than thirty five fields in which AHP has been applied. The other two review papers (Xu, 1986; Zahedi, 1986) are also noteworthy.

AHP has been so popular perhaps for the following reasons :

- its ability to handle inconsistency in judgments,
- its ability to incorporate intangible or non-quantifiable criteria in the decision making process, and
- its ease of use.

Despite its advantages as a powerful MCDM methodology, a major drawback in the use of AHP is the amount of work required to make all the necessary pairwise comparisons. For example, if we have a problem of determining overall weights of 10 alternatives with respect to 5 criteria, then a total of 235 comparisons must be made. In realistic situations, this number may be quite large. Harker (1987a, 1987b) presented two possible remedies to reduce the labor in filling out the matrices. But in course of reducing comparisons, his incomplete pairwise comparison matrix technique, in turn, invokes computations of derivatives which is also arduous for a large size problem. Millet and Harker (1990) extended Harker's (1987b) 'stopping rule' technique by introducing a 'global stopping rule'. Lim and Swenseth (1993) presented an iterative procedure in order to reduce the number of pairwise comparisons. In some sense, this iterative technique is also a 'stopping rule' technique. Weiss and Rao (1987) presented a balanced incomplete block design (BIBD) technique in order to reduce the number of comparisons. But this approach is not practically efficient because of its inherent complexity.

In this connection, Saaty's suggestion (1990b) of clustering alternatives into groups according to a common attribute appears to be more appropriate. But so far no work has been done to verify its applicability. This paper intends to fill up this gap. In Section 2, we have described the idea of clusterization in an algorithmic form. In order to show the applicability of the clustering procedure, we have considered the problem of choice of the best transport aircraft from several ones based upon a number of criteria. The problem has been solved by traditional AHP and by the clustering procedure. The clustering procedure may be called Clustered AHP. The results of the two methodologies are compared and illustrated in Section 3. From the published AHP literature, we have considered five other problems (Arbel, 1983; Saaty, 1979; Saaty and Gholamnezhad, 1982; Sinuany-Stern, 1988; Vachnadze and Markozashvili, 1987) and solved them by the clustering procedure. The results are compared with the actual results in Section 4. Concluding remarks are provided in Section 5.

*Present affiliation: Department of Computer Science and Information Systems, Birla Institute of Technology and Science, Pilani - 333 031, India.

2. Procedure for clustering alternatives

It has been mentioned that for a large number of alternatives, the task to complete all the necessary pairwise comparisons among the alternatives may be quite tedious. The need for reducing the number of pairwise comparisons is two-fold :

- to minimize the labor and time in constructing pairwise comparison matrices, and
- to obtain greater consistency.

In order to reduce the number of comparisons, Saaty (1990b) suggested a clustering technique which is valid when the number of alternatives is quite large (more than seven), say 20, and evaluation scores of alternatives on the criteria are widely dispersed. In the following, the technique is presented in an algorithmic form :

Step 1: Construct the decision hierarchy of inter-related decision elements by breaking down the decision problem at hand.

Step 2: Adopting some (subjective) suitable scale, obtain absolute ranking of all the alternatives with respect to a certain criterion.

Step 3: Make clusters of alternatives having closure magnitudes. Each cluster consists of atmost 's' elements. The smallest element of the largest cluster is included as the largest element in the next cluster and so on.

Step 4: Find the priority weights of all the alternatives belonging to each of the clusters. To pull together and make commensurate of the weights of the alternatives, divide the relative weights of all the elements in the second cluster by the weight of the common element and then multiply by its weight in the first cluster. Repeat the process for the remaining clusters. Steps 2, 3, and 4 are to be repeated for all the criteria.

Step 5: Obtain the global ranking of all the alternatives by applying the usual principle of hierarchical composition.

Theorem 1: In the clustering procedure, the maximum number of pairwise comparisons necessary for a discrete choice problem involving n alternatives is given by

$$[s(n-1)+p(p-s+1)]/2,$$

assuming not more than 's' elements are compared simultaneously and p ($1 \leq p \leq s$-1) is the number of elements in the last cluster.

Proof : Let $\xi_1, \xi_2, \ldots, \xi_m$ be the m clusters obtained by decomposing the set of n alternatives. The mth cluster consists of only p alternatives, where $1 \leq p \leq s$-1. The first cluster contains s elements. Then each of the remaining contains s-1 elements. One alternative from the first cluster is to be added to the second cluster, and one from the second cluster to the third cluster, and so on, thereby making sizes of the pairwise comparison matrices s (except the last one, where $p < n - 1$).

The number of clusters consisting of exactly s - 1 elements is equal to m - 2, which implies

$$s+(s-1)(m-2)+p=n,$$

or,

$$m-2=(n-s-p)/(s-1)$$

Hence, the maximum number of pairwise comparisons necessary is

$$[(n-s-p)/(s-1)+1]\times s(s-1)/2 + (p+1)p/2 = [s(n-1)+p(p-s+1)]/2$$

Corollary 1: The minimum number of pairwise comparisons saved is given by

$$[(n-1)(n-s)-p(p-s+1)]/2$$

where n is the total number of alternatives and the last cluster contains only p ($1 \leq p \leq s$-1) elements.

Proof: By Theorem 1, the maximum number of comparisons necessary is

$$[s(n-1)+p(p-s+1)]/2.$$

Without clustering, the total number of comparisons required for n alternatives is equal to n(n-1)/2. Therefore, the minimum number of comparisons saved is

$$n(n-1)/2-[s(n-1)+p(p-s+1)]/2 = [(n-1)(n-s)-p(p-s+1)]/2$$

Definition: The ratio of the number of direct pairwise comparisons to the total number of comparisons when the alternatives are clusterized, is known as cluster efficiency of the hierarchy.

Corollary 2: The cluster efficiency of a hierarchy is of order n/s.

Proof: The number of direct pairwise comparisons for n alternatives is equal to n(n-1)/2. By Theorem 1, after clusterization, the maximum number of comparisons

$$= [s(n-1) + p(p-s+1)]/2, 1 \leq p \leq s-1$$

Therefore, the cluster efficiency of the hierarchy

$$= [n(n-1)/2]/[s(n-1) + p(p-s+1))/2]$$

$$\geq [n(n-1)/2]/[s(n-1)/2]$$

$$= n/s$$

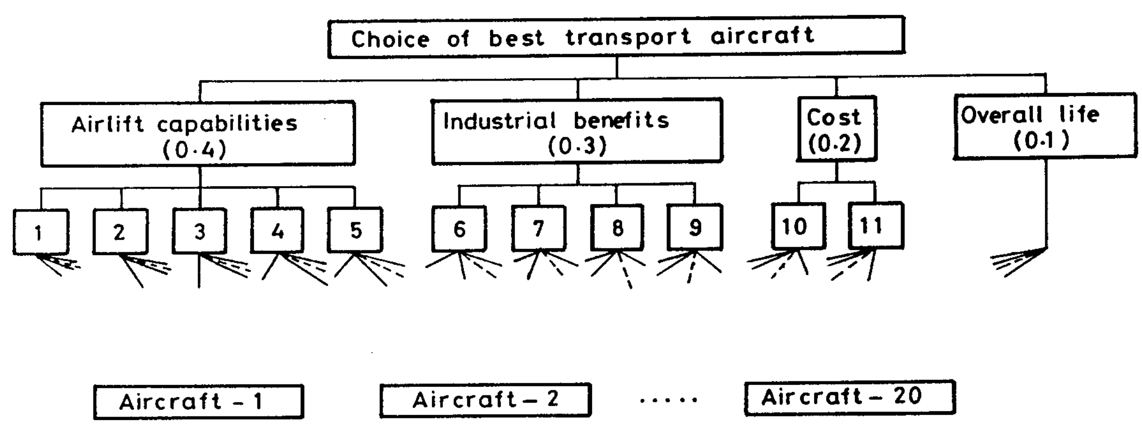
It may be noted that cluster efficiency gives an idea about the magnitude of the number of comparisons to be saved in the clustering procedure. More number of comparisons are saved at the increase of the cluster efficiency.

Remark 1: For the sake of generality of the foregoing decomposition scheme, we have kept the maximum number of elements in a cluster as a variable, namely s, rather than a fixed number. Consequently, the formulas for the reduction in pairwise comparisons are expressed parametrically in terms of the variable s. However, from the application point of view, the reasonable value of s is 7, because one can compare 7±2 elements simultaneously without any confusion (Miller, 1956). Further, in favor of the value 7 for s, Saaty (1977) writes, "... *using the consistency index C the number 7 is a good practical bound on n, a last outpost, as far as consistency is concerned*".

Remark 2: The clustering procedure can also be performed by comparing clusters pairwise. Then local weights of the alternatives belonging to a particular cluster are determined; these weights are multiplied by the weight of the concerned cluster (Saaty, 1990a).

3. Choice of the best transport aircraft

The complex airlift problem has been investigated by many military researchers (Quade, 1978). Presence of multiple conflicting objectives, many decision makers, unavailability of



1. Aircraft requirements (0.25)
2. Aircraft characteristics (0.4)
3. Airbase and support characteristics (0.4)
4. Route structure (0.1)
5. System policies (0.15)
6. Distribution of works among industries (0.1)
7. Creation of jobs (0.15)
8. Transfer of technology (0.6)
9. Regional distribution of industrial activities (0.15)
10. Initial cost (0.7)
11. Operating cost (0.3)

Figure 1. Hierarchy for the transport aircraft choice problem

Table 1. One pairwise comparison matrix for the twenty alternatives

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	9	4	2	4	5	3	2	1	1	3	7	3	2	2	1	1	2	3	6
2		1	1/2	1/5	1/3	1/2	1/3	1/5	1/9	1/5	1/3	1	1/3	1/6	1/6	1/7	1/7	1/4	1/3	1/2
3			1	1/2	1	2	1	1/2	1/4	1/2	1	2	1	1/2	1/2	1/3	1/3	1	1	2
4				1	2	2	2	1	1/2	1	2	4	2	1	1	1/2	1/2	1	2	3
5					1	1	1	1/2	1/4	1/2	1	2	4	1/2	1/2	1/3	1/3	1/2	1	2
6						1	1/2	1/3	1/5	1/3	1/2	2	1/2	1/3	1/3	1/4	1/4	1/2	1/2	1
7							1	1/2	1/4	1/2	1	3	1	1/2	1/2	1/3	1/3	1	1	2
8								1	1/2	1	2	4	2	1	1	1	1	1	2	3
9									1	2	4	8	4	2	2	2	2	3	3	6
10										1	2	5	2	1	1	1/2	1/2	2	2	4
11											1	2	1	1/2	1/2	1/3	1/3	1/2	1	2
12												1	1/2	1/3	1/3	1/6	1/6	1/3	1/3	1
13													1	1/2	1/2	1/3	1/3	1	1	2
14														1	1	1/2	1/2	2	2	4
15															1	1/2	1/2	2	2	4
16																1	1	2	3	5
17																	1	2	3	5
18																		1	2	3
19																			1	2
20																				1

market mechanisms to determine the relationship between a proposed system's cost and its military worth of effectiveness make any defense decision complex (Ng, 1980).

Any defense decision calls for consideration over political, social, and economic life of citizens. Transport aircraft choice is not an exception. Particular choice of transport aircraft

Table 2. Decomposed matrices

	1	9	16	17
1	1	1	1	1
9		1	2	2
16			1	1
17				1

	16	4	5	10	14	15
16	1	2	2	2	2	2
4		1	1	1	1	1
5			1	1	1	1
10				1	1	1
14					1	1
15						1

	4	3	7	11	13	18	19
4	1	2	2	2	2	1	2
3		1	1	1	1	1	1
7			1	1	1	1	1
11				1	1	1/2	1
13					1	1	1
18						1	2
19							1

	19	2	5	6	12	20
19	1	3	1	2	3	2
2		1	1/3	1/2	1	1/2
5			1	1	2	2
6				1	2	1
12					1	1
20						1

Table 3. Priority weights of 20 alternatives by Method 1 and Method 2

	Attribute 1		Attribute 2		Attribute 3		Attribute 4		Attribute 5		Attribute 6	
Alt.	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2
1	0.1013	0.1237	0.0103	0.0085	0.1384	0.1551	0.0341	0.0345	0.0137	0.0085	0.0153	0.0101
2	0.0102	0.0077	0.1423	0.1629	0.0169	0.0145	0.0341	0.0345	0.0331	0.0285	0.1032	0.1010
3	0.0308	0.0283	0.0670	0.0536	0.0711	0.0775	0.0247	0.0295	0.0652	0.0599	0.1443	0.2019
4	0.0554	0.0511	0.0270	0.0239	0.0346	0.0300	0.0341	0.0345	0.0598	0.0599	0.0299	0.0215
5	0.0291	0.0214	0.0670	0.0536	0.1125	0.1030	0.0264	0.0295	0.1004	0.1101	0.0299	0.0215
6	0.0214	0.0159	0.0271	0.0239	0.0273	0.0214	0.0122	0.0126	0.1035	0.1387	0.0100	0.0050
7	0.0324	0.0283	0.0122	0.0101	0.0132	0.0102	0.0354	0.0345	0.1258	0.1748	0.0376	0.0239
8	0.0565	0.0511	0.1040	0.1027	0.0549	0.0515	0.0305	0.0345	0.0747	0.0676	0.0673	0.0642
9	0.1214	0.1740	0.0353	0.0268	0.0840	0.0877	0.0178	0.0181	0.0409	0.0376	0.0147	0.0101
10	0.0627	0.0511	0.1754	0.2587	0.0840	0.0877	0.0763	0.0689	0.0174	0.0128	0.0444	0.0341
11	0.0306	0.0256	0.0288	0.0268	0.0840	0.0877	0.0763	0.0689	0.0137	0.0085	0.0309	0.0194
12	0.0134	0.0094	0.0073	0.0045	0.0234	0.0174	0.0609	0.0689	0.0202	0.0152	0.0969	0.1010
13	0.0317	0.0283	0.0073	0.0045	0.0234	0.0174	0.0939	0.0782	0.0681	0.0599	0.0157	0.0101
14	0.0607	0.0511	0.0782	0.0686	0.0118	0.0079	0.1112	0.1112	0.0197	0.0152	0.0523	0.0384
15	0.0598	0.0511	0.0165	0.0138	0.0118	0.0102	0.0948	0.0782	0.0197	0.0152	0.1000	0.1010
16	0.0948	0.1022	0.0703	0.0602	0.0226	0.0174	0.0535	0.0689	0.0376	0.0266	0.0234	0.0147
17	0.0948	0.1022	0.0349	0.0239	0.0199	0.0159	0.0545	0.0689	0.0196	0.0152	0.0149	0.0101
18	0.0440	0.0391	0.0294	0.0268	0.0753	0.0877	0.0790	0.0689	0.0303	0.0285	0.0102	0.0050
19	0.0320	0.0256	0.0491	0.0403	0.0753	0.0877	0.0201	0.0222	0.0658	0.0599	0.0102	0.0050
20	0.0167	0.0126	0.0106	0.0085	0.0155	0.0112	0.0299	0.0327	0.0658	0.0599	0.1488	0.2019

may have an impact on domestic employment, regional economic activity, technology transfer, etc. So, at the time of choice of the aircraft, the following gross criteria are of utmost importance from the defense point of view :

i) satisfaction of military requirements,
ii) maximization of industrial benefits,
iii) cost, and
iv) overall life of the aircraft.

To be more specific, each of the foregoing criteria can be splitted into several subcriteria. For example, the criterion 'satisfaction of military requirements' can be divided into five subcriteria : a) airlift requirements, b) aircraft characteristics, c) route structure, d) airbase and their support characteristics, and e) system policies. A comprehensive description of the criteria is available in Ng (1980).

Thus, it is clear that the transport aircraft choice problem is characterized by a multiplicity of incommensurable criteria. Fig. 1 depicts the salient factors of the problem in a hierarchical form.

From the Indian perspective, the weights of the criteria and subcriteria are calculated in consultations with a number of Indian Airforce officials. After calculating the weights of the criteria and subcriteria (weights are shown in Fig. 1 itself), the relative ranking of twenty types of transport aircraft with respect to one criterion at a time has been determined. Table 1 shows one comparison matrix for the alternatives constructed for the criterion 'aircraft characteristics'. There are altogether 12 such 20×20 matrices for 12 criteria and subcriteria. The decomposed matrices obtained from the matrix in Table 1 are shown in Table 2. The weights of all the alternatives are calculated by both traditional AHP and the algorithm presented in Section 2. The procedure is repeated for all the criteria. The results are shown in Table 3.

Table 3. Continued

Alt.	Attribute 7		Attribute 8		Attribute 9		Attribute 10		Attribute 11		Attribute 12	
	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2
1	0.1106	0.1172	0.0237	0.0154	0.1022	0.1186	0.0699	0.0719	0.0230	0.0179	0.0470	0.0508
2	0.0090	0.0044	0.0380	0.0308	0.1353	0.1768	0.0586	0.0719	0.0217	0.0090	0.0470	0.0508
3	0.0263	0.0201	0.0164	0.0154	0.0638	0.0593	0.0398	0.0470	0.0395	0.0358	0.0595	0.0562
4	0.0514	0.0485	0.1050	0.1122	0.0545	0.0530	0.0282	0.0260	0.0698	0.0717	0.0503	0.0562
5	0.1516	0.1516	0.0201	0.0154	0.0539	0.0530	0.0390	0.0470	0.0217	0.0090	0.0595	0.0562
6	0.0227	0.0149	0.1050	0.1122	0.0730	0.0767	0.0183	0.0130	0.0395	0.0358	0.0686	0.0648
7	0.0651	0.0622	0.0783	0.0739	0.0631	0.0593	0.0803	0.0719	0.0395	0.0402	0.0470	0.0508
8	0.0130	0.0077	0.0076	0.0044	0.0252	0.0161	0.0183	0.0130	0.0729	0.0841	0.0503	0.0508
9	0.0330	0.0243	0.0380	0.0308	0.0108	0.0063	0.0282	0.0260	0.0328	0.0320	0.0739	0.0746
10	0.0330	0.0243	0.0237	0.0154	0.0312	0.0182	0.0183	0.0103	0.0230	0.0179	0.0866	0.0746
11	0.0810	0.0865	0.0103	0.0081	0.0252	0.0161	0.0183	0.0130	0.0810	0.0841	0.0470	0.0508
12	0.0546	0.0585	0.0783	0.0739	0.0388	0.0297	0.0183	0.0130	0.0698	0.0743	0.0595	0.0562
13	0.0227	0.0149	0.0602	0.0466	0.0257	0.0161	0.1210	0.1310	0.0217	0.0090	0.0115	0.0107
14	0.1306	0.1650	0.1504	0.2245	0.0257	0.0161	0.0897	0.0826	0.0395	0.0358	0.0232	0.0254
15	0.0093	0.0044	0.0390	0.0308	0.0794	0.0884	0.0596	0.0719	0.0328	0.0320	0.0232	0.0254
16	0.0203	0.0126	0.0602	0.0466	0.0116	0.0063	0.0282	0.0260	0.1193	0.1486	0.0686	0.0646
17	0.0130	0.0077	0.0201	0.0154	0.1179	0.1534	0.0744	0.0719	0.0623	0.0608	0.0273	0.0269
18	0.0472	0.0384	0.0103	0.0081	0.0127	0.0068	0.0773	0.0719	0.0810	0.0841	0.0273	0.0269
19	0.0927	0.0970	0.0103	0.0081	0.0275	0.0161	0.0744	0.0719	0.0395	0.0464	0.0269	0.0269
20	0.0130	0.0077	0.1050	0.1122	0.0224	0.0136	0.0398	0.0470	0.0698	0.0717	0.0961	0.1002

Method 1 (M - 1) = Saaty's AHP, Method 2 (M - 2) = Clustered AHP

Remark 3: By a series of consultations with some officials working in the Indian Airforce, the data in the pairwise comparison matrices were considered. The entries in the pairwise comparison matrices depend upon the number of alternatives, i.e., deletion of some alternative(s) from the set of alternatives may alter the strength of preference of an alternative over another. Presumably, the difference of the two preference ratios will not be sufficiently high. For this reason, we have retained the same preference ratio of two alternatives in the decomposed matrices. Clearly the problem can also be solved by taking different judgments in the decomposed matrices. The entries in the lower triangular part of a pairwise comparison matrix are the reciprocals of the corresponding entries in the upper triangular part of the same matrix.

Now the global ranking of all the alternatives is obtained for both the methods. At first we have obtained the ranking for 4 attributes only (in discrete MCDM problems the words 'attributes' and 'criteria' are used interchangeably). Then the process is repeated adding one attribute at a time. Each time Pearson's rank correlation coefficient is calculated. The detailed results are shown in Table 4. It is worth noting that, in all the cases, the rank correlation coefficients are greater than 0.9. For the transport aircraft choice problem, comparing all the alternatives with respect to all the 12 attributes using Saaty's AHP takes 2280 pairwise comparisons, whereas the Clustered AHP takes only 683 comparisons (thereby saving 1597 comparisons).

Following Freund (1992), a statistical test has been performed for the rank correlation coefficients to test the variability of the rank-

Table 4. Global ranking of alternatives for various number of alternatives

	A.N. = 4		A.N. = 5		A.N. = 6		A.N. = 7		A.N. = 8		A.N. = 9		A.N. = 10		A.N.=11		A.N.=12	
Rank	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2
1	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
2	10	10	16	10	10	10	10	10	16	10	10	10	16	10	10	10	10	10
3	2	2	10	16	16	2	16	16	10	16	16	2	10	2	2	2	10	2
4	16	16	2	2	2	16	2	2	4	4	2	16	2	16	16	16	2	16
5	4	4	4	4	4	4	4	4	7	7	4	7	7	7	7	7	4	20
6	8	20	20	20	7	9	9	9	2	2	7	4	4	4	4	20	7	4
7	20	8	7	9	17	1	8	20	20	20	6	6	6	6	20	4	20	7
8	6	6	13	6	1	7	7	1	6	6	5	20	20	20	6	6	5	1
9	7	7	17	7	13	17	20	8	5	9	20	9	5	17	8	8	1	6
10	13	13	6	17	9	20	5	7	8	5	8	1	17	5	5	17	6	17
11	3	3	9	13	8	8	17	6	9	8	17	8	8	1	17	5	8	9
12	5	19	8	8	5	13	1	17	13	1	9	17	13	8	13	1	17	5
13	19	5	1	1	19	19	6	15	19	19	1	5	1	9	3	3	13	8
14	17	18	18	15	20	5	19	19	1	13	13	19	9	13	1	9	3	3
15	18	17	15	18	18	6	13	18	17	17	3	3	3	19	9	13	9	13
16	12	12	19	19	6	18	18	13	3	3	19	18	19	3	19	15	18	19
17	9	15	12	3	3	15	3	3	18	18	18	13	18	18	15	19	19	18
18	15	9	3	12	15	3	12	12	12	12	12	15	12	15	12	12	12	15
19	1	1	5	5	12	12	15	15	15	15	15	12	15	12	18	18	15	12
20	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
R.C.C.	0.9939		0.9699		0.9548		0.9669		0.9879		0.9504		0.9834		0.9849		0.9594	

ings obtained by the two procedures (Saaty's AHP and Clustered AHP). Let

H_0 : the two rankings are significantly different
H_1 : the two rankings are not significantly different

At 1% level of significance, the value of the statistic $z = 2.575$. The computed value of $z = r \times \sqrt{n\text{-}1}$ for various number of attributes are shown in Table 5. From this table, we observe that the computed z-value does not fall within the critical region for any number of attributes. So, we must reject the null hypothesis H_0. Therefore, the two rankings are not significantly different.

4. Five problems from AHP literature

Apart from the aircraft choice problem, we have also considered five other problems from

Table 5. z-value for various number of attributes

No. of attributes	4	5	6	7	8	9	10	11	12
Computed value of z	4.3323	4.2277	4.1618	4.2146	4.3062	4.1427	4.2866	4.2931	4.1819

the published AHP literature. To show the applicability of the clustering procedure, we have solved these problems by clustering procedure and compared the results with the actual studies.

Problem 1: High level nuclear waste management problem

In nuclear industry, how to perform safe disposal of 'high level' waste, some of which remain radio active for hundreds of thousands of years, is a major problem. Because of toxicity and long half-lives, management of such nuclear waste is the most challenging problem in radio-active management. Saaty and Gholamnezhad (1982) made a detailed discussion on this problem. They considered the following five options by which disposal can be done:

a) geological disposal using conventional mining techniques,
b) very deep hole,
c) island disposal,
d) subseabed disposal, and
e) disposal into space.

These five strategies are judged on eight criteria, namely, (i) state of technology, (ii) health, safety, and environmental impacts, (iii) cost, (iv) socio-economic impacts, (v) lead time, (vi) political considerations, (vii) resource consumption, and (viii) aesthetic effects.

The overall weights and ranking obtained by Saaty and Gholamnezhad and the same obtained by the Clustered AHP are shown in Table 6. It may be noticed that both the rankings of alternatives are exactly same.

Problem 2: Ranking of sixteen sports teams

Sinuany-Stern (1988), in his paper, predicted the ranking of 16 soccer teams participating in the Israeli National League. The evaluation was based on six criteria : the facility, the coach, the players, the fans, the previous season's performance, and the current performance. The overall weights and the ranking of the teams obtained by him and those obtained by the Clustered AHP are shown in Table 7.

By applying the Clustered AHP, we have also solved three other problems, viz., "A university budget allocation problem (Arbel, 1983)", "U.S.-OPEC energy conflict (Saaty, 1979)", and "A relay race team formation (Vachnadze and Markozashvili, 1987)". The results are similar to those obtained by the respective authors. Details are omitted due to space limitations.

Table 6. Overall weights and rankings of the alternatives in nuclear waste management problem

	Saaty's AHP			Clustered AHP		
Alternatives	Weights	Rank	P.C. Reqd.	Weights	Rank	P.C. Reqd.
1.Geologic disposal	0.3000	1	162	0.3509	1	99
2. Very deep hole	0.1720	3		0.1459	3	
3. Island disposal	0.1580	4		0.1307	4	
4.Subseabed disposal	0.1390	5		0.1120	5	
5. Space disposal	0.2280	2		0.2590	2	

P.C. Reqd. = Pairwise Comparisons Required.

Table 7. Overall weights and ranking of 16 soccer teams

	Saaty's AHP			Clustered AHP		
Teams	Weights	Rank	P.C. Reqd.	Weights	Rank	P.C. Reqd.
1	0.0507	9	645	0.0446	9	272
2	0.1138	2		0.1304	2	
3	0.0894	5		0.0841	5	
4	0.0640	7		0.0665	7	
5	0.1328	1		0.1495	1	
6	0.1003	3		0.1191	3	
7	0.0664	6		0.0771	6	
8	0.0269	15		0.0206	14	
9	0.0446	10		0.0368	10	
10	0.0245	16		0.0193	15	
11	0.0608	8		0.0571	8	
12	0.0323	13		0.0254	13	
13	0.0956	4		0.0830	4	
14	0.0284	14		0.0175	16	
15	0.0392	11		0.0349	11	
16	0.0357	12		0.0300	12	

5. CONCLUSIONS

When applying AHP to a large-scale discrete choice problem, a large number of pairwise comparisons appear as an intriguing problem. Many suggestions are proposed to reduce the number of comparisons. But none has emerged as particularly fruitful from the application point of view. Saaty's (1990b) proposal of clustering alternatives seems to be a better remedy. But no work has been done to verify its applicability. In this paper, an attempt has been made to fill this gap based upon experiments on a series of real world problems. It has been shown that in the clustering procedure, the number of comparisons required is much less than that required in the unified approach and the rankings that result are sufficiently close to the standard AHP with all the pairwise comparisons. This fact is substantiated by the computed rank correlation coefficients and the performed statistical test.

REFERENCES

Arbel, A. (1983) A university budget allocation problem : a priority based approach, Socio-Economic Planning Science, 17, 181-189.

Freund, J.E. (1992) Mathematical Statistics, Prentice Hall, Englewood Cliffs, N.J., USA.

Harker, P.T. (1987a) Alternative mode of questioning in the analytic hierarchy process, Mathematical Modelling, 9, 353-360.

Harker, P.T. (1987b) Incomplete pairwise comparisons in the analytic hierarchy process, Mathematical Modelling, 9, 837-848.

Lim, K.H. and Swenseth, W.A. (1993) An iterative procedure for reducing problem size in large-scale AHP problems, European Journal of Operational Research, 67, 64-74.

Miller, G.A. (1956) The magical number seven plus or minus two : some limits on our capacity for processing information, The Psychological Review, 63, 81-97.

Millet, I. and Harker, P.T. (1990) Globally effective questioning in the analytic hierarchy process, European Journal of Operational Research, 48, 88-97.

Ng, K.Y. (1980) On the choice of a military transport aircraft fleet via worth assessment procedure : a case study, IEEE Transactions on SMC, SMC -12, 894-899.

Quade, E.S. (1978) "Military Systems" in: J.J. Moder and S.E. Elmagbraby (eds) Handbook of Operations Research: Models and Applications, Von Nostrand, New York, 2, 503-534.

Saaty, T.L. (1977) A scaling method for priorities in the hierarchical structures, Journal of Mathematical Psychology, 15, 234-281.

Saaty, T.L. (1979) The US-OPEC energy conflict: the pay-off matrix by the analytic hierarchy process, International Journal of Game Theory, 8, 225-234.

Saaty, T.L. (1990a) The Analytic Hierarchy Process, RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, PA 15213, USA.

Saaty, T.L. (1990b) How to make a decision: the analytic hierarchy process, European Journal of Operational Research, 48, 9-26.

Saaty, T.L. and Gholamnezhad, H. (1982) High level nuclear waste management problem: analysis of options, Environmental and Planning B, 9, 9-26.

Sinuany-Stern, Z. (1988) Ranking of sports teams via the AHP, Journal of Operational Research Society, 39, 661-667.

Vachnadze, R.G. and Markozashvili, N.I. (1987) Some applications of the analytic hierarchy process, Mathematical Modelling, 9, 185-191.

Vargas, L.G. (1990) An overview of the analytic hierarchy process and its applications, European Journal of Operational Research, 48, 2-8.

Weiss, E.N. and Rao, V.R. (1987) AHP design issues for large-scale systems, Decision Sciences, 18, 43-61.

Xu, S. (1986) A survey of the theory and applications of the AHP, Journal of System Engineering, 1.

Zahedi, F. (1986) The analytic hierarchy process – a survey of the method and its applications, Interfaces, 16, 96-108.

Evaluating the Effectiveness of Shoot-Look-Shoot Tactics in the Presence of Incomplete Damage Information

Yossi Aviv

Graduate School of Business
Columbia University, NY

and

Moshe Kress

Center for Military Analyses,
P.O.Box 2250 (T1),
Haifa Israel[1] and
Department of Industrial Engineering
University of Rhode Island
Kingston, RI

ABSTRACT

The advent of new technologies, and in particular the development of efficient detection sensors, have produced advanced long-range and accurate weapon systems.

One of the major requirements for operating these weapon systems effectively and efficiently is to properly assess the damage that was inflicted upon the targets by previously delivered rounds.

This assessment process is commonly called Bomb Damage Assessment (BDA).

This paper presents a number of shoot-look-shoot (SLS) tactics and evaluates their efficiency in situations where the availability of damage information is not certain. The evaluations are performed with respect to the expected number of kills criterion.

It is shown that certain SLS tactics are superior to others and that a certain simple tactic may be the best choice in terms of fire efficiency and operational convenience.

Key Words: *Damage Assessment, Shoot-Look-Shoot, Fire Efficiency*

INTRODUCTION

The advent of new technologies, and in particular the development of effective detection sensors and efficient fire-control devices, have resulted in advanced long-range and accurate weapon systems. HELLFIRE, MILAN, SLAM and GBU-15 (see, e.g. [7], [9] and [11]) are a few examples which represent the wide range of these types of weapon systems. In general, these systems usually seek to engage clusters of point targets such as tanks, artillery pieces and missile launchers. Once such a cluster has been detected, a single target is then acquired out of this cluster and a round of ammunition is delivered towards it. Such munitions, which are usually referred to as Precision Guided Munitions (PGM), are accurate, lethal and very expensive.

In combat there are usually two assets that are in short supply with respect to these types of weapon systems: one is ammunition and the other is time. Because of the high cost of precision-guided munitions on one hand, and weight limitations on the carrying platform (e.g. helicopters) on the other, the number of munitions available to a particular system, such as AT missiles, may be limited. The shooter may have to carefully consider the effectiveness or utility of each delivered round.

Time is a major factor to consider in situations where certain missions are contingent on others or when the targets are maneuverable and an extended engagement may provide them time to hide. In such cases the time frame allotted for completing a mission may be constrained. Furthermore, staying at the same position for an extended period of time may cause the weapon system to become vulnerable to the opponent's fire. Therefore, the objective in such instances is to complete the mission as fast as possible.

To operate efficiently both in terms of ammunition expenditure and time, one of the important tasks of a shooter is to properly assess the damage that was inflicted on the targets by previously delivered rounds. Such an assessment may save both ammunition and time by preventing the engagement of previously killed targets. Hence, damage assessment is a major issue to consider when evaluating the effectiveness of PGM systems. The importance of damage assessment was notable in the Gulf War ([3], [4], [5], [6]) where costly munitions were wasted because of poor damage assessment that led to incidences of multikill on one hand, and uncompleted missions on the other.

Shoot-Look-Shoot (SLS) is a firing tactic which comprises both fire and damage assessment. In the presence of only a single shooter, this tactic represents a sequential engagement where the shooter may occasionally assess the damage inflicted on a certain target before acquiring and shooting at other targets. A deterministic analysis of some general SLS tactics is given in [1] and [2].

The purpose of this paper is to introduce several SLS tactics for a single shooter. We define effectiveness criteria for these tactics and construct probabilistic models that represent them. These selected SLS tactics are analyzed with respect to these criteria and their relative effectiveness is evaluated using typical scenarios.

It is shown that the performance of a certain variant of the Persistent Shooting tactic (which is discussed in detail in Sec-

1-Correspondence address: Moshe Kress, Center for Military Analyses, 32 Leonard Bodwell Road, Narragansett, RI 02882

APPLICATION AREAS:
Land Warfare, Air Warfare

OR METHODOLOGY:
Stochastic Processes, Dynamic Programming

tion 4) appears to be very close to optimality. According to this tactic, which is called HBPS, once a shooter ceases to engage a certain target it will never return to engage it again. The number of rounds expended at a certain target is bounded by a number that is updated as the engagement progresses. This tactic minimizes the number of instances of fire redirection and it is relatively simple to implement. Therefore it is recommended as a practical and efficient tactic even though it is not optimal in the mathematical sense [10].

In the next section we describe the basic combat situation, introduce notation and define the criteria according to which the SLS tactics are evaluated. In Section 3 the basic model with no damage assessment provisions is discussed. Section 4 describes and evaluates the four variants of the Persistent Shooting Tactic and the operational advantages of this family of shooting strategies are discussed. The performance of these tactics are compared to the results of an optimal strategy in Section 5. Section 6 contains concluding remarks and summary.

2. Description, Notation and Criteria

We consider a situation where a single shooter is engaging a cluster of m homogeneous point-targets. During the firing process we assume that the targets do not fire back at the shooter and therefore the latter is invulnerable. We also assume that the cluster of targets remains unchanged throughout the engagement, that is, no targets disappear from the engagement area and no new target enters into it.

The engagement process comprises three stages: acquisition, firing, and damage assessment. When a specific target has been detected by the weapon system and the conditions are such that immediate shooting towards it is possible, then it is said that the target has been **acquired.** The acquisition stage is usually time consuming and therefore it is a major factor in evaluating the effectiveness of the weapon. A common phenomenon is that of **false acquisition** when either a non-target or a previously killed target is mistakenly acquired. Throughout the paper we will consider false acquisitions of one type only, namely, acquiring killed targets. Once a target has been acquired, a round of ammunition is fired upon it. The result of the shooting can be either a hit or a miss. If the round hits the target it may either kill it or cause no damage. In other words, we assume that the shooting process causes no cumulative damage to targets.

A killed target may appear alive to the shooter if there are not enough signs to indicate otherwise. A killed target which also appears as killed to the shooter is called an **evidently killed (EK) target.** In the third stage of the engagement process the damage that was caused by the shooting is assessed. We assume that a previously killed target that was not EK at the time of kill may not appear to the shooter as killed later on unless it is shot upon again and is "killed" once more. In other words, there is no "record" for previously killed targets and therefore a necessary condition for detecting a kill, following a round of fire, is that the particular round was potentially lethal.

The damage assessment is designed to evaluate the impact of the engagement on one hand, and to minimize the number of incidences of false acquisitions on the other. Damage assessment is performed by inspecting the targets. We assume that there is no "type II" error, that is, a live target will never appear to the shooter as killed. This is a reasonable assumption since a kill indication is usually obtained under very rigid conditions.

In conclusion, following delivery of a round, a target may be in one of the three possible states (i) undamaged, (ii) killed with no signs to that effect (K) and (iii) evidently killed (EK).

Denote the single-round kill probability of a target by p_k and let $p_{s/k}$ denote the conditional probability that a killed target will be recognized as such by the shooter. Thus, $p_s = p_{s/k}\, p_k$ is the probability that a target is evidently killed by a single-shot. The random variables X_n and Y_n represent the number of killed (K) targets and evidently killed (EK) targets, respectively, following the delivery of n rounds. We define T_i as the random variable that counts the number of rounds needed to kill i targets.

The basic criterion according to which the various SLS tactics are evaluated is $E(X_n)$ - the expected number of killed targets following the delivery of n rounds.

However, the models that are presented in this paper may be suitable also for evaluation of other effectiveness measures such as: $M_i(q)$ - the number of munitions needed to obtain a kill level of i targets with probability of at least q; the probability for a minimum number (r) of killed targets by a given number (n) of rounds ($Pr[X_n \geq r]$); and the expected number of rounds needed to obtain damage level r.

In this paper we focus on four SLS tactics: (1) Basic Persistent Shooter (BPS), (2) Fixed Bound Persistent Shooter (FBPS), (3) Dynamic Bound Persistent Shooter (DBPS) and (4) Heuristics for Bounded Persistent Shooter (HBPS). These tactics, which are described in Section 4, are compared to each other and are evaluated against two benchnmarks: (a) the basic tactic that takes into consideration no damage assessment, a tactic that is usually described by Urn Models [8], and (b) The Greedy Shooting (GS) tactic - which is a globally optimal tactic under some reasonable assumptions [10]. We start off, in the next section, by describing the shooting tactic without damage assessment.

3. The Basic Model - No Damage Assessment

We consider a situation where n rounds are randomly and independently delivered towards m targets. The result of each shot is unknown to the shooter and therefore $p_{s/k}=0$. In other words, for each round the shooter randomly selects a target out of the cluster of m targets. His choices are independent of each other and in particular, the decision to engage a certain target does not depend upon its state (killed or alive).

Let $X = \{X_n; n=1,2,3,..\}$ be a stochastic process with a state space $\{0,1,..m\}$. X_n indicates the number of killed targets following the shooting of n rounds. Clearly, this process is a Markov Chain with an initial state $X_0 = 0$. Note that since no provisions for damage assessment are made, the random variable Y, which represents the EK targets, has no meaning here.

The transition probabilities for the Markov Chain are given by

$$P_{ij} = \begin{cases} p_k \dfrac{m-i}{m} & \text{if } j = i+1,\ 0 \leq i < m \\ 1 - p_k \dfrac{m-i}{m} & \text{if } j = i,\ 0 \leq i \leq m \\ 0 & \text{otherwise} \end{cases} \quad (3.1)$$

The probability, $P_n(x)$, that x targets are killed by n rounds satisfies:

$$P_n(x) = P_{n-1}(x) \cdot \left(1 - p_k \frac{m-x}{m}\right) + P_{n-1}(x-1) \cdot p_k \frac{m-x+1}{m}. \quad (3.2)$$

It is easily seen [8] that the expected number of killed targets is:

$$E[X_n] = m\left[1 - \left(1 - \frac{p_k}{m}\right)^n\right]. \quad (3.3)$$

Table 3.1 presents the expected values $E[X_n]$ as a function of the number of rounds n and for the case where the number of targets m=10. The expected number of kills is computed for 3 representative values of the kill probability: p_k = 0.2, 0.5, 0.8 and 1. The effect of multiple kills is best noted for the case where p_k=1. Even though the kill is certain for each round, the expected number of kills, for n=10 rounds, is only about 6.5. That is, there are in this case, on average, 3.5 incidences of multiple kills.

Next consider the number of rounds needed to obtain a certain level of damage.

Let M_i, i=0,..,m-1, be a random variable which represents the number of rounds needed

Table 3.1 Expected Number of Killed Targets—Basic Model m = 10

n	p_k = 0.2	p_k = 0.5	p_k = 0.8	p_k = 1.0
5	0.96	2.26	3.41	4.10
10	1.83	4.01	5.66	6.51
20	3.32	6.42	8.11	8.78
30	4.55	7.85	9.18	9.58
40	5.54	8.72	9.64	9.85
50	6.36	9.23	9.85	9.95

to kill the (i+1)-th target, given that i targets have already been killed. Clearly,

$$E[M_i] = \frac{m}{p_k(m-i)}. \tag{3.4}$$

The expected number, $E[T_i]$, of rounds needed to kill i targets is given by

$$E[T_i] = E[M_0] + E[M_1] + \cdots + E[M_{i-1}], \tag{3.5}$$

or

$$E[T_i] = \frac{m}{p_k}\sum_{j=0}^{i-1}\frac{1}{m-j}. \tag{3.6}$$

Since $E[M_i]$ is monotone increasing in i, it follows that $E[T_i]$ is a convex function of i. The operational interpretation of this property is that the marginal expected number of munitions needed is increasing as the damage level requirement (number of killed targets, i) gets higher.

The simple Urn Model presented in this section represents a situation where the entire engagement is random and no decisions are made throughout it. Such shooting tactics may result in incidences of multiple kills, a phenomenon which may cause waste in both munitions and time. This elementary shooting tactic has been described as a back-drop for the SLS tactics which are discussed in detail in the next sections. These tactics incorporate damage assessment and decision making.

4. The "Persistent Shooter"

4.1 The Basic Persistent Model (BPS). The persistent shooter selects the first target to engage out of a cluster of m targets at random and keeps engaging this target (delivering rounds on it) as long as the target is not evidently killed (EK). Once the target is EK, the shooter selects, at random, a new live target to engage. Evidently, the efficiency of this shooting tactic strongly depends on the conditional probability $p_{s/k}$ that a kill is evident. For example, if $p_{s/k}$ = 0 then the maximum possible number of killed targets, under this tactic, is 1.

As in Section 3, let X_n be a random variable that counts the number of killed targets by n rounds. Define Y_n as the number of EK targets after shooting n rounds. Under the assumption that $p_{s/k}$ is constant throughout the engagement, Y_n is a Markov Chain. Clearly, the number of EK targets Y_n can never exceed the number of killed (K) targets X_n. For the Persistent Shooter tactic, X_n can exceed Y_n by at most one target. Hence

$$X_n - 1 \le Y_n \le X_n$$

It can be easily shown that the random variable Y_n has a truncated binomial distribution. That is,

$$P(Y_n = i) = \begin{cases} \binom{n}{i} p_s^i (1-p_s)^{n-i} & \text{if } n \le m \text{ or} \\ & \text{if } (n > m \text{ and } i \le m) \\ \sum_{j=m}^{n} \binom{n}{j} p_s^j (1-p_s)^{n-j} & \\ & \text{if } n > m \text{ and } i = m, \end{cases} \tag{4.1}$$

where $p_s = p_k p_{s/k}$.

However, the parameter that is most relevant to the shooter is the number of killed targets, X_n. Here, the process $\{X_n\}$ is not a Markov Chain; the transition probability from i killed targets to i+1 kills depends on the status (K or EK) of the i-th killed target. The probability distribution function of X_n is derived directly in the following way.

There are three possible states out of which one can arrive at the situation where X_n=x.

State a: Y_{n-1} = x-1 and X_{n-1} = x-1, in which case the transition to X_n = x is with probability p_k.

State b: Y_{n-1} = x-1 and X_{n-1} = x, in which case the transition is with probability 1.

State c: Y_{n-1} = X_{n-1} = x, in which case the transition is with probability (1-p_k).

Therefore,

$$\begin{aligned} P(X_n = x) &= p_k \cdot P(X_{n-1} = x-1, Y_{n-1} = x-1) \\ &+ P(X_{n-1} = x, Y_{n-1} = x-1) \\ &+ (1-p_k) \cdot P(X_{n-1} = x, Y_{n-1} = x). \end{aligned} \tag{4.2}$$

Equation (4.2) applies for the case where $x \leq m-1$. Since $x = m$ is an absorbing state a slight modification of (4.2) is needed to account for that property. For the sake of brevity, this modification is omitted here.

We can obtain now a recursive formula for the probability distribution of the number of killed targets X_n following the firing of n rounds.

$$P(X_n = x) = \begin{cases} (1-p_k)^n & \text{if } x = 0 \\ (1-p_k) \cdot P(X_{n-1} = x) + pk \cdot P(Y_{n-1} = x-1) & \text{if } 1 \leq x \leq m-1 \\ (1-p_x) \cdot P(X_{n-1} = x) + p_k \cdot [P(Y_{n-1} = m-1) + P(Y_{n-1} = m)] & \text{if } x = m. \end{cases} \quad (4.3)$$

Since the probability distribution function (pdf) of Y_n is known from (4.1), the pdf of X_n in (4.3) can now be easily computed.

In particular, the expected number of killed targets is given by

$$E[X_n] = (1-p_k) \cdot E[X_{n-1}] + p_k \cdot E[Y_{n-1}] + p_k \cdot [1 - P(Y_{n-1} = m)]. \quad (4.4)$$

If $n \leq m$ (in which case $E[Y_{n-1}]=(n-1)p_s$ and $P[Y_{n-1}=m]=0$), then it can be shown that,

$$E[X_n] = np_s + (1-p_{s/k}) \cdot [1-(1-p_k)^n]. \quad (4.5)$$

Table 4.1 shows the expected number of targets killed, in a cluster of m=10 targets, by a persistent shooter. The ranges of kill probabilities and kill-recognition probabilities are $p_k = 0.2$, 0.5 and 0.8 and $p_{s/k} = 0.1$, 0.5 and 0.9, respectively.

Notice the tradeoffs between p_k and $p_{s/k}$. From the practical point of view, the realistic range of p_k in actual combat is 0.2 - 0.8. Kill detection however may have a larger range of probabilities; an advanced sensor may have a kill detection rate as high as 90% while the unassisted eye of a shooter may detect no more than 10% of the killed targets. Under these bounding assumptions, the advantage of high $p_{s/k}$ - low p_k over the reverse is evident as n increases. For example, if n=50 then the expected number of killed targets is 8.41 if $p_k = 0.2$ and $p_{s/k} = 0.9$. This value is only 4.89 if $p_k = 0.8$ and $p_{s/k} = 0.1$.

Table 4.2 shows the number of rounds to be delivered in order to obtain a 90% chance for a specified minimum level of damage. Here, $p_k=0.5$ while $p_{s/k}$ assumes two values 0.5 and 1.0. For example, if the mission's objective damage level is to kill 5 out of the 10 targets in the cluster, then if $p_{s/k}=0.5$, the number of rounds required is $M_5(0.9) = 27$, while if $p_{s/k}=1.0$, then $M_5(0.9) = 14$.

These two values are compared with the number of rounds, $M_5(0.9) = 19$, needed in the Urn Model for this damage level objective and the same kill probability $p_k=0.5$.

From Table 4.2 it can be seen that, for the case discussed here, the choice between the random Urn Model and the Basic Persistent Shooting tactic depends on the required damage level and on the effectiveness of the damage assessment process (measured by $p_{s/k}$); for small values of required damage level the Urn Model is superior even for relatively high capabilities of damage assessment. However, if

Table 4.1 Expected Number of Killed Targets for BPS Tactics m = 10

n	$p_k = 0.2$			$p_k = 0.5$			$p_k = 0.8$		
	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$
5	0.71	0.84	0.97	1.12	1.73	2.35	1.30	2.50	3.70
10	1.00	1.45	1.89	1.40	3.00	4.60	1.70	4.50	7.30
20	1.29	2.49	3.70	1.90	5.49	8.61	2.50	8.20	9.99
30	1.50	3.50	5.48	2.40	7.74	9.89	3.30	9.77	10.00
40	1.70	4.50	7.12	2.90	9.17	10.00	4.10	9.99	10.00
50	1.90	5.48	8.41	3.40	9.77	10.00	4.89	10.00	10.00

Table 4.2 Number of Rounds Required for BPS to Obtain a Certain Damage Level with 90% Confidence

Damage level	$p_{s/k}$ = 0.5	$p_{s/k}$ = 1.0	Urn Model
1	4	4	4
2	11	7	7
3	17	9	11
4	22	12	15
5	27	14	19
6	32	17	25
7	37	19	32
8	42	21	41
9	47	24	56
10	52	26	89

the required damage level is high (9 or 10 targets) then the Persistent Shooter tactic is superior, even for less than moderate damage assessment capabilities. In the latter case the effect of the "random" multi-kills in the Urn Model overrides the effect of the "single target" multi-kill that may occur in a persistent shooting.

4.2 Fixed Bound on the Number of Munitions (FBPS). A notable operational property of the Persistent Shooting tactic is the relatively small number of instances where fire is to be redirected between targets; there are at most m-1 such instances. This property is important in cases where the transition from engaging one target to engaging another is demanding both in terms of time (e.g. acquisition time) and resources (e.g. wasted ammunition in the fire adjustment process). However, as was shown in Table 4.1 above, this tactic may be of little use if the damage assessment effectiveness, $p_{s/k}$, is low, or even moderate. One way to reduce the possibility of wasting ammunition due to poor damage assessment is to limit the number of rounds per target. This may be done by setting an upper bound on the number of rounds shot at each target. The problem of determining this bound is an optimization problem where the tradeoff is between possible high rates of multi-kills - in the case where the bound is high - and terminating the engagement of a certain target prematurely, while it is still alive - in the case where the bound is low.

Let n and v be the number of rounds available and the number of targets that have not been engaged yet, respectively. Define $R_u(v,n)$ as the expected number of targets killed by a persistent shooter with an upper bound u, and denote by $u^*(v,n)$ the optimal upper bound which maximizes $R_u(v,n)$. Clearly, for the case of a single target we have:

$$u^*(1, n) = n; \quad R_u(1, n) = 1 - (1 - p_k)^n. \tag{4.6}$$

Define:

A_i - The target is declared EK right after the i-th shot, i = 1,..,u,

A^c - The target is not EK after firing upon it all of its u rounds, where the upper bound u is a number between 1 and n.

Let $P(K/A^c)$ denote the probability that a target is killed (K) after being shot upon by u rounds, even though it is not evidently killed (EK). It can be shown that:

$$P(A_i) = (1 - p_s)^i p_s, \tag{4.7}$$

$$P(A^c) = (1 - p_s)^u, \tag{4.8}$$

$$P(K|A^c) = 1 - \left(\frac{1 - p_k}{1 - p_s}\right)^u. \tag{4.9}$$

The expected kill $R_u(v,n)$ can now be derived recursively by:

$$R_u(v, n)\text{-}\sum_{i=1}^{\bar{u}}(1 - p)_s^{i-1} p_s[1 + R_u(v - 1, n - i)]$$

$$+ (1 - p_s)^{\bar{u}}\left[1 - \frac{1 - p_k}{1 - p_s}\right)^{\bar{u}} + R_u(v - 1, n - \bar{u})], \tag{4.10}$$

where $\bar{u} = \min(n, u)$

The optimal value of u, denoted by u^*, may be obtained by solving

$$\max_{u=1,\ldots,n} \{R_u(m, n)\} \tag{4.11}$$

The following two properties are fairly obvious:

Property 4.1

$$u^*(v, n) \geq \left[\frac{n}{v}\right], \qquad (4.12)$$

where [x] is the largest integer not greater than x.

Proof. Clearly the upper bound u cannot be smaller than [n/v] since otherwise at least v rounds are redundant and therefore the engagement is not executed in its maximal possible effectiveness. However, this bound may be equal to [n/v] (with a possible "waste" of n - v[n/v] rounds) when, for instance, p_k is high but $p_{s/k}$ is low.

Property 4.2. The FBPS tactic is superior to the BPS tactic.

Proof. This follows directly from the observation that

$$\max_{u=1,\ldots,n} \{R_u(m, n)\} \leq R_n(m, n). \qquad (4.13)$$

Table 4.3 presents optimal upper bounds for the FBPS tactic when both the kill probability p_k and the kill-detection probability $p_{s/k}$ are equal to 0.5. Next to each such bound u we indicate, in parentheses, the expected number of killed targets, $R_u(m,n)$

It can be seen, in the above example, that for large values of n, the optimal upper bound may be strictly larger than the average (# of rounds)/(# of targets). Also notice that when comparing the expected number of kills using FBPS to those using BPS (as presented in Table 4.4 below), an improvement is evident. For example, in the case of n=20 and $p_k = p_{s/k} = 0.5$ these expected values are 5.49 and 7.68 for the BPS and the FBPS tactics, respectively, which represents a 40% improvement. For low values of $p_{s/k}$ the improvement is much higher. For example, if n=40, $p_k = 0.5$ and $p_{s/k} = 0.1$ the expected values are 2.90 and 9.41 for BPS and FBPS, respectively (225% improvement!).

Table 4.3 Optimal Upper Bounds for the FBPS Tactic ($p_k = p_{s/k} = 0.5$)

m / n	1	3	5	10
5	5 (0.97)	2 (2.12)	1 (2.50)	1 (2.50)
10	10 (1.00)	4 (2.79)	2 (3.85)	1 (5.00)
15	15 (1.00)	6 (2.95)	4 (4.57)	2 (6.46)
20	20 (1.00)	9 (2.99)	5 (4.84)	2 (7.68)

Table 4.4, which is similar to Table 4.1, presents the expected values for FBPS.

4.3 Dynamic Bound on the Number of Rounds (DBPS). Suppose now that the upper bound on the number of rounds allocated to a given target may be updated and changed as the shooter proceeds from one target to another. Clearly, this relaxed condition can only improve the performance of the shooter, as compared to the FBPS tactics where this upper bound remains fixed. This improvement is due to the ability to respond to **actual** outcomes in the engagement process and to optimize accordingly.

Determining the upper bound is a multistage problem which is once again solved by means of dynamic programming. As before, let $R^*(v,n)$ denote the expected number of kills under the optimal DBPS tactics when the shooter has n rounds and v targets have not been engaged yet.

The dynamic programming problem is stated as follows

$$R^*(v, n) = \max_{u=1,\ldots,n} \left\{ \sum_{i=1}^{u} (1 - p_s)^{i-1} p_s [1 + R^*(v-1, n-i)] + (1 - p_s)^u \left[1 - \left(\frac{1 - p_k}{1 - p_s} \right)^u + R^*(v-1, n-u) \right] \right\} \qquad (4.14)$$

with $R^*(1,n) = 1 - (1-p_k)^n$.

Table 4.3 presents the optimal (dynamic) upper bounds on the number of rounds to be allocated to the **first** out of v remaining targets. The figures in parenthesis, next to the upper bound values, are the optimal expected number of kills - $R^*(v,n)$. To interpret the figures in Table 4.3 suppose that the shooter has yet to engage n=10 targets with n=20 remaining rounds. To the first of these 10 targets it should allocate at most 3 rounds. Suppose that it delivered these 3 rounds with no EK indication (or a kill was evident only by the third round).

Table 4.4 Expected Number of Killed Targets for FBPS Tactic m = 10

n	$p_k = 0.2$			$p_k = 0.5$			$p_k = 0.8$		
	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$
5	1.00	1.00	1.00	2.50	2.50	2.50	4.00	4.00	4.00
10	2.00	2.00	2.00	5.00	5.00	5.00	8.00	8.00	8.00
20	3.62	3.71	3.91	7.55	7.68	8.71	9.62	9.64	9.99
30	4.93	5.15	5.70	8.80	9.25	9.90	9.93	9.98	10.00
40	5.97	6.41	7.26	9.41	9.78	10.00	9.99	10.00	10.00
50	6.81	7.36	8.47	9.71	9.94	10.00	10.00	10.00	10.00

Table 4.5 Optimal Upper Bounds for DBPS Tactic ($p_k = p_{s/k} = 0.5$)

n	m = 5	m = 6	m = 7	m = 8	m = 9	m = 10
10	3 (3.93)	2 (4.28)	2 (4.50)	2 (4.67)	2 (4.83)	1 (5.00)
11	3 (4.12)	2 (4.52)	2 (4.81)	2 (5.00)	2 (5.17)	2 (5.33)
12	3 (4.28)	3 (4.74)	2 (5.09)	2 (5.32)	2 (5.50)	2 (5.67)
13	3 (4.41)	3 (4.93)	2 (5.33)	2 (5.63)	2 (5.83)	2 (6.00)
14	4 (4.52)	3 (5.10)	3 (5.54)	2 (5.90)	2 (6.15)	2 (6.13)
15	4 (4.62)	3 (5.24)	3 (5.74)	2 (6.13)	2 (6.45)	2 (6.66)
16	5 (4.69)	4 (5.37)	3 (5.91)	3 (6.34)	2 (6.71)	2 (6.98)
17	5 (4.75)	4 (5.47)	3 (6.07)	3 (6.54)	2 (6.93)	2 (7.26)
18	5 (4.80)	4 (5.56)	4 (6.20)	3 (6.73)	3 (7.15)	2 (7.51)
19	6 (4.84)	5 (5.64)	4 (6.32)	3 (6.89)	3 (7.35)	2 (7.74)
20	6 (4.88)	5 (5.70)	4 (6.42)	3 (7.03)	3 (7.54)	3 (7.95)

The shooter is now to engage n=9 targets with n=17 rounds. The number of rounds to be delivered to the first of these 9 targets should not exceed 2 (8th row, 5th column). Suppose that this target has become EK by the first round. There are now 16 rounds to engage 8 targets, so the next target should be engaged by at most 3 rounds, and so on. . .

Table 4.6 shows expected kills for the DBPS tactic.

Note that the improvement of the Fixed Bound PS (FBPS) tactic over the Basic PS (BPS) tactic is much more significant than the improvement of the Dynamic Bound PS (DBPS) tactic over the FBPS tactic. For example, if n=20, m=10 and $p_k = p_{s/k} = 0.5$ then the BPS tactic results in expected kills of 5.49, where FBPS results in 7.68 and DBPS results in 7.95. For values of p_k and $p_{s/k}$ further to the extreme, the difference between FBPS and DBPS is even less notable.

Table 4.6 Expected Number of Killed Targets for DBPS Tactic m = 10

n	$p_k = 0.2$			$p_k = 0.5$			$p_k = 0.8$		
	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$
5	1.00	1.00	1.00	2.50	2.50	2.50	4.00	4.00	4.00
10	2.00	2.00	2.00	5.00	5.00	5.00	8.00	8.00	8.00
20	3.62	3.74	3.93	7.56	7.95	8.80	9.63	9.81	9.99
30	4.94	5.24	5.75	8.84	9.37	9.91	9.94	9.99	10.00
40	6.00	6.49	7.33	9.46	9.85	10.00	9.99	10.00	10.00
50	6.84	7.49	8.53	9.76	9.97	10.00	10.00	10.00	10.00

We can conclude that by imposing upper bounds on the PS tactic, one can considerably improve its effectiveness. However, in many cases it may be impractical to rely on an algorithmic procedure to determine the upper bounds.

Next we present a simple and straightforward heuristic for DBPS which can be used for determining the upper bounds. Using the realistic examples in this section, it is shown that its effectiveness may be almost as good as that of the DBPS tactics.

4.4 Heuristics for Determining Upper Bounds for the PS Tactics (HBPS). A simple rule for determining the upper bound on the number of rounds to be allocated to the next target is to divide the number of rounds available by the number of targets yet to be engaged and take the smallest integer greater than or equal to that number. In other words, when v targets are to be shot upon by n munitions, then the upper bound on the first target to be engaged is $[n/v]^+$, where $[x]^+$ is the smallest integer not smaller than x. This tactic represents an "average" principle in allocating munitions. Hence, after terminating the engagement of a certain target - either by evidently killing it or by expending all the rounds that were allocated to it, the shooter recalculates the updated "average" and sets it as the new upper bound.

The expected number of killed targets under this tactics is given recursively by

$$R_A(v, n) = \sum_{i=1}^{[n/v]^+} (1 - p_s)^{i-1} p_s [1 + R_A(v - 1, n - i)] + (1 - p_s)^{[n/v]^+} \left[1 - \left(\frac{1 - p_k}{1 - p_s} \right)^{[n/v]^+} + R_A(v - 1, n - [n/v]^+) \right] \quad (4.15)$$

Table 4.7 shows the expected kills when the HBPS tactic is applied.

Table 4.8 summarizes the analysis so far. It compares the performance of the four aforementioned Persistent Shooting tactics and the shooting tactic with no damage assessment (Urn Model). The comparison is applied to the case of m=10 targets, n=10, 20, 50 rounds and five representative pairs of p_k and $p_{s/k}$.

From Table 4.8 we can draw the following operational conclusions:

(a) If the number of available rounds (n) is not greater than the number of targets (m), then each round should be directed towards a different target. It can be shown [10] that this tactic is optimal. Evidently, the expected number of killed targets is independent of $p_{s/k}$ and it is equal to np_k.

(b) Compared to the no-damage Urn model and the basic PS tactic, each one of the three bounded PS tactics improve the effectiveness of the engagement considerably.

(c) Although DBPS is mathematically superior to FBPS, and HBPS is just a heuristic which is independent of the probability parameters of the engagement, these three bounded PS tactics perform similarly.

(d) Considering the relatively ease by which HBPS can be implemented, we conclude that it is the most efficient bounding tactic if one chooses to adopt a "persistent" strategy in engaging the targets - a strategy that has several operational advantages in terms of target acquisition and fire control.

Clearly, the HBPS tactic is not optimal. Thus, in view of the above, a practical question that may be asked is: how close is HBPS to an optimal strategy?

5. Comparing the HBPS tactic to an Optimal Tactic.

In [10] it is shown that the *Greedy Shooting* (GS) tactic - a tactic in which each round is shot upon the least previously engaged, non-EK target - is optimal with respect to the expected kills criterion. This tactic however is very impractical in actual combat since it requires excessive control measures. Specifically, a shooter must label the targets and then maintain and update a list of these labels, according to the status of the corresponding targets. Moreover, the GS tactic requires frequent redirection of fire which imposes additional constraint on the shooter in terms of acquiring the targets and setting up the weapon system for the new target. Note that the HBPS tactic requires some effort of target labeling too, however it is considerably less demanding than the optimal GS tactic. In

Table 4.7. Expected Number of Killed Targets for HBPS Tactic m = 10

n	$p_k = 0.2$			$p_k = 0.5$			$p_k = 0.8$		
	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$	$p_{s/k} = 0.1$	$p_{s/k} = 0.5$	$p_{s/k} = 0.9$
5	1.00	1.00	1.00	2.50	2.50	2.50	4.00	4.00	4.00
10	2.00	2.00	2.00	5.00	5.00	5.00	8.00	8.00	8.00
20	3.62	3.74	3.89	7.56	7.92	8.39	9.63	9.79	9.89
30	4.94	5.21	5.53	8.84	9.24	9.47	9.94	9.98	9.99
40	6.00	6.41	6.82	9.46	9.72	9.79	9.99	10.00	10.00
50	6.84	7.34	7.75	9.75	9.89	9.91	10.00	10.00	10.00

Table 4.8. A Comparison of Expected Number of Killed Targets m = 10

n	$p_k/p_{s/k}$	BPS	FBPS	DBPS	HBPS	URN
10	0.2/0.1	1.00	2.00	2.00	2.00	1.83
	0.2/0.9	1.89	2.00	2.00	2.00	1.83
	0.5/0.5	3.00	5.00	5.00	5.00	4.01
	0.8/0.1	1.7	8.00	8.00	8.00	5.66
	0.8/0.9	7.3	8.00	8.00	8.00	5.66
20	0.2/0.1	1.29	3.62	3.62	3.62	3.32
	0.2/0.9	3.70	3.91	3.93	3.89	3.32
	0.5/0.5	5.49	7.68	7.95	7.92	6.42
	0.8/0.1	2.50	9.62	9.63	9.63	8.11
	0.8/0.9	9.99	9.99	9.99	9.89	8.11
50	0.2/0.1	1.90	6.81	6.84	6.84	6.36
	0.2/0.9	8.4	8.47	8.53	7.75	6.36
	0.5/0.5	9.77	9.94	9.97	9.89	9.23
	0.8/0.1	4.89	10.00	10.00	10.00	9.85
	0.8/0.9	10.00	10.00	10.00	10.00	9.85

Table 5.1. A HBPS-GS Comparison of Expected Number of Killed Targets m = 10

n	$p_k/p_{s/k}$	HBPS	GS	n	$p_k/p_{s/k}$	HBPS	GS
	0.2/0.1	1.00	1.00		0.2/0.1	4.94	4.95
	0.2/0.9	1.00	1.00		0.2/0.9	5.53	5.80
5	0.5/0.5	2.5	2.5	30	0.5/0.5	9.24	9.49
	0.8/0.1	4.00	4.00		0.8/0.1	9.94	9.94
	0.8/0.9	4.00	4.00		0.8/0.9	9.99	10.00
	0.2/0.1	2.00	2.00		0.2/0.1	6.00	6.01
	0.2/0.9	2.00	2.00		0.2/0.9	6.82	7.43
10	0.5/0.5	5.00	5.00	40	0.5/0.5	9.72	9.91
	0.8/0.1	8.00	8.00		0.8/0.1	9.99	9.99
	0.8/0.9	8.00	8.00		0.8/0.9	10.00	10.00
	0.2/0.1	3.62	3.63		0.2/0.1	6.84	6.86
	0.2/0.9	3.89	3.94		0.2/0.9	7.75	8.64
20	0.5/0.5	7.92	8.04	50	0.5/0.5	9.89	9.99
	0.8/0.1	9.63	9.63		0.8/0.1	10.00	10.00
	0.8/0.9	9.89	10.00		0.8/0.9	10.00	10.00

particular, if n > m, then there are only m-1 incidents of redirection of fire in the HBPS tactic. This number is at least m-1 for the GS tactic.

Table 5.1 shows the expected killed targets for both HBPS and the optimal GS tactics.

It is evident from Table 4.8 that for the spectrum of real-world cases analyzed here, the HBPS tactic performs, in terms of expected kills, very close to the optimal GS tactic.

6. Summary and Conclusions

As advanced, precise and expensive weapon systems (e.g. Precision Guided Munitions - PGM) are introduced into the battle field, questions of their operational efficiency and optimal utilization become more prevalent. Moreover, as the range of weapon systems increases, the problem of fire control, and in particular the effect of damage assessment capabilities, becomes crucial.

In this paper we presented several Shoot-Look-Shoot tactics that may apply to PGM systems in situations where damage information is not necessarily complete.

The Basic Persistent Shooting (BPS) tactic simply instructs the shooter to stick to a certain target and engage it as long as the target is not evidently killed. Two variations of this tactic (FBPS and DBPS) were introduced and were shown to perform better than the basic one. Out of these two tactics emerges a simple tactic - the HBPS - that can be easily implemented in combat. This tactic appear to perform close to optimality in a range of realistic scenarios.

Recall that a number of simplifying assumptions were made in the construction of the models. Specifically, we assumed no cumulative damage in the shooting process, ignored "Type II" errors (that is, declaring a live target as killed) and required that the event "kill" (K) is necessary, at each round, for the event "evident kill" (EK). A natural extension of the analysis presented in this paper may be obtained by considering situations where these assumptions are relaxed.

REFERENCES

[1] Anderson, L., B., "A Heterogeneous Shoot-Look-Shoot Attrition Process", **IDA paper P-2250**, Institute for Defense Analyses, Alexandria VA 1989.

[2] ———, "Heterogeneous Point Fire and Area Attrition Process that Explicitly Consider Various Types of Munitions and Levels of Coordination", **IDA paper P-2249**, Institute for Defense Analyses, Alexandria VA 1989.

[3] Cohen, E. A., The Air War in the Persian Gulf", **Armed Forces Journal**, V. 130, No. 11, June 1993, pp 10-14.

[4] Dacus, J. R., "Bravo Company Goes to War", **Armor**, Sept-Oct 1991, pp 9-15.

[5] GAO/PEMD-87-22, "Current and Future Capabilities of Anti Tank Weapons, Chapter 2: Weapon Effectiveness", 1987, pp 24-30.

[6] House Armed Services Committee, "Conduct and Lessons of the Persian Gulf War", US Government Printing Office, 1992.

[7] JANES - Weapon Systems 1988-89.

[8] Johnson, N. L., and S. Kotz, ***Urn Models and Their Application***, John Wiley & Sons,1977.

[9] Lee, R. G., Garland-Collins, T. K., Johnson, D. E., Archer, E., Sparks, C., Mess, G. M. and A. W. Mowat, ***Guided Weapons***, Brassey's Defence Publishers, 1988.

[10] Manor, G., and M. Kress, "Optimality of the Greedy Shooting Strategy in the Presence of Incomplete Damage Information", Working Paper, CEMA, 1996.

[11] Rybicki, J. F., "Emerging Technologies and the AirLand Battle", MILTECH, Oct 1990, pp 25-30.

MORS
MILITARY OPERATIONS RESEARCH SOCIETY

Subscriptions, Monographs & Workshop Reports Order Form

SHIP TO:	**Return form to:**
Name: ______	**Military Operations Research Society**
Organization: ______	**101 S. Whiting Street, #202**
Address: ______	**Alexandria, VA 22304-3418**
______	Phone: 703-751-7290
Phone: ______	FAX: 703-751-8171
All orders must include a phone number.	E-mail: morsoffice@aol.com

Please note: **All prices are for US orders only! For INTERNATIONAL prices please call the MORS office.**

PHALANX Bulletin subscription (Published Quarterly)	1 year—$20	2 years—$35		
Military Operations Research (MOR) Journal subscription	1 year—$40	2 years—$75		
SAVE$$** Combined both subscriptions and **SAVE$$ *PHALANX* and *MOR* Journal	**1 year—$55**	**2 years—$100**		

DESCRIPTION		**QTY**	**Freight**	**PRICE**	**TOTAL**
Simulation Technology 1997, Sessions 1,2 & 3	(1988)			$ 10.00	
Human Behavior and Performance as Essential Ingredients in Realistic Modeling of Combat—MORIMOC II, Vol 1 & 2	(1989)			$ 55.00	
MORIMOC III	(1990)			$ 10.00	
MORS/ITEA *Emphasizing the "E" in T&E*	(1991)			$ 10.00	
Simulation Validation Workshop—SIMVAL II	(1992)			$ 10.00	
Emphasizing the Cost in COEA—MORS/SCEA Mini-symposium	(1993)			$ 10.00	
ITEA/MORS *How Much Testing is Enough?*	(1994)			$ 10.00	
Joint Requirements Oversight Council Process Workshop (JROC)	(1994)			$ 10.00	
Simulation Data and Its Management Workshop (SIMDATAM)	(1993 & 1995)			$ 10.00	
Quick Response Analysis Methodology (QRAM)	(1996)			$ 15.00	
MOR Analyst's Handbook, Vol I: *Terrain, Unit Movement, and Environment*			$ 5.00	$ 20.00	
MOR Analyst's Handbook, Vol II, Area 1: *Conventional Weapons Effects (Ground)*			$ 3.00	$ 5.00	
SAVE$$ Combine *Military OR Analyst's Handbook*, Vol I and Vol II, Area 1			$ 5.00	$ 25.00	
Warfare Modeling			$ 5.00	$ 35.00	
Military Modeling for Decision Making, Third Edition			$ 5.00	$40.00	
Sales TAX applicable on BOOKS delivered in Virginia only! Sales tax DOES NOT apply to freight or subscriptions.			**VA Residents**	**4.5% TAX**	
❑ CASH ❑ CHECK ❑ VISA ❑ MC ❑ AMEX			TOTAL ENCLOSED		$

Credit card # ______

Expiration date: ______ Billing zip code: ______

Name on Credit Card (print): ______

Signature: ______

TOTAL $

CALL THE MORS OFFICE FOR CREDIT CARD ORDERS

About Our Authors

MULTIPOLAR NUCLEAR STABILITY: INCENTIVES TO STRIKE AND INCENTIVES TO PREEMPT

Jerome Bracken

Jerry Bracken is a consultant and Adjunct Professor of Operations Research at Yale University. He has contributed to both nuclear and conventional modeling. He recently co-edited (with Moshe Kress and Rick Rosenthal) the book Warfare Modeling published by MORS and John Wiley & Sons, Inc.

GAIN-SHARING, SUCCESS-SHARING AND COST-BASED TRANSFER PRICING: A NEW BUDGETING APPROACH FOR THE DEPARTMENT OF DEFENSE (DoD)

Francois Melese

Francois Melese teaches mangerial economics, budgeting, and decision techniques to DoD managers at the Defense Resources Management Institute (DRMI) located at the Naval Postgraduate School in Monterey. His publications include articles on the defense industry, military medicine, the role of benefits in military compensation, and organizational incentives. He surfs in the summer, skis in the winter, and plays tennis and volleyball when his aging body allows.

MODELING LOSS EXCHANGE RATIOS AS INVERSE GAUSSIAN VARIATES: IMPLICATIONS

D. H. Olwell

LTC Dave Olwell is an assistant professor in the Department of Mathematical Sciences at the United States Military Academy at West Point. A USMA graduate, he also holds a Ph.D. in Statistics from the University of Minnesota. He consults extensively for Army activities, and is finishing his fourth book. After five years in Thayer Hall, his greatest professional ambition is an office with a window.

CLUSTERIZATION OF ALTERNATIVES IN THE ANALYTIC HIERARCHY PROCESS

R. Islam, M.P. Biswal and S.S. Alam

Dr. R. Islam received his M.Sc. degree in Applied Mathematics from Calcutta University in 1988. Then he received his Ph.D. in the same subject from the Indian Institute of Technology, Kharagpur in 1996. Presently, he is engaged in teaching several courses in Computer Science and Information Systems at the Birla Institute of Technology and Science, Pilani, India. His research area includes multiple criteria decision making, genetic algorithms, and fuzzy neural networks.

Dr. M.P. Biswal received his B.Sc. (Hons) and M.Sc. degrees in Mathematics from Utkal University, Bhubaneswar, India in 1979 and 1981 respectively. He received his Ph.D. in Mathematics (Operations Research) from Indian Institute of Technology, Kharagpur, India in 1986–87 and joined in a Postdoctoral fellowship for a year in the same department. Then he joined as a faculty member in Mathematics in 1988. Presently, he is an Assistant Professor and has published about twenty-six research papers in international journals of repute. His research area includes operations research, fuzzy programming, multi-objective optimization, numerical analysis, stochastic modeling and computer software.

Born on June 19th, 1945 at Madina of Arambagh district of West Bengal in India, Syed Alam completed his M.Sc. in Statistics from Patna University and Ph.D. in Statistics from Indian Institute of Technology, Kharagpur. He joined as a Lecturer in the Department of Mathematics of this institute in 1976. He became an Assistant Professor in 1984, an Associate Professor in 1990 and Professor in 1996. Since 1976 he has been engaged in teaching and research in this institute. He has published over forty papers in international journals of repute. He has guided nine students towards a Ph.D. degree. He has also published two books.

EVALUATING THE EFFECTIVENESS OF SHOOT-LOOK-SHOOT TACTICS IN THE PRESENCE OF INCOMPLETE DAMAGE INFORMATION

Yossi Aviv and Moshe Kress

Yossi Aviv is completing his Ph.D. in Management Science and Operations Research at Columbia School of Business. His primary interests are in production and operations management with special interest in logistics and supply-chain management. He has worked as a consultant in defense and industrial related projects. His secondary research interests include military applications of operations research.

Moshe Kress is a senior researcher at the Center for Military Analyses in Israel, heading the Logistics Research and Combat Modelling Branch. He is also an adjunct professor at the Technion—Israel Institute of Technology—where he teaches military operations research. His current research interests are operational logistics, combat modelling and multi-criteria decision making. He will seldom miss a folk-dancing opportunity.

Editorial Policy and Submission of Papers

EDITORIAL POLICY

The title of our journal is *Military Operations Research.* We are interested in publishing articles that describe *operations research* (OR) methodologies used in important *military* applications. We specifically invite papers that are significant military OR applications. Of particular interest are papers that present case studies showing innovative OR applications, apply OR to major policy issues, introduce interesting new problem areas, highlight educational issues, and document the history of military OR. Papers should be readable with a level of mathematics appropriate for a master's program in OR.

All submissions must include a statement of the major contribution. For applications articles, authors are requested to submit a **letter** to the editor—exerpts to be published with the paper—from a **senior decision-maker** (government or industry) stating the benefits received from the analysis described in the paper.

To facilitate the review process, authors are requested to categorize their articles by **application area** and **OR method**, as described in Table 1. Additional categories may be added. (We use the MORS working groups as our applications areas and our list of methodologies are those typically taught in most OR graduate programs.)

INSTRUCTIONS TO *MILITARY OPERATIONS RESEARCH* AUTHORS

The purpose of the "instructions to *Military Operations Research* authors" is to expedite the review and publication process. If you have any questions, please contact Ms. Corrina Ross, MORS Communications Manager (email: morscm@aol.com).

General

Authors should submit their manuscripts (3 copies) to:

Dr. Gregory S. Parnell, Editor, *Military Operations Research*
Military Operations Research Society
101 South Whiting Street, Suite 202
Alexandria, VA 22304

Per the editorial policy, please provide:

- authors statement of contribution (briefly describe the major contribution of the article)
- letter from senior decision-maker (application articles only)
- military OR application area(s)
- OR methodology (ies)

TABLE 1: APPLICATION AREAS & OR METHODS

Composite Group	APPLICATION AREA	OR METHODOLOGY
I. STRATEGIC & DEFENSE	Strategic Operations Nuclear Biological Chemical Defense Arms Control & Proliferation Air & Missile Defense	**Deterministic Operations Research** Dynamic Programming Inventory Linear Programming
II. SPACE/C41SR	Operational Contribution of Space Systems C41SR Operations Research & Intelligence Information Warfare Electronic Warfare & Countermeasures Unmanned Systems Military Environmental Factors	Multiobjective Optimization Network Methods Nonlinear Programming **Probabilistic Operations Research**
III. JOINT WARFARE	Land & Expeditionary Warfare Littoral Warfare/Regional Sea Control Power Projection, Planning, & Execution Air Combat Analysis & Combat ID Special Ops/Operations other than War Joint Campaign Analysis	Decision Analysis Markov Processes Reliability Simulation Stochastic Processes Queuing Theory
IV. RESOURCES	Mobility & Transport of Forces Logistics, Reliability, & Maintainability Manpower & Personnel	**Applied Statistics**
V. READINESS & TRAINING	Readiness Analytical Support to Training & Mission Rehearsal Battlefield Performance, Casualty Sustainment, & Medical Planning	Categorical Data Analysis Forecasting/Time Series Multivariate Analysis Neural Networks
VI. ACQUISITION	Measures of Effectiveness Test & Evaluation Analysis of Alternatives Cost Analysis Decision Analysis	Nonparametric Statistics Pattern Recognition Response Surface Methodology
VI. ADVANCES IN MILITARY OR	Modeling, Simulation, & Gaming Revolution in Military Affairs (Long Range/Strategic Planning) Computing Advances in Military OR	**Others** Advanced Computing Advanced Distributed Systems (DIS) Cost Analysis Wargaming

Length of Papers

Submissions will normally range from 10–30 pages (double spaced, 12 pitch, including illustrations). Exceptions will be made for applications articles submitted with a senior decision-maker letter signed by the Secretary of Defense.

Paper Electronic Submission with Figures, Graphs and Charts

After the article is accepted for publication, an electronic version of the manuscript must be submitted in Microsoft Word or WordPerfect. For each figure, graph, and chart, please include a camera-ready copy on a separate page. The figures, graphs, and tables should be of sufficient size for the reproduced letters and numbers to be legible. Each illustration must have a caption and a number.

Mathematical and Symbolic Expressions

Authors should put mathematical and symbolic expressions in Microsoft Word or WordPerfect equations. Lengthy expressions should be avoided.

Approval of Release

All submissions must be unclassified and be accompanied by release statements where appropriate. By submitting a paper for review, an author certifies that the manuscript has been cleared for publication, is not copyrighted, has not been accepted for publication in any other publication, and is not under review elsewhere. All authors will be required to sign a copyright agreement with MORS.

Abbreviations and Acronyms

Abbreviations and acronyms (A&A) must be identified at their first appearance in the text. The abbreviation or acronym should follow in parentheses the first appearance of the full name. To help the general reader, authors should minimize their use of acronyms. If required, a list of acronyms can be included as an appendix.

Footnotes

We do not use footnotes. Parenthetical material may be incorporated into a notes section at the end of the text, before the acknowledgment and references sections. Notes are designated by a superscript letter at the end of the sentence.

Acknowledgments

If used, this section will appear before the references.

References

References should appear at the end of the paper. The references should be unnumbered and listed in alphabetical order by the name of the first author.

Appendixes

If used, this section will appear after the references.

MILITARY OPERATIONS RESEARCH SOCIETY, INC.

COPYRIGHT ASSIGNMENT AGREEMENT

THIS AGREEMENT, made by and between ____________________, hereinafter ("Assignor(s)") and MILITARY OPERATIONS RESEARCH SOCIETY, INC., hereinafter Assignee).

Assignor(s) (*for U.S. Government employees: to the extent legally permissible*) hereby grants, transfers, assigns, and conveys to Assignee, its successors and assigns, the entire title, right, interest ownership and all subsidiary rights in and to an article entitled ____
____
____ (hereinafter the "Work"), including, but not limited to, the right to secure copyright registration of the Work and to any resulting registration in Assignee's name as claimant. Notwithstanding the foregoing, Assignor(s) reserves all proprietary rights other than copyright, such as patent rights; a royalty-free right to reproduce the Work; and the right to prepare derivative works based upon the copyright in the Work.

Assignor(s) hereby confirms that Assignor(s) owns the entire title, right and interest in the work, including the right to reproduce, prepare derivative works based upon the copyright in the Work, and distribute the Work, whether or not the Work constitutes a "work made for hire" as defined in 17 U.S.C. Section 201(b); Assignor(s) agrees that no rights in the Work are retained by Assignor(s) except as stated above. Assignor(s) agrees to execute any documents that might be necessary to perfect Assignee's ownership of copyrights in the Work and to registration.

Assignor(s) represents that the Work has not been copyrighted or published; that it is not being submitted for publication elsewhere; and, if the Work is officially sponsored by the U.S. Government, that it has been approved for open publication.

This Agreement constitutes the entire agreement between the parties hereto; this Agreement supersedes any prior oral or written agreement or understanding between the parties; and, **in the case of a Work Made for Hire, this Agreement has been signed by the Assignor(s)'s employer.** This Agreement shall only be effective if and upon the date that the Work is accepted by Assignee for publication in *Military Operations Research*, the Journal of the Military Operations Research Society or PHALANX, the Bulletin of the Military Operations Research Society.

ASSIGNOR(S)	ASSIGNOR(S)
By: ____________________	By: ____________________
Title: ____________________	Title: ____________________
Date: ____________________	Date: ____________________

MILITARY OPERATIONS RESEARCH SOCIETY, INC. (ASSIGNEE)

By: ____________________

Title: ____________________

Date: ____________________

MILITARY OPERATIONS RESEARCH SOCIETY BOARD OF DIRECTORS

OFFICERS

President
Dr. Jerry A. Kotchka*
(314) 777-9990
E-mail: jkotchka@mdc.com

Vice President Finance and Management
Dennis R. Baer*
(703) 312-2149
E-mail: drbaer@aol.com

Vice President Meeting Operations
Susan M. Iwanski*
(516) 346-9138
E-mail: sue_iwanski@atdc.northgrum.com

Vice President Professional Affairs
Dr. Robert S. Sheldon*
(703) 648-8268
E-mail: bs@s3i.com

Secretary
CAPT Lawrence L. Dick*
(703) 602-2791
E-mail: leesmail@pipeline.com

Past President
Frederick E. Hartman*
(202) 298-7166
E-mail: hartmanfe@aol.com

Executive Vice President
Richard I. Wiles*
(703) 751-7290
E-mail: morsone@aol.com

Vice President Administration
Natalie S. Addison
(703) 751-7290
E-mail: morsvpa@aol.com

*** - Denotes Executive Council Member**

OTHER DIRECTORS

Col Thomas Allen, USAF, AFSAA
LTC(P) James E. Armstrong, USA, AD*, USMA
Mary T. Bonnet, AFSAA
Dr. Alfred G. Brandstein, MCCDC
Dr. Yupo Chan, AFIT
Dorn Crawford, AD, US ACDA
Dr. Henry C. Dubin, USA OPTEC
James B. Duff, PRC, Inc
Helaine G. Elderkin, FS, AD, Computer Sciences Corporation
Brian D. Engler, Systems Planning and Analysis, Inc.
Christine A. Fossett, AD, GAO
Maj Mark A. Gallagher, USAF, USSTRATCOM
Priscilla A. Glasow, AD, MITRE
Dr. Dean S. Hartley, Data Systems R&D Program
Dr. Glen H. Johnson, US ACDA
RADM Pierce J. Johnson, USNR, Naval Reserve Readiness Command
COL James L. Kays, USA, USMA
Kerry A. Kelly, AD, USSTRATCOM
COL Kenneth Konwin, USAF, DMSO
MAJ Willie J. McFadden, USA
CDR Kirk Michealson, USN, OSD PA&E
Dr. Julian I. Palmore, University of Illinois
Dr. Gregory S. Parnell, FS, AD, Virginia Commonwealth University
Anne M. Patenaude, SAIC
Dr. Roy Rice, Teledyne Brown Engineering
Dr. Patricia A. Sanders, OSD PA&E
Edward A. Smyth, The Johns Hopkins University
Dr. Cyruss Staniec, Logicon
Dr. Stuart H. Starr, AD, MITRE
Joseph A. Tatman, AD, TASC
Harry J. Thie, AD, RAND
LCDR Katie P. Thurman, USN, XO, NRD
Howard G. Whitley III, USA CAA
James L. Wilmeth III, AD, SETA Corp.
LTC Mark A. Youngren, USA, AD, Naval Postgraduate School

*** - Denotes Advisory Director**