Discussion Paper No. 851

INCENTIVES, THE BUDGETARY PROCESS, AND INEFFICIENTLY LOW PRODUCTION RATES IN DEFENSE PROCUREMENT

by William P. Rogerson*

September 1989

*I would like to thank Kathleen Hagerty for helpful comments and discussions. This work was supported by the Lynde and Harry Bradley Foundation, NSF Grant SES-8906751 and the Rand Corporation through funds provided by Program Analysis and Evaluation Office of the Secretary of Defense.

1. Introduction

A number of different studies by Congress, the military, and outside sources have documented the fact that production rates in plants producing major weapons systems tend to be very low relative to the capacity of the plants. The most often cited study is by the Congressional Budget Office (CBO) [1987b]. They examined production rates for 40 weapons systems over the time period 1983-1987. This was the peak period of the Reagan defense buildup. Thus evidence of under utilization during this period would be particularly compelling. The study concluded that 20 of the 40 systems were being produced at unacceptably low rates.[1] Only 18 of the 40 systems were being produced at more than 50% of the maximum production rate for the existing production facility. It was estimated that a 50% increase in the production rate for the twenty systems identified as being produced at unacceptably low rates would have decreased average cost by 10% to 15%. In a different study Rand researchers have summed up the situation as follows.

> "The full, planned production rate (for which the production line was designed to be efficient) is seldom achieved or, if achieved, rarely long maintained . . . (some aircraft, for example, have been produced at rates of less than one or two per month) . . . No major Air Force program has been procured to the original plan since 1969, and the other services display no greater stability. [Dews and Birkler 1983, page 3].

See Air Force Systems Command [1983] and Gansler [1989, chapter 5, for similar conclusions.

Therefore there appears to be a consensus within the defense community that production of weapons systematically occurs in plants which are of a large scale relative to the levels of output actually produced. The

important consequence of this phenomena which this paper will focus on is that this means that existing output levels could be produced more cheaply in plants of a smaller scale.[2]

The existing policy debate over "low production rates" seems to have largely ignored this aspect of the problem. The almost universal policy prescription advocated by existing analyses is to increase production rates. Given a fixed defense budget this implies that fewer types of weapons will be purchased.[3] Thus the existing debate focuses on the idea that in their long range planning Congress and the military together may have made the mistake of deciding to produce too many different types of weapons at too low a quantity. To correct this mistake one would close down some plants in the short run and then develop fewer types of weapons in the longer run.

This may or may not be true. However the issue of whether production occurs at efficient scale or not is really quite separate from the issue of whether the right weapons systems are chosen. Take Congress's and the military's preferences over the quantity-variety choice as given. Then whether these are correct or not, one would presumably like to have the output choices arising from these preferences to be produced at minimum cost. Thus the current debate over low production rates which focuses on whether output choices are being made correctly ignores an important aspect of the phenomena of low production rates. This is that defense production appears to systematically occur at inefficiently high scales. The purpose of this paper is to highlight this as a separate issue, to describe a theory of why it occurs, and to suggest a number of possible policy approaches to dealing with it which follow from the theory.

The theory of this paper is based on the fact that Congress does not make technical design choices over production facilities itself. Rather it delegates these decisions to the military because Congress does not have sufficient time or expertise to make these decisions itself. Congress of course retains control of the decision over how many units of each weapons system to buy. Therefore on a formal level, the decision-making process leading up to the production of a new weapons system is modelled as a two-person game where the military first chooses scale and then Congress decides what quantity to purchase (if any). The military's goal is to maximize the number of units produced. The key idea is that by increasing capacity, the military lowers marginal costs and thus increases the amount that Congress will buy (so long as it buys any). Thus the military will expand capacity until production is so inefficient that Congress is indifferent between buying and not buying the system. This maximizes the number of units purchased.

In the resulting equilibrium, more output is produced than Congress would ideally want. However there is another problem as well. The military cannot simply order Congress to increase output until all social surplus vanishes. Rather it must manipulate the production technology so that Congress will rationally want to order more units. In the resulting equilibrium the scale of production is too large given output. That is, the same output could be produced more cheaply using a lower-scale plant. Thus inefficient production is essentially an unintended by-product of the military's attempts to expropriate the social surplus arising from weapons programs by inducing Congress to increase quantities purchased.

Perhaps the most important policy prescription to flow from this analysis is simply to emphasize the fact that Congress cannot rely on the military to choose a cost-minimizing production technology even if the military's goal is to maximize military preparedness and Congress and the military agree on what constitutes military preparedness. Therefore there is a need for Congress to critically evaluate whether correct capacity choices are being made. Of course, direct monitoring can never provide a perfect solution. As stated above, decisions regarding the configuration of production facilities are delegated to the military to some extent because Congress has neither the time nor expertise to make these decisions. Nonetheless, Congress does perform some oversight and, naturally enough, concentrates its efforts on areas where it suspects the military will have incentives to make choices other than those Congress would make itself. The point of this paper is that scale choice is such an area.

The alternative to direct monitoring is to attempt to implement institutional changes which somehow reduce the military's incentives to inflate capacity. This paper suggests three possible approaches to doing this -- pre-commitment to fixed budgets, creation of inter-service rivalry through overlapping service jurisdictions and greater Congressional control of the projected procurement rates and overall budget levels that the military uses for internal planning purposes. The latter approach seems particularly promising.

In addition to the above policy implications, this paper sheds light on the issue of why the military has been slow to adopt flexible manufacturing technology for defense production. The basic idea of this paper is that the military attempts to manipulate Congressional decisions by pre-committing to production technologies which limit Congressional choice. By its very nature flexible manufacturing technology tends to defeat this purpose by allowing more flexibility both in regards to production rate and product choice.

This paper's model is not closely related to the bulk of formal principal-agent models of procurement for two reasons.[4] First, on a conceptual level it considers the relationship between Congress and the military and implicitly assumes that defense contractors are perfectly controlled agents of the military. Most formal procurement models analyze the relationship between the military services and defense contractors.[5] Although I believe this latter relationship is also interesting, I think that a full understanding of the procurement process requires models which explicitly address the fact that important incentive and informational problems exist within the defense bureaucracy. Second, on a more technical level the bulk of formal principal-agent models of procurement assume that the principal can precommit to some fixed policy. The goal is then to solve for the optimal precommitment. In order to create a problem simple enough to solve, very extreme assumptions on the nature of the problem must be made. The goal of this paper is much more basic. It is simply to show that an incentive for the agent (the military) to choose excess capacity exists when no precommitment by the principal (the Congress) is possible. On the other hand, the model in which the result is derived is extremely general compared to the sorts of models required in the standard approach.

The paper's model is most closely related to a previous paper of my own [Rogerson 1989] in which it is shown that a similar type of model can explain the apparent bias of procurement policy towards quality and away from quantity.

The paper is organized as f 'lows. Section 2 examines the evidence for inefficient production in more detail. A simple econometric analysis of the CBO [1987b] data is performed which supplies additional evidence. Section 3 presents and analyzes the basic one-period model. Section 4 shows that the key ideas generalize to the more realistic case where weapons production occurs over many periods and Congress chooses both a rate of production and total quantity. Section 5 shows how the formal model can be interpreted as applying to the choice of fl 'bility. Sections 6 and 7 explore possible methods for improving the military's incentives. Section 8 considers the issue of flexible manufacturing technology. Section 9 briefly speculates on how the model might generalize to consider conflicts within the military. Finally, Section 10 distinguishes this paper's perspective from that of most existing analyses of low production rates.

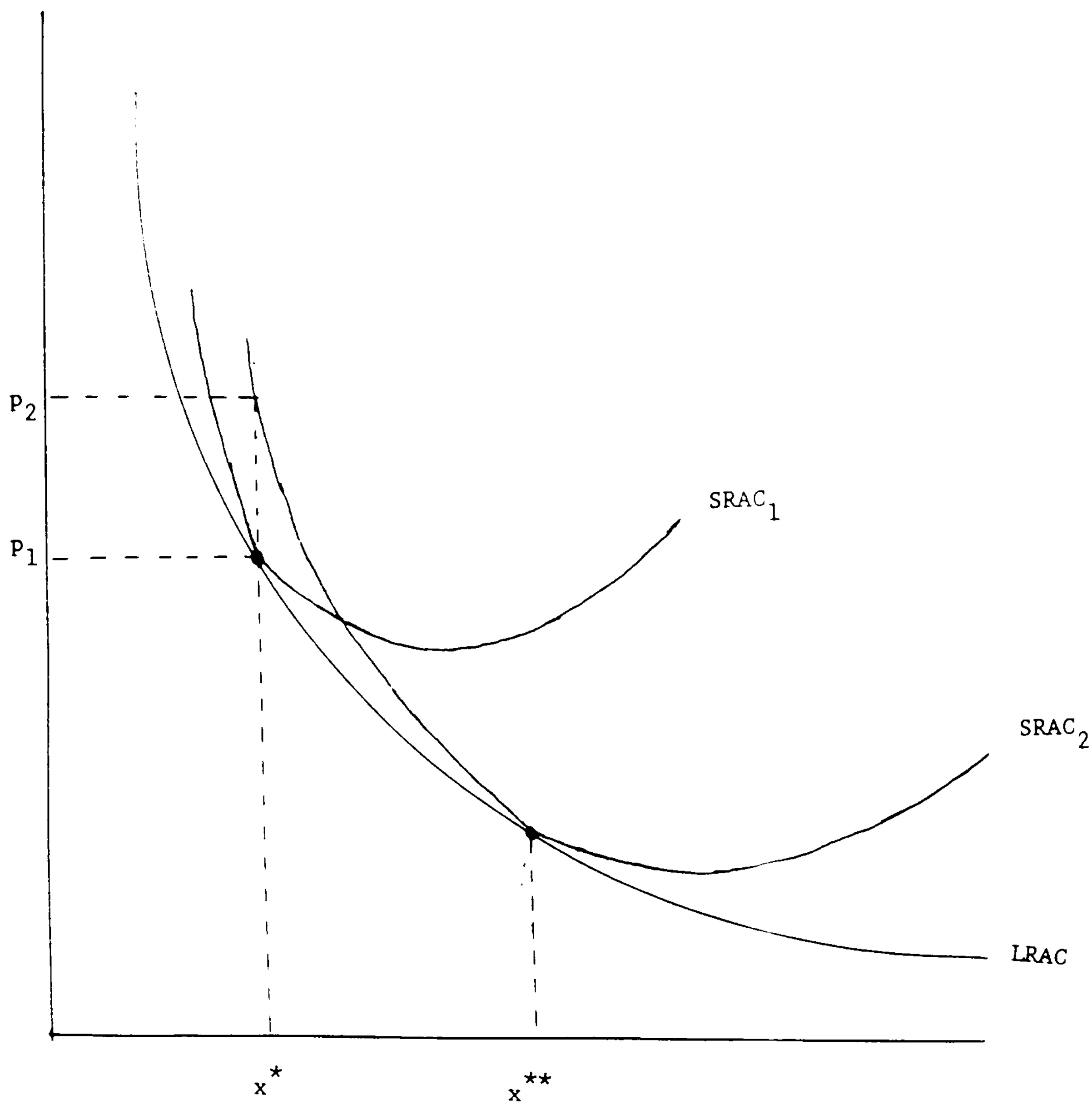
2. Production Off The Long Run Cost Curve

The basic stylized fact that this paper attempts to explain is that production of major weapons systems occurs off the long run cost curve in the sense that the scale or capacity of the production facility is too large given the existing output. The purpose of this section is to briefly describe the quantitative evidence supporting this stylized fact and then attempt to strengthen it somewhat through a simple econometric analysis of the CBO [1987b] data.

First it will be useful to illustrate the stylized fact on a simple graph. This is done in Figure 2.1. The long run average cost curve, labeled LRAC, is drawn to be declining everywhere since this is probably the case over all relevant ranges of production for most weapons systems. Each possible

Figure 2.1

Production Off The Long Run Cost Curve



production facility has a short run average cost (SRAC) curve tangent to LRAC at one point. Larger scale production facilities are tangent to LRAC further to the right. The short run average cost curves for two production facilities are drawn in the figure. These are denoted by $SRAC_1$ and $SRAC_2$. Suppose that x^* is the current level of production. Then the cost minimizing technology is technology 1. This would yield average costs of p_1. If a larger scale technology were used average costs would be higher. For example if technology 2 were used then average costs were be p_2. The stylized fact which this paper attempts to explain is that precisely this situation occurs -- i.e. -- technology 2 is used to produce x^*.[6]

At the outset it should be noted that extremely good evidence does not exist or is at least not publicly available. Extremely good evidence would consist of careful engineering analyses of alternative production technologies which conclude that existing levels of output could have been more cheaply produced by using smaller facilities. Existing quantitative evidence focuses on establishing two properties.

(1) Declining Short Run Average Cost

Many studies claim that average costs would decline dramatically if output rates were increased using the current production facilities. One can interpret the CBO [1987b] conclusion cited in the introduction that 18 of the 40 systems studied were produced at unacceptably low rates as a statement that short run average costs were declining dramatically for these 18 systems. The CBO [1987b] also directly estimated the slope of short run average cost for some systems as reported in the introduction.

Gansler [1989] and Air Force Systems Command [1983] supply this type of evidence as well.

(2) Excess Capacity

Many studies claim that existing production facilities are operated at extremely low levels of output relative to their maximum possible outputs. The best quantitative evidence of this sort is once again from the CBO [1987b] and it will be described in more detail below.

The problem with both of these types of evidence is that they do not directly address the question of inefficient scale choice. That is, both properties are actually consistent with operation on the long run cost curve. First consider property (1), declining SRAC. So long LRAC is declining then so will SRAC, even when scale choice is efficient. This is illustrated in Figure 2.1. Observe that $SRAC_1$ is declining at x^*. In order to interpret property (1) as supporting the existence of excessive scale one must therefore draw a subjective and judgmental conclusion. Namely, one must conclude that short run average cost is declining <u>so</u> dramatically that it seems very implausible that the existing technology is the cheapest method of producing the existing output.

Now consider property (2), excess capacity. Once again, it is not directly obvious how this relates to the issue of excessive scale. In particular, one would expect that even an efficiently sized plant would not be run at 100% of maximum capacity. Thus one must somehow know the optimal utilization rate in order to use existing utilization rates to infer the existence of productive inefficiency. The CBO [1987b], for example, essentially simply states that in its judgment the existing utilization rates averaging 43% of maximum capacity are too low to possibly be optimal. The only evidence cited is that commercial manufacturing typically exhibits much higher utilization rates of perhaps 70%[7]. Although their conclusion seems plausible it is not supported by objective analysis of cost functions. In particular, as for property (1), it is possible that the observed objective evidence is consistent with operation on the long run cost curve. One compelling explanation, for example, is that there are large indivisibilities in defense production. Thus it is efficient to construct a plant capable of producing perhaps 150 aircraft per year even if only 20 aircraft per year will be produced.

This paper will now perform a simple econometric analysis of the CBO [1987b] data establishing properties (1) and (2) to help strengthen the case that production occurs off the long run cost curve. In particular, it will be shown that properties (1) and (2) continue to hold to about the same extent for all weapons systems analyzed by the CBO regardless of their production levels. Thus the production facility for an aircraft being produced at a rate of 8 per year will generally have no greater excess capacity than that for an aircraft being produced at a rate of 150 per year. Similarly SRAC is no more likely to be severely decreasing in either case.

Before presenting the empirical analysis, its theoretical significance will be explained. First consider property (1). Suppose that production of all airplanes occurred on the LRAC curve. Then so long as LRAC was convex and so long as relatively similar technologies were used for all aircraft, we would expect the slope of SRAC to be flatter for aircraft being produced at larger levels of output. For example, in Figure 2.1, if x^* and x^{**} are both being produced efficiently then the slope of SRAC is flatter for x^{**} than x^*.

Therefore we would expect property (1) to be less true for aircraft being produced at higher outputs. Now consider property (2). Once again suppose that production of all aircraft occurs on the LRAC curve. As well, suppose the existence of large amounts of excess capacity is explained by indivisibilities. Then once again we would expect excess capacity to become less evident for aircraft being produced at larger output levels. For example, consider the very simple case where the smallest possible plant can produce 150 aircraft per year. Then we would expect to see excess capacity diminish to zero as output rates increased to 150.

Therefore it seems plausible that if production is occurring on the LRAC curve, then both properties (1) and (2) will tend to diminish for systems produced at higher outputs. In particular, suppose that the production rate does not appear to be correlated with either excess capacity or slope of SRAC over the cross-section of programs analyzed by the CBO. Then this would provide some extra evidence that production is not occurring on the long run cost curve.

This fact will now be demonstrated. In order to attempt to control in a rudimentary fashion for the effects of different types of production technologies, the sample of weapons systems will be divided into two groups. The first group is aircraft and helicopters and the second group is missiles. From now on the group "aircraft and helicopters" will simply be referred to as "aircraft". There are 19 observations in the first group and 16 in the second for a total of 35. Three of the 40 systems analyzed by the CBO were neither a missile nor an airplane so were discarded. The other two discarded data points were for systems which had not yet entered full-rate production.[8]

Tables 2.1 and 2.2 present the CBO [1987b] data for each group. Table 2.2 also presents the averages for each group and the average across both groups. The first column is simply the average annual production rate in number of units over the 1983-87 time period. It will be donated by R. The second column is what the CBO terms the "minimum economic rate." This is defined as "the point on the cost schedule below which unit costs rise at an excessive rate"[9] and is simply a number reported by program managers for each system. No systematic or uniform method for determining this rate exists and the rate is thus a highly suspect number as the CBO itself admits.[10] Nonetheless it is the only available measure for all 35 systems of how steep the short run average cost curve is at the current output level. Thus it will be used. The third column presents the maximum output rate for each system given the current production facility. This number is once again derived from DoD reports but it is more reliable since a more precise definition of it exists. The fourth and fifth columns present the current output rate as a percentage of the minimum and maximum output rates. These will be called, respectively, the minimum utilization rate and the maximum utilization rate and will be denoted, respectively, by U_{MIN} and U_{MAX}. Finally the sixth column is a binary variable which records whether the current rate is greater or less than the minimum economic rate. This is denoted by U_{BIN} and will be called the binary utilization rate.

$$(2.1) \qquad U_{BIN} = \begin{cases} 100, & U_{MIN} \geq 100 \\ 0, & U_{MIN} < 100 \end{cases}$$

Table 2.1

Production Rates for Aircraft*

System	Annual Rate (R)	Minimum Economic Rate	Maximum Rate	U_{MIN}**	U_{MAX}**	U_{BIN}**
E-2C Hawkeye Aircraft	7	6	18	117%	39%	100
SH-2F Seasprite Helicopter	8	6	48	133%	17%	100
A-6E Aircraft	8	12	72	67%	11%	0
C-2 Greyhound Aircraft	8	8	9	100%	89%	100
P-3C Aircraft	8	16	24	50%	33%	0
EA-6B Prowler Aircraft	9	6	24	150%	38%	100
KC-10 Tanker/Cargo Aircraft	9	8	24	113%	38%	100
CH-53 Super Stallion Helicopter	12	12	24	100%	50%	100
C-5B Transport	15	4	24	375%	63%	100
EH-60 Quickfix Helicopter	17	24	48	71%	35%	0
F-14A Aircraft	21	12	96	175%	22%	100
SH-60B LAMPS Helicopter	23	24	60	96%	38%	0
B-1B Bomber	31	24	48	129%	65%	100
AV-8B Aircraft	34	36	72	94%	47%	0
F-15 Aircraft	41	120	144	34%	28%	0
F/A-18 Aircraft	84	84	145	100%	58%	100
UH-60 Black Hawk Helicopter	85	96	144	89%	59%	0
AH-64 Apache Helicopter	117	72	144	163%	81%	100
F-16 Aircraft	155	108	324	144%	48%	100
Aircraft	36	36	79	100%	46%	63

* This is from CBO [1987b], pages 10-11. Annual rates are averages for 1983-87.

$$**U_{MIN} = \frac{\text{Annual Rate}}{\text{Min. Econ. Rate}} \times 100\%$$

$$U_{MAX} = \frac{\text{Annual Rate}}{\text{Max. Rate}} \times 100\%$$

$$U_{BIN} = \begin{cases} 100, & U_{MIN} \geq 1 \\ 0\ , & U_{MIN} < 1 \end{cases}$$

Table 2.2
Production Rates for Missiles *

System	Annual Rate (R)	Minimum Economic Rate	Maximum Rate	U_{MIN}**	U_{MAX}**	U_{BIN}**
MX	17	21	48	81%	35%	0
Ground Launched Cruise Missile	99	120	600	83%	17%	0
Tomahawk	186	300	540	62%	34%	0
Phoenix	222	240	420	93%	53%	0
Harpoon	284	360	660	79%	43%	0
Standard Missile 2 (Extended Range)	296	360	480	82%	62%	0
Patriot	485	240	840	202%	58%	100
Standard Missile 2 (Medium Range)	552	480	844	115%	65%	100
Laser Maverick	1300	1800	3600	72%	36%	0
HARM	1460	3240	6480	45%	23%	0
Sparrow	2015	1200	3804	168%	53%	100
Sidewinder	2122	2400	8400	88%	25%	0
IIR Maverick	2205	6000	10800	37%	20%	0
Stinger	3539	1800	11520	197%	31%	100
Hellfire	6131	1500	6720	409%	91%	100
TOW 2	15482	21600	30000	72%	52%	0
Aircraft Average	36	36	79	100%	46%	63
Missile Average	2275	2604	5360	87%	42%	31
Aircraft and Missile Average	1060	1210	2493	88%	43%	49

* This is from CBO [1987b], pages 10-11. Annual rates are averages for 1983-87.

$$**U_{MIN} = \frac{\text{Annual Rate}}{\text{Min. Econ. Rate}} \times 100\%$$

$$U_{MAX} = \frac{\text{Annual Rate}}{\text{Max. Rate}} \times 100\%$$

$$U_{MIN} = \begin{cases} 100, & UB_{MIN} \geq 1 \\ 0\ , & U_{BIN} < 1 \end{cases}$$

It is straightforward to test whether property (2) holds constant over various production rates. This is because U_{MAX} is an inverse measure of excess capacity. The test of whether property (1) holds constant over various production rates is less direct and probably therefore less reliable. The minimum economic rate is defined to be a point where the slope of SRAC becomes quite steep. Therefore U_{MIN} can perhaps be interpreted as an inverse measure of the slope of SRAC. One might feel that U_{MIN} is transmitting more information than is available regarding relative slopes. That is, one might interpret the CBO data as simply reporting a binary variable. This is whether the slope is "high" (i.e. -- $U_{MIN} < 100$) or "low" (i.e. - $U_{MIN} \geq 100$). Therefore an alternate inverse measure of slope is U_{BIN}.

Therefore we would expect all three utilization rates to fall as the production rate increases if production occurs on the long run cost curve. As a formal statistical matter we are testing the hypothesis

H_0: U_x is constant over output rates

against the alternative

H_1: U_x falls as output rates rise

where $x \in \{MIN, MAX, BIN\}$.

Table 2.3 reports the results of running the linear regression

$$(2.2) \qquad U_x = \alpha + \beta R .$$

Table 2.3

Regression Results

Regression x	Group	α	β	R^2	t_β*	$\bar{U}_x$	$(\alpha/\bar{U}_x) \times 100\%$
MAX	Aircraft	38.5	.1819	.14	1.69	45.2	85%
	Missiles	40.8	.0012	.06	.92	43.6	94%
	Both	43.5	.0009	.02	.75	44.5	98%
MIN	Aircraft	118.7	.0607	.00	.15	121.0	98%
	Missiles	107.2	.0046	.04	.74	118.8	90%
	Both	115.6	.0036	.02	.74	119.5	97%
BIN	Aircraft	58.8	.1192	.01	.42	63.2	93%
	Missiles	29.9	.0006	.00	.18	31.3	96%
	Both	50.5	-.0018	.01	-.57	48.6	104%

*Significance levels are as follows for the regressions:

	$t_{.05}$	$t_{.10}$	$t_{.25}$
Aircraft	1.74	1.33	.69
Missiles	1.76	1.35	.69
Both	1.65	1.28	.67

Results for $x \in \{MIN, MAX, BIN\}$ are reported. As well, for each x, the results of three separate regressions for aircraft, missiles, and aircraft and missiles combined are reported. Thus there are nine regressions. The first five columns are self explanatory. The sixth column labeled "t_β" is the t-statistic for β. The seventh column labeled $\bar{U}_x$ is the average value of U_x. The final column is the percentage of the average explained by the intercept term.

The results generally support the conclusion that production rates have no significant effect on utilization rates. This is for three reasons. First, the values of β are not significantly different from 0 for any of the regressions. One cannot reject the null hypothesis that $\beta = 0$ with 95% confidence for any of the regressions. The regression using U_{MAX} for aircraft has β significant at a 90% level. However all the other regressions yield insignificant β's at the 90% level as well. Second, the magnitudes of the estimated β's are very small. Thus production rate differentials have a very small impact on utilization rates even if they are viewed as statistically significant. This can be illustrated as follows. For any given regression let $\bar{R}$ and $\bar{U}_x$ denote the mean of R and U_x. Then, according to the regression equation, a program with extremely low output will exhibit a utilization rate equal to $(\alpha/\alpha+\beta\bar{R})$ x 100% of the average program. This equals $(\alpha/\bar{U}_x)$ x 100% because the regression line goes through the means. These figures are reported in the last column of Table 2.3. The lowest percentage is 85% and all others are over 90%, with many over 95%. Thus as the production rate increases from 0 to the average rate there is only a small percentage impact on the utilization rate, according to the regression results. Third, all of the regressions have extremely low R^2 values, indicating that the production rate has very little explanatory power.

A number of other remarks should be noted about these regressions. First, the U_{BIN} variable appears to have the least positive correlation with R of the three utilization variables. The values of β are not significantly different from 0 at even the 75% level for all of these regressions. Thus the probability of production occurring below the minimum economic rate appears to be totally unrelated to the level of production.

Second, the values of β for the missile regressions are always less than the corresponding values of β for the aircraft regressions. Thus the evidence that missile production occurs off the long run cost curve is greater. This is consistent with other aspects of the data. Namely, with reference to the bottom rows of Table 2.2, missile production appears to be less efficient than aircraft production by almost any measure. The average aircraft program operates at 100% (46%) of its minimum economic (maximum) rate while the average missile program operates at only 87% (42%) of its minimum economic (maximum) rate. While 63% of the aircraft programs operated above their minimum economic rate, only 31% of missile programs did so. These average statistics suggest that missile production is less efficient and this relative ranking is perfectly consistent with the relative rankings yielded by the regression results of this paper. This consistency can perhaps be interpreted as supporting the validity of the regression approach.

Third, the regressions which use aircraft and missile data together always yield lower values of β and lower t-statistics than do either of the regressions run on missiles or airplanes separately. There is a very clear economic reason for this. As explained above, the missile programs as a group exhibit lower utilization rates of all sorts than do the aircraft programs. However, the missile programs are on average much larger. From Table 2.2 the average aircraft program produced 36 units per year while the average missile program produced 2,275 units per year. Therefore when these two data sets are combined, the fact that the larger missile programs exhibit lower utilization rates pushes down the value of β. In fact for the case of U_{BIN} the value of β becomes mildly negative.

How one interprets this result depends upon whether one believes that a major difference in production technologies exists between missiles and aircraft which means that capacity utilization figures are non-comparable. If one believes that such a difference exists then the combined regressions should be ignored. However,if one believes that production technologies across the groups are no more variable than those within the groups, then the combined regressions strengthen the conclusions of this paper. On an economic level we observe that missile production occurs at very high volumes relative to aircraft production. Thus, the long run cost curve should have begun to "flatten-out" considerably at the rates missiles are produced. Therefore if production is occurring on the long run cost curve we should observe fewer reports of extremely steep short run average cost curves or large amounts of excess capacity. However precisely the reverse is true. It is for this reason that the combined regressions report lower values of β.

In conclusion, the evidence that inefficient scale choices are made in defense production is not perfect. This section has attempted to supply additional quantitative support for this contention by showing that the existence of extremely steep SRAC curves and large amounts of excess capacity are essentially unrelated to production volume. It is difficult to interpret these phenomena as being consistent with operation on the long run cost curve. The remainder of this paper will simply take the stylized fact of inefficiently large scale choice as given and attempt to explain it.

3. The Single Period Model

A. The Model

This section will demonstrate the key results and ideas of this paper in a model which abstracts away from the complication that Congress in reality must choose both an annual rate of production and a total quantity of production. Formally it will be assumed that all production occurs in a single year and thus annual rate and total quantity are the same.[11] Section 4 will show that the key results of this section all still hold in a more realistic model where Congress chooses both the annual rate of production and the total quantity of production.

A weapons program will be completely described by two non-negative numbers (x,s) where x denotes the number of units procured and s denotes the scale of production. The social value (in dollars) of x units of the weapon is given by $V(x)$. Both Congress and the military agree on $V(x)$. The cost of producing x units given the scale s is denoted by $C(x,s)$. For any fixed scale, C will be assumed to be a regular well-behaved cost function. Thus each s represents a choice of a particular production technology and it is assumed that a continuum of technologies are available indexed by $s \in [0,\infty)$. As will be explained below, an assumption on $C(x,s)$ will be made which allows one to naturally interpret s as a measure of scale of production.

Note that both V and C are assumed to be non-random. One can show that the major conclusion of this paper that capacity is chosen too large still holds true when V and C are assumed to be known only probabilistically at the time of scale choice. However the analysis is considerably simpler and clearer for the non-random case. Therefore all formal models of this paper will assume that V and C are non-random. In particular this means that the need surge capacity is not considered. One would model the need for surge capacity by assuming that V depends on what type of (if any) war is occurring and that this is only known probabilistically at the time of scale choice.

The various regularity assumptions which will be made about V and C will now be presented and discussed where necessary. For the most part they simply guarantee that various maximization problems have unique well-behaved solutions. Assumptions 3.1-3.3 state the smoothness and concavity properties of V and C.

Assumption 3.1

(i) $V(x)$ is twice continuously differentiable over $(0,\infty)$

(ii) $V(x)$ is strictly increasing over $[0,\infty)$

(iii) $V(0) = 0$

Assumption 3.2

(i) $C(x,s)$ is twice continuously differentiable over $(0,\infty)^2$

(ii) C is strictly increasing in x over $[0,\infty)$ for every s

(iii) $C(0,0) = 0$

(iv) $C(x,s) > 0$ for $(x,s) \neq (0,0)$.

Assumption 3.3

$V(x) - C(s,x)$ is globally strictly concave over $(0,\infty)^2$.

A first-best program maximizes social surplus. It is formally defined as follows.

<u>**Definition**</u>:

A weapons program is first best if it solves the following program.

$$(3.1) \qquad \underset{x,s \geq 0}{\text{Maximize}} \quad V(x) - C(x,s)$$

It will be assumed that a unique first-best program exists and that it is strictly preferred to producing nothing.

<u>**Assumption 3.4**</u>

A unique first-best weapons program denoted by (x^*,s^*) exists. Furthermore

$$(3.2) \qquad V(x^*) - C(x^*,s^*) > 0 \ .$$

A scale is second best given an output if it minimizes the costs of producing that output.

<u>**Definition**</u>:

A scale is second best given x if it solves the following problem.

$$(3.3) \qquad \underset{s \geq 0}{\text{Minimize}} \quad C(x,s)$$

It will be assumed that a unique second best scale exists for every $x > 0$.

<u>Assumption 3.5</u>

A unique second-best scale given x exists for every $x > 0$. Let $\phi(x)$ denote this value. Furthermore $\phi(x) > 0$ for all $x > 0$.

Note that by Assumption 3.2 - (iii) and (iv), $s = 0$ is the second-best scale when $x = 0$.

The short run cost curve for a given scale $\hat{s}$ is simply $C(x,\hat{s})$. By Assumption 3.2 - (iv) and 3.5 the long run cost curve is also well-defined. Let $L(x)$ denote the long run cost curve. It is given by

$$L(x) = \begin{cases} 0 & , x = 0 \\ C(x,\phi(x)), & x > 0 \ . \end{cases} \tag{3.4}$$

The above assumptions guarantee that $L(x)$ is smooth over $(0,\infty)$. However, it may jump at zero.

It will be useful to define the second best output over the non-negative and positive orthants. These will be called, respectively, the second-best and interior second-best outputs.

<u>Definition</u>:

An output is second best (interior second best) given s if it maximizes the following function over $x \geq 0$ ($x > 0$).

$$V(x) - C(x,s) \tag{3.5}$$

Assumption 3.6

A unique interior second best output given s exists for every $s > 0$. Let $\psi(s)$ denote this value.

It will also be assumed that there exists a capacity level $\bar{s}$ such that the optimal interior second best output yields negative social welfare for values of $s > \bar{s}$. This simply means that large enough capacity levels would result in it being ex ante preferable to produce nothing. (If one had to build a plant able to produce one billion F16's then it might be better not to build the plant and have none.) For expositional convenience it will also be assumed that the interior maximum produces strictly positive social welfare for values of s less than $\bar{s}$. However this is not actually necessary for the results.

Assumption 3.7

There exists an $\bar{s} > 0$ with the following property.

$$V(\psi(s)) - C(\psi(s), s) \overset{>}{\underset{<}{=}} 0 \iff s \overset{<}{\underset{>}{=}} \bar{s} \tag{3.6}$$

Let $\bar{x}$ denote the interior second-best output at $\bar{s}$ -- i.e. --

$$\bar{x} = \psi(\bar{s}) \tag{3.7}$$

This completes the technical regularity assumptions. Now the assumption which allows one to interpret s as a "scale" parameter will be presented and explained.

Assumption 3.8

$$(3.8) \qquad C_{xs}(x,s) < 0$$

According to Assumption 3.8 increasing the scale of production lowers the marginal cost of production. This is intuitively reasonable. Intuitively a large scale plant is one where greater fixed costs are incurred in order to reduce marginal costs. Support for the contention that this is a reasonable definition of scale can be found by examining how the second-best scale choice varies with x. The second-best scale choice is determined by the first order condition.

$$(3.9) \qquad C_s(x,s) = 0 .$$

Total differentiation yields

$$(3.10) \qquad \frac{ds}{dx} = \frac{C_{sx}}{-C_{ss}} .$$

Since the denominator must be positive by the concavity assumption (Assumption 3.3), it can be seen that a larger scale is optimally chosen for a larger output if and only if C_{sx} is positive.

B. Equilibrium

The decision-making process leading up to the adoption of a new weapons program will now be described. Three fundamental assumptions about the nature of this process will be made.

First, it will be assumed that the military only considers military value when deciding between two programs and ignores cost. That is, the military prefers one weapons program (x,s) to another $(\hat{x},\hat{s})$ if and only if

$$(3.11) \qquad x > \hat{x}$$

regardless of whether

$$(3.12) \qquad V(x) - C(x,s)$$

is greater or less than

$$(3.13) \qquad V(\hat{x}) - C(\hat{x},\hat{s}).$$

Second, it will be assumed that the decision-making process is sequenced as follows. First, the military chooses a scale of production. Then Congress chooses a level of production taking the scale decision as given. This assumption reflects the fact that the military has greater technical expertise than Congress. Thus Congress must delegate the determination of a production technology to the military. Given the technology selected by the military Congress can calculate the costs of producing various levels of output. By choosing a funding level for the program it then determines how many units will be produced. However it is incapable of determining whether other production technologies might have resulted in lower costs or not.

Obviously both of these assumptions are somewhat extreme. In reality military planners may care about more than the success of their program and Congress may be able to evaluate some aspects of the choice of production technology. Nonetheless it certainly seems plausible that planners care more about the success of their programs than whether social value is maximized and that Congress has less expertise than the military regarding the effects of plant scale on production cost. This paper makes extreme versions of both of these assumptions to clearly illustrate their effects in the simplest possible model.

Third, it will be assumed that the defense contractor (who is perfectly controlled by the military) initially pays for the production facility when it is built. Then Congress pays for the facility as part of the production cost if and only if it purchases any units of the system. That is, Congress has the option of purchasing zero units and paying zero dollars. However, it pays for the entire production facility if it purchases any units.

There are two reasons for making this assumption. First, this is to a large extent the way that the procurement process is organized. Second, if it was assumed that Congress always paid for facilities capital investments authorized by the military regardless of eventual production decisions there would be another more obvious incentive for the military to increase capacity above the first best. The incentive for the military to choose excess capacity identified by this paper is totally separate and unrelated to this. The clearest method of demonstrating this is of course to simply assume that Congress follows the policy of paying zero if it does not purchase any units.

The incentive for creation of excess capacity which depends on Congressional funding of facilities capital regardless of production will now be briefly described. See Rogerson [1989] for a formal model and a more complete institutional discussion. The basic idea is that Congress will ignore facilities capital costs when making its adoption decision if it has already paid these costs. Thus the military can increase the probability that Congress will adopt a program by decreasing variable costs through employing greater amounts of sunk facilities capital expense. As I argue in Rogerson [1989], a major reason why Congress follows the policy of only paying for facilities capital which is used is probably precisely to avoid this incentive problem. The point of this paper is that the incentives for creation of excess capacity identified here are very different and still exist even when Congress only pays for facilities capital which is used.

The above three assumptions result in the following structure to the game between Congress and the military. As usual, it is most convenient to work backwards. Given the military's choice of s, the Congress will then choose a value of x to maximize

$$V(x) - C(x,s) \tag{3.14}$$

so long as it can achieve greater than zero surplus. Otherwise it will choose $x = 0$. Let $\xi(s)$ be the correspondence denoting Congress's choice given s. By the assumptions in Part A above, ξ is single valued for all values of s except $\bar{s}$. For smaller values it is the interior second best output and for larger values it is zero. For $\bar{s}$, both $\bar{x}$ and 0 are optimal. Formally

$$(3.15) \qquad \xi(s) = \begin{cases} \psi(s) & , \; s < \bar{s} \\ \{0, \bar{x}\}, & s = \bar{s} \\ 0 & , \; s > \bar{s} \end{cases}$$

At the beginning of the game the military therefore chooses s to maximize V(x) realizing that s affects x as described above.

Formally, then, an equilibrium weapons program is described as follows.

Definition:

An equilibrium weapons program solves the following problem.

$$(3.16) \qquad \underset{x,s}{\text{Maximize}} \quad V(x)$$

$$(3.17) \qquad \text{subject to } x \in \xi(s)$$

C. Analysis

Since V is strictly increasing in x the military's problem boils down to choosing s to maximize $\xi(s)$. Thus in order to describe capacity choice, one needs to describe the behavior of $\xi(s)$. Recall that $\xi(s)$ equals the interior solution, $\psi(s)$, for $s \leq \bar{s}$ and equals zero for $s \geq \bar{s}$. Proposition 1 describes the critical feature of $\psi(s)$ for the purposes of this paper. Namely, it is strictly increasing. The reason for this is very simple and intuitive. Increases in scale result in lower marginal costs. This results in a larger interior maximum.

Proposition 3.1:

(3.18) $\psi'(s) > 0$

proof:

The interior maximum for any s is determined by the first order condition

$$(3.19) \qquad V'(x) = C_x(x,s) \ .$$

Total differentiation yields

$$(3.20) \qquad \frac{dx}{ds} = \frac{C_{xs}(x,s)}{V''(x) - C_{xx}(x,s)} \ .$$

This is positive by Assumptions 3.3 and 3.8.

QED.

The nature of the equilibrium program is now clear. The military can induce Congress to buy more units by choosing a larger scale of production and thereby lowering marginal cost. However, production at scales larger than $\bar{s}$ is so inefficient that Congress would rather cancel the program. Thus the optimal course of action for the military is to increase scale up until this point of indifference. This is stated as Proposition 3.2.

Proposition 3.2:

The unique equilibrium weapons program is $(\bar{x},\bar{s})$.

<u>proof</u>:

As above.

QED.

A technical remark about this proposition should be noted. At the capacity $\bar{s}$ Congress is actually indifferent between choosing zero or $\bar{x}$ units of production. The definition of an equilibrium program in (3.16) and (3.17) implicitly assumes that Congress will choose the output level preferred by the military when it is indifferent between two output levels. However this is simply a technical convenience routinely made in the agency literature to guarantee existence of a solution. If one preferred one could think of the military choosing a capacity level slightly below $\bar{s}$ which results in Congress strictly preferring a positive level of output.

The welfare properties of the equilibrium program can now be analyzed. This is done in Proposition 3.3.

<u>Proposition 3.3</u>

(i) The equilibrium capacity and output are both strictly greater than the first best -- i.e. --

(3.21) $\bar{x} > x^*$

(3.22) $\bar{s} > s^*$

(ii) The equilibrium output is second best given the equilibrium capacity -- i.e. --

(3.23) $\bar{x} = \psi(\bar{s})$

and

(3.24) $V(\bar{x}) - C(\bar{x},\bar{s}) \geq V(0) - C(0,\bar{s})$

(iii) The equilibrium capacity is strictly greater than the second best-capacity given output -- i.e. --

(3.25) $\bar{s} > \phi(\bar{x})$

<u>proof</u>:

Parts (i) and (ii) follow immediately. Part (iii) is an immediate consequence of the concavity of the problem. The general result which this depends on is stated and proven in Appendix A.

QED.

These results can be very clearly illustrated on a graph of average costs. Let AC(x,s) denote the average cost of producing x units given scale s. Let AL(x) denote the average long run cost curve. Now refer to Figure 3.1. The long run average cost curve is shown to be declining, since this is probably true over the range in which most weapons would be purchased. The quantity x^* is the optimal quantity. The capacity s^* is the optimal capacity to produce x^* at. This means that $AC(x,s^*)$ is just tangent to AL(x) at x^* as drawn. The quantity $\bar{x}$ is actually purchased and this is greater than x^*. Let s′ denote the optimal scale technology to produce $\bar{x}$ at. Therefore AC(x,s′) is drawn to be tangent to AL(x) at $\bar{x}$. However this is <u>not</u> the technology that is used in equilibrium. Rather, a larger scale technology $\bar{s}$ is used. Therefore $AC(x,\bar{s})$ is drawn to be tangent to $AL(x)$ at a point to the right of $\bar{x}$, which is labeled x'.

The first-best program would have been produced at an average cost of AC^*. If the actual quantity chosen was produced efficiently it would be produced at a lower average cost, AC'. Even this would have been a socially inferior choice because the marginal costs would exceed the marginal benefits. However, the actual outcome is even worse, because $\bar{x}$ is produced using an inefficiently large technology. This results in average costs of $\overline{AC}$. ($\overline{AC}$ may or may not be larger than AC^*.)

Therefore there are two inefficiencies in equilibrium. First, the quantity produced is too large even if it were produced efficiently. Second, an efficiently large scale production technology is used.

4. Multiple Periods of Output

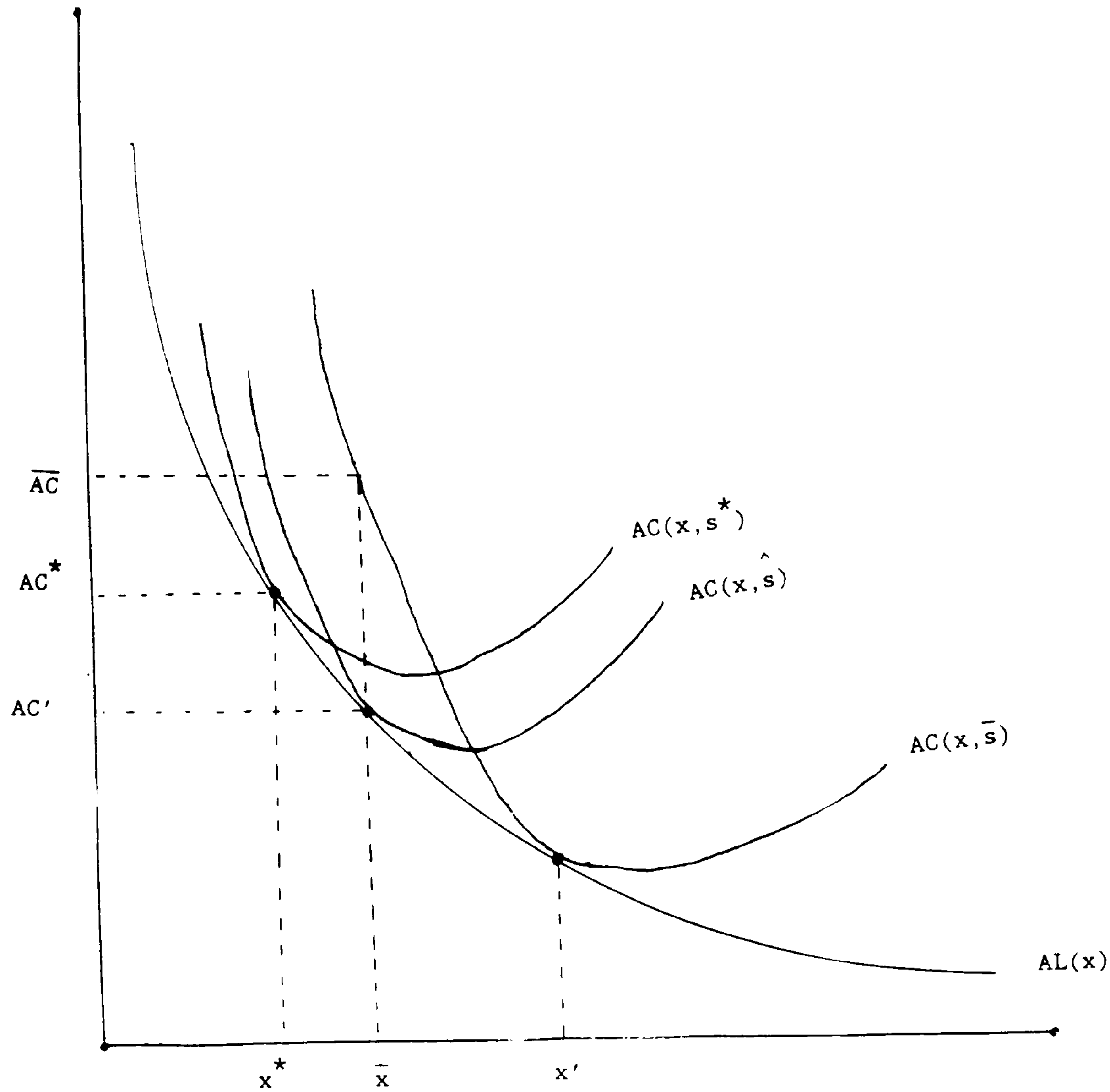
A. Regularity Assumptions

The purpose of this section is to analyze the case where Congress chooses both an annual rate of production and the number of years of production in response to the military's choice of scale. Define an output plan to be an ordered pair on non-negative numbers (x,n) where x denotes the annual rate of production and n denotes the number of years of production. The fact that n must be an integer will be ignored for this analysis. For any output plan, the total quantity of production, denoted by y, is given by

$$(4.1) \qquad y = nx \ .$$

Figure 3.1

The Equilibrium Weapons Program



A weapons program is a three tuple of non-negative numbers (x,n,s) where (x,n) is an output plan and s is the scale of production. Let $V(x,n)$ denote the present value of the military benefits of the output plan (x,n) and let $C(x,n,s)$ denote the present value of the production costs of the output plan (x,n) given the scale s.

Regularity assumptions and associated definitions very analogous to those of Section 3 will now be made. Since they are essentially the same, they will be presented with a minimum of discussion. The definitions of a first-best program, second-best scale, second-best output plan and interior second-best output plan are exactly analogous to those in Section 3 so will not be formally presented.

<u>Assumption 4.1</u>

(i) $V(x,n)$ is twice continuously differentiable over $(0,\infty)^2$.

(ii) $V(x,n)$ is strictly increasing in both variables over $[0,\infty)^2$.

(iii) $V(0,n) = V(x,0) = 0$

<u>Assumption 4.2</u>

(i) $C(x,n,s)$ is twice continuously differentiable over $(0,\infty)^3$.

(ii) C is strictly increasing in its first two variables over $(0,\infty)^2$ for every s.

(iii) $C(0,n,0) = C(x,0,0) = 0$

(iv) $C(x,n,s) > 0$ if $(x,n) >> 0$ or $s > 0$.[12]

<u>Assumption 4.3</u>

$V(x,n) - C(x,n,s)$ is globally strictly concave over $(0,\infty)^3$.

Assumption 4.4

A unique first-best weapons program, denoted by (x^*,n^*,s^*) exists. Furthermore, the first-best program is strictly preferred to producing nothing -- i.e. --

$$(4.2) \qquad V(x^*,n^*) - C(x^*,n^*,s^*) > 0$$

Assumption 4.5

First every $(x,n) >> 0$, a unique second-best scale denoted by $\phi(x,n)$ exists. Furthermore $\phi(x,n) > 0$. The long run cost curve, denoted by $L(x,n)$ is thus given by

$$(4.3) \qquad L(x,n) = \begin{cases} 0 & , \quad x = 0 \text{ or } n = 0 \\ C(x,n,\ \phi(x,n)), & (x,n) >> 0 \end{cases}$$

Assumption 4.6

For every $s > 0$, a unique interior second-best output plan exists denoted by $\psi(s) = (\psi_x(s),\ \psi_n(s))$.

Assumption 4.7

There exists an $\bar{s} > 0$ with the following property.

$$(4.4) \qquad V(\psi(s)) - C(\psi(s),s) \underset{<}{\overset{>}{=}} 0 \iff s \underset{>}{\overset{<}{=}} \bar{s}$$

These seven assumptions correspond to the seven regularity assumptions made in Section 3. As in Section 3 let $(\bar{x},\bar{n})$ denote the interior maximum at $\bar{s}$ -- i.e. --

$$(4.5)\qquad (\bar{x},\bar{n}) = \psi(s) \ .$$

Let $\xi(s)$ denote the correspondence determining the set of Congress's optimal choices of output given capacity s. By the above, ξ is single-valued for all values of s except $\bar{s}$. For smaller values it is the interior maximum and for larger values it is zero. For $\bar{s}$, both $(\bar{x},\bar{n})$ and (0,0) are optimal. Formally,

$$(4.6)\qquad \xi(s) = \begin{cases} \psi(s) & , \quad s < \bar{s} \\ \{(0,0),\ (\bar{x},\bar{n})\}, & \quad s = \bar{s} \\ (0,0) & , \quad s > \bar{s} \ . \end{cases}$$

Two additional regularity assumptions will be required for the multi-period case which were not required for the single-period case. Define the interior second-best output rate and production period in the natural way.

Definition

(i) An interior second-best value of x given (n,s) is a solution to

$$(4.7)\qquad \underset{x > 0}{\text{Maximize}} \quad V(x,n) - C(x,n,s) \ .$$

(ii) An interior second-best value of n given (x,s) is a solution to

$$(4.8) \qquad \underset{n > 0}{\text{Maximize}} \quad V(x,n) - C(x,n,s) .$$

The first additional regularity assumption is that unique second-best values exist.

<u>**Assumption 4.8**</u>

(i) A unique interior second-best value of x exists for every $(n,s) >> 0$. Let $f(n,s)$ denote this value.

(ii) A unique interior second-best value of n exists for every $(x,s) >> 0$. Let $g(x,s)$ denote this value.

Two more pieces of notation will be useful to present the second additional regularity assumption.

$$(4.9) \qquad \begin{aligned} v^f(n,s) &= V(f(n,s),\ n) \\ v^g(x,s) &= V(x,g(x,s)) \end{aligned}$$

The function $v^f(n,s)$ is the military benefit which would result if Congress was told it must produce the weapon for n years using scale s and could choose x itself. The function $v^g(x,s)$ is interpreted similarly. The second additional regularity assumption is then as follows.

<u>**Assumption 4.9**</u>

$$(4.10) \qquad (i) \quad \frac{\partial v^f}{\partial n}(n,s) > 0$$

$$(4.11) \qquad \text{(ii)} \ \frac{\partial v^g}{\partial x}(x,s) > 0$$

This assumption will first be economically interpreted. Then the mathematical reason for its necessity will be explained. Consider the following artificial problem. Hold scale fixed at some level s throughout the following. Suppose that military was told that it would be allowed to choose the annual rate of production for some new weapon. Given the rate, then Congress would choose the number of periods to produce the weapons for. What output rate would the military select, given its goal is to maximize military preparedness generated by the given weapons system? If the military choose a larger value of x in all likelihood Congress would choose a smaller value of n. Formally, it is likely that $g_x < 0$. Knowing nothing more than this, the effect of increasing x on military value is thus ambiguous. The direct positive effect of increasing x will be counteracted by the indirect negative effect of a decrease in n. Assumption (4.9) - (ii) states that the net effect is positive. That is, it states that the military would always prefer to increase the annual output rate if Congress were to delegate this decision to the military. Of course a parallel example can be used to interpret Assumption (4.9) - (i). It states that the military would always prefer to increase the number of years of production if this decision were delegated to the military.

On an economic level this assumption therefore means that the military unambiguously likes to try to lever either of Congress's two decision variables upwards. The military knows that if it can directly force Congress to raise one of the variables that Congress may compensate to some extent by reducing the other. However the compensation is only partial. The mathematical role of this assumption should now be clear. It will be assumed (in Assumption 4.10, below) that an increase in scale lowers the marginal cost of producing x and n and thus amounts to a direct effect on each variable for Congress to increase it. It is likely that both effects generate compensating reductions in the other variable. Assumption 4.9 guarantees that the direct effects dominate and it is therefore desirable to try to increase the variables.

A number of remarks should be noted about this. First, as formally stated, Assumption 4.9 is required to hold globally. Of course it is really only required to hold in the "relevant range" of values of (x,n,s). It may be that the military would not want to demand that 10,000 F16's be produced each year even if Congress would delegate this decision to the military. However it seems much more plausible that the military would select a higher rate than that currently chosen. Second, when Assumption 4.9 is not true there will still in general be an optimal scale choice from the military's point of view and the military's choice will only equal the first best by coincidence. However to prove that the military unambiguously likes to increase scale one must assume that the military unambiguously likes to try to lever Congress's choices of x and n higher since this is the economic effect of an increase in scale. Third, Assumption 4.9 is <u>not</u> sufficient to imply that Congress chooses a higher level of both x and n when s increases. It may be that one of the two variables will fall.[13] Thus the result that the military desires to increase capacity will be shown without demonstrating that Congress's choice of x and n are both monotone increasing in s. In Appendix C it is shown that a slightly different set of assumptions imply that x and y increase when s is increased. This provides the basis for a different proof of the result of this section. However, even under the assumptions in Appendix C, it is possible for n to rise or fall when scale increases. If x increases by a larger (smaller) percentage than y, then n will fall (rise). Therefore the results of Appendix C suggest that it is plausible that the effect of an increase in scale will be to raise Congress's choice of x and y; however, the effect on n may be positive or negative.

B. The Definition of Scale

In addition to the regularity assumptions, an assumption which defines the economic properties of scale must be made. This will now be presented and discussed.

Assumption 4.10

(4.12) (i) $C_{xs} < 0$

(4.13) (ii) $C_{ns} < 0$

The important point to note is that the concept of scale becomes more complicated in the generalized model of this section. Intuitively one thinks of a larger scale of plant as making it cheaper to produce larger quantities of output. However, there are two dimensions to the output decision in this model. First, one can increase the flow of output. Second, one can increase the number of periods the flow is produced for. Assumption 4.10 requires that a larger scale plant make it cheaper to expand quantity in either sense.

As in the simple model, this assumption can be interpreted as requiring that the second-best scale increase with quantity. Total differentiation yields

$$(4.14) \qquad \frac{\partial \phi}{\partial x} = - \frac{C_{sx}}{C_{ss}}$$

and

$$(4.15) \qquad \frac{\partial \phi}{\partial n} = - \frac{C_{sn}}{C_{ss}} .$$

Since the denominator is positive by the concavity assumption (Assumption 4.3), Assumption 4.10 implies that the second-best scale of plant increases if either x or n are increased.

In order to develop a better economic interpretation of scale it will be useful to create an example where C(x,n,s) is given more economic structure. Suppose in particular that C(x,n,s) is given by

$$(4.16) \qquad C(x,n,s) = N(s) + R(x,s)\,\Gamma(n)$$

where

(4.17)

(i) Γ is twice continuously differentiable over $[0,\infty)$

(ii) $\Gamma(0) = 0$

(iii) $\Gamma'(n) \geq 0$

(iv) $\Gamma''(n) \leq 0$

The function $\Gamma(n)$ is interpreted as the present discounted value of receiving one dollar per year for n years. The standard discounting formula using a constant rate of interest satisfies (4.17). The function N(s) is interpreted as the non-recurring costs. There are costs which are incurred if any production occurs regardless of the total quantity or rate of production. The function R(x,s) is interpreted as the annual recurring cost of producing x units of output.

Now Assumption 4.10 can be interpreted in the context of (4.16). First consider Assumption 4.10 - (i). From (4.16),

$$C_{ns} = R_s(x,s)\ \Gamma'(n)\ . \tag{4.18}$$

Since Γ' is positive, Assumption 4.10 - (i) is therefore equivalent to

$$R_s < 0. \tag{4.19}$$

That is, an increase in scale must lower the recurring costs of production. This is consistent with the idea that a larger scale technology is one which involves greater capital investments initially and lower variable costs each year.

Now consider Assumption 4.10 - (ii). From (4.16).

$$C_{xs} = R_{xs}(x,s)\ \Gamma(n). \tag{4.20}$$

Since $\Gamma(n)$ is positive, Assumption (4.10) - (ii) is equivalent to

$$(4.21) \qquad R_{xs} < 0.$$

That is, an increase in scale must lower the marginal recurring cost.

C. Analysis

The result that the military will increase scale as much as possible will now be proven. Since the definition of an equilibrium weapons program is essentially the same as in the previous section with (x,n) substituted for x, it will not be formally redefined here. For any $s \leq \bar{s}$ Congress chooses the output plan $\psi(s)$. Let v(s) denote the military value which results if the military chooses s. This is defined by

$$(4.22) \qquad v(s) = V(\psi(s))$$

It is clearly sufficient to show that v(s) is increasing in s. In this case the military will choose s as large as possible. This is proven below.

Proposition 4.1

$$(4.23) \qquad v'(s) > 0$$

proof:

See Appendix B.

QED.

Proposition 4.2 draws the immediate conclusion.

Proposition 4.2

The unique equilibrium weapons program is given by $(\bar{x},\bar{n},\bar{s})$.

proof:

By Proposition 4.1 the military will increase s so long as Congress is willing to choose the interior maximum output plan. The capacity $\bar{s}$ is the largest such capacity.

QED.

Proposition 4.3 sums up the welfare consequences.

Proposition 4.3

(i) The equilibrium capacity is strictly greater than the first-best -- i.e. --

$$\bar{s} > s^{*}. \tag{4.24}$$

(ii) The equilibrium output plan is second-best given the capacity -- i.e. --

$$(\bar{x},\bar{n}) = \psi(\bar{s}) \tag{4.25}$$

and

(4.26) $V(\bar{x},\bar{n}) - C(\bar{x},\bar{n},\bar{s}) \geq V(0,0) - C(0,0,\bar{s})$.

(iii) The equilibrium capacity is strictly greater than the second-best capacity given the output plan -- i.e. --

(4.27) $\bar{s} > \phi(\bar{x},\bar{n})$.

proof:

Only (iii) requires any proof. This follows from the concavity of the optimization problem as explained in Appendix A.

QED.

5. Incentives to Avoid Flexible Technologies

A. Introduction

There are at least two ways in which a technology can be flexible. First, it may be that a given plant could be easily adapted to produce a different weapons. This will be called design flexibility. Second, it may be that a given plant could produce at a wide range of output rates with very little variation in average recurring cost. This will be called rate flexibility. These are both obviously desirable properties, other things being equal. The purpose of this section is to show that the basic type of model developed in the previous sections can also be used to show that the military may strictly prefer less flexible technologies.

Section B will consider design flexibility and Section C will consider rate flexibility.

B. Design Flexibility

In Section 4 there was no need to distinguish carefully between sunk and non-sunk expenses. However, this becomes more important when one wishes to explicitly model the issue of design flexibility. This is because a plant that can be easily converted to produce a variety of weapons systems essentially exhibits very few sunk costs. Sunk facilities capital expenditures are a non-recurring cost because they must be paid even if production occurs only for one year. However, non-sunk facilities capital expenditures generate a recurring cost equal to the cost of capital each year so long as production occurs. This is because the expenditure for the program can be avoided when production ceases and the non-sunk capital is transferred to some other use.

This will now be explicitly modeled. For simplicity it will be assumed that no other sorts of non-recurring costs except facilities capital expenditures exist. Furthermore, in order to focus on the flexibility issue it will be assumed that the scale of plant remains fixed at some level s. The scale variable will be suppressed in the following notation since it is held fixed. Let F denote the facilities capital cost of the plant and let $A(x)$ denote the annual non-capital cost of producing x units.[14]

Now the nature of flexibility will be described. It will be assumed that Congress is constantly beginning a variety of other weapons programs in addition to the program being analyzed. It will be assumed that if production on the program being analyzed is cancelled, that the facilities capital of value F would save Congress $(1-\delta)F$ dollars on one of the other programs where $\delta \in [0,1]$. Thus δ is an inflexibility parameter. If δ equals 0 the facilities capital is perfectly flexible in the sense that the plant can be costlessly adapted for use on some other program. If δ equals 1, the plant has no alternate use at all. Thus the sunk facilities capital costs are δF and the non-sunk facilities capital costs are $(1-\delta)F$.

Two more simplifying assumptions will be made. First, it will be assumed that the capital does not depreciate through use. Second it will be assumed that Congress calculates the present value of costs using an instantaneous interest rate of r. Let $C(x,n,\delta)$ denote the presented discounted value of producing the output plan (x,n) given δ. It is given by

$$C(x,n,\delta) = F + A(x)\,\Gamma(n) - (1-\delta)Fe^{-rn} \tag{5.1}$$

where

$$\Gamma(n) = \frac{1-e^{-rn}}{r} \tag{5.2}$$

Note that the cost is reduced by the salvage value of the plant at the end of production. This is the third term of (5.1). Equation (5.1) can be rewritten as

$$C(x,n,\delta) = \delta F + \{r(1-\delta)F + A(x)\}\,\Gamma(n). \tag{5.3}$$

Therefore in terms of the concepts of non-recurring and recurring costs, the non-recurring costs are given by

(5.4) $$N(\delta) = \delta F$$

and the recurring costs are given by

(5.5) $$R(x,\delta) = r(1-\delta)F + A(x).$$

Thus the non-recurring costs equal the sunk facilities capital expenditures. The recurring costs equal the non-capital costs plus the annual cost of capital for the non-sunk facilities capital. The important point to note is that the non-sunk capital costs are recurring because they can be avoided by shutting down production.

Now suppose that δ is a choice parameter. In particular suppose that the military chooses δ when the plant is constructed. Then Congress chooses (x,n) given the military's choice of δ. Two points should be noted about this model.

First, the socially optimal level of δ is clearly always $\delta = 0$. This is because a greater fraction of the facilities capital can then be transferred to a new program when the old program is finished. Formally, this is true because $C_\delta > 0$.

Second, the military will have incentives to increase δ above zero. Differentiation of (5.3) yields

(5.6) $$C_{x\delta} = 0$$

(5.7) $$C_{n\delta} = -rF\Gamma'(n) < 0.$$

Thus, δ is a simple form of scale parameter. Increases in δ decrease the marginal cost of n and leave the marginal cost of x unaffected. Therefore, it is straightforward to apply the results of Section 4 and show that (given the same regularity assumptions), the military will want to increase δ as much as possible. In particular it will increase δ until Congress is indifferent between buying the program and not buying it. If Congress will buy the program for all values of $\delta \in [0,1]$ then the military will choose $\delta = 1$.

In conclusion, making a technology more flexible essentially transforms sunk costs into recurring costs. On a mathematical level this increases the marginal cost of producing for more periods and thus amounts to a decrease in scale. Thus the analysis of Section 4 shows that the military will prefer less flexible technologies so long as Congress will still buy the system.

C. Rate Flexibility

On an intuitive level one production technology can be thought of as exhibiting more rate flexibility than another one if production is efficient over a greater range of output rates For example, in Figure 5.1 the technology with average recurring cost curve AR_1 is more flexible than the technology with average recurring cost curve AR_2. Both technologies can produce the output $\hat{\hat{x}}$ at a cost of $\hat{p}$. However technology 1 can produce a range of smaller outputs at the same average cost while technology 2 cannot.

Suppose that both technologies were available to produce a weapons system and both had identical non-recurring costs. It would clearly never be strictly preferable to choose technology 2. This is because costs are never lower with technology 2 while they may be lower with technology 1. Thus technology 1 allows a more flexible response to changing conditions. The purpose of this section is to argue that the same type of model as in the previous section can be used to show that the military may well have an incentive to purposely avoid rate flexible technologies. That is, in terms of Figure 5.1, the military would strictly prefer technology 2 over technology 1.

The reason is very simple. The idea underlying the model of the previous sections is that the military can force Congress to procure a larger number of units by precommitting to a technology which penalizes Congress for procuring low quantities and rewards it for procuring high quantities. Very flexible technologies would frustrate this ability.

This idea can be very clearly illustrated by considering the following simple example. For simplicity it will be assumed that n is fixed and cannot be changed. Therefore the variable n will be suppressed and the cost curve is given by

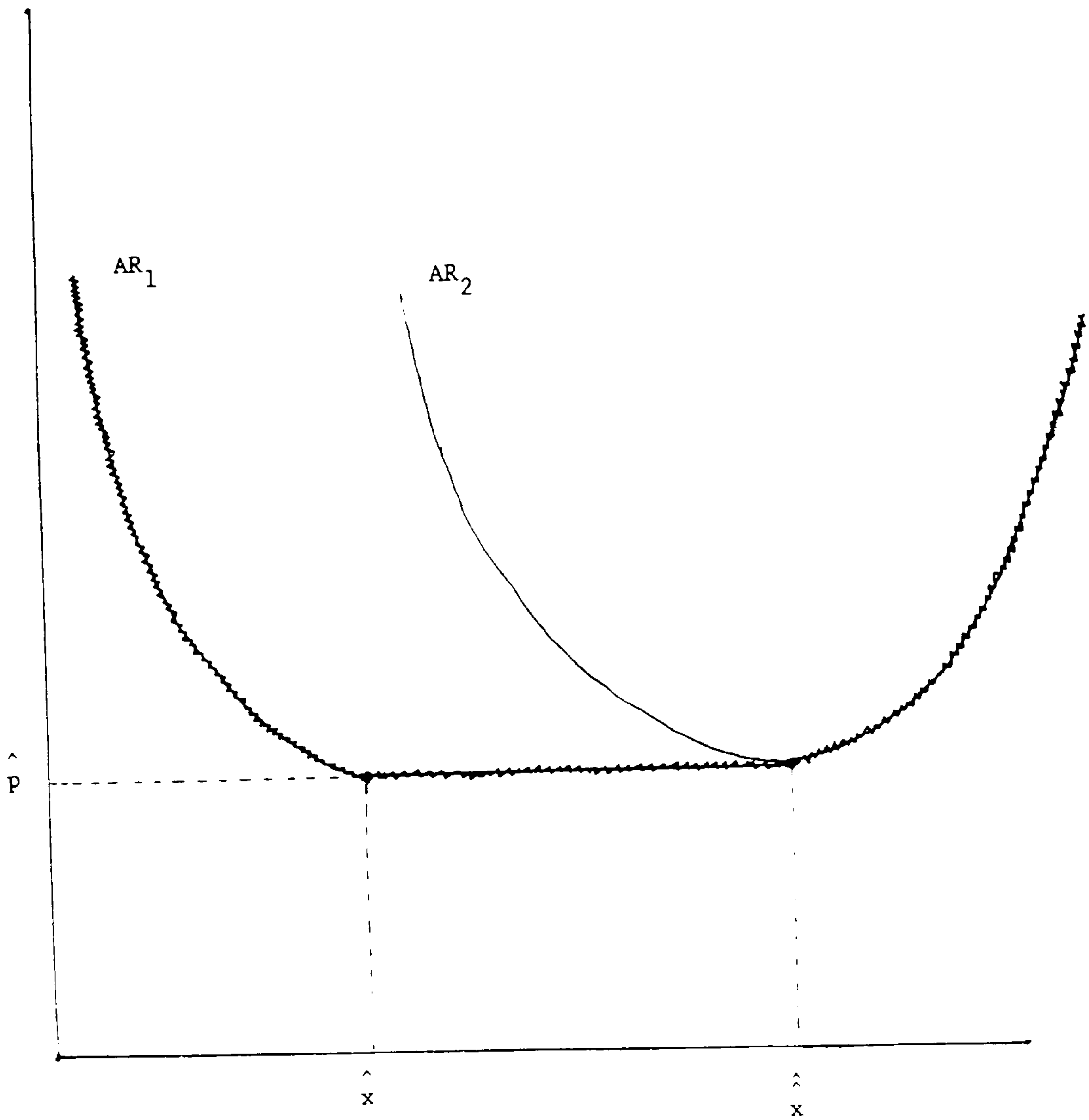
$$(5.8) \qquad C(x,s) = N(s) + R(x,s)\,\Gamma.$$

The value function is V(x). Therefore this is equivalent to the one period model of Section 3. Assume that the problem is well-behaved as modeled in Sections 3. This is illustrated in Figure 5.2. The curve V(x) denotes the value of x units and L(x) is the long run cost curve. The curve B(x) is the cost that would be incurred to produce x units if the military chose a capacity which induced Congress to select x. Formally B(x) is defined by

$$(5.9) \qquad B(x) = C(x,\psi^{-1}(x)).$$

Figure 5.1

Rate Flexibility



The first-best output is x^*, where marginal cost equals marginal cost equals marginal benefit. If the military chooses the first-best capacity, s^*, then Congress will choose the first-best output, x^*. Therefore

$$(5.10) \qquad B(x^*) = L(x^*) \ .$$

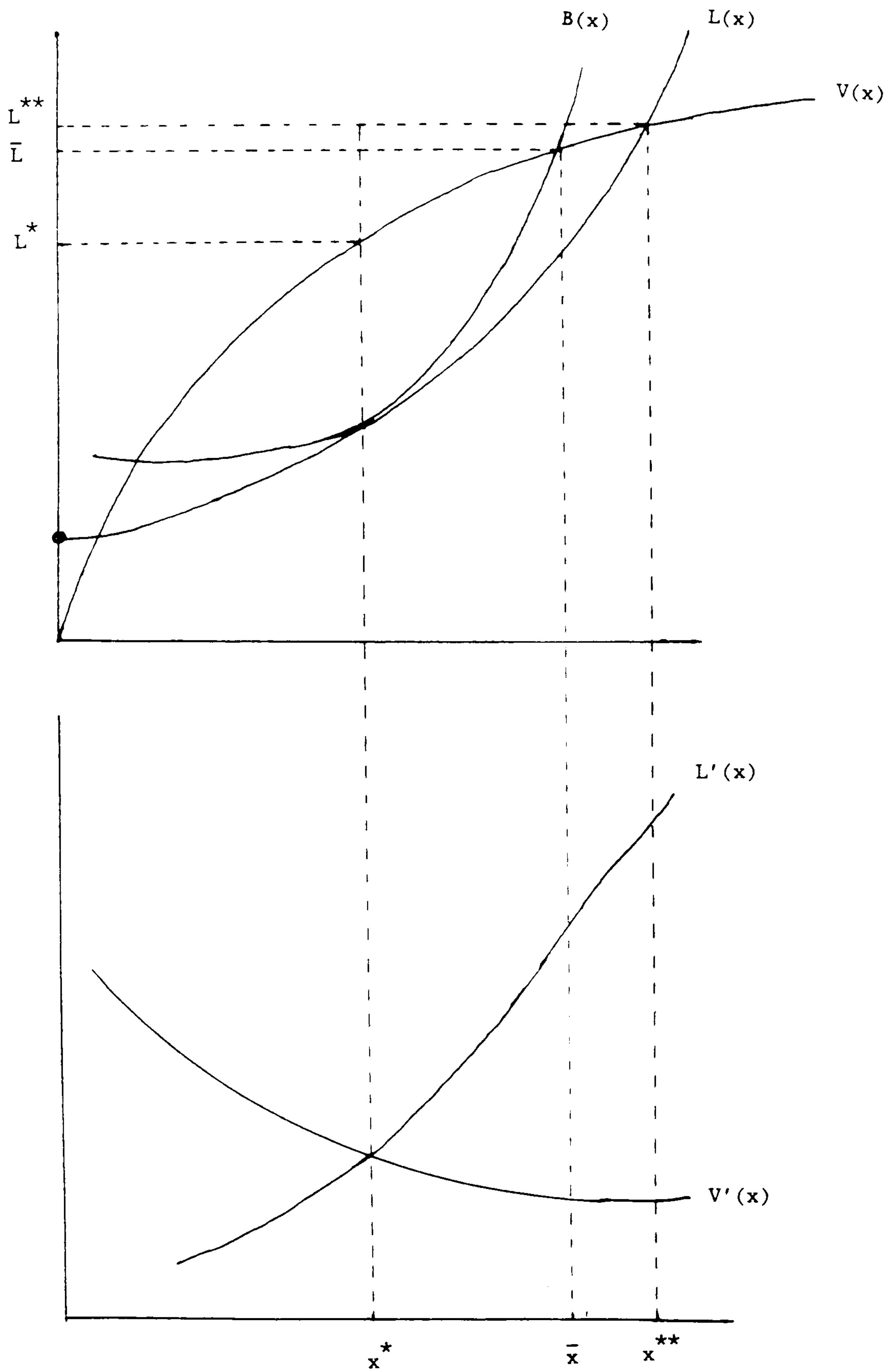
However in order to induce Congress to choose higher outputs, the military must choose higher capacities. Furthermore, as was shown in Proposition 3.3, the resulting weapons program will have a capacity greater than the second-best capacity given the output. Therefore B is above L to the right of x^*. A similar argument shows B is also above L to the left of x^*. Therefore B is tangent to L at x^* as drawn.

The equilibrium output is $\bar{x}$. This is the highest output that Congress can be induced to select because higher outputs would result in costs exceeding benefits and the entire program would be cancelled. Note in particular that it would be possible to produce greater outputs and have total costs be less than total benefits. In Figure 5.2 the point x^{**} is the largest such output. However these points are not attainable because the equilibrium weapons program involves inefficient production -- i.e. -- production off the long run cost curve. Thus, given the inefficient production, $\bar{x}$ is the greatest attainable output.

Now suppose that the military had the option of making any production technology more inflexible. Formally assume that the military can choose any cost function D(x) to present to Congress so long as there exists an s such that

Figure 5.2

Incentives To Adopt Tailor-Made Inflexibilities



$$(5.11) \qquad D(x) \geq C(x,s)$$

for every x. That is, the military selects a scale just as before which determines the best attainable technology $C(x,s)$. However it can now alter the technology to make costs of particular outputs rise if it wants to.

It is clear that the following choice is optimal for the military. Let s^{**} denote the scale of plant which is second-best given x^{**}. Then the military will choose s^{**} and make the costs of selecting any $x < x^{**}$ prohibitively high. Formally, it will choose $D(x)$ as follows

$$(5.12) \qquad D(x) = \begin{cases} \infty, & x < x^{**} \\ C(x,s^{**}) & x \geq x^{**} \end{cases}$$

Faced with this cost function, Congress will choose x^{**} units of output. This must be optimal for the military because it can never induce an output higher than x^{**}.

The welfare consequences of the above thought experiment of allowing tailor-made inflexibilities are perhaps somewhat surprising. In particular there is a Pareto-improvement since society is equally well off and the military is strictly better off. The source of this improvement is of course the removal of production inefficiency -- i.e. -- production now occurs at an efficient scale given the output selected.

The results of the above example were particularly dramatic because n was assumed to be fixed. Since tailor-made inflexibilities provide perfect control of x, the military thus had perfect control of all decision variables. Thus it was able to achieve the desired weapons program with no inefficiency.

In a model where Congress chose both n and x the results would not be so extreme since the military would still have to use s in order to influence n. Thus some inefficiency would still result. Nonetheless, the major point of this section would still be true. Namely, certain types of rate inflexibilities can be of value to the military to the extent that they restrict Congress's choices in fashions desired by the military.

6. Improving Incentives: Fixed Budgets

Perhaps the major policy implication of this paper is simply to highlight the fact that Congress cannot rely on the military to choose cost-minimizing scales for production facilities even if the military's only goal is to maximize military preparedness to the best of its ability. Therefore Congress should attempt to directly monitor this choice to the extent possible. Of course Congress has neither the time nor expertise to fully and completely monitor all aspects of the decision-making process which occurs regarding the design of production facilities. Nonetheless, Congress does perform some oversight and, naturally enough, concentrates its efforts on areas where it suspects the military will have incentives to make choices other than those Congress would make itself. An implication of this paper is that scale choice is such an area.

Since Congressional monitoring can never be perfect (or there would be no need for delation in the first place) an important issue is whether it is possible to solve this problem by adopting institutional changes which somehow reduce the military's incentives to increase scale. This is the subject of Sections 6 and 7. Section 6 describes one possible approach -- precommitment to fixed budgets. It will be argued that this is not likely to be a practical solution. Nonetheless it is very interesting to consider because it sheds light both on the logical structure of the incentive problem in the formal model and the practical problems which any proposed solution must face. This will also lay the groundwork for the analysis of an alternate approach described in Section 7 which seems more feasible.

Suppose that Congress was able to precommit to a fixed budget level for a program regardless of the scale chosen by the military. In this case the military would choose the second-best program given the budget level.[15] Therefore if Congress were to precommit to a budget exactly sufficient to fund the first-best weapons program, the military would choose the first-best program.

The above analysis makes it tempting to view precommitment to fixed program budgets as offering a complete solution to the problem of provision of excess capacity. There are a number of problems with this solution, however. First, in order to calculate the first-best budget, Congress must in general know the entire function $C(x,n,s)$. In this case Congress would not need to delegate decision-making authority to the military. It could simply instruct the military to choose s^*. Second, for a major weapons system, procurement will occur over a ten or even twenty year period. It is hard to believe that Congress could precommit ten years in advance to anything. Furthermore, many factors in the environment will change between the time capacity is chosen and the quantity decisions are made. This means that Congress would in reality have to precommit to a budget <u>rule</u> -- i.e. -- a rule describing what the budget will be each year as a function of the environmental factors. However, describing the set of all contingencies in an objectively verifiable fashion and the budget level for each one would probably be an impossible task.[16]

Thus precommitment to fixed budget levels clearly does not provide a complete solution. Nonetheless this point is still interesting for a number of reasons. First, it highlights the two key problems which any proposed solution must face -- Congress's inability to precommit and Congress's lack of information and expertise especially at the planning stage. Second, the fact that choosing a fixed budget does not generate the first-best outcome when Congress is not perfectly informed does not mean it should not be used if precommitment was possible. After all, the previous sections show that the alternative of waiting for the military to precommit to a capacity choice does not yield a first-best outcome either. This is an interesting question for future research.[17]

7. Improving Incentives: The Planning Process and Budget Projections

A. Introduction

A semi-public formalized planning process exists within the DoD which must be followed for every weapons program. In particular, a plan will always exist for any proposed or currently produced weapons system outlining future production rates and projected costs. These plans must be approved by senior officials within the DoD and the military services. Projections five years into the future are created in great detail because these are used to create a set of complete projected defense budgets five years into the future. This set of budgets and the plans supporting them are referred to as the Five Year Defense Plan (FYDP). However less detailed projections are created fifteen years into the future for use in planning purposes.[18]

A striking characteristic of this planning process is that the planned levels of procurement are almost never achieved. This is because the projected levels of production imply an agg-egate defense budget much higher than Congress would ever approve. Thus actual levels of production (both in terms of rates per year and total quantity) must invariably be scaled back at the time real budget decisions are made.

The purpose of this section is to argue that this characteristic is consistent with the theory of this paper. More importantly, it will also suggest a possible solution to the incentive problem of excess capacity through greater regulation of the planning process.

Section B will outline some of the evidence regarding the planning process. Section C will explain the result using this paper's theory. Then Section D will explore policy implications.

B. Evidence

The fact that the DoD and military services plan to procure weapons systems at rates and total quantities much higher than will ever be achieved has been widely documented. This section will describe a variety of studies drawing this conclusion.

The Air Force conducted a major internal study of its own planning procedures in the early 1980's and drew the following conclusion.

> "An historical review shows that during the last half of the 1970's, Air Force budget projections made early in the budget process were optimistic, and the level of funds ultimately appropriated were significantly below original expectations . . . Program cost growth often results directly from instability created by reductions in budget authority that detour programs from optimum schedules and

> production rates. Thus the study identified program instability as a major problem." [Air Force Systems Command 1983, pages E56-E57]

In our a study of 40 major weapons systems the Congressional Budget Office concluded the following.

> "Compared with plans established in 1983, production from 1983 through 1987 averaged about 85 percent of plans and, for many systems, amounted to two-thirds or less of plans. These slowdowns or 'stretchouts' of production occurred even though funding for the Department of Defense increased in real terms during three of the five years from 1983 through 1985." [CBO 1987a]

The Wall Street Journal printed an article in 1987 [Corrington 1987] describing the situation at that time. It clearly points out that the DoD five year plan was projecting budgets much larger than anyone believed would actually come to pass. In fact at that time, the DoD internal budget projections for planning purposes had the defense budget growing faster than the President's own proposal (for 3% annual real growth) to Congress. However most observers were predicting that even the President's proposal was much too high.

Finally see Kaufmann and Korb [1989, pages 12-15, especially Table 8] for evidence that the same type of optimistic planning is still occurring.

C. An Evaluation

Suppose that senior officials within the DoD and/or the relevant military service had decided to increase scale for a given weapons program in accord with the theory of this paper. Prior to the establishment of any production facility a detailed plan of future production rates would have to be created. Policy makers could follow one of two approaches in creating the production plan.

To describe these approaches it will be useful to recall some notation from Section 4. Recall that $(\bar{x},\bar{n},\bar{s})$ is the equilibrium program. It has the property that $\bar{s}$ is too large a scale given $(\bar{x},\bar{n})$. However there are a continuum of programs for which $\bar{s}$ would be the second-best scale. For example, we could simply increase $\bar{x}$ until the second-best scale increased to $\bar{s}$. Let $(\lambda\bar{x},\bar{n})$ for some $\lambda > 1$ be an output plan such that the second-best scale equals $\bar{s}$. The two approaches can now be explained.

<u>Approach #1</u>: Correctly predict that $(\bar{x},\bar{n})$ will be the output plan. However incorrectly state that $\bar{s}$ is the second-best scale given the output plan.

<u>Approach #2</u>: Incorrectly predict that $(\lambda\bar{x},\bar{n})$ will be the output plan. However correctly state that $\bar{s}$ is the second-best scale given the output plan.

A key difference between these approaches lies in the budget projections underlying them. Under Approach #1 the projected budget is correct. However under Approach #2, the projected budget is higher than the actual budget will turn out to be. This is because the DoD justifies a higher scale by formally predicting higher levels of production than will occur. This means that if

Approach #2 is followed that budget projections under the FYDP will be higher than actual budgets turn out to be.

The stylized facts are of course consistent with the theory that policy makers follow Approach #2. The interesting question this raises is why they do not follow Approach #1. It may be, of course, that they are totally indifferent and just happen to use Approach #2 for historical reasons. It may also be that the very naive, straightforward theory that the DoD believes it will increase its budget by asking for more than it is likely to get explains the choice. However this section will advance a slightly more subtle theory which plays a role in the policy analysis of the next section.

The theory relies on the assumption that decision-making within the DoD cannot be organized so that "secret" orders are easily transferred from senior officials to program managers without Congressional knowledge. Therefore the formal messages conveyed through the semi-public planning process are likely to have some impact on program managers' decisions. In particular, on a formal level, DoD planners ask program managers to try to create the lowest cost production facility given the anticipated production level described in the production plan for the weapon. In order to follow Approach #1, senior officials would have to send a "secret" message to the program manager, his technical staff and perhaps the technical staff of the defense contractor that the scale should actually be chosen to be larger than the cost minimizing level. The advantage of Approach #2 is that it can use the formal communication channels which exist as part of the planning process to transmit instructions regarding scale choice to the program manager and other officials involved in technical planning of the production facility. On a large program there are perhaps hundreds of people involved in designing the production facility. The public announcement of a production rate much larger then will actually be produced serves as a remarkably simple and effective public communication to all these parties telling them to design a plant of a scale larger than would be cost-minimizing given the actual production levels which will occur.

In fact the remarkable simplicity of Approach #2 might mean that it would be used even if DoD was somehow able to send "secret" orders that Congress could never gain access to. This is because "scale" in reality is not a simple one-dimensional variable. Rather, the scale of a plant is determined by a large variety of complicated engineering and design decisions. Therefore even if senior military officials were as fully informed as their technical staffs, it might be much simpler to issue the order "build a plant that would be efficient to produce $\lambda\bar{x}$ units per year" than to actually describe the engineering details of the plant. Furthermore, it is actually quite likely that senior Dod and military officials are not familiar with many of the technical details of the production technology. Thus it might be impossible for them to exactly describe the production facility. In this case the approach of simply specifying an overly large quantity and asking that the facility be designed to produce that level efficiently might be the only feasible way of asking for an overly large scale.[19]

D. Policy Implications

If the theory of Section C is correct, the DoD is to some extent a prisoner of its own formal communication channels. In particular, if it was forced to make more accurate projections of production rates, it would inevitably end up choosing more efficient scales. Therefore a possible approach for solving the problem of excess capacity would be to institute procedures which somehow required projected production levels used in the planning process to be more realistic.

In terms of formal principal agent models one can think of the above point in the following way. The simplest model would have Congress as the principal and the military as a single agent all of whose decision-making processes were internal and non-verifiable. In this case, requiring that formal planned levels of output be realistic would be of no value. Regardless of any formal statements about projected output, the agent would choose scale in a strategic fashion. In terms of the notation used above, even if the agent was required to project an output plan of $(\bar{x},\bar{n})$, it would still choose scale equal to $\bar{s}$. That is, when the agent's decision-making processes are all internal and non-observable then formal statements of plans are irrelevant. However, now suppose that the military itself consists of a principal and an agent. That is, there is a hierarchy of decision-makers. Furthermore suppose that the military agent chooses scale to be second best given the announcement of output by the military principal, and that no other secret communications are possible. Then Congress can gain leverage over the military's scale choice by controlling the formal announcement of projected production levels.

A question which should immediately arise in the reader's mind is how this proposal differs from the fixed budget proposal of Section 6. Announcing a fixed budget or a fixed production rate for a program are relatively similar policies. In particular, the same two problems exist with either policy as identified in Section 6.

(i) precommitment

Congress cannot easily precommit to any future budget or production rate.

(ii) informational requirements

Congress must not know the entire production function or else there would be no need for any delegation. However, calculation of the optimal budget or production rate requires knowledge of the production function.

First consider the precommitment question. The proposal of this section is not for Congress to precommit itself to anything. Rather it simply requires that the formal planned output rates be chosen realistically. Whether Congress can precommit to actually choose these rates is irrelevant. This is because the program-level decision-makers are assumed to be relatively non-strategic and therefore simply attempt to choose a cost-minimizing scale given the planned output levels.

Now consider the informational requirements question. There is more validity to this point. Congress should not be assumed to know the entire production function for if it did there would be no need for delegation. However, in order to calculate the optimal production rate for a program Congress must know the production function $C(x,n,s)$. There are two responses to this issue.

First, there is clearly a sense in which the proposal of this section would produce better output decisions than the solution of Section 6. This is because the proposal of Section 6 actually requires precommitment to the initial position. Thus Congress must actually implement a given budget precommitment even if its guess turns out to be wrong. However the proposal of this section simply uses Congress's guess to determine a scale choice. It is true that a bad guess will yield a bad scale choice. However the actual choice of output can be made given all information available at the time regardless of the initial guess. To put this point more simply, Congress might feel it has enough information to guess at a probable output level for the purposes of determining scale, even though it does not have enough information to predict what output it will definitely produce.

Second, one possible method of implementing the proposal of this section would be to place no restrictions at all on individual program plans. Rather Congress could simply require that when aggregated together, the program plans produced a FYDP with annual budgets which did not exceed specified levels. Presumably Congress may be able to guess at what levels of overall budget authority it will be willing or able to approve even if it cannot guess about individual programs.

To summarize the above discussion, the difference between the proposal of Section 6 and this section does <u>not</u> depend on specifying budgets vs. production levels. Either proposal could be implemented to some extent through either control device. The critical difference is that the proposal of Section 6 requires Congressional precommitment to future actions while the proposal of this section only requires Congressional control of formal messages within the defense hierarchy.

Therefore the basic proposal of this section is that Congress play a more activist role in determining the planned production rates which form the basis for scale decisions within the DoD. It could do this by one of two policies. First, it could require that the out-years of the FYDP have budget levels not to exceed pre-specified ceilings. Second, it could more actively analyze projected production plans prior to the construction of production facilities and attempt to guarantee that they are not wildly unrealistic.

E. Discussion

The idea that the DoD can some how increase its eventual budget allocations by persistently using plans which imply unrealistically high budgets is not new to this paper. In fact, this is widely accepted within the procurement community. However, the accepted explanation for why this is so is quite different from the explanation advanced by this paper. Although it is rarely explicitly articulated, the following explanation appears to be generally accepted. Convincing Congress to raise the defense budget requires the DoD to convincingly argue to Congress, the press, and the public that pressing military needs exist which require increased expenditures. Part of an effective lobbying effort then requires the DoD to prepare plans consistent with this story. Thus the preparation of unrealistic budgets is simply interpreted as a lobbying or public relations device.

If unrealistically high budget projections are simply viewed as a lobbying device there is no particular harm in allowing them to occur. In fact one might argue that allowing them is good to the extent that it allows the military to make its best case to Congress and the public. Thus the current Congressional practice of focusing primarily on the current budget year and allowing future plans to be unrealistic seems to be relatively harmless or even mildly beneficial given the generally accepted explanation.

The contribution of this paper is to suggest a different theory of why these unrealistic projections occur. Namely, high budget projections cause overly large scale decisions which in turn imply that a fully rational Congress will choose larger output levels independent of any lobbying or public relations effects.

An important implication of this paper's theory is that allowing the DoD to create plans employing unrealistically high budget projections is <u>not</u> benign or harmless. High budget projections do not simply have a lobbying effect. They affect real resource decisions today which will impact the choice set available to Congress in the future.

A number of brief concluding comments will now be noted about this result. First the two theories are not mutually exclusive. Probably both are true. However the accepted theory suggests at best mild benefits from allowing unrealistic projections while this paper's theory predicts large harm. Therefore if both theories are true, the policy implications outlined above are probably still valid.

Second, a method for empirically determining the relative significance of the two competing theories exists. The accepted theory predicts that overly large production plans should occur for programs in all stages of development. This paper's theory suggests that overly large projections should occur primarily during the period prior to construction of the production facility. Therefore if one examined the severity of the over-projections for a number of programs over time, the "average" level of over-prediction over all periods could be ascribed to the generally accepted theory, while the excess level of over-prediction (relative to the average level) for programs prior to commencement of production could be ascribed to this paper's theory.

Third, if the policy suggestion of this section were adopted and Congress began regulating the future overall budget levels and/or procurement quantities used by the military for planning purposes, this would move Congress closer to the process of multi-year defense budgeting which has been advocated elsewhere [CBO 1987a, Gansler 1989]. True multi-year budgeting requires Congressional commitment to future budgets. The idea of this section is that perhaps some of the benefits can be achieved by the more modest program of regulating projected procurement quantities.

Finally it should be stressed that this policy will only work to the extent that incentives for excess scale are felt at the "top" of the military while actual technical decisions on capacity and plant configuration are made at the "bottom" by relatively non-strategic program managers and their staff. If senior officials within the military actually make the scale decision themselves or if program managers themselves make strategic decisions then the policy of this section will be ineffective.

8. Flexible Manufacturing Technology

The automation revolution in manufacturing often referred to as flexible manufacturing technology (FMT) or computer aided manufacturing has not yet arrived in defense production. Rather, what will be referred to as the standard manufacturing technology (SMT) is still used. In an extremely insightful article two Rand researchers sum up the situation as follows:

> "Today's defense manufacturing technology is still characterized by the kind of inflexible production line, pioneered by Henry Ford, that reached maturity in World War II. This production line is set

up to produce a single design, in large quantities, over long periods of time. Although production lines have been progressively automated since 1960, the kind of automation adopted in the defense sector has done little to increase flexibility, and the procurement culture seems to have changed very little. Both the government buyer and the contractor seem to regard the specialized, optimized production line, designed for high rates of output, as the norm." [Dews and Birkler 1983, page 1]

Dews and Birkler [1983] go on to describe three key features of the new technology in more detail.

(i) It is much more efficient than the SMT for low production rates. In particular they argue that production rates on most aerospace programs are low enough that using FMT would be more efficient.

(ii) It can be easily adapted to produce different designs or products.

(iii) Efficient production can occur over a relatively broad range of output rates.

In terms of the definitions introduced in this paper, Dews and Birkler are claiming that FMT is a lower scale technology than SMT and that it exhibits greater levels of both design and rate flexibility.

As the above quote indicates, they feel that no one, including the military, seems to be in much of a hurry to employ the new technology. They do not describe any reasons for this apparent reluctance other than perhaps institutional inertia.

This paper of course supplies a theory which explains precisely why the military might be reluctant to adopt such a technology. According to this paper's theory each of the characteristics (i) - (iii) supplies a possible reason for this reluctance. First, it may be the case that plants using the FMT are of lower scale than those using the SMT. That is, it may be efficient to use plants employing the FMT for lower levels of production and plants using the SMT for higher levels of production. Sections 3 and 4 suggest that the military might well prefer to use the SMT so long as Congress would still purchase the system, even though the FMT would be more efficient. Second, design flexibility is itself a form of lower scale as explained in Section 5. Therefore the fact that the FMT offers design flexibility might well be viewed as a negative by a military planner wanting to guarantee a long production run for his program. Third, it may be the case that FMT is capable of producing even large levels of output as efficiently as the SMT. However, if the FMT exhibits more rate flexibility in the sense that it would permit Congress to reduce the quantity purchased without significantly raising average cost, then the analysis of Section 5 suggests that the military may well prefer the SMT.

The example of FMT also illustrates another important point. Namely, it is not the case that increases in scale are always synonymous with increases in capital intensity. It could easily be the case that highly capitalized automated production facilities can operate efficiently at lower rates of output than can less highly capitalized facilities using older technology. Thus the prediction of Sections 3 and 4 that the military will choose inefficiently large scales of production is not necessarily a prediction that it will choose too high a capital intensity. In fact if Dews and Birkler are correct, precisely the reverse may be true.

9. Conflicts Within the Military

The formal model of this paper assumed that the military was a single monolithic rational decision maker. In reality, of course, the military is a highly fragmented bureaucracy where individuals or groups within the bureaucracy may wield substantial power and may have goals diverging from maximization of overall military preparedness. In particular it is widely believed that the individual military services are powerful actors whose goals at least to some extent are to maximize their own contributions to military preparedness.[20] The purpose of this section is to briefly speculate on how this divergence of preferences might manifest itself in the model of this paper.

In particular, it will be assumed that there are n + 1 actors, n military services each controlling a single program and the DoD overseeing all n programs. It will be assumed that each service's goal is to maximize its own contribution to military preparedness and the DoD's goal is to maximize overall military preparedness. This is obviously somewhat simplistic but it clearly captures the basic idea that individual services are more likely to have a narrower focus on their own activities than are central decision-making authorities within the DoD.

Two extreme examples will now be considered to illustrate the range of possible outcomes. First, suppose that the overall defense budget is absolutely fixed. However, budget allocations to individual programs may vary. It is clear, based on the reasoning of Section 6, that the DoD will want to choose first-best scales for all the programs. This is because the overall defense budget is fixed and the DoD desires to maximize overall defense preparedness. However the individual services will in general want to increase the scale of their programs above the first-best. This is because they can increase the number of units purchased of their own program by doing so. Therefore, in this first extreme example, the military services have a stronger desire to increase scale because of their desire to increase their budgets at the expense of other services.

Now consider a second example where two of the services are responsible for identical programs. Therefore only one of the two programs will actually be produced. Suppose the services choose scale and suppose that their equilibrium choices can be modeled as a Nash equilibrium in scale choice. Then it is clear that the unique equilibrium will have both choose the first-best scale. However if the DoD was allowed to choose scale it would of course choose a much higher level in general. Therefore in this extreme example, competition among substitutes for the same mission results in the military services choosing lower scales than would the DoD.

These examples illustrate the two key economic factors which create a divergence in the preferences of DoD and the services. Holding the choices on all other programs fixed, the DoD will generally want to increase scale on a given program less than will the military service which runs the program. This is because the individual military service will value budget increases which occur at the expense of other services while DoD will not. However, the equilibrium scale choices made by the military services may well be smaller than the "collusive" choice which the DoD would make for all services at once.

Theoretically, either effect could predominate as the above two examples illustrate.

One final point should be noted about this result. The second example above illustrates the fact that overlapping service jurisdictions can play a beneficial incentive role to the extent that military services control the scale decision. Overlapping service jurisdictions are said to occur when two or more of the military services share and compete for responsibility for the same military function or mission. Such overlaps are widespread. Traditional military analysts have noted these overlaps and cited them as examples of inefficient organization within the military. [Stubbing 1986, chapter 7] In a world with no information or incentive problems this point would be perfectly valid. However incentive problems do exist and competition is a powerful method to improve incentives. Thus a certain amount of overlap may in fact be optimal given the existing incentive problems.

The use of overlapping service jurisdictions is therefore another method for improving incentives in addition to those described in Sections 6 and 7. However this device is already widely used and it is not clear that there is any room for further increases in inter-service overlaps. Thus this paper's argument should be viewed as an explanation of why service overlaps may exist rather than as a call for their greater use. Whether there would be a role for greater overlaps is a question beyond the scope of this paper.

10. Too Little Output or Too Much Capacity?

The fact that defense plants operate at extremely low output rates relative to their capacities is widely recognized in the defense community. Not surprisingly a large policy debate has arisen over what the correct response to this problem should be. However I believe that this debate has been largely focused on the wrong issue or at least has ignored one of the key issues. With the notable exception of Dews and Birkler [1983], which has been described above, all of the published analyses that I am aware of focus almost exclusively on the suggestion that production rates on existing plants be increased. Given a fixed overall defense budget this implies that fewer types of weapons would be purchased. For example the CBO [1987] suggested a detailed procurement program which involved increasing the production rates of some systems by cancelling others currently in production and delaying the introduction of some new systems. They summarized the basic idea motivating their suggestion as follows.

> "The military services seek to develop and acquire too many different weapons systems simultaneously. Budget limitations then force program managers to cut their annual purchases to uneconomically low quantities." [CBO 1987, page 1]

Many other studies including those by Air Force Systems Command [1983], Gansler [1989], Kaufmann [1986], and Kaufmann and Korb [1989] all contain the suggestion that weapons systems should be produced at higher rates.

This paper suggests a very different perspective on this issue. Conceptually, one can think of two problems which might exist with the procurement process. First, there may be an agency problem. By this it is meant that the military chooses scale too large in order to expand Congress's output choice as described in this paper. When no agency problem exists, the military simply chooses scale to minimize the production cost of the output plan Congress will select. The key observable outcome when this problem exists is that scale is too large given the observed output choices. That is, production occurs off the long run cost curve. Conversely, if no agency problem exists, production occurs on the long run cost curve in the sense that scale is chosen to minimize production cost. The second problem which might occur is that Congress and the military may make bad choices in the sense that they do not choose output plans which maximize military benefits given a fixed budget. This is the problem identified by the existing analyses. This problem may occur because Congress simply makes mistakes, because voting equilibria yield non-optimal results, or because individual Congressmen have goals such as maximizing employment in their district which conflict with maximizing military value. When this problem exists Congress will choose output plans which do not maximize military value given the mount of money spent.

The important point to note is that production at inefficient scales (i.e. -- off the long run cost curve) occurs if and only if there is an agency problem and is unrelated to the issue of whether Congress makes bad choices. In particular suppose Congress made bad choices but the military was a good agent. We would expect to see a bad output plan but it would be produced in plants of efficient scale.

To put this point another way, suppose existing analyses are correct and Congress does make bad decisions. However suppose by some miracle Congress suddenly begins to make good decisions. Perhaps this might mean that production rates on some existing programs would be increased as existing analyses suggest. Now suppose that time passes and new programs are begun. Suppose Congress continues to make good decisions. If an agency problem exists then we will observe excess capacity in these new programs even though Congress makes good decisions.

Therefore from the perspective of this paper, although existing analyses are focused on an important question, they are not focused on a question directly relevant to the issue of excess capacity.

Finally it should be noted that it is by no means clear that it would be desirable to choose increased production rates of a smaller number of systems given existing production facilities. The existing policy analysis often cites the following course of events as evidence the production rates are too low.

(1) Production plans are initially created for a high rate of output.

(2) A plant is built consistent with the above estimate.

(3) Congress then chooses to fund a lower rate of output and average costs are higher as a consequence.

However, this reasoning is implicitly assuming that the original production plan was somehow ideal. This paper has argued that a more straightforward interpretation of the above course of events is that the military's initial production plan is purposely chosen to be much higher than it expects production to be. This is because the original announcement's only function is to generate an overly large capacity.

Furthermore the fact that average costs would fall if output rates were increased means very little. Average cost is probably falling over the entire relevant range of production for most weapons systems. Thus average costs will be declining even under the optimal output plan. See Figure 3.2 and the discussion surrounding it for a formal discussion of this fact. To prove that the output rate is too low, one must show that the marginal military benefit of increasing output exceeds the marginal cost.[21] This of course involves the military benefits function as well as the cost function and is inherently more difficult to prove in any objective fashion.

In conclusion it may or may not be the case that it would be desirable to produce larger amounts of a smaller number of weapons systems given existing production facilities. However even if this is true, it sheds very little light on why we observe production of weapons programs systematically occurring at inefficiently high scales. The contribution of this paper to the policy debate is to highlight this as a separate issue, to describe a theory of why it occurs and to suggest a number of possible policy approaches to dealing with it.

Appendix A

Suppose that $F(x_1, \ldots, x_n, y)$ is a twice continuously differentiable globally concave function. Let x denote $(x_1, \ldots, x_n)$. Assume that a unique value of x maximizes F(x,y) for every y. Let $\phi(y) = (\phi_1(y), \ldots, \phi_n(y))$ denote this value. Assume that a unique value of y maximizes F(x,y) for every x. Let $\psi(x)$ denote this value. Finally assume that a unique maximum to F(x,y) exists at (x^*, y^*). Let $\gamma(y)$ denote the value of y which maximizes F given that x equals $\phi(y)$ -- i.e. --

$$\text{(A.1)} \qquad \gamma(y) = \psi(\phi(y)) \ .$$

Let G(y) denote the value of F(x,y) given that x is chosen optimally.

$$\text{(A.2)} \qquad G(y) = F(\phi(y), y)$$

Lemma:

G(y) is single-peaked at y^*. That is,

$$\text{(A.3)} \qquad G'(y) = 0 \iff y = y^*$$

proof:

From (A.2)

$$\text{(A.4)} \qquad G'(y) = \sum_{i=1}^{n} F_i \phi_i'(y) + F_{n+1} \ .$$

However $F_i = 0$ for $i=1, \ldots, n$. Therefore

$$(A.5) \qquad G'(y) = F_{n+1}(\phi(y), y) \ .$$

If $G'(y)$ equals zero this therefore implies that $F_{n+1}(\phi(y), y)$ equals zero. However then $y = y^*$.

QED.

Theorem:

$$(A.6) \qquad y \overset{>}{\underset{<}{=}} y^* \iff y \overset{>}{\underset{<}{=}} \gamma(y)$$

proof:

Consider some $\hat{y} > y^*$. Now suppose for contradiction that

$$(A.7) \qquad \hat{y} < \gamma(\hat{y}) \ .$$

It will be shown that

$$(A.8) \qquad G(\hat{y}) < G(\gamma(\hat{y}))$$

which contradicts the fact that $G(y)$ is single-peaked. To see this, first note that by definition

$$(A.9) \qquad G(\hat{y}) = F(\phi(\hat{y}), \hat{y}).$$

Since $\psi(\phi(\hat{y}))$ is the unique maximizing choice of y given $\phi(\hat{y})$ it must always be true that

$$(A.10) \qquad F(\phi(\hat{y}),\ \psi(\phi(\hat{y}))) > F(\phi(\hat{y}),\ \hat{y})$$

so long as

$$(A.11) \qquad \psi(\phi(\hat{y})) \neq \hat{y}\ .$$

However $\psi(\phi(\hat{y}))$ is $\gamma(\hat{y})$ by definition. Therefore (A.7) implies (A.11. Therefore (A.10) is true and can be rewritten as

$$(A.12) \qquad F(\phi(\hat{y}),\ \gamma(\hat{y})) > F(\phi(\hat{y}),\ \hat{y})\ .$$

Finally, by the definition of G(y),

$$(A.13) \qquad G(\gamma(\hat{y})) \geq F(\phi(\hat{y}),\ \gamma(\hat{y}))\ .$$

Combining (A.9), (A.12), and (A.13) yields (A.8).

Exactly the same procedure is used to derive a contradiction when $\hat{y} < y^{*}$.

QED.

Appendix B

Proof of Proposition 4.1

It is straightforward to show that Assumption 4.9 - (ii) implies that

(B.1) $$V_x(V_{nn} - C_{nn}) - V_n(V_{xn} - C_{xn}) < 0 .$$

Similarly, Assumption 4.9 - (i) implies that

(B.2) $$V_n (V_{xx} - C_{xx}) - V_x(V_{xn} - C_{xn}) < 0.$$

Now consider the first order conditions determining $\psi(s)$.

(B.3) $$V_x - C_x = 0$$

(B.4) $$V_n - C_n = 0$$

Total differentiation yields

(B.5) $$M \begin{bmatrix} d_x \\ d_n \end{bmatrix} = \begin{bmatrix} C_{xs} \ ds \\ C_{ns} \ ds \end{bmatrix}$$

where

$$\text{(B.6)} \qquad M = \begin{bmatrix} V_{xx} - C_{xx} & V_{xn} - C_{xn} \\ V_{xn} - C_{xn} & V_{nn} - C_{nn} \end{bmatrix} .$$

Rewrite (B.5) as

$$\text{(B.7)} \qquad \frac{dx}{ds} = \frac{1}{\det(M)} \left\{ C_{xs}(V_{nn} - C_{nn}) - C_{ns}(V_{xn} - C_{xn}) \right\}$$

$$\text{(B.8)} \qquad \frac{dn}{ds} = \frac{1}{\det(M)} \left\{ C_{ns}(V_{xx} - C_{xx}) - C_{xs}(V_{xn} - C_{xn}) \right\} .$$

Now, v'(s) can be written as

$$\text{(B.9)} \qquad v'(s) = V_x \frac{dx}{ds} + V_n \frac{dn}{ds} .$$

Substitute (B.7) and (B.8) into (B.9) and reorganize to yield

$$v'(s) = \frac{C_{xs}}{\det(M)} \left\{ V_x (V_{nn} - C_{nn}) - V_n (V_{xn} - C_{xn}) \right\}$$

(B.10)

$$+ \frac{C_{ns}}{\det(M)} \left\{ V_n (V_{xx} - C_{xx}) - V_x (V_{xn} - C_{xn}) \right\} .$$

The expressions in brackets are negative by (B.1) and (B.2). The values of C_{xs} and C_{ns} are negative by assumption. Finally det(M) is positive by the concavity assumption. Therefore $v'(s) > 0$.

QED

Appendix C

The purpose of this Appendix is to show that a slightly different set of assumptions can be made to prove the result of Section 4. It will be convenient to view Congress as choosing x and y instead of x and n. Then an output plan is an ordered pair (x,y) and a weapons program is an ordered triple (x,y,s). Let U(x,y) denote the benefits of the output plan (x,y) and D(x,y,s) denote the cost of the program (x,y,s). In terms of the notation of Section 4 these are defined by

$$(C.1) \qquad U(x,y) = V(x,\frac{y}{x})$$

and

$$(C.2) \qquad D(x,y,s) = C(x,\frac{y}{x},s) \ .$$

The assumptions used in this Appendix differ in two major ways from those used in Section 4. These differences will now be described. The first difference regards Assumption 4.9. A much simpler and arguably more plausible assumption can be made instead of Assumption 4.9. Recall that the problem that Assumption 4.9 addressed was that a direct increase in one of Congress's two decision variables in all likelihood would cause a decrease in the other variable. That is, x and n were in some sense substitutes. However it is probably very plausible that x and y are complements in the sense that an increase in one of the two would cause Congress to want to increase the other. That is, supposed Congress was told it must increase the total number of F16's it was going to procure. In all likelihood the optimal response would be to (weakly) increase the annual production rate. Because of this, Assumption 4.9 can be replaced by the assumption that x and y are complements in the sense described above. This will now be formally stated.

Assumption C.1

(i) Suppose a unique solution to the following problem exists for every $(y,s) >> 0$.

$$\max_{x > 0} \; U(x,y) - D(x,y,s) \tag{C.3}$$

Let $f(y,s)$ denote the solution.

(ii) f is weakly increasing in s.

Although Assumption C.1 is not mathematically weaker than Assumption 4.9 it is perhaps more plausible. At a minimum it involves observables while 4.19 involved unobservable utility levels. Furthermore, Assumption C.1 will allow one to prove monotonicity of Congress's decision variables in s. Recall that this was not possible in Section 4.

The second major difference in assumptions regards Assumption 4.10. It is replaced by Assumption C.2.

Assumption C.2

(C.4) (i) $D_{xs} < 0$

(C.5) (ii) $D_{ys} < 0$

This seems very similar to 4.10 in that both assume that the two cross partials are negative. However Assumption C.2 is actually stronger. To see this, differentiate (C.2) to yield

$$(C.6) \qquad D_{ys} = C_{ns}/x$$

$$(C.7) \qquad D_{xs} = C_{xs} - C_{ns}\ y/x^2$$

From (C.6), Assumptions 4.10 - (ii) and C.2 - (ii) are equivalent. However, from (C.7), Assumption C.2 - (i) is stronger than Assumption 4.10 - (i) given that $C_{ns} < 0$. That is, $D_{xs} < 0$ implies $C_{xs} < 0$ but the reverse is not true.

As in Section 4 it is useful to consider an example where D(x,y,s) is given more economic structure in order to provide a better economic interpretation of the assumption defining scale. Assume that C(x,n,s) is given by (4.16). Therefore D(x,y,s) is given by

$$(C.8) \qquad D(x,y,s) = N(s) + R(x,s)\ \Gamma(\tfrac{y}{x}).$$

Differentiation yields

$$(C.9) \qquad D_{ys} = R_s(x,s)\ \Gamma'(\tfrac{y}{x})/x$$

$$(C.10) \qquad D_{xs} = R_{sx}(x,s)\ \Gamma(\tfrac{y}{x}) - R_s(x,y)\ \Gamma'(\tfrac{y}{x})\ y/x^2 .$$

Since $\Gamma' > 0$, Assumption C.2 - (ii) is equivalent to

$$R_s(x,s) < 0 \tag{C.11}$$

as in Section 4. However, the assumption that

$$R_{sx}(x,s) < 0 \tag{C.12}$$

is not sufficient to guarantee that Assumption C.2 - (ii) is true. The first term of (C.10) is negative by (C.12). However, the second term is negative by (C.11). Thus their difference may be positive or negative.

The problem is that changing x while holding n fixed is very different than changing x while holding y fixed. In the letter case, an increase in x means a shorter period of production. This "extra" effect which did not exist in Section 4 complicates the nature of D_x and the effect of s on it.

It can be shown, however, that the following somewhat stronger assumption on R is sufficient for C.2 - (i) to hold. Let AR denote the average recurring cost function. The assumption is then that

$$AR_{xs} \le 0. \tag{C.13}$$

That is, the slope of the average cost curve is more negative as scale increases.

Two points should be noted about this assumption. First, (C.13) is stronger than (C.12) given that (C.11) is also being assumed. To see this, note that

$$\text{(C.14)} \qquad AR_{xs} = \frac{R_{xs}x - R_s}{x^2} .$$

Given (C.11), then (C.13) implies (C.12). Second, (C.13) is still a fairly plausible definition of scale. For example, suppose the same U-shaped average cost curve is shifted to the right and down as scale increases. Then (C.13) will be satisfied so long as the average cost curve is convex. In fact, (C.13) must be true at the minimum point of the average cost curve in order that increases in scale move the minimum point to the right. Thus, although (C.13) is clearly stronger than (C.12) it may still be plausibly satisfied by "well-behaved" cases.

Proposition C.1 will not show that (C.13) is sufficient for Assumption C.2 - (i) to be true.

Proposition C.1

Suppose that (C.11) and (C.13) are true. Then Assumption C.2 - (i) is true.

proof:

The assumptions in (4.14) imply that

$$\text{(C.15)} \qquad \frac{\Gamma'(n)n}{\Gamma(n)} \leq 1.$$

Substitution of (C.15) into (C.10) therefore implies that

$$D_{xs} \le \Gamma(\tfrac{y}{x})\left\{\frac{R_{sx}(x,y)x - R_s(x,y)}{x}\right\} . \tag{C.16}$$

Then (C.14) implies the result.

QED.

This completes the discussion of the two major changes in assumptions. In summary the approach of this appendix allows one to replace Assumption 4.9 by a (perhaps) more plausible and simpler assumption. However it also requires a stronger assumption about the nature of scale. Thus the approach of this appendix is not clearly more or less general than that of Section 4.

Proposition C.2 will now present the monotonicity result. For purposes of presenting the proposition the necessary regularity properties will be presented as Assumption C.3.

Assumption C.3

(i) V and C are twice continuously differentiable over the strictly positive orthant.

(ii) The function

$$V(x,y) - C(x,y,s) \tag{C.17}$$

is strictly globally concave over $(x,y) \in (0,\infty)^2$ for every s.

(iii) A unique maximum over $(o,\infty)^2$ exists to (C.17) for every $s > 0$.

Let

$$\psi(s) = (\psi_x(s), \psi_y(s)) \tag{C.18}$$

denote this interior maximum.

Proposition C.2

Given Assumptions C.1 - C.3 then $\psi_x(s)$ and $\psi_y(s)$ are increasing in s.

proof:

The first order conditions are

$$V_x - C_x = 0 \tag{C.19}$$

and

$$V_y - C_y = 0 . \tag{C.20}$$

Totally differentiate and reorganize to yield

$$\begin{bmatrix} d_x \\ \\ d_y \end{bmatrix} = \frac{1}{D} \begin{bmatrix} (V_{yy} - C_{xy}) & -(V_{xy} - C_{xy}) \\ \\ -(V_{xy} - C_{xy}) & (V_{xx} - C_{xx}) \end{bmatrix} \begin{bmatrix} C_{xs} d_s \\ \\ C_{xy} dy \end{bmatrix} \tag{C.21}$$

where D is the determinant of

$$\text{(C.22)} \qquad \begin{bmatrix} V_{xx} - C_{yy} & V_{xy} - C_{xy} \\ V_{xy} - C_{xy} & V_{yy} - C_{yy} \end{bmatrix} .$$

By Assumption C.3 - (ii), D is positive and the two diagonal terms of the matrix in (C.21) are negative. By Assumption C.2 the two terms C_{xs} and C_{xy} are negative. Therefore the result is true if the off-diagonal terms of the matrix can be shown to be non-positive. This follows from Assumption C.1.

QED.

It is now straightforward to complete the analysis and show that the military will increase capacity until Congress is indifferent between accepting or rejecting the entire program. The equilibrium capacity level is greater than the first-best level. By Assumption the equilibrium output plan is second-best given the capacity. However, by using the proposition in Appendix A, it is immediate that the equilibrium capacity is too large relative to the second-best capacity.

References

Air Force Systems Command [1983], The Affordable Acquisition Approach, Andrews Air Force Base, Maryland.

Baron, David and David Besanko [1987], "Monitoring, Moral Hazard, Asymmetric Information, and Risk Sharing In Procu ment Contracting," Rand Journal of Economics, 18, 509-532.

Carrington, Tim [1987], "The Pentagon's Spending Specialists Boldly Go Where Congress Isn't Remotely Likely to Follow," Wall Street Journal, August 21, page 32.

Congressional Budget Office [1987a], Assessing the Effectiveness of Milestone Budgeting.

Congressional Budget Office [1987b], Effects of Weapons Procurement Stretch-Outs on Costs and Schedules.

Dews, Edmund and John Birkler [1983], "Reform In Defense Acquisition Policies: A Different View," Rand Paper P-6927, The Rand Corporation, Santa Monica, California.

Gansler, Jacques [1987], Affording Defense, MIT Press, Cambridge, Massachusetts.

Kaufmann, William [1986], A Reasonable Defense, The Brookings Institution, Washington, D.C.

Kaufmann, William and Lawrence Korb [1989], The 1990 Defense Budget, The Brookings Institution, Washington, D.C.

Laffont, Jean-Jacques and Jean Tirole [1986], "Using Cost Observation to Regulate Firms," Journal of Political Economy, 94, 614-641.

Lewis, Tracy [1986], "Budget Competition in Project Funding," mimeo, UC Davis.

McAfee, R.P. and J. McMillan [1986], "Bidding for Contracts: A Principal Agent Analysis," Rand Journal of Economics, 17, 326-338.

Riordan, Michael and David Sappington [1986], "Designing Procurement Contracts," mimeo, Boston University.

Rogerson, William P. [1989], Reimbursement of Defense Contractors For Facilities Capital Investment: Theory and Practice," mimeo, Northwestern University.

Rogerson, William P. [1990]. "Quality vs. Quantity in Military Procurement," American Economic Review, forthcoming.

Stubbing, Richard [1986], The Defense Game, Harper and Row, New York.

Weitzman, Martin [1974], "Prices vs. Quantities," Review of Economic Studies, XLI, 477-491.

Notes

1. A "minimum economic rate" was calculated for each system given the existing production facility. Twenty of the 40 systems were being produced at less than this rate.

2. Because of the need for surge capacity (i.e. -- extra capacity to be used in the event of war), it may be that the optimal scale is somewhat larger than that which would minimize production cost of the planned peacetime rate. The above studies conclude that capacity is too large even considering this factor.

3. This solution is advocated by Air Force Systems Command [1983], CBO [1987b], Gansler [1989], Kaufmann [1986], and Kaufmann and Korb [1989]. A notable exception, however, is Dews and Birkler [1983]. Their conclusions will be described in more detail in Section 8.

4. See, for example, Baron and Besanko [1987], Laffont and Tirole [1986], McAfee and McMillan [1986], and Riordan and Sappington [1986].

5. An exception is Lewis [1986] which considers a model of Congressional-military interactions. The model is focused on different issues than the model of this paper.

6. As will be the case for all the formal models of this paper, the above simple graphical presentation ignores the issue of surge capacity -- i.e. -- the need for extra capacity to be used in the event of war. It is straightforward to adapt the arguments of this section to a more realistic model where this is considered. The simpler model is used primarily for expositional clarity.

7. CBO [1987b], page 12.

8. The three discarded systems are the M1 tank, the Bradley fighting vehicle, and the Multiple launch rocket system. The two missile systems still in low rate initial production at the time of the CBO study were the Tomahawk and AMRAAM.

9.CBO [1987b], page 9.

10. CBO [1987b], page 12.

11. After reading Section 4 it will be apparent to the reader that the model of this section can also be interpreted as one where Congress chooses an annual rate of production and the number of years of production is held fixed.

12. Throughout this paper the notation $v >> 0$ will be used to denote that all elements of the vector v are strictly greater than zero.

13. It is straightforward to show that the concavity of the problem implies that at least one of the two variables must rise when s rises.

14. If the scale variable was not being suppressed we would write F(s) and A(x,s).

15. Formally, a second-best program given the budget level B is a solution to the following problem

$$\max_{x,n} \quad V(x,n,s)$$

$$\text{Subject to} \quad C(x,n,s) \leq B$$

One coula also specify the budget on a year-by-year basis and modify the above definition in the obvious way.

16. See Rogerson [1990] for a formal model of this idea.

17. Perhaps an analysis similar to Weitzman's [1974] prices vs. quantities paper could be done.

18. Gansler [1989], page 98.

19. This latter argument assumes that military decision-makers are not fully informed about the production function. Thus they would "guess" at the optimal scale choice given their limited information. This is a slight departure from the formal model of previous sections which assumed that military decision-makers w