SHOULD A PRODUCTION RATE VARIABLE BE INCLUDED IN MILITARY AIRCRAFT LEARNING CURVES? EMPIRICAL EVIDENCE FROM THE UNITED STATES AIR FORCE

Caleb J. AHERN, Jonathan D. RITSCHEL, Edward D. WHITE, Brandon M. LUCAS, Robert D. FASS, Shawn M. VALENTINE

Department of Systems Engineering and Management
Air Force Institute of Technology, Dayton Ohio, United States of America

Learning curves are a core analytical method employed by cost analysts to estimate weapon system production costs. This study examines United States Air Force aircraft programs and compares the traditional (e.g., Unit Theory) learning curve model to a production rate learning curve model. While there are some previous studies examining production rate models, one novelty of this research is the size of the dataset, which comprises the largest military examination to date. The results suggest the production rate model outperforms the traditional learning curve model. Additionally, the analysis identifies the post-Initial Operational Capability (IOC) time period as the preferred milestone in the life-cycle to employ the production rate model.

Key words: *learning curve, production rate, cost estimating, cost modeling, resource allocation, defense budgets*

1. INTRODUCTION

Defense cost analysts employ a multitude of techniques to estimate the cost of a weapon system. One of the most widely accepted and utilized techniques is the learning curve. Learning curves are traditionally used to estimate recurring costs in a production process (Mislick and Nussbaum, 2015). While learning curves have been studied along a multitude of dimensions (Boone et al., 2021; Moore et al., 2022), empirical examinations incorporating a production rate (PR) variable in defense programs is sparse. The purpose of including a PR variable is to capture cost reductions that are realized through economies of scale (Government Accountability Office, 2020). However, the mixed results in the extant literature has led to some debate on whether the PR variable should be employed in cost estimates. Therefore, we examine the evidence in United States Air Force (USAF) aircraft to shed light on the issue.

To the best of our knowledge, this study utilizes the largest USAF learning curve dataset ever collected. The robust dataset enables comparisons between traditional (e.g., Unit Theory) learning curve models and a PR learning curve. Thus, the purpose of this article is two-fold: 1) To discern *if* a PR model is preferred in USAF programs and 2) To determine *when*, in the acquisition process timeline, a PR model should be employed.

Military aircraft are expensive. Unfortunately, cost growth has historically plagued these programs (Jones et al., 2023). Improving the toolkit for defense cost analysts is one small step towards better estimates and reducing future cost growth. This has implications for defense resource management in the form of better informed decisions and the resultant improvement in resource allocation.

2. LEARNING CURVES AND THE PRODUCTION RATE MODEL

Learning curves are routinely used to estimate recurring manufacturing costs in a multitude of industries (Womer, 1979). Defense cost analysts widely adopted these learning curve methodologies after Wright's (1936) and Crawford's (1947) seminal studies.

Wright's (1936) analysis of World War I aircraft production costs revealed a mathematical relationship between the quantity produced and the amount of labor hours (or cost) needed to complete the task. More specifically, Wright found that as the quantity of units produced doubled, the *cumulative average cost* decreased by a constant percentage. This insight became known in the learning curve vernacular as Cumulative Average Theory or the Wright Curve.

Crawford (1947) subsequently found a similar relationship in his study using World War II aircraft production costs. The USAF commissioned his study to validate Wright's theory (Mislick and Nussbaum, 2015). Crawford found that as the quantity of units produced doubled, the *individual unit cost* decreased by a constant percentage. This insight became known in learning curve vernacular as Unit Theory.

Mathematically, the formulation of Wright's and Crawford's curves are the same. Differences lie solely in the interpretation of the variables themselves. This paper employs Unit Theory (or the Crawford curve) as it is the more common technique used by US defense cost analysts in estimating recurring aircraft costs. Equation 1 denotes the traditional Unit Theory model.

$Y = AX^b$ Equation 1

Where:

Y = the cost of unit X

A = the theoretical cost of unit one (T1)

X = the unit number

b = a constant representing the slope of the learning curve

Learning curves capture the expected unit cost decrease as additional quantities are produced. But it is also reasonable to expect unit costs to decrease as the production rate increases (Government Accountability Office, 2020). This is known as the PR effect.

The PR effect is in addition to the learning curve effect. It captures cost reductions realized through economies of scale (Government Accountability Office, 2020). Banks et al. (2016) attribute the PR effect to large fixed costs inherent in defense systems. The Government Accountability Office (2020) identifies specific examples where economies of scale can occur, such as quantity discounts, reduced ordering, processing, shipping, receiving, and inspection costs.

Regardless of the specific source, the impact of the PR effect is similar to the learning curve effect in that unit costs decrease as quantities increase. Where the PR effect differs from learning curves is that it lacks "memory" (Banks et al., 2016). In other words, the PR only affects a specific lot's unit cost. The PR impact does not carry over to the next production lot (Banks et al., 2016).

Large et al. (1974) is the first known defense report that explicitly models the PR effect. Their equation has been replicated by multiple outlets since that time, with slight changes in the PR variable nomenclature. See Equation 2.

$Y = AX^bR^c$ Equation 2

Where:

Y = the cost of unit X

A = the theoretical cost of unit one (T1)

X = the unit number

b = a constant representing the slope of the learning curve

R = production rate (quantity per time period or lot)

c = rate coefficient

In summary, Figure 1 depicts the traditional Unit Theory learning curve on the left and the PR model on the right. Incorporation of the rate variable results in lower costs, as the slope of the curve is steeper. The empirical comparison of these two models with USAF data is the genesis of this paper.

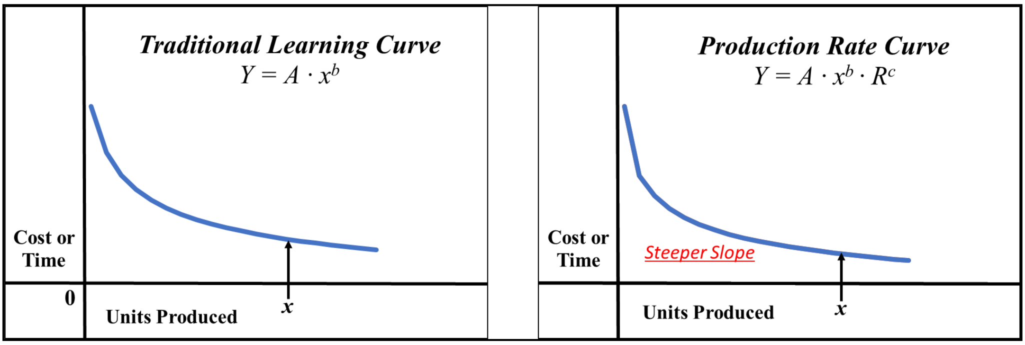

Fig. 1 Unit Theory and PR Model

2.1. Previous PR Studies

Although Large et al. (1974) are not the first to discuss the PR theory, their mathematical modeling of the concept consistently appears in the literature as a seminal contribution. Somewhat ironically, Large et al. (1974) found that a generalized estimating equation was not reliable in their dataset of 29 aircraft. Rather, while they believed the PR theory generally holds, they concluded that in any specific instance the effect depends upon the circumstances leading to a rate change. As a result, they suggest that each case must be evaluated individually prior to usage.

In contrast, subsequent studies by Smith (1976) and Congleton and Kinton (1977) both found statistically significant relationships between production rate and direct labor hours. Both of these studies were limited in scope, but they inspired Bemis (1981) to examine the PR model further.

Bemis (1981) used Selective Acquisition Reports (SARs) from the 1950s to 1970s for his investigation. His empirical analysis on historical and on-going aircraft resulted in a recommendation for PR to be considered in defense cost estimates.

Moses (1990) acknowledged these inconsistent findings within the literature and attempted to discern a set of conditions under which a PR model is preferred. He found neither model dominated under all conditions. Rather, the PR model outperformed when there were large fixed costs, a growing production rate trend, or large variability in period-to-period production rates.

A more recent study by Arena et al. (2008) found some evidence of higher production rates reducing costs in fixed-wing aircraft. While Banks et al. (2016) noted that PR effects were strongest in high quantity systems such as missiles. Notably, their study was limited by the small dataset of 11 defense systems. Lastly, Boone et al. (2021) were not directly testing for a PR model, but their findings are still relevant to this research. Boone's suggested modification to the Wright curve results in a steeper slope during early production lots. This idea of a steeper curve is commensurate with the PR effect.

In summary, the previous PR study results are mixed. The majority of studies point to PR being an important consideration in future cost estimates. But many of these studies suffer from small datasets and/or the analyses were conducted decades ago. This endeavor seeks to fill that gap by utilizing the most robust, modern military aircraft dataset collected to date.

3. DATA AND METHODS

3.1 Data

The data is sourced from Contractor Cost Data Summary Reports (CDSR), or DD 1921-2s (progress curve reports), through the Cost and Economics Division (FZC) of the Life Cycle Management Center (LCMC) at Wright-Patterson Air Force Base, Ohio. The 1921-2 report captures recurring costs by unit or lot for selected reporting elements (Department of Defense, 2021). The original dataset included 158 aircraft programs with 813 production lots. Table 1 shows the final dataset after applying an exclusion criteria screening. The red font indicates removal of lots or programs.

Table 1 Dataset

	Number of Lots	**Number of Programs**
Original Dataset	813	158
Less than 3 Units per Lot	46	0
Missing Data	49	26
Lack of Sequential Lots	125	56
Non-Aircraft	53	8
Final Dataset	**540**	**68**

The first exclusion criterion removes aircraft that had less than three units in a production lot. This criterion is consistent with Arena et al. (2008) who suggest a minimum lot size of three is needed to consider a PR variable as part of the learning curve model. The second exclusion criterion removes 26 programs that have missing data. The third criterion removes 56 programs that did not have sequential production lots or had less than 4 sequential lots. Sufficient sequential lots are necessary to model learning curves. The fourth and final exclusion criterion removes eight non-aircraft programs. These rotary wing and missile systems are outside the aircraft-centric focus of this analysis. The final dataset, therefore, consists of 68 programs comprised of 540 lots.

The 1921-2s report actual expenditure costs. These are referred to as Then Year expenditure (TY exp) dollars (Office of the Secretary of Defense, Cost Assessment and Evaluation [OSD-CAPE], 2021). Best practices necessitate normalization to Constant Price (CP) dollars prior to modeling and analysis (OSD-CAPE, 2021). Thus, the OSD published Raw Inflation Rates for aircraft procurement (3010 appropriation) were utilized to convert the 1921-2 TY exp data into CP dollars.

3.2 Method - Comparing Models

The first investigative question seeks to determine if a PR model is preferred to the traditional learning curve. To discern this requires several steps. First, the *traditional learning curve model* ($Y = AX^b$) from Equation 1 is transformed into log-log space and Ordinary Least Squares (OLS) regression is performed on each program. Because the 1921-2 data is in lot format, rather than individual units, the Y variable is the Average Unit Cost (AUC) and the X variable is the Lot Midpoint.

Next, the Absolute Percentage Error (APE) is calculated from the regression model output. The APE is calculated as shown in Equation 3. Note that the predicted AUC comes from the regression model while the actual AUC originates from the dataset.

$$APE = \left| \frac{AUC\ (Actual) - AUC\ (Predicted)}{AUC\ (Actual)} \right| \quad \textbf{Equation 3}$$

The APE data is then used to calculate a median APE (MdAPE) and mean APE (MAPE). These metrics will be used in subsequent tests (to be discussed) when comparing the traditional and PR models.

Lastly, the OLS regressions, APE, MdAPE, and MAPE calculations are duplicated for each program, but this time employing the *PR model* ($Y = AX^bR^c$) from Equation 2. With both the traditional and PR model data compiled, testing between the two approaches can occur.

Model comparisons are conducted via two methods. The first method utilizes a confidence interval. The MdAPEs from the traditional and PR models are differenced. Then a histogram is created from the differenced MdAPEs and its' associated confidence interval is calculated. If the confidence interval does not contain zero, then the traditional and PR models are different.

The second method employs the Wilcoxon Signed Rank test. The Wilcoxon Signed Rank test is a non-parametric test for matched or paired data like the Sign test, but also considers the magnitude of the observed differences (LaMorte, 2017). Like the Sign test, the hypothesis for the Wilcoxon Signed Rank test uses the sample's locations (often interpreted as the median for similar distributions) of the difference scores. The hypothesis for the Wilcoxon Signed Rank test is:

H_0: the median of the population of differences between the paired data = 0

H_a: the median of the population of differences between the paired data ≠ 0

In other words, the Wilcoxon Signed Rank test determines whether there is a statistical difference between the traditional and PR models.

3.3 Method – Acquisition Timeline

The second investigative question seeks to determine when, in the acquisition process timeline, the PR model should be employed. The Initial Operational Capability (IOC) date is chosen as the point of demarcation for the analysis. The IOC was chosen because it is a focal point in the acquisition timeline. IOC denotes the point where a system is ready to perform its intended mission in an operational environment with an initial quantity of assets (AcqNotes, 2023).

Table 2 illuminates the mapping process employed to answer this question. The lot where the PR model becomes statistically significant is mapped in 10 percent increments (positive or negative) from the IOC date.

Table 2 IOC% Mapping Example

IOC% Model Example (0% is IOC)							
IOC %	**-30**	**-20**	**-10**	**0**	**10**	**20**	**30**
Program #1	X						
Program #2				X			
Program #3		X					
Program #4					X		
Program #5				X			
Program #6		X					
Program #7							X
Sub-Total	**1**	**2**	**0**	**2**	**1**	**0**	**1**

A histogram is then compiled from the data in Table 2. Descriptive statistics from the histogram provide the mean, median, and 95% confidence interval. Lastly, a Wilcoxon Ranked Sign test is performed to determine whether the PR model occurs at IOC vice a point in time before (or after) IOC.

4. RESULTS

4.1 Comparison of Traditional and PR Models

There were 68 programs modeled with the traditional, $Y = AX^b$ learning curve construct. The data was transformed into log-space and OLS regression was run. Each program's prediction value for the dependent variable was then compared to its original value to calculate the APE within each program. Next, the process was duplicated for the PR model. A sample from the calculations are shown in Table 3.

Table 3 Sample Calculations

Program	Traditional		PR			
	MAPE	MdAPE	MAPE	MdAPE	MAPE Difference	MdAPE Difference
Pgm. 1	1.8%	1.6%	1.8%	1.7%	**7%**	**0%**
Pgm. 2	4.3%	2.4%	4.4%	2.3%	**-7%**	**2%**
Pgm. 3	2.0%	1.9%	1.7%	1.8%	**-6%**	**-17%**
Pgm. 4	5.3%	4.0%	3.5%	3.1%	**-23%**	**-33%**
Pgm. 5	2.1%	2.1%	2.2%	1.9%	**-6%**	**1%**

Analyzing the full set of 68 programs showed that the PR model reduced the MdAPE by an average of 10% compared to the traditional learning curve model. The associated confidence interval had a lower bound of -19% and an upper bound of -1%. Note that the confidence interval did not contain zero, suggesting that there are differences between the models. Table 4 has these summary statistics for the MdAPE. An examination of the MAPE has similar results. The PR model reduced the average MAPE by 23% in comparison to the traditional learning curve model.

Table 4 MdAPE Differences

N	Mean	Std Dev	Lower CI	Upper CI
68	-10%	38%	-19%	-1%

Next statistical significance is discerned via the Wilcoxon Ranked Sign test. The null hypothesis is that there is no difference between the traditional and PR curve, while the alternative states a difference exists. An alpha of 0.05 is used. See Figure 2.

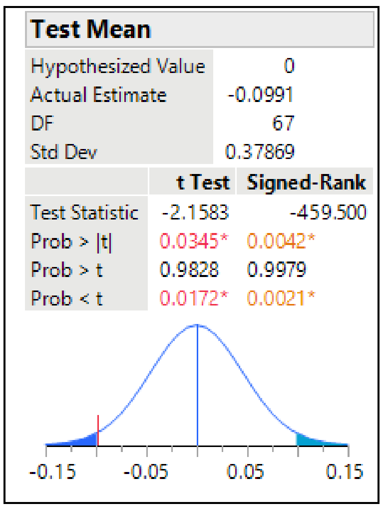

Fig. 2 Wilcoxon Test of MdAPE Differences

The two-tailed test returns a p-value of 0.0042. This result is well below the 0.05 threshold. The evidence suggests that there is a statistically significant difference between the PR and traditional models. Combining the results of the MdAPE analysis with the Wilcoxon results indicates that a PR effect is present in USAF military aircraft. Therefore, USAF cost analysts should consider employing a PR variable in their models.

4.2 The Acquisition Timeline

With the appropriateness of a PR model established, the next question is *when* in the acquisition timeline it is best utilized. This is determined by mapping the lot number where the PR model became preferred to the program's IOC date. IOC was chosen because it is a focal point in the acquisition process.

IOC is mapped in intervals of +/-10%. This mapping process necessitated development of a mapping rule. If a program falls between the interval 0.01 and 4.99 it maps to the smaller interval (e.g. 24.68% is in the 20% interval); alternatively, a program between 5.00 and 9.99 maps to the larger interval (e.g. 25.10 is in the 30% interval).

The 68 programs from Table 1 comprised the initial dataset. However, several data issues arose. These included unavailable IOC dates and IOC dates that were outside the production window. After excluding programs with these issues, the final dataset consisted of 36 programs.

Figure 3 shows the resultant histogram from the mapping process. Recall that zero denotes the IOC data and the programs are binned in +/-10% increments. Results show that the median PR value is ~20% after IOC, with a mean of 13.8% post IOC. The associated 95% confidence interval ranges from 2% to 26% past IOC.

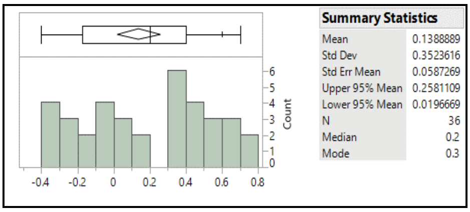

Fig. 3 Histogram of IOC%

Next a Wilcoxon Ranked Sign test is performed to determine if the PR occurs at IOC vice a point in time past (or before) IOC. An alpha of 0.05 is used. Figure 4 shows a p-value of 0.0285, which indicates there is a statically significant difference.

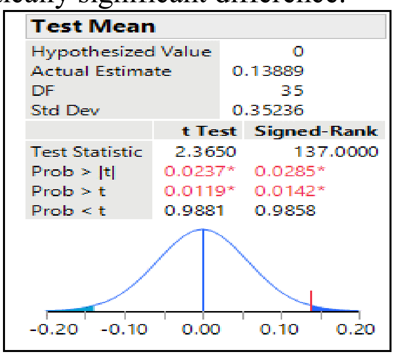

Fig. 4 Wilcoxon Ranked Sign Test of IOC%

The results from Figure 3 and 4 suggest that including a PR variable in the learning curve model should occur once the production process has ramped up and manufactured the initial quantity necessary to attain an IOC. To be conservative, defense cost analysts are cautioned that ~20% past IOC is where economies of scale are expected to be realized.

5. CONCLUSION

Learning curves are a core analytic method employed by defense cost analysts. They are ubiquitously recommended in defense cost estimating textbooks, manuals, guides, and best practices (Mislick and Nussbaum, 2015; Government Accountability Office, 2020; OSD-CAPE, 2022). Published support for modifying this core technique to include a PR variable that captures economies of scale has grown over time (Kunc et al., 2018; Government Accountability Office, 2020; OSD-CAPE, 2022). Yet previous empirical investigations into the efficacy of the PR modified model have been sporadic and are mired in small-scale datasets.

This paper sought to resolve the ambiguity regarding practical applicability of the PR model in military aircraft. Employing the largest known dataset to data, the results provide optimism that the PR model can be utilized. The analysis of 68 USAF military aircraft suggests the PR model is preferred to the traditional model. More specifically, the MdAPE error was reduced an average of 10% by utilizing the PR model. Due to the large production cost of military aircraft, the impact of a 10% deviation is not trivial. It often translates to millions of dollars per aircraft. These impacts pertain to the recurring flyaway cost level of analysis.

While the results are promising, caution is also advised due to several limitations. First, the dataset did not allow for analysis of different aircraft platform types. In other words, it could be that bombers, fighters, trainers, or cargo planes have unique characteristics that impact the generalized results. Second, there are likely to be unique, program specific effects that have impacts on recurring costs. These include contract specifications, tooling, or labor considerations and they should be deliberated upon prior to the decision to employ a PR model in favor of a traditional learning curve model. In other words, practitioners should take care to know their program first, and incorporate those unique characteristics in their estimating methodology. That caution is not unique to this analysis, rather it represents best practices in the cost estimating community.

In summary, ameliorating cost growth in defense programs is only achievable when cost estimators are equipped with the best available tools. The findings of this paper support guidance suggesting a PR variable is important to consider when developing an aircraft learning curve model. Additionally, practitioners are advised to only employ the PR model post-IOC. These enhancements to current practices can lead to better informed decisions and improved resource allocation in the defense arena.

REFERENCES

[1] AcqNotes (2023) Acquisition Process: Initial Operational Capability. Available at: https://acqnotes.com/acqnote/acquisitions/initial-operational-capability

[2] Arena M, Younossi O, Brancato K, Blickstein I and Grammich C (2008) *Why Has the Cost of Fixed-Wing Aircraft Risen? A Macroscopic Examination of the Trends in US Military Aircraft Costs Over the Past Several Decades*. RAND report MG696, Santa Monica, USA.

[3] Banks W, Cosgray R and Lawless (2016) Observation of Production Rate Effect in Historical Costs of Weapon Systems. ICEAA Online. Available at: http://www.iceaaonline.com/wp-content/uploads/2016/06/AC01-ppt-Observation-Historical-Weapons.pdf

[4] Bemis J (1981) A Model for Examining the Cost Implications of Production Rate. *The Journal of Defense Systems Acquisition Management* 4(2): 84-94.

[5] Boone E, Elshaw J, Koschnick C, Ritschel J and Badiru A (2021). A Learning Curve Model Accounting for the Flattening Effect in Production Cycles. *Defense Acquisition Research Journal*, 28(1): 72-98.

[6] Congleton D and Kinton D (1977) *An Empirical Study of the Impact of a Production Rate Change on the Direct Labor Requirements for an Airframe Manufacturing Program*. Master's Thesis, Air Force Institute of Technology, USA.

[7] Crawford J (1947) Crawford-Strauss Study. Report for the Air Material Command, Dayton Ohio, USA.

[8] Department of Defense (2021) DoD Manual 5000.04: Cost and Software Data Reporting, 7 May 2021.

[9] Government Accountability Office (2020) *Cost Estimating and Assessment Guide: Best Practices for Developing and Managing Program Costs*. GAO-20-195G, Washington DC, USA.

[10] Jones S, White E, Ritschel J and Valentine S (2023) Schedule and Cost Estimations Through the Decades: Are They Improving? *Journal of Cost Analysis and Parametrics* 11(1): 87-99.

[11] Kunc W, Barth S, Hunt C and Titrault (2018) *Joint Agency Cost Estimating Relationship (CER) Development Handbook*, Washington DC, USA.

[12] LaMorte W (2017) Nonparametric tests. Wilcoxon Signed Rank Test. Available at: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/BS704_Nonparametric6.html.

[13] Large J, Hoffmayer K and Kontrovich F (1974) *Production Rate and Production Costs*. RAND report R-1609-PAE, Santa Monica, USA.

[14] Mislick G and Nussbaum D (2015) *Cost Estimation: Methods and Tools*. New Jersey: John Wiley & Sons.
[15] Moore S, Ritschel J, White E, Koschnick C, Lucas B and Valentine S (2022). Step-Down Functions in Airframe Learning Curves: Do They Exist? *Journal of Cost Analysis and Parametrics*, 10(3): 32-43.
[16] Moses O (1990) *Extensions to the Learning Curve: An Analysis of Factors Influence Unit Cost of Weapon Systems*. Naval Post-Graduate School report NPS-54-90-016, Monterey, USA.
[17] Office of the Secretary of Defense, Cost Assessment and Program Evaluation [OSD-CAPE] (2021) *Inflation and Escalation Best Practices for Cost Analysis*. Washington DC, USA.
[18] OSD-CAPE (2022) *DOD Cost Estimating Guide*, Washington DC, USA.
[19] Smith L (1976) *An Investigation of Change in Direct Labor Requirements Resulting From Changes in Airframe Production Rates*. PhD Thesis, Oregon University, USA.
[20] Womer N (1979) Learning Curves, Production Rate, and Program Costs. *Management Science* 25(4): 312-319.
[21] Wright T (1936) Factors Affecting the Cost of Airplanes. *Journal of the Aeronautical Sciences* 3(4): 122-128.

Disclaimer: *The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.*