The Changing Cost of Defence Systems

A Thesis Submitted in Fulfilment of the Requirements for
the Degree of *Doctor of Philosophy, 2024.*

Timothy John Jefferis
Department of Management Science
Strathclyde Business School
University of Strathclyde, Glasgow

Limited amounts of new data were created during this study. All new data, together with pre-existing data that appear not to be available electronically are either reproduced in this report or are openly available from the University of Strathclyde KnowledgeBase at
https://doi.org/10.15129/6f1f8d5e-ec5a-4dd1-8e4a-edc06be39a83 (School Fees)
https://doi.org/10.15129/3fd130a0-be32-4509-98de-af5259f34b72 (Aircraft Data)

For which if you, intending to build a tower, sitteth not down first, and counteth the cost, whether he have sufficient to finish it? Lest haply, after he hath laid the foundation, and is not able to finish it, all that behold it begin to mock him, Saying, This man began to build, and was not able to finish.

Luke 14:28 – 30 (King James Bible).

Abstract

Ever since humankind began to undertake building projects and other complex activities, there has been a need to forecast the resources required, to ensure that they will all be available. Advances in the scope and complexity of projects make this task significantly more challenging, especially in areas such as Defence where the customer requirements drive the scope and complexity of projects to the maximum level that can (probably) be delivered.

This study explores the existing literature on changes in the cost of Defence Systems over time and notes that, in addition to a pervasive lack of accessible data, many previous authors have tended to rely on variation in the parameters of the products (such as in-service date, speed and mass) to explain changes in their costs. In some cases this appears to be justifiable where, for example, all of the Systems being compared contain similar technologies and were designed and manufactured in a similar context. However, other cases, such as when it is asserted that all tactical aircraft ever produced can be compared without considering the context in which they were produced, seem hard to defend.

This literature review generates three key questions that the rest of this research sets out to answer. These are:

1. What did Defence Systems cost in the past?
2. How should we compare Defence Systems over time?
3. How should we expect the prices of Defence Systems to change over time?

To shed light on these questions three case studies, down selected from a longer list of potential options, are examined in detail. These are:

- UK Independent School fees
- UK based Formula 1 team running costs
- UK Defence System costs

The two non-Defence studies shed light on changes in the processes delivering goods and services in manpower intensive and technology intensive enterprises, which are then compared with the results from the Defence case study. In each case consistent results are found which suggest that common contextual factors are at work that drive the cost changes in the processes underlying all three domains. This analysis also includes a careful examination of how costs should best be compared across time.

Considering the processes used to produce products and their wider context, rather than the products themselves, a novel approach to forecasting the cost of future Defence Systems is outlined. This examines the likely duration of a development project, depending on high-level measures of project challenge and also examines the size of the team likely to be required. Data on the project Challenge Ratings of post war UK and US military and civil aircraft development projects are used to predict project duration and a good fit is found. Insufficient data is currently available to understand the drivers of increases in required team sizes and so the future work necessary to examine this dimension of the new approach is outlined.

Contents

List of Figures

List of Tables

Abbreviations

AWS	Amazon Web Services
CAP	Common Agricultural Policy
CPI	Consumer Price Index
EU	European Union
FIA	Federation Internationale de l'Automobile
FY	Financial Year
GDP	Gross Domestic Product
GNP	Gross National Product
GPA	Grade Point Average
JSF	Joint Strike Fighter
ONS	Office for National Statistics
PC	Personal Computer
PFI	Private Finance Initiative
PPP	Public Private Partnership
RPI	Retail Price Index
TRL	Technology Readiness Level
US / USA	United States of America

Preface and Acknowledgements

This thesis represents the culmination of a decade of part time research, analysis and writing. Although the impact of the Covid-19 pandemic on working practices had some influence on the duration of this activity, the main challenge was extracting a coherent narrative from the huge volume of interesting and potentially relevant material that is available in archives and elsewhere. This process means that examination of topics such as the change in remuneration of Church of England bishops and of English Premier League footballers over time, the costs of 18th and 19th Century civil engineering projects and detailed historical costs of many different Defence Systems have had to be postponed to allow this thesis to actually be completed.

I have been fortunate to benefit from the courses that the University of Strathclyde's Business School makes available to Post Graduate Researchers which are, according to colleagues, more comprehensive and useful than those offered by a number of other institutions. Despite the best efforts of those teaching the Research Philosophy course I still struggle with the idea of a constructivist world view, so this work takes a firmly realist viewpoint, assuming that there is a real world that can be (at least partly) observed and whose features can be agreed upon.

As my background in this area is in practice rather than pure research I am aware of some of the practical difficulties in justifying the approach that should be taken in certain areas of this type of analysis, such as the adjustment of historic costs to a common economic baseline. In Chapter 6 I have therefore written on this topic at greater length than might strictly be necessary, in the expectation that the extra content will be useful to practitioners.

The completion of this work and the emergence of a coherent narrative is due, in no small part, to the efforts of my main supervisor Professor Tim Bedford and also to my other supervisors Professor Graeme Acherson and Dr Andrew Perchard. I am extremely grateful to them for their encouragement, feedback and insights throughout this process.

Lastly I want to also express my thanks to my family and friends for their continued understanding and support during this somewhat protracted enterprise.

Chapter 1 - Foreword

Warfare seems to have been present in human culture as far back as the archaeological record can take us. By 8,000 BC 'Jehrico's inhabitants had surrounded themselves with an extensive elliptical wall, nearly four meters tall, two meters thick at the base, and incorporating a massive circular tower (Lee (2016, p.15)). If this evidence is rejected as some assert (p.16) that the 'wall was intended for flood control', there is strong evidence that Hamoukar (which sits on a tributary of the Euphrates) was destroyed in warfare around 3,500 BC. It 'was encircled by a thick mud-brick wall over three meters high' and the city was 'substantially if not completely destroyed by fire' and 'several thousand sling bullets lie among the wreckage' (p.17).

Since this time, determining 'adequate levels of military spending and sustaining the burden of conflicts have been among key fiscal problems' Eloranta (2023). There is a considerable literature examining spend on military activity and as Eloranta's survey shows, it is possible to make quantitative assessments of these values in absolute terms and also as a proportion of either Gross Domestic Product (GDP) or of state spending from the Roman period onward. Tracking changes in these values over time, in the light of changes in technology, geopolitics, the scale of warfare and national and global economies offers the opportunity for fascinating examination of problems such as 'How much should be spent on Defence?'. However, in this work, we focus on a small part of this broader question, specifically how and why the cost of major Defence Systems (such as aircraft and ships) have changed over time and how we might expect these effects to behave in the future[1].

As we shall see in Chapter 2 two of the classic publications on this topic (Augustine (1997) and Kirkpatrick and Pugh (1983)), amongst others, agree that that the prices paid for military aircraft have shown significant growth in recent decades. Having defined some relevant terminology and clarified that we are interested in the overall project cost, rather than predicting variations from the estimated budget, we then consider the Defence specific literature in more detail in Chapter 3.

That section of this work leads us to observe that a number of specific issues arise from the literature about the most appropriate approaches to comparing the cost (or price) and quality of different systems, which are more fully considered in Chapters 4, 5 and 6. Additionally, we observe that these studies all subscribe to a realist ontology, meaning that there is a single reality that can be agreed upon by different viewers. Hence, for example, they hold that it is valid to suggest that there can be a single value for the costs incurred in producing a given Defence System (say HMS Queen Elizabeth), for a defined scope of the System, and that this can be meaningful to different observers.

This work will also adopt this world view, because adopting a constructionist ontology that would deny that there could be agreement on a single view of the cost of a warship, would

[1] This focus also means that this work does not attempt to address the knotty problem of measuring (or otherwise valuing) Defence Outputs.

make quantitative analysis meaningless.

Although never explicitly discussed, it appears that a number of authors have adopted a positivist epistemology, which assumes that the construction of models based upon measurable product data (usually parameters such as in-service date, speed and mass) will allow inductive reasoning to generate conclusions about the causes of the growth of the growth in the cost of relevant Defence Systems. For example Augustine (1997) tacitly assumes that date of an aircraft is sufficient data to determine its cost. The design of the study undertaken by Large et al (1976) shows similar thinking as (page v) previous work had found that 'the characteristics that best explain variations in cost among airframes are airframe unit weight and maximum speed' and sought to find 'additional characteristics that would make an estimating model more flexible and hence better able to deal with characteristics peculiar to an individual aircraft'. As it turned out 'that effort was not productive' as the 'variations in cost that are not explained by weight and speed are not explained by any other objective parameters tested'. More recently Johnstone (2020) revisited Augustine's work with a new data set that normalised for the directly measurable factors of production numbers, system weight and also capability, as measured by the composite Aircraft System Performance metric. A correlation between the performance and the cost of aircraft was observed (as one might expect) and although causation is not specifically claimed, it is still implied as the analytical focus is still on the measurable parameters. If one accepts that the assumption that System performance (or other approaches) can provide a reliable measure of System utility, then this implies the possibility of there being a correct[2] choice between procurement options, which can be evaluated analytically (see Markowski et al (2023) for example).

Whilst the positivist focus on measurable product parameters is likely to be an excellent approach in circumstances where the other factors are (essentially) identical, (e.g. comparing two technologically and temporally similar projects), the less similar the projects and hence the less similar the other factors, the less satisfactory the strictly positivist stance will be. In response to such situations (unlike Large et al) some authors infer additional parameters to capture the impact of aspects not directly reflected by measurable performance or programme parameters, but which they require to be present to explain the observed changes in cost. Webb (1990) contains a particularly clear example of this. Also the additional parameter ('complexity') is vaguely defined and hence its value for a historic project must be determined by back calculating what value must have applied for the measurable parameters to have yielded the recorded project costs.

In a similar vein, Hess and Romanoff (1987) recognise that the technological challenge of a project influences the likely cost and produce a Technology Index (p.3) which relates 'the time of appearance of an aircraft design to its level of performance, which is interpreted as a measure of its level of technological sophistication'. To forecast the likely cost of a future project, the expected system performance must be calculated and also assumptions have to be

[2] i.e. That the procurer can weigh the performance and cost of candidate options and select the best choice with no concern that its eventual utility may depend on the roles it may or may not be called upon to perform in an uncertain future.

made about the likely general rate of technological advance, to determine how challenging the system performance will be compared with the general state of the art.

Needless to say, using either approach to forecast the likely costs of a future project that is not technologically and temporally similar to an existing data point contains significant challenges, as these approaches will generally require extrapolation from known data, in order to predict what might happen in the future. Later in this work we invent the concept of project Challenge Ratings, where the difficulty of technology and design aspects of a project is assessed against a fixed 5 point scale. Whilst this approach does not provide the same level of granularity as, say, measuring an achieved speed, provided the scale categories are clearly described they should be just as objective and observable as performance parameters.

For this research I shall be adopting elements of a Critical Realist approach, which was popularised by Roy Bhaskar in the 1970s (see Bhaskar (1978 (for example)). A very convenient summary of the main points of this approach was produced by Bygstad and Munkvold (2011) who were exploring its application to Information Systems. The initial premise is (p.2) that reality is stratified into ‘three domains; the real, the actual and the empirical. The real domain consists of structures of objects, both physical and social, with capacities for behaviour called mechanisms. These mechanisms may (or may not) trigger events in the domain of the actual. In the third layer these events may (or may not) be observed, in the empirical domain’.

They further report (p.2 – p.3) that critical realism ‘combines a realist ontology with an interpretive epistemology’ however this ‘does not imply a judgemental relativism; since a real world does exist critical realism holds that some theories approximate reality better than others, and that there are rational ways to assess knowledge claims’. It follows ‘that critical realism does not aim to uncover general laws, but to understand and explain the underlying mechanisms’. ‘The basic objects and mechanisms are usually not observable’, for example, ‘while we may observe buyers and sellers agree on prices and volumes, the underlying market mechanism is unobservable. Thus contrary ‘to positivist research, the aim of critical realism in not to investigate regularities at the level of events, but rather to uncover and describe the mechanisms that produced these events… thus, instead of aiming to generalize at the level of events, critical realism methodology rests on abstract research, which aims at a theoretical description of mechanisms and structures, in order to hypothesize how the observed events can be explained’.

In this context, the ‘methodological question is; how do we identify mechanisms, since they are not observable? As Bhaskar puts it, “theoretical explanation proceeds by description of significant features, retroduction to Research Methods and Philosophy possible causes, elimination of alternatives and identification of the generative mechanism or causal structure at work”’. The technique of retroduction is explained in the following terms, ‘we take an empirical observation and hypothesize a mechanism that might explain that particular outcome’.

For this study, two main insights are taken from the Critical Realist approach:

- One should accept that there may well be important aspects of the process of developing and producing Defence Systems (and other goods and services) which can be inferred to exist, but whose operation cannot be directly observed.
- One should not always expect direct proof of causation and should be satisfied with the most plausible explanation that fits the known observations.

In this thesis we consider the literature relevant to the cost of Defence Systems and also that relating to the key enabling techniques of comparing the cost of different Defence Systems across time. We consider a range of potential case studies and then select for detailed development UK Independent School fees and the cost of running a UK based Formula 1 motor racing team, as an initial qualitative assessment suggests that the processes determining their costs are likely to be influenced by similar factors to those affecting Defence Systems. Further investigation suggests that the mechanisms and structures generating the observable effects in these cases do indeed have relevant similarities to those influencing the design and development of Defence Systems.

The likely structures, mechanisms and effects identified are finally considered in terms of their ability to provide a plausible explanation of the observed outcomes of the System development projects and hence to provide an alternative Cost Forecasting approach to the unsupported extrapolation involved in forecasting future project costs using current approaches.

In reviewing the current literature three main research questions are identified, each of which has a number of subtopics. These are laid out in Table 1.1.

Question Number	Subtopic Number	Description
1		**What did Defence Systems cost in the past?**
	1.1	What was the scope and amount of these costs?
	1.2	In what context were they incurred?
	1.3	What was delivered for these costs?
2		**How should we compare Defence Systems over time?**
	2.1	How can we compare their cost?
	2.2	How can we compare their quality?
	2.3	How have contextual changes affected outcomes?
3		**How should we expect the prices to change over time?**
	3.1	How will cost of systems change over time?
	3.2	How will customer willingness to pay change?

Table 1.1 - Cost of Defence Systems - Research Questions and Subtopics

Overall this work makes four original contributions to knowledge.

1. By focussing on the processes that produce goods and services and the context in which they operate, rather than the details of the products themselves, two matrices are created that identify similar goods and services according to
    a. The labour intensiveness of the processes and the scarcity of the outputs.

   b. Whether stakeholders can directly measure the impact of changes in the process's input and also its output quality.

2. Data sets relating to a broad range of goods and services are considered and the matrices are used to identify those which are most similar to Defence Systems. It is not intuitively obvious which goods and services should be most similar and this approach provides a well-defined process for determining similarity.

3. The data sets relating to UK Independent School fees, the cost of running a UK based Formula 1 team and the cost of developing and producing UK Defence Systems are developed into full case studies and relevant similarities are identified.

4. A novel approach to forecasting the cost of Civil and Military Aircraft development projects is proposed and it is shown to successfully forecast the duration of such projects.

Chapter 2 - Introduction - Augustine and the Starship Enterprise

This Chapter introduces the background to research about the cost of Defence Systems and starts by considering two pieces of literature that were published in the early 1980s and which are probably the best known work on the cost of Defence Systems. We then discuss the vocabulary used by different authors and make choices about which to employ in this research. The Chapter concludes with a discussion of the different types of management decision that can be informed by Cost Analysis and also by providing a reminder that Defence Systems represent only one of the components of Fighting Power that are necessary to produce military effect.

2.1 - Introducing Augustine's Laws

In 1983 Norman R Augustine, a leading figure in the US aerospace industry, who was later to become Chairman and CEO of Lockheed Martin, produced the first edition of *Augustine's Laws*. According to the Library of Congress catalogue data this is a collection of anecdotes about management in general and industrial project management in particular and it is organised into a series of 52 essays. For the purposes of this study the latest edition (the sixth) (Augustine (1997)) was consulted, but the main thrust of the arguments remains unaltered between the editions. About one-third of the way through this volume the essay entitled *The High Cost of Buying* is reached, which opens with the assertion that 'It can be shown that the unit cost of certain high-technology hardware is increasing at an exponential rate with time[3]' (Augustine 1997, p.104). Graphs are provided that illustrate this point for commercial airliners, bomber aircraft and tactical aircraft, the last of which is redrawn below as Figure 1.

Augustine's legend to this chart (Augustine 1997, p.105) asserts that the 'unit cost of tactical aircraft has increased in a very consistent manner ever since the beginning of the aviation age. The rate of climb is a factor of four every ten years, There is no ceiling in sight'. He also asserts that the same inexorable trend can be shown to apply to helicopters and even ships and tanks, although the last two are reported to only increase by a factor of two every ten years. Whilst this is an interesting observation, the real impact is when Augustine compares this growth rate with the rate of increase in the US Defence Budget and US Gross National Product (GNP). His graph (Augustine 1997, p.106), redrawn in simplified form below as Figure 2 shows that in 2054 the whole budget will only suffice to purchase one aircraft.

This leads him to assert (Augustine 1997, p.107) that 'In the year 2054, the entire defense budget will purchase just one aircraft. This aircraft will have to be shared by the Air Force and Navy 3½ days each per week except for the leap year, when it will be made available to the Marines for the extra day'. He labels this as Law Number XVI. This is a stark image and the corollary is even starker, once this state is reached it will, presumably, be impossible to sustain an air force, as no new planes will be available at a price that any nation can afford. This situation would presumably also result in the extinction of the military aerospace industry.

[3] This creation of a general law, based upon the observed data is a positivist position.

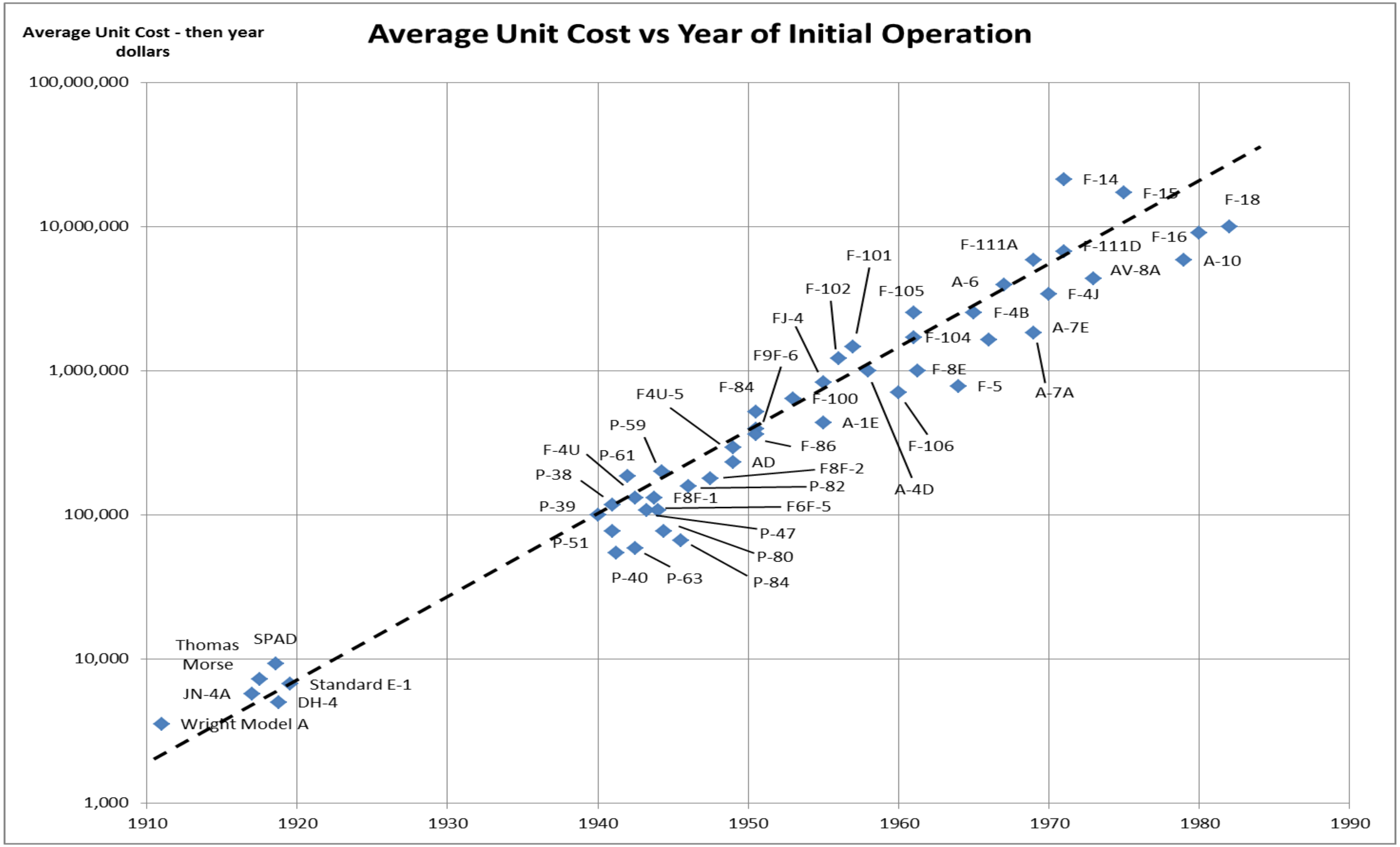

Figure 1 - Augustine's Law XVI - Cost of Tactical Aircraft

(See Table 15 for source data).

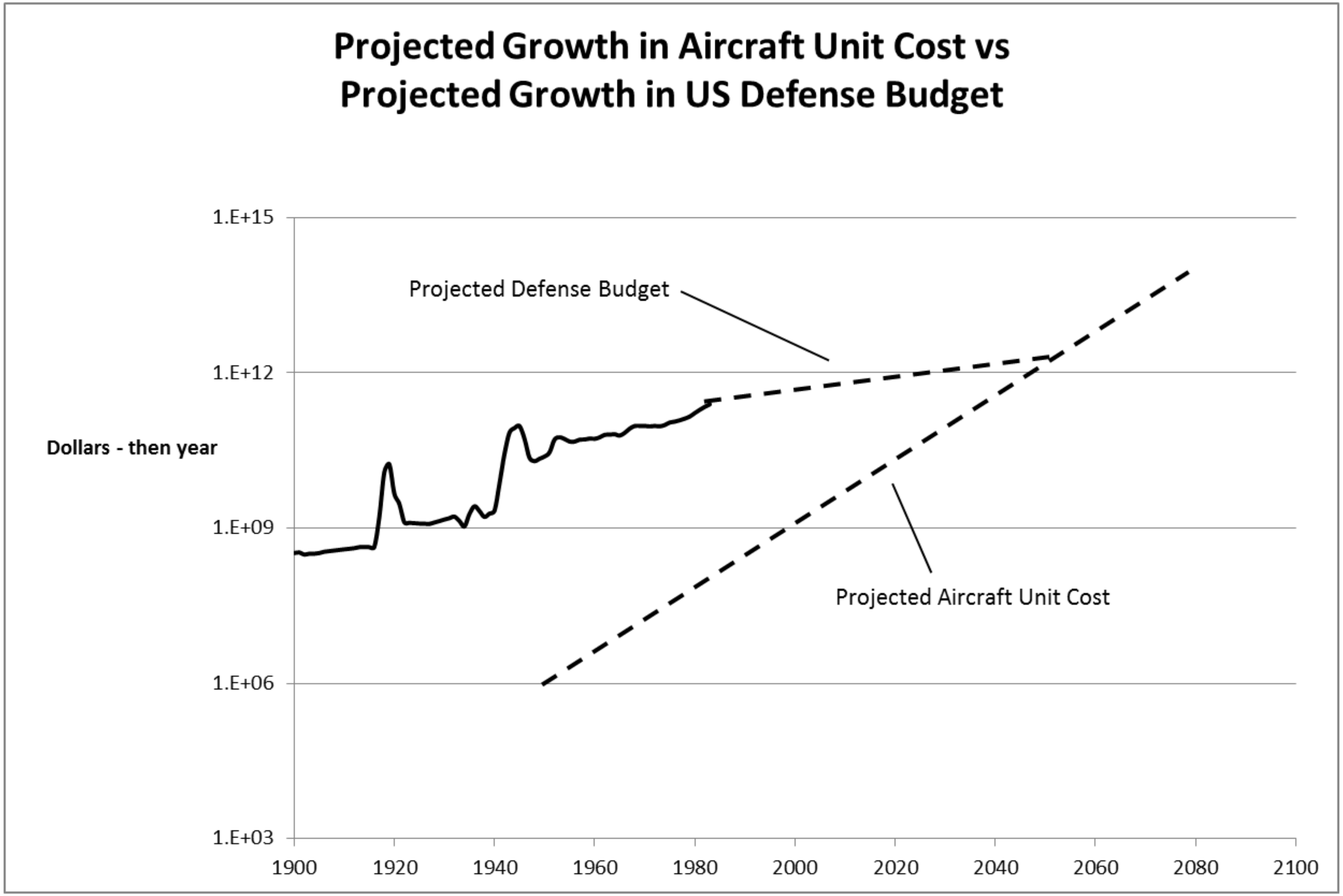

Figure 2 - Projected Growth in Aircraft Unit Costs vs Projected US Defense Budget

Such prophecies of doom are always well received, as John Stuart Mill (1828) observed 'not the man who hopes when others despair, but the man who despairs when others hope, is admired by a large class of persons as a sage'. In his 2011 book *The Rational Optimist* Matt Ridley (Ridley, 2010) devotes a whole chapter (p.279 – 311) to a history of pessimism and attempts to show that man's perennial drive to look on the down side has, so far, been misguided. It is not clear how seriously Augustine expects us to take his conclusions as his final thought is to cite the maxim that he attributes to Lenin that 'quantity has a quality all its own'. This suggests that aircraft unit costs cannot continue to rise until the air force consists of a single platform, which is, essentially, the same point that Stein's law[4] makes about the trend being unsustainable. Additionally, in two other places[5] in *Augustine's Laws* he illustrates the absurd conclusions that would result from assuming that current trends would continue. We may therefore suspect that whilst Augustine has produced a compelling piece of writing, originally he expected the conclusion to be taken with a pinch of salt.

2.2 - Introducing the Starship Enterprise

In their paper *Towards the Starship Enterprise – are the current trends in defence unit costs inexorable?* Kirkpatrick and Pugh (1983) come to similar conclusions to Augustine's work even though they appear to have written without sight of his publication. Their Figure 1 contains 'unit production cost'[6] and 'date of first production delivery' data for 19 post-Second World War UK military aircraft. In this case the data have been normalised to refer to the 'average cost' for producing the first hundred units and the costs are all presented at September 1980 prices, which also means that some method has been used to adjust the historic costs to a common basis.

Kirkpatrick and Pugh's paper does not attempt an Augustinian type comparison of aircraft cost with the Defence Budget, so we are simply left with their calculations that the data adjusted to September 1980 prices exhibit an annual growth of 8.3% per annum. As this equates to an increase by a factor of 2.2 per decade this is less alarming than Augustine's factor of 4 per decade, but we cannot tell how much of this caused by the exclusion of inflation and normalisation for production quantities and how much is a reflection of differences between UK and US designs and procurement processes.

Kirkpatrick and Pugh's second figure is similar in structure to Figure 1 in this work, but is wider in scope covering some 140 military aircraft from the UK and US with in service dates ranging from 1910 to 1980. Sadly these data points are not labelled and so, unless supporting information can be discovered, it is possible that, apart from a few readily identifiable outliers, the identity of these points will never be established with any certainty. Their observation on the trend that appears in this figure is 'that the upward trend in unit cost has continued since the beginning of military aviation'. In order to broaden the basis of the paper,

[4] Stein's law states 'If something cannot go on for ever, it will stop'. (Stein 1989).
[5] Law XIV shows that by 2015 100% of aircraft weight will be composed of electronics and Law LI shows that by 2076 the US government will employ more workers than there are in the US.
[6] It is unclear whether prices or costs are actually being represented. As with Augustine the term used by the authors will be followed in this text.

they also make reference to other writers who have observed the increasing unit costs of both combat aircraft and also other types of weapon system.

Having completed a review of the production costs of military aircraft the authors then, in a step away from the positivism of Augustine, discuss the factors that might be driving the perceived increases, to which we shall return in Chapter 2. The title of the paper comes from their observation about the logical conclusion of the current trend continuing unchecked, which would result in ‘the Defence of the UK being entrusted to the *Starship Enterprise* in synchronous orbit for a century or more’.

2.3 - General Study Approach

This current study aims (inter alia) to discover whether the observed inexorable rise in the cost of Defence Systems (such as aircraft) is real and if so what is driving it and what (if anything) may be expected to occur in the future to divert the trend away from universal unaffordability. The data available about Defence Systems can be somewhat limited (both in number of data points available and the clarity of what each data point represents), so, to broaden the data sets available for analysis, this work will also consider some other relevant goods and services, to examine what insights can be gained from them.

Assuming that advances in areas of theory outlined above can be realised, the maximum value will be generated if the insights can be embedded into the processes that are commonly used to estimate the likely cost of Defence Systems. This would, in turn, allow more informed decisions to be made in the management of programmes aiming to deliver Defence Systems in the UK and internationally. The rest of this work therefore aims to ensure that, as far as possible, the enablers for the advancement of theory are also considered through the lens of their application to practice.

2.4 - Discussion of Vocabulary

Before considering the existing literature on relevant topics there are three areas of scope and definition that need to be clarified. The first of these is to ensure that a clear and appropriate vocabulary exists to discuss concepts. As we shall see, an initial review of relevant literature reveals inconsistencies in the terms used by past authors to describe different features of cost changes. It will therefore be useful to consider what terms might be employed consistently through the rest of this study, without meanings applied by previous authors potentially causing confusion.

For complex public procurements, such as Defence Systems (i.e. usually bespoke or customised designs, built for a public sector customer), there are two main ways in which headline cost changes are considered. The first of these is exemplified in Pugh’s (1986) comparison of the cost of HMS Havock from 1894 (p.73) with that of a Type 21 frigate from 1974 (p.88), which is an examination of change in cost of (arguably similar) warships over the 80 years between the launch of the two vessels. The second is illustrated by the headline ‘Edinburgh’s tram system, opens - £375m over budget and three years late’ (Guardian (2014)) which refers to cost change (and schedule slip) between the budget being set and the project being completed.

In Arena et al (2008) we find clear definitions and names for these two distinct types of cost change. The authors name the first as cost escalation, which represents 'the general changes in price, typically for a similar item or quantity, between periods of time' (p.9) whilst the second they call cost growth, being 'the difference between actual and estimated costs' (p.9). Hove and Lillekvelland (2016) also follow this approach in specifically differentiating between 'cost escalation' which 'occurs between generations (intergenerational) and within generations (intragenerational)' and 'cost growth' which 'is the rise in costs from the time a project is started to the time of acquisition' (p.208).

From the title 'Intergeneration Equipment Cost Escalation' we might expect Davies at al (2011) to follow Arena et al and Hove and Lillekvelland's definitions, but closer inspection reveals that (p.8) under the heading 'Intergenerational Cost Escalation' the definition given is for 'intergenerational cost growth' which is listed as the 'change in cost between one platform of military equipment and the next generation of a similar platform of military equipment'. This is very similar to the previous definitions of cost escalation. Further examination of the paper reveals that Davies et al (2011) actually use various terms interchangeably to refer to this phenomenon.

Gansler and Lucyshyn (2013) also distinguish between 'the growth between estimate and achievement' (p.1) and the issue of 'the unit cost of DoD weapons systems, which has also increased significantly over time'. However their observation on this second issue that 'for example, the unit cost of high-performance aircraft programmes, which has grown at an exponential rate' (p.1) illustrates that they use the term growth to refer to both effects.

Authors who are only interested in one of these effects have more choice in selecting which vocabulary to use. Kirkpatrick and Pugh (1983) and Augustine (1997) are only interested in the change in the cost of military equipment over time and so are able to use a number of terms interchangeably. Kirkpatrick and Pugh do use the term escalation, but only to specifically refer to the process by which the continuing attempts of two rival nations to produce a technologically superior aircraft drive up the unit costs for each nation. Lehtonen and Anteroinen (2016) have similar interests and favour the term growth when they examine (p.280) 'The unit cost growth of military equipment' which 'denotes the per piece cost of military equipment, such as fighter aircraft, which is intended for the same purposes, over time and even over generations'. They only use the term escalation where quoting or paraphrasing other authors who use it.

Looking back to earlier publications we find other terms also in use. For example, in illustrating intergenerational cost differences the UK Ministry of Defence in the *Statement on the Defence Estimates* (UKMOD (1981, p.45)) caption their figure 'change in cost in real terms' and in examining the increase between estimate and project completion Novick (1970) favours the term 'cost overrun'.

Although one could continue collecting different vocabulary until all the potentially relevant publications had been exhausted, the conclusion that different authors use different terms to describe the same concepts is already clear. It will therefore be necessary to make reasoned choices about the vocabulary to be used to discuss this research, rather than relying on the choices of previous authors. To describe the differences between the initial budget and schedule and what is actually achieved we shall use Novick's terminology and adopt the terms 'cost overrun' and 'schedule overrun' as these carry the clear implication that the expected limits have exceeded. In considering the difference in cost over time between different systems fulfilling similar roles we shall generally use the terms 'cost growth' and 'schedule growth', except in specific cases where some driving factor(s) have been identified that are causing costs and or schedules durations to escalate, when the term 'escalation' will be used.

Devising a scheme of terminology that would satisfy and be adopted by all authors in this field would be desirable and could be addressed by a number of different approaches. For example, from a theory approach, academic research might unearth compelling arguments to support one particular set of terminology choices, similarly, from a practice approach, an authoritative standardisation body might impose a solution. It is almost certain that any such approach would require consideration of a range of factors that are outside the scope of this current research. Therefore, having made choices for this work, development of this topic must be left as an opportunity to be addressed in future work, by an appropriately positioned author.

At this stage it is also convenient to review the usage of the terms 'cost' and 'price' (all the following definitions are from Bannock et al (2003)). Cost or opportunity cost is the value 'that must be given up to acquire or achieve something' typically we shall encounter it as meaning the expense incurred for creating a product or service a company sells. Price, on the other hand is 'What must be given in exchange for something. Prices are expressed usually in terms of a quantity of money per unit of … a good or service'. We shall usually encounter it as the amount a customer must pay for a product or service, with the difference between the price paid and costs incurred representing the contractor's profit'.

Authors tend to interchange cost and price and this work will aim to be specific about what is included in each articulation of cost and price, where this is known. Another feature of Defence Systems, is that on occasions the government will purchase some subsystems directly and then supply them to the prime contractor for integration into the overall system as Government Furnished Equipment (GFE). In such cases the cost to government will be the total for all the inputs, not just those incurred by the prime contractor.

2.5 - Cost Analysis Aims

The second topic that requires clarification is exactly what class of insights we wish Cost Analysis to provide to relevant decision makers. The traditional goal is to directly forecast the cost and duration of a whole project, or a certain project phase, so that appropriate budgets can be set and progress can be monitored. The usual approaches to generating such outcomes include (Office of the Secretary of Defense (2020, p.59)) Analogy, Build Up, Extrapolation of Actuals and Parametric. These techniques revolve around using relevant outturn[7] data from

[7] i.e. What actually occurred.

previous similar projects to populate one or more models that will provide a forecast of the eventual cost and schedule. This class of analysis does not consider the accuracy of the initial cost and schedule estimates for the previous projects being considered, simply what their eventual outcome was.

There is, however, an alternative class of analysis, which aims to assist management by forecasting the likely cost and schedule overruns on a project, by examining the divergence of previous similar projects from their initial forecasts. In 2002 Mott MacDonald[8] produced a report for the UK government entitled ‘Review of Large Public Procurement in the UK’, which examined the management factors that had been identified as materially contributing to cost and schedule overruns. The projects surveyed were divided up into six distinct classes of product where member projects would be expected to behave similarly.

These were:

Standard Buildings

Non-Standard Buildings

Standard Civil Engineering

Non-Standard Civil Engineering

Equipment / Development[9]

Outsourcing

Within each class a subdivision was made depending on the procurement route[10] as it was perceived that this could affect project behaviour. Whilst some of these classes, such as Standard Buildings (office buildings, hospitals and prisons) and Standard Civil Engineering (mainly new road building) have immediate appeal as homogeneous groups, others appear more challenging. For example Equipment / Development contains only two members, the Faslane Ship Lift and the Defence Fixed Telecommunications System, whose main connection appears to be that they do not fit into any other class. Similarly the diversity of projects in Non-Standard Civil Engineering class, which includes surface and underground rail lines, nuclear power stations and the Thames Barrage, do not have obvious elements in common. The challenge of identifying which past projects are similar to a current project will be a recurring theme in this research.

Despite the challenges with some of the classes, this report resulted in guidance being produced by HM Treasury (2013) laying down a method to address the problems of ‘optimism bias’, by providing a structured approach to assess the likely impact of systematic management failings in identifying and valuing the likelihood and impact of project risks.

[8] A consulting firm.
[9] It is worth noting that the Mott MacDonald data for Equipment / Development was rather sparse as it covered one traditional and one Private Finance Initiative (PFI) / Public Private Partnership (PPP) project.
[10] Whether a traditional approach was being followed or a PFI / PPP used.

This work by Mott Macdonald and the guidance that flowed from it was not the first identification of 'Optimism Bias.' In the project management classic 'The Mythical Man Month', Brooks (1975), discusses a number of reasons why software projects always tended to be over budget and behind schedule. Early in the paper he cites 'the first false assumption that underlies the scheduling of systems programming' which is that 'all will go well, i.e. that each task will only take as long as it 'ought' to take' (p.14), which is the sort of fallacious reasoning that underlies Optimism Bias. The second fallacy that he identifies (p.16) is the implication of the selection of the man-month as a unit of effort, which is that 'the men and months are interchangeable'. He continues to observe that the men and months 'are interchangeable commodities only when a task can be partitioned among many workers with no communication among them. This is true of reaping wheat or picking cotton; it is not even approximately true of systems programming.' The opposite extreme is a task that 'cannot be partitioned because of sequential constraints' (p.17) where the 'application of more effort has no effect on the schedule', a clear example of this is the bearing of a child, which 'takes nine months, no matter how many women are assigned'.

Most tasks are susceptible to some partitioning, but fall short of the extreme of complete man and month interchangeability. Brooks identifies these as (p.17) 'tasks that can be portioned but which require communication among the subtasks' and rightly notes that 'the effort of communication must be added to the amount of work to be done'. If each part of the task has to communicate with every other part, 'then the effort increases as n (n-1) / 2' (p.18), thus dividing a project into five tasks will require ten times as much communication as a project divided into two tasks. Brooks also reminds us that 'each worker must be trained in the technology, the goals of the effort, the overall strategy and the plan of work', this effort cannot be partitioned and so this additional effort which 'varies linearly with the number of workers' must also be considered. Whilst Brooks was considering software development within a single company, these two issues that he raises will be of even greater significance when applied to multi-national projects, particularly if some participants need to undertake significant learning before they can contribute effectively. The difficulties in co-ordinating the necessary communication across multiple sites in multiple nations is noted by other authors such as Hersleb and Grinter (1999) who offer insights into how such difficulties may be mitigated.

In more recent work, rather than attempting to identify specific factors that might impact the accuracy of cost and schedule forecasts, Flyvberg attempts to take a broader approach. In his paper 'From Nobel Prize to Project Management: Getting Risks Right' Flyvberg (2006) he lays out a history of investigations of why project estimates tend to be inaccurate. He observes (p.6) that various commonly suggested technical explanations for the observed effects (e.g. 'unreliable or outdated data' and 'the use of inappropriate forecasting models') appear unlikely as these should produce effects where one would expect 'the distribution of inaccuracies to be normal or near-normal with an average near zero', as one might expect for random measurement errors or other forms of noise. His analysis finds that actual 'distributions of inaccuracies are consistently and significantly non-normal with averages that are significantly different from zero', which leads to the conclusion that 'the problem is bias and not inaccuracy as such' (p.6). He also optimistically argues that 'if imperfect data and models were main explanations of inaccuracies, one would expect an improvement in accuracy over time, because in a professional setting errors and their sources would be recognized and addressed, for instance, through referee processes with scholarly journals and similar expert critical reviews' (p.6). Interestingly he does not consider whether censoring of the data might impact the observed results. For example, if a contractor is delivering a fixed price contract, and happens to complete the work under budget, the whole of the contract value will still be paid, thus resulting in zero chance of the work being completed under budget, but a finite probability of it being over budget.

Flyvberg's (2006) solution to this problem is Reference Class Forecasting which 'does not try to forecast the specific uncertain events that will affect the particular project, but instead places the project in a statistical distribution of outcomes from the class of references projects' (p.8) which he explains (p.8) 'consists of regressing forecasters' best guess towards the average of the reference class and expanding their estimate of the credible interval toward the corresponding interval for the class'. The author does acknowledge (p.9) that 'choosing the right reference class of comparative past projects becomes more difficult when managers are forecasting initiatives for which precedents are not easily found, for instance the introduction of new and unfamiliar technologies'. However he offers the opinion that (p.9) 'most projects are both non-routine locally and use well-known technologies. Such projects are, therefore, particularly likely to benefit from the outside view and reference class forecasting'. Whilst many Defence projects are 'non-routine locally', the more demanding of them tend not to use 'well-known technologies' and hence it is unclear how good a fit they are to Reference Class forecasting.

As this work does not seek to consider cost and schedule overruns we shall only be considering Reference Class Forecasting and other approaches of that ilk in so much as they contribute to the debate of how similar a current project needs to be to past projects, for their results to provide useful insights.

2.6 - Components of Fighting Power

Finally we need to consider the actual purpose of Defence Systems. Joint Doctrine Publication 0-01, titled 'UK Defence Doctrine' (Ministry of Defence (2022)), articulates the three components of fighting power, which are Moral, Conceptual and Physical (p.24). Figure 3, which is reproduced from that document lays out all the relevant sub-elements.

From this we observe that whilst the equipment that Defence Systems represent is undoubtedly an important element of the Fighting Power that delivers Defence Outcomes, it is a single element of the Physical component. Without the other Physical elements and also without the Conceptual and Moral components they are simply rather expensive ornaments.

Having explored the background to this topic and defined the scope that we aim to explore further, we may now move on and investigate the relevant literature in more depth.

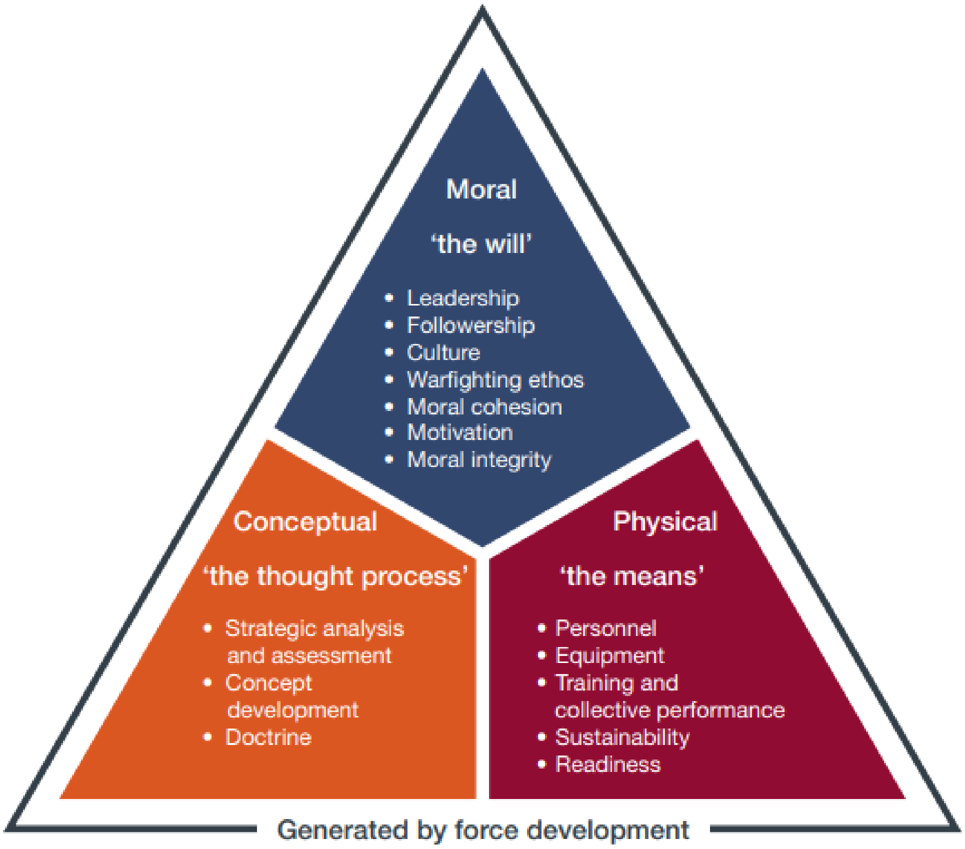

Figure 3 - Components of Fighting Power

Chapter 3 - Introduction - Augustine and the Starship Enterprise

3.1 - A Critique of Augustine's Law XVI

As noted in the Introduction there is a suggestion that when originally writing Augustine may not have expected his 2054 deadline to be taken seriously, despite this, the idea has been widely cited (e.g. Economist (2010), Sound and Vibration Magazine (Smallwood (2012)), Aerospace Today…and Tomorrow an Executive Symposium (AIAA (2015)), Augustine (2015)). Some more recent citations have tended to focus on the articulated cost growth either as a benchmark against which to test other data series (Johnstone (2020) and Hartley (2020)) or as a useful context against which to market the benefits of new engineering approaches (Siemens (2023)). Recently Hartley and various co-authors (Brauer et al (2021), Hartley (2022), Markowski et al (2023)) have explored the concept of Augustine Investments, which are a 'distinct category of technology-intensive military investments' (Markowski et al (2023), p.293). Starting from Augustine's initial concept this recent developments explores these weapon systems as a new class of investment system and investigates how well their economic behaviour is explained by different theoretical models.

With this continuing level of interest in Augustine's work it is therefore important to understand the strengths and weaknesses of its construction. The first observation is that there is a lack of data transparency because Augustine does not cite any sources for his data or tabulate the data used to generate any of his figures. Thus, to create Figure 1 it was necessary read the relevant values from the published graph. This means that the captured values are somewhat approximate. We shall see that a lack of data transparency is a common issue with published work in this area.

There are also two data consistency issues that make Augustine's arguments less easy to interpret (see Eskew (2000), Jefferis (2014) and Johnstone (2020) for example). Firstly, the data plotted is the *average* unit cost, which means that aircraft with longer production runs and consistent production rates from year to year will appear cheaper, as the learning effects of the extra numbers produced and efficiencies of consistent numbers will lower the overall unit cost (see Wright (1936) for a discussion of aerospace labour learning and Henderson (2015) for a more general discussion of the effects of experience). Thus, for example, the SPAD[11] and P-51 Mustang with total production numbers of 8,472 (Smithsonian (2015)) and 15,586 (mustangsmustangs.com (2015)) respectively, benefit in comparison with the F-18 with a total run of about 1,480 aircraft for the classic version[12] (McDonnell Douglas F/A-18 Hornet (2015)). Secondly, consideration of Figure 1 shows that there are no data points for aircraft entering service between 1920 and 1940. It is possible that Augustine happened not to be able to locate any data for any relevant aircraft, such as the Boeing Peashooter which entered service 1932, but, given the completeness of data elsewhere the omission is puzzling.

In addition to these issues with data transparency and consistency, there are two fundamental challenges relating to the interpretation of the results. The first is that all of the aircraft unit

[11] A French aircraft manufacturer. Name in full is 'Société Pour L'Aviation et ses Dérivés'

[12] That is F-18 A, B, C and D variants.

costs[13] are expressed in 'then year dollars', this being the cost or price actually paid when the aircraft were purchased. If one were to construct a Wright Model A[14] today, inflation in the costs of the necessary materials and labour means that the total outlay would be significantly higher than the few thousand dollars indicated in Figure 1. Clearly it would be very desirable to have a consistent, defendable approach that allowed the cost or price of Defence Systems to be compared at a common baseline. We shall return to this topic in Chapter 6.

The second fundamental challenge relates to the selection of the aircraft included in the dataset (Jefferis (2014)) and whether it is really credible to include the Wright Model A and the F-18 in the same data set. Both are capable of sustaining powered flight, but from engineering and organisational viewpoints the other similarities in production of these craft are minimal, therefore, it is unclear whether they should legitimately be regarded as comparable. Augustine's clear expectation that a general law can be derived that predicts the cost of all tactical aircraft, simply by observing when it appeared, seems to be a positivist view. If we were to consider the development of military aircraft as an evolutionary process, then Augustine's view would be firmly in the 'phyletic gradualism' camp, which hold that 'new species arise from the slow and steady transformation of entire populations' (Eldredge and Gould (1972, p.84)), which in this case would translate as 'new types of aircraft arise from slow and steady developments across the whole population of [US] military aircraft'. Thus Augustine would (presumably) argue that all these aircraft can be included in the same Reference Class (to borrow Flyvberg's term), as they are part of the same continuous development process.

The alternative view (Eldredge and Gould (1972, p.84)), would be in the 'punctuated equilibria' camp, where (relatively) long periods of stasis are disturbed by 'rapid and episodic events of speciation', i.e. long periods where there is little design innovation are occasionally interrupted by short periods featuring the creation of new types of aircraft design, which then become the norm for the following stasis period. As we shall later see, deciding which data points are sufficiently similar to be included in the same class of data is a fundamental challenge in this area.

From this consideration of Augustine's work, we observe that it is desirable to have:

A transparent and consistent data set

A meaningful way to account for the changing value of money over time

A defendable approach to test for similarity (e.g. membership of a Reference Class)

The challenges in each of these areas make it more difficult to ascertain from Augustine's work how much of the perceived increase is caused by inconsistencies in the data set, how

[13] It appears likely that Augustine is quoting the cost to the US Government, which is the contractor's selling price, rather than the cost to the contractor. Without further information it is not possible to determine which costs or prices are actually being quoted so Augustine's terms will be used.
[14] The Military Flyer, the US's first military aircraft, was a modified Wright Model A.

much by changes in the value of money, how much by shorter production runs reducing the learner effects and how much by increases in the underlying unit costs.

3.2 - A Review of the Starship Enterprise

In their paper *Towards the Starship Enterprise – are the current trends in defence unit costs inexorable?* Kirkpatrick and Pugh (1983) address some of the issues that were identified with Augustine's work. Their first figure contains 'unit production cost'[15] and 'date of first production delivery' data for 19 post- Second World War UK military aircraft. In common with Augustine, there are data transparency issues, with the data points simply being presented on a graph, rather than being tabulated and in this case the points on the graph are not even labelled with the relevant aircraft types[16]. Even though detective work examining the years in which the first production deliveries were made might allow guesses to be made about the identity of some points, this would introduce another potential source of uncertainty.

On a more positive note, the authors have attempted to address the consistency issue identified above. To remove the effects of learning acting over varying numbers of production aircraft, the data have been normalised to refer to the 'average cost' for producing the first hundred units. Provided that this transformation was based on access to actual production costs, rather than being inferred from higher-level data, then this feature is an excellent move towards reducing extraneous variation within the dataset.

They have also dealt with the changing value of money over time by presenting the production costs at September 1980 prices. As the original (then-year) costs are not provided, this transformation means that one cannot tell, from this paper, the original cost values. Additionally, there is no discussion in the paper of the method used for this transformation. The remaining author (D Kirkpatrick (2015) pers. comm. 8th May) believed that a GDP[17] deflator had been used, but was not certain. This thought is consistent with the opening paragraph of the paper which refers to 'the real unit production cost of UK aircraft' (Kirkpatrick and Pugh (1983), p.16). Officer and Williamson (2010) suggest that the real price measures the worth of 'a subject (a commodity) against the cost of a bundle of goods and services that in principle is fixed though in practice varies over time.' However, they define a commodity as a 'good or service, usually purchased by a consumer. In this category are such items as bread, a restaurant meal, an automobile, a tax paid, or a charitable contribution' which suggests that military aircraft are not commodities. They would, instead favour the use of historic opportunity cost, which 'measures a subject (generally a project) against a bundle of consumer goods and services (via the CPI[18] or RPI[19]) or a bundle of all goods and services (using the GDP deflator.) For a project owned by a person or household, the CPI or RPI is preferred. For investment and government projects, the GDP deflator is

[15] Again, it is unclear whether prices or costs are actually being represented. As with Augustine the term used by the authors will be followed in this text.
[16] It is possible that the authors faced issues with the security classification of the data that constrained them to this approach.
[17] Gross Domestic Product
[18] Consumer Price Index
[19] Retail Price Index

more appropriate' (Officer and Williamson (2010)).Thus, it appears probable, that Kirkpatrick and Pugh are measuring the cost of UK military aircraft relative to the bundle of all goods and services that make up the GDP over time.

There is no discussion in Kirkpatrick and Pugh's paper about how the rate of increase in the cost of military aircraft is expected to compare with that seen for the bundle of goods making up the GDP, nor do they attempt to make an Augustinian comparison with their affordability compared with the Defence Budget. As noted above, according to their calculations the data adjusted to September 1980 prices exhibit an annual growth of 8.3% per annum, which equates to an increase by a factor of 2.2 per decade and hence is less alarming than Augustine's factor of 4 per decade. However, one cannot tell how much of this caused by the exclusion of inflation and normalisation for production quantities and how much is a reflection of differences between UK and US designs and procurement processes[20].

Their observation on the trend that appears in the figure is 'that the upward trend in unit cost has continued since the beginning of military aviation'. In order to broaden the basis of the paper, they also make reference to other writers who have observed the increasing unit costs of both combat aircraft and also other types of weapon system.

Having completed a review of the production costs of military aircraft the authors then discuss the factors that might be driving the perceived increases. They suggest that there are five vicious circles that drive increases in the costs of military aircraft. The main driver is the perceived relative effectiveness of 'Red[21]' and 'Blue[22]' aircraft. As 'Blue' produces aircraft that are perceived to be of increased effectiveness, 'Red' will counter the increased threat to 'Redland' by advancing their technology and hence producing more effective aircraft. This will, in turn, produce an increase in the perceived threat to 'Blueland' and need for more advanced technology, which will, in turn, lead to more effective 'Blue' aircraft and so the cycle will repeat. The authors assume that these more effective aircraft will usually require greater development and production costs and (implicitly) that these will become less affordable as the budgets will not increase by a commensurate amount. This situation then leads to other escalatory vicious circles kicking-in to make the affordability of future aircraft progressively worse. As the development costs increase, greater time will be taken in analysing and scrutinising the options, to ensure that the budget is well spent, which will increase the duration and hence cost of the programme. The increased duration will also mean that greater steps in technology need to be taken (if the aircraft is to remain relevant) which will increase cost, duration and risk. On the production side, the increased costs will result in fewer aircraft being built, which will reduce the cost savings through learning and will also undermine the case for investment in production facilities, which will also impact on the unit cost. Here the authors are qualitatively discussing a number of effects that might be relevant to critical realist view of this process through understanding the context of the

[20] Johnstone (2020) with a different US data set to Augustine, but normalising to 2018 economic conditions and for the 100th aircraft produced found a value of 6.8% per annum, which reduced to 4.4% per annum when the effect of weight growth was removed.
[21] Generic potential adversaries are denoted as 'Red'.
[22] One's own forces and those of close allies are denoted as 'Blue'.

programme. However, their expectation that all of the influencing factors can be directly observed still leaves them with a least one foot firmly in the positivist camp[23].

The balance of their paper is concerned with identifying engineering and management reasons that could be driving up the production costs and discussing what might be done to mitigate the rate of increase. The overall theme is that the technical requirements for (say) 'a combat aircraft's ability to penetrate enemy defences at high speed and low altitude' tend to drive engineering solutions that are close to the limits of technology and that 'the approach to the limit is associated with a rapid escalation in the cost of airframes'. Similar themes are observed in the costs of engines and avionics. Although the authors do not specifically articulate the point, they are tacitly assuming the solution is to push current technologies further towards their limits, rather than adopting a new approach. The authors complete their work by recognising, as Augustine does, that a single aircraft, no matter how capable, (their Starship Enterprise) cannot be everywhere at once and that there will be some balance between aircraft effectiveness and aircraft numbers. Wisely they do not attempt to determine exactly where this will be in future.

Kirkpatrick and Pugh's viewpoint (as with Augustine) implicitly assumes that new generations of aircraft will evolve by slow and steady development of current types. If it proved possible to produce a more effective aircraft by utilising different technologies, rather than simply demanding greater performance from the current technologies, then it might, conceptually, be possible to break out from the vicious circles outlined above. To date the different technologies that have been utilised to increase aircraft effectiveness have tended to require significant bespoke development, hence increasing aircraft costs, however this could change if it was possible to utilise technologies that had already been de-risked.

3.3 - Contribution of Earlier Authors

The theme of increases in the costs of defence equipment had been discussed in various places before 1983, with the first instance being attributed by Augustine (1997, p.108) to Calvin Coolidge. Two other sources relevant to this study are cited by Kirkpatrick and Pugh. The first of these, UK Ministry of Defence's Statement on the Defence Estimates 1981 (UK MOD (1981), p.45) shows the relative costs of a number of different defence systems of different generations. The caption indicates that the changes indicated are the 'change in cost in real terms'. There is also a note that 'successive generations of weapons increase in sophistication, have greater operational effectiveness and may often be of different size. Thus these comparisons can only give broad indications' (UK MOD (1981) p.45). The data are represented as the ratio between two broadly similar weapon systems, for example, the Harrier GR1, introduced into service in about 1971 is listed as costing three times as much as a Hunter introduced in 1957, similarly Hawk from 1978 cost about 1.5 times as much as the Gnat from 1964. As the data is quoted as being in 'real terms' one must assume that some (unspecified) transformation has been applied to the raw data to (attempt to) bring them to a common baseline. Additionally, a quick review of all the cost ratios quoted indicates that

[23] Given that both authors' backgrounds tended to be focussed on the practical aspects of cost forecasting, their desire to be positivist enough to create some general laws, but critical realist enough to accept that the direct description of the end product is insufficient to explain the cost changes seems a very reasonable position..

some rounding may have been applied as the ratios are either integers or integer and a half. Whilst we may accept these calculations as being (broadly) factually correct, it is unclear what the authors of this document intended the readers to conclude from it.

The other relevant source cited by Kirkpatrick and Pugh (1983) is the work of Seymour Deitchman (1979) whose book *New Technology and Military Power: General Purpose Military Forces for the 1980s and Beyond* attempts to map the changes to (US) general purpose military forces[24] since World War 2 and to sketch out what changes may be necessary to keep them relevant into the 1980s and beyond. For each of the three services (US Air Force (p.60), US Army (p.81) and US Navy (p.106)) relevant weapon costs and the relevant budget from 1945 to 1979 are graphed in a very similar format to Figure 1 in this document.

In the case of the US Air Force the 'Combat Aircraft and Weapons Costs' for eleven aircraft and four weapons are plotted against years[25], together with the budget for relevant years. It is unclear exactly what is included in the costs quoted, but the author specifies (p.61) that the costs 'have not been corrected for inflation', so one assumes that they are then-year costs. The note attached to the actual graph (p.60) indicates that it shows 'costs in current dollars', which might be interpreted as being 1979 prices, but is presumably actually the nominal, then year, costs. Whilst this uncertainty does not particularly undermine Deitchman's arguments about increases in weapon system costs, it again makes it difficult to reliably infer the original data values.

The US Army graph is similarly constructed and contains data for anti-tank systems, 155mm artillery and tanks. The data are somewhat sparse, with points being present for three tanks, two artillery pieces and five anti-tank systems. This means that whilst Deitchman is able to fit a straight line to each data set, the predictive value of such an enterprise may be limited. The graph carries a note that 'costs and dates are approximate because sources and bases of 'buy' vary; costs in current dollars; all weapon systems with basic ammunition load'.

The US Navy graph, which has a similar format and carries the same 'current dollars' note represents, we are told (p.105) 'unit acquisition cost[s]'. The content of this diagram is quite eclectic, as it contains separate data series for aircraft carriers, escorts, ship weapon systems and aircraft. It is reassuring to note that the F-4 and F-14 aircraft which appear on both the US Air Force and US Navy graphs show roughly similar costs in each case. As the data is presented in a similar format to Augustine and Kirkpatrick and Pugh it comes as no surprise to discover that none of Deitchman's data is tabulated, so, as usual, any values required must be read off the relevant graph.

In addition to these standard graphs, Deitchman also presents information about the costs of subsystems, including M60 tank fire control systems (p. 244) as well as a tabulation of cost growth in the F-111's avionics (p.242 – 243). The cost discussions are rounded off with an

[24] Which Deitchman defines as (p. xvii) as 'essentially, all of our military forces other than intercontinental nuclear forces'.
[25] Presumably year of entry into service, but this is not specified.

examination of trends in the costs of new generations of Defence Systems against those for new iterations of existing systems. As the data sets each contain only three or four points there is uncertainty about exact values of cost growth in the subject populations, but some relatively consistent upwards trends are visible. Deitchman (p.252) observes that the 'average difference in slope between these two sets of curves is about a factor of 2.5. That is, over a period of time it is less than twice as expensive to improve the capabilities of major systems by continually improving their subsystems than it is to buy wholly new systems incorporating new technology in all their parts'. However, as Kirkpatrick and Pugh have noted, attempting to push existing technology too far can become prohibitively expensive and so Deitchman's data may be censored by the US military declining to undertake upgrades to existing systems, where designing and building a totally new system would be likely to be more cost effective. It is also worth of note that at no point does Deitchman suggest that the new system would be anything other than an evolution of the current systems.

3.4 - The Cost of Seapower

In 1986 Pugh (of Kirkpatrick and Pugh fame) produced his magnum opus 'The Cost of Seapower'. He opens this (p.1) with a number of good questions such as 'What are we to conclude from knowing the cost of, say, a Nelsonian frigate?' and 'How are we to grasp the meaning of such sums of money when our grandfathers tell us that, even in their lifetimes, £1 has declined in value from what was once a good part of a week's housekeeping for a family to what is now unblushingly charged for a single snack?'. The author's approach (p.3) is to appeal to a 'guiding star' that 'making things, buying and selling, getting and spending all remain fundamentally the same whoever does them and whatever the scale upon which they are done'. This approach, which would see no difference between the Wright Brothers building a Wright Flyer in their Dayton workshop and the construction of a modern F-18 aircraft, would appeal to Augustine, but as it potentially represents the inclusion of an arbitrarily large number of warships in a single data set, we shall treat it with some caution.

In Chapter 3 (the first two chapters are concerned with economic and military history) he attempts to establish how one might compare the cost of different ships over time. His example is to compare the cost of HMS Havock (p.73) which cost the Admiralty £34,254 in 1894 with the cost of £29M[26] for a Type 21 frigate at 1974 prices (p.88). He rightly observes that when HMS Havock managed 26.7 knots on trial that this result is still directly comparable with the performance of a modern frigate as distance and time are absolute measures, whose values are constant through time, but that the price of Havock had (p.73) 'a substantially different value even by 1900 – let alone after the nine decades separating that purchase from the present'[27].

He rightly observes that 'we must examine the changing value of money if we are to analyse and understand the costs of warships'. The first element he examines (p.73) is the connection between 'the quantity of money, its speed of circulation and the volume of goods on sale'.

[26] It is unclear how Pugh has arrived at the cost. The outturn cost for these vessels is reported (HC Deb 23 Oct 1989) as ranging from £14.4M to £27.7M.

[27] Pugh fails to note that whilst HMS Havock and the Type 21 might have comparable speeds, the invention of technologies such as radars, helicopters and wireless make the newer vessels more capable.

(Although he does not identify it, this relates to the Quantity Theory of Money (see, for example, Begg et al (1984))). This explains how, with a relatively constant supply of goods, the price may vary in response to the prevailing economic situation.

However, price may also be affected by changes in demand for a good. Pugh illustrates this by examining (p.74 – 75) the consumption the coal and other energy sources in the UK between 1947 and 1982. The cheapness and convenience of oil meant that there was substitution of oil for coal during the 1960s, which reduced demand for coal, resulting in the less economically viable mines being closed and investment concentrated in those with better prospects, which reduced their production costs. These changes in the coal industry, coupled with the substitution of gas and nuclear for oil following the oil crises of the early 1970s then 'reduced the rise in the average cost of fuel and light to about two-thirds of that which it would have been had the pattern of consumption remained unchanged'.

The final economic point that Pugh discusses (p.76) is the phenomenon of increasing productivity. He asserts that 'a large proportion of technical innovations have the objective of enabling existing goods to be made more cheaply or of increasing the value of a product without increasing the cost of its manufacture'. This may be correct in situations where there is a long production run of similar products, Pugh cites examples of the automation of block making for sailing vessels[28] and the substitution of welding for riveting in World War II merchant ship construction and Kronemer and Henneberger (1993 p.30) find continued increases in US aircraft manufacturing productivity between 1972 and 1991. It is, however, unclear whether this assertion holds in the case of military aviation, where a significant proportion of the innovation has been devoted to increasing performance.

Pugh comments that to 'measure the productivity of a firm we must determine its costs so that we can compare them with the value of its output'. In considering the costs of a manufacturer he observers that the costs incurred are for 'paying for energy, materials and labour it consumes'. However the costs of energy and materials consumed by the manufacturer are also composed of 'partly wages and partly payments to other suppliers' and the payments to other suppliers are similarly composed, thus if 'we trace costs back far enough we must find that all costs are labour costs'. Although theory suggests that economic rents and profits should also appear somewhere in the mix, it is very possible that the majority of the expense is eventually attributable to labour costs. Building upon this basis Pugh then (p.77) asserts that the average wages in high technology manufacturing will be similar to those in industry in general because 'for every £1 spent for labour directly engaged upon manufacturing tasks as much as £4 to £6 will be spent elsewhere'. Hence at most 20% of the costs of manufacturing will be connected with the product being constructed, with the balance being tied to general industrial costs. As it has already been argued that all cost are mainly composed of wages Pugh concludes that if we 'consider a complete industry, its costs will vary in proportion to the generality of wages throughout the nation'. Pugh supports this argument by graphing (p.78) the average wages in British Industry from 1971 to 1981 against

[28] In 1808 the Portsmouth Dockyard Block Mills produced 130,000 blocks and it was claimed that 10 unskilled men using the machines could replace 110 skilled block makers (Coad and Guillery (2003, p.1))

the Aerospace Index[29] and demonstrating that their growth rates are similar. He then compares the increases in wages with the RPI and observes that wages grow, on average, at 2% above the RPI per annum, which represents the productivity growth of industry. However, it is important to note that as the RPI is the basis for the comparison, this only represents the productivity growth in industries producing for consumers, it says nothing about the productivity of non-consumer industries[30].

Pugh then discusses the productivity of defence industries, starting from the basis that the 'changing productivity of an individual industry can usually be assessed from a similar comparison of the cost of the inputs to it with the value of the outputs from it' (p.78). However, as we 'lack the independent valuation of its products such as is provided to consumer goods by the prices paid for them on the open market' valuing the outputs is problematic. Although attempts have been made by using performance parameters as a surrogate for value (see for example Bongers and Torres (2014)) Pugh instead prefers to 'infer it from indirect evidence'. He recalls that we have demonstrated the close link between the inputs cost of all industries and makes what he claims is a 'small extrapolation' to conclude that the productivities of all the industries of a nation will increase at broadly the same rate', baring industry specific innovations. Thus he concludes that 'defence industries experience improvements in productivity at broadly the same rate as industry in general'. Where the content of products to be compared remains broadly similar over time, then this is probably a reasonable position to take as one would expect that the inputs consumed in each case would be similar. However where the value[31] of the product changes markedly (say between different generations of military aircraft) this is often related to the utilisation of markedly different inputs, so it is unclear whether Pugh's assumption would be valid.

After some passing comments about the impact of automation (and other features of the substitution of capital for labour) and some pertinent observations about the terminological difficulties of using the phrase 'real cost' to mean something other than the price paid, Pugh moves on to consider how his previous economic arguments should be used to adjust historic costs / prices to a common set of economic conditions.

He observes (p.84) that one way of uplifting the historic cost of HMS Warrior from 1860 would be to uplift the costs of all of the inputs to current prices, which would provide 'the cost of building today an exact replica with the materials and methods of 1860'. Whilst this would provide the current cost of building a direct replica this would be a challenging task as the availability and hence cost of materials such as oak, aluminium and wrought iron have significantly changed since 1860[32]. Whilst a more representative value might be obtained by adjusting the inputs according to the relative weights used in shipbuilding in the years in question, this would be labour intensive and possibly not very satisfactory, as it would rely on the content of an 'average ship' built in that year. Finally we reach (p.86 – 87) Pugh's suggested approach which is to use the RPI as the cost deflator as this, he asserts measures

[29] It is implied, but not stated that this is an Input Cost Index for the Aerospace Industry.
[30] Examining a Producer Price Index would give a view for all producers.
[31] For defence products the military value of an item may not be directly connected with its financial value.
[32] It is likely that there would be similar issues in sourcing craftsmen skilled in working with historic materials.

‘what could have been done with the money’. By applying this approach (p.88) Pugh then concludes that at July 1974 prices HMS Havock of 1894 would cost £0.41M, which is quite modest, compared to the £29M paid for a Type 21 frigate[33].

The final innovation that Pugh brings to the literature is the concept of a ‘specific cost’. He reports, in discussing estimation methods (p.106) that ‘top-down methods usually regard cost as the product of ‘size’ and ‘complexity’. Size is almost invariably described in terms of weight’. However (p.108) ‘Complexity is a more difficult quantity to define. Indeed the only certain thing about it is that it is measured as cost per unit weight’, which leads to the observation that it is better described as ‘specific cost’. So ‘top-down estimating methods are no more that calculating the price of a bag of apples from its weight and the price per lb’. However the ‘trick is to establish the specific cost’.

3.5 - Common Themes

So, having considered these publications that may be regarded as being the foundation of scholarship in the area of Defence System costs, there are a small number of common themes that emerge:

a. Data Transparency and Data Consistency: Exact data values are often not reported and, in some cases, there is a lack of clarity about what the data are meant to represent. This makes replication of the reported results more challenging.
b. Changing Value of Money: There is often no discussion of the options for uplifting historical costs to a common set of economic conditions, nor what result the selected approach is assumed to generate.
c. Changing Quality or Value of Systems Produced: There is little discussion of potential methods of adjusting for changes in the Quality or Value of different generations of Defence Systems.
d. Conclusion: The authors all agree that there is an (apparently inescapable) exponential increase in unit costs.

In their 2016 examination of capability factors as a potential explanation for the growth in the unit costs of Defence Systems Lehtonen and Anteroinen (2016) produced a wide ranging survey of the relevant literature. In addition to Kirkpatrick and Pugh, who they characterise (p.281) as being the ‘most productive authors in the field’, they provide an analysis of authors who have undertaken ‘empirical studies on unit cost growth’ (p.283). Of these the work by Kvalvik and Johansen (2008) is unusual in only using and referencing entirely public domain data sources and in actually tabulating the data. However the authors are cautious about the provenance of the data as ‘even media and diverse websites, such as Wikipedia, are cited as sources’ which means that ‘there cannot be a common definition of what items are included in the price or cost’. They contrast this with the data sources used by Pugh who ‘collected an extensive database from public sources, covering 66 equipment types and subtypes, with 1400 individual observations. The database itself is proprietary and from what is publicised, one cannot verify any particular data point independently’. They compare Pugh’s extensive

[33] We shall revisit this comparison in Chapter 6, where we consider different approaches to uplifting historic costs.

but unverifiable data with that used by Arena et al (2006 and 2008), which they characterise as intensive. The 2006 study covers the manufacturing costs for 37 ship classes, drawn from US budget documents and the 2008 study covers 116 aircraft systems divided into seven types. It is worth noting that although the work by Arena cites its data sources, the actual data used is not tabulated in the report. Lehtonen and Anteroinen also reference a Swedish report (Nordlund (2011)) which contains a mixture of private and public domain data, but which unfortunately normalises all the data to conceal the values of the private data.

They compare the structure and results of the five ‘Unit Cost Growth Studies’ mentioned above and also include Davies et al (2011). Each of these measure the Cost Growth between different generations of Systems. In order to characterise each study they examine four different metrics:

a. Whether the original data are accessible and how well documented they are.
b. The approach used for uplifting the data to common economic conditions.
c. The scope and coverage of the data
d. The growth rates observed by the studies.

The similarities between these measures and the common themes identified above from the 1980s literature suggest that there has not been a significant change in the basic structure of these empirical studies over the past 30 years[34].

Having examined in some depth the literature that established this field of study and the 2016 survey of Lehtonen and Anteroinen, we may execute a slightly less detailed survey of the remaining work, with the focus being on the general direction of intellectual travel and the identification of novel contributions and/or disagreements with the orthodoxy.

Considering the work of Arena et al (2008) (with similar themes appearing in Arena et al (2006)) we find a thorough and lucid discussion of the issues at hand. The authors make a firm initial distinction between the two different types of question that are usually asked about the variation in cost of a Defence product. The first relates to cost growth, which they term cost escalation, which is ‘the general changes in price, typically for a similar item or quantity, between periods of time’ whilst the second covers cost overruns, which they name, cost growth, being ‘the difference between actual and estimated costs’.

The exclusion of Cost Overruns from this current study means that the publications such as Novick’s (1970) work examining cost overruns from the Roman Empire to (relatively) modern times will not feature in this section, nor need we consider the US Government publications relating to Selected Acquisition Reports and Nunn-McCurdy breaches (see GAO (2011) for example). Such works would only appear as a potential source of data on the final cost of various projects.

Jones and Woodhill (2009, p.2) who, when discussing Defence Inflation, write that there are ‘three reasons why costs may change: changes in quality, changes in quantity and pure price

[34] This observation is borne out by detailed examination of the works cited.

change due to economic phenomena'. Arena et al (2008) choose instead to initially identify two separate groups of factors that influence the cost of a project and hence can drive cost growth, these are 'economy-driven variables' and 'customer-driven variables' (p. xvi). Their first category ('economy–driven') covers variables whose behaviour is determined by the overall economy, rather than Service or Government actions because 'these variables are beyond the ability of the Services to control' (p.17). For example 'aircraft manufacturing wage rates increase over time because of overall changes in wages and prices throughout the economy, as well as changes in the prevailing wages manufacturers must pay to retain skilled workers' (p.17). In addition to considering wage costs the authors also examine changes in 'the other two economic factors in aircraft cost growth' (p.24), namely equipment and materials. Finally, in this group, mention is made of an area over which the government might aspire to have some influence, namely the overheads, fees and profits charged by the contractors.

After a thorough examination of how these factors have varied over time (in the US) and hence how they have impacted on various aircraft programmes, the authors move on (p.33) to consider the 'Customer-Driven Factors'. They regard these as having two aspects as the customers 'decide on the number of aircraft they wish to purchase' (Jones and Woodhill's quantity changes) and they also 'determine the characteristics they want these aircraft to have' (Jones and Woodhill's quality changes).

The first aspect of quantity that the authors investigate (p.34) is the impact of the total number of aircraft purchased and the Cost Improvement yielded by the Leaner Effects (see Wright (1936), for example) and then (p.36) the impact of the procurement rate (i.e. the number bought each year). Generally high procurement rates are seen as a good thing as it 'can help spread fixed overhead costs over more aircraft, thus reducing the average unit cost'. Similarly 'higher procurement rates may result in greater and more efficient use of existing plant and tooling, also helping to reduce unit costs' and also 'higher procurement rates may also lead to more efficient use of labor through specialisation'.

Finally they move on to address differences in aircraft characteristics (Jones and Woodhill's quality changes), which it is (implicitly) assumed can be represented by the aircraft's configuration. The initial discussion (p.39) concerns the need to recognise that different versions of the same aircraft type are often produced, for example with the F-15 A and B models being followed in production by the C and D models. It is necessary, the authors reasonably suggest, to take account of such changes in production standard when examining the Learner effects, as production of the changed portions will probably not benefit from the improved learning of the non-changed portions. Finally consideration is given to the impact of the Base Technical Characteristics on the cost of aircraft. Arena et al (2008) conclude after some detailed analysis that (as one might expect) larger, more powerful and more complex aircraft are generally more expensive.

It is of note that although Arena et al (2008) identify economy-driven factors and customer-driven factors, in the latter class, they make no distinction between those which the contractor can attempt to influence and those which are externally imposed. It is possible that this distinction is unnecessary in the case of US projects as past procurements (with the possible exception of F-35) have been entirely driven by US requirements and processes, which could mean that the degree of external imposition of requirements has been consistent over time. In the UK this has probably not been the case as there has been involvement in various multi-national procurements, each with different international partners, with more or less well aligned requirements and also with different management arrangements.

As the UK MOD's Chief Economic Advisor, Neil Davies was well placed to comment on Defence Inflation. He and his co-authors (Davies et al. (2011, p.7) write that 'A measure of "defence inflation" has recently been developed and is published annually by the MoD. This measures yearly change in the cost of defence (goods and services, pay etc.) with quality and quantity held constant, however, the measure excludes the increased cost associated with the purchase of equipment with changed capability'. Thus one might assume that any cost variation above that predicted by the defence inflation measure must be caused by capability changes or as Jones and Woodhill would have it 'quality changes'.

A number of authors (recent examples including Arena et al (2008), Davies et al (2011), Lehtonen and Anteroinen (2016)) have followed this logic and examine how differences between the growth in the inputs and that of the finished products can be explained by changes in the physical and performance characteristics of the products[35]. It appears that all these authors assume that the unexplained cost growth must be caused by these observable physical differences (rather than, for example, structural changes in the producing industry). They all find correlation between some of these characteristics and the unit costs. However, as the overwhelming trend is for more recent products to be larger, more complex and to have higher performance than their predecessors[36], it is unclear whether the changed characteristics are the cause of the cost increases, or are simply correlated with some common driving factors[37].

3.6 - Dissenting Views

Kirkpatrick (2003) was one of the first authors to question whether the orthodoxy of expecting the cost of Defence Systems to rise at the same rate as general goods was realistic. He built on the economics of Tournament Theory which suggests that in some circumstances 'rewards can usefully be based upon the relative performance of economic agents, rather than on their absolute performance' (Bannock et al (2003)), so, for example, a horse that wins a race by running 10% faster than its nearest rival is worth significantly more than the 10% speed advantage would suggest. This approach is particularly relevant in warfare where finishing a close second is generally an unsatisfactory outcome[38]. Kirkpatrick cites examples of similar goods, which include 'equipment for professional sportsmen and sportswomen competing for individual titles' and 'talented footballers or fund managers recruited to rival teams or banks' and observes (p.14) that 'the unit costs of all these categories are increasing

[35] Arena et al (2008) also make allowances for changes in production rates.
[36] Possibly because customers tacitly assume that this represents a quality improvement.
[37] The critical realist viewpoint aims to investigate what the underlying causes might be and hence whether correlation and causation are likely to be related.
[38] As Pyrrhus of Epirus discovered winning by a small margin may also be unsatisfactory.

rapidly, just like those of defence equipment'. This is a step towards identifying one of the potential underlying mechanisms causing the observed effects.

The recent work by Hartley and his co-authors suggests that the driving factor behind 'Augustine Weapon Systems' (Markowski et al (2023, p.294)) is (quoting Augustine) new technology options open 'vast new capability vistas which are then crammed into each new generation of a product'. This results in 'growing system complexity…[which is] in large part, responsible for equipment cost escalation'. However these (p.295) 'Augustine weapons systems are not merely functionally more complex and capable and thus more costly bundles of product attributes relative to military systems embodying previous vintages of military technology. They may also embody transformative, frame-breaking technologies that open vast new capability vistas, which are subsequently funnelled into new generations of innovative military equipment.'

The literature surveyed so far shows general agreement that the costs of Defence Systems are growing at a faster rate than goods in the general economy, that there is some correlation between increases in system physical parameters and the escalation in costs, but, apart from some very specific cases, there is no direct evidence of a significant causal link. Additionally no clear evidence has been discovered suggesting that adjusting the physical characteristics and performance requirements downwards would still produce a useful Defence System, or that this would necessarily significantly affect the programme cost, indeed the concept of Augustine Weapon Systems suggests that this might be counter-productive in the long run.

3.7 - Initial Observations and Research Questions

Reflecting on the current literature, overall there appear to be three distinct questions that are raised, namely:

What did Defence Systems cost in the past?

How should we compare Defence Systems over time?

What price changes should we expect to see over time?

The various authors have set out to make contributions in one or more of these areas, with each aiming to provide a contribution relevant to the audience that they were writing for. For example, the work of Augustine, which was well received by the general industry audience that it was aimed at, appears thin by academic standards. Similarly whilst the work of Kirkpatrick and Pugh (being an academic paper, rather than a short essay) is able to be more thorough, their frustrating lack of detail on the data sources may well be driven by commercial or military confidentiality requirements.

The three questions identified above are very broad and so in Table 3.1 they have been decomposed into a number of subtopics. These subtopics represent specific challenges that must be addressed in the practice of understanding the cost of past Defence Systems and using these data as a basis for forecasting the likely cost of future Systems. Additionally, they represent detailed areas where this research and future work can make a contribution to theory. If future research were to be directed at the area of a single subtopic there would be value in producing a further level of decomposition, however this approach appears suitable for providing structure to this current work.

As we progress through this research, reference will be made back to the fundamental questions and related subtopics identified here. Later in this work, where it is relevant to draw conclusions about the overall contributions being made to theory, reference will be made back to the questions and subtopics identified here.

Question Number	Subtopic Number	Description
1		**What did Defence Systems cost in the past?**
	1.1	What was the scope and amount of these costs?
	1.2	In what context were they incurred?
	1.3	What was delivered for these costs?
2		**How should we compare Defence Systems over time?**
	2.1	How can we compare their cost?
	2.2	How can we compare their quality?
	2.3	How have contextual changes affected outcomes?
3		**How should we expect the prices to change over time?**
	3.1	How will cost of systems change over time?
	3.2	How will customer willingness to pay change?

Table 3.1 - Cost of Defence Systems – Research Questions and Subtopics

We have now considered the broad thrust of literature about the cost of Defence Systems and observed that a positivist viewpoint pervades much of the literature, which leads some authors to expect to be able to explain the observed changes in the cost of Defence Systems by referring to observable system parameters and where this is not possible to invent parameters to explain the discrepancies. We now move on to consider more general literature and changes in the cost and price of goods and services and also the detailed options for uplifting historic costs and prices to a common economic baseline.

Chapter 4 - Literature Review 2 - General Changes in Cost of Goods and Services

As Defence authorities are often criticised about the rate of increase in its costs, relative to other goods and services in the economy, it is now timely to consider broader material that has been written about the behaviour of various goods and services in the wider economy and how one may expect their costs and prices to vary over time.

4.1 - Positional Goods and Scarcity

In his book of 1976[39] *Social Limits to Growth* Fred Hirsch notes (p.2) that to 'a hungry man, the satisfaction derived from a square meal is unaffected by the meals that other people eat or, if he is hungry enough, by anything else they do…In technical terms it is a pure private good'. Conversely (p.2) that 'the quality of the air that the modern citizen breathes in the center of a city depends almost entirely on what his fellow citizens contribute towards countering pollution…In technical terms, it is close to a pure public good'[40]. However he contends that these extreme cases are (p.3) 'relatively few in number' and that often in private consumption 'the satisfaction or utility it yields' is influenced by 'the consumption of the same goods or services by others'.

For example, (p.3) 'the utility of expenditure on a given level of education as a means of access to the most sought after jobs will decline as more people attain that level of education[41]' because the 'value to me of my education depends not only on how much I have but also on how much the man ahead of me in the job line has[42]'. Similarly whilst one might measure the characteristics of a car in terms of maximum speed or of a country cottage in terms of spaciousness the 'satisfaction derived' from them 'depends on the conditions in which they can be used, which will be strongly influenced by how many other people are using them'. Congestion caused either by too much traffic attempting to use a fixed road network or by building too many country cottages in a given area are both forms of (p.3) 'social congestion, which in turn is a major facet of social scarcity'. Scarcity can be produced by 'physical limitations of producing more of them' and also by the 'absorptive limits on their use'.

Later (p.27) Hirsch divides the economy into the 'material economy' which is 'defined as output amenable to continued increase in productivity per unit of labor input' and 'embraces production of physical goods as well as such services as are receptive to mechanization or technological innovation without deterioration in quality as it appears to the consumer' and the positional economy (p.27) which 'relates to all aspects of goods, services, work positions, and other social relationships that are either

[39] For this work the third printing, dated 1978, was consulted.
[40] The Peace and Security provided by effective National Defence is also a pure public good.
[41] This is similar to the diminishing returns from investment in Defence technology referred to by Kirkpatrick and Pugh (1983).
[42] Whilst the concept of education as being a way of accessing more sought after jobs could be valid in some cases, it is not clear from the context whether Hirsch also allows that education can have other value, such as personal satisfaction and growth.

(1) scarce in some absolute or socially imposed sense or

(2) subject to congestion or crowding through more extensive use'[43].

Hirsch then explains (p.29) what 'happens when the material pie grows, while the positional economy remains confined to a fixed size'. This means that 'goods and services sharing some or all of the characteristics of positional goods attract an increasing proportion of family expenditure as family income rises'. Prominent examples he cites are expenditure on education, vacation housing and a variety of personal services'. He then (p.31) starts to examine the 'potential social waste' caused by the increased demand for positional goods and consequent price increases, equating it to the audience at a concert where everybody stands up to get a better view and ends up with nobody having a better view however (p.49) 'at the start of the process some individuals gain a better view by standing on tiptoe, and others are forced to follow if they are to keep their position. If all do follow, whether in the sightseeing crowd or among job-seeking students, everyone expends more resources and ends up in the same position'.

Hirsch coined the term 'positional goods' and then used the concept to examine 'potential social waste' and to suggest some of the challenges government faces in minimising this problem and hence delivering social justice, however (p.190) he observes that 'the transition to a just society is an uncertain road strewn with injustice'. As alluded to by the feedback diagram created by Kirkpatrick and Pugh (1983), military systems that compete with the enemy are similar to the standing concert goers, with all producers straining to gain a slight advantage, by getting slightly higher onto their technological tiptoes. This indicates that, according to Hirsch's definition they are subject to congestion[44], which suggests that they are Positional Goods and hence, following Hirsch's reasoning, it should come as no surprise that as nations become more prosperous spending on some Defence Systems will tend to rise more rapidly than other goods and services (to deliver a constant outcome relative to potential adversaries). Although Defence expenditure might be seen as 'social waste' from some viewpoints, there are clear hazards in attempting to agree and enforce the multi-lateral agreement that would be required to enable risk-free reductions in Defence spending.

A variety of other authors have also sought to define how similar goods should be characterised, often to meet the specific arguments that they are interested in exploring. Schneider (2007) has surveyed various different definitions. In addition to the work of Hirsch he (p.62) considers the definition of Matthews that positional goods 'are defined as those to which access is a function of an individual's income relative to other people's' and observes that this is a tighter definition than that suggested by Hirsch as it excludes (for example) leadership positions, whose availability to individuals may depend (to some degree) on the personal qualities of the individual concerned, rather than simply their wealth. He then contrasts this with the view of Frank who chooses to frame his definition as 'goods that are sought after less because of any absolute property they possess than because they compare favourably with others in their class' (Frank (1999, p.1 – 2) illustrates this with a discussion

[43] As will be shown later it is by no means clear that these two classes span the whole economy.
[44] Unless some novel approach can be found to bypass the congestion.

of the features and potential utility of the Viking-Frontgate Professional Grill)[45]. This approach would confine 'non-positional goods' to those where the quality differences, if any, were regarded as being unimportant by consumers.

A third definition advanced by Ugo Pagano is also considered. His work (Pagano (2007, p.29)) is based on the view that Hirsch included 'two types of goods' in his definition of positional goods. The first type is those whose supply was 'limited by their natural scarcity' whereas the second includes goods 'like power and status whose supply was limited by their social scarcity'. Items in the first category that cannot be reproduced are scarce but (he asserts) 'positional only in the weak sense that the relative positions of the individuals matter to acquire them' and that 'these goods could be consumed independently of the behaviour of other individuals and, indeed, more easily without their interference[46]'.

In the second category the positional nature is 'much stronger: in the act of consumption, individuals must necessarily divide themselves into two different groups of "positive" and "negative" consumers'. For example 'it is impossible for somebody to consume "prestige" or "social superiority" if others do not consume the same "inferiority" similarly no 'soccer team in a tournament can consume three points of advantage if another team is not consuming three points of disadvantage'.

From this standpoint he then defines (p.30) private goods as being 'characterized by the fact that other individuals consume a *zero amount* of what a certain individual consumes' and indeed the other individuals 'are excluded from consumption of these goods'. This is a clear distinction from public goods where such an exclusion cannot be made and 'indeed in the case of a pure public good all the agents will consume the *same positive amount'.* Finally he considers positional goods, such as 'status and power', observing that 'when some individuals consume these goods other individuals must be included in consumption of related negative quantities'. A pure positional good can be defined as a good such that an agent consumes the same but negative amount of what another agent consumes'.

These definitions lead Pagano (p.30) to offer observations about the behaviours we might expect these different types of goods to drive, for example with public goods 'we have the standard under-investment problem in their supply (and in their abatement when they are public bads)' whereas 'in the case of positional goods we have a problem of over investment'. Which is likely to be driven by all 'the agents [trying] to consume positive amounts of these goods and include other individuals in the negative consumption'. Which will mean that '"positional consumption" is much harder, and sometimes more violent, than competition for "private" goods'. Echoing Hirsch he also observes that it is 'also wasteful because individual efforts do often offset each other' which can result in people ending 'up

[45] Applying this definition to Defence Systems might appeal to Kaldor (1982) who (p.5) wrote 'As it becomes more and more difficult to achieve 'improvements' the hardware becomes more complex and sophisticated. This results in increases in the costs of individual weapons. But it does not increase military effectiveness'.
[46] He also dismisses many of the concerns of other authors about difficulties in the equitable distribution of such scarce resources writing (p.29) 'an egalitarian consumption of these goods is not impossible and it is indeed a likely outcome where there are no relevant differences in the social standing, the relative wealth and the preferences of the different individuals'.

with the same outcome that they would have achieved, if they had not dedicated any effort to the improvement of their relative position'.

These definitions reveal an interesting dichotomy, that the peace and security delivered by effective Defence is a pure public good (from a national standpoint) as every resident of a nation 'will consume the same positive amount' of peace and security. However, the individual systems that are used to deliver the Defence Outputs that generate peace and security are positional goods, viewed from the standpoint of competing nations, both in terms of Hirsch's congestion and also in terms of Pagano's definition that by purchasing a system that will deliver some operational advantage one is seeking to (have the option of) inflicting operational disadvantage on a potential adversary. This is the basis of Kirkpatrick and Pugh's primary feedback mechanism, where both Blueland and Redland are seeking to impose operational disadvantage on their adversary.

4.2 - The Cost Disease

Returning to Hirsch's division of the economy into 'material' and 'positional' goods and services it is not clear that this division covers all of the economy. For example Baumol and Bowen (1966) identify that live arts performances are not amenable to the productivity gains seen in manufacturing, but, according to the definitions above, it does not appear that they are generically either scarce or subject to crowding[47]. Additionally the suppliers may also positively attempt to prevent their outputs becoming positional goods so, for example, the Metropolitan Opera (p.173) (in New York) did not 'raise its [ticket] prices to the levels that traffic [would] bear'. This is because 'it is presumably not morally acceptable to turn the Metropolitan Opera House into an institution analogous to an exclusive restaurant in terms of the magnitude of its prices and the economic class of its clientele'.

Having examined trends in the input costs (by comparing string quartet performances and car manufacturing) the authors then examine influences on the pricing of tickets, which they observe (p.172) 'ticket prices have in general lagged behind costs of performance'. They identify three basic influences that can be expected to restrict the rate of increase of ticket prices, namely:

1. The 'disinclination of individual arts organizations to raise their prices, on moral grounds
2. The place of the arts in the ticket purchaser's hierarchy of necessities; and
3. The forces of competition'.

They also note that in the case of higher education there had then been substantial price rises, mitigated to some degree by the availability of scholarships and other forms of fee reductions for less well-off students. They report that it seems to have been agreed that education is something that (p.174) 'is a career requisite [for] a substantial segment of the population, who are willing to pay whatever price is demanded of them' but few people would put the arts into this same category. In the case of live performance there is stiff competition from films,

[47] Naturally one-off performances by popular artists will have an intrinsic scarcity.

television and record / radio performances which are ‘close and low-priced substitutes for live performance’.

The relative rise in costs in areas of the economy that are not amenable to productivity improvements has been entitled ‘Baumol’s Cost Disease’ a phrase (as reported by Baumol (Baumol (2012, p.xii))) coined by Professor Alice Vandermeulen, presumably in her communication ‘A Remission from Baumol’s Disease: Ways to Publish More Articles’ (Vandermeulen (1968)) which explores ways for the labour content of the publication of academic papers to be reduced[48].

The identification of the lack of labour productivity increases in some areas of the economy as a disease (and hence as something that needs to be cured) appears to be reflected by Baumol who later gave them the somewhat pejorative name of ‘the stagnant sector’ when compared with the ‘progressive sector’ where significant productivity increases can be realised (Baumol (2012, p.xx)). However, he does not indicate that the predicament of the ‘stagnant sector’ is any fault of its operators, but seeks to ‘explain why the costs of some labour-intensive services – notably health care and education – increase at persistently above-average rates’.

Baumol explains (p.19) the cost disease by reference to an (in)famous newspaper headline ‘Nearly Half of UK Student Grades are Below Average’[49] and observes that the nature of averages this statement must always be true. Similarly, ‘if the prices of all commodities are not rising at the same pace, then some must be increasing at a rate above average’ which means ‘their inflation adjusted – or real - prices must be rising’. ‘The list of those items whose real costs are rising remains roughly constant, decade after decade…the items in the rising-cost group generally have a handicraft element – that is, a human element not readily replaceable by machines...which makes it difficult to reduce their labour content.’ This is clearly reflected in the car worker vs musician example from Baumol and Bowen where, provided the car workers continue to increase their productivity and the musicians do not, the cost of the musical performance will continue to increase (relative to cars, assuming equality of wage increases).

As long as productivity continues to increase, these cost increases will persist. But even more important, as some years ago, Robinson is reported (Baumol (2012, p.xviii)) as noting that the productivity in almost industries is increasing, hence ‘all industries must be growing less costly in the amount of human labour they require’ and this labour is ‘the real cost humanity incurs in producing a commodity’. Baumol then recognises that (p.xx – xxi) the ‘declining affordability of stagnant-sector products makes them politically contentious and a source of

[48] This did not seek to reduce the effort in the initial authoring, simply the publication process.

[49] It has not been possible to locate this headline, however it may refer to the incident when the head of OFSTED stated that he found it unacceptable that ‘one in five children in primary schools at the age of 11 are leaving primary school without the national average (attainment)’ reported by Curtis (2012). It appears that median performance was initially accepted as a target, with the recognition that not all would meet this standard; however, this was then adjusted by politicians to become the minimum requirement. Clearly, unless the subject population can be made completely homogenous this will never be achieved.

disquiet for average citizens'. However 'as productivity grows' and hence the labour required to produce all the commodities in the economy decreases 'our ability to pay for all of these ever more expensive services' will increase. However if 'governments cannot be led to understand the ideas presented here, then their citizens may be denied vital health, education, and other benefits[50] because they appear to be unaffordable, when in fact they are not'.

Much of the research in this area has been concerned with taking a view on what outcome public policy should aim for and how this might best be achieved. Fortunately, in this case, interest is solely in identifying what the future cost trends for Defence systems (and other complex systems) are likely to be and what the causes might be (and hence what could be done to influence the future trend) and not what the optimum total annual spend on Defence should be.

4.3 - Further Insights from Literature Review

A key element of this current research is understanding the drivers of the cost and price of various goods and services over time and so the main points of interest are what this literature tells us about Positional and other goods.

Initially let us consider Hirsch's material economy, which is identical to Baumol's 'Progressive Sector' as both are defined by showing[51] (significant) progressive increases in the productivity per unit of labour input. One would expect items in this class to show real-terms reductions in cost if products are compared on a like-for-like basis. The converse is Baumol's 'Stagnant Sector' which is those areas that do not show the same (consistent) level of progressive productivity increases. This includes labour intensive areas such as live arts performances and healthcare.

Having considered the expected changes in the input costs of goods and services, we must now come to examine the impact of scarcity and congestion. For many items that show progressive productivity gains the prices will fall in real terms as they will track the costs. However, due to direct scarcity or congestion driving relative scarcity,[52] the prices that individual or government consumers are willing to pay for (perceived) high quality products may exhibit significant growth.

[50] Such as Defence
[51] Or being capable of showing
[52] i.e. The price of higher quality goods and service being increased, due to congestion with lower quality offerings undermining their perceived value.

Chapter 5 - Selection of Case Studies

5.1 - Characterising Potential Case Studies

Before examining the potential case studies it is timely to consider how we might go about determining which of them show relevant similarities to the development and production of Defence Systems. Considering the two dimensions identified at the end of Chapter 3, degree of productivity growth and also the degree of scarcity / congestion experienced by a good or service, we may construct a novel 2 x 2 Boston Consulting style matrix with axes reflecting these two dimensions (See Figure 4).

This approach produces four different categories for which names are required. The right hand half of the matrix represents Goods and Services that are amenable to higher levels of Productivity Growth (Hirsch's Material Economy), where prices fall, in real terms, relative to average wages, and the left half represents those which are subject to lower rates of Productivity Growth (e.g. Baumol and Bowen's String Quartet). The bottom half of the matrix represents those Goods and Services where scarcity does not have a significant impact on the value to the consumer, whereas the top half represents those where, for whatever reason, the absolute or relative scarcity have significant consumer value.

It now remains to choose names to identify each of the four categories. The bottom right category represents Goods and Services that show high productivity growth and where

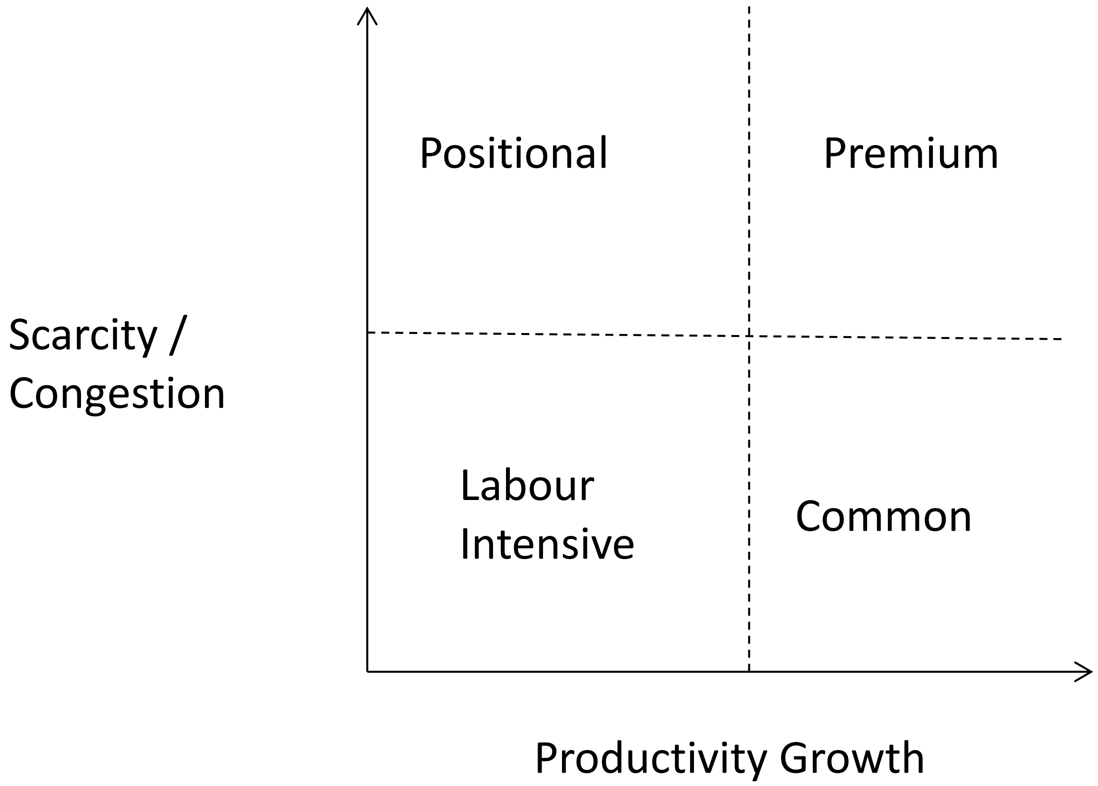

Figure 4 - Goods and Services Productivity Growth vs Scarcity Matrix

scarcity does not impact on consumer value. Many routine items in the economy meet these criteria, especially where there are many suppliers, many consumers and low barriers to entry. Therefore it is reasonable to identify this category as *Common* goods and services. Now moving to the bottom left category one finds that this represents goods and services where scarcity does not impact on consumer value, but where there is lower productivity growth, for example live arts performances. The distinguishing feature of these items is the level of human inputs, so let us identify them as *Labour Intensive*.

The next issue is a name for the categories where the scarcity has a high impact on their value to the consumer. Those with high productivity growth would be *Common* goods, if some aspect of the product or its context did not introduce a scarcity related value premium. The designation *Premium* is therefore chosen for this category. The final category represents goods and service which exhibit a low productivity growth (and hence would fit Hirsch's definition of Positional). Additionally they would tend to also satisfy Matthews' definition of Positional, as their scarcity and price growth (relative to the rest of the economy) will tend to favour those with greater resources to devote to the items. Therefore the name *Positional* may be selected for this category. Many of the Tournament Good that were previously discussed (such as star footballers or fund managers) will fall into this category, due to their natural scarcity. Where other goods, such as equipment for individual sportsmen and women, are not necessarily naturally scarce, it is likely to be in the manufacturer's interest to create scarcity to ensure that they remain in the *Premium* category.

As the Augustine Weapon Systems discussed by Hartley and his co-authors are constantly striving to embody potentially frame breaking technologies and hence pushing the boundaries of what system complexity can be fitted into a given space, weight and power envelope, it is inevitable that these will show low productivity growth as a limited amount of the learning from one system will be applicable to its successor. Additionally the technical and financial challenge of producing such systems will put them firmly into the *Positional* quadrant.

The discussion in Chapter 2 and Chapter 3 has provided some initial insights into how we should expect the cost and price of Defence Systems to change over time (Question 3 in Table 3.1). Although we have not yet considered how the cost of producing the Systems is likely to[53] change (Subtopic 3.1) we have observed that when a nation is threatened[54], then it would be expected that they want to own the best Defence Systems that they can. The positional nature of these Systems means that better quality is expected to cost more, but consistent with Kirkpatrick's view of them as Tournament Goods, a slight quality advantage can provide a significantly different outcome in combat. Thus we observe that it is absolutely reasonable for nations to pay as much as can be afforded for these Goods (Subtopic 3.2). However, as Figure 3 reminds us, the physical Systems are but one of the components necessary to produce effective fighting power. Although the literature tends to assume that 'top trumps approach[55]' to assessing quality is valid, we currently have no direct evidence to

[53] Or ought to
[54] Or perceives that they need to deter threats.
[55] Similar to the game of Top Trumps the higher quality system will have better values for its relevant performance characteristics.

support this (Subtopic 2.2).

In the rest of this Chapter we consider a number of case studies to determine which of them might support our understanding of the factors driving the cost of Defence Systems. To aid this we shall consider which of the quadrants of Figure 4 each of the different potential case studies falls into. Whilst being a useful initial step, this may not capture the whole range of factors that provide the context for each data set, therefore Table 5.1 will also be completed for each case study, to more broadly capture the context and potential mechanisms and effects surrounding the data collected for each study.

Element	Description
Outcomes	The end results of the activity and how the relevant stakeholders value them
Output Quality Measures	Ways in which the quality of the direct outputs of the activity can be (usefully) measured[56]
Input Quality, Cost and Volume Measures	Absolute or relative changes in the input factors to the activity
Environment	The general structure of the environment and its rate and degree of change

Table 5.1 - Contextual factors for each data set

Globally there are relatively few projects to design and produce Defence Systems and with those that do occur there are challenges[57] assembling and analysing relevant data. By their nature the public availability of the data necessary to adequately characterise a project is often limited and, additionally, the constant striving for operational advantage (through improved product quality) makes it challenging to compare their outputs. It was therefore decided to construct case studies, to allow the examination of goods and services where data availability and quality measurement challenges are less severe, to provide insight into the mechanisms and effects that might be influencing the cost of Defence Systems.

Price (or cost where relevant) and contextual information were therefore gathered to allow the compilation of data sets covering eleven different types of goods and services as it was initially unclear which would be most appropriate to develop into full case studies. The details of each data set are briefly considered below, together with an examination of the changes over time of the quality of the inputs and the outputs. The results are then summarised and additionally the position of each case study is plotted on a Productivity Growth vs Scarcity matrix. In assessing the changes in input quality an assessment is made of whether the impact of quality changes can be directly measured or whether it is simply assumed that the quality changes must have a positive impact. In the case of changes to the output quality a judgement is made of whether the quality changes can be measured directly, or are reliant on the perceptions of consumers and other stakeholders. Additionally it is noted

[56] Relevant to Topic 2.2.
[57] Often relating to security or commercial restrictions.

whether the quality is most appropriately measured in absolute or relative terms[58].

A variety of data sets was selected, firstly to provide coverage of different areas of the Productivity / Scarcity and Quality Grids (see Figure 12 and Figure 13) and secondly to provide some redundancy should any of the data sets not be suitable for further development.

Finally the best data sets with relevant similarities to the Defence Systems data are selected for further development as detailed case studies. The categorisations applied in this Chapter suggest that relevance to Defence Systems can be measured by similarities in the context of production (in terms of Productivity Growth and Scarcity / Congestion) and also how the input and output quality of the products are assessed. The usefulness of these measures is confirmed by the results of the fully developed case studies.

5.2 - UK Consumer Staples

The UK Office for National Statistics (ONS) produced a number of data series, as part of their Retail Price Index collection, which illustrate how the average prices of various consumer staple goods (mainly food and drink) have changed between 1914 and 2004 (ONS (2019)) and there are various additional series that update some of the goods with data up to 2018. Although there are variations in the items covered, as consumer tastes have changed, data is present for a sufficient proportion of the period to allow complete or near complete series to be extracted for standard weights of Streaky Bacon, Flour, White Bread, Tea, Sugar, Milk and Butter. In principle such a dataset ought to be useful as the nature of many of the products is, essentially, unchanged across the 90 years covered.

However, on further investigation, it becomes clear that for significant sections of the period covered the prices were not set by market forces. For example, the European Union[59] (EU) price setting and import tariff mechanisms were introduced in 1968[60] covering a wide range of agricultural products as part of the Common Agricultural Policy (CAP), which aimed to promote European self-sufficiency in food production and to support agriculture. The approach to this was for the EU to set prices and intervene to support the prices at which agricultural products could be sold within the Union’s borders. Production quotas were introduced later to support this effort.

Quotas and target prices were not invented by the EU, as, for example, from 1937 to 1968 the International Sugar Council attempted to regulate global supply and demand in line with the series of International Sugar Agreements, by setting export quotas and global prices. The UK also imposed a range of tariffs on imported commodities before joining the EU and hence being subject to the EU-wide CAP approach. For example, in the 1930s – 1950s imports of food from the British Empire (and later the Commonwealth) were tariff free, whereas those from other nations were not. Additionally, during WW1 and WW2 there was government control over the price of many foods. Thus one observes that, for these items, in the UK, for

[58] Absolute quality is represented by direct measures of a product or service, such as the speed of an aeroplane, whereas relative quality is measured or perceived relative to other options on the market.
[59] Previously the European Economic Community
[60] But did not impact the UK until it joined the EU on 1 January 1973.

much of this period market forces did not consistently apply and in many cases prices were set at a level that bureaucrats or politicians considered ‘reasonable’. Therefore, whilst the data does correctly report the prices paid by UK consumers, it is unclear how these are connected to the cost of production and any trends in that measure. Care would therefore have to be taken in deciding what economic conclusions could be drawn from this dataset.

Overall the farming and processing activities that produce consumer staples still yield goods that are similar to those available in 1914. Provided that the goods meet relevant safety standards there are no clear direct quality measures that can be applied, as options such as nutritional value appear less relevant to consumers than their perceptions of quality. One might argue whether a loaf of bread produced using the Chorleywood bread process is of the same quality as a traditionally kneaded artisanal loaf, and the relative qualities of these two products, as judged by the consumers, would be evident in the relative consumption preferences over time. However, as the data simply covers the average price we have no visibility of any such subtleties.

We do, however, observe that process improvements appear to have been implemented, at least in UK agriculture. For example, available data shows that the number of UK agricultural labourers has fallen from a high of 892,000 in 1923 and by 2018 had stabilised at about 180,000. The same data also shows that productivity roughly doubled between 1953 and 2018 (Zayed and Loft (2019)). Whilst each labourer may not work any harder than in 1914 the progression from horse power to tractors and their associated machinery represents a clear input quality improvement that has resulted in demonstrable productivity benefits.

Zayed and Loft (2019) also indicate that overall the most significant change in activity has been the reduction in orchards, whose area has fallen by about 80% since 1951. The authors remark (p.7) that this ‘is frequently seen as an indication of the industrialisation of agriculture’ and presumably the difficulties in introducing such industrialisation into orchards.

The context of this data set is summarised below in Table 5.2.

Element	Description
Outcomes	Although the details of production and presentation may have evolved, these products still fill the same needs as their predecessors at the start of the data series.
Output Quality Measures	The data represents ‘average price’ and hence there is no way to measure the perceived relative quality of products that consumer buying choices might reflect.
Input Quality, Cost and Volume Measures	Whilst the quality of the labourer may not have changed, mechanisation and other advances have increased the quality of some inputs and also the overall productivity.
Environment	The CAP and other government initiatives provided varying incentives and disincentives to

	farmers during this period. In recent years concerns about ethical production have started to become relevant.

Table 5.2 - Context for UK Staples Data Series 1914 – 2004

5.3 - Model T Ford

The next data set concerns the Model T Ford, the first mass market car, whose absolute quality remained very similar over time, but where there were significant, constant improvements in the production process. The specification of the Model T was not entirely static over time as various minor adjustments were made, but there was only one step change in quality[61], but even this did not make the old versions obsolete, simply rendering the newer versions slightly more desirable (i.e. improving the relative quality).

As may be seen from Figure 5 the nominal price of the Model T actually fell over time. There was a rise in the period 1918 – 1920, part of which is likely to have been caused by post-war inflationary pressures, but part of which may also be due to the vehicle specification change mentioned above. Given that wages rose over the period of production, the overall reduction in the nominal price indicates the great progress that was made in reducing the direct assembly labour and also, presumably, the labour embodied in the components. The price of this car also fell in real terms, relative to average prices in the economy and also relative to wages.

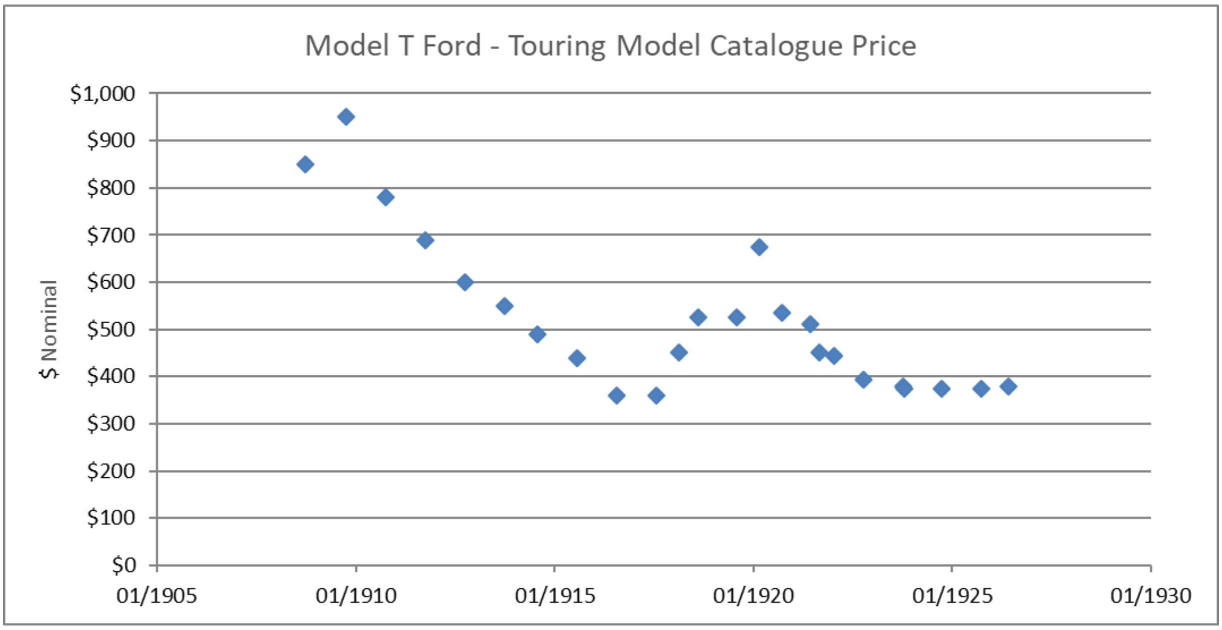

Figure 5 - Model T Ford - Touring Model Catalogue Price

Notes: Source – MFTCA (2023), Values - $ Nominal

Model T sales peaked in 1923, but then fell as the relative quality of the Model T, as perceived by consumers, fell, as competitors launched new models. Its production was discontinued in 1928 as the reduced level of sales meant it was no longer profitable. It was

[61] The fitting of demountable wheels and a starter as standard from March 1920.

replaced with the Ford Model A. The overall context for the Model T Ford price data is shown in Table 5.3.

Element	Description
Outcomes	The end products provided affordable transport to a broad range of consumers.
Output Quality Measures	There were various minor specification changes, which improved the absolute quality of the product. However, public perception in the mid-1920s was that the relative quality was falling, when compared with newer competitor models.
Input Quality, Cost and Volume Measures	Although process improvements may have reduced scrap and wastage, the volume of raw materials is likely to have remained broadly similar. Cost of wages increased, but the significant reductions in labour hours per car provided direct benefits.
Environment	As this data series ends before the recession of 1929, the wider US economy experienced significant real growth in wages and GDP throughout this period.

Table 5.3 - Context for Model T Ford Production

5.4 - Periodicals – Times Newspaper and Beano Comic

We next come to two similar data sets from the publishing field that contain cover price and circulation data from the Times[62] newspaper and also for the Beano comic. In each case, although the page count and physical production quality has improved over time, the essential nature of the product, in terms of what is supplies to the reader, has remained constant. It is also notable, from a business model viewpoint, that both publications involve a significant human effort to create and distribute each issue, but the marginal cost of producing an additional copy, given the previous investment, is very low.

The Times was first published in 1785 as the Daily Universal Register, becoming the Times on 1 January 1788. Initially the cover price was 2½d, rising to 3d in 1785. At this time there were only four pages in a standard issue and part of the cover price was driven by the Stamp Duty, which was levied at 1d per whole newspaper sheet, with additional levies on advertisements. This was raised over the years finally reaching 4d per copy in 1815 before being reduced to 1d per copy in 1836 and finally being abolished in 1855.

By examining scans of historic copies of the Times, that are available in a number of on-line repositories, it is possible to track increases in the cover price. Coupling this with the daily circulation, it would be possible to gain a view on how the sales income has changed over time. However, as the existence of free newspapers such as Metro and Evening Standard illustrates, the advertising income can form a significant proportion of income and so it

[62] Of London.

would require some fairly broad assumptions to arrive at an assessment of how the Times' production costs have changed over time.

For the Beano comic it has also been possible to recover the data relating to the cover price from on-line copies and also from other sources. It has been possible to gather some circulation data from public domain sources and also some was supplied by the publisher (DC Thompson). The Beano was first published on 30 July 1938, when it was priced at 2d. The first price rise was on 15 October 1960 when the price rose to 3d. The circulation of issue 1 of the Beano was 443,000 with the peak circulation being achieved in April 1950 at 1,974,072 copies. It is interesting to note that this is approximately double the peak circulation achieved by the Times newspaper.

As it is understood that the Beano carried no external advertising[63] until 1988, it would be possible to calculate the income at different points throughout its life, however, as its publisher produces a wide range of publications, it is possible that there was cross-subsidisation between different products. This situation would again require some careful assumptions to navigate.

It would be possible to devise a measure of absolute production quality, based on metrics such as number of pages, quality of paper and number of coloured pages. However, this might not be a useful exercise, as it appears unlikely that this is the main basis upon which potential readers value these publications. It is possible that the relative production quality, compared with rival publications, might form part of the decision making, however, it appears likely that the relative content quality will be a more important factor. Additionally, in recent years, non-print media have also started to provide significant competition as they can provide the relevant information, entertainment and advertiser exposure in a (potentially) more convenient form.

Over time there have been a range of improvements in the printing industry that aim to reduce production times and also to reduce the resources devoted to typesetting and the preparation of art work for printing. Although these improvements in the quality of the input factors produce demonstrable benefits, the same improvement is unlikely to be evident in the actual creation of writing and artwork.

Element	Description
Outcomes	These products provide information and entertainment to the purchasers and exposure for the advertisers.
Output Quality Measures	Whilst creation of absolute production quality measures would be possible, the relative content quality, as perceived by potential readers is likely to be more relevant.
Input Quality, Cost and Volume Measures	Demonstrable benefits have been realised from the improvement in production times

[63] i.e. It only promoted the products of DC Thompson, its publisher.

	and costs and also in reducing the resources devoted to typesetting and the preparation of artwork. However, the actual creation of writing and artwork is not subject to such process improvements.
Environment	For most of this period the environment was stable, with competition from other printed media representing the main challenge. However the appearance of the internet as a widespread source of news and entertainment in about 2000 introduced an additional form of challenge.

Table 5.4 - Context for Production of Times and Beano

5.5 - UK House Prices

The Nationwide Building Society produces an index of UK house prices which has values for each quarter since Q4 of 1952 (Nationwide (2023)). Figure 6 shows that, apart from dips during the recessions of the early 1990s and the late 2000s, there have been consistent increases in the price of houses, in nominal terms. Despite these hiccups in growth, overall house prices have risen significantly when compared with average prices in the UK economy, having risen four-fold, in real terms, since the mid-1950s.

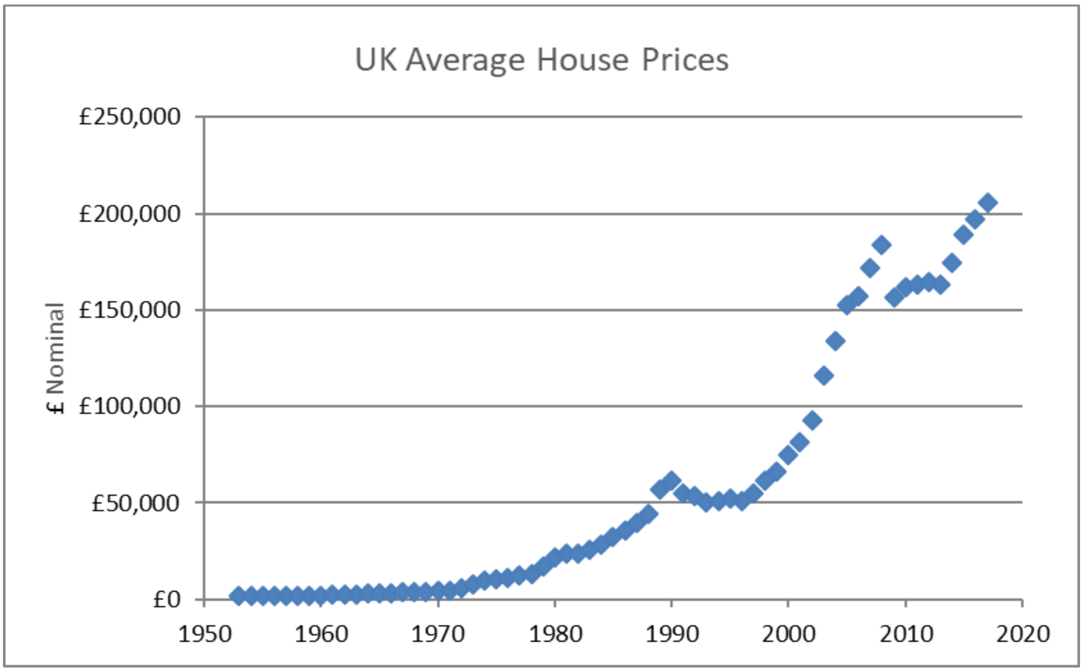

Figure 6 - UK Average House Prices

Notes: Source – Nationwide (2023), Values £ Nominal

Whilst the quality and volume of the housing stock of the UK does change over time, this is a slow process, with the annual increase in stock having been between 0.54% and 1.0% for the period between 2001 and 2017 (Ministry of Housing, Communities and Local Government (2017)). Thus, presumably, the price increases shown in Figure 6 mainly represent increases due to the changing value of money, the increasing premium that a purchaser is willing to pay for one of these scarce items and also, to some degree, intrinsic change in quality.

There have been various improvements in the processes for manufacturing building materials and for the processes of building houses. However, despite these demonstrable improvements, the main influences on the price of houses has been their perceived quality (i.e. value), which is probably driven by increasing individual wealth and also scarcity of houses in desirable and convenient locations.

Element	Description
Outcomes	The end products provide accommodation for UK residents.
Output Quality Measures	Traditional absolute quality measures relate to floor area, number of rooms and facilities, with more recent metrics including setting, local facilities and greenness. However, the most meaningful measure (in terms of house selling prices) is the market price, for a given house in a given location.
Input Quality, Cost and Volume Measures	Process improvements have been made that allow houses to be built more speedily than before and which minimise the cost of labour and materials. However new house construction represents a small part of the UK housing market and the land value also represents a significant proportion of the price.
Environment	The UK is a relatively densely populated nation and has planning laws that restrict the building of houses. This results in scarcity / congestion of land for housing, which tends to result in house prices being driven by this rather than the building costs.

Table 5.5 - Context for UK House Price Data

5.6 - Super Bowl Tickets

Also subject to scarcity are tickets for what in 2024 is billed as the ‘Ultimate Football Experience’ (Ticketmaster, 2023), the Super Bowl. The first Super Bowl was staged on 15 Jan 1967 at the Los Angeles Memorial Colosseum when the Green Bay Packers beat the Kansas City Chiefs 35-10. Uniquely it was broadcast simultaneously by NBC and CBS[64] and was the only Super Bowl that was not a sell-out, with some 33,000 of the 94,000 seat capacity being unfilled. Since then the popularity of this end of season game has risen significantly, with the cost of tickets rising very consistently over time, in 50 years increasing

[64] The two main US television networks.

from $10 in 1967 to $2,800 in 2018[65]. In real terms this is an increase of at least an order of magnitude.

As each game since 1968 has been a sell-out, despite the significant increase in price, one must assume that the scarcity of the tickets has been sufficient to allow the market to bear the increased cost. Additionally, as wages only rose by a factor of 10 between 1967 and 2018, unless the staffing at the event is now twenty eight times higher than in 1967, or other running costs have increased significantly, one must assume that additional profits are being made from ticket sales.

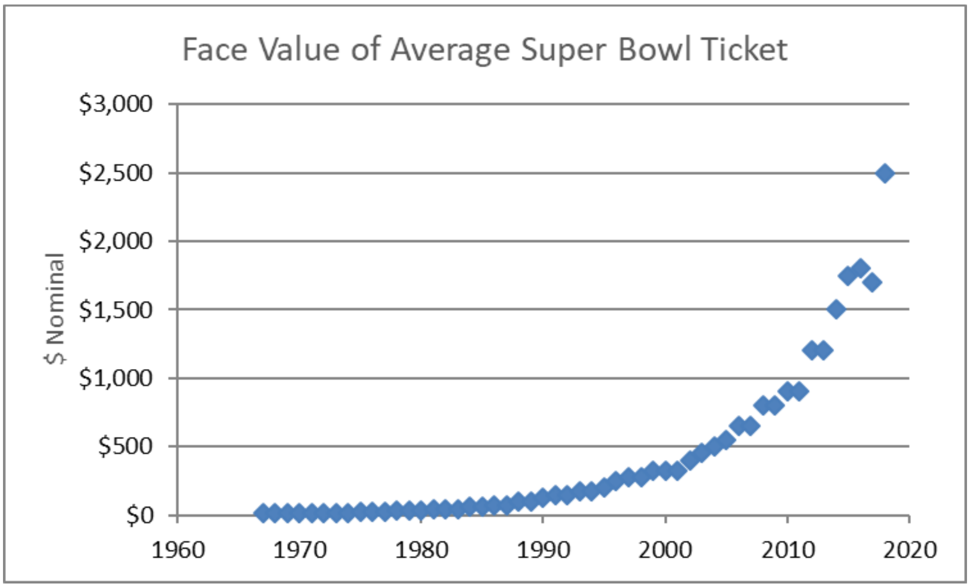

Figure 7 - Face Value of Average Super Bowl Tickets

Notes: Source – Dallas News (2011), with additional data from more recent news stories. Values - $ Nominal

Element	Description
Outcomes	The product allows the holder to witness this sporting event in real life.
Output Quality Measures	Whilst the volume of marketing and the quality of the half time show[66] have improved over the period in question, this is still a ticket to watch a game of US football. The ability to still sell tickets at these higher prices suggests that the perceived quality must be sufficient for the numbers available.
Input Quality, Cost and Volume Measures	Whilst the average wages in the US have increased by an order of magnitude over this period, it is not clear that the quality of volume of the inputs should have changed significantly.

65 Values taken from on line news article (Dallas News (2011)) and other more recent sources.
66 Early Super Bowls tended to feature University marching bands and other performances of that ilk. However, since the appearance of Michael Jackson at half time show for Super Bowl XXVII major stars have performed.

Environment	Although there have been recessions during this period, the US has, overall seen sustained growth.

Table 5.6 - Context for Super Bowl Average Ticket Price

5.7 - Stradivarius Violins

A good whose quality and content are (essentially) unchanging over time is represented by the auction prices of Stradivarius violins where the items in question have never, since their original manufacture varied in quality, number and content[67]. Thus, presumably, the price increases in nominal terms shown in Figure 8 represent the increasing premium that a purchaser is willing to pay for one of these items, rather than reflecting any intrinsic change in quantity or perceived quality.

Considering this Figure, we note that the vertical scale is logarithmic, to ensure that all the data points could be accommodated with reasonable legibility, Figure 9 shows the same data, but has a linear scale for consistency with other graphs in this section. We remark that from an average price of a little over £100 in the 1820 – 1850 period there has been a significant increase in price, as an auction price of £1M is now not unexpected. There is a reasonable size population represented here, as it covers 36 individual violins and 47 transactions.

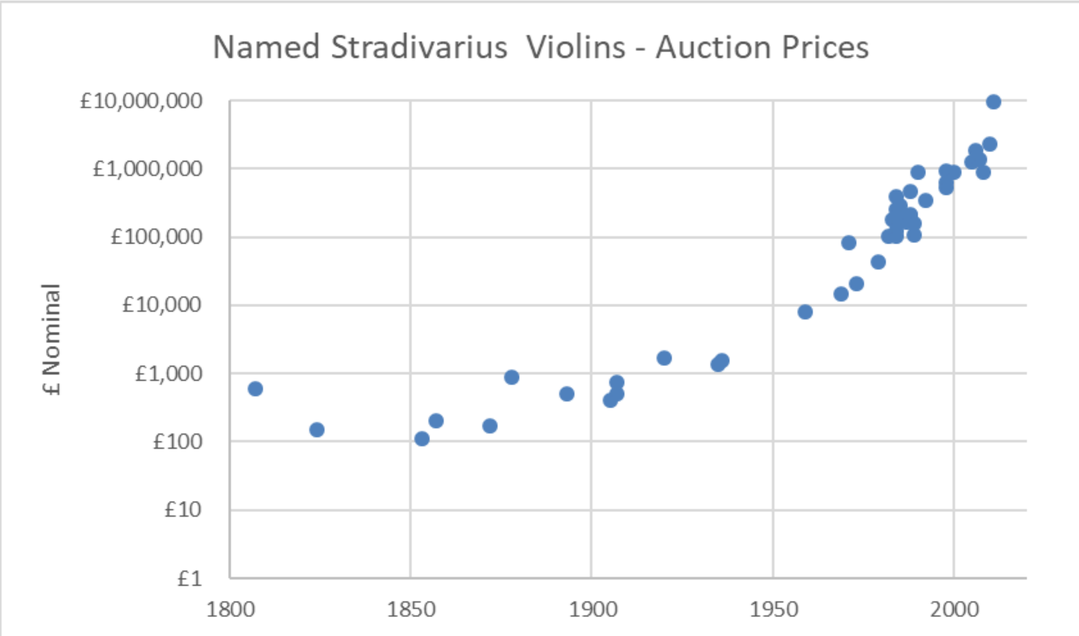

Figure 8 - Named Stradivarius Violins - Auction Prices – Log Scale

Notes: Sources – See Annex A, Values - £ Nominal

[67] Apart from standard maintenance, which should not change the fundamental nature of the instrument.

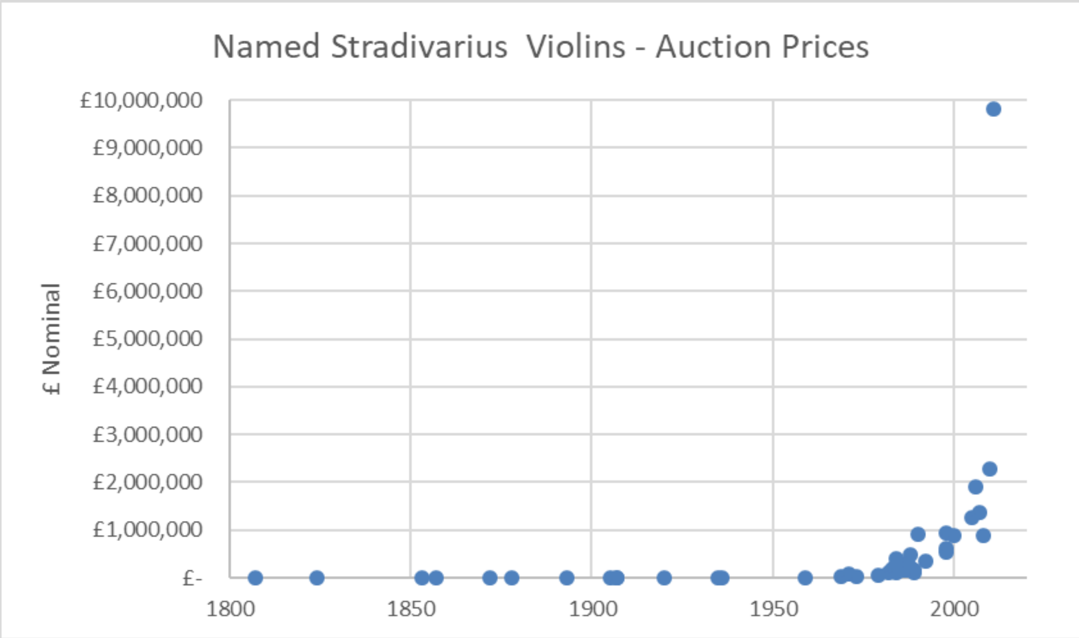

Figure 9 - Named Stradivarius Violins - Auction Prices - Linear Scale

Notes: Sources – See Annex A, Values - £ Nominal

An initial consideration might be that auction prices have simply risen to match the average price rises in the economy, however, analysis shows that prices have actually increased, in real terms, by almost two orders of magnitude sine the 1825 – 1840 period. Compared with average wages, over the same period, the prices have increased by about one order of magnitude. However, it may also be shown that when compared with general wealth available (as measured by UK nominal GDP) the price rises are much less significant.

This tends to suggest that, whilst average working members of the population have been priced out of the Stradivarius market, those individuals, corporations and other bodies who have access to national-levels of resources have been willing to spend the same proportion of their wealth on a Stradivarius since (say) 1850. It is also worth noting that there is dispute over whether Stradivarius violins are intrinsically better than other high quality violins, however their reputation for quality seems to be sufficient to attract well-resourced people and organisations to purchase them.

Element	Description
Outcomes	The owner has a Stradivarius violin to admire. On occasions some of them may be played.
Output Quality Measures	The violins (and other stringed instruments) produced by Antonio Stradivari are widely regarded as being the highest quality ever produced, although there is little evidence from Physics to support this assertion[68]. This reputation for quality is sufficient to

[68] See Gough (2000), for example, for a discussion of this topic.

	attract well-resourced individuals and corporations willing to pay significant sums for them.
Input Quality, Cost and Volume Measures	This represents one of the ultimate scarce commodities. There are a finite number of these and they are offered for sale infrequently.
Environment	There has been general growth in the economy over this period and the reputation of these violins has also grown.

Table 5.7 - Context for Stradivarius violin auction prices

5.8 - New York Subway Fare

A very different class of service is considered in the next data set - a single journey on the New York subway. Like the production of staple goods it has provided the same consumer outcome throughout its history, although there have arguably been some improvements in the quality of the outcome and the efficiency with which it is delivered.

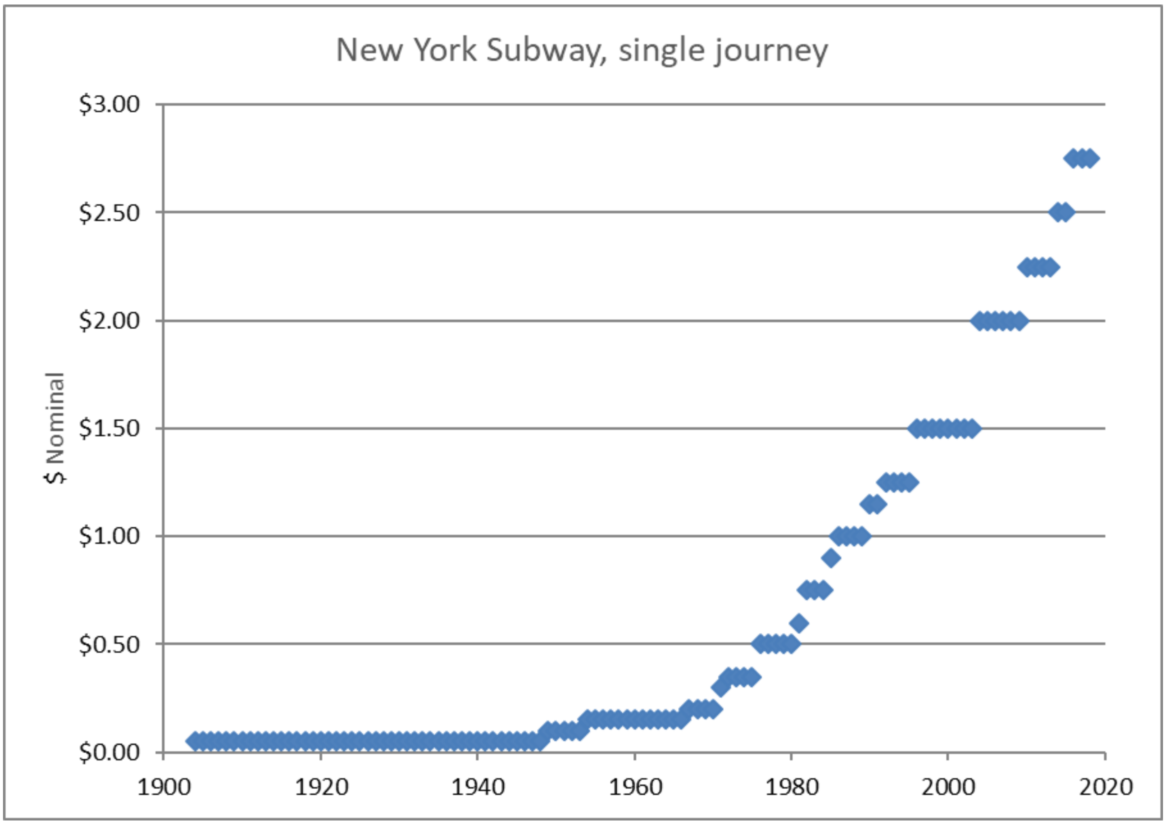

Figure 10 - New York Subway, single journey price

Notes: Source – Business Insider (2015), individual points confirmed by news reports. Values - $ Nominal

The subway system was opened in 1904 and until 1940 was operated by two private companies. In that year operation was taken over by the City (Cudhay (1995, p.118)) and since that date has been operated by the Board of Transportation (later the Transit Authority). In nominal terms the expected monotonic increase in fares is observed, see Figure 10, and in this case we note that for the first 44 years there were no changes in the nominal price of a journey[69], which naturally resulted in a real-terms fall in price. Since that point increases have broadly kept pace with rises in skilled wages. This suggests that either there have been minimal reductions in the human inputs required to deliver this service, or that the reductions that there have been are matched by increases in the costs of other inputs.

Element	Description
Outcomes	This service provides relatively speedy and affordable transport to a broad range of consumers.
Output Quality Measures	The absolute quality of the journey has not shown significant improvement. However, relative to other options for travelling around New York, providing relatively speedy and affordable transit through a busy city, is still sufficient, relative to other options.
Input Quality, Cost and Volume Measures	Although there have been incremental improvements in safety[70] and comfort[71], through improved car construction and signalling, there have been no fundamental changes in the inputs to this service.
Environment	Whilst private companies were running the subway the commercial agreement did not allow them to raise fares. Since operations were taken over by the city in 1940 regular fare rises have been required to (broadly) match revenue to operating costs.

Table 5.8 - Context for New York subway fare

5.9 - US College Education (University of Pennsylvania)

The University of Pennsylvania grew from a civic desire articulated by Benjamin Franklin to 'create a college to educate future generations of Philadelphians' (Friedman (1996)) to which end he produced a pamphlet in 1749 outlining his vision for 'a school to be known as the 'Publick Academy of Philadelphia'' (Friedman (1996)).

Department	Date Founded
The Department of Arts	1755
The Department of Medicine	1765
The Department of Law	1789 (Re-established 1850)
The Towne Scientific School	1872
The Department of Philosophy	1882

Table 5.9 - Foundation Dates of University of Pennsylvania Departments

Notes: Source – University of Pennsylvania (1994, p.11)

[69] This was a result of the 5c fare being written into the contract between the City and the operators. Initially the operators used this provision to resist calls for lower fares, but as operating costs rose their inability to raise fares resulted in underinvestment in the system.
[70] Composite construction (i.e. wood and metal) cars were banned from the subway in 1916 (Davis (1985)).
[71] Air conditioning in cars was introduced in 1967 (Davis (1985)).

The University developed over the years, with the dates of establishment of the earliest departments being shown in Table 5.9.

The University published an annual catalogue that reflected (inter alia) the details of the academic staff, the degrees on offer and the tuition fees and other costs. This cost data has been compiled by the University's University Archives and Record Center for the years 1900 – 2016 (Lloyd and Heavens (2016)) and, additionally, for the period up to 1923 the catalogues are also available on-line to refer to. Earlier data, (dating from 1828 to 1899) have been compiled by the author from the on-line copies of historic catalogues which have been digitised by the University of Pennsylvania (University of Pennsylvania (2019a)) and are available online.

The longest run of data is for the Undergraduate Arts degree, which formed the basis of the original Liberal Arts education offered by the institution and data are available from 1828 through to 2018. The fees for PhDs do not appear until the formation of the Department of Philosophy in 1882, but are available from then until 2016 (with gaps for 1921, 1922 and 1982 where, for some reason, the relevant data appears not to have been included in the catalogue). The catalogues often also give an indication of the likely cost of board and lodging either in the city or later in the University accommodation. From 1997 the University website also includes relevant data, which allows cross checking with the print data up to 2016 and for 2017 and 2018 provides the only available data.

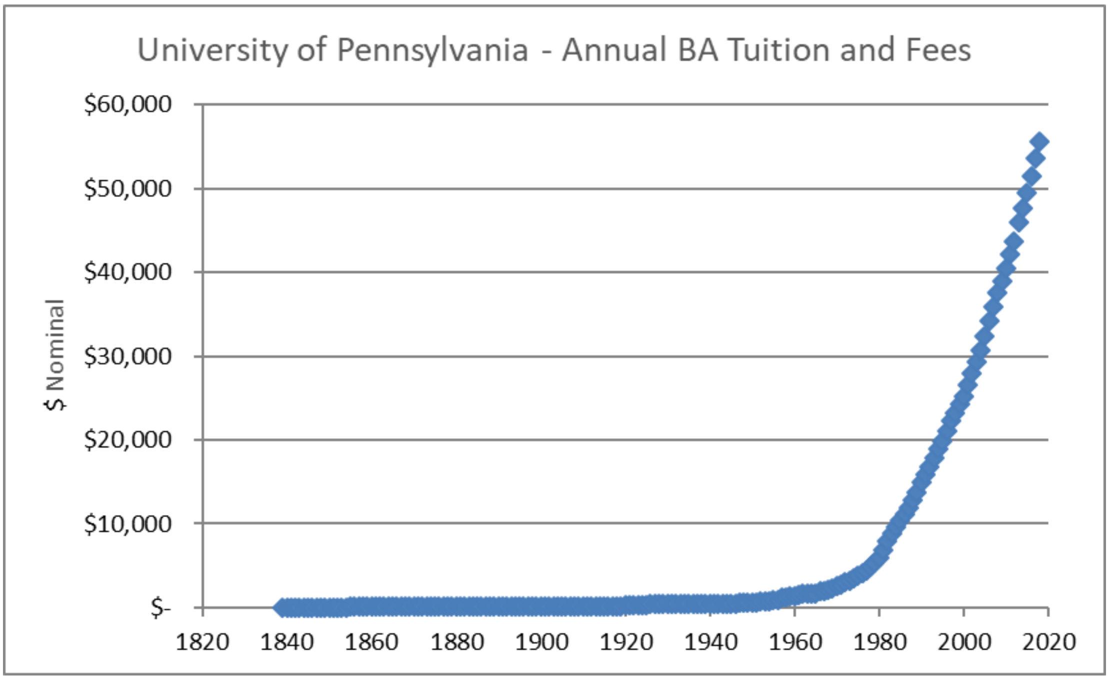

Figure 11 - University of Pennsylvania, annual BA nominal fees.

Notes: Source – See Annex B, Values - $Nominal

It is challenging to measure the output quality of a college education in absolute terms. The Grade Point Average (GPA) achieved by a student should give an indication of their relative academic position in their class, as assessed by the assignments and examinations. However, it is by no means certain that this grading represents anything meaningful in absolute terms.

Fortunately for graduates there seems to be general acceptance that a high GPA from a prestigious university represents a high quality education, which in turn is assumed to make them an excellent prospect in the jobs market. As we shall later discuss (see Chapter 7), there is a similar general assumption that more expensive facilities and more teaching staff represents a higher quality offering in an independent school.

Element	Description
Outcomes	The service provides the opportunity for the student to receive an education, which may give them an advantage in the jobs market.
Output Quality Measures	There tends to be a perception that a degree with a high GPA from a university with a good reputation must represent relatively good quality.
Input Quality, Cost and Volume Measures	There have been many innovations that reduce the burden of college administrative activities. The essential nature of a taught degree has not changed over time, however there have been various technological and pedagogic innovations that have been widely adopted. These increase the cost and volume of inputs when compared with the early 19th century. It is generally assumed that these changes improve the quality of the education available.
Environment	During this period the number of degrees awarded by US institutions has risen significantly. Thus, whilst the possession of any degree was once a mark of education, the perceived quality of the degree will now be more important.

Table 5.10 - Context for US College Education Fees

5.10 - UK Independent School Fees

In the United Kingdom there are a range of schools that operate outside the state funded and administered sector. Generically these are termed independent schools, to show that they are independent of government funding and (to some degree) regulation. Despite many of the earliest foundations having been designed to provide free education for deserving scholars, the vast majority of pupils are now charged fees for attendance.

Historically there were two main types of school, preparatory schools, which educated pupils aged 7 – 13 and senior schools which provided education for 13 – 18 year olds. In addition there was a distinction between public senior schools, where admission was open to all irrespective of home location, religious denomination or parental occupation[72], and private schools where there might be some restrictions on admissions. Such distinctions are now generally moot with many schools offering education for 3 – 18 year olds.

[72] Although naturally, still subject to the ability to afford the fees.

The data concerning the fees charged by UK independent schools have the advantage that about 90 years of data is available covering the fees charged, with additional information on the sector being available for the most recent 45 years. Fee data has been captured for the years 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2010 and 2018 with a small number of data points having been discarded where errors have been identified. The general context for the data is outlined in Table 5.11, with more detailed discussion in Chapter 7.

Element	Description
Outcomes	The end products provides an education that is perceived to be high quality, together with other advantages to the participants.
Output Quality Measures	The results of public examinations provide a guide to the relative performance of pupils at independent schools. However these do not capture any of the additional benefits, nor do they make allowance for the varying ability levels of pupils.
Input Quality, Cost and Volume Measures	Technological and process innovations have reduced the resources required for administration and other support activities. There is, however, a general perception that more teachers and better facilities will always tend to produce a higher quality education.
Environment	Over the period considered there have been significant increases in the direct and indirect costs of staff. Additionally, there have, more recently been requirements for schools that are charities to demonstrate public benefit.

Table 5.11 - Context for UK Independent School Education Fees

5.11 - Formula 1 Team Costs

Formula 1 is a class of racing sanctioned by the Federation Internationale de l'Automobile (FIA) and is also a multi-billion dollar, world-wide industry. For various historical, economic and geographic reasons a significant number of Formula 1 teams are based in the UK and are registered as UK companies. To meet the requirements of UK law they are required to file annual accounts, most of which can be accessed on line, which report relevant business information such as turnover and employee numbers. This allows a view of the trends in different teams to be compiled that, in the best cases, run from the early 1980s to the present time.

The Formula 1 business is notable for the requirement to regularly produce small numbers of technology intensive cars, which have evolved to include very significant integration challenges. The scale of these technical challenges and the rapid rate of change tends to result in higher levels of human involvement than would be expected for a manufacturer of mass market vehicles.

Element	Description
Outcomes	The team and their drivers aim to perform competitively, ideally winning the championship.
Output Quality Measures	Qualifying times, race results, and championship points measure output quality, relative to other teams. Absolute comparisons between seasons are complicated by changes to the permitted car specifications.
Input Quality, Cost and Volume Measures	Good drivers, sufficient, skilled engineers and high quality facilities.
Environment	The general technologies underpinning Formula 1 are well understood by the teams. However, FIA keeps adjusting the allowed specification for cars, which means that different designs are required each season.

Table 5.12 - Context for Formula 1 Team Costs

5.12 - Defence System Costs

Defence Systems represents a potentially hugely diverse field. Historically it included arms and armour for soldiers, castles and other defences, sailing ships and dockyards. As technology has advanced the focus is now more on the technologically advanced systems, such as aircraft and warships that are used to deliver Defence outcomes.

As will be further explored in Chapter 7, Defence Systems have two roles, firstly they are designed to deter potential aggressors, thus avoiding the huge costs of going to war. Should deterrence not be effective, then the Defence Systems (in conjunction with the other elements of fighting power) must outperform those of the aggressors, in order to ensure that victory is achieved with minimum casualties to one's own forces.

As has been discussed in the earlier chapters of this work, advances in technologies, increases in system complexity and other factors appear to be driving up the relative costs of these systems. The design and development of such systems tends to be manpower intensive and not amenable to significant automation. If considerable numbers of a given system design are to be built, then effort can be made to make the manufacturing process more efficient, by substituting capital for labour.

Element	Description
Outcomes	The deployment of the end products, combined with personnel and infrastructure to deliver peace and security.
Output Quality Measures	The quality of systems can be measured either directly, with measures such speed and payload, or indirectly, with measures such as advantage over potential aggressor systems.
Input Quality, Cost and Volume Measures	It is generally assumed that the embodiment of new technologies, that require many

	technical staff and costly facilities to design, test and integrate, will represent an increase in input and hence output quality.
Environment	For most of the period since 1945 there have been general or specific threats that have necessitated the on-going development of Defence Systems in the UK. During this period the resources available to support such developments have varied according to the economic situation and government priorities.

Table 5.13 - Context for Defence System Costs

5.13 - Comparing the Data Sets

We now come to compare the data sets, in terms of the insights that they are likely to provide to the analysis of Defence Systems. Rather than seeking technical similarities or numerical similarities, we instead seek similarities in the processes and the effects that these produce (i.e. the context of production). Based on the discussion of different types of goods and services in Chapter 4 and also at the start of this Chapter we revisit Figure 4 and plot each set on the Productivity Growth vs Scarcity chart (see Figure 12).

This approach has been taken as the opportunity for productivity growth will have an influence on the rate of change of the cost of inputs to the processes[73] and the scarcity / congestion dimension informs the willingness and necessity of customers to pay the prices charged (Subtopics 2.3 and 3.2 in Table 2). On the output side, productivity growth is also likely to be relevant to the System Cost Change (Subtopic 3.1). Picking case studies that are similar to Defence in these areas are expected to provide relevant insights.

Although the relative positions of some of these items within a sector could be debated, it is believe that each is plotted in the correct sector, or on the correct border[74], as applicable. From this we find that the Defence Systems fall into the ‘Positional’ quadrant where low productivity growth is experienced and there is a high level of either intrinsic or congestion driven Scarcity. It therefore appears that detailed case studies about other data sets in this sector might produce useful insights. UK Independent School Education, Formula 1 and Defence Systems all viable, with US College Education being discarded as the available dataset only covers a single institution and the literature indicates difficulties in meaningfully expending the dataset.

[73] Depending on the good or service being produced these may be development and / or production processes.
[74] Stradivarius violins and Super Bowl tickets are intentionally placed on the border between the two Productivity Growth sectors, as each is strictly limited in number. Hence productivity is not a meaningful concept.

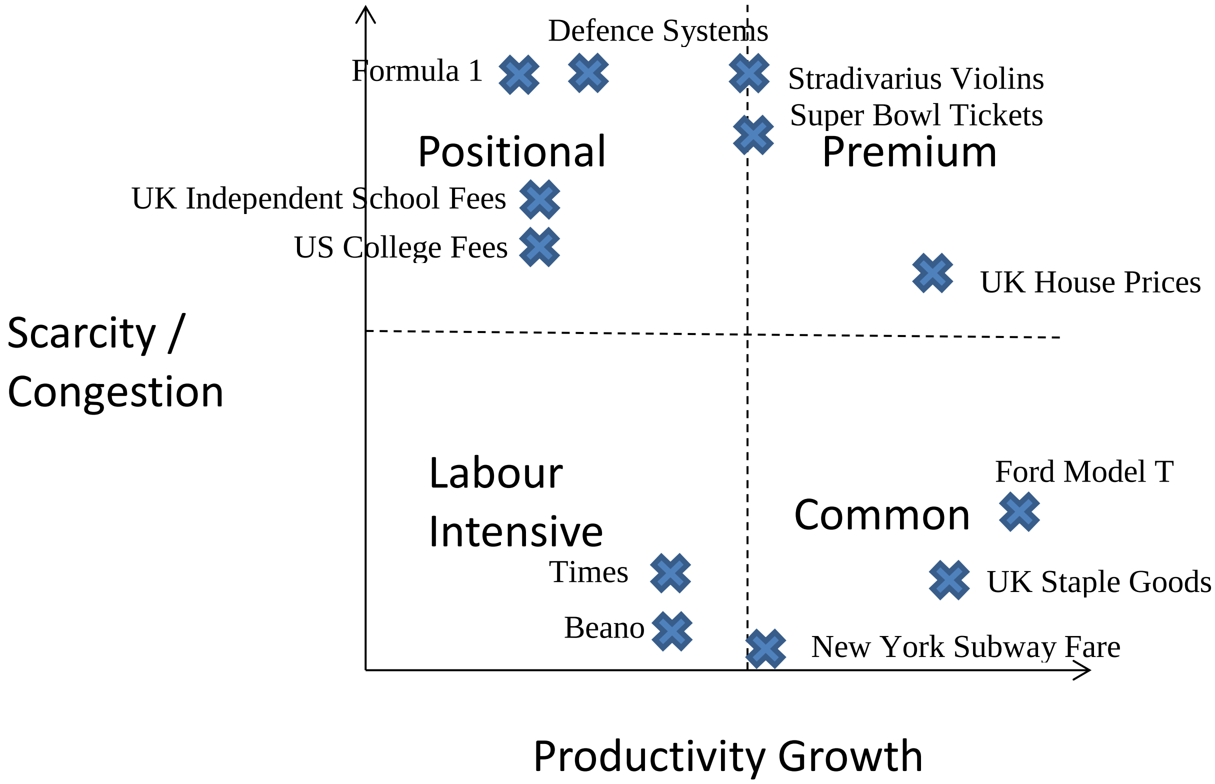

Figure 12 - Data Sets Plotted on Productivity Growth vs Scarcity Matrix

Additionally, as captured in the context table for each case study, it appears likely that the (often) unobservable mechanisms affecting procurement and project management decisions may well be connected with measurements or perceptions of the quality of the inputs and the outputs, as described above. Where the output quality or value of the good or service can be directly measured (and there is recent data available) then the stakeholders are likely to aim to maximise this value. However, where output quality measures are infrequently collected, or where their direct measurement is not tractable, then stakeholders will need to come to an understanding of what changes in input quality are likely to drive improvements in output quality.

In some cases, direct measurement of output quality can be achieved and this is simplest where there is direct competition, such as during Formula 1 races and also during wars. As we shall see in Section 7.4.2 direct measurement of the quality of education (in terms of the outcomes that parents value) does not seem a realistic prospect and so in that context it is expected that the focus will be on input quality measures, even though their connection to output quality may be tentative. A similar situation applies to Defence Systems in peacetime, when their aim is to deter potential adversaries, but as the deterrent effect cannot be directly measured, the focus is likely to be on the quality of inputs to the Systems (and the other components of Fighting Power). Similarities in the quality field between potential case studies and Defence Systems are likely to provide insights into the issues of comparing quality over time and willingness to pay (Subtopics 2.2 and 3.2 in Table 3.1).

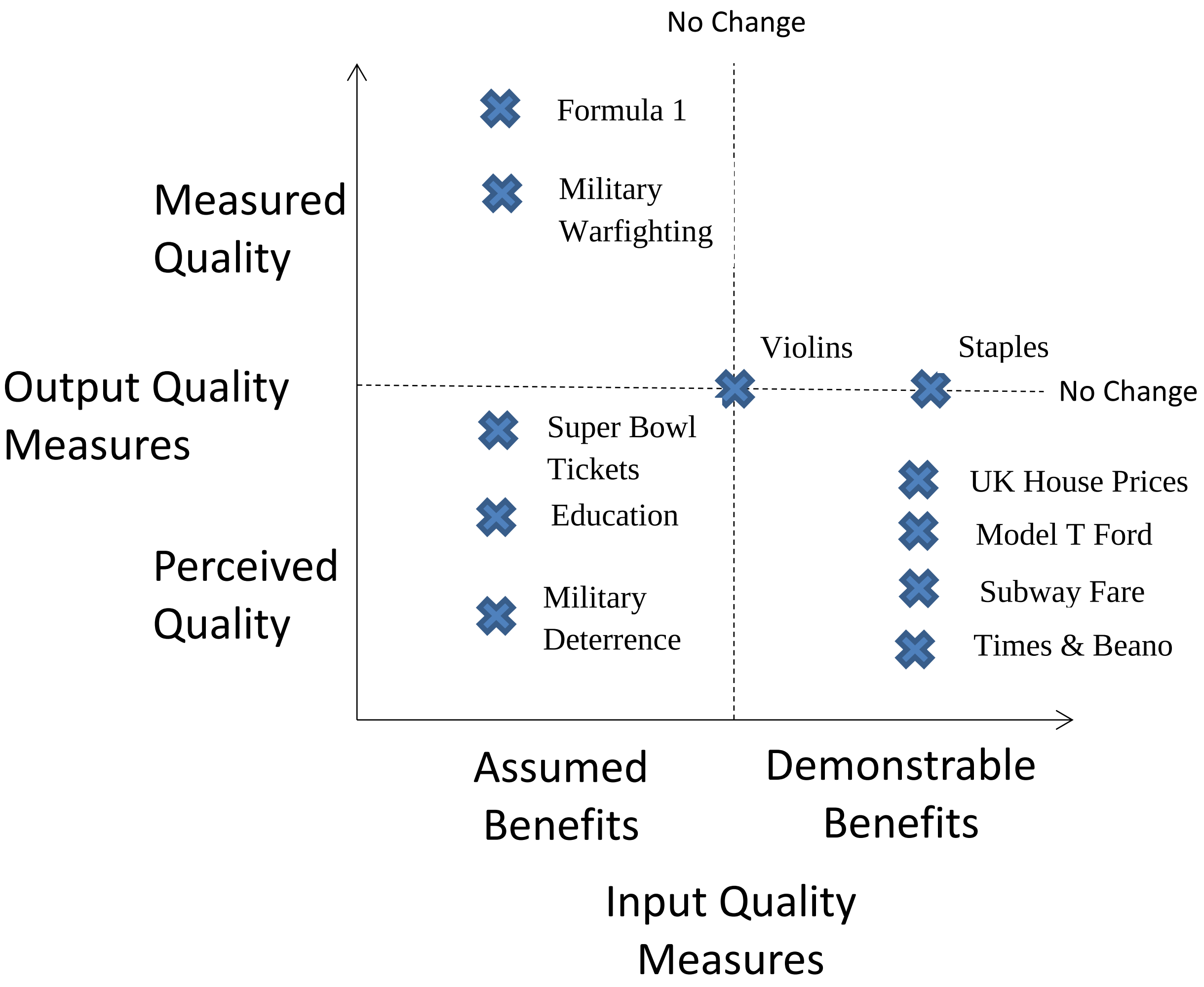

Figure 13 - Data Sets plotted on Quality Grid

In Figure 13 we note that the Defence, Formula 1 and Education (which includes UK Independent Schools and US Colleges) all contain significant areas where the benefit of changed inputs is assumed and the output quality measures cover both perceived quality and demonstrated quality. From this quality point of view we observe that Military Deterrence and Education fall into the same quadrant (Perceived Quality / Assumed Benefits) and that Military Warfighting and Formula 1 also fall into the same quadrant (Measured Quality / Assumed Benefits). This confirms that Formula 1 and Education also appear to be similar to Defence Systems from a quality viewpoint and hence should provide insights to the Defence Systems field.

As the review of the previous literature concerning the cost of Defence Systems has illustrated, there is a lack of agreement about which of the different ways in which historic costs can be uplifted to a common baseline is most appropriate, in a given set of circumstances. Before we engage on the case studies we shall thoroughly explore this topic, to ensure that the meaning of the different approaches is thoroughly understood, so that the best approaches can be chosen. This will provide useful insights for practitioners wrestling with this problem and also captures the current state of knowledge relevant to subtopic 2.1 in Table 3.1.

Chapter 6 - The Changing Value of Money

In considering how the cost and price of Defence Systems and other goods and services have varied in the past, we have observed that one of the key enablers is the ability to understand how the value of money has changed over time and hence how a historic cost or price may be compared with that of a current offering. Initially we shall examine how such changes affect individuals, before considering their role at the national level. This chapter is completed by an examination of the impact of different potential approaches on the values applied to historic expenditure.

6.1 - Bishop Fleetwood - Value of Money and the Individual

Initially it is useful to consider how this subject affects individuals and one of the earliest writers on this subject was Bishop William Fleetwood who published 'Chronicon Preciosum' anonymously in 1707, although a later edition (Fleetwood (1745)) is the point of reference for this research. The Bishop was posed a question by a correspondent, specifically (p.1): 'The statutes of a certain College (to the Observations of which everyone is sworn, when admitted Fellow) vacating a Fellowship, if the Fellow has an Estate in land of Inheritance or a perpetual Pension, of Five Pounds per Annum, I desire you would be pleased, to give me your Answer to the following Questions[77]; when I have told you, that the college was founded between the years 1440 and 1460…Whether He, who is actually possessed of an Estate, of Six Pounds per Annum, as Money and Things go now, may safely take that Oath, upon Presumption, that VI *l* [78] now is not worth what V *l*[79] was, when that Statute was made?'.

Or, to paraphrase, 'If I have an income of more than £5 the statutes require me to give up my Fellowship. I actually have an income of £6 per annum, but can I keep my Fellowship by arguing that my current income would have been worth less than £5 two hundred and sixty years ago, in the period 1440 – 1460?'

Fleetwood (1745, p.7) sets the foundation for his analysis by observing that 'A Pound (for Instance) will buy either more, or less Corn…now, than it would in H. VI Time[80]. A Pound is therefore of more or less Value now, than it was then; and the Value of a Pound is truly a Pound, and not its mere Name. It is not therefore the same Thing now, that it was in H VI Time'. In other words, in the context of this question, a pound should be defined in terms of the goods that it will purchase, rather than by its name. He justifies this by asserting that (p.9) 'The Founder intended the same Ease and Favour to Those who should live in his College 260 Years after his Decease, as to Those who lived in his own Time'. Hence 'They who lived in his Time, might, with V Pounds, purchase so much Bread, so much Drink, Meat, Cloth, Firing, Books and other Necessaries, or Conveniences: I know not exactly how much, nor is it Material: I only say, the Founder intended I might keep such an Estate, as would suffice to

[77] Three questions are actually asked, but the first two are not relevant to this current discussion.
[78] i.e. £6
[79] i.e. £5
[80] i.e. The time of King Henry VI, when the College in question was founded.

procure the same Bread, Drink, Meat, Cloth, Books etc., as the other might have procured for V Pounds, 260 Years ago'.

To support this argument he also cites the example (p.10) of 'indigent and virtuous scholars', who were given an allowance by some Founders, towards taking their degree. However, the cost of taking a degree has risen five times above what it was 260 years ago, so the allowances have become 'so very inconsiderable, that they signify little or nothing towards' the relevant costs but the same amount 'in those early Days would (with a little Help of Friends) have been sufficient to the intended purpose'. He then concludes 'this is a clear Proof, that Regard both may and must be had, to the different Value of Money at different Times' and that 'the Founder's visible Intention is better answered by such Regard, than it would be by a strict and obstinate Adherence to the bare Letter of the Statute'.

So we observe that Fleetwood structures the problem in the following terms:

1. The Founder intended to allow Fellows of the college a certain standard of living ('the same Ease and Favour').
2. One may characterise the standard of living in terms of a fixed basket of goods and services.
3. The current value of the original £5 should be set by examining the current cost of a representative £5 basket of goods in the Founder's time.

This problem formulation assumes that the basket of goods required to deliver a given standard of living does not vary over time (Fleetwood has it consisting of Bread, Drink, Meat, Cloth, Firing, Books and other Necessaries, or Conveniences). Whilst this may not have been an unreasonable working assumption for Fleetwood to make, as he was writing before the Industrial Revolution, it may not hold true in the 20th Century, as the goods available for consumers to purchase have changed and hence the contents of a representative basket of goods might also change.

6.2 - Price Indices

In order to quantify the type of change over time that Fleetwood describes price indices have been created. The guidance for UK politicians (House of Commons Library (2009), p.3) on how to adjust for inflation defines a price index as 'a series of numbers used to show general movement in the price of a single item, or a set of goods[81], over time'. It is further observed that in general 'any price index must consist of a set of prices for goods, and a set of corresponding weights assigned to each good in the index.

The same document (p.4) indicates that the UK Consumer Price Index (CPI) and the UK Retail Price Index (RPI) are both 'a composite measure of the price change of around 650[82] goods and services on which people typically spend their money' and that a perfect consumer price index 'would be calculated with reference to all consumer goods and services, and the prices measured in every outlet that supplies that supplies them'. As this would be somewhat

[81] The original document indicates that 'goods' should be taken to encompass both goods and services.
[82] Number covered in 2009, this number has since been increased.

challenging the 'CPI/RPI use a representative *sample* of goods and the price data collected for each is a *sample* of prices'. In 2017 (ONS (2017)) prices were collected for the 700 representative items whose 'movements are taken to represent the price changes for all goods and services covered by the index, including those for which prices are not specifically monitored.' Locally monitored prices are collected 'around the middle of each month' by price collectors, who note about 100,000 prices for about 520 items consisting of specified types of goods and service. The price collectors visit shops in around 150 locations throughout the UK, with most local shops being visited in person. Prices for the remaining goods and services are collected centrally by ONS.

Rippy's 2014 work on the history of the US Consumer Price Index illustrates that consumption patterns (and hence what might reasonably be expected to represent average consumer spending) have changed relatively rapidly in the 20th Century, unlike in the years preceding Fleetwood's time. For example (p.10) between 1919 and 1934 the spending on Fresh and Frozen fish significantly increased and so these items were added to the basket that the index was based upon, to ensure that it reflected consumer activity and similarly the invention of synthetic fabrics allowed the cost of clothing to fall, which required a reweighting of that element of the basket. External factors can also have an immediate influence on the index, for example (p.12) in the early months of 1942 'because the federal government issued rationing orders on new automobiles and new tires and tubes, the Cost of Living Division[83] dropped these two items entirely from the index'. They were replaced with used automobiles, retreaded tyres and an increased weight for 'streetcar and bus fares' (p.13).

Despite the challenges of collecting fully representative data, those compiling the US index appear to have made best efforts to represent the real cost of living[84]. In the UK this was not always true as (O'Donoghue et al 2004, p.39) the weights used for the Cost of Living Index 1914 – 1947 'were influenced by a highly subjective assessment of what constituted legitimate expenditure for a working class family; beer was completely excluded and the weight used for tobacco was much less than the actual proportion of expenditure on tobacco'.

In the UK (ONS 2011, p.2 - 4) the measurement of cost of living has been attempted since 1914 with an index of consumer prices which was replaced in 1947 by the RPI. In 1996 a Harmonised Index of Consumer Prices (HICP) was also introduced to meet European Union requirements. The RPI and HICP (or CPI) are similar, but have different exclusions and scope. The CPI is based on all spending by all institutional and private households and derives its spend data from the Household Final Monetary Consumption expenditure in the National Accounts, whereas the RPI excludes institutional households, the top 4% of households by income and pensioner households with three quarters of their income coming from the state pension and benefits, as these are judged not to be representative of general spending. The RPI's weighting data comes from a household survey known as The Living Costs and Food Survey (p.8).

[83] Of the US Bureau of Labor Statistics
[84] In terms of keeping the basket of goods current and representative.

Attempts to retrospectively construct equivalent measures to CPI and RPI (and in some cases compare them to wages) have been undertaken in a number of cases, see for example Phelps Brown and Hopkins (1955) and (1956), O'Donohue et al (2004) and Clark (2011), all of whom examined the UK, with the most ambitious being Clark, whose data stretch back to 1209. It is rare for researchers in this field to undertake primary research, but instead to refer to the limited number of authors that have consulted primary sources, as consulting primary sources can be rather time consuming. See for example Thorold Rogers' *Six Centuries of Work and Wages* (1890) and his work *A History of Agriculture and Prices in England (1259 – 1792)* which runs to seven volumes (see for example vol.6 1887) and (according to Phelps Brown and Hopkins (1955), p.195) 'drew chiefly on the college and estate accounts in the muniment rooms of some of the Oxford and Cambridge colleges, and the farm bailiffs' rolls and monastic accounts in the Public Record Office.' The main reason for the drive to reuse existing data, rather than gathering new primary data is articulated by Phelps Brown and Hopkins ((1955), p.195) who write that 'some apology is due for piecing and patching with secondary sources, when the primary materials are probably there for more detailed and solid work' but they conclude that 'the results which can be won from what is immediately accessible seem worth setting out for the sake of their grand perspective'. These authors (Phelps Brown and Hopkins (1956)) also cite Beveridge who (with others), produced a weighty tome on Prices and Wages in England from the Twelfth to the Nineteenth century (Beveridge (1939)[85]). Although Beveridge only produced a single volume (of a planned series) it is apparently an improvement on Thorold Rogers (p.xxi) 'Roger's work, while it has no rival, is both incomplete and imperfect…and the quality of the later volumes is inferior of that of the earlier ones. Yet even these earlier volumes are based in material much poorer than has now been discovered'.

Just as Phelps Brown and Hopkins reference those who have consulted primary sources, when writing about UK wages Clark (2011) references (for example) Levi (1885) who undertook detailed collection of primary data to support his work *Wages and Earnings of the Working Classes* and combined it with secondary census data to support his analysis.

In the UK, since 1963 the UK Office of National Statistics has produced regular data on wages and earnings, but all of the authors covering prices and wages before this have had to address uncertainties, make assumptions and interpolate between available data. This means that the perceived value and acceptance of results can change over time. For example Phelps Brown and Hopkins (1956, p.306) concluded somewhat tentatively that, based upon available data, the comparison of real wages against the price of consumables showed them in the period following 1510 'a Mathusian crisis, the effect of a rapid growth of population impinging on an insufficiently expansive economy; such as perhaps we see also in the fall that set in again around 1750, until this time a commercial and industrial revolution came to save Britain from the fate of Ireland'. However by 2007 Maddison (p.253) contends that the use of 'the real wage index to be a representative picture of living standards' is flawed and asserts that these 'results have now been almost universally rejected as a proxy for the

[85] The majority of the 1939 printing of this work were apparently destroyed by an air raid. The version consulted for the research was the 1965 reprinting.

movement of real GDP per capita' and hence standard of living. Other approaches, such as the examination of probate inventories indicate that with 'very few exceptions, each generation of descendants from the mid-seventeenth to the late eighteenth century left behind more and better possessions' and hence had a higher standard of living.

Difficulties can also exist with interpreting more recent data. After WW2 the US experienced what appeared to be (Rippy, (2014, p.21)) 'a slowdown in aggregate growth and an accelerating rate of increase in the CPI'. A committee led by George Stigler, reporting to Congress in 1961, observed (quoted by Rippy, (2014, p.21)) 'It is often stated that the Consumer Price Index measures the price change of a fixed standard of living based on a fixed market basket of goods and service. In a society where there are no new products, no changes in the quality of existing products, no changes in consumer taxes, and no changes in relative prices of goods and services, it is indeed true that the price of a fixed market basket of goods and services will reflect the cost of maintaining (for an individual household or an average family) a constant level of utility. But in the presence of the introduction of new products, and changes in product quality, consumer tastes, and relative prices, it is no longer true that the rigidly fixed market basket approach yields a realistic measure of how consumers are affected by prices. If consumers rearrange their budgets to avoid the purchase of those products whose prices have risen and simultaneously obtain access to equally desirable, new, low-prices products, it is quite possible that the cost of maintaining a fixed standard of living has fallen despite the fact that the price of a fixed basket has risen.

'These findings argue' (Rippy (2014, p.22)) that 'the CPI should measure the change in the cost of maintaining a constant utility (by adjusting consumers' preferences and spending behaviour) through time as relative prices change'. This recognises that the requirement for CPI is to measure the cost of generating a given outcome (i.e. a given level of utility) without specifying how consumers should meet this level of utility. This highlights one of the challenges in producing a Defence Price Index, in that whilst it is possible to measure changes in the costs of Defence Inputs, measuring the utility of the outputs is problematic.

We have observed that CPI and similar indices should measure the cost to consumers of maintaining a constant level of utility, which may (in some cases) be related to the cost of maintaining a constant standard of living. Whilst meeting this goal can be achieved relatively simply if the type and volume of goods purchased by consumers over time does not change (as might be assumed in the time of Fleetwood) steps must be taken, as mentioned by Rippy, to allow for changes in the type and relative balance of goods consumed.

In constructing a CPI or similar index there are a number of options about how one can deal with the situation when a product available in the first sample period (say January) is unavailable in the second sample period (say February) and is replaced with a product with different qualities, such as size, performance or other relevant characteristics. Such an occurrence is referred to as a quality change. Discussion of the possible treatments of this type of event appears in the current UKI CPI guidance (ONS (2017)) which notes that when particular products disappear from the market, 'care is taken to ensure that replacements are of broadly comparable quality so that price comparisons are not distorted'. How this is achieved varies, depending on the specific circumstances, and full details of quality adjustment methods are found in the *Consumer Price Indices Technical Manual* (ONS(2014)). There we find, in Section 8.2 (p.49), three quality adjustment methods outlined: direct comparison, direct quality adjustment and imputation.

The direct comparison is a simple process, but it can only be applied when there is a new product that is directly comparable meaning that 'it is so similar to the old one that it can be assumed to have the same base price', such as 'a garment identical except that it is a different colour'. Thus one may obtain 'a replacement which may be treated as essentially identical', and hence allows one to assume that 'any difference in price level between the new and the old product is entirely due to price change and not quality differences'.

Where the replacement product is (ONS (2014a, p.49) 'of a different quality or specification' then direct quality adjustment is the preferred method. In this case an 'attempt is made to place a value on the quality, or specification difference' and then the price of the new good is adjusted to account for the difference. The adjustments of this type were applied in October 1995 in the UK when *The Units of Measurement Regulations 1995*[86] required that pre-packaged goods be sold in metric rather than imperial quantities. In this case 'in each outlet the nearest equivalent new size of the product priced in that outlet was found, and an adjustment made to the base price pro rata for the change in weight'. This approach is similar to Pugh's specific cost concept (Pugh (1986, p.108)) where he asserts that one can track the cost per ton of warships over time and hence make a direct quality adjustment to reflect changes in warship size, independently of other factors.

In more complex cases, where a good is described by multiple qualities / specifications an adjustment for quality can only be made where it is possible to value the changes in specification separately. This can be accomplished using Hedonic Regression (ONS (2014a, p.50)) which is 'a technique that uses a set of ordinary least squares regressions to relate the price of an item to its measurable characteristics'. A common application of this and the illustrative example cited by ONS is for personal computers (PCs). The 'hedonic regressions are calculated on the basis of a single month's data, using unweighted regressions based on price data collected from retailers' web sites. This can then be used to calculate the effects of quality changes between an old PC and a new PC replacing it in a retailer's range of products. The regression of the last month in which the old PC was available is used to calculate the expected price of the old PC and also what the expected price of the new PC would have been, had it been available that month. The ratio of these two expected prices represents, according to the regression calculations, the effect of the quality changes. This factor is then used to adjust the observed price of the new PC downwards to remove the effect of the quality change. The remaining difference in the two prices observed for the old PC and the new PC is due to price change, rather than quality changes[87].

[86] Available at www.legislation.gov.uk/uksi/1995/1804/made, accessed 23 March 2018.

Despite the relative simplicity of the approach there are a number of necessary assumptions, which are outlined by Gilbert (2013). Quoting original work by Rosen he outlines (p.4) a primary assumption, namely that a stable price function exists, so that the price of a good is completely determined by the characteristics. This also implies other assumptions which include:

> Completeness – All possible products within the product space are available for sale.
>
> Availability – At any given time within a single market, all products are available to all consumers – that is, that we are looking at a single unified market
>
> Market Power – No consumer of producer has market power – that is all participants in the market are price takers.

Whilst these assumptions may well be true for consumer PCs, it is by no means certain that the physical characteristics of a Defence System determine its price, as we have previously discussed in critiquing overly positivist viewpoints. Additionally the single unified market assumption and the completeness assumption also appear to be questionable in such cases.

The final approach to quality adjustments is impution (ONS (2014a p.52) which is applied if 'the replacement product is of a different quality or specification, and no information is available to quantify the difference'. In this case the base price for the item is imputed assuming that the price change between the base and now is the same as for the elementary aggregate for that item[88]. This means that the price of the new item is adjusted to ensure that its introduction does not change the value of the index. Whilst this may be an acceptable approach where prices for many similar items are being collected, such an assumption would be an extremely bold step where there are very few items being considered.

6.3 - Gross Domestic Product and Price Changes

Whilst the CPI and related series attempt to examine the impact of price changes on individuals, we must now take a more national-level view to provide the context for Defence spending. Although, in ancient Athens, for example (Peck, (2001, Section 3.1)) some private individuals took responsibility for building and manning warships on behalf of the state and in mediaeval England individual landowners would build castles, the high and apparently increasing cost of such items means that the state has generally assumed responsibility for national defence. It is therefore now necessary to consider how the changing value of money impacts on national governments and their spending and how this may be measured.

Gross Domestic Product (GDP) is one of the main measures of economic activity of a nation and has three economically equivalent definitions (ONS (2014b, p.10)). An even fuller set of GDP definitions and guidance on how it should be calculated is provided in the System of National Accounts (2008)[89]. This sets down international standards for the preparation and presentation of national accounts and is co-authored by the United Nations, the World Bank, the European Union, the International Monetary Fund and the Organisation for Economic

[88] That is, its price has behaved in an identical fashion as those of similar items.
[89] This is the latest version of a document originally published in 1953.

Co-operation and Development. Its aim is to provide a standard approach to the preparation of national accounts, specifically 'how to compile measures of economic activity in accordance with strict accounting conventions based on economic principles' (United Nations (2009 p.1)). The same document (p.34) states that 'GDP derives from the concept of value added. Gross value added is the difference between output and intermediate consumption. GDP is the sum of gross value added of all resident producer units plus that part (possibly the total) of taxes on products, less subsidies on products, that is not included in the valuation of output' Additionally 'GDP is also equal to the sum of the final uses of goods and services (all uses except intermediate consumption) measured at purchasers' prices, less the value of imports of goods and services' and also 'GDP is also equal to the sum of primary incomes distributed by resident producer units'. These are known, respectively, as the production approach, the expenditure approach and the income approach (UK Treasury (2014, Section 2.2)). The GDP can (Section 2.3) 'be expressed in terms of either current or constant prices'. A current price GDP value shows the 'value of transactions in the prices relating to the period being measured', whilst constant price 'figures express value using the average prices of a selected year', which is known as the base year. Therefore over a given period the change in GDP at constant prices will measure 'how the quantity or volume of goods has changed' whereas the current price values will include prices changes as well as volume changes. 'The ratio of the current and constant price series is therefore a measure of price movements, and this forms the basis for the GDP deflator'. Thus the GDP deflator is the equivalent of a price index for the whole economy.

As the ONS states (2007, p.71), in 'order to reflect changes in real values of inputs and outputs, measures of productivity should take quality changes in both into account. This is usually achieved by ensuring that the price indices used for deflation are adjusted for these quality changes'. We have already seen how quality adjustments are made in CPI and a similar approach is used for other indices.

Similarly the ONS (2007, p.74) reports that when measuring GDP 'volume is regarded as having the dimensions of quantity and quality'. This means that 'for growth in volumes to be measured correctly, then changes to both these dimensions must be taken into account.' For the National Accounts, which cover the whole economy, 'volumes must be aggregated and therefore expressed in a common metric; and since this metric cannot be tonnes, litres or another physical measure the metric used is economic value in the prices of a price-base period'.

When it becomes necessary to estimate these volumes there are two possible approaches (p.74):

1. 'Extrapolation or quantity revaluation – where the number of units in the current period is multiplied by the unit price in the base period'.
2. 'Deflation – nominal values in the current period are deflated to constant prices using price indices'.

Clearly the first approach is more straightforward method, but can only be used where the characteristics of the product being considered do not change over time. Thus is may be applied to homogenous products such as tons of grain or Kilowatt hours of electricity produced, but is not suitable for other, heterogeneous, products, whose volumes must be derived through deflation.

6.4 - Challenges in Constructing Price Indices

Thus we observe that price indices are important both in attempting to evaluate the impact of price changes on the standard of living of an individual and also in terms of separating the effects of economic growth on GDP from the effects of price increases. The work of Nordhaus (1996) suggests that there may be significant issues in constructing price indices because during 'periods of major technological change, the construction of accurate price indexes that capture the impact of new technologies on living standards is beyond the practical capability of official statistical agencies' (p.29). The main problem arises because 'most of the goods we consume today were not produced a century ago' and thus one has to attempt to 'compare the services of horse with automobile, of Pony Express with facsimile machine, of carbon paper with photocopier' etc. He acknowledges that undertaking 'a complete reckoning of the impact of new and improved consumer goods on living standards is an epic task' therefore his study is confined to 'exploring the potential bias in estimating prices and output in a single area – lighting'.

Nordhaus then constructs a data series covering the input costs and efficiency of different lighting technologies from an ancient wood fire through various lamps to modern light bulbs based both on previous authors and also some personal experimentation. He then (p.39) moves onto some theory related insights, such as the observation that 'traditional price indexes measure the prices of goods that consumers buy rather than the prices of the services that consumers enjoy' however, to measure the true cost of living 'we clearly should focus in the outputs rather than on the inputs'. He then considers (p.41) how technology change can cause traditional price indices to be in error and find three sources, incorrect weights, improvements in efficiency and incorrect linking of new goods.

He is relatively unconcerned about the first problem as the expenditure weights can be directly observed (at least in theory) and are not affected by the use of traditional (input based) rather than true (output based) prices. In the second case, if the efficiency of the production function[90] improves over time 'this will lead to a decline in the service-good price ratio…which will be entirely missed by traditional price indexes'.

The third problem applies when new goods are introduced if 'new goods are introduced for which the service-good price ratio is lower [for the new good] at the time that the new good is introduced'[91]. Typically (p.41) 'the statistician will simply assume that the products deliver the same quantity of service characteristics per dollar of spending and hence the price index

[90] i.e. The efficiency with which the inputs are transformed to useful outputs.
[91] i.e. The new good is more efficient at transforming the inputs to useful outputs.

will be accurate if the shadow price of the service characteristic for the new good is the same as that for the old at the date when the new good is introduced into the price index.'

As an illustration of the impact of these problems Nordhaus then examines the example of the price of lighting between 1883 and 1993. In 1883 Edison introduced his first electric light and (p.43) 'priced it at an equivalent price to gas light'; at this point (Nordhaus reports) the cost of kerosene and gas lighting was also similar, which provides a convenient base year from which to track changes in the cost of lighting. In the intervening 110 years the cost of gas and kerosene have risen by a factor of 10 (in nominal terms) the cost of gas light has risen by a factor of 3 whilst the cost of electric light has fallen by a factor of nearly 100.

Later (p.51) he calculates that 'the traditional price of light has risen by a factor of between three and five in nominal terms since 1800. This is not bad compared to all consumer prices …which have risen tenfold over the same period' (based on the traditional calculation methods). However, the traditional price of light has risen 'by a factor of between nine hundred and sixteen hundred relative to the true price'. This is mainly because (p.55) by design 'price indexes can capture the small, run-of-the-mill changes in economic activity, but revolutionary jumps in technology are simply ignored by the indexes'.

In order to assess the potential impact of technology improvements not being correctly captured Nordhaus (p.58) takes the then current consumption bundle and divides it into three categories:

> Firstly sectors that have experienced 'run of the mill' changes, where changes in technology have been 'relatively small and price indexes are likely to miss relatively little of the quality change or the impact of new goods'. Examples include 'home consumption of food (e.g. potatoes), most clothing (e.g. cotton shirts), personal care (e.g. haircuts), furniture' etc.
>
> Secondly 'seismically active sectors', which have experienced 'both major changes in the quality of goods and the provision of new goods, but where the good or service is still recognisably similar to its counterpart at the beginning of the 19th century'. Examples include housing, watches and private education.
>
> Thirdly 'tectonic shift' sectors, where 'the entire nature of the production process has changed radically' and hence where the changes are so vast that 'the price indexes do not attempt to capture the qualitative changes'. Examples include medical care, telecommunications, transport and electronic goods.

Nordhaus calculates that (about the time of writing) respectively these sectors represented 28%, 36% and 37% of consumption[92]. He calculates that the change in the cost of lighting has been overstated in price indices by 3.7% per annum and uses this as a top estimate of the overstatement in the 'tectonic shift' sector and then has other factors proportionally lower. Based upon this value he estimates that whilst the generally accepted factor by which real US

[92] It is assumed that the total of 101% is a result of rounding effects.

wages have increased since 1800 is 13, the actual results may be between 40 and 190. (This also means that price increases measured by the GDP deflator and elsewhere may have been overstated by a similar proportion).

Nordhaus valued lighting in terms of the brightness and duration of the outputs as these are well defined measures that can be compared over time. For some Defence Systems similar measures that are meaningful and consistent over time may be constructed (e.g. the speed and payload of transport aircraft), however, for other Systems (e.g. Combat Aircraft) outputs can only be measured in terms of a performance advantage against a likely adversary, across a Defence campaign. This is particularly problematic as such direct measurements of performance only occur very infrequently.

As has been previously observed price indices such as the GDP deflator are commonly used as methods for determining the current value of historic Defence Systems and other goods, services and projects. Having examined how such indices are constructed, and some of the problems that may be inherent in them, it is now worth considering a range of approaches for making comparisons between current and historic costs.

6.5 - Measuring Worth

In a search for a well attested approach to determining the optimum methods for valuing the worth of historic costs, the first port of call for many researchers is the website www.measuringworth.com which is curated by Lawrence Officer and Samuel Williamson. The data and tools that it provides sprang from their 2006 work 'Better Measurements of Worth'. In that paper they wonder 'How can we determine the worth of something, especially its historical worth?' observing that 'the price, even deflated for inflation, is not enough' to address questions such as 'Was Carnegie richer than Gates? Did Babe Ruth make more than Barry Bonds? Was the cost of a loaf of bread more then than now?'

The authors open their discussion by examining whether one can actually measure 'worth' in a generalised way and observe (p.87) that the usual approach is to 'use the language of "opportunity cost"' and say that 'the worth of something is "the cost of the most valuable forgone alternative"', which means that 'the worth of something can only be measured in terms of something else'. This raises the issue that 'the most valuable alternative may be different for different things for one individual and is most likely different for one thing among different people'. They therefore aim to 'state the cost of the thing and let everyone "compute" the most valuable alternative that this money would buy for him or her'.

The paper presents different indices to measure worth, with the context for the question determining which is likely to be most representative (p.88) as 'measuring the worth of a loaf of bread differs substantially from measuring the worth of the interstate highway system' or indeed of a military aircraft. The authors identify three different classes of item for which historic worth may be calculated, commodities, incomes and projects. Projects are 'an investment, such as construction of a canal or installation of a cable network, or a government expenditure, such as the financing of Medicare or a war' and so will be of most interest to this research.

Officer and Williamson's initial (2006) approach has been slightly expanded and clarified in their later work 'Measures of Worth' (2012), where they identify four ways of examining the cost of a project over time. The first of these is Labor[93] Value, which values the historic cost of the project in terms of the number of labor hours its cost represents at the relevant historic rates. The later value is then generated by valuing the labor hours at the rates applicable in the later time-frame. The second approach is the historic opportunity cost, which values the project in terms of the volume of goods in the economy that it related to when it was undertaken and then uplifts the costs to the of those goods to the later-time frame using the GDP deflator. The third measure is the economy cost, which examines the historic cost in terms of the % of GDP that it represented then and then applies that to the relevant GDP of the later year. The fourth measure is the contemporary opportunity cost, which measures the historic cost in terms of how many consumer bundles it would purchase and then applies this number to the later-year cost of consumer bundles.

Of these potential approaches it is worth noting that the third measure is free of the difficulties outlined above in calculating accurate price indices, as may be the first option and fourth options, provided that nominal, rather than 'real' labor rates and bundle costs are used. The second approach requires a GDP deflator to be calculated, which as we have seen may not correctly capture the impact of technology changes.

Officer and Williamson illustrate these calculations by uplifting the historic cost of the Empire State Building from 1931 to 2010 economic conditions, which gives the following results:

'The Empire State building, a giant of a structure in its day, was built at a cost of $40,948,900. This may seem inexpensive in today's terms when we compare its cost using the GDP deflator and determine a contemporary cost of $507million.

Alternatively, the cost in terms of the goods and services the average household implicitly gave up would be about $1.25 billion in today's money, and the "labor" value of the building was $2.15 billion in today's production worker wages.

Finally, if you want a current-dollar indicator of how important the building was compared to other projects in New York City when the Empire State Building was completed, then a number close to $8.1 billion is the best number.'

It is notable that the highest and lowest of these results vary by a factor of 14, which is consistent with Nordhaus' estimate for the difference between indicated and actual growth in real wages. Returning to the example of HMS Havock and the Type 21 frigate from Chapter 2, if the cost of HMS Havock was uplifted to 1972 values, in line with the fraction of GDP that it consumed, then the uplifted cost would be £1.62M, compared with the £29M for a Type 21. Although this value is four times higher than the value calculated using the RPI, it is still significantly less costly[94].

---

[93] The original spelling of the US based authors is preserved here.
[94] But arguably also significantly less capable.

Having identified a number of different approaches to allow the comparison of costs / prices from different points in history and noting the variation in the uplifted costs of the Empire State Building, it is now useful to consider a specific example of the impact that varying these approaches would have on apparent trends over time. Given the overall topic of this research, using the aircraft time data from Augustine's Law seems an excellent candidate.

Unfortunately it is unclear what data sources Augustine based his chart upon, and the data values used have not been published. It has therefore been necessary to extract data from the published version of his chart. This was achieved by digitising the image and then loading it into the Microsoft Paint software package, which allows the cursor positon to be recorded. The image (from first point to last) was 542 pixels in the date direction and 431 pixels in the cost direction. In some cases there was uncertainty about which pixel was the best fit to a point, but it is believed that no point will be out by more than 2 pixels in any direction. This gives a potential error of 0.5% on cost and an error of +/- 6 months on date.

Of the possible approaches, uplifting the costs by comparing them with the value of the consumer bundles option was excluded, as this is more relevant to individuals or groups of individuals and in this case the costs under examination were incurred by a nation. The three remaining approaches yield the following results yield the following results when used to uplift the costs to 1982 (the date of Augustine's final point):

Name	Base Year[95]	Cost	1982 labour value (Skilled Labour Rates).	1982 historic opportunity cost (GDP Deflator)	1982 economy cost (% of GDP)
Wright Model A	1911 Q2	$3,450	$209,000	$30,500	$333,000
JN-4A	1917 Q2	$6,047	$229,000	$35,500	$336,000
Thomas Morse	1917 Q3	$7,481	$283,000	$43,600	$415,000
Spad	1918 Q3	$9,620	$290,000	$48,200	$420,000
DH-4	1918 Q4	$4,983	$150,000	$25,000	$218,000
Standard E-1	1919 Q2	$6,661	$167,000	$32,600	$282,000
P-39	1939 Q3	$100,000	$1,690,000	$641,000	$3,580,000
P-51	1940 Q4	$73,182	$1,180,000	$464,000	$2,380,000
P-38	1940 Q4	$110,530	$1,780,000	$701,000	$3,590,000
P-40	1941 Q2	$52,522	$767,000	$312,000	$1,360,000
P-61	1942 Q1	$189,785	$2,380,000	$1,040,000	$3,820,000
F-4U	1942 Q1	$127,159	$1,600,000	$699,000	$2,560,000
P-63	1942 Q3	$59,045	$741,000	$325,000	$1,190,000
P-47	1942 Q4	$108,338	$1,360,000	$596,000	$2,180,000

[95] The data from 'Measuring Worth' does not provide granularity beyond a given year. The information about which Quarter best represents an aircraft's Base Year has therefore been ignored for purposes of uplift to 1982 economic conditions.

F8F-1	1943 Q2	$129,731	$1,430,000	$683,000	$2,140,000
F6F-5	1944 Q1	$106,191	$1,090,000	$546,000	$1,580,000
P-59	1944.Q2	$201,534	$2,070,000	$1,040,000	$3,000,000
P-80	1944 Q3	$77,594	$798,000	$399,000	$1,160,000
P-84	1945 Q3	$65,097	$663,000	$326,000	$954,000
P-82	1946 Q1	$158,489	$1,510,000	$703,000	$2,330,000
F8F-2	1947 Q3	$175,178	$1,460,000	$701,000	$2,340,000
AD	1949 Q1	$222,754	$1,650,000	$845,000	$2,730,000
F4U-5	1949 Q2	$288,981	$2,140,000	$1,100,000	$3,540,000
F9F-6	1950 Q2	$431,303	$3,010,000	$1,620,000	$4,810,000
F-86	1950 Q2	$360,182	$2,510,000	$1,350,000	$4,010,000
F-84	1950 Q2	$516,468	$3,600,000	$1,940,000	$5,750,000
F-100	1952 Q4	$656,737	$3,880,000	$2,260,000	$5,970,000
FJ-4	1955 Q2	$835,101	$4,400,000	$2,770,000	$6,550,000
A1E	1955 Q2	$414,373	$2,180,000	$1,370,000	$3,250,000
F-102	1956 Q1	$1,172,102	$5,860,000	$3,750,000	$8,710,000
F-101	1957 Q3	$1,487,352	$7,170,000	$4,610,000	$10,500,000
A-4D	1958 Q1	$1,020,048	$4,610,000	$3,090,000	$7,080,000
F-106	1960 Q1	$725,888	$3,090,000	$2,140,000	$4,470,000
F-104	1960 Q4	$1,743,329	$7,410,000	$5,140,000	$10,700,000
F-105	1960 Q4	$2,492,021	$10,600,000	$7,350,000	$15,300,000
F-8E	1961 Q1	$1,000,000	$4,150,000	$2,920,000	$5,940,000
F-5	1964 Q4	$786,416	$2,940,000	$2,210,000	$3,840,000
F-4B	1965 Q3	$2,592,944	$9,330,000	$7,140,000	$11,700,000
A-7A	1966 Q3	$1,610,262	$5,540,000	$4,320,000	$6,610,000
A-6	1966 Q4	$3,856,620	$13,300,000	$10,300,000	$15,800,000
F-111A	1968 Q4	$5,851,152	$18,000,000	$14,600,000	$20,800,000
A-7E	1969 Q2	$1,850,297	$5,370,000	$4,410,000	$6,070,000
F-4J	1970 Q2	$3,423,598	$9,410,000	$7,740,000	$10,600,000
F-111D	1970 Q4	$6,591,216	$18,100,000	$14,900,000	£20,500,000
F-14	1971 Q3	$20,433,597	$51,800,000	$44,000,000	$58,500,000
AV-8A	1974 Q1	$4,344,412	$8,620,000	$7,800,000	$9,380,000
F-15	1975 Q2	$17,090,652	$30,700,000	$28,100,000	$33,800,000
A-10	1979 Q2	$5,851,152	$7,580,000	$7,410,000	$7,440,000
F-16	1980 Q1	$8,877,197	$10,500,000	$10,300,000	$10,400,000
F-18	1982 Q2	$10,200,482	$10,200,482	$10,200,482	$10,200,482

Table 6.1 - Augustine Data uplifted to 1982 economic conditions

Notes: Values calculated by using tools at www.measuringworth.com

From this table we note that each method of treating the historical costs gives relatively similar results for the most recent decade (e.g. the lowest AV-8A value is 83% of the highest value). However, when we examine the data from 1911, the lowest value is less than 10% of the maximum value. The lower value, which is $30,500 is generated by uplifting the historic cost by the GDP deflator, which simply revalues the historic cost according to the general inflation in the economy. On the other hand, the higher value of £333,000 is generated by determining the percentage of GDP of 1911 that the historic cost represented and then

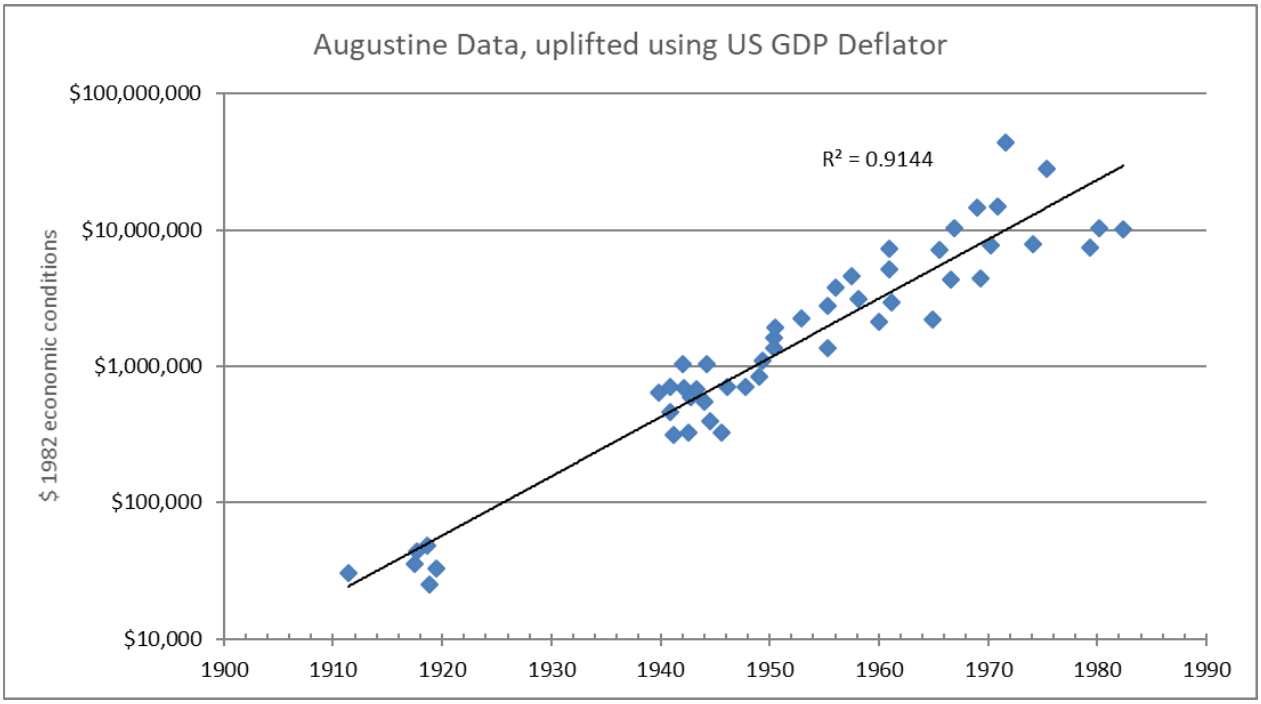

Figure 14 - Augustine Data uplifted using US GDP Deflator

Notes: Values from Table 6.1. Uplifted using US GDP Data from www.measuringworth.com

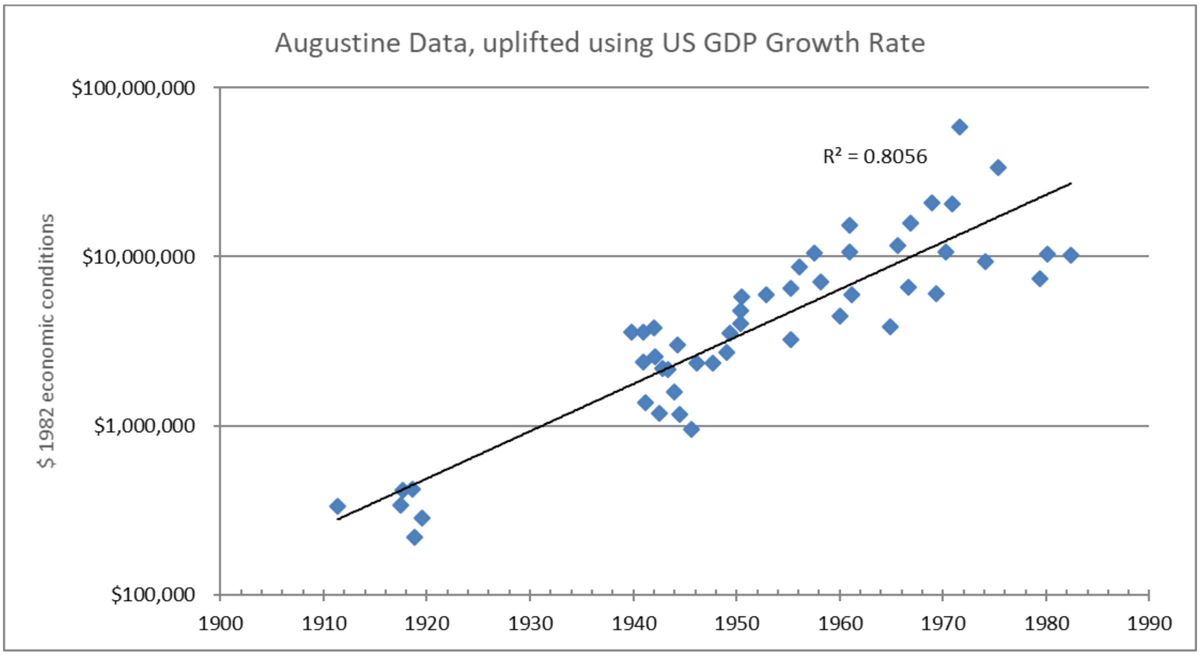

Figure 15 - Augustine Data uplifted using US GDP Growth Rate

Notes: Values from Table 6.1. Uplifted using US GDP Data from www.measuringworth.com applying this same percentage to the GDP of 1982. The difference between the results of these two approaches is caused by the growth in GDP which in nominal terms has grown from $34Bn in 1911 to $3.3 Trillion in 1982.

As one may observe from Figure 14 and Figure 15 these two different approaches generate significantly different predicted growth rates and impacts, with the GDP Deflator approach predicting that the $100,000,000 aeroplane would be reached in about 1996 and the $1Bn aeroplane being expected in about 2020. On the other hand the percentage of GDP values predicts the $100,000,000 aeroplane in about 2001, but does not expect aircraft to cost $1Bn until about 2036. Thus, whilst it is perfectly feasible and indeed likely that the cost of past aircraft will inform the cost of future similar projects, it is essential that the correct method of uplifting the historic costs to a constant set of economic conditions be applied. Different approaches can vary the perceived cost of historic aircraft by more than an order of magnitude, which in turn influences the prediction of future costs and hence a detailed comparison of the competing approaches is required to ensure that the most appropriate is selected.

6.6 - Preferred Price Uplift Approach

Considering the four potential approaches identified by Officer and Williamson, we observe that using the GDP Deflator attempts to show the current value of the goods that were historically foregone to fund the investment. This gives a worth that is lower than the other approaches, part of this may be because economies expand over time, making the historic goods foregone a less significant part of the economy and part of this effect may be because the GDP Deflator over estimates the increase in prices over time (see Nordhaus). Even if it were possible to calculate an accurate GDP Deflator, it is not clear how relevant this measure would be as it simply answers the question: *If the price of the historic item had increased at the same rate as an average good in the economy what would it cost now?* In Nordhaus' study of lighting we have seen that the cost of gas and kerosene (and the cost of light produced by these means) has increased over time, but that the cost of electric light has significantly decreased. It is therefore unclear what, if anything, applying the average price increase in the economy should represent, other than some arbitrary scaling factor. Clearly if we had an expectation about how the price of a given good should behave over time and it failed to follow this path, this would be of interest, but there appears to be no prima face evidence to suggest that the price of Defence Systems should follow the average price rises in the economy.

Although the comparison of labour hours does not suffer from the same 'price index' issues as the GDP deflator does, it does have some weaknesses. It might be valid to undertake such a comparison (and expect it to be meaningful) if the production methods have not changed significantly, however, most industries are constantly striving to devise production methods that are cheaper, often through requiring less labour. So, if one were to build an Empire State Building today it would, no doubt, take less labour effort than the original did (for example through increased mechanisation and prefabrication) which would make a meaningful comparison difficult.

Making a comparison with the value of household bundles foregone in each time period is meaningful and has no obvious structural weaknesses, as it uses then-year values, rather than relying on a price index. The calculation shows how many households the government could have supported, had it not undertaken the project, and then uses this value to calculate a current value, in terms of household bundles. Although this knowledge might be useful in making social comparisons, its value to determining the changing worth of Defence Systems is not clear.

The final option, measuring investment as a percentage of GDP appears much more promising. As the calculation is made in the nominal values of relevant years there are no issues with using price indices to uplift prices from one year to another. Also this measure is meaningful for government purchases of all kinds, (including Defence Systems), as it represents the proportion of the national income in that year that the country committed to delivering the subject project. This is, at a national level, a measure of affordability and translates meaningfully across time. This approach would give an appropriate measure of the cost for most government projects, even if they were not positional or tournament goods.

Chapter 7 - Case Studies

7.1 - Introduction to Case Studies

The preceding chapters have explored the literature relating to the cost of Defence Systems and also that relating to the changing value of money over time. They also initially examined a number of data sets that might form the basis for detailed case studies and selected the three that are further developed in this chapter. This background sets the context within which we examine the case studies which cover the observed increases in the cost and price of:

UK Independent School education

Running a UK based Formula 1 racing team

Designing and building UK Defence Systems

and also consider the structures and mechanisms that help define the context that produces these observed effects.

Education and Formula 1 operate on annual cycles, with the opportunity to make adjustments in-year, so management decisions can have a relatively immediate impact. In Defence it currently takes at least a decade to design and build a major new system such as a ship or an aircraft and even designing, testing and implementing minor changes across an equipment fleet can take several years. Additionally, once delivered, many Defence Systems are retained in service for 20 – 40 years, so the consequences of poor decisions may be persistent. The long term impact of decision making in Defence procurement therefore should provide an incentive to make good initial decisions, however, given the challenge in making decisions to plan for an uncertain future, there may be an incentive to make what would be regarded as 'safe' decisions, rather than risking an innovative approach failing to deliver.

These decisions include, for example, the balance between system procurement cost and its military effectiveness, the balance between support solution flexibility and cost and most importantly the balance between project technical risk and potential military effectiveness. Providing the evidence required to support such decision making requires, inter alia, the ability to understand and forecast (with appropriate accuracy) the design, development, manufacturing and in-service costs and timescales of different candidate options. Before considering the specific case studies we shall examine the background to cost forecasting.

7.2 - Forecasting Challenges

Forecasting the cost and / or duration of a planned project is usually undertaken by using some combination of bottom up and top down approaches. In the bottom up approach the cost and / or duration of each individual element of the project will be forecast and then aggregated up to form an overall estimate. In the top down approach the project is considered as a whole and forecasts made at that level. Naturally, in the initial stages a top down approach is likely to be used, but as details about candidate solutions emerge, then a greater emphasis on bottom up forecasts can be expected.

The work of Daniel Kahneman and Amos Tversky relating to heuristics and biases (see Tversky and Kahnmen (1974) for example) shines an important light on how humans tend to make decisions. One specific element of their very broad work relevant to forecasting is the planning fallacy (Kahneman, (2011, p.245 – 254)) which shows that people tend to underestimate costs and risks and overestimate likely benefits. Kahneman's personal example of this relates to a team developing new curriculum elements (and supporting materials) for the Israeli Ministry of Education. The team initially estimated that it would take between 1.5 and 2.5 years to complete the work, however when they gathered information on the performance of past teams undertaking similar work, they discovered that 40% of the teams never completed their tasks and those than did took 7 to 10 years. The performance of the past teams may be considered a 'reference class' whose performance was (probably) relevant to Kahneman's team.

For activities with a short decision cycle (such as Independent Education and Formula 1) it may be very reasonable to expect the forecast for the expenditure next year to use the expenditure in the current year as a baseline, with adjustments being made to account for the expected future situation. For Defence Systems, where time horizons are much longer and the technological and environmental changes from the last project may be significant, then the question of which past projects is it appropriate and useful to regard as being 'similar' does not have a clear answer. Indeed, this conundrum is a statement of the Reference Class Problem that was fist articulated by Venn (1888, p.224 - 226) when he considered the probability of a fifty year old man living to be sixty one. Starting from the general question he gradually adds further specifics, culminating in a fifty year old man named John Smith, who has consumption and lives in the North. Venn then observes (p.226) that it 'is obvious that every individual thing or event has an indefinite number of properties or attributes observable in it, and might therefore be considered as belonging to an indefinite number of different classes of things'. In defining a useful reference class, the challenge is that the 'class must be broad enough to be statistically meaningful, but narrow enough to be truly comparable with the specific project.' Flyvbjerg (2006).

7.3 - Defence System Forecasting

We shall now consider the two main types of top-down forecasting approach and how the reference class is currently employed in them. The first type is Historic Trend Analysis where the aim is to directly forecast the cost of the project, together with other relevant features such as duration, based on the trends in the reference class. As Shermon (2011) discusses, there are at least three different approaches that are applied to deliver this outcome, each of which makes explicit or implicit assumptions about the relevant content of and applicability of the reference class.

The first of these is the Complexity vs Time approach which was developed by Darryl Webb (see Webb (1990)), which uses the proprietary parametric cost forecasting software PRICE H to infer the 'Manufacturing Complexity' for past production projects, where the production costs and other relevant project parameters are known. This ‘Complexity’ is a representation of the ‘technology and productivity’ difficulty in manufacturing a part or system and so is some undefined mix of technology, manufacturing difficulty and contractor productivity. It is not clear whether Webb regards this as the sole effect in operation, or an amalgamation of all the effects. Webb (p.67) does give specific examples of changes in complexity, such as ‘the manufacturing cost complexity values for electronics that are in common use today…are representative of dual-in-line flatpack, and leadless packaging methods and therefore do not represent the higher cost, lower physical volume, and higher cost trends representative of future packaging concepts’ and also that other changes in ‘cost and technical complexity over time’ are due to ‘the technology transition from steel to aluminium, to titanium and then to composites’. Whilst these technology changes are likely to influence product cost, the lack of a meaningful definition for ‘Complexity’ is problematic.

The likely Complexity of a new item to be manufactured can only be estimated by comparison with similar items, whose Complexity is known (having been reverse-engineered from known production costs), or inferred from long-term trends in complexity. This approach is satisfactory (and widely used) for estimating the cost of new parts in a manufacturing environment with low rates of innovation and it is also likely to be satisfactory for many areas of Formula 1, where, at the whole system level, only the rate of innovation over the forthcoming year needs to be estimated, which is likely to be relatively predictable. However, for Defence Systems, the need to rely on long-term trends in complexity for forecasting the cost of novel systems is deeply unsatisfying. Directly forecasting what trends in complexity might happen over a 20 year programme would be infeasible and the alternative of assuming that observed historic trends are inevitable without satisfactorily explaining what might be driving them and hence how they might be avoided is distinctly unappealing.

Webb (1990, p.73 - 75) shows that for 61 fighter and attack aircraft from the Fokker E.1 of 1915 to his projection for the Eurofighter of 1999, the complexity of almost all aircraft conform very closely to the complexity trend line (Webb (1990, Figure 4)). Webb does recognise that such a trend requires explanation and offers the following mechanism: ‘all systems depend on the development of subsidiary technologies and must wait until new advancements in materials, processes, and components allow increased performance and cost complexity at the system level. In the reverse, obsolete technologies’ i.e. those with low complexity ‘are not desirable as a new fighter or attack aircraft must be competitive in performance or it is nor produced’. Which is equivalent to saying ‘the customer wants fighter and attack aircraft to be as good as they can be (as they compete directly with the enemy) and there is an average rate at which technology, design and production develops’. This is, essentially, a rehearsal of Augustine’s views, but with the introduction of Complexity as a positivist explanation for the relentless cost growth.

Although Shermon (2011, p.56) does not favour the second approach, which focusses on the changes over time in production cost per unit weight, its simplicity has much to recommend it. Pugh (1986, 2007, etc) was particularly interested in warships and aircraft and these have the advantage that (for a given class of design) accepted design practice and the laws of physics will tend to result in similar proportions of the total weight being dedicated to systems (e.g. engines) no matter what the exact size of the platform. This approach appears to provide a potentially useful approach to constructing a reference class for low innovation Defence System production costs.

Shermon does favour the third approach, which is multi-variate analysis over time, in which stepwise regression is undertaken to generate a line of arbitrarily good fit to any given data set. Although Shermon (p.56 – 57) explains the process he is not explicit about how the variables to be regressed should be picked. He does observe that they need to be independent, but uses the phrase ‘experience has shown that…’ in justifying a variable choice. This leaves the potential for the problems identified in Chapter 2 with Hove and Lillekvelland’s approach, where variables and relationships may be identified, such as the weight and cost of different generations of armoured vehicles increasing over time, and the assumption made that the correlations are causative, whereas both might be driven by (say) an unobserved desire to increase crew protection and platform performance, which resulted in progressively more complex and better armoured platforms being developed. However, delivering the required protection and performance in a lighter-weight vehicle would tend to increase production costs above those for a vehicle of the expected weight, rather than the achieved lower weight corresponding to a reduced cost. Additionally, as arbitrarily many variables may be regressed until a desired goodness of fit has been achieved there is also a risk of over-fitting the model. This positivist approach that simply seeks patterns in the observed data is intellectually unsatisfactory where extrapolation is required, as it implicitly assumes that all mechanisms not directly captured by the selected variables are unimportant. Being able to hypothesise which mechanisms are producing effects, and hence what might happen in the future would be a far more desirable result.

The second forecasting type is Reference Class Forecasting which was devised by and is commercially promoted by Professor Bent Flyjberg and is powered by the Oxford Global Projects Database, which contains huge volumes of data, mainly about infrastructure projects. It includes, for example 2,522 projects on rail, road, bridge and tunnel new build projects (Department for Transport (2020, p.3)). The aim of this approach is not to actually forecast the likely cost of the project, but to assess the likely difference between the estimate at different stages of the project and the eventual cost and benefit delivery. Whilst this is clearly of value to a governance process, if it can be made to work, it relies on the assertion that estimates in a given reference class will always be inaccurate, will always be similarly inaccurate and that better data cannot ever resolve this problem. Whilst there is good evidence from the Planning Fallacy to illustrate that human nature is always to be over optimistic, it is counter intuitive to insist that one cannot learn from experience. Additionally, it appears that as projects become more complex there ought to be more opportunity for estimates to diverge from eventual performance. Thus, as the literature suggests that it provides reliable insights, we assume that either the Reference Class only includes projects of similar levels of complexity and management competence, operating in similar environments, or that estimates are consistently accurate, but management reporting consistently lower values, in the interests of entryism or other organisational ‘game playing’.

These constraints of this approach are summarised by the Department of Transport (2020, p.6) who cite Awojobi and Jenkins (2016), Batselier and Vanhoucke, (2016) and Walczak and Majchrzak, (2018) to conclude that 'The effectiveness of RCF depends on the similarity of the reference class. If the project fits well into the reference class, the resulting uplift from the RCF will provide a more reliable estimate of the cost of the project. Moreover, the effectiveness of RCF is influenced by the size of the projects and the size of the reference class; projects need to be sufficiently large and the reference class should include enough projects. Only if these criteria, similarity, project size, reference class size, are met will RCF outperform other methods.'

The updated Oxford Global Projects Database contains data from 1851 – 2019 (the 1851 building is presumably the Crystal Palace). The Department of Transport document investigates the segregation of projects by geographical location and other factors, but implicitly assumes that all projects in a given field (e.g. roads) are similar and comparable. As the projects covered are all civil engineering, a field not generally noted for rapid technological innovation, this may well be a fair assumption.

The large number of civil engineering projects worldwide allow them to conclude that in 'practical terms, any data is better than no data and a reference class comprising 20-30 past, similar projects is robust to derive meaningful insights. Moreover, as with the RCF analysis below, once data are pooled, they can be analysed to statistically test for similarities between subtypes of projects in the reference class or other characteristics, e.g. size, cost, timelines, location which might show statistically significantly different risk profiles.' We shall later examine what size of Reference Class it might be realistic to construct for a Defence System project. Although this technique would be compatible with a critical realist viewpoint, with the cost and schedule overruns being the observable effects, it is currently applied with a positivist epistemology, with the goal being to use inductive reasoning to infer the general behaviour of similar projects, based on the observed performance of those in the reference class. Despite the value that it might provide to the practice of cost forecasting, there is no structured attempt to investigate the mechanisms that are likely to be driving the observed effects.

In each of three case study areas, the costs and / or prices of the activities producing the relevant outputs, and hence delivering the outcomes for stakeholders, exhibit growth rates above the average for the economy and in each case this is causing affordability concerns to those involved in that field. For example, the comment from Jencks and Riesman (1968) 'if we extrapolated current [cost] trends sufficiently far into the future, the entire GNP would be devoted to higher education' is analogous to Augustine's Law covering the increasing cost of military aircraft, which was discussed in Chapter 1. However education and Defence are both vital national issues and Barack Obama's comment on education '*if you think education is expensive, wait until you see how much ignorance costs in the 21st century'* could equally well be applied to Defence.

To provide a consistent structure, for each case study we consider, as far as possible, we build upon the categories identified in Table 5.1. For convenience this Table is reproduced below as Table 7.1.

Element	Description
Outcomes	The end results of the activity and how the relevant stakeholders value them
Output Quality Measures	Ways in which the quality of the direct outputs of the activity can be (usefully) measured
Input Quality, Cost and Volume Measures	Absolute or relative changes in the input factors to the activity
Environment	The general structure of the environment and its rate and degree of change

Table 7.1 - Case Study Elements

Within this framework we consider how changes in the environment directly and indirectly drive changes in the cost and volume of inputs to the activity. These changes may also affect the quality of the input factors and hence may, or may not, alter the output quality and the outcomes. It is expected that these elements will provide insights into some of the underlying mechanisms that are likely to be driving some of the observed effects.

In each case we consider which of the available historic data points are comparable (i.e. part of the same Reference Class), as they represent 'similar' enterprises, activities and / or outputs and whether we can predict this similarity in advance, or only retrospectively. To address this in a structured manner we build upon literature about the cost of Education to come up with three specific requirements for data points to be (potentially) members of a consistent reference class.

In each study the aim is to understand the main mechanisms driving increases in costs and prices and hence to ascertain whether these are inevitable or whether steps could be taken to mitigate their scale and impact.

7.4 - UK Independent School Fees

7.4.1 - Introduction

This section assumes that, in general, the outcome that parents (and others responsible for funding children's schooling) desire is an education where their investment is perceived to provide good value for money. As we cannot directly observe this calculation, we start by exploring the factors that parents appear to value in an independent school education and review how the resources that parents may be able to dedicate to education have changed over time.

Having considered the two main competing theories about what might be driving growth in the cost and price of education we then examine how far the available evidence suggests that each affects the level of UK Independent School fees. Prompted both by concerns in the literature and evidence from the historic data, the historic outcomes of different categories of fees and of school are examined, to determine which may be regarded as being similar and hence can be combined into a coherent subset.

For relevant subsets of the data, we then examine exogenous and endogenous changes in the costs and volumes of inputs to the education process and consider how far the mechanisms proposed by the two main theories explain the observed changes.

7.4.2 - Valuing Education

Historically little academic thought was given to the value of education, the 'wealthy were assumed to consume more non-compulsory education than the less well off, just as they consumed more of other goods. Education was also classified as a status good, consumed by the middle and upper classes to signal higher social standing'. (Machin and Vignoles (2005, p.4)). Since the development of the human-capital theory economists suggest that individuals 'invest in human capital, such as schooling, because human capital makes a person more productive and this gain in productivity is reflected in higher wages. Thus it is argued that individuals primarily make investments in schooling and other forms of human capital to earn a return' (p.4).

However, any assumption that human capital can entirely be captured by academic achievement is problematic. For example in the introduction to 'School choice in an established market' (Gorard (2019)) the author refers (p.5) to personal experience of working in a fee-paying school in Wales. 'The school was in a poor state of repair with falling pupils rolls and increasing debt. Although it supposedly selected students at intake by academic ability, participated in the Assisted Places Scheme, was a member of the Headmasters' Conference and had become one of the most expensive schools on Wales with a very low pupil to teacher ratio, nevertheless the public examination results were poor – worse than the national average for all schools at A level in fact'. Given that the direct output quality measure (A level results) was poor, one must assume that the parents and the Local Education Authority saw value in the education provided that was not captured by the examination outcomes, but which made it worth paying for.

In his survey of the state of the art of research into the Economics of Education Burgess (2016) identifies that a major advance (p.19) ‘in the economics of education in recent years is the recognition that there are multiple valuable skills’ when previously work tended to focus ‘on a single measure of ability, implicitly intellectual ability, often discussed as IQ’. Considering a range of authors he identifies terms such as ‘grit’ and ‘conscientiousness’ as being valued, with various authors finding correlations between these personal effectiveness traits (however they are named) and labour market outcomes. This matches the conclusions of Green et al (2017) where (p.37) they find ‘that simple surveys, while they commonly show the importance of academic motives…do not reveal all parents' motives. Preferences for an exclusive peer group, and a focus on confidence and aspiration building, are revealed only through in-depth interviewing techniques.’

We may suspect that these expectations are not new. When the Romans established schools in Britain (Lawson and Silver (1973, p.1)) in A.D. 78 these were designed ‘to romanise the sons of native chieftains’. Latin was the only written language and its knowledge ‘would be confined to relatively small groups: tribal notables, officials, some craftsmen and traders in the towns, a few wealthy villa owners in the countryside, and, from the third century, the leaders of scattered, mainly urban, Christian communities’. The vast majority of the population remained ‘Celtic and illiterate’. Although not investigated until relatively recently, it appears likely that choices between competing education options have always made by considering the whole breadth of the outcomes, the skills, experience, knowledge and other advantages that different options would impart. Although the weight given to different aspects of this decision will vary depending on the environment, one can imagine an aspirational Celtic parent valuing the opportunity for their child to receive a high quality education, in a stone built Roman school and to mix with the ‘right sort of people’ just as a current parent might look at independent schools.

7.4.3 - Parental Resources

Despite the perceived benefits that an independent school education is perceived to impart, we assume that for most parents (and other responsible adults) who are likely to send children to an independent school, the available resources will be finite and hence changes in their disposable income may affect the level of school fees that they can afford (or are willing) to pay. As the vast majority of funding for UK independent schools comes from the tuition fees, it is likely that a reduction in incomes, will reduce the number of pupils and hence the income available to independent schools.

Green et al (2017, p.37) identify that (at the time of their study) nearly half of children at independent schools come from families in the top 10% of the income distribution. Therefore, Figure 16, taken from Jenkins et al (2016a), which shows growth in the 90th centile of income may be significant in examining the changes in available resources.

We note that between 1980 and 1990 the mean income and the 90th centile incomes both rose by about two-thirds, in real terms. However, given that the 90th centile income was (obviously) much higher in 1980 than the mean, the increase in cash terms will have been significantly more for the 90th centile. This presumably resulted in a significant increase in the disposable income available to those interested in funding independent school fees. Although the same rate of growth is not maintained, there are fairly consistent increases through to 2008. However, after this point incomes fall significantly, presumably as a result of the global recession. Having considered the expectations and resources of parents (and others with similar responsibilities) and identified the lack of comprehensive direct output quality measures, we now move on to examine the changes that are evident in the costs and volumes of the inputs to the education process.

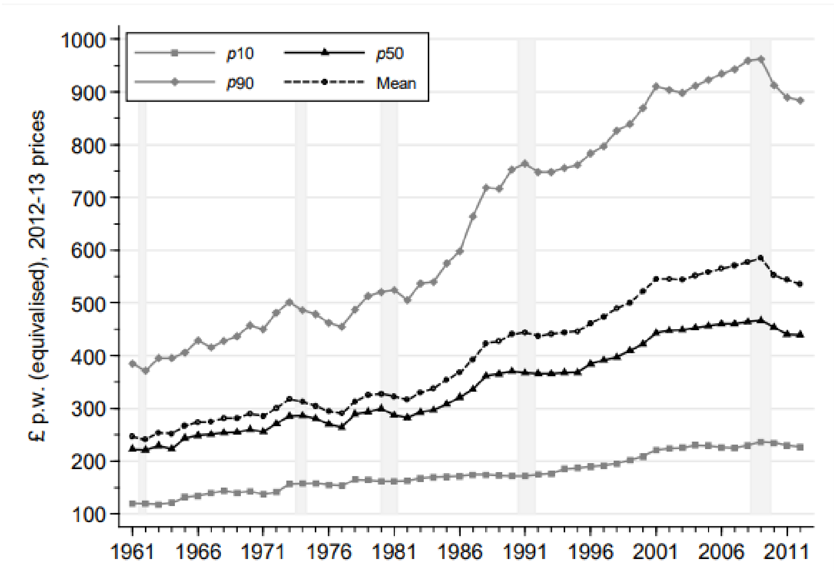

Figure 16 - UK Changes in Real Income Levels 1961 – 2012

Notes: Source – Reproduced from Jenkins et al (2016a)

7.4.4 - Educational Inputs Factors

Literature examining the steadily increasing cost and price of education is mainly rooted in work relating to US college education. This field is naturally of interest to academics who work in such institutions and hence as Kimball and Luke (2018, p.30) observe, quoting Howard Bowen, ‘scholars have long studied “the seemingly inexorable tendency for institutional cost per student…to rise faster than costs in general over the long term”’.

There is similar interest in the steadily increasing fees that UK independent schools charge and the ability of parents to afford them. ‘The fee rises, in spite of the surge they have triggered in both parents working full-time, are driving many… families out of the private school market’ (Dunnett (2018)). However, there appears to have been minimal systematic study in this area.

It is assumed that, as that vast majority of the schools covered by the data are not run for profit, the fees charged are set at the level necessary to cover running costs, to make investments for the future and to provide a prudent level of reserves, hence that profiteering is not a significant driver for fee increases.

7.4.5 - Cost Disease and Revenue Theory of Cost

Broad discussion of why the cost and price of goods and services provided by some parts of the economy have consistently risen by an above-average rate has been occurring for at least five decades. In their examination of the economics of the performing arts Baumol and Bowen (1966) discuss (p.162) the US economy's productivity gains which they class as 'truly remarkable'. For most of the twentieth century 'output per man-hour (the amount of goods and services yielded by one hour of labor) has gone up at a steady rate of about 2½ per cent per year compounded'. These 'continuous and cumulative' improvements mean that 'output per man-hour has doubled approximately every 29 years'.

The factors they identify as being responsible for this increase include 'new technology, an increasing capital stock, a better educated labor force [and] economies of large-scale production', many of which are of very limited relevance to live artistic performances. If the audience are content to have a televised performance then gains are to be found as an orchestral performance on television takes (p.163) 'less than twice the man-hours of a live performance' and can then reach 'an audience of 20 million instead of the 2,500 persons who occupy a concert hall, thus yielding an increase in productivity of four hundred thousand percent'.

They identify that, even in the performing arts, there had been some relevant innovations such as where 'air conditioning has made year-round operation possible for many groups that formerly had an enforced summer hiatus', improvements in air travel have 'speeded tours and decreased the cost of travel', administrative activities have 'benefitted materially from the availability of new types of office equipment' and whilst there have also been (p.164) 'improvements in lighting facilities, in the methods used to shift scenery and in a few other peripheral areas' they observe that 'the basic character of performance itself has stayed much the same. The playing of an instrument or the acting of a role remains today largely what it has been for centuries'. 'It still requires as many minutes for Richard II to tell his "sad stories of the death of kings" as it did on the stage of the Globe Theatre' and similarly no one has yet worked out how to decrease 'the human effort expended at a live performance of a 45 minute Schubert quartet much below a total of three man-hours'.

To illustrate the impact of these differences they imagine (p.167 – 168) an economy 'divided into two sectors, one in which productivity is rising' (say producing automobiles) and another in which it is constant (say the performance of Haydn trios). They then imagine that in the automobile production 'output per man-hour is increasing at an annual rate of 4 per cent, while the productivity of the trio players remains unchanged year after year'. They further imagine that 'the workers in the automobile industry recognize the changes which are taking place and persuade management to agree to a matching rise in wages'. Therefore each year the auto worker's wages 'goes up by 4 percent, but his labor output increases by the same percentage' hence 'labor costs per unit …remain absolutely unchanged' and this process can 'continue indefinitely in our imaginary world, with auto workers earning more each year' but with 'costs per car remaining stationary'.

So, in this situation of economic growth, what are the prospects for the string trio? There are two extremes to the potential outcomes (p.168) it might be that the performer's wage 'remains constant year after year, so that none of the economy's prosperity rubs off on him. His ability to buy goods and services does not increase at all'. Which results in him becoming increasingly impoverished, which will in turn 'presumably discourage some people from becoming musicians and will encourage movement into the automobile industry'. If on the other hand the string players succeed in getting their wages raised in line with those of the auto workers, then (p.169) if 'in a forty hour week the string player provides just as many performances as he did the previous year, but his wage is 4 per cent higher, the cost per performance must have risen correspondingly. Thus so 'long as the musicians are successful in resisting erosion of their relative incomes, the cost per performance must continue to increase along with the performer's income' hence 'rising costs will beset the performing arts with absolute inevitability'.

Just as the playing of an instrument and the acting of a role today retain remarkable similarities with performances in years gone by, so the teaching of a school class today would be very recognisable to teachers from 50 or 100 years ago. Recent years have seen a proliferation of technology in the classroom and similar innovations have improved the speed and accuracy of many administrative tasks, but the essential 'performance' of a teacher effectively imparting of knowledge to a class has certain immutable qualities. Thus, the economic challenges relating to the lack of productivity improvement opportunities articulated above (sometimes referred to as the Cost Disease) could explain some or all of the increases in education costs and hence prices.

In considering the performance of Haydn trios, without discussing the quality of the performance, Baumol and Bowen are (presumably) implying that there is no quality change, each performance is identical and interchangeable with every other performance. If a cost saving were desired, amateur musicians could be engaged at a lower rate, but having less time to practice (and perhaps less ability) the performance is likely to be of lower quality.

The main alternative to the Cost Disease theory is Bowen's[96] Revenue Theory of Cost (as quoted by Kimball (2014, p.889)), which is specific to education and which has five rules:

1. 'The dominant goals of institutions are educational excellence, prestige and influence…
2. In quest of excellence, prestige, and influence, there is virtually no limit to the amount of money an institution could spend for seemingly fruitful educational ends…
3. Each institution raises all the money it can. No college or university ever admits to having enough money…
4. Each institution spends all it raises…
5. The cumulative effect of the preceding four laws is toward ever increasing expenditure. The incentives inherent in the goals of excellence, prestige, and influence

[96] This is Howard R Bowen, not William G Bowen who worked with Baumol in the 1960s.

are not countered within the higher education system by incentives leading to parsimony or efficiency.'

The result of Bowen's theory is that colleges / universities (and possibly independent schools) would naturally end up engaged in an 'arms race' that consumes ever increasing resources, but arguably contributes more to the status of individual managers rather than educational excellence (i.e. that there is no demonstrable increase in output quality and hence no demonstrable improvement in the outcomes). The disconnection between increasing spend and the outcomes achieved is similar to the argument advanced by Kaldor (1982) in criticising the 'Baroque Arsenals' of the military as being decadent.

As Kimball and Luke (2018) articulate, there is no agreement about whether one or both of the Cost Disease and the Revenue Theory of Cost consistently apply to US college education (and related fields). They then proceed (p.36 – p.43) to examine the challenges that previous authors have found in validating either theory and in particular issues that exist within the datasets that have been compiled for this purpose. Of particular relevance are the following challenges:

a. Difficulties in disentangling Education costs from those for accommodation, food services, bookshops, hospitals and auxiliary enterprises that US colleges tend to engage in.
b. The wide variety of internal and external factors that could drive cost changes and the differences in how these may influence different public and private colleges.
c. Challenges in measuring and valuing changes in the quality of education.

It seems eminently reasonable that if the Cost Disease and the Revenue Theory of Cost could apply to US college education, then they might also apply to independent school fees. Independent schools are likely to prove a simpler proposition to analyse and hence might provide insights for colleges and other more complex sectors.

Bearing in mind Kimball and Luke's challenges, data to investigate these propositions would need to be:

Challenge Number	Requirement
1	Relatively specific to the educational activities of independent schools
2	Drawn from institutions that are relatively homogenous (and hence may be expected to have common reactions to internal and external influencing factors)
3	Address the articulation and measurement of quality changes.

Table 7.2 - Kimball and Luke's Challenges

We have two main, public domain data sets that should allow us to test how far observable increases in the cost and volume of inputs and possibly hence changes in the quality of education (as investment in improved pupil to teacher ratios and facilities are supposed to deliver) might be reasonable explanations of what is driving some or all of the observed fee increases.

7.4.6 - Whitaker's Almanack Data Set

The first collection of publicly available data that we shall consider was published between 1929 and 2018 and gives a broad overview of the sector during this period. Whitaker's Almanack[97] reported the year founded, number of pupils, annual fees and the current head teacher for a number of the UK's better known independent schools. For example, an entry from 1930 reads:

Name of School	F'ded	No. of Boys	Annual Fees[98] (Boarding)	Annual Fees (Day)	Headmaster
Eton College, Windsor	1440	1125	£230		Rev C.A.Allington

Table 7.3 - Extract from Whitaker's Almanack 1930

This[99] data has been captured for the years 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2010 and 2018 and may be found at https://doi.org/10.15129/6f1f8d5e-ec5a-4dd1-8e4a-edc06be39a83. A small number of data points have been discarded where errors have been identified. Whitaker's did not report this data in 2002 and when it was reinstated in the 2003 edition termly fees were reported, rather than annual fees. These later data points were converted to annual fees by multiplying by three (as there were three terms in the academic year). Additionally, so that dissimilar fees were not accidentally grouped together, each data point has been identified according to which of the following types of school and fee it fell into:

School and Fee Categories	Fee Identifier
Boarding fee at a school that only accepts boarding pupils	Only Boarding
Boarding fee at a school that accepts both boarding and day pupils	Dual (Boarding Element)
Day fee at a school that accepts both boarding and day pupils	Dual (Day Element)
Day fee at a school that only accepts day pupils	Only Day

Table 7.4 - School and Fee Categories and Fee Identifiers

This categorisation was based on whether annual fee values are reported for boarding and / or day pupils. Where doubt existed historical and news sources were consulted to confirm that

[97] The data for this analysis was transcribed from hard copy versions of Whitaker's Almanack at Bristol Central Library. See Whitaker's Almanack (1980)
[98] In some cases Whitaker's quotes a range of fees as being applicable. In such instances the mid-point of the range has been taken for this analysis.
[99] The data on schools belonging to the Society of Headmasters of Independent Schools which is reported in 1970 and 1980 was not collected as these schools fall outside the scope of reporting in other years, which is based on the members of the Headmasters' Conference and the Girls' Schools Association.

transition from one school category to another had occurred, rather than the presence or absence of a data entry simply representing a typographical or administrative error. Similar sources were consulted to confirm the closure dates of schools that are no longer extant.

This dataset covers 830 named schools, however 118 of these have to be excluded. The reasons for these exclusions and number of schools so affected are shown in Table 7.5.

Exclusion Reason	Number Affected
Duplicate[100] (amalgamated into single entry)	12
All fee data missing	6
History of school unclear	13
Republic of Ireland School	4
UK State School (never independent)	77
Special Type (fees not comparable)	6

Table 7.5 - Reasons and Numbers of Whitaker's schools excluded from final data set.

For each remaining school, from examining the available data and the history of the school, a further categorisation was undertaken, as shown in Table 7.6.

School History Category	Number
Open throughout period, full data series	135
Open throughout period, data incomplete	257
Founded during period	66
Closed during period	145
Chose to become state controlled	121

Table 7.6 - School History Categorisation of Whitaker's data

Notes: Twelve schools appear in both the 'Founded during period' and 'Closed during period' categories. Thus the total of categories in Table 7.6 is 724, rather than the expected 712.

Before exploring these series further, we shall attempt to remove any biases that might be introduced by non-homogeneity in the schools covered by the data. Initially we return to the first criterion raised above (see Table 7.2) and consider how far these schools focus on their roles as UK educational institutions and how far they are involved in auxiliary enterprises. We do this by testing the homogeneity of the sector in terms of the proportion of their annual income generated by non-educational activities (mainly endowment income[102], non-educational enterprises and overseas franchising arrangements). The majority of non-educational enterprises tend to be making use of the school's estate (e.g. sports facilities and accommodation) in the evenings / weekends / school holidays.

So far no relevant time series has been identified that addresses this, so the annual accounts filed with the UK Charity Commission[103] for the financial year ending in 2019[104] (i.e. before the impact of COVID-19) have been consulted for the 108 English and Welsh[105] schools that

[100] Where, for example, a name change has occurred.
[102] Including that from all sorts of investments.
[103] https://www.gov.uk/government/organisations/charity-commission
[104] See Eton College (2019) for example.
[105] Charities registered in Scotland have a separate regulator.

appear regularly in the Whitaker's data[106] and which also allow this school specific information to be gleaned from the accounts[107].

Figure 17 shows that the majority of these schools generated less than 10% of their annual turnover from activities other than running independent schools in the UK. The outliers that exceed this value tend either to be long established, with generous endowments (e.g. Eton College) or to be particularly active in overseas activities (e.g. Dulwich College). The maximum value of 25% (for Eton College) is low when compared to some US colleges where, for example, at the University of Pennsylvania[108] (in 2018 – 19) 63% of the turnover came from running healthcare businesses and a further 6% from investments.

Although this snapshot provides no specific evidence about the composition of school turnover in earlier years, it appears likely that some spread would still have existed. For example, in the 19th century Eton College enjoyed significant income from investments, from ownership of land and from church livings (see Royal Commission on The Revenues and Management of Certain Colleges and Schools (1864) (Appendix p.94)) whilst other institutions with a shorter pedigree and less illustrious benefactors would have been less well resourced. The opportunities for generating turnover from non-educational activities are likely to have been more limited earlier in the 20th century, which suggests that non-educational income would mainly have accrued to schools with significant endowments. Based on this admittedly limited evidence we shall assume that, unlike some US colleges, the income and expenditure of most UK independent schools are very closely related to their

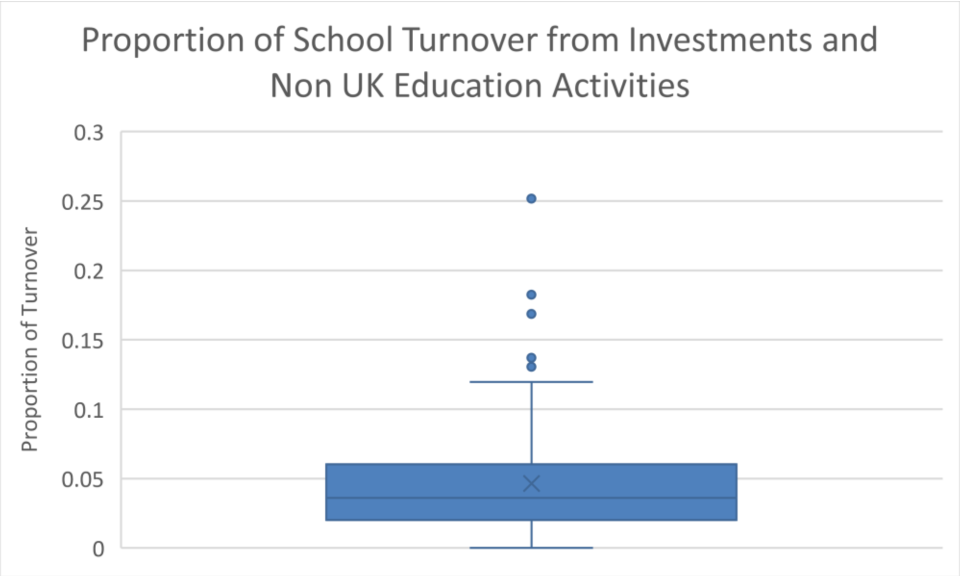

Figure 17 - School turnover from investments and non-UK education activities 2018 - 19

educational activities, thus the first difficulty articulated by Kimball and Luke (separating education costs from those relating to other aspects of the enterprise) should be less impactful

[106] These appear in all of the sampled years.
[107] A number of schools have therefore been excluded where, for example, the same charitable organisation runs disparate schools and hence the necessary values cannot be directly extracted from the accounts.
[108] See University of Pennsylvania (2019b).

for most UK independent schools. This focus on education also makes it likely that increases in cost will be reflected as price increases relatively quickly.

The second criterion in Table 7.2 concerns the homogeneity of the institutions considered and how the range of internal and external factors could drive cost (and hence price) changes differently in different institutions. For this analysis we shall assume that the type of school will be the most significant influence on the observed prices[109]. Although the factors driving (say) the cost (and hence price) of a boarding education should be similar whether the school is entirely boarding or has a mixture of boarding and day pupils, Figure 18 to Figure 21 (which reflect the identifiers from Table 7.4) indicate that schools having both boarding and day pupils appear to charge higher fees to the day pupils, possibly to subsidise the price of the boarding element, which tends to be lower than for schools with only boarding pupils.

These Figures show that whilst the absolute levels of fees have increased significantly between 1930 and 2018 the relative positions of the four different fee types have remained steady. Specifically we observe that the boarding fees at schools that only have boarding pupils have always tended to be higher than those at schools that have both boarding and day pupils. Similarly, the day fees at schools that accommodate both boarding and day pupils tend to be more expensive than those which solely have day pupils.

An additional question about potential bias in the data rests on the observations that, whilst some schools operated continuously throughout the period under consideration, some were founded during the period, some closed and some converted to state control (see Table 7.6). To include the maximum number of data points in the analysis we need to understand whether the sizes and fees charged by these categories of schools are similar to those of continuing schools (i.e. those that operated independently throughout the period).

Figure 22 to Figure 25 show, for schools that closed during the subject period, the fees charged in a given year, plotted against how long before closure that year was. The fees are plotted as a ratio with the mean of the fees charged by continuing schools of that type in the relevant year. These figures show that there is no significant time trend and that the fees charged by closing schools appear to be broadly in line with the continuing cohort. As there must, presumably, have been some compelling reasons for these schools to have closed, Figure 26 to Figure 28 examine how the sizes[111] of these schools have changed over time and illustrate that the majority of these schools are below average size and some seem to have been this smaller size for many years.

[109] It would be possible to examine the impact of location on the prices charged by schools, but that is not pursued in this case.
[111] i.e. number of pupils

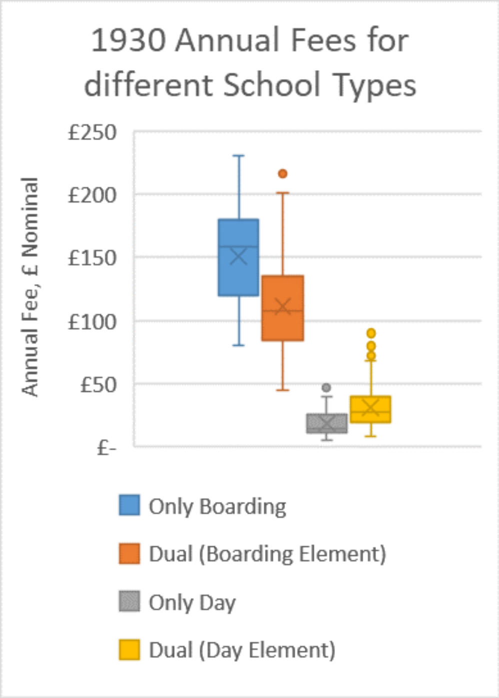

Figure 18 - 1930 Annual Fees for different School Types (Whitaker’s data)

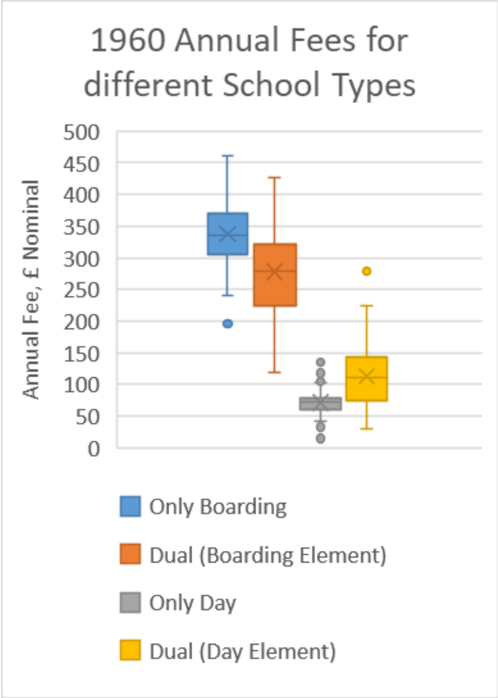

Figure 20 - 1960 Annual Fees for different School Types (Whitaker's data)

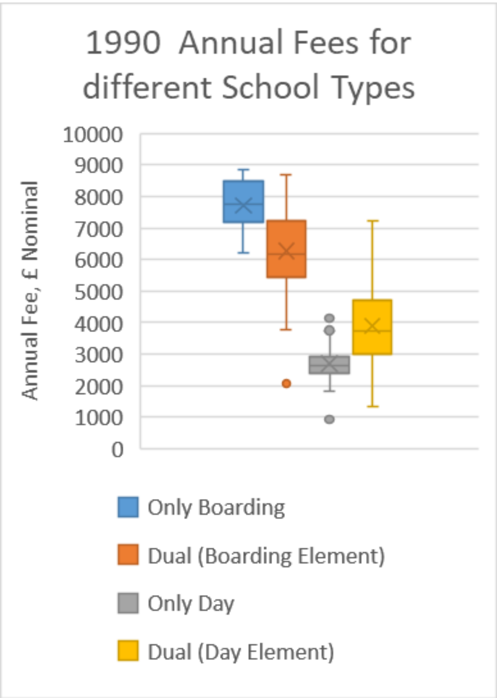

Figure 19 - 1990 Annual Fees for different School Types (Whitaker's data)

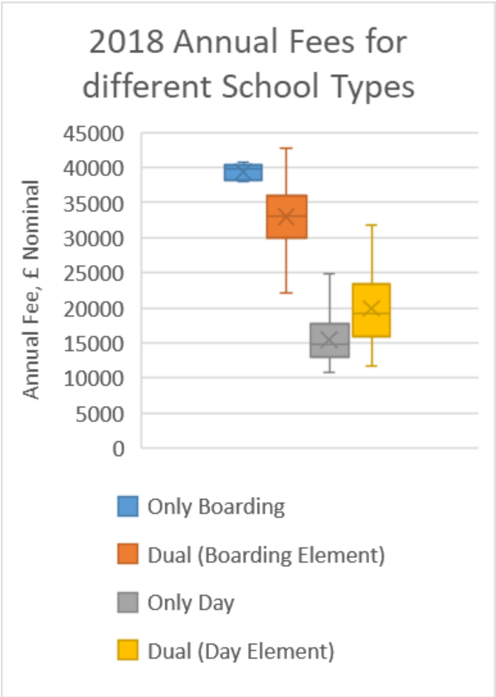

Figure 21 - 2018 Annual Fees for different School Types (Whitaker's data)

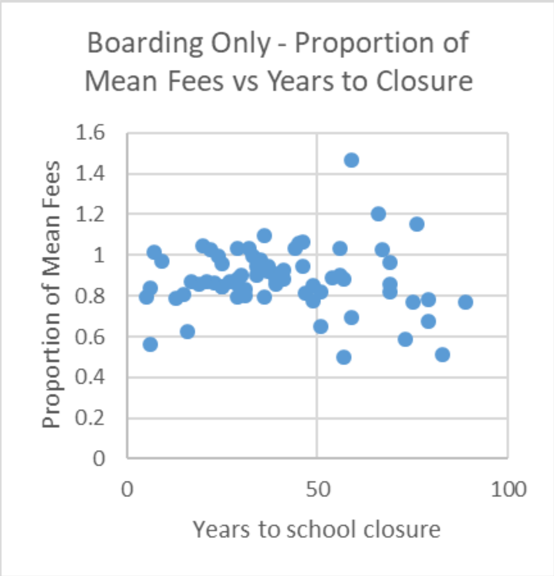

Figure 22 - Boarding Only - Fee Ratio (Closing Schools) vs Years to Closure

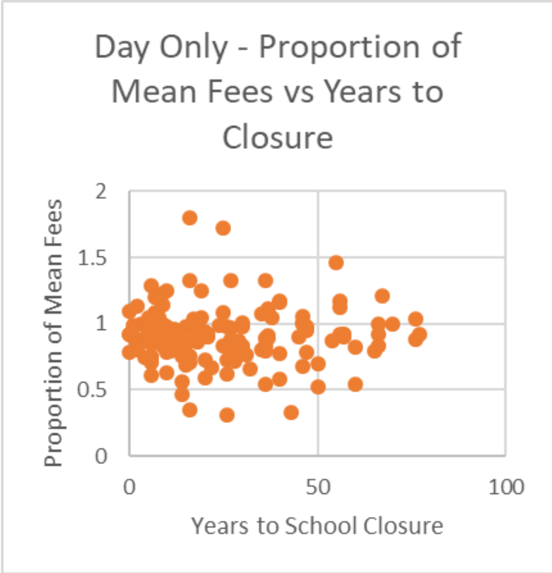

Figure 24 - Day Only - Fee Ratio (Closing Schools) vs Years to Closure

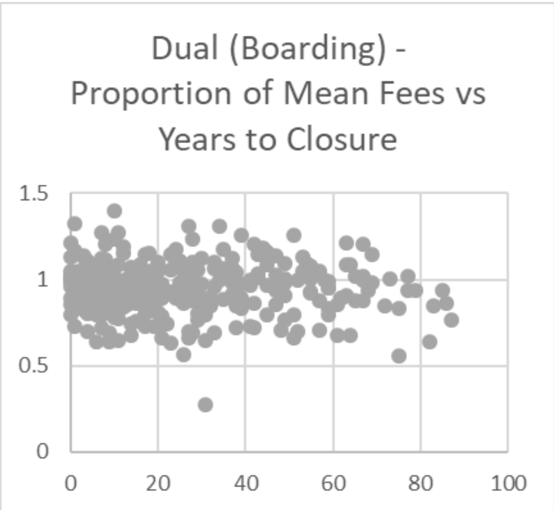

Figure 23 - Dual (Boarding) - Fee Ratio (Closing Schools) vs Years to Closure

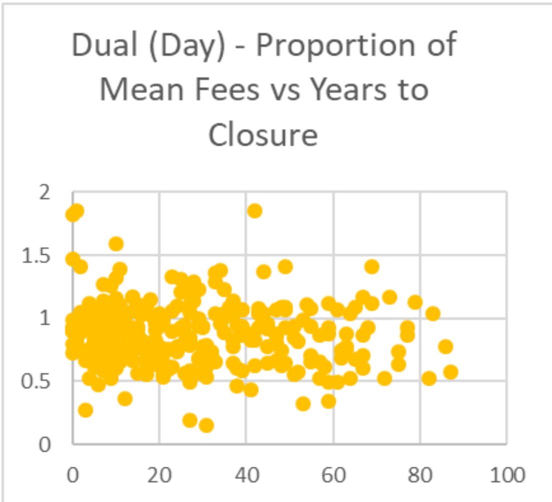

Figure 25 - Dual (Day) - Fee Ratio (Closing Schools) vs Years to Closure

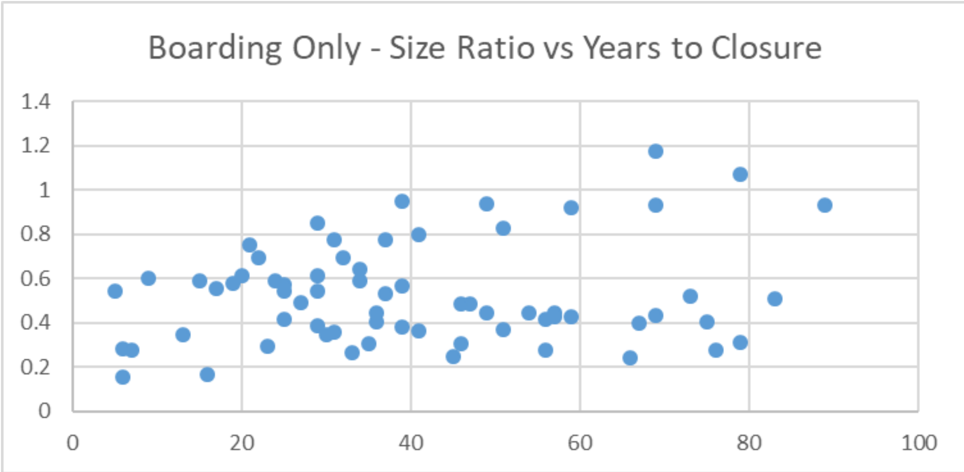

Figure 26 - Boarding Only Schools, Size Ratio with Mean vs Years to Closure

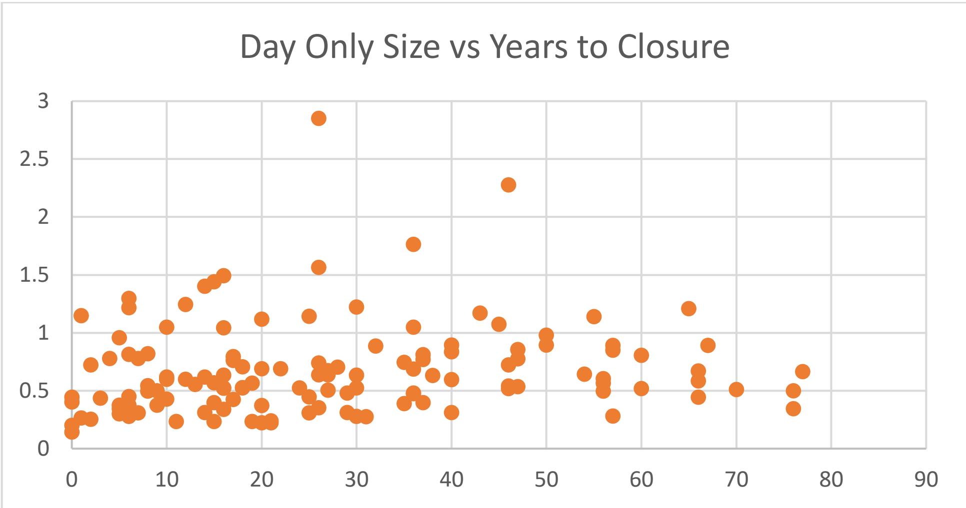

Figure 27 - Day Only Schools, Size Ratio with Mean vs Years to Closure

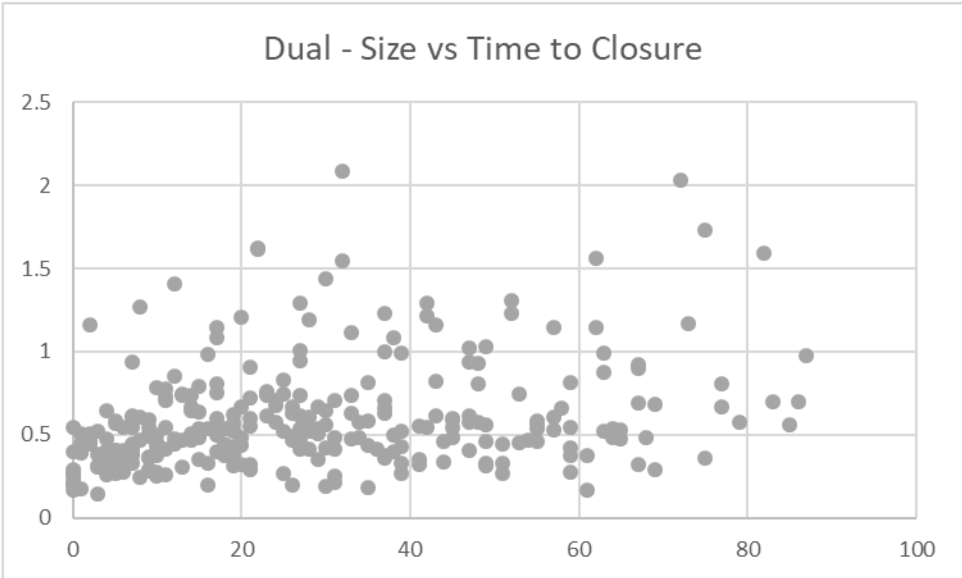

Figure 28 - Dual Schools, Size Ratio with Mean vs Years to Closure

Inspection of Figure 22 to Figure 28 shows that a significant number of data points are available relating to schools that (eventually) closed. Far fewer points are available to examine the fees and size of schools that opted for state control, or which were founded since 1929. Table 7.7 shows this relative difference in numbers of data points.

	1930	1940	1950	1960	1970	1980	1990	2000	2010	2018
Closed BO[113]	10	12	14	14	14	2	2	0	0	1
Closed DO[114]	7	12	13	16	19	17	27	20	15	0
Closed D[115]	27	145	33	36	39	38	63	22	0	0
State BO	0	0	0	0	0	1	0	0	0	0
State DO	40	65	9	10	9	5	6	6	8	0
State D	15	25	5	3	6	2	3	1	1	0

[113] Boarding only
[114] Day only
[115] Dual (i.e. Boarding and Day pupils)

Founded BO	0	1	1	1	2	1	0	0	0	0
Founded DO	0	0	0	0	0	0	7	13	24	26
Founded D	0	0	0	2	4	3	9	9	22	22

Table 7.7 - Number of data points for different school categories

Notes: Each row shows schools of a given History Category (see Table 7.6) and a given Fee Identifier (see Table 7.4), for those schools that were not open as Independent Schools throughout the period covered by this data. Thus, for example, the first row is schools that Closed during the period covered and which only took Boarding pupils.

Although the available data suggests that schools that chose to convert to state control were charging significantly below average fees (see Figure 29), the heavy weighting towards schools that converted in 1944 / 45 (following the 1944 Education Act) means that there are only a handful of schools where the data ranges over more than 15 years. As the focus of this analysis is on the long-term trends that drive fee increases, rather than the factors that drove decisions whether or not to change to state control or to found a new school, we shall exclude these categories of school and base the rest of the analysis on the data covering continuing schools and the schools that closed.

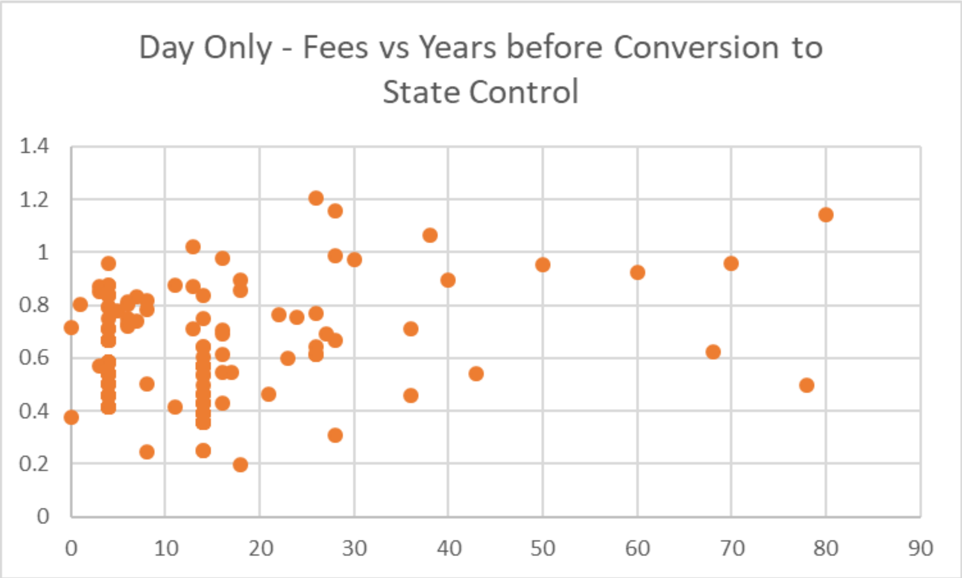

Figure 29 - Day Schools, Fees Charged vs Years before Conversion to State Control

Notes: For Independent Schools that only accommodated Day Pupils, this chart shows the fees charged by those which converted to State control. The fees are shown as a proportion of the average charged by all schools that only accommodated Day Pupils and are plotted against the number of year before conversion to State control.

We thus arrive at a data set that covers 537[116] schools, where each data point is categorised according to whether the school admitted boarding and / or day pupils at that time. Although a number of these schools closed during the period under consideration, it appears that all of the fees charged are broadly comparable and so no major biases should be present in the data. (Note – The third criterion from Table 7.2, of addressing quality changes, only appears in the second data set (see below)).

We may now consider the broad trends that emerge from these data.

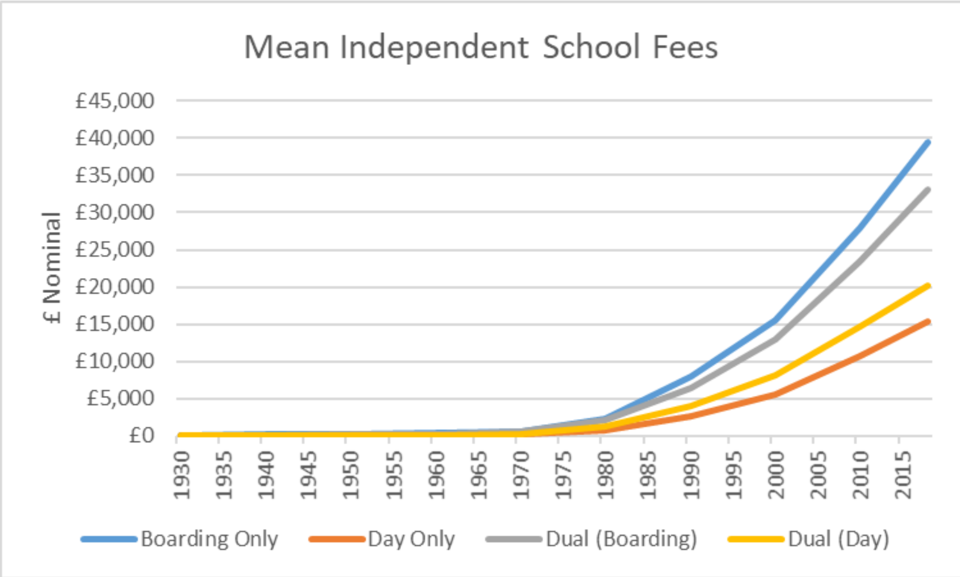

Figure 30 - Mean Independent UK School Fees

We observe from Figure 30 that, as one might expect, there have been significant increases in the nominal values of UK independent school fees between 1930 and 2018. Some of this increase will have been caused by inflation and other changes in the value of money over time, which will have affected the whole economy. It is therefore useful to look for ways of visualising the fee changes in real terms. We start (Figure 31) by comparing fees with average prices in the UK economy by uplifting the values according to the GDP deflator.

This shows that, in real terms, boarding fees fell gradually between 1930 and 1950, but then rose gradually so that they had regained the 1930 level (in real terms) by 1980. The day fees were static from 1930 to 1950, then rose gradually through to 1980. However, after 1980 both types of fee showed rapid growth. Given that the day fees should broadly represent the price of the educational elements, with the boarding fees additionally containing the price of the board and lodging elements, one must assume that productivity gains were made in the board and lodging elements between 1930 and 1950, when boarding fees fell in real terms and day fees remained relatively static.

[116] Representing the numbers from Table 7.6 that were open throughout the period covered, plus those that closed during the period covered.

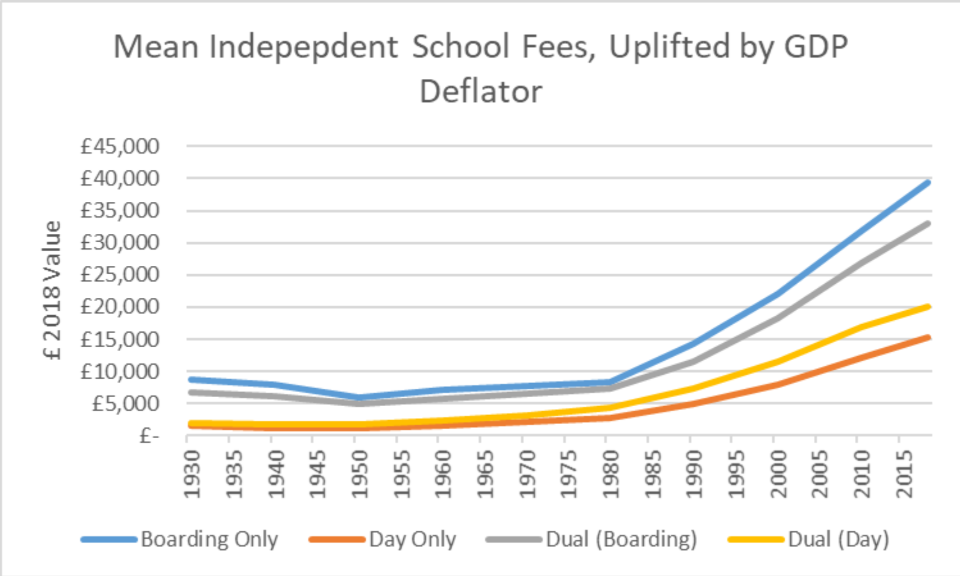

Figure 31 - Mean Independent School fees (Uplifted by GDP Deflator)

If changes in UK independent school fees, relative to average prices in the economy, are driven by Baumol's Cost Disease, as suggested above, then we might expect to see some relationship between them and earnings in general[117]. Relative to UK average wages the productivity gains in the boarding sector are clear, with real-term decreases between 1930 and 1980, whereas prices of day education rose over the same period (see Figure 32). For clarity Figure 33 shows the same results, but as a ratio with their values from 1930.

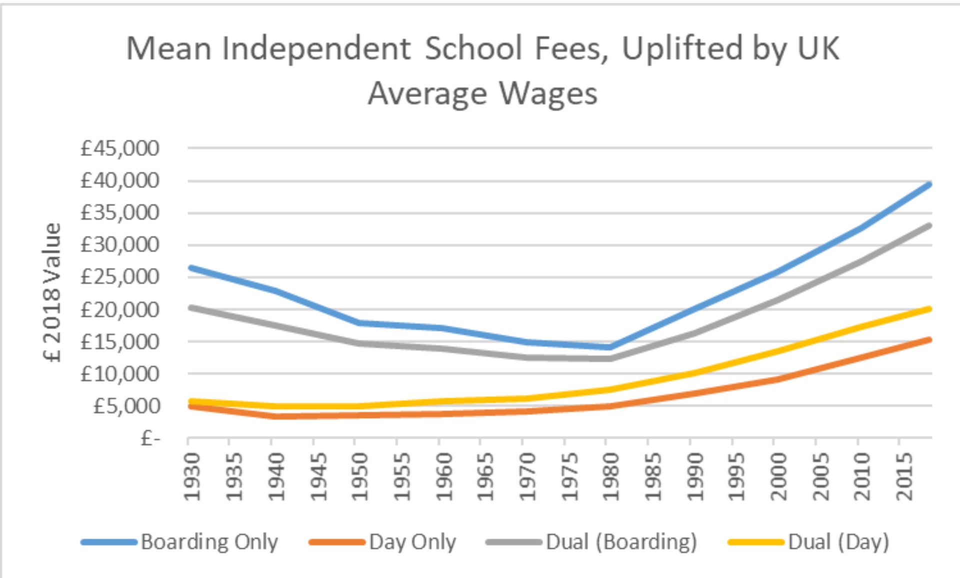

Figure 32 - Mean Independent Schools Fees (Relative to UK Average Earnings)

We thus observe that, assuming no quality changes, there appear to have been productivity gains in board and lodging elements up to 1980, which were not reflected in the price of the

[117] Independent Schools' Council (2001, p.3) reports that 'Salaries of teachers and other staff are the largest items of school expenditure, accounting for up to 80 percent of a typical day school's costs and around 60 percent in a boarding school

education elements. This shows that the education sector is (apparently) capable of delivering efficiencies in some circumstances, indicating that any suggestion of sectoral incompetence is certainly not universally true. Although this data set only contains one set of points per decade, the consistent slope of the lines either side suggests that whatever change may have occurred probably did occur round 1980. The data set discussed below starts in 1977 and so should reflect the most recent phase of price growth noted above.

It should be noted that whilst the Whitaker's data only covers secondary schools (i.e. for pupils aged 11 – 18 or 13 – 18) the second data set, from Independent Schools Council, also encompasses preparatory schools, some of which may have nursery departments (i.e. for pupils in the 0 – 11 or 0 – 13 age ranges).

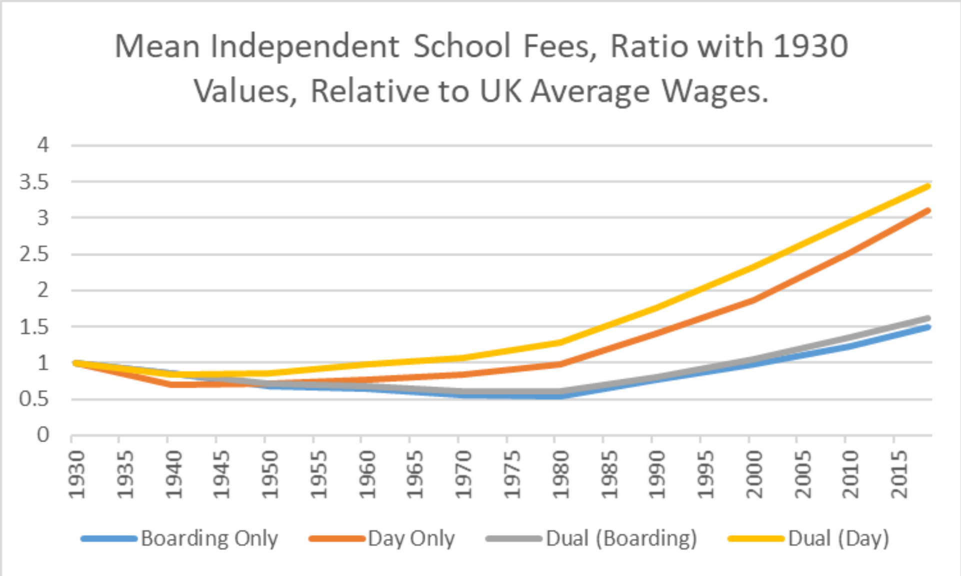

Figure 33 - Mean Independent Schools Fees - Ratio with 1930 Values, Relative to UK Average Wages.

7.4.7 - Independent Schools Council Annual Census Data

The second time series has been produced since 1974[118] by the Independent Schools Council (ISC) (the trade association for UK independent schools) and is contained in their annual census of the sector. This contains aggregated data about the member independent schools, including structure, size, costs and fees charged. The original data is available on the ISC website, see, for example, Independent Schools Council (2019).

In 1974 the first ISC[119] Census was produced, which provided data on a few topics, including school sizes, pupil to staff ratios[120] and spend on new buildings and equipment. In 1975 data was first reported covering average fees in various different classes[121] of school, but it is not intended to investigate those here, and the focus will be on average fees, the staffing ratios and spend on new and improved buildings and equipment, as these are relevant to

[118] Although content has varied over time.
[119] The ISC was initially named the Independent Schools Information Service (ISIS).
[120] Which are often cited as a surrogate measure of quality of education.
[121] As defined by their membership of different trade associations, such as the Headmasters' Conference (HMC) and the Girls' Schools Association (GSA).

examination of the Cost Disease and Revenue Theory of Cost questions. In order to focus on actual changes in the cost of education, without potential complications from variations in the cost of board and lodging and in the pricing differential between boarding and day fees at schools that have both types of pupil, the analysis of this data will be confined to schools that only cater for day pupils.

From 1977 data is published covering the fees at day schools, although until 1987 this separately identifies data from only schools belonging to the Head Masters Conference (HMC) and the Girls' Schools Association (GSA). From 1987 a sector-wide average is also produced and values for this back to 1977 have then been calculated as a constant proportion of the GSA average data.

There have been two method changes in the ISC's calculation of average fees. The first, in 1992 was a weighting change from average per school to average per pupil and the second in 2007 – 2008 introducing greater granularity taking account of the differences in fees charged between juniors, seniors and sixth formers. These are shown in the figures below as separate data sets.

Initially (from 1974) the staffing ratios are only reported in weighted form (with each 6th form pupil counting as two). From 1980 the raw pupil to teacher ratios are also reported, so, by assuming that the proportion of pupils who are in the 6th form remained constant from 1974 to 1980, the raw values from 1974 to 1979 can be back calculated. We note that, on average, for schools covered by the ISC Census, there has been a 35% improvement in this ratio between 1974 and 2022. There are two relatively minor scope changes in this series. Until 2013 all staff and pupils are included in the ratios. From 2014 onwards nursery pupils and staff are excluded and from 2020 teachers who have senior leadership roles are only included when they are teaching. For this analysis, these changes have been ignored and it has been assumed that all the pupil to teacher ratio data represent a single coherent series.

Starting in 1974 the ISC reports the amount that schools spent on new buildings and equipment and also on improvements to existing buildings and equipment. From 2003 information is also included on spending on new / refurbished boarding accommodation. Further granularity is introduced from 2007 when the amount spent on IT is separately itemised and additionally equipment new purchases and improvements are separately listed. The narrative that accompanies the Census data does not make it clear whether the extra categories of expenditure simply provide extra granularity on the previously reported scope of spend, or whether it covers additional scope. For the purposes of this analysis, it will be assumed that the totals reported all have a consistent scope and that the extra categories simply provide additional granularity. The extra expenditure categories represent about 30% of the total facilities spend in 2013 (when the data is last reported) and so some uncertainty in these values should be borne in mind.

We finally come to consider teacher pay. It is common for independent schools to offer pay rates a set percentage above those in the state sector. It has therefore been assumed that annual percentage increases in the pay of teachers in the independent sector will be identical to those in the state sector. The data on state sector teacher pay is drawn from the School Workforce statistics Department for Education (2023) and earlier versions at Bolton (2008).

From these various data sources we can extract the following:

Relative increases in the pay of teachers

Pupil to Teacher Ratios

Spend on Facilities

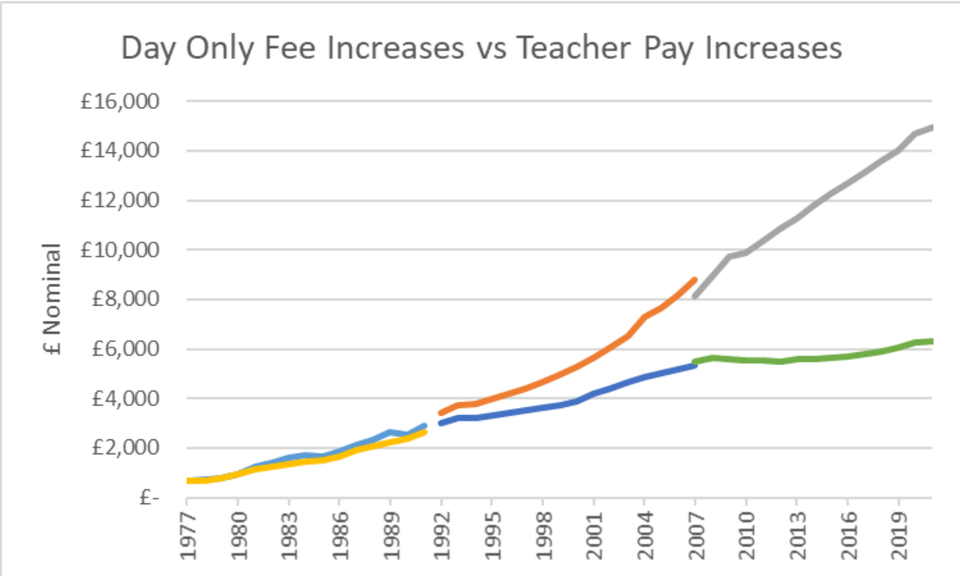

Figure 34 - Day Schools Observed Fees vs Projected based on Teacher Pay Increases

Notes: The lines on this chart are divided into three sections corresponding with the three different approaches to fee calculation outlined in the preceding text. The upper line represents the observed average fees charged, with the lower line representing the expected increases, based on increases in the pay of teaching staff.

Figure 34 shows two sets of lines. The upper set are the observed average fees for Day Only schools, with the lower set being the values we would expect to observe if the increases since 1977 were purely driven by increases in the pay bill for staff[122]. This shows that between 1977 and 1991 the increases might well be mainly driven by the Cost Disease, if no other efficiencies are being introduced. The difference between what was observed and what would be expected from pay increases is slightly less than 12%, given the common 1977 starting point. However, after this point there is clear deviation between the observed and expected results.

If the quality improvements, in terms of better pupil to teacher ratios are also considered, then there is now no significant difference between the fees that were observed and those which

[122] Assuming that teachers and other staff costs form the majority of school running costs and that non-teaching staff pay increases match those for teachers.

would be expected from the increases in teacher pay and in pupil to teacher ratios, until 1995, when the observed difference is 7%.

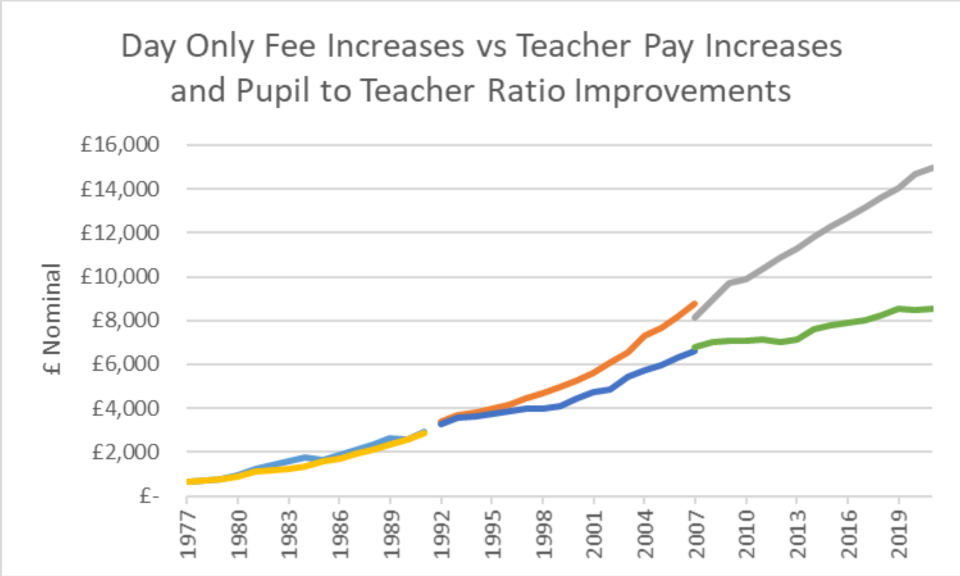

Figure 35 - Day Schools Observed Fees vs Projected based on Teacher Pay Increases and Pupil to Teacher Ratio

Notes: The lines on this chart are divided into three sections corresponding with the three different approaches to fee calculation outlined in the preceding text. The upper line represents the observed average fees charged, with the lower line representing the expected increases, based on increases in the pay of teaching staff and changes to the pupil to teacher ratio.

Finally (see Figure 36) also including the impact that increased investment in facilities would be expected to drive, there is little difference between the expected and observed fee levels until 2007 or 2008, however, after this point, there is a marked divergence[123]. Given that this date marked the start of a global financial crisis which reduced income (see Figure 16) it is possible that reductions in pupil numbers may have curtailed discretionary spending (say on Facilities) and also increased the fees charged. Between 2007 and 2022 there was a 10% reduction in average school size, but further work with the records of individual schools would be required to test the level of connection between these two occurrences.

Similarly changes in employment costs and additional regulatory requirements may all play a factor in explaining what has happened, as may the impact of the Charities Act (2006) which introduced a requirement for schools registered as charities to demonstrate a level of 'public good' in order the maintain their charitable status. Without further information it is not possible to determine whether some or all of them combined are sufficient to explain the divergence observed.

Although a wide range of factors will influence the fees charged by individual independent schools in the UK, it is reasonable to conclude that until 2007 / 2008 the Cost Disease, the

[123] After 2012 ISC no longer publishes data on investment in facilities and equipment, which is why the series ends.

Revenue Theory of Cost and Quality Improvements[124] were all significant drivers of the increases in UK independent school fees.

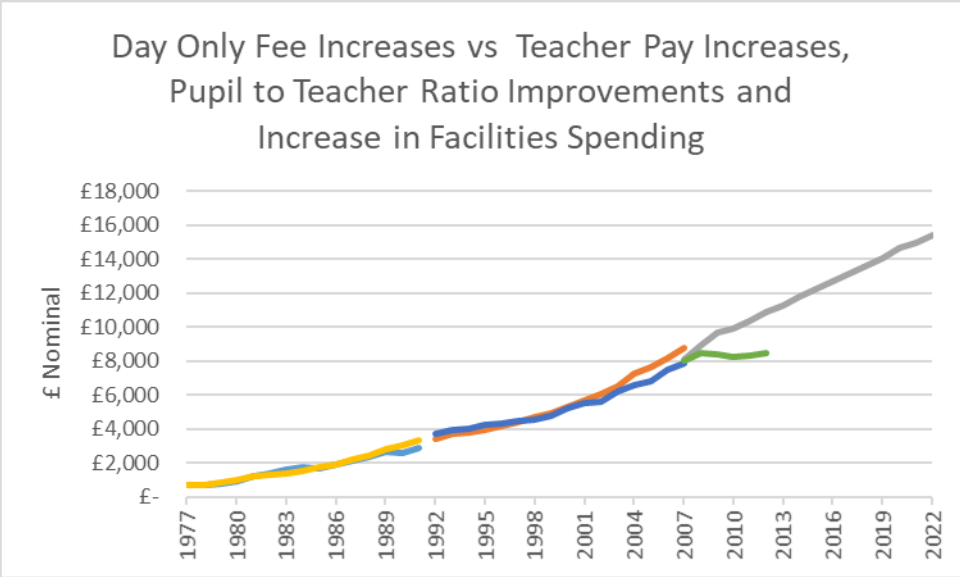

Figure 36 - Day Schools Observed Fees vs Projected based on Teacher Pay Increases, Pupil to Teacher Ratio and Facilities Spend

Notes: The lines on this chart are divided into three sections corresponding with the three different approaches to fee calculation outlined in the preceding text. The upper line represents the observed average fees charged, with the lower line representing the expected increases, based on increases in the pay of teaching staff and changes to the pupil to teacher ratio.

It appears that other factors have come into play since this point, because the observed fees diverge significantly from those expected after this point. Further detailed work on the financial records of individual schools is required to validate the impact of these three factors in earlier years and also to determine what additional factor(s) are now conspiring to push fees up.

7.4.8 - Valuing Quality

Returning to the data used to assemble Figure 17, we may also investigate (for FY[126] 2018 – 2019) the perceived quality of the education generated by different levels of fees. We shall assume that teaching activities and premises[127] spending add to the quality of the education and hence that prospective parents might be willing to pay higher fees to access these (supposed) quality benefits. This may provide insights about why high fee and lower fee independent schools can coexist.

Examining Figure 37 and Figure 38 we observe a clear correlation between the Net Fees per Pupil and the Spend on Premises (i.e. Buildings and Facilities) and a similar relationship between Net Fees per Pupil and Spend on Teaching. We might expect increased Spend on Teaching to equate with an improvement in the pupil to teacher ratio. However, shows that

[124] Assuming that improving the pupil to teacher ratio does represent a quality improvement.
[126] Financial Year
[127] e.g. Facilities costs.

this relationship is less clear-cut. This may be the result of different schools having different approaches to delivering teaching value, but may also be affected by the number of part time teachers. The data represents total headcount, rather than Full Time Equivalent teachers, thus variations in the proportions of part time teachers between institutions will cause scatter.

Recalling Figure 26 to Figure 28, where schools that closed were likely to be of below average size, it appears likely that there is a minimum viable school income (and hence number of pupils). Figure 37 and Figure 38 suggest that above this minimum level there is a wide range of viable fee levels, which will equate to varying levels of facilities and teaching resources. It is notable that for a lower fee level of £15,000 an average spend on Premises of about £2,000 and on Teaching of about £9,000 per pupil is achieved. However, increasing the fees paid results by one-third (to £20,000 per annum) results in average spends of about £4,000 and £11,000 respectively. This suggests that paying higher fees, if they can be afforded, allows schools to spend more on teaching and particularly on facilities. It is possible that the extra investment, particularly in facilities genuinely improves of the quality of the education on offer, however, it is equally possible that this is simply evidence of Bowen's Revenue Theory of Cost at work.

It is worth noting that (as Figure 18 to Figure 21 show) even though a range of fees are charged, no schools charge fees that are (say) double those of close competitors, even if the market would stand this. We have no direct evidence about the cause of this restraint, but it may be driven by the lack of a profit motive in charitable organisations, such as independent schools.

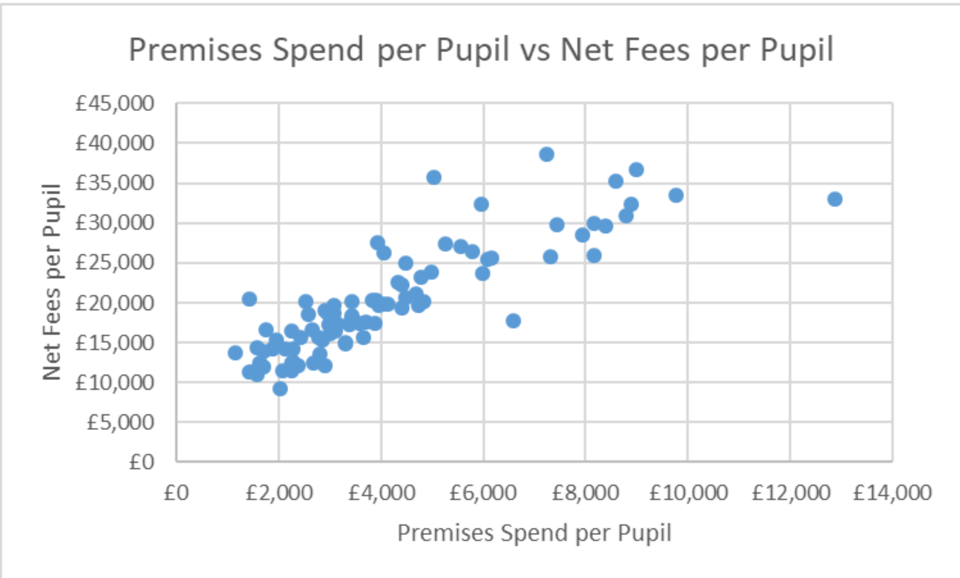

Figure 37 - Premises Spend per Pupil v Net Fees per Pupil (Academic Year 2018 - 19)

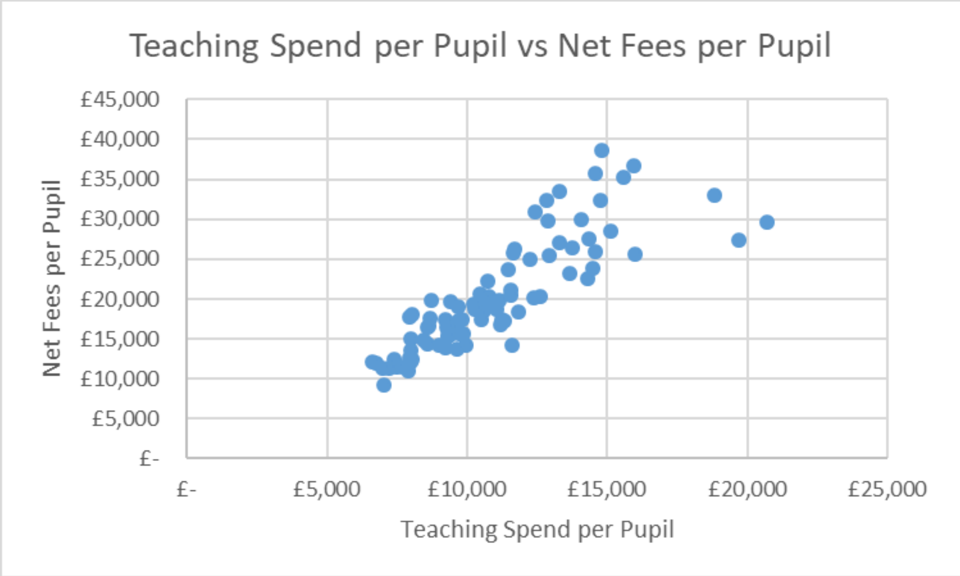

Figure 38 - Teaching Spend per Pupil vs Net Fees per Pupil (Academic Year 2018 - 19)

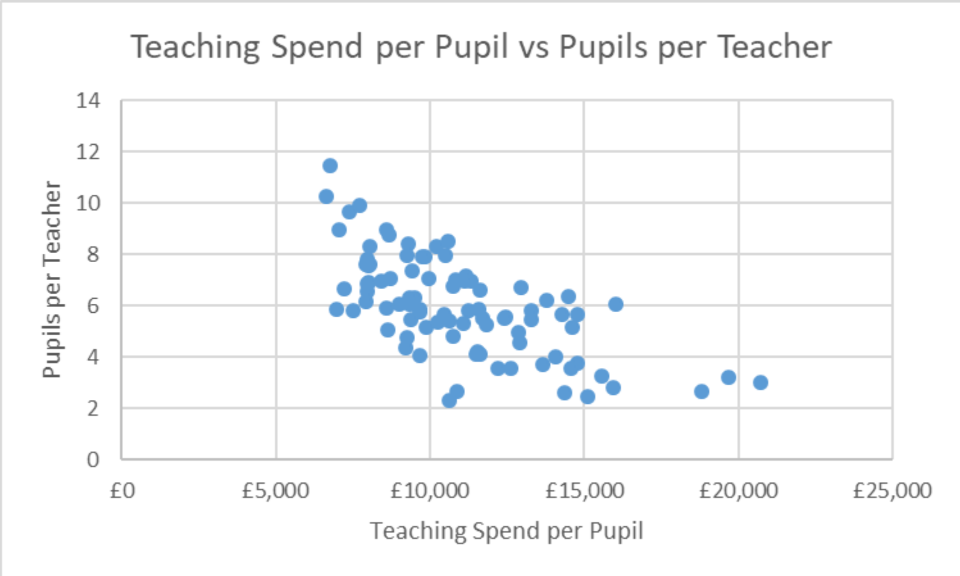

Figure 39 - Teaching Spend per Pupil vs Pupils per Teacher

It must be recalled that education offers little in the way of direct measures of output quality, with quality assessment tending instead to focus on the cost and volume of the inputs. Additionally, the decision making that parents undertake to decide whether to pay for an independent school education is not observable. We can, however, observe that those paying for education currently appear content to accept that increases in the cost and volume of the inputs do truly represent an increase in the output quality and hence the outcomes and so all is well, provided that a sufficient number of them remain able to afford the fees. Naturally, if fewer can afford the fees and hence the viability of some schools is threatened, then there may be a need for increased evidence of how effectively these increases in input quality and volume translate into output quality.

7.4.9 - Conclusions

We have been able to construct data sets that are sufficiently homogenous to make an initial investigation of some of the factors that have been suggested as drivers of inflation in the price of education. Although further work to construct more granular data sets would allow more detailed investigation of specific factors, those currently available have been sufficient for this task.

We have observed that impact of the Cost Disease on teachers' earnings appears to be a significant contributor to inflation in UK independent school fees. If more granular data had been available it is likely that similar observations could have been made about its impact on the earnings of support staff.

We have also observed that the Revenue Theory of Cost may also be driving up spending on facilities and, to some extent, teaching. However, as the actual quality of education delivered cannot be measured or tested, and the education choices that might depend on this metric cannot be observed, we cannot construct a relationship to connect these factors. However, we can observe the effect that that parents who pay the school fees and the school management who make these investments appear united in agreeing that they do contribute to quality and hence are worth paying for. Naturally if the increasing quality (and hence price) starts to make the increased fees unaffordable to so many parents that schools cannot attract viable numbers of pupils, then this situation would have to be carefully investigated.

We have also observed that, subject to a minimum income level being met, a broad range of fees seem to produce viable business conditions for independent schools. We further observed that no schools charge fees significantly higher than their near competitors, even though their market position might make this possible. It is possible but not proved that their charitable status and consequent lack of a profit motive leads to this position.

It is possible that the impact of the Charities Act (2006) and other Safeguarding and regulatory requirements are responsible for a proportion of the additional increases in fees after about 2008, in addition to the impact of the global financial crisis. Further work with detailed, school specific, data would be required to investigate these points.

Considering the broader context, we have observed that parents wish pupils to receive a 'quality' education, to given them an advantage in life[128] and also that schools wish to provide this sort of education, as it will tend to increase the number of pupils attending. However, the actual scope and definition of what defines a 'quality' education cannot, as far as we can see, easily be easily observed and will be influenced by a range of social and economic factors. The literature indicates that the perceived quality cannot be simply measured in terms of academic performance and although we may also observe that some specific changes have occurred over time (e.g. learning Latin no longer appears to be a necessary part of a high quality education) significant additional research would be required

[128] It is also possible that some parents may feel that they personally gain value from being seen to send their children to independent schools that are perceived to provide a high quality education.

to build up anything better than an anecdotal view of the whole range of relevant effects at work.

Frustrating as this lack of a clear approach to observing and measuring quality is for the researcher, it also affects parents and schools in determining the value and value for money of relevant educational offerings. Based on the observed behaviour of the school fees it appears plausible that the need for parents and schools to be able to find some surrogate measure for the quality of education has resulted in a tacit agreement that pupil to teacher ratios and investment in facilities are appropriate measures of quality. We have no way of verifying this conclusion, however it is consistent with the observed relationship between these factors and increases in fees.

We also noted that there are additional, exogenous, mechanisms that will influence the fees, such as the overall pay increases that schools need to award to teachers to attract the right numbers and quality of staff.

7.5 - Factors Influencing the Cost of Running a Formula 1 Team

7.5.1 - Introduction

In this section we examine how the costs of running a Formula 1 motor racing team have increased over time. As the requirements for comparability for educational institutions raised by Kimball and Luke (data specificity, homogeneity and consideration of quality, (see Table 7.2)) appear to have broad applicability, we set out to also consider them with regard to the Formula 1 team enterprise data.

Having briefly examined the history of Formula 1 as a sport, we note that with the exclusion of privateer[129] teams in 1981, each team now has to generate sufficient publicity and other benefits for their sponsors and / or owners. Teams that are owned by a mass market motor manufacturer may have additional ways of generating value (e.g. through technical insights), compared with a team that is simply supported for its ability to generate exposure for the sponsors. However, it appears that the responses from owners and / or sponsors to poor performance are similar, no matter what the ownership model.

The main sets of financial and other enterprise related data are drawn from the annual accounts filed by relevant UK companies with Companies House. Therefore, common standards apply to all the data points from a given year and although accounting standards do change over time, we will expect to see high levels of consistency across the time period covered. This will help ensure that specific and consistent comparisons relating to Formula 1 are being made between the different companies.

We finally come to consider how we can articulate and conveniently measure quality in the performance of Formula 1 teams. The aim of every Formula 1 team is to have one of their drivers win the Drivers' Championship and for their Team to win the Constructors' Championship and hence we choose the end of season rank achieved as the ultimate measure of quality.

We find that performance in race qualifying across the season is a good indicator of the likely ranking of a given driver / car combination at the end of the season. As we are interested in the performance of the team, rather than how good they are at talent spotting good drivers, it is necessary to then pick apart the relative contribution of the driver and the car to their combined performance. We next examine the very significant quality changes that are evident across the whole sport before finally examining the relationship between changes in the cost and volume of inputs to the Teams and the changes in output quality and hence season performance.

7.5.2 - History of Formula 1 Racing

As Judde et al (2013, p.411 - 412) report the 'first Grand Prix motor race was held at Le Mans by the Auto Club de France in 1906'A closed circuit was used because the French authorities had banned road racing following the first day of the 1903 Paris – Madrid race,

[129] Those not building their own cars, but purchasing them from another manufacturer.

which had resulted in multiple deaths[130] of both spectators and racers. In 1946 or 1947 a specification for Formula 1 cars was written by the Commission Sportive Internationale, part of the Federation International de l'Automobile (FIA). There were a number of Formula 1 races in 1947 – 1949, but it appears the first race that was part of the Drivers' World Championship was the British Grand Prix at Silverstone in May 1950. The first Constructor's Championship was awarded in 1958.

Teams entering cars in Grands Prix have fallen into three categories:

Team Category	Definition
Privateer	Purchases race cars from a manufacturer and then prepares and supports them in house
Works (Independent)	Designs, builds and races its own cars (and does not manufacture cars to sell to privateers or the general public)
Works (Manufacturer)	Designs, builds and races its own cars and also designs and builds cars for sale to privateers and / or the general public

Table 7.8 - Categories of Formula 1 Team

Privateer teams were either owned by or supported by wealthy individuals, with the best known being Rob Walker Racing (funded by the heir to Johnny Walker whisky) which won 8 Grands Prix between 1958 and 1968. Apart from occasional prize money, the outcomes of such teams provided no return to their backers, other than the interest and enthusiasm generated by an exciting, if somewhat expensive hobby. Since 1981 only works teams (that have designed and built their own cars) are eligible to enter Formula 1 races.

Judde et al (2013) observe that the sponsorship of Formula 1 developed from 1968 when 'Imperial Tobacco branding appeared on the cars of Team Lotus' (p.412) and accelerated in 1978 when television coverage which had previously been 'fragmented' instead 'proliferated globally'. This led to 'an arms race that saw the budget of the world champion team increase from US$5 million in 1980 to US$40 million in 1990' and 'US$300 million in 2000'. This changed environment also illustrated the potential for mass market motor manufacturers to benefits from the 'positive externalities' (p.413) of a successful in-house Formula 1 team.

This opportunity has led to 'manufacturer teams beginning to displace the independents' (p.413), as the independents need, in the long run, to break even, whereas a large manufacturer may be willing to provide an annual subsidy because of the value of the publicity and technical insights gained. However, it should also be noted that the owners of manufacturer teams[131] have shown themselves willing to abandon Formula 1 when overall finances become tight.

[130] There appears to be no agreement about the actual number of casualties, with numbers between 6 and 12 being quoted.
[131] And other teams owned by organisations whose main reason for existence is not motorsport.

Successful teams, irrespective what type of organisation funds them, receive significantly more media coverage than those with poorer performance, with Table 7.9, which Judde at al (2013, p.415) have extracted from the motorsport trade press, indicating that during the 2006 season the three most successful teams obtained 65% of the race coverage. This presumably means that their sponsors / owners were rather more satisfied than those of the remaining eight teams, whose average race coverage was only 4.3% each.

Team	Type	Race Exposure %	Season Points %
Renault	Manufacturer	29.8	29.34
Ferrari	Manufacturer	22.5	28.63
McLaren	Independent	13.1	15.67
Honda	Manufacturer	9.1	12.25
Williams	Independent	5.0	1.57
BMW	Manufacturer	5.0	5.13
Toyota	Manufacturer	4.6	4.99
Red Bull	Independent	4.5	2.28
Toro Rosso	Independent	3.2	0.14
Midland F1	Independent	1.8	0.00
Super Aguri	Independent	1.3	0.00

Table 7.9 - Race Exposure for 2006 Season Teams and Overall Points Gained

Notes: Source - Judde et al (2013)

Although Table 7.9 refers to the whole season, this will also be broadly applicable at an individual race level and exposure will be heavily weighted towards teams that achieve a high finishing position. Based on this observation, we shall assume that finishing in the top three (i.e. a podium finish), at least once in the season, may be regarded as being successful, in terms of generating some sponsor / owner value. We shall assume that this applies both to manufacturer teams and independent teams and that the value of technical insights gained are not sufficient for a manufacturer to support a team that does not also generate positive publicity. This assumption is key to addressing the requirement that the enterprises being compared are relatively homogenous[132] and so may be expected to respond to internal and external factors in a similar way.

The Formula 1 teams that are based in the UK and whose racing activities are organised as Limited Companies are required to file annual accounts with Companies House, which may

[132] Challenge 2 in Table 7.27.

be accessed through the relevant website (Companies House (2023)). Table 7.10 shows the range of accounts data that have been discovered, with continuing enterprises being shown in a single column, irrespective of name changes. To ensure that the data are all comparable and relevant to Formula 1, broader engineering activities (such as Williams Advanced Engineering) have been excluded.

In addition to financial performance, the accounts also report changes in the volume and cost of the inputs devoted to running each team. Additionally, for much of the period covered, there is visibility of the numbers of employees in technical roles and also in administrative and support roles. We thus have data that should be specific to the motor racing activities and also homogeneous in its coverage of the changes in the volume and cost of input factors.

Team	Williams	Tyrrell	McLaren	Benetton	Jordan	Stewart
Dates	1982 - 2019	1983 – 1998	1982 - 2021	1993 - 2001	1993 – 2005	1997 - 1999
Team		BAR		Renault	Midland	Jaguar
Dates		1999 - 2005		2002 – 2011	2006	2000 - 2004
Team		Honda		Lotus	Spyker	
Dates		2006 - 2008		2012 – 2015	2007	
Team		Brawn		Renault	Force India	
Dates		2009		2016 – 2020	2008 - 2016	
Team		Mercedes		Alpine		
Dates		2010 - 2020		2021		

Table 7.10 - UK Based Formula 1 Teams, Accounting Data Availability

Note – This table shows the availability of annual accounts at Companies House (2023) for UK based F1 teams. Each column represents a single team, with successive rows representing changes in ownership.

7.5.3 - Articulating and Measuring Quality in Formula 1

The scientific, technical and business challenges inherent in Formula 1 and also its global appeal have stimulated the production of a broad literature that is directly based on or refers to various aspects of Formula 1. This diversity is illustrated in Table 7.11, which summarises the topics of the first fifty Formula 1 related publications found on Google Scholar on 22 May 2023:

Topic	Proportion of Publications
Design & Technology	50%
Business (History and Processes)	14%
Politics / Sociology	12%
Other topics (e.g. Medicine)	12%
Economics	8%
Statistics and Analysis	4%

Table 7.11 - Topics for Formula 1 Related Academic Publications

Notes: Source – Google Scholar 22 May 2023.

This indicates that whilst there is quite a broad range of interest in Formula 1, about half of the papers are about detailed design and technology topics, which are not directly relevant to this study. Of the remainder, nearly one third are about the business aspects of the sport, however this is a very broad field and only a small fraction are about the business of running an effective Formula 1 team. We shall return to these later once we have considered how one can best measure successful outcomes in Formula 1, across a number of seasons, and the output quality factors that produce successful outcomes.

The number of publications examining the performance aspect of the sport in terms of its statistics is relatively small. Those that do undertake analysis of this kind appear to fall into two categories, those using Formula 1 performance as an interesting topic, with accessible data sets, to generate interest in the exploration of new analysis techniques and those using relatively traditional techniques to explore what new information can be wrung from the data sets.

A good example of the first sort of publication was produced by Amazon Web Services (AWS) in their Machine Learning Blog series, titled ‘The fastest driver in Formula 1’ (Smedley et al (2020)). This focusses on the detail of the analysis processes followed and the Amazon produced Tools used for the analysis, rather than in-depth discussion of the data or the results. In some quarters this was regarded as a missed opportunity with Carvalho (2022) suggesting it might represent (p.5) a ‘publicity stunt’ connected to Amazon’s sponsorship of Formula 1. Despite this disappointment we note that the analysis has the advantage of simplicity, solely using the qualifying lap times of all drivers from 1983 – 2020, which is a straight-forward data set. It is assumed that the best overall lap time for each driver qualifying for each Grand Prix is used, rather than the best from each of the three qualifying session for each race, however this is not explicitly stated and it has not been possible to identify the exact source of the data used.

In a similar vein, but with broader goals, Patil et al (2023) analyse a broad range of factors (captured as 22 different variables) that affect both individual Formula 1 race results and also the points that are earned during a season. (These data points are drawn from the 2015 – 2019 seasons). However, to appeal to the audience at the Artificial Intelligence and Cognitive Science conference where it was presented, the focus is more on the approaches used, rather than the specific results generated. Despite this focus, the authors do draw solid conclusions, some of which intuitively appear to be correct (e.g. a better grid position is correlated with a better finishing position, having fewer accidents is correlated with a better finishing position) and some that are not, such as tyre choice influencing the overall expected points tally for the season. There is no obvious connection here, as tyres will be chosen to maximise the performance on a given day, so it appears likely that being in a position to select tyres that allow faster speeds to be achieved will lead to better finish positions, however the tyres selected are probably an indicator of an advantageous situation, rather than being a direct cause[133]. In addition to the speed advantage of having selected the correct tyres, an incorrect selection will result in a time penalty as the driver will have to stop in the pits to change tyres.

As the work of Carvalho (2022) was undertaken as a Data Analytics Project it also contains a full discussion of the techniques applied, but also includes a detailed discussion of the data utilised and how the analysis was structured. The data used by this author, which is translated into 14 analysis variables, entirely relates to driver performance and experience. The former encompasses in-race performance, including positions on the starting grids, finishing positions and number of retirements, whilst the latter measures previous racing experience and the outcomes of this experience in terms of Championships won. Notably the author collects no data about the team or the car and indeed criticises (p.8) Eichenberger and Stadelmann (2009) for suggesting that some part of the performance of a driver depends on the car that they are driving. As we shall see, denying the role of the car in the performance of a driver / car combination does not seem to be a sustainable position. It is notable that none of the techniques focussed authors discussed above consider the impact of the car, whereas all of the authors below, who are more focussed on the insights that the analysis generates, do explicitly consider its contribution.

The first ever paper focussed on wringing additional information from the Formula 1 statistics appears to have been produced by Eichenberger and Stadelmann in 2009. Here the authors collected data covering the running and results of all Formula 1 races from 1950 to 2006 and used a multiple regression approach to establish the best drivers of all time, in terms of position held at the conclusion of a race. This is a bold effort and requires 10 data elements for each of the 768 races covered, plus a host of dummy variables to account for different models of car, technical failure rates and other relevant factors. Multiple regressions are used to determine values for the Driver Effect for each of the 302 drivers covered (drivers who never scored a championship point in their career are excluded). This Driver Effect is used to rank the drivers in terms of their contribution to their successes. The authors investigate the robustness of their results and are able to conclude that experience has a positive effect on

[133] This is an example of where a positivist viewpoint can engender confusion between correlation and causation.

driver performance, but that this is small compared with impact of a driver's talent. They also test for evidence of self-matching, where teams with good cars only employ good drivers and good drivers only drive good cars but find no evidence that this is a significant effect.

Whilst using similar analysis techniques and agreeing that both driver and car contribute to performance, Phillips (2014) disagrees with Eichenberger and Stadelmann's use of finishing position as the metric for level of success, and suggests that points scored should be used, as this adds greater weight to the value of a podium finish. This approach is further extended by (for the purposes of analysis) awarding fractional points for positions down to 16th. The results of this analysis indicated that driver performance contributed 39% to the overall outcomes, with the team contributing the remaining 61%. In a later version of the model Phillips (2019) additions are made to the model to allow for age and experience effects and also the relative level of competition in different years, as this affects the level of challenge a given driver will experience.

Building on the work of the previous authors, Bell et al (2016) adopt a random effects model and also attempt to consider the impact of different race weather conditions and the layout of different tracks. The change in method allows the influence of the Team over time to be considered in addition to its specific performance in a given season. Bell et al choose to use the expected results in an individual race as their dependent variable, which (as one might predict) produces greater uncertainty than the whole season performance considered by Phillips. This analysis indicates that, on average, the team's contribution to the overall performance is 86%, with the remaining 14% being contributed by the driver.

More recent work by Rockerbie and Easton (2022) evaluates the contribution of the driver as being between 10% and 20 % depending on the season being considered. According to their analysis, the remaining non-random variation is explained by team effects and also team-driver interaction effects. They also discover that spending more on the team budget and also on the driver salaries improves performance, albeit with diminishing effects as spend increases. However, these authors do appear sometimes to be in danger of neglecting the connection to causation, for example in assessing the impact of driver salary they state 'hence a driver must be paid an additional $10.5 million to improve by one [finish] position'. When they probably mean 'a driver worth an additional $10.5 million must be hired to improve results by one position'. Increasing the pay of the current driver is unlikely to significantly change their performance.

Having considered the analytical literature, it is worth also covering a newspaper article, whose contents are often cited, which appeared in the New York Times in 2009 (Spurgeon (2009)). In this the author cites comments from Nico Rosberg (a F1 driver who retired in 2016) who opines that the 'difference between each driver in Formula 1, from the best to the worst, is about 0.3 seconds a lap' but that from 'the best car to the worst car, I guess it's two seconds or one and a half seconds'. From which thoughts Rosberg concludes that the relative importance is '20 percent driver and 80 percent car'. The assessment of the contribution of the driver is echoed by comments from Frank Williams (who then owned the Williams team) who felt that ‘if your car is a quarter of a second, half a second off the front of the grid and you have the world’s best driver, then he’ll make a difference’.

From this literature we note that there are choices to be made about whether to assess quality as being the performance of a driver or driver and car combination for individual races or across a season and also whether to measure finishing position or the points earned. For this current case study we shall choose the whole season performance for a given driver and car combination, as this measure is less prone to random variations than individual race results. We shall also choose to assess finishing position, rather than points earned, as no points normalisation then needs to be undertaken to account for variations in the scoring scheme over time.

We shall now examine how far some direct comparisons of driver performance support or refute the findings in the literature. Overall one may characterise the aim of a Formula 1 driver in a race as ‘going as fast as possible, without breaking down or crashing’. We shall consider technical failures and human related failures later and so here we shall confine ourselves to examining the potential for speed. Ideally the speed achieved in a race should be measured, however, there can be many factors that will influence the best lap time recorded. Although still subject to random variation, the times achieved in qualifying should be a clearer measure of the potential speed of a driver / car combination.

When discussing performance Rosberg instinctively though in terms of differences in seconds per lap. However the lap times of different circuits can vary significantly, in 2022 the fastest pole qualifying time was at the Austrian Grand Prix at Spielberg and the slowest was at the Singapore Grand Prix on the Marina Bay circuit. The differences in times (1:04.984 vs 1:49.412) are caused by the differences in length (4.318 km vs 4.928 km) and configuration of the tracks (fast racing circuit vs street circuit) (Formula 1 (2023)). In this case, the choice has therefore been made to compare performances as a percentage of the pole qualifying time for a given Grand Prix.

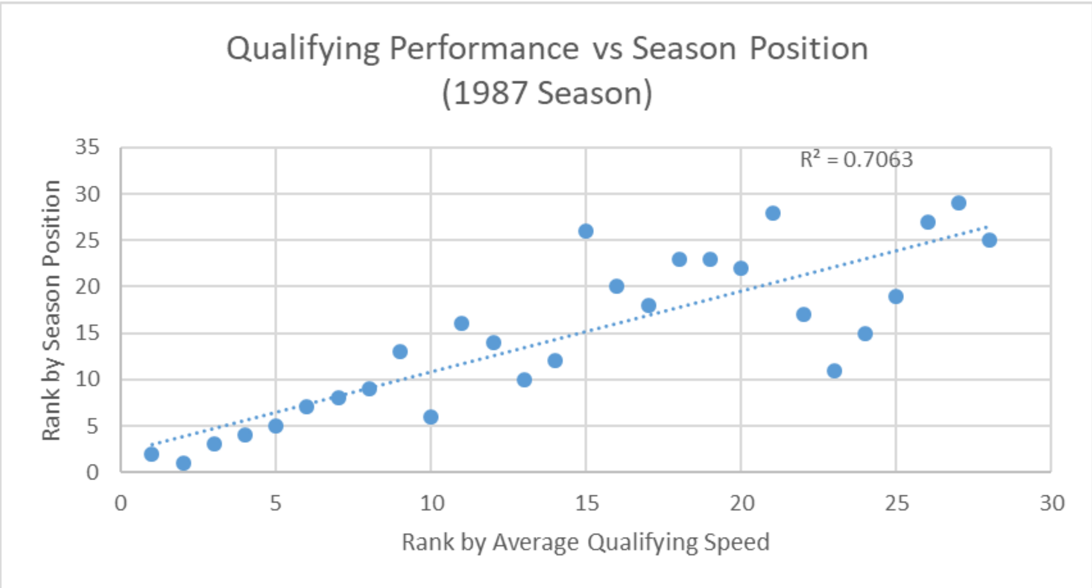

Figure 40 - Qualifying Performance vs Season Position (1987 Season)

Given the evidence from Table 7.9 that performance, in terms of season position, is related to race coverage and hence sponsor / owner happiness we need to examine whether potential speed (as measured by qualifying performance time) is sufficiently related to season performance to be useful. Comparisons are made for several seasons where detailed qualification performance data was to hand.

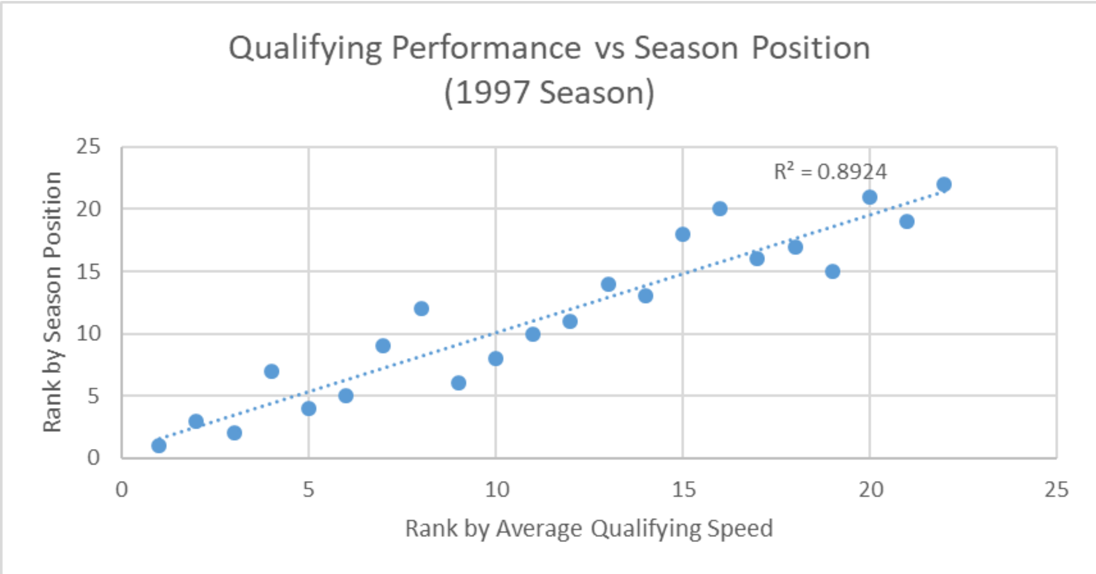

Figure 41 - Qualifying Performance vs Season Position (1997 Season)

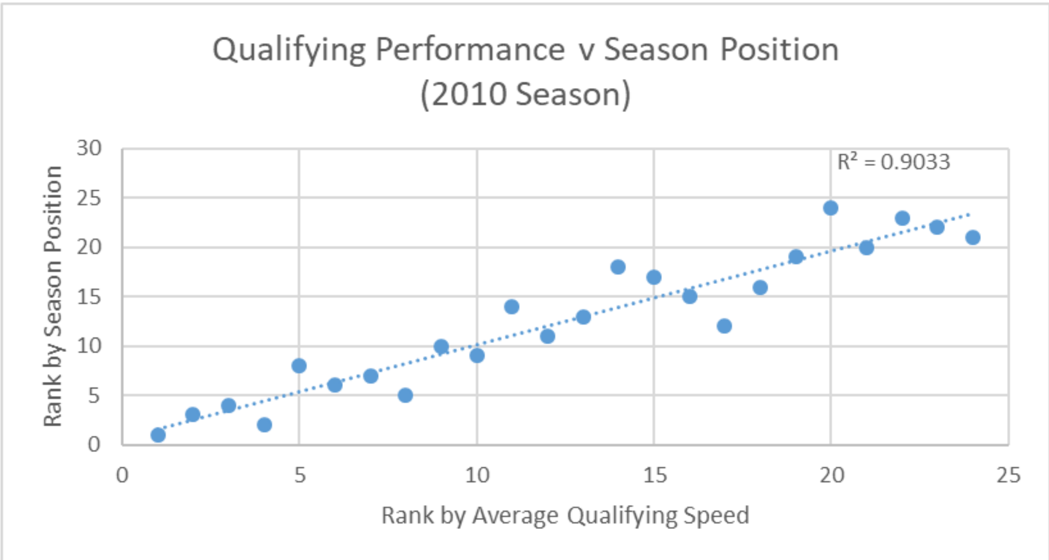

Figure 42 - Qualifying Performance vs Season Position (2010 Season)

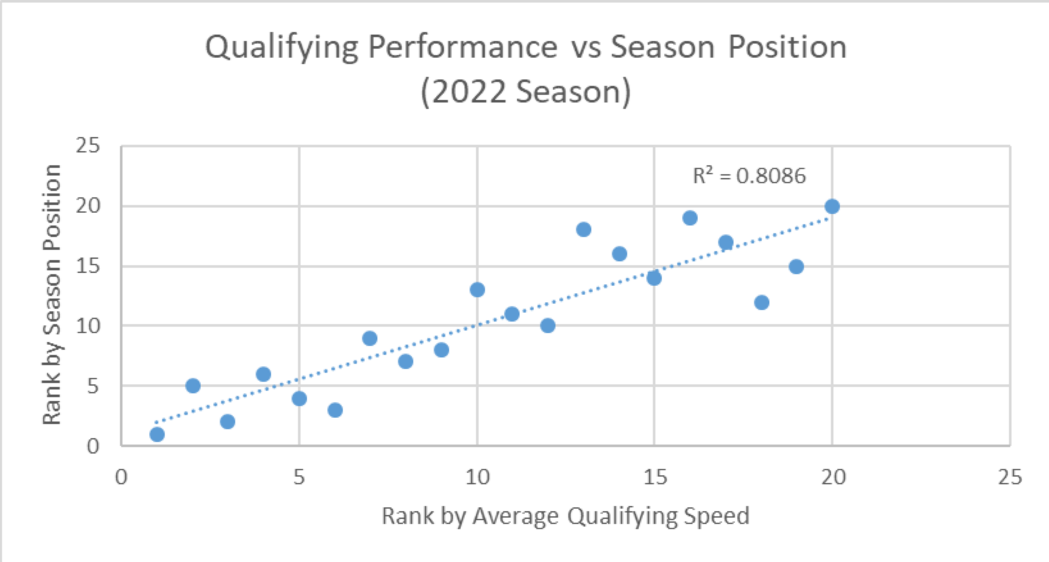

Figure 43 - Qualifying Performance vs Season Position (2022 Season)

The information in Figure 40 to Figure 43 indicates that the average qualifying performance of a given driver / car combination across a season is a good indicator of their likely position in the end of season rankings. This indicates that qualifying performance is a relevant output quality measure as it is related to the outcome measure of season performance of a given driver / car combination. It is interesting to note that the R^2 value increases until 2010, but then is lower for 2022 and further work to examine this relationship in other years could provide insights.

Having found average qualifying performance to be a reasonable surrogate for overall season performance we now explore how specific qualifying performance comparisons between drivers correspond to the literature.

Having collected all the qualifying times for the 1996 season (Formula 1 (2023)) we can compare the qualifying speeds achieved by the two Williams drivers Damon Hill and Jacques Villeneuve. As they were driving nearly identical cars we assume that differences in performance are caused by driver effects and random factors in each race. Averaged across the season, Hill's times, as a proportion of the pole position time, were 0.43% quicker than those of Villeneuve. After that season, Hill, despite having been World Champion, left Williams and drove for Arrows, whose car in 1997 was much less competitive than the Williams (Arrows came 8th in the Constructors' Championship and Williams came 1st). Across the 1997 season Hill was, on average, 2.19% slower in qualifying than Villeneuve, who had remained at Williams.

Assuming the same relative driver performance that we have observed from the 1996 data, then we infer that the Arrows car was 2.63% slower than the Williams and Hill's performance reduced the deficit to 2.19%. So, in this case, the driver contributed 0.43% and the car 2.63% of the performance difference, giving a split of 16% and 84% respectively. These are measured as fractions of the performance difference, not of some notional baseline.

The preceding example featured two drivers at the height of their powers[134] and so we expected to observe (and did observe) a small difference in their performance. By examining instances when three different drivers have driven for a given team in one season we can examine the impact of the changes on the qualifying times (again as a fraction of the pole position time).

In some cases we find that a driver change had a noticeable impact on the performance (where 1% would equate to about 0.65 seconds per lap at Spielberg and over 1 second per lap at Marina Bay). This occurred at Ferrari in 1999 when Mika Salo temporarily replaced Michael Schumacher, due to injury. Other examples include the Stewart team in 1998 when Jan Magnussen was sacked and replaced with Jos Verstappen and also the Footwork team in 1995 when Gianni Morbidelli was temporarily replaced by Max Papis for sponsorship reasons.

Team	Driver 1	Driver 2
Ferrari 1999	Michael Schumacher	Mika Salo
Av Qualification Time	1.0069	1.0175
Stewart 1998	Jan Magnussen	Jos Verstappen
Av Qualification Time	1.0493	1.0387
Footwork 1995	Gianni Morbidelli	Max Papis
Av Qualification Time	1.0390	1.0603

Table 7.12 - Significant Impact of Driver Changes, measured as difference in Average Qualification Time

The general quality of the pool of drivers is illustrated by Mika Salo, even though he was not good enough to have a regular team drive in 1999, he was only about 1% slower than Schumacher in qualifying. Similarly, Jos Verstappen, who had been working as a temporary test driver for Benetton, was about 1% faster than Jan Magnussen in the Stewart. We also note that Max Papis, who was apparently employed because of the extra sponsorship he delivered, rather than his driving ability, was over 2% slower than Gianni Morbidelli.

There have been other occasions when driver changes have illustrated that the two drivers exhibited very similar abilities. In 2019 Alexander Albon was promoted from the Toro Rosso team to drive for Red Bull, with Pierre Gasly moving in the opposite direction. The most notable feature of this move was that it shows that the Toro Rosso was 4% to 5% slower than the Red Bull. Two other examples also confirm that similar performances can be expected from competent drivers. In 2006 Robert Kubica, who was the Sauber test driver, replaced Jacques Villeneuve due to injury and in 2009 when Sebastien Bourdais was sacked by Toro Rosso and replaced with the reserve driver Jaime Alguersuari.

[134] Villeneuve was champion in 1997.

Team	Driver 1	Driver 2
Red Bull 2019	Pierre Gasly	Alexander Albon
Av Qualification Time	1.0149	1.0191
Toro Rosso 2019	Alexander Albon	Pierre Gasly
Av Qualification Time	1.0695	1.0603
Sauber 2006	Jacques Villeneuve	Robert Kubica
Av Qualification Time	1.0162	1.0157
Toro Rosso 2009	Sebastien Bourdais	Jaime Alguersuari
Av Qualification Time	1.0209	1.0201

Table 7.13 - Limited Impact of Driver Changes, measured as difference in Average Qualification Time

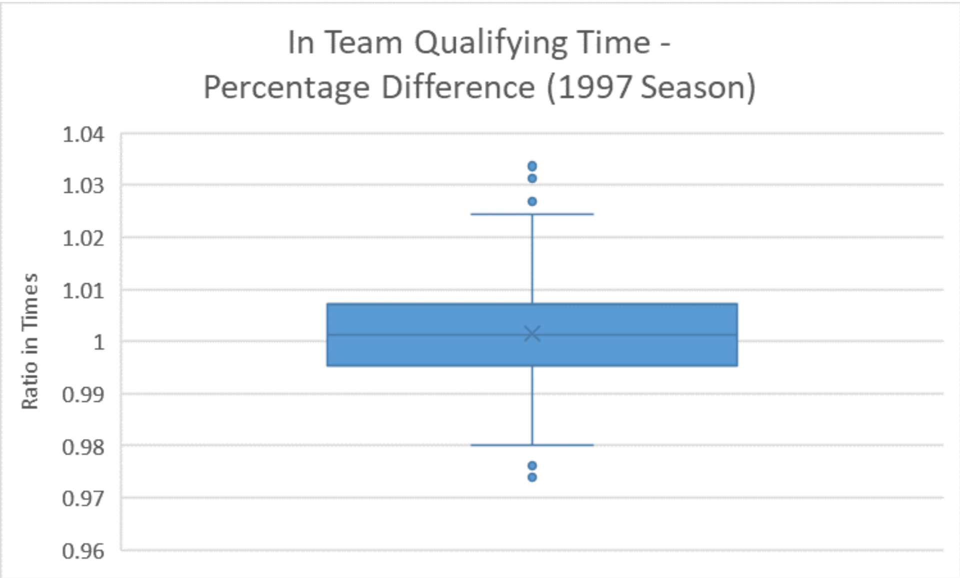

Figure 44 - In Team differences in qualifying times (1997 Season)

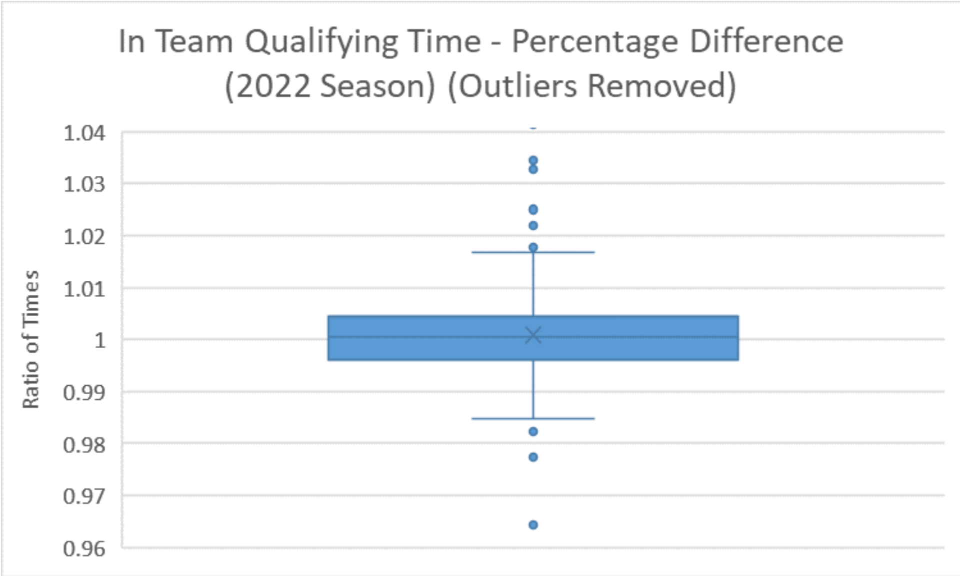

Figure 45 - In Team difference in qualifying times (2022 Season)

Figure 44 and Figure 45 show the differences in qualifying times for each pair of drivers in each team. Despite some outliers in 2022 (which have been removed to make the graphs comparable), it appears that between 1997 and 2022 the variation in qualifying times between team mates has reduced, suggesting an improvement in driver quality and / or a reduction in the contribution of the driver's skill to performance.

Considering the previous literature and direct evidence, there is nothing to suggest that a brilliant driver can make a poor car regularly competitive. However consistent variations in performance between competent drivers of 0.5% seem perfectly reasonable. This would make a difference of 0.3 or 0.4 seconds per lap at a quick circuit, which confirms the views expressed by Frank Williams and Nico Rosberg.

In the somewhat limited literature covering Formula 1 performance, we note that most authors agree that the driver and the team both contribute to the in-race performance and hence the performance across the season. There is also agreement that a considerable array of variables could be included in attempting to build models of driver and team performance over time. Authors make different choices about whether to use race position or season position as the measure of success and whether the actual rank should be considered or, the points earned for the performances (as these are weighted in favour of higher placed finishers). Where performance metrics across multiple years are to be derived then consideration has to be given to the weight that should be given to the performance and experience of the team and driver in previous years.

This case study only seeks to examine the performance of a driver / car combination in a given year, relative to the other competitors, to identify which teams produced competitive entries and the resources that they invested to deliver this. (As noted in the discussion of Table 7.9, competitive is taken to be achieving at least one podium finish during the year). We additionally examine certain trends over in the Formula 1 environment, to explore whether these are affecting the difficulty in delivering a competitive partnership.

7.5.4 - UK Based Formula 1 Teams in Context

Literature about UK based Formula 1 companies generally falls into two categories, the first of which covers the development and sustainment of the economic cluster titled 'Motor Sport Valley' which refers to the collection of 'manufacturers and suppliers stretching through a 100-mile arc in England from Cambridge to Surrey, centred around Northamptonshire' (Henry et al (2021, p.468)). The identified characteristics were a 'group of technologically advanced SMEs geared to small batch production, at very high rates of product turnover within a vertically disintegrated production system, financed by global investment' (p.468). The name appears to be likening it to Silicon Valley, with the president of Ferrari having remarked in 1992 (as reported by Henry et al (1996, p.27)) 'In Italy we are cut away from the Silicon Valley of Formula One that has sprung up in England'.

Motor Sport Valley has been studied in depth and became 'the staple diet for many undergraduate economic geographers across the world as it entered a series of major disciplinary text books as a striking and engaging example of clustering in the face of "flat world" arguments for the impact of recent periods of globalisation' (Henry et al (2021, p.467)). The absolute dominance of UK based teams is illustrated in their performance in the World Constructors' Championship, with Ferrari being the only non-UK based team to have ever won the Championship[135].

Although examination of the factors that have generated and sustained this cluster of world-beating companies is fascinating, we must forgo further discussion and instead focus on understanding the factors and decisions that appear to have shaped the fortunes of individual teams.

Our next step is to briefly consider the way in which the Formula 1 environment, in which all the teams operate, is created and adjusted by the FIA. In examining the Competitive Balance in Formula 1 Judde et al (2013), drawing on earlier authors, suggest the 'spectators derive greater utility from observing contests with an uncertain outcome, and that the more evenly talent is distributed among the teams, the less certain the outcome…will be' (p.416). Without this, 'perennially unbalanced contests would eventually cause fan interest to wane and industry revenues to fall'. Some motor racing series, such as NASCAR, address this challenge by forcing cars to meet detailed specifications, including using 'standardized chassis, gearbox and engine packages' (p.417) so that each car should, in essence, have the same performance. However, in Formula 1 each manufacturer designs and builds their car to a relatively broad specification, which can result in significant performance differences which, as we have seen, not even the most skilled driver can hope to overcome.

Mastromarco and Runkel (2009) have examined Formula 1 rules changes and remarked on 'the tendency to frequent rule changes. In the past 15 years there were on average 9 changes per season.' They recognise that some of these were introduced to improve safety, but that others 'have less to do with driver safety' (p.3003). The FIA will consider and introduce rule changes with the aim of maximising (p.3004) 'its broadcasting revenue which is positively

[135] The 1969 winners Matra International were mainly UK based, being a joint venture between the French manufacturer Matra and the UK based Tyrrell Racing.

correlated with fan interest'. The fan interest is 'increasing the (total or average) performance of the racing teams and in competitive balance between the teams' and as we have previously seen the teams' interest is in maximising their points count and television coverage. The authors find that rule changes can operate in two different ways, either as 'a uniform reduction in the teams' abilities or by a decline in the discriminatory powers of the contest' (p.3004) i.e. increasing the influence that chance has on the outcome.

The authors find that, as one might expect, a season with a low competitive balance will tend to increase the regulation changes at the start of the following season. They also find that regulation changes are initially successful in restoring the competitive balance, however this effect is likely to wear off as teams adapt to the new rules.

Marino et al (2015) consider how Formula 1 teams should best respond to this constantly changing environment. Responding to a lack of consensus in the literature about how to fully describe exogenous environmental changes, the authors suggest examining the rate and the magnitude of changes. The former 'represents the frequency of changes in the environment, and reflects time lags between successive shifts, whereas the latter represents the size of the discontinuity of a new (technological) paradigm' (p.1081).

In considering the balance between exploration and exploitation of new technologies, they note that (p.1082) prior 'research has shown that exploration is relevant for firm performance because it enables firms to discover new knowledge…that better fits the new environment in which they operate'. Some previous authors have assumed that exploration and exploitation can be undertake simultaneously, whilst others feel that scarcity of resources to conduct the two tasks must result in a trade-off. If the former applied, then more exploration and hence more potential for knowledge capture would always be better. In the latter case the potential value generated from exploration would form an inverted U shape, as beyond a certain point greater investment in exploration would reduce the ability to exploit what had already been discovered.

Marino et al agree with the inverted U shape, but contend that technical resource is not the only limiting factor. The complexity and interconnectedness of the design and hence the duration and risk of integrating any newly devised technology will also be important. Once sufficient exploration has been undertaken to assess the likelihood of a technology being derisked, integrated and exploited in a timely fashion, then further exploration may be of limited value. There is no value in too much exploration generating potential design changes that are too challenging to be effectively integrated into the design.

Following quantitative analysis that confirms their supposition, they also examine (p.1087 – 1091) this concept in terms of the 2009 Formula 1 season, which featured the introduction of the Kinetic Energy Recovery System (KERS). Ross Brawn, Technical Director of Brawn F1 is quoted at the start of the 2009 season as saying 'the theoretical advantage of having KERS is perhaps 2 or 3 tenths of a second per lap but you lose in terms of weight, packaging, and torsional moment at the back. In any case, there is no clear decision on KERS. I think it will take a while before we can eliminate the disadvantages. There will be a number of KERS versions, and perhaps more versions that we haven't thought of so far. For us the system only starts to work when we overcome the disadvantages.' Assessing that KERS was too unknown and too difficult to integrate, the Brawn and Red Bull teams decided not to include KERS in the design of their 2009 cars. Following this decision these two teams were dominant, especially in the first half of the season when they won all of the races and scored 21 of the 27 available podium finishes.

There are also a number of other authors, such as Aversa et al (2015), Jenkins et al (2016b) and Boyns (2021) that use Formula 1 activities as case studies to business patterns, company structures and the like in the past and present. Although the specific examples cited in such works are interesting and potentially insightful, there is always the challenge of ensuring causality without having a relevant counterfactual example. Fortunately such topics are not central to this analysis and so whilst some of them contain useful contextual material we may safely park concerns about correlation and causation.

Before considering the performance of individual teams in responding to the changing Formula 1 environment, it is useful to examine some of the quality changes that have occurred across the whole Formula 1 enterprise. As Figure 46 shows the casualty rate in Formula 1 seasons and also in the Indianapolis 500 race was rather alarming until about 1980. Concerted efforts to improve both the safety of cars and also of the tracks that they race on have yielded very clear improvements.

Figure 47 illustrates that part of this improvement is due to advances in medical care. We see that since 1980, provided a casualty survives the accident, then they have a very good chance of not dying. Additionally, since the mid-1990s, better driver equipment and other improvements have also reduced the number of immediately fatal accidents. Mastromarco and Runkel, (2009, p.3003) reflect on these figures, articulating that in 1950s and 1960s one in ten F1 accidents was serious or fatal and that by 2009 this had fallen to one in 300.

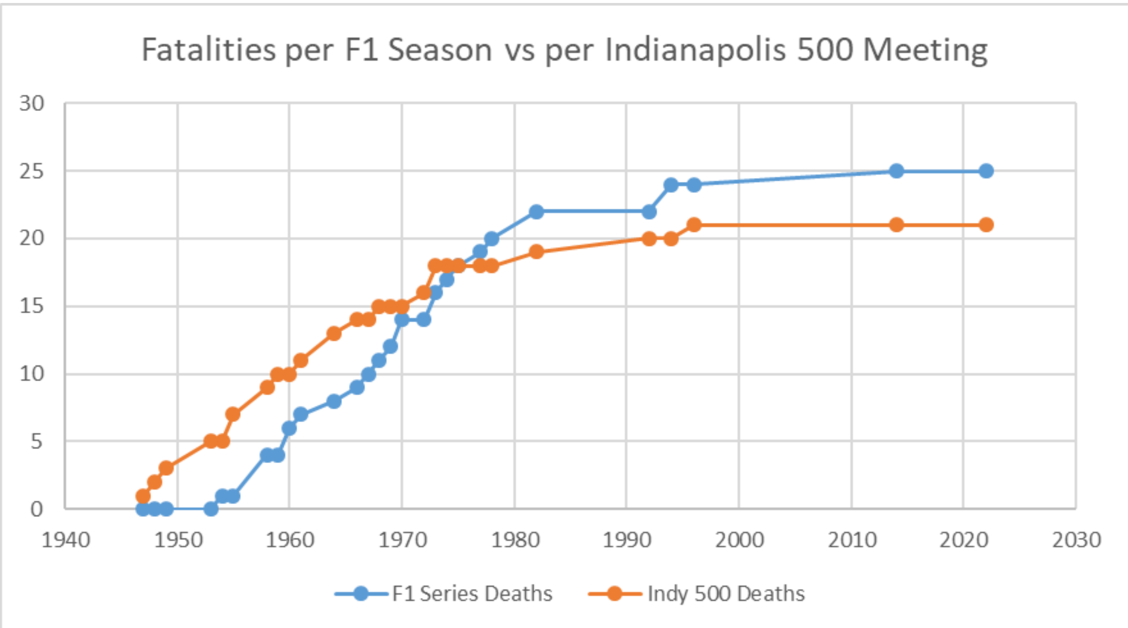

Figure 46 - Fatalities per Formula 1 Season vs per Indianapolis 500 Meeting

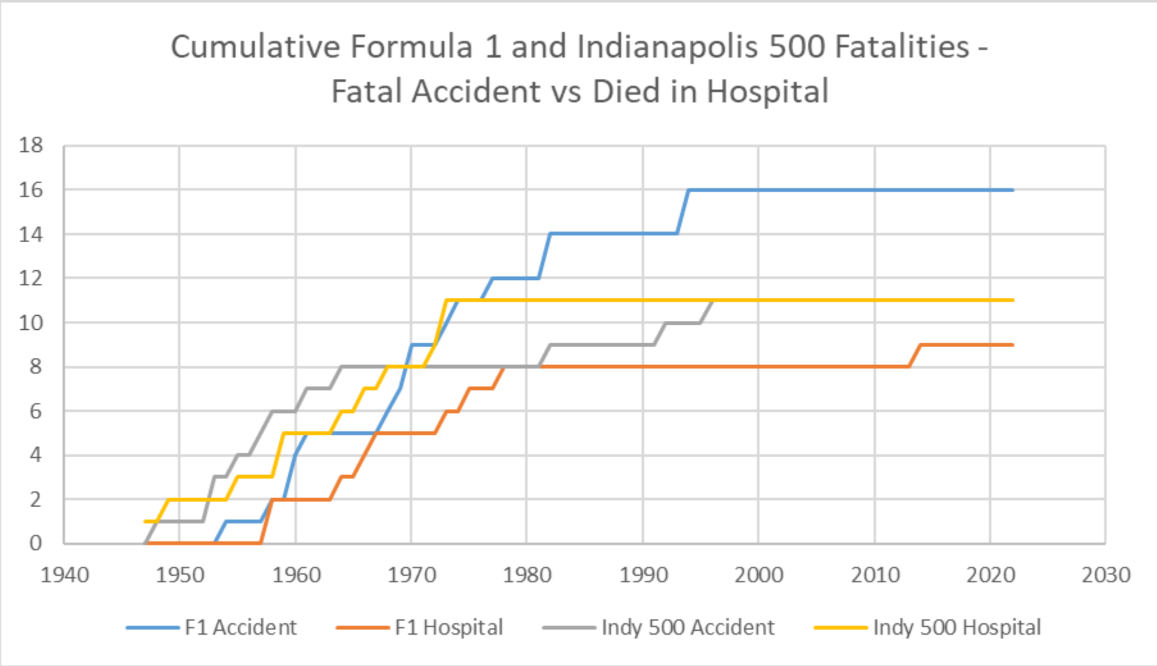

Figure 47 - Cumulative Formula 1 and Indianapolis 500 – Fatal Accidents vs Died in Hospital

There have also been industry-wide changes in system reliability, with Figure 48 illustrating that whilst Human failures (driver errors, pit crew errors, etc.) have remained constant in the 8% to 10% of starts range, the number of starts that end in technical failure have fallen from the 30% to 40% range that applied from the 1950s through to the end of the 1980s and are now also in the 8% to 10% range.

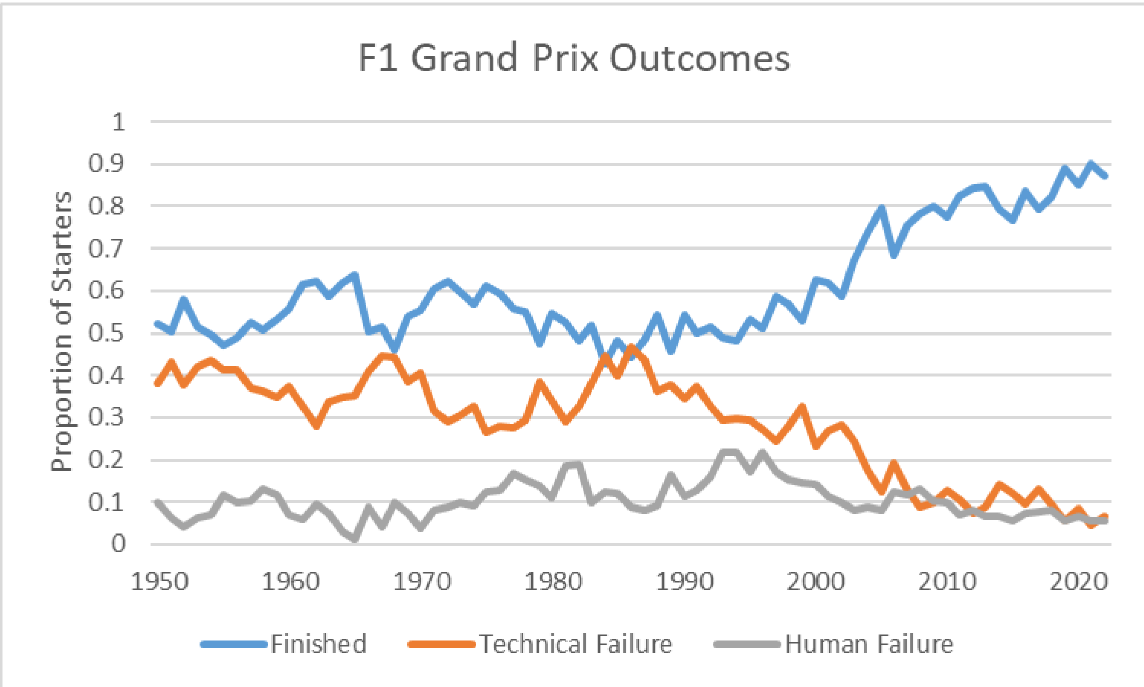

Figure 48 - Formula 1 Grand Prix Outcomes

This improvement in reliability did not occur uniformly across all of the teams, as Figure 49 shows. In this diagram the cumulative numbers of expected technical failures per team is plotted, assuming that the season-wide average applied to each team. Additionally, the actual numbers observed per team is plotted.

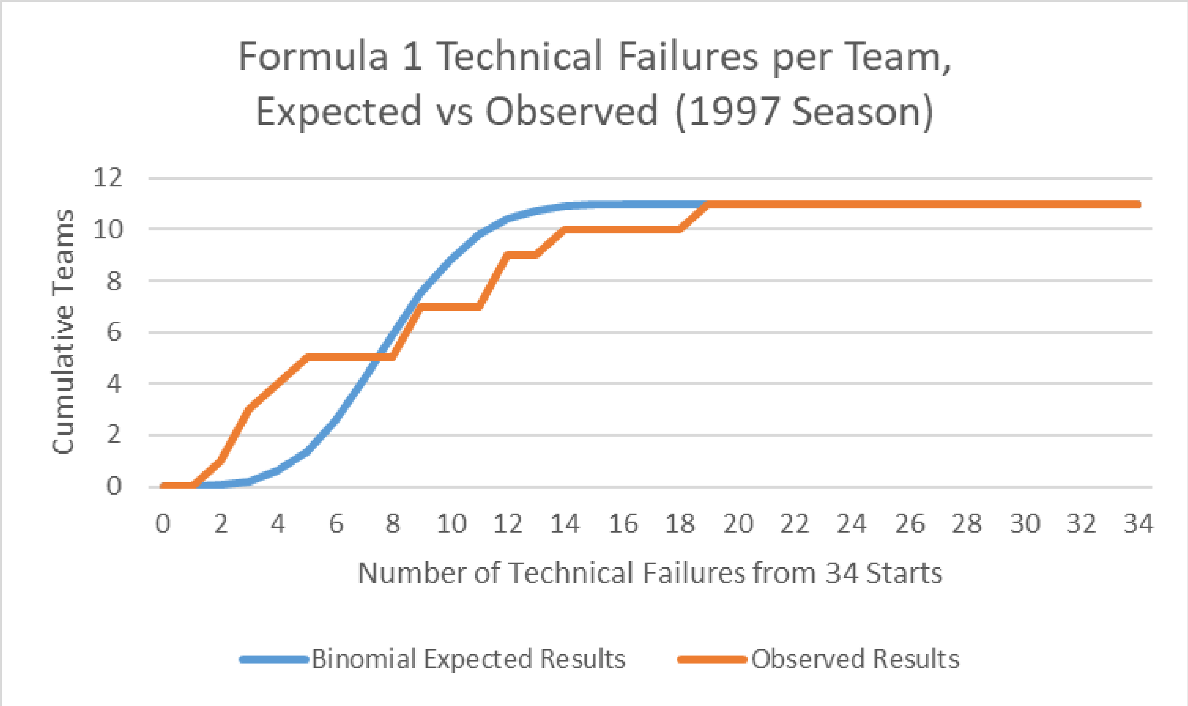

Figure 49 - Formula 1 Technical Failures per Team, Expected vs Observed (1997 Season)

Notes: Source Table 7.14

As we observed from the results of the 1997 season, shown in Table 7.14 there was a tendency for better performing teams to show better reliability. This appears to be necessary for success, as evidenced by the showing of the top three teams, but not sufficient as Sauber exhibited very good reliability, but their driver / car partnerships were insufficiently fast to gain many scoring finishes. Similarly some teams were both slow and unreliable. Minardi's two best finishes were 9th positions, but in each case they had been lapped by the leaders.

Team	Technical Failures	Championship Position
Williams	3	1
Ferrari	4	2
Benetton	2	3
McLaren	9	4
Jordan	5	5
Prost	9	6
Sauber	3	7

Arrows	12	8
Stewart	19	9
Tyrrell	14	10
Minardi	12	11

Table 7.14 - Technical Failures v Constructors' Championship Position – 1997 season.

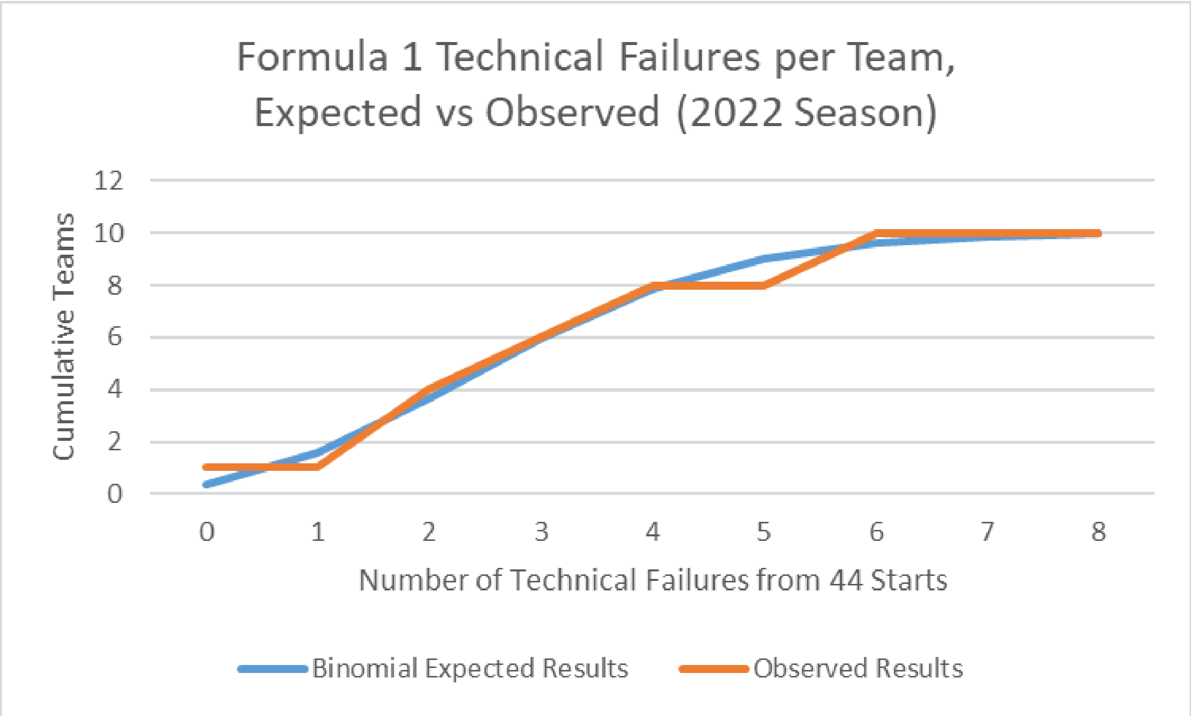

Figure 50 - Formula 1 Technical Failures per Team, Expected vs Observed (2022 Season)

Notes: Source Table 7.15

By the time of the 2022 season the reliability situation has changed significantly. We see in Figure 50 that there is very little deviation from the random occurrences line. Also Table 7.15 confirms that there now appears to be no relationship between car reliability and overall season position.

Team	Technical Failures	Championship Position
Red Bull	2	1
Ferrari	4	2
Mercedes	0	3
Alpine	6	4

McLaren	2	5
Alfa Romeo	6	6
Aston Martin	2	7
Haas	2	8
Alpha Tauri	4	9
Williams	3	10

Table 7.15 - Technical Failures v Constructors' Championship Position – 2022 season.

Before considering the final element of this case study, the resources consumed and performance of individual teams, it is worthwhile summarising progress to date.

We have noted the opinion of different sources on the relative contribution of the driver and the car to the results achieved by a Formula 1 team. Based upon this and informed by the specific examples explored above, we note that variations in the performance of competent drivers can be expected to amount to 0.5% of the best lap time, whereas the variations in the performance of cars from different manufacturers can be four or five times this level. Therefore, whilst the skill of a brilliant driver may make the difference between a car being competitive and actually winning, it will not make a poor car competitive.

Based on Table 7.9, which indicated that the top three teams over the season receive significantly more race coverage, we choose to regard a podium finish (i.e. first, second or third) as being a competitive outcome. Although this assessment is slightly arbitrary, it provides a consistent measure that is easy to apply.

We have also noted the constant rule changes that the FIA imposes on Formula 1, partly in the interests of safety, but also to improve the competitive balance by reducing the value that previously successful teams can transfer into the current season.

We have also noted the industry-wide quality improvements in terms of safety and reliability and note that the other main business of the teams is making quality improvements that do, directly or indirectly, make their car more likely to be successful in individual races and across the season. In the previous Education case study, we found that there was an assumption that increases in the volume and cost of teaching staff and facilities represented an increase in output quality. In this case we shall be able to look for direct evidence of the impact of increases in the volume and quality of technical staff and other resources.

7.5.5 - Team Specific Analysis

We will therefore consider the annual spend of Formula 1 teams, the staff they employed and the results that they achieved. Data from before the period outlined in Table 7.10 is difficult to source and may lack the necessary context to make it meaningful. For example, Boyns (2022) cites the following figures (each reported in separate Times newspaper articles):

Year	Annual Cost of Formula 1 Team, £ Nominal
1939	£250,000
1953	£250,000
Late 1950s – early 1960s	£36,000
1968	£150,000

Table 7.16 - Public Domain Estimates for Annual Cost of Formula 1 Team.

It appears unlikely that all of these figures are correct, however we have no basis upon which to value each. There are other sources, such as a 1977 interview with Bernie Eccleston (Thames TV (2015)) (when he was boss of Brabham) in which he quotes an annual cost of £800,000 to £1 million to run a Formula 1 team, but again we lack to context to understand what is included in this and hence how we may compare it with other figures.

In the interests of consistency and comparability we therefore limit ourselves to the accounting data outlined in Table 7.10. These data sources consistently provide data on the 'cost of sales' and often the number of employees for the Formula 1 teams listed.

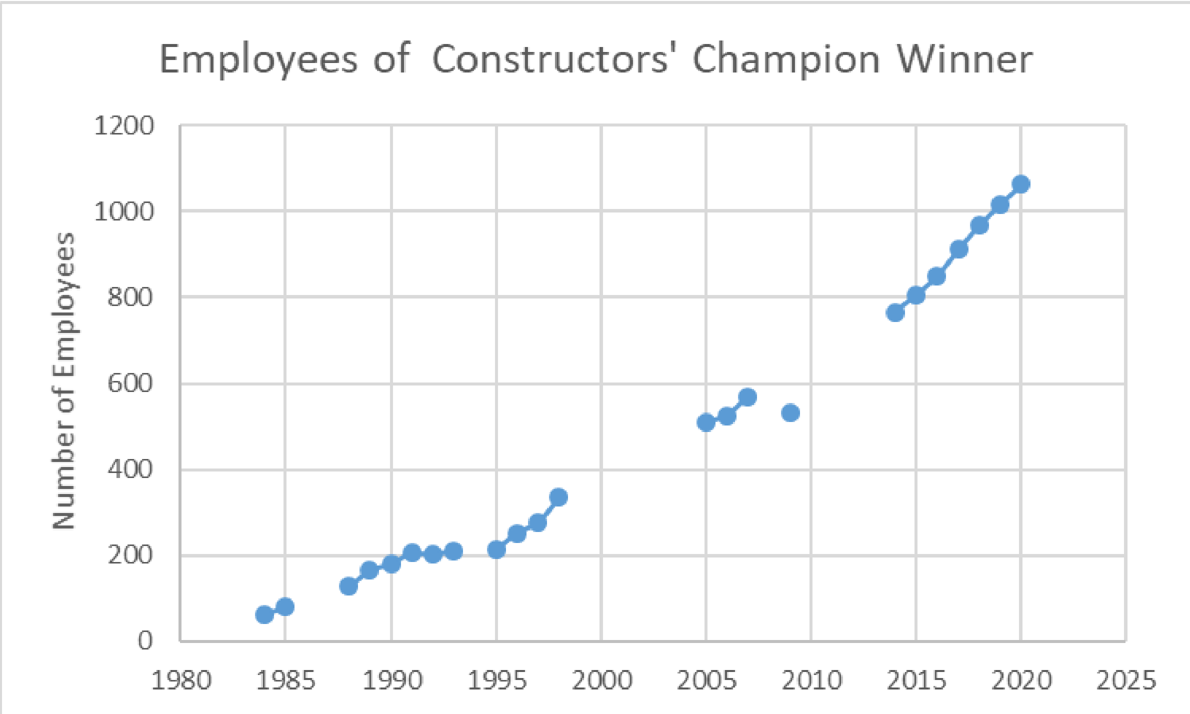

Figure 51 - Employees of Constructors' Championship Winner

Notes: Sources – Published Company Accounts

We may therefore plot Figure 51 that marks the growth of the size of the most successful Formula 1 teams. As this growth in employees occurs following the introduction of significant television revenues in 1978 it was probably fuelled by the increase in money in the sport. However, unlike education, Formula 1 is continually competitive and has very clear measures of success, so we may hypothesise that in a well-managed team, each of the employees will be judged to add value. Any freeloaders would be taking up resources that could be used to make the car more competitive and so would be replaced. This view is reinforced by Figure 52 where we note that the ratio between technical and administrative staff is maintained so that technical staff account for about 85% of the organisation.

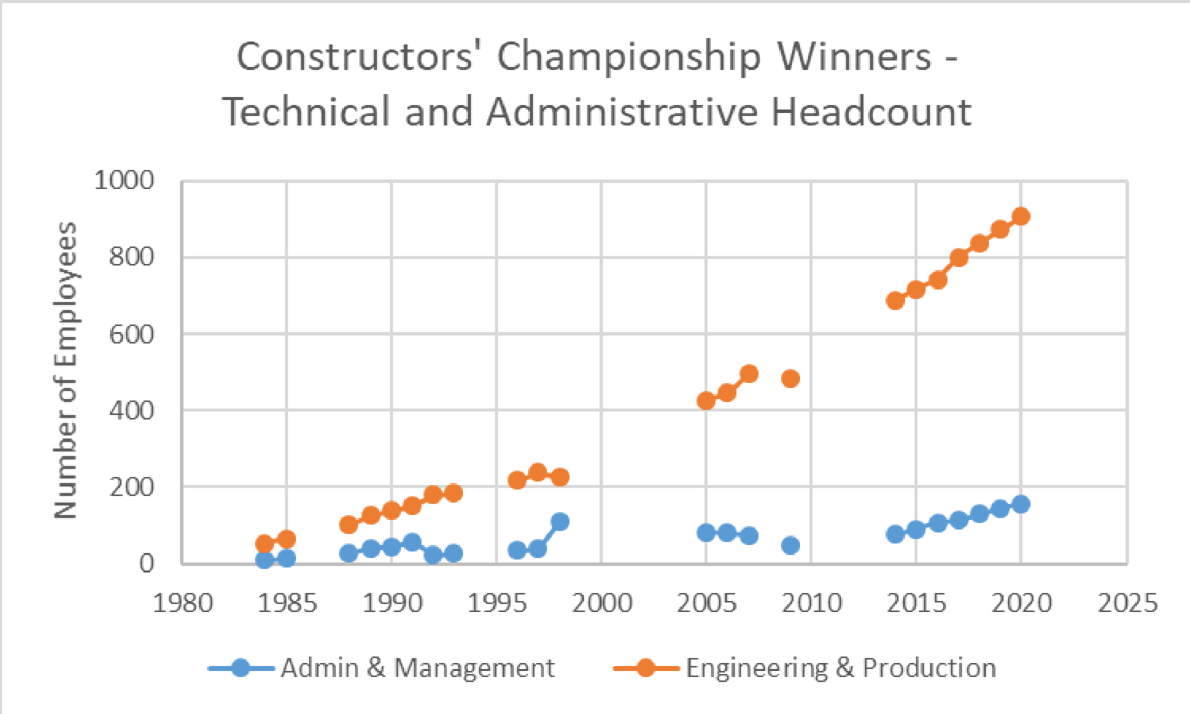

Figure 52 - Constructors' Championship Winners - Technical and Administrative Headcount

Notes: Source – Published Company Accounts

Figure 53 illustrates the ratio between the per capita employment costs for the Constructors' Championship Winner and average UK wages, in each year where the data is available. Whilst there is some variation over the period plotted, there is no long-term trend, which suggests that this industry is not experiencing growth in staff costs that are significantly different from average wage growth in the UK economy.

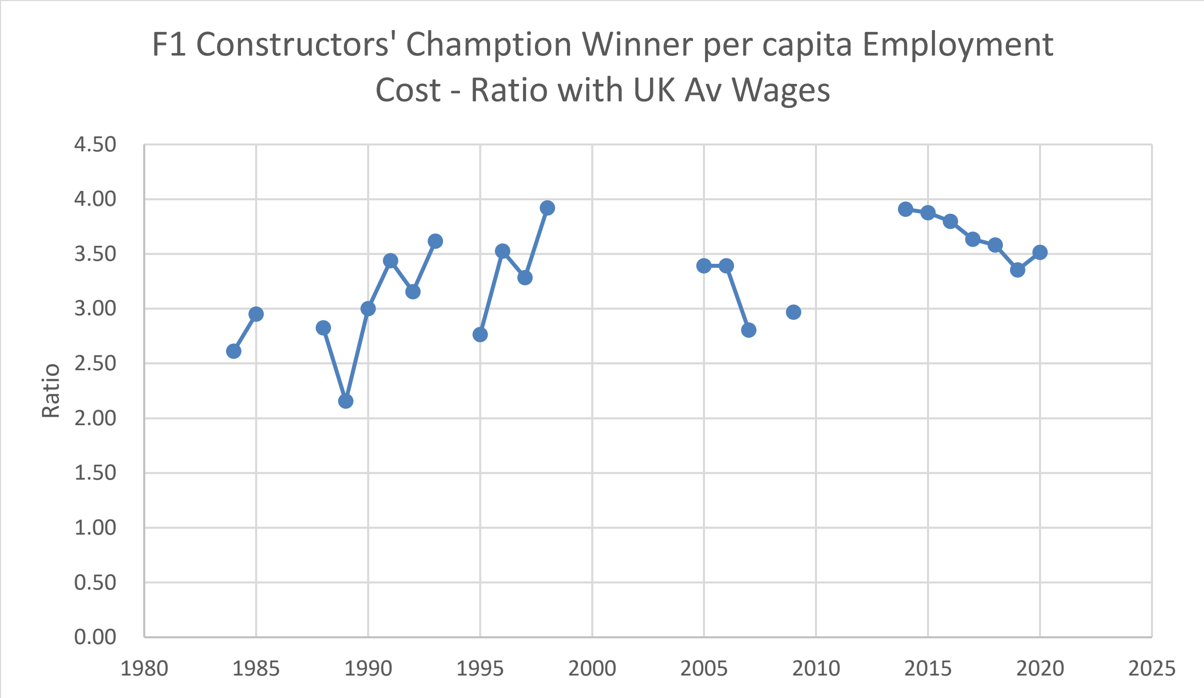

Figure 53 - Constructors' Championship Winner - Per Capita Employment Costs - Ratio with UK Average Wages.

Notes: Data reflecting the changes in wages of individuals employed by Formula 1 teams over time is are not available. Instead it has been assumed that (excluding step changes such as increases in employers' National Insurance contributions) the ratio between employment cost and average wages will indicate whether Formula 1 teams are having to increase salaries to attract and retain the required number of skilled employees.

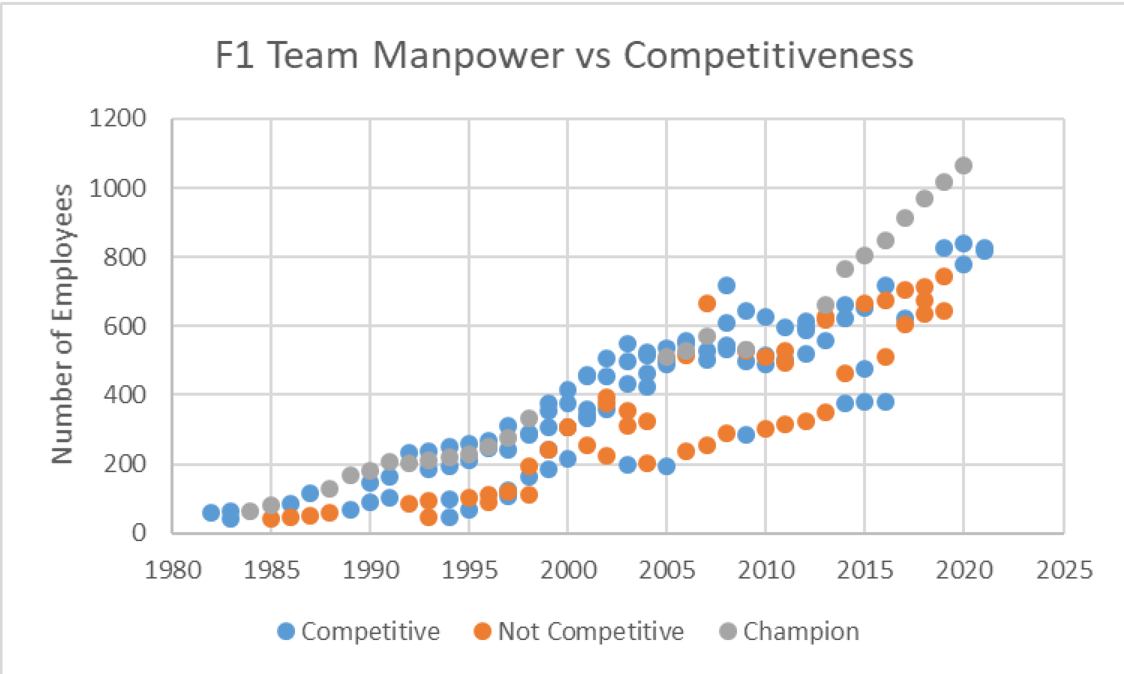

Figure 54 - F1 Team Employee Numbers - Showing Team Competitiveness

Notes: Data points designated as 'Champion' indicates that Formula 1 team won the Constructors' Championship and those designated as 'Competitive' indicate that the team scored at least one podium finish. 'Non Competitive' indicates that the team achieved no podium finishes during the season.

Figure 54 illustrates how the size of all the Formula 1 teams (for which we have data) has increased over time. We observe that it appears likely that smaller teams will perform poorly, however, as such a team is likely to be smaller because it has a limited budget, it is impossible to infer whether it is the lack of staff, the lack of budget or a combination of both factors that leads to suboptimal outcomes. We may also remark that, whilst low staff levels seem to contribute to poor performance, high staff levels are no guarantee of a good outcome. In 2007 Honda had 15% more staff than McLaren, but achieved no podium finishes, having only managed three podium finishes from 18 races in 2006. In 2007 McLaren managed 24 podium finishes from 17 races and Honda managed none (although McLaren was excluded from the Constructors' Championship due to an espionage scandal).

Based on the evidence that we have of actual growth in staff numbers, it appears likely that there has been a considerable increase in the minimum numbers required to run a Formula 1 team. Although we cannot be certain without a counterfactual example, it appears likely that this is driven by increases in the minimum quality and complexity necessary for a car to be potentially competitive. The example of the Brawn team in the 2009 season also illustrates that making good decisions about the overall system design can also be of vital importance in producing a good car.

We may dig slightly more deeply into the growth in the employee numbers at different Formula 1 teams. Table 7.17 to Table 7.19 show the increase in employees for McLaren, Williams and Tyrrell (and its successors, including most recently Mercedes) as these are the teams with the longest series of company data.

Team	Period	Employee Numbers – annual growth rate
McLaren	1982 – 2021	7.1%
Tyrrell / Mercedes	1983 – 2020	9.2%

Table 7.17 - Formula 1 Teams growth in employees 1982 - 2021

Team	Period	Employee Numbers - annual growth rate
McLaren	1982 – 1990	15.2%
Tyrrell / Mercedes	1983 – 1990	12.0%

Table 7.18 - Formula 1 Teams growth in employees 1982 - 1990

Team	Period	Employee Numbers - annual growth rate
McLaren	1990 – 2021	5.1%
Tyrrell / Mercedes	1990 – 2020	8.6%
Williams	1990 – 2019	5.3%

Table 7.19 - Formula 1 Teams growth in employees 1990 - 2021

The data covering the McLaren team and the Tyrrell team (and its successors) starts in the early 1980s. This allows us to observe the impact of the increase in available funding through the 1980s, when the number of employees grew by over 10% per annum. From 1990 onwards we also have data from the Williams team and observe that both McLaren and Williams sustained a growth rate of over 5% per annum through to about 2020. Figure 54 illustrates that between 2014 and 2020 the Mercedes team (successor to Tyrrell), had a run of success and that this was coupled with significant growth in the team size. Figure 52 reminds us that this growth has been achieved through recruiting additional technical staff, rather than the administrative functions becoming over inflated. Each of the Teams had successful periods and lean periods between 1990 and 2020 and based on this admittedly small sample, we observe that to remain current, given the technology advances and rule changes, it appears that an annual growth in technical employees of about 5% per annum is required.

Table 7.20 shows that the growth in the cost of employees, when compared with average UK wages, has grown slowly since 1990, indeed the fractions of a percent reported by McLaren and Williams may well represent increases in National Insurance and other overheads that accrue to the employment of staff.

Team	Annual increase up to 1990	Annual increase since 1990
McLaren	6.1%	0.4%
Tyrrell / Mercedes	4.0%	1.25%
Williams	n/k	0.2%

Table 7.20 - Formula 1 Teams real cost per employee growth

We thus observe that despite having significantly increased the number of employees Formula 1 teams do not appear to have had (at least since 1990) to offer significant wage increases (above the UK average) to support an annual staff increase of at least 5% per annum.

In Figure 55 we observe the vast increase in Formula 1 team budgets (in nominal terms) rising from £3.3m in 1982 to a high of £333m in 2019, which equates to an annual rate of increase of just over 13%. Even for if we consider that the budget of Mercedes might be an outlier, the McLaren budget of £216.7, in 2019 still indicates an annual growth rate of about 12%. As this increase includes the effects of wage increases and other macro-economic factors it is also useful to consider Figure 56 which shows the same data but in real terms (normalised for growth in average wages).

In real terms, the rate of inflation is between 7.2% (for McLaren) and 8.5% (for Mercedes) per annum, which is very similar to the growth rate in staff over the same period (see Figure 54) which is between 7.5% and 8% per annum.

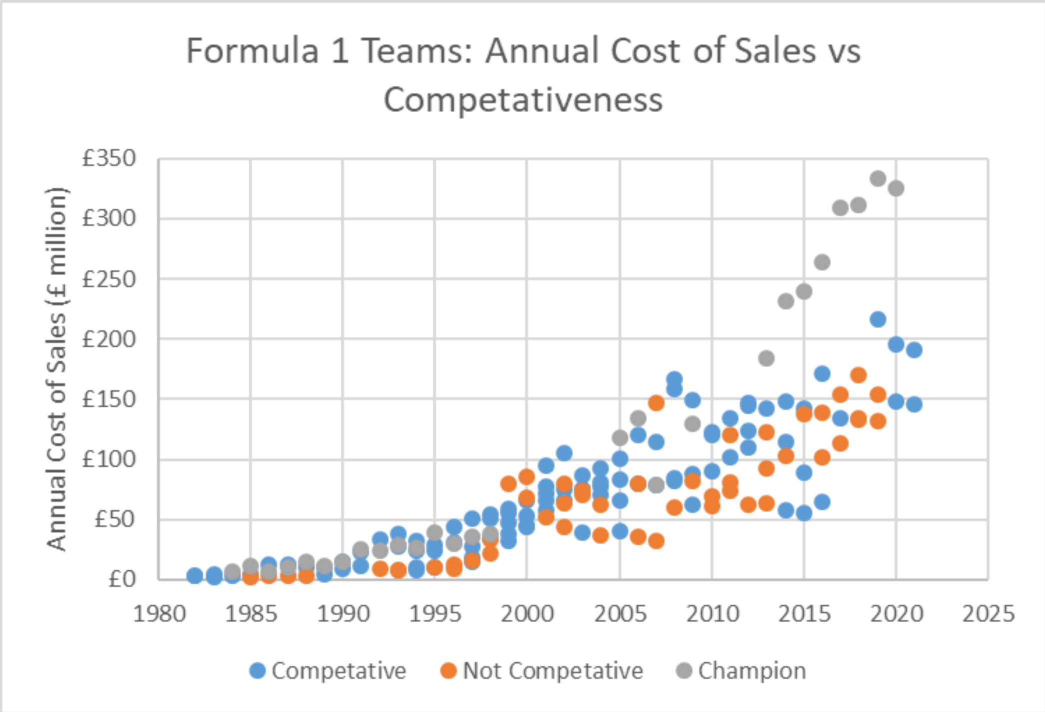

Figure 55 - F1 Teams - Annual Cost of Sales vs Competitiveness

Notes: Data points designated as ‘Champion’ indicates that Formula 1 team won the Constructors’ Championship and those designated as ‘Competitive’ indicate that the team scored at least one podium finish. ‘Non Competitive’ indicates that the team achieved no podium finishes during the season.

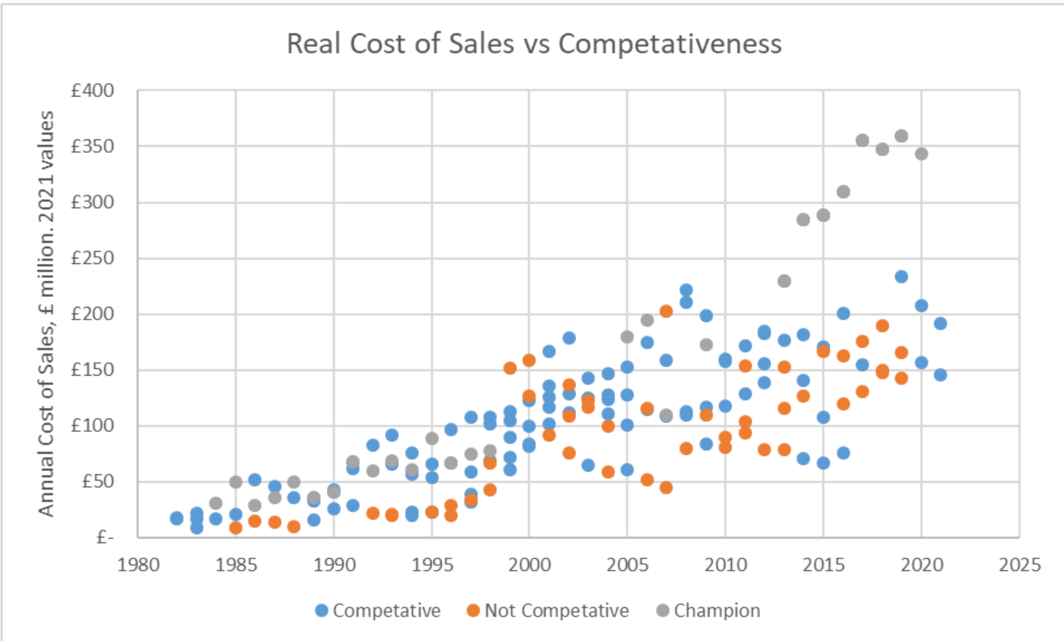

Figure 56 - F1 Teams - Real Cost of Sales (2021 conditions) vs Competitiveness

Notes: Historic values have been uplifted to 2021 conditions by the increase in UK average wages. Data points designated as ‘Champion’ indicates that Formula 1 team won the Constructors’ Championship and those designated as ‘Competitive’ indicate that the team scored at least one podium finish. ‘Non Competitive’ indicates that the team achieved no podium finishes during the season.

7.5.6 - Formula 1 Case Study Conclusions

We therefore conclude that most of the growth in the cost of Formula 1 team activities has been driven by the increasing staff numbers. Given the natural business incentives for effective use of resources (which may not have applied in the previous Education case study) we may assume that the increases in staff and budget are broadly required to dependably produce racing cars that are safe and reliable and which often have a chance of being competitive.

It is not possible to determine what proportion of the staff increases are driven by the need to respond to changes in the environment (driven by FIA rule changes and other factors) and what proportion by the need to remain competitive against other teams who are also attempting to maximise their performance. Further detailed work focussing on the detailed business records of individual companies would be required to shed light on this.

It is notable that the finding from the Independent Schools case study, that once a minimum resource level is reached, then a range of fee values and associated levels of spending can be supported, does not apply in Formula 1. It appears that high levels of investment in staff and presumably matching levels of facilities, are required to enable a competitive car to be produced, but whilst these are necessary, they are not sufficient, as a number of teams have demonstrated that whatever the level of investment a poorly performing car can always be produced.

Having considered the insights that may be gained from the available data, we now consider the context that is giving rise to the observed effects. Whilst the overall Formula 1 enterprise is hugely complex, it appears that there are two main sets of structures and mechanisms at work influencing the design of the cars. Firstly there are those that have the effect of defining the solution space within which the Teams’ design solutions must fit, which include the FIA and its rules and also general technological advances. Secondly there are the Teams themselves (and the other enterprises that have a share in the development, construction and support of the cars), whose structures and mechanisms aim to produce the highest quality car and related support that they can, within the confines of the solution space and available resources.

Whilst we can observe certain effects, such as FIA rule changes, we have no way of determining the relative quality of the cars produced by the different Teams, until the formal Grand Prix races, which are then very effective in providing measurements of the car / driver / support team performance. The immediate and absolute clarity of racing results means that there is no hiding place for mediocre designs and so this should provide an environment in which well managed innovation ought to thrive.

7.6 - Factors Influencing the Cost of Defence Systems

7.6.1 - Introduction

As we have previously observed, in Formula 1 the time frames for decision making are relatively short, with major environmental changes usually appearing annually (e.g. in the FIA pre-season rule changes), performance feedback being gathered every two weeks (during the season) and a new car being designed each year. The short design life of the cars means that a team is only directly penalised for poor design decisions for a single season, with the opportunity to remedy them in time for their new car the following season. Additionally, as each team only runs two cars, implementing minor modifications mid-season is a straight forward task. Similarly, in education, although a pupil may attend a school for up to 7 years, major management decisions are also generally made on an annual basis and so rapid adjustments in response to environmental changes can be made.

In Defence, considering the current US led F-35 programme, we find (Bolkcom (2009, p.2 – 3)) that this first 'emerged in late 1995 from the Joint Advanced Strike Technology (JAST) program, which began in late 1993'. In November 1996 it was announced that 'Boeing and Lockheed Martin had been chosen to compete in the 1997 – 2001 concept demonstration phase'. In October 2001 the US Department of Defense 'selected a team of contractors led by Lockheed Martin to develop and produce the JSF' (Joint Strike Fighter). The JSF, now designated as F-35, achieved Initial Operational Capability with the US Marine Corps in July 2015 (Drew (2015)). Along similar lines, the UK's aircraft carrier Queen Elizabeth was ordered in July 2007 (BBC (2007)) and achieved Initial Operating Capability in January 2021 (Ministry of Defence (2021)).

Thus we observe that complex Defence projects, such as a major warship or a new design of aircraft, can easily take 15 - 20 years to deliver. Even designing, testing and implementing modifications across a large fleet can take years. Also, unlike Formula 1 cars, once delivered major Defence systems are currently often required to be in service for 20 – 40 years and so there should be a huge incentive to make good decisions in the initial design process.

Before attempting to measure the output quality of the delivery of Defence Systems it is important to understand their purpose. According to UK Defence doctrine (UK Ministry of Defence (2022, p,52)) 'Defence's fundamental purpose is to protect the people of the UK, prevent conflict and be ready to fight our enemies'. The quality of Defence Systems may be clearly measured in a 'fighting our enemies' situation, in terms of maximising losses to the enemy whilst minimising losses to ones' own forces. However, in the situation of 'preventing conflict' (i.e. deterrence) the situation is less clear. Deterrence is defined as (p.49) 'the convincing of a potential aggressor that the consequences of coercion or armed conflict would outweigh the potential gains. This requires the maintenance of a credible military capability and strategy with the clear political will to act'. It is also observed, quoting Jim Mattiss, former United States Secretary of Defence (p.49) that 'Deterrence exists in the mind of an adversary'.

Thus we observe that whilst in war fighting the output quality of Defence Systems will become obvious through direct completion (similar to Formula 1), in the case of deterrence, the quality is whatever potential adversaries perceive it to be, just as (in the main) the output quality of education is whatever the stakeholders perceive it to be.

In the cases of education and also airliners (which are discussed below), measuring the output quality is difficult, however, we have the advantage that they operate in a commercial environment and, in both cases, a perceived lack of quality will be demonstrated by a lack of interest from end users (parents and airlines respectively). Any such lack of consumer satisfaction will tend to result in market forces putting the school or aircraft manufacturer out of business. In the case of Defence manufacturers market forces may not apply, as such companies may be regarded as national strategic assets and hence protected from competition. Additionally, the government commissioning a given class of Defence System will often specify the required performance and other characteristics of the product, which tends to absolve the manufacturer for some of the blame if an unsuccessful product is produced.

This difficulty in objectively determining the likely quality of Defence Systems in an uncertain future, the lack of timely feedback and the potentially non-market nature of the Industry makes it very difficult to determine whether suitable Systems are being produced and whether they represent good value for money. This conundrum, coupled with the above-inflation increases in the costs of new Systems leaves Defence open to accusations that their Systems are overly expensive and in the extreme that they ‘represent decadent technology, the know-how of older industries elaborated beyond what is useful along a technological dead-end’ (Kaldor (1981, p.160)). Similar, but less politically strident points are made by Franck (1992). This case study cannot determine whether the right Defence Systems have been bought by the UK, but it can attempt to examine what has driven the price increases in areas where data is available.

However, before diving into the detailed data, it is worth recapping what our critical realist view of the other two case studies has yielded and how we might expect the environment surrounding Defence Systems to behave. Drawing on the Formula 1 view, it would be reasonable in this case to expect there to be structures and mechanisms that yield a defined solution space within which a new Defence System must fit. In this case (as with Formula 1) the availability of new technology will help define this space but the rest of the definition will be much more broadly founded than just FIA rules. For example, national political aspirations and geopolitical imperatives will all be factors as will a range of economic factors. In Formula 1 the team owners and / or sponsors ideally require a car (and relevant support) to be delivered that is very competitive and hence generates championship points and sponsor exposure. As the continued existence of the Team depends on satisfying the owners and / or sponsors, there is a strong incentive for the Team to produce the best car (and relevant support) that it can. In Defence the aim of the government customer will be to receive the best solution that fits within the specified solution space, subject to the resources available to deliver it. The aim of the Defence contractor will presumably be to make a return for their shareholders, whilst delivering something that is arguably close enough to the best potential solution.

However, as we shall observe, the difficulties encountered in determining the quality of an education reoccur in the Defence context, as it may not be clear what features and performance measures of a system may be most important in an uncertain future. As we shall see, whereas parents and schools have agreed that pupil to teacher ratios and investment in facilities are the appropriate ways to measure quality of education, in Defence, there is a general agreement that comparing system performance metrics against those of potential competitor systems represents the most pragmatic approach to measuring quality.

7.6.2 - Initial Exploration of Defence Systems Data

Initially we shall consider some relatively straight forward Defence Systems data, the procurement costs of UK submarines. Figure 57 shows the cost to UK MOD (and previously the Admiralty) of various submarines, as reported in the annual Statement on the Defence Estimates (and its predecessor, the Statement Explanatory of the Navy Estimates). Given the source of the data it appears likely that the data do represent the direct cost to MOD of each submarine, including relevant overheads) thus satisfying the ‘specificity’ criterion from Table 7.2. These historic costs have been adjusted to a common baseline using the percentage of GDP spent (Figure 58) and also by UK average earnings (Figure 59). We observe that, even having been adjusted for changes in the value of money, there is a very significant range for values recorded, particularly for values after 1960.

Figure 60 shows that attempting to extract insights by adjusting for the size of the vessels (in terms of the surfaced displacement) reduces the variation from over two orders of magnitude to less than one, however, it would be difficult to argue that all of these data points appeared to represent a homogeneous collection of submarines. However, given that this data set includes nuclear submarines, traditional diesel electric submarines, midget submarines, hydrogen peroxide powered vessels and HMS X.1, a prototype submersible commerce raider significant variation is not unexpected, but could not necessarily be predicted in advance. (Wage rates rather than percentage of GDP were used to normalise the costs in this case, as we are looking for similarities between the vessels, such as similar numbers of labour hours per ton. If percentage of GDP was used instead, the shape of the graph would be similar, but differently scaled).

Figure 61 shows that when the data set is confined to standard diesel electric vessels a much more satisfactory outcome is achieved[136]. The older results still exhibit scatter, but the results since 1955 are relatively tightly grouped. However, given that these only represent three of the fourteen classes of standard diesel electric submarine, it is unclear whether this apparent convergence is real, or simply an artefact of the paucity of data points.

[136] In terms of generating a homogenous data set.

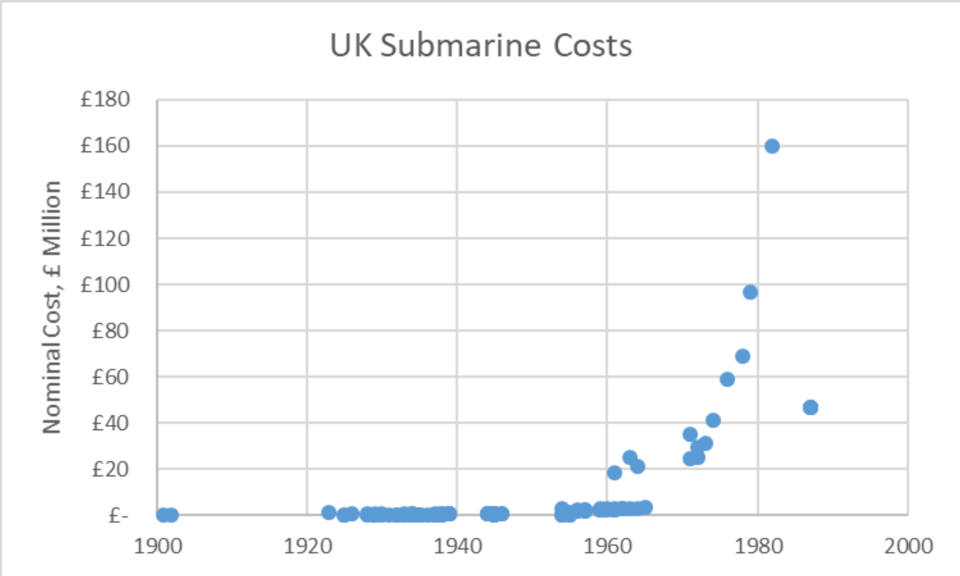

Figure 57 - UK Submarine Costs

Notes: Source – Relevant Statements on the Defence Estimates and Statements Explanatory of the Navy Estimates.

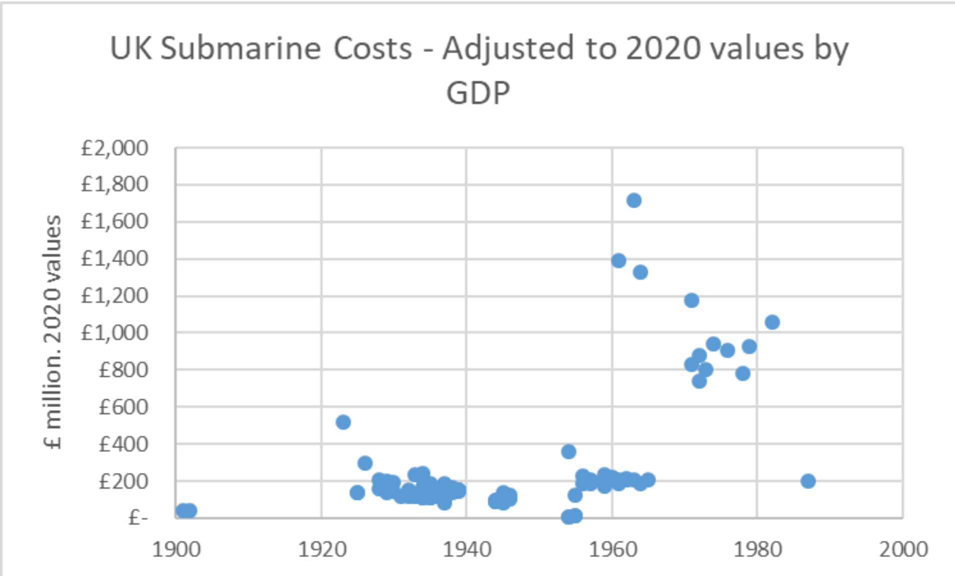

Figure 58 - UK Submarine Costs - Adjusted to 2020 values by GDP increase

Notes: Historic costs uplifted to 2020 values by rate of GDP increase. This results in the plotted cost of each submarine being the same proportion of GDP in 2020 as it was originally. Source – Relevant Statements on the Defence Estimates and Statements Explanatory of the Navy Estimates.

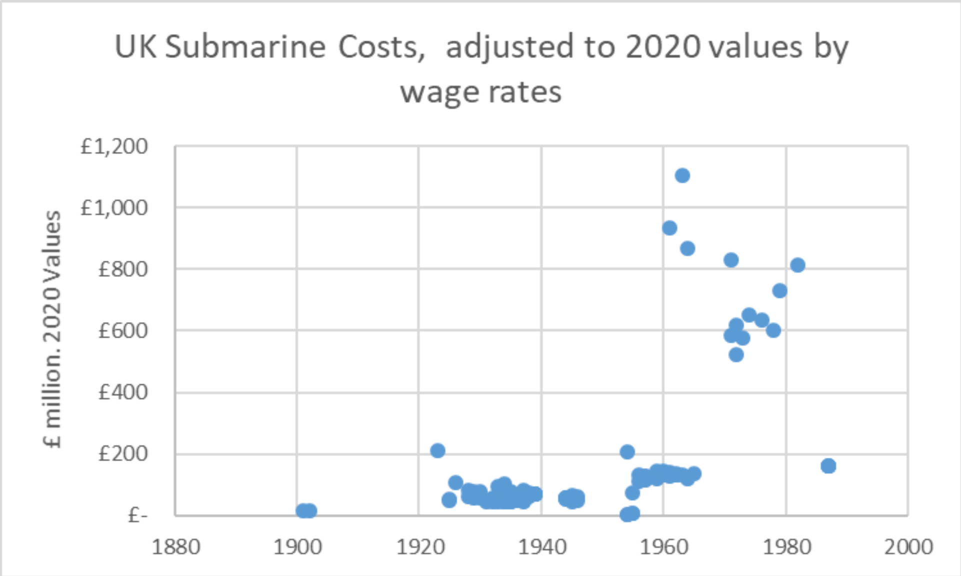

Figure 59 - UK Submarine Build Costs - Adjusted to 2020 by UK wage rates

Notes: Historic costs uplifted to 2020 values by rate of UK wage increase. This results in the plotted cost of each submarine representing the same number of labour hours in 2020 as it did originally. Source – Relevant Statements on the Defence Estimates and Statements Explanatory of the Navy Estimates.

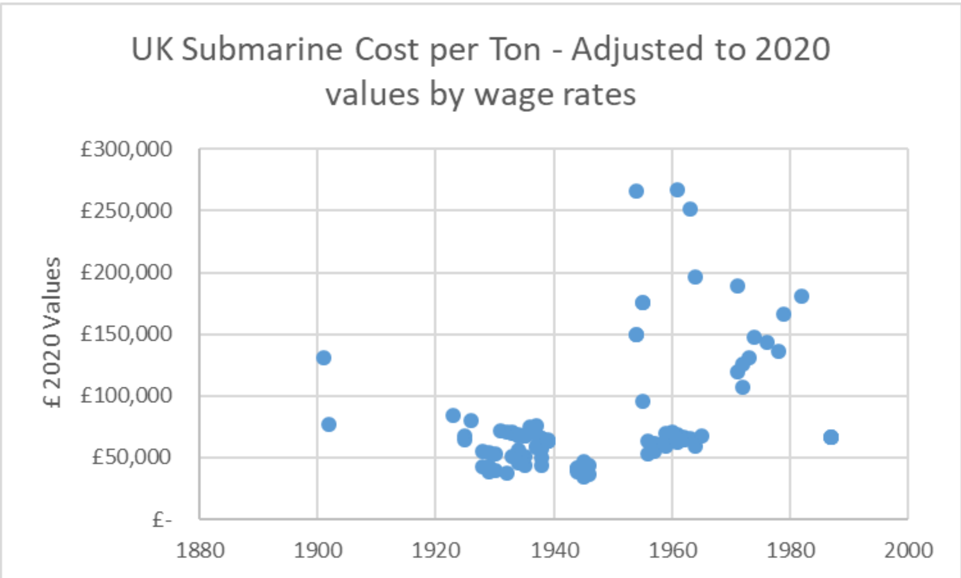

Figure 60 - UK Submarine Costs per Ton (2020 Rates)

Notes: Historic costs uplifted to 2020 values by rate of UK wage increase. This results in the plotted cost of each submarine representing the same number of labour hours in 2020 as it did originally. Source – Relevant Statements on the Defence Estimates and Statements Explanatory of the Navy Estimates.

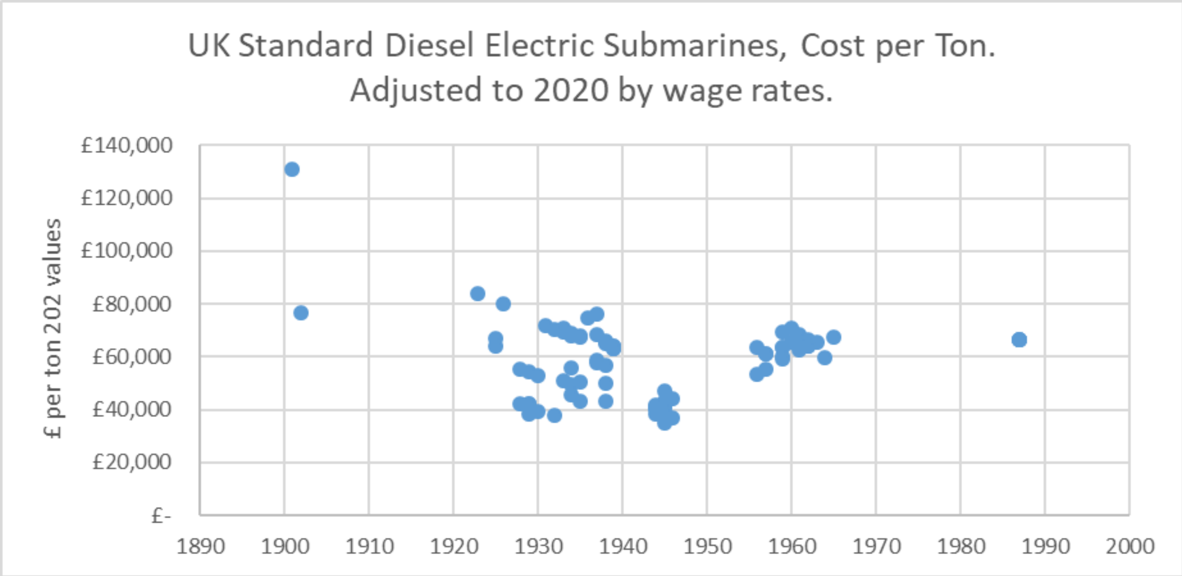

Figure 61 - UK Standard Diesel Electric Submarine Cost per Ton (2020 Rates)

Notes: Source – Relevant Statements on the Defence Estimates and Statements Explanatory of the Navy Estimates.

Another potential cause of variation in Figure 61 is that the eighty four submarines covered were spread across five different builders, four of which were in the private sector and one of which was a naval dockyard. If the averages for each class are plotted the variation is slightly reduced and we obtain the results shown in Figure 62. Excluding the first point (on the assumption that construction of the first submarine in the UK might not be representative of long-term averages) then we end up with a fairly constant set of values over time. Although having only 13 data points might be a challenge with respect to overfitting of a trend line, if a time related variation was apparent, in the current case, a (roughly) horizontal line just below the £60,000 per ton line appears a fair long-term average.

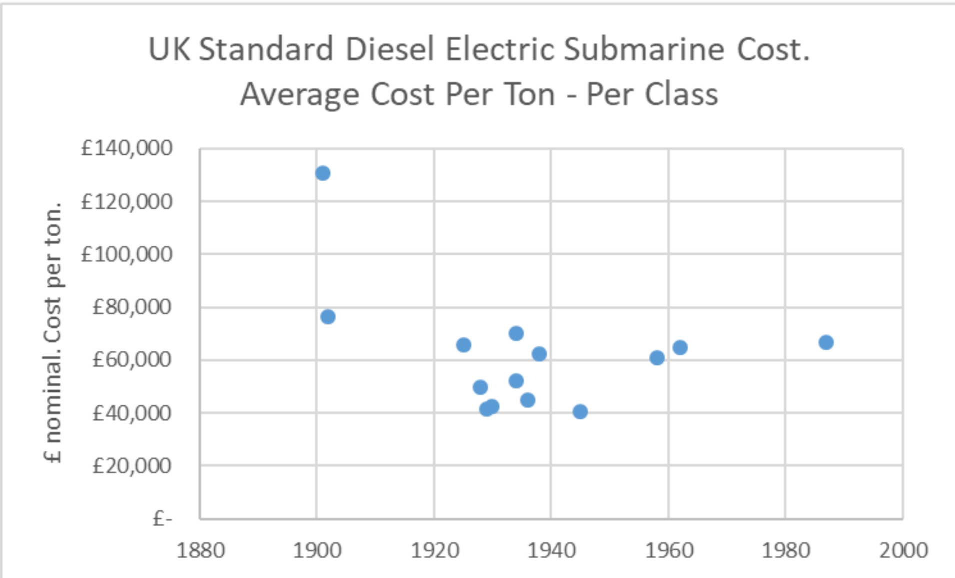

Figure 62 - UK Standard Diesel Electric Submarine Cost per Ton, Class Averages (2020 Rates)

Source – Relevant Statements on the Defence Estimates and Statements Explanatory of the Navy Estimates.

From this examination of Submarine build costs, we observe that, at constant labour rates, there is no time-trend evident, provided that we do only consider traditional diesel electric designs, however, once nuclear, hydrogen peroxide and novel designs are considered, then considerable variation is evident. The only trend that this variation reflects is that over time new technologies have been invented that enable the building of more expensive submarines that are more complex and are assumed to be more capable in relevant ways.

Having demonstrated that, once changes in labour rates are allowed for, the cost per ton for diesel electric submarines appears to be relatively constant, we may also consider how their affordability has changed over the years. We assess this by uplifting the historic cost per ton by the percentage of GDP that had to be expended in the year of manufacture to purchase it. This yields Figure 63, which illustrates that these submarines became more affordable up to about 1940, after which they appear to have reached a steady state of affordability.

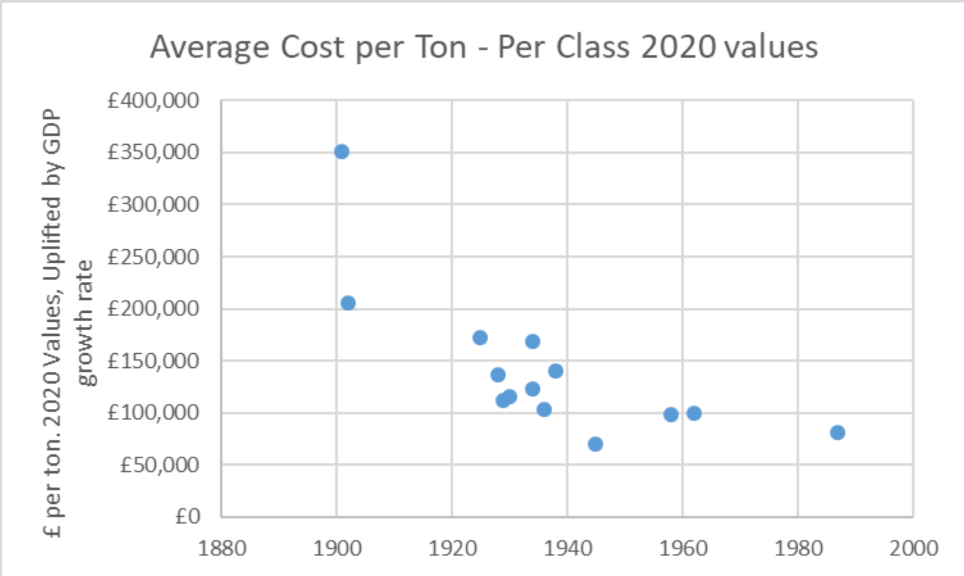

Figure 63 - UK Standard Diesel Electric Submarine Average Cost per Ton - Per Class (Relative Affordability)

Source – Relevant Statements on the Defence Estimates and Statements Explanatory of the Navy Estimates.

We should also note that it was not obvious at the outset that diesel electric submarines would from a coherent reference class. There have been many incremental and step changes in the equipment fits and levels of sophistication in these vessels as well as in the efficiency of ship building and the structure of the UK ship building industry over the period covered and so the presence of changes in the average cost per ton over time would have been completely feasible.

As an illustration of this point we may use the unit cost data quoted by Hartley (2023, Table 9) to examine the difference in unit production cost for different types of UK military aircraft. In Figure 64 the points labelled ‘B’ are bomber aircraft, with the others being fighters.

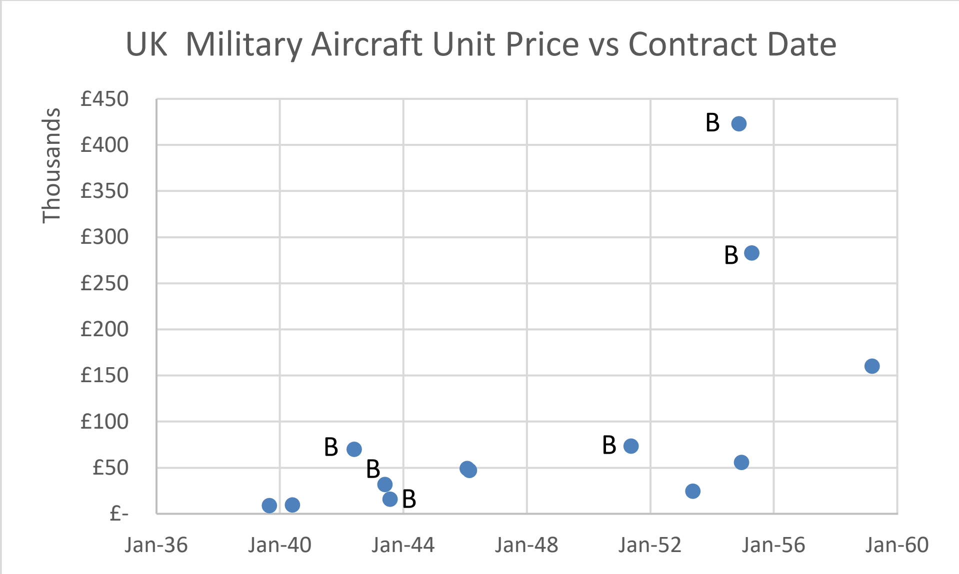

Figure 64 - UK Military Aircraft Unit Prices vs Contract Date (1959 economic conditions)

Notes: Points labelled 'B' represent bomber aircraft, those without labels are fighter aircraft. Source Hartley (2023 (Table 9))

This shows a very considerable variation in the unit price, but as bomber aircraft tend to be significantly larger than fighter aircraft this might be expected to drive some variation. This aspect is explored in Figure 65 where we observe significant variations in the cost per kg for both fighter and bomber aircraft.

This figure illustrates that for the data points before January 1944 a similar level of variation in the cost per unit weight is seen to that in the submarines in Figure 62. After this point all of the aircraft are jet powered, rather than propeller driven, and hence appear to be part of a different reference class. Additionally there appears to be more variation in cost for the jet powered aircraft, however, due to the small number of data points we cannot be sure whether this is intrinsic variation, whether it is due to the first instances of a new technology being more costly, or some other factor. This general infrequency of projects is a continual challenge in analysis of Defence Systems, also, as Systems become more expensive there is a greater chance of them being produced multi-nationally, which again may affect the homogeneity of the reference class. We also note that more recent Systems, that are regarded as being more capable and hence of higher quality, cost more in overall and in per unit weight terms.

Consideration of the production costs for submarines and aircraft shows that platforms embodying significantly different technologies should (probably) form different Reference Classes. Whilst the corollary of this (that projects with similar technologies can form a single reference class) is the basis for Reference Class Forecasting, it fails to answer the question of how similar a future project would need to be for us to be confident that it would fall into the same Reference Class. Whilst this may be immaterial for industries with well-tried approaches and little technological innovation (for example, many road building projects) it represents a significant issue for Defence where gaining military advantage through technological innovation is (often) a key theme. If we are not to have to fall back on the assertions of Augustine and of Webb that an underlying tide of technological innovation exists, that can be relied upon to progress at a constant rate and which will have a consistent effect on the cost of Defence Systems, then an alternative approach is required.

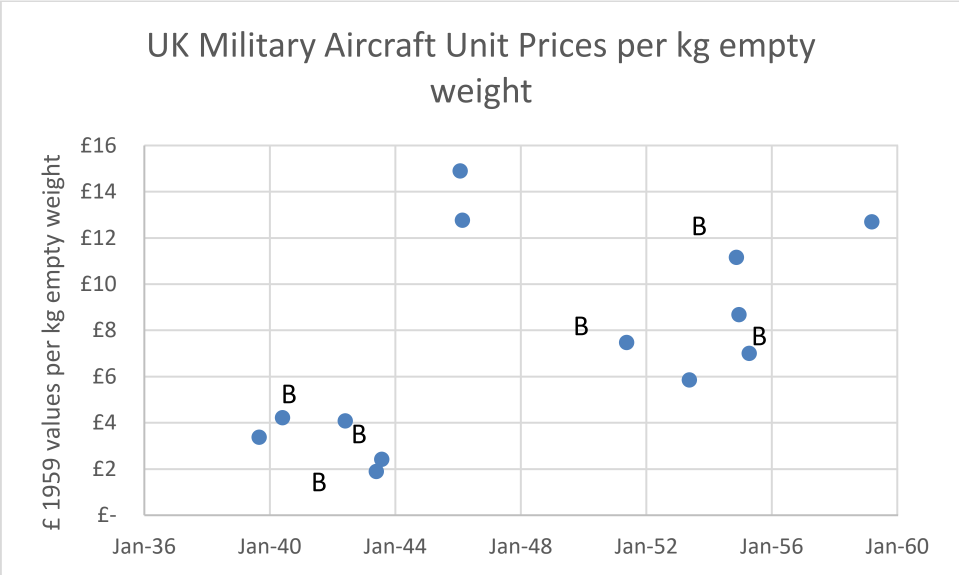

Figure 65 - UK Military Aircraft Unit Prices per kg vs Contract Date (1959 economic conditions)

Note: Data points marked with a ‘B’ represent bomber aircraft.

In the Education and Formula 1 case studies, we noted that a significant proportion of the cost increases were made up of increases in the number of staff and the cost of the staff, with a small proportion relating to the non-staff costs of running and maintaining facilities and also bought-in elements. If we apply this logic to the development cost of (say) a new military aircraft, then the project cost will, to a first approximation, be driven by the annual cost of running the project team and the number of years over which the team is employed. We now investigate whether an approach to forecasting the team size and project duration might be constructed that does not require a detailed assessment of the state of the potential technologies to be embodied in a future project.

7.6.3 - Defence Systems – Project Duration

As there is more accessible data on the duration of different Defence System projects we shall start by considering this aspect and examine the team size later. An unpublished report by DARPA contains an image (Figure 66), which was cited by Kozloski (2013) as providing evidence of how the burden of project governance and oversight is delaying military projects, when the same constraints do not apply to civil projects.

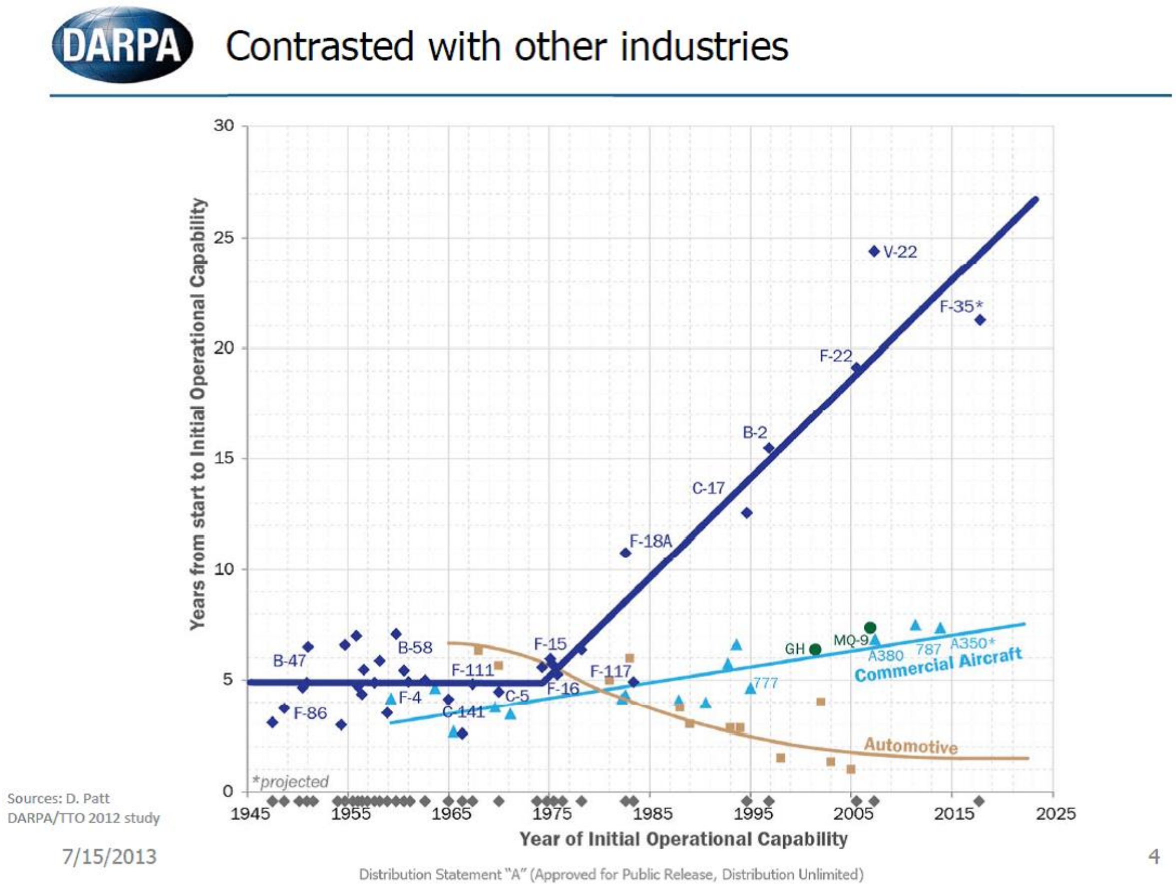

Figure 66 - DARPA Chart on development project durations.

Notes: Source – Reproduced in full from Kozloski (2013), who apparently reproduced it from a DARPA study. It has not been possible to identify the original source.

Whilst this might be a possible explanation, it is also worth considering what other factors have changed over this period illustrated. For example, might it be possible that recent aircraft projects such as the F-35, F-22 and V-22 contain more technical and design challenges and risks than earlier aircraft? We may be certain that the recent aircraft contain more electronics and software than the F-86 Sabre and all of them also have unusual design characteristics, related to their operational roles. Similarly could there be different factors in play that influence the development timescales of civil aircraft when compared with their military counterparts?

In order to further investigate changes in programme duration over time, data has been collected on 19 British post-war military fast jet and trainer aircraft, on 19 US post-war fast jet aircraft, 13 post-war European airliners and 14 post-war US airliners[137].

For each of these aircraft the following dates have been identified:

Date when initial design studies were commenced

[137] https://doi.org/10.15129/3fd130a0-be32-4509-98de-af5259f34b72

Date when commitment was made to finalise design details and produce aircraft

Date when platform entered service (Initial Operating Capability or equivalent)

Plotting the data for the military aircraft (Figure 67 and Figure 68) we observe that as Figure 66 suggests there has been significant growth in the duration of military aircraft projects, however, there is no evidence of a knee in the line, with a consistent increase appearing more likely. It should also be noted that the paucity of data points since 2000 makes it difficult to determine the current shape of the line.

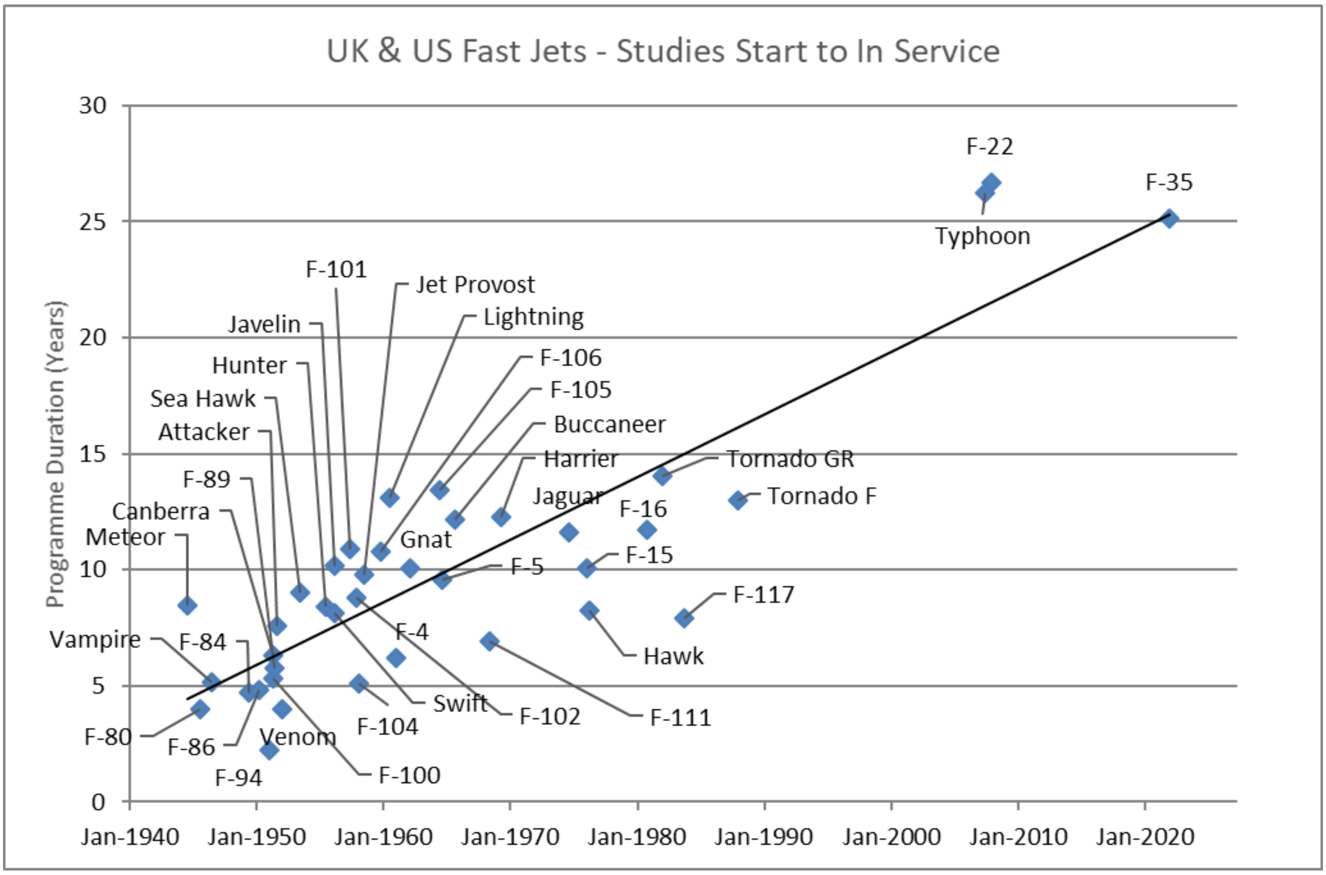

Figure 67 - UK & US Fast Jets & Jet Trainers - Studies Start - In Service Duration

Notes: Source data at - https://doi.org/10.15129/3fd130a0-be32-4509-98de-af5259f34b72

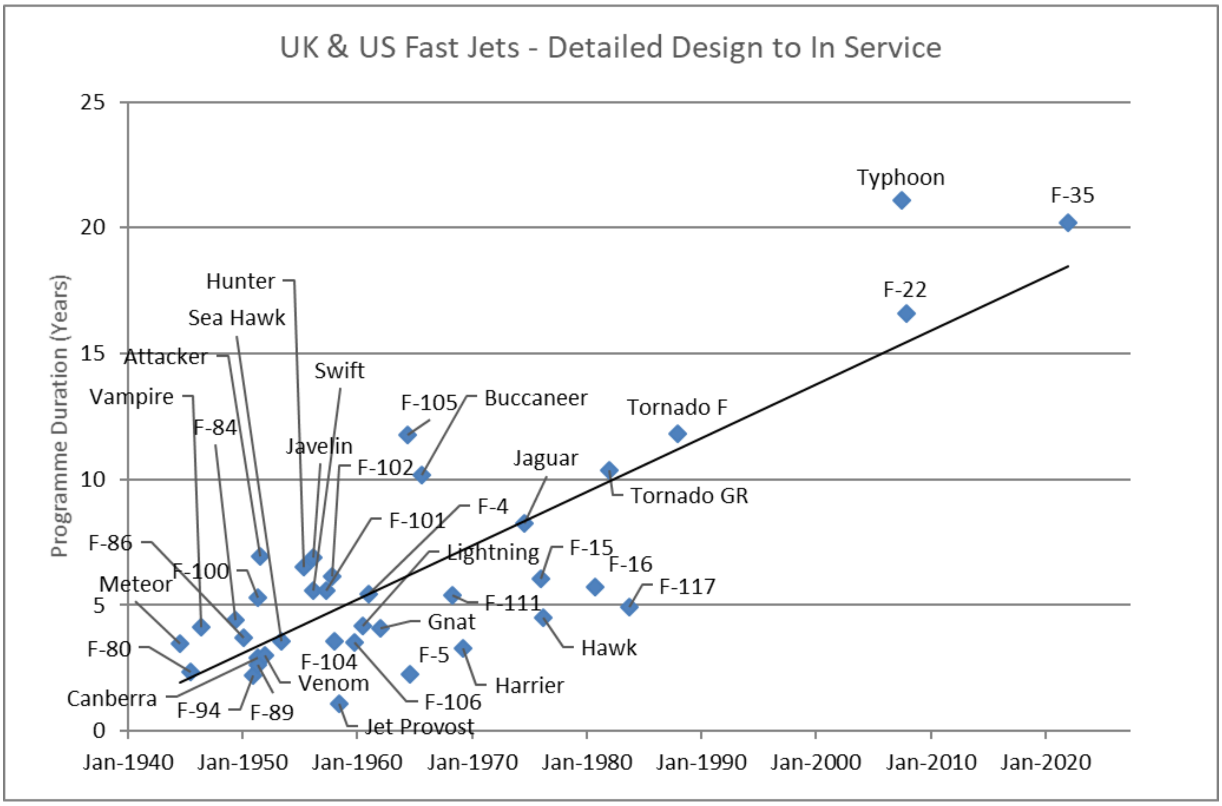

Figure 68 - UK & US Fast Jets and Jet Trainers - Detailed Design - In Service Duration

Notes: Source data at - https://doi.org/10.15129/3fd130a0-be32-4509-98de-af5259f34b72

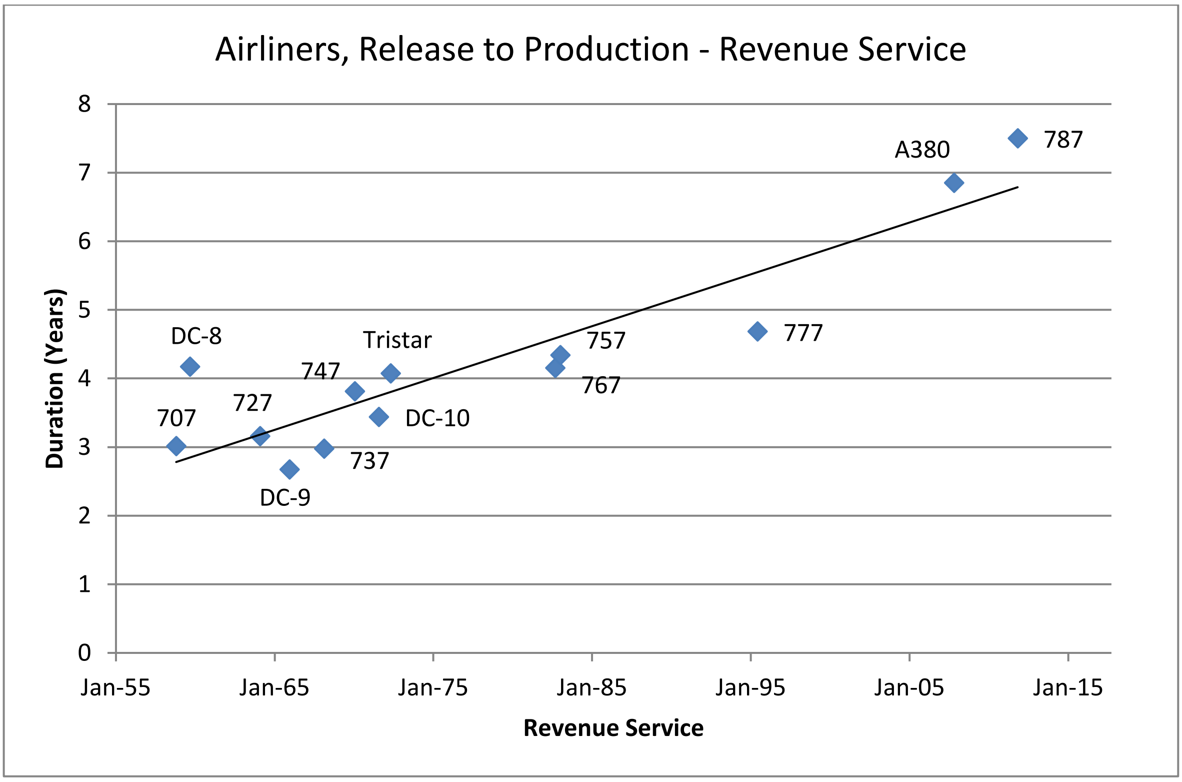

Figure 69 - Airliners Release to Production – Revenue Service Duration

Notes: Source data at - https://doi.org/10.15129/3fd130a0-be32-4509-98de-af5259f34b72

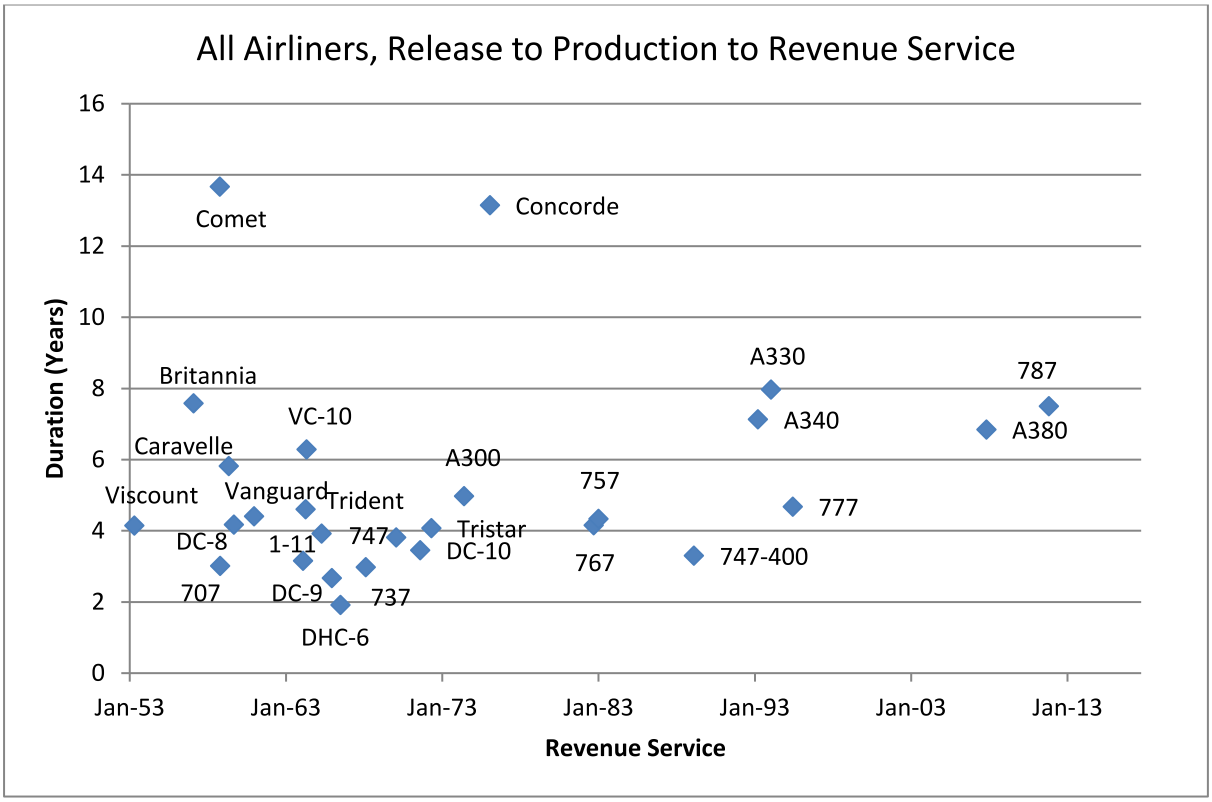

Figure 70 - All European and US Airliners. Release to Production - Revenue Service Duration

Notes: Source data at - https://doi.org/10.15129/3fd130a0-be32-4509-98de-af5259f34b72

If one assumes that the trend lines are broadly correct, this implies that the overall programme length has increased by about 2.5 years per decade and the detailed design and production schedule has increased by about 2 years per decade. In the unlikely event that this trend continued indefinitely, then one would, eventually, end up designing a military aircraft with no vision of what its role would be or what technology would be available to make it function. This would significantly increase programme risks.

Before digging further into the military aircraft projects it is worthwhile considering the trends in civil aircraft programme durations (Figure 69 and Figure 70). Figure 69 which covers post war US developed airliners (plus the A380 as it appears on the original DARPA graph) indicates an increase in programme length of about 9 months per decade, which is very similar to the value shown in Figure 66. However, airliners were also being developed in Europe over this period and Figure 70 also includes these additional projects. It is evident that this latter Figure shows significantly more variation in development durations and also that Comet and Concorde are significant outliers. The Comet was the first jet powered airliner and Concorde was the first supersonic airliner and both of these innovations proved to have significant programme challenges that took some time to resolve. This context, coupled with the presence of the 747-400 and DHC-6 at the other end of the time scale (both of which are a reworking of existing, proven designs) suggests that (unsurprisingly) the level of technological and / or design challenge in a new aircraft design could influence how long the development process will take.

One standard approach to assessing technology challenge is the Technology Readiness Scale, which was first conceived by NASA researcher Stan Sadin in 1974 (NASA (2010)) but was 'not formally defined until 1989' presumably in Sadin et al. (1989). The original seven point scale was designed to assist NASA with technology management whilst suffering 'budget limitations in the post-Apollo period' which 'played havoc with the new technology development process'. It allowed management to ensure that feasibility studies of candidate technologies were pursued to the required level, so their readiness for adoption into a live programme could be guaranteed when required.

Technology Readiness Levels (TRLs) can be an excellent way of assessing the current understanding of individual technology elements, however, as Olechowski et al (2020) indicate in their survey of perceived weaknesses of the traditional TRL approach, the lack of consideration of the challenges of integrating new technologies into highly complex systems via relevant interfaces can be a major issue (identified as Challenges A1 and A2). This is a particular issue for aircraft development, where each subsystem has to function within itself, to be integrated into a functioning whole aircraft system of systems and to achieve this without compromising the weight and power constraints necessary for flight. The authors also identify (Challenges B5 and B7) that TRLs do not address the risks inherent in progressing to higher TRLs, nor the likely time and effort required to achieve such goals. This latter shortfall is identified as being particularly relevant to the use of TRLs in planning and review.

The recognition of the integration issue was one of the drivers for the work of Sauser et al (2006), where the concept of a System Readiness Level (SRL) that incorporates both Technology Readiness Levels and Integration Readiness Levels (IRL) is explored. Although the concept is simple, the need to generate an overall SRL by building up from the TRL and IRL for individual components makes it more suitable for a detailed bottom up approach, rather than a broad-brush assessment of future programmes. Alexander (2017) agrees, writing that in the 'early conceptual stages of a development project, design and performance information typically applied in traditional parametric cost and schedule models is usually very limited. Key attributes of such models often focus on subsystem- or unit/assembly-level characteristics or performance metrics that have not yet been determined in these preliminary stages'.

Searching for better approaches, Alexander used the NASA Technology Cost and Schedule Estimating (TCASE) tool as the basis for his study, as it contains records from over 2,900 projects from 'fourteen wide-ranging technology areas and a broad scope of applications and systems that are relevant across the scientific, military and intelligence sectors'. The tool had been established 'partially in response to the NASA cost community's findings from the 2011 Cost Symposium' which concluded that 'there is no known good method to estimate the cost of Technology Readiness Level advancement that is supported by actual data' and its development is outlined by Cole et al (2013). The diversity of projects covered in TCASE meant that Alexander was limited in the number of parameters where sufficient data was available to undertake reliable modelling. Despite this he was able to produce relationships between project cost, project duration, System Hierarchy level and the TRL Improvement level[138].

NASA conducts research and development at a number of different system levels and so in TCASE each project is identified according to its System Hierarchy level, which is defined as follows (Cole et al (2013, Table 3))

No.	Tier	Definition	Example
1	System	An integrated set of constituent elements that are combined in an operational or support environment to accomplish a defined objective	A spacecraft or launch vehicle stage
2	Subsystem	A portion of a system	A satellite's propulsion system or a launch vehicle's propulsion system
3	Assembly	A set of components (as a unit) before they are installed to make a final product	A satellite's thruster or launch vehicle's engine turbo-machinery
4	Component / Part	A portion of an assembly	A satellite's propellant valve or a launch vehicle's engine injector
5	Hardware / Material	An item or substance used to for a component	Alloy, polymer, screws, bolts, pipes, semiconductor chips

Table 7.21 - NASA System Hierarchy Definitions

Although System Hierarchy is a potentially important discriminator for studying technology projects in general, in this current case the situation is simplified, as we are only interested in System Hierarchy Level 1 – that of the whole aircraft system.

Although not used by Alexander the NASA model also includes top down and bottom up assessments of the difficulty likely to be encountered over the course of a technology maturation project and both scales have potentially relevant content. The top-down approach identifies the Research and Development Degree of Difficulty (R&D3) and is reproduced below as Table 7.22.

R&D3	Definition	Probability of Success
1	A very low degree of difficulty is anticipated in achieving research and	≥ 95% - 99%

[138] Simply a measure of the number of TRLs that the project advanced the technology.

	development objectives for this technology.	
2	A moderate degree of difficulty should be anticipated in achieving R&D objectives for this technology.	≥ 90%
3	A high degree of difficulty anticipated in achieving R&D objectives for this technology.	≥ 80%
4	A very high degree of difficulty anticipated in achieving R&D objectives for this technology.	~ 50 – 60%
5	The degree of difficulty anticipated in achieving R&D objectives for this technology is so high that a fundamental breakthrough is required.	≤ 30%

Table 7.22 - NASA Research and Development Degree of Difficulty definitions.

The Advancement Degree of Difficulty (AD2) is used by NASA to make a bottom up assessment of individual projects to assess how confident those responsible for planning the project are that the proposed approach(es) will generate successful outcomes. Where the path is well defined and low risk, then there can be high confidence in a successful outcome in a timely fashion. However, where greater uncertainty exists, then it may be necessary to pursue multiple approaches either in tandem or in sequence. Table 7.23 shows how NASA has chosen to structure this scale and although not all of the elements are relevant to the assessment of whole aircraft development programmes (e.g. generally such a programme would not be started with only a 20% chance of success), the concepts of capturing the level of uncertainty in the development path might well be relevant to determining the duration of aircraft development programmes.

AD2	Definition	Risk	Category	Success Chance
1	Exists with no or only minor modification being required. A single development approach is adequate.	0%		Guaranteed Success
2	Exists but requires major modifications. A single development approach is adequate.	10%		
3	Requires new development well within the experience base. A single development approach is adequate.	20%		
4	Requires new development but similarity to existing experience is sufficient to warrant comparison across the board. As single development	30%	Well Understood (Variation)	Almost Certain Success

	approach can be taken with a high degree of confidence.			
5	Requires new development but similarity to existing experience is sufficient to warrant comparison in all critical areas. Dual development approaches should be pursued to provide a high degree of confidence for success.	40%	Known Unknowns	Probably will Succeed
6	Requires new development but similarity to existing experience is sufficient to warrant comparison only on a subset of critical areas. Dual development approaches should be pursued in order to achieve a moderate degree of confidence for success. (Desired performance can be achieved in subsequent block upgrades with high confidence).	50%		
7	Requires new development but similarity to existing experience is sufficient to warrant comparison in only a subset of critical areas. Multiple development routes must be pursued.	70%		
8	Requires new development where similarity to existing experience base can be defined only in the broadest sense. Multiple development routes must be pursued.	80%	Unknown unknowns	High Likelihood of Failure (High Reward)
9	Requires new development outside of any existing experience base. No viable approaches exist that can be pursued with any degree of confidence. Basic research in key areas needed before feasible approaches can be defined.	100%	Chaos	Almost Certain Failure (Very High Reward)

Table 7.23 - NASA Advancement Degree of Difficulty definitions

For their top down approach (R&D3) NASA adopted a five point rating, and that scale has been followed for the construction of novel Challenge Ratings for the Technology and Design elements of an aircraft development programme. The two scales build on the elements of R&D3 and AD2 respectively, with the aim of capturing the level of uncertainty about the eventual paths to technology maturation and to design integration and programme delivery that existed at the start of the programme.

All of the projects assessed by the NASA process were undertaken by NASA, who are world leaders in their field. For this study we recognise that different aircraft are developed and built by different contractors. The definitions for this new Challenge Rating approach therefore make reference to company or site specific experience, as well as that applying to the Industry in general. The Challenge Ratings have also, specifically, been designed so that they may be assessed before detailed work on a project has started and hence are more top-level than the specific TRL assessments undertaken by NASA.

Challenge Rating	Technology Challenge	Design Challenge
1	This company (site) have successfully fielded a system using this technology in this domain	A straight forward modification of an existing aircraft to accommodate a new engine, radar etc.
2	This technology is well proven and understood by this contractor, however there is work to do in repackaging, scaling (etc.)	A new aircraft design with no / few design challenges, alternatively a more involved modification programme.
3	All the relevant technologies have previously been demonstrated. But not in this domain, so significant risks could remain.	A new aircraft design, which generally incorporates proven technology and design elements, but with one novel / challenging aspect.
4	Whilst significant elements of the technology are novel and/or challenging, there is a relatively good understanding of potential paths to address likely risks.	This aircraft is novel to this design organisation. Although all the principles have been demonstrated elsewhere, challenges remain in making it a functional design.
5	There is theoretical evidence that the technology will work. However, there is uncertainty over whether / how it can be made to work acceptably in practice.	This new aircraft is novel in significant structural, operational and / or aerodynamic ways.

Table 7.24 - Technology Challenge and Design Challenge Scales

For each of the data points shown in Figure 67 to Figure 70 an assessment of their Technical Challenge and Design Challenge has been made. Where relevant, the values at the start of the programme and at the point where the decision to commit to detailed design work and manufacture were separately assessed. Figure 67 and Figure 68 do not separately identify each of the aircraft covered, so the complete list appears in Table 7.25.

Gloster Meteor	Hawker Harrier	Convair F-102
De Havilland Vampire	SEPECAT Jaguar	Lockheed F-104
Supermarine Attacker	Panavia Tornado GR	Republic F-105
Hawker Sea Hawk	Panavia Tornado F	Convair F-106
English Electric Canberra	Hawker Siddeley Hawk	General Dynamics F-111
Hawker Hunter	Eurofighter Typhoon	McDonnell Douglas F-4
Gloster Javelin	Lockheed F-80	Northrop F-5
Supermarine Swift	Republic F-84	McDonnell Douglas F-15
BAC Jet Provost	North American F-86	General Dynamics F-16
Folland Gnat	Northrop F-89	Lockheed F-117
Blackburn Buccaneer	Lockheed F-94	Lockheed Martin F-22
English Electric Lightning	North American F-100	Lockheed Martin F-35

Table 7.25 - List of UK and US Military Aircraft studied

Gaining broad agreement on the Technology and Design Challenge Ratings for each of these projects at the start of the project and also when the decision was made to finalise the design and proceed to manufacture could be a very time consuming task. However, at present the main aim is to assess whether this approach appears to give broadly correct results and so an initial assessment of all the military and civil aircraft programmes was made by the author and reviewed by an expert colleague (G Bishop, personal communication, March 2017). The review showed up two typographical errors, but otherwise did not disagree with the preliminary values assigned[139].

Returning to the airliners covered in Figure 70 we now replot the same points according to the sum[140] of the Technology and Design Challenge Ratings, rather than the entry into service date. This yields the results shown in Figure 71, where we observe that there is a clear trend for the more challenging airliners to have taken longer to deliver. During the process of allocating the Challenge Ratings to the different aircraft there was debate about the A330 and A340 aircraft, as these were designed in tandem, to maximise commonality and also to reduce development costs. It was not clear whether, compared with two separate projects, this move would have increased the Challenge Ratings of these two aircraft, left it the same or actually reduced it. In this case the ratings were left the same as for two independent programmes, however, the results suggest that the Challenge Rating should have been increased.

Figure 72 and Figure 73 illustrate that there may be some differences between the US and European context for airliners, as a slightly better correlation is achieved by separating the two sets of data points. This difference may be coincidental, or may be due to a range of factors, including programme managerial and organisational complexity (rather than just technical complexity). For example, European governments have tended to be more hands-on

[139] This data is available at https://doi.org/10.15129/3fd130a0-be32-4509-98de-af5259f34b72
[140] It has been assumed that each programme covered was well managed, and hence delays in technology development were resolved before proceeding to detailed design work. In this way it is assumed that technology and design challenge ratings add separately to the overall duration. In cases where technology development and detailed design activities are undertaken in parallel a multiplicative approach might be more relevant.

in their promotion of and involvement in aircraft development, whereas, until the Boeing 787 programme, US efforts tended to be more straight-forward to manage.

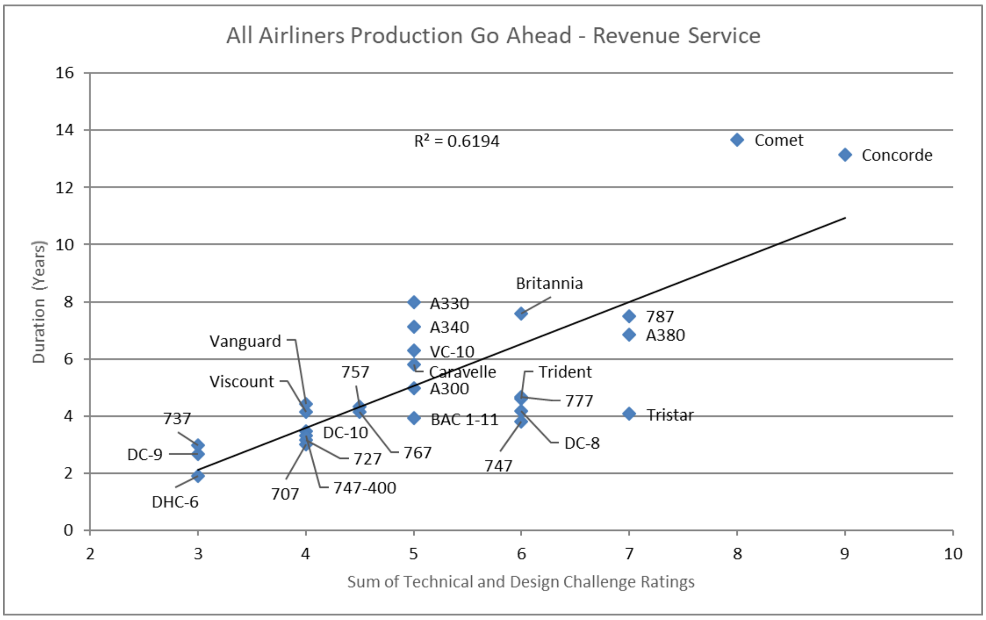

Figure 71 - All Airliners Production Go Ahead to Revenue Service Duration vs Sum of Technical and Design Challenge

Notes: Technical and Design Challenge Rating as defined in Table 7.24

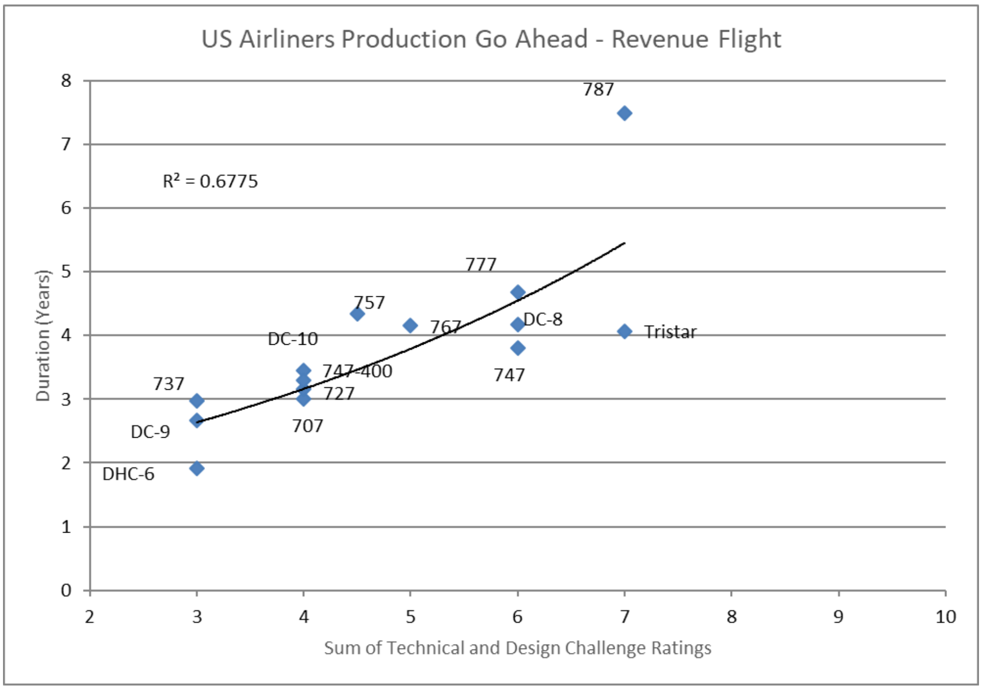

Figure 72 - US Airliners Production Go Ahead - Revenue Service Duration vs Sum of Technical and Design Challenge

Notes: Technical and Design Challenge Rating as defined in Table 7.24

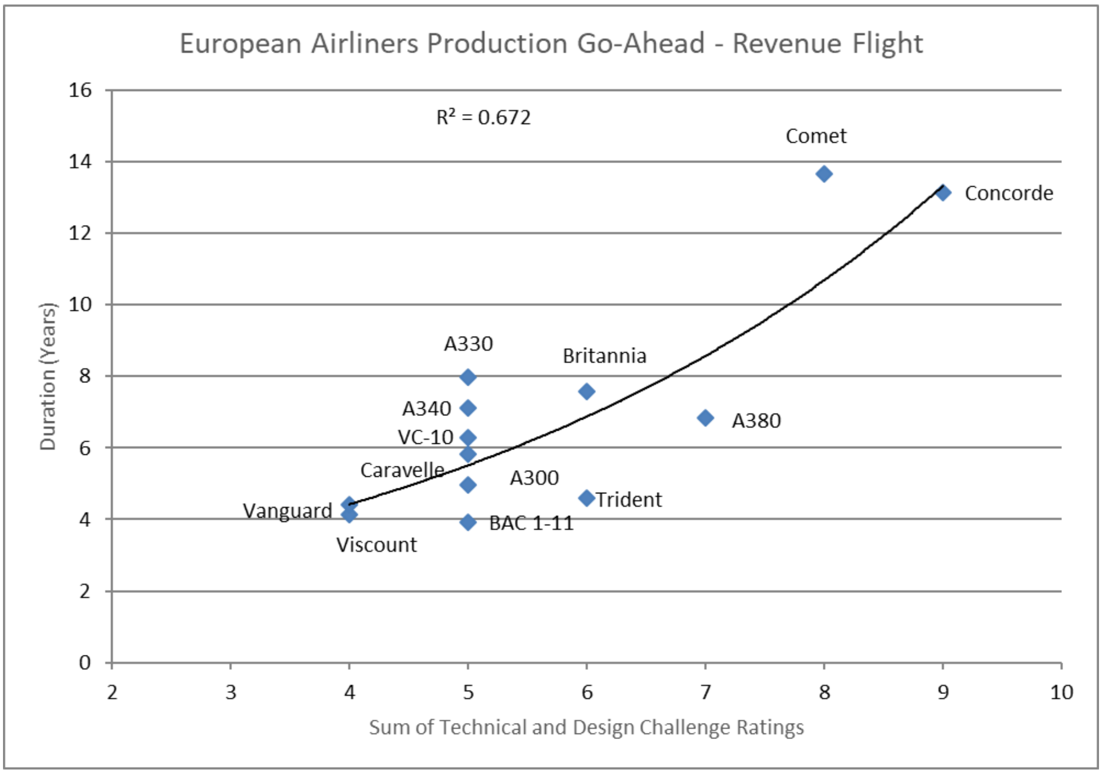

Figure 73 - European Airliners Production Go Ahead - Revenue Service Duration v Sum of Technical and Design Challenge

Notes: Technical and Design Challenge Rating as defined in Table 7.24

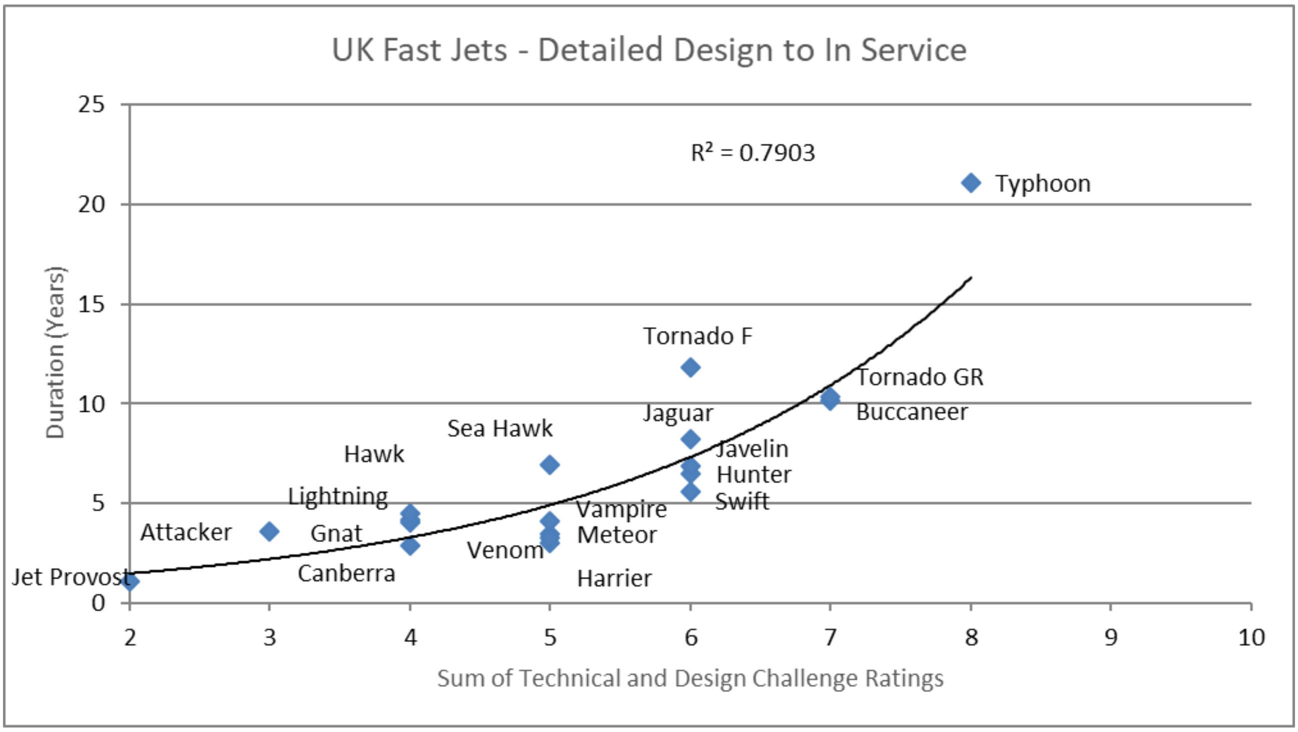

Figure 74 - UK Military Aircraft Detailed Design - In Service Duration

Notes: Technical and Design Challenge Ratings as defined in Table 7.24

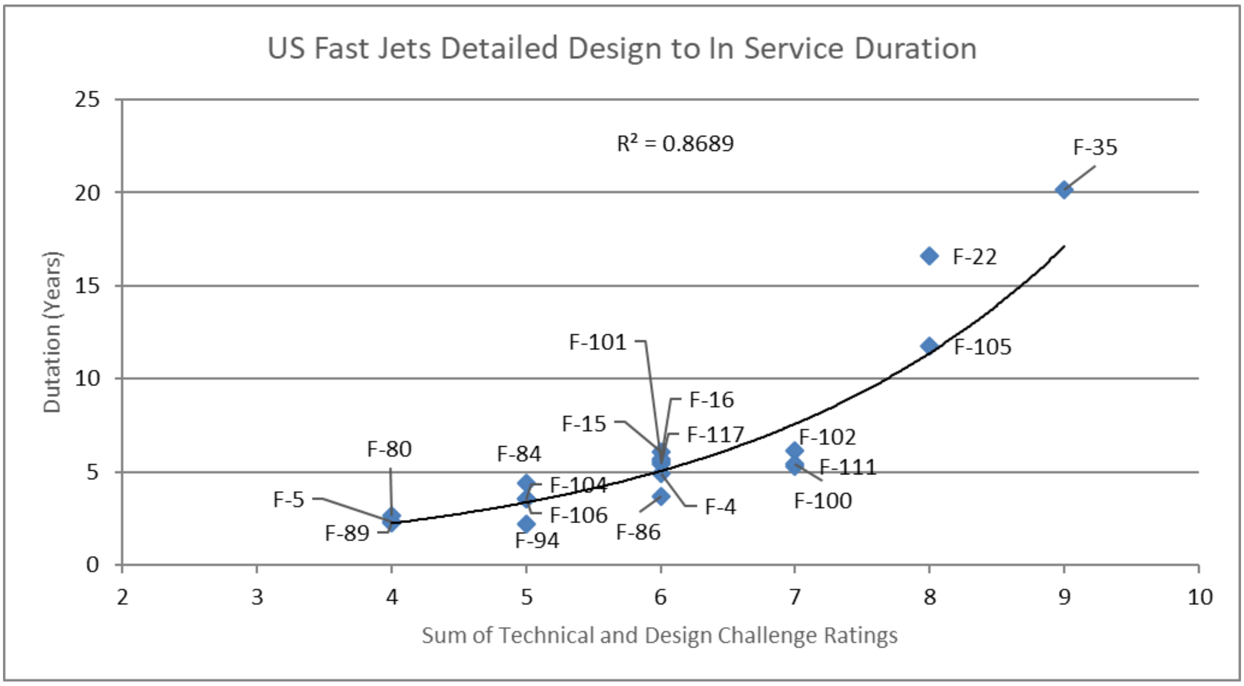

Figure 75 - US Military Aircraft Detailed Design - In Service Duration

Notes: Technical and Design Challenge Ratings as defined in Table 7.24

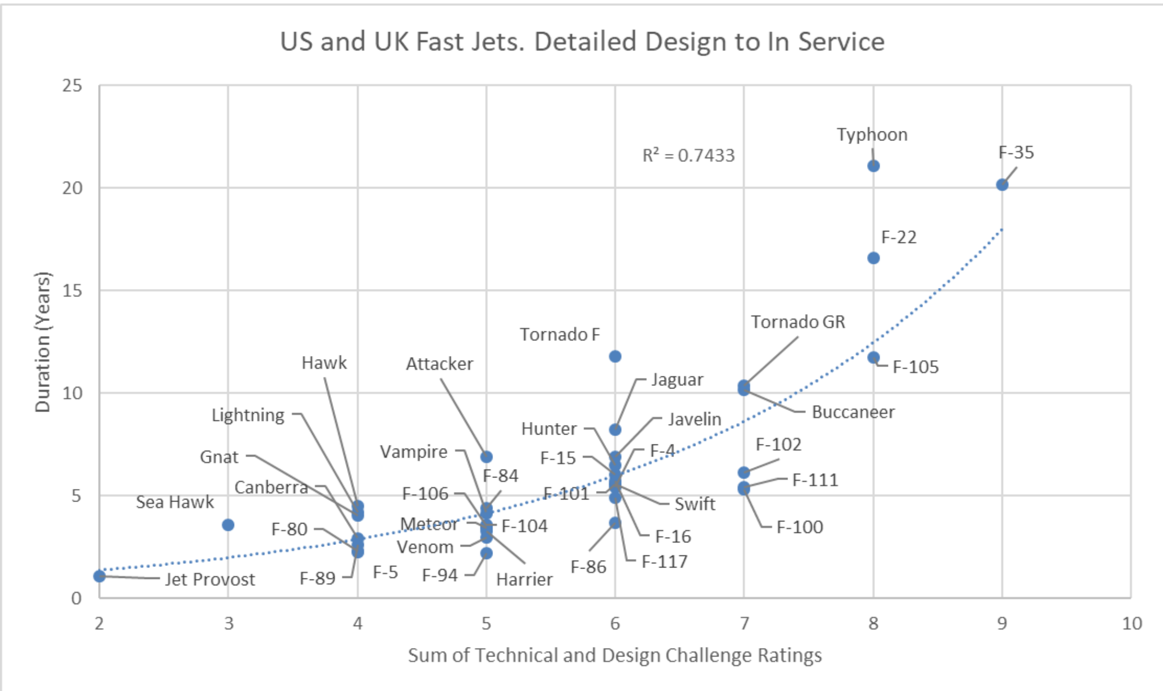

Figure 76 - All Military Aircraft Detailed Design - In Service Duration

Notes: Technical and Design Challenge Ratings as defined in Table 7.24

Having observed that this approach appears to give sensible results and potentially useful insights for civil airliners, we now come to examine whether this approach gives useful insights in the case of military aircraft programme durations.

Figure 74 and Figure 75 show good correlation between the sum of the technical and design Challenge Ratings for different military aircraft programmes and the programme duration, when UK and US programmes are separated. However, Figure 76 shows that the correlation for all programmes is slightly less good and for more complex projects there is a tendency for UK projects to be above the trend line. This suggests that, as with the Civil Airliners, there may be environmental or contextual factors in play which have different impacts depending on the nation owning the programme.

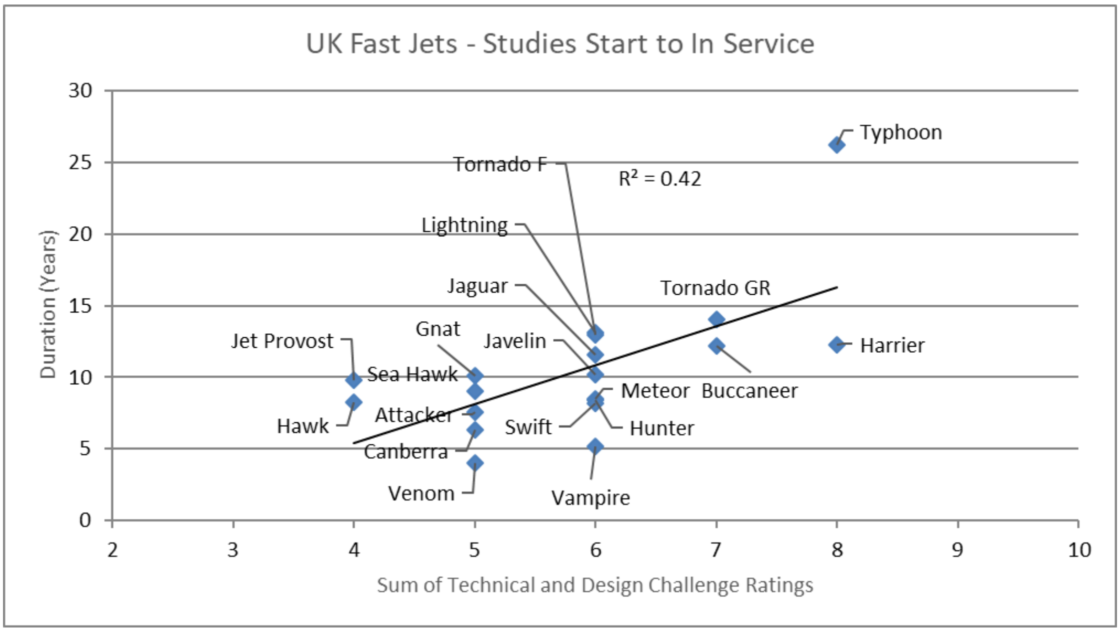

Figure 77 - UK Military Aircraft Studies Start - In Service Duration

Notes: Technical and Design Challenge Ratings as defined in Table 7.24

If one considers the duration of the whole programme, from when initial technology studies started, throught to In Service, then we find the results at Figure 77. A significant increase in the variability in durations is observed here, some of which is due to protracted technology development activties, for example, on the Lightning and the Harrier these each took nine years, but were very successful in derisking the final product. However, the early phases of a programme will also include the development and agreement of the requirements that the platform is to meet and other management activities, which can be very time consuming, particularly for multi-national programmes. The Challenge Rating scale does not currently capture uncertainty from non-technical sources and so, as we observe, may require an additional mangement / programme context dimension to improve forecasts of the total programme duration.

Having created and tested a promising approach for forecasting the duration of an aircraft development programme from making the decision to complete detailed design work through manufacturing to initial entry into service, we now consider how the size of the project teams (specifically the design office staff) have changed over time.

7.6.4 - Development Effort

It has been possible to source data on the design effort invested in past programmes undertaken by some UK companies[141]. Best efforts have been made to collect the design and development effort, without including the effort invested in atually constructing prototype aircraft. For most of the platforms the hours shown are the total to first flight, however for P1A & B and also for the Tornado ADV the totals include the flight testing and related support, but not the manufacturing hours. This data comes from company sources (F Berry, Personal Communication (August 2023)).

Name	Start	End	Total MH	Total Adj	FTE Years[142]
Canberra	1 Jan 1945	13 May 1949	755,000	755,000	503.2
P1.A & B[143]	1 Sep 1948	31 Dec 1959	3,127,000	3,127,000	2084.7
Jet Provost Mk1	1 Oct 1952	26 Jun 1954	165,000	165,000	109.7
Jaguar	1 Apr 1965	12 Oct 1969	1,987,000	1,420,000[144]	946.4
Tornado IDS	1 May 1970	1 Aug 1974	14,492,000	8,051,000[145]	5520.8
Tornado ADV	1 Jan 1975	27 Oct 1979	10,259,000	5,699,000	3908.0

Table 7.26 - Design Man Hours for various UK Aircraft Programmes (rounded to nearest thousand)

Notes: 'Total MH' represents the total Man Hours of design effort recorded against the project. 'Total Adj' represents the total Man Hours, adjusted to the value that would have been incurred on a UK only programme (See following text for the basis of this calculation). 'FTE Years' is the translation of the 'Total Adj' value into Full Time Equivalent (FTE) Years of effort, assuming 1500 effective working hours per annum. Data Source – BAE Systems Internal Records, via F Berry

The data in Table 7.26 has been adjusted where indicated to allow for the inefficiencies that a multi-national project introduces imposes, when compared with a national project (see Hartley and Braddon (2014)) for a discussion of the nature of these inefficiencies. These authors identify the 'square root rule' as being one approach to estimating the impact of additional partners, which means that the total effort involved in a two nation programme will be √2 great than the effort for a similar single nation programme. Similarly the total effort involved in a three nation programme will be √3 greater than for a single nation programme. The impact of this in Table 7.26 is that the bi-national Jaguar programme total is adjusted downwards by a factor of 1.4 (approximately[146] √2) and that for the two Tornado programmes by 1.75 (approximately √3). In addition to the data in this table, we also have different Challenge Ratings for each programme, ranging from 2 for the Jet Provost to 7 for the Tornado IDS and also different weights (larger aircraft tend to have more space for the fitting of complex sub systems).

As we might expect, with so many different potential dimensions and data from only six aircraft development projects, it is hard to draw any firm conclusions. For example, Figure 78

[141] All of which are now part of BAE Systems.
[142] Assuming 1500 productive hours per annum.
[143] Prototypes for the English Electric Lightning
[144] Assuming that a bi-national project consumes x1.4 as much effort as a national project with the same scope
[145] Assuming that a tri-national project consumes x1.8 as much effort as a national project with the same scope
[146] Approximate values are used as the original research only provides approximate values

illustrates how the design effort in notional Man Years has changed over time. However these data points represent diverse programmes with the last three being multi-national.

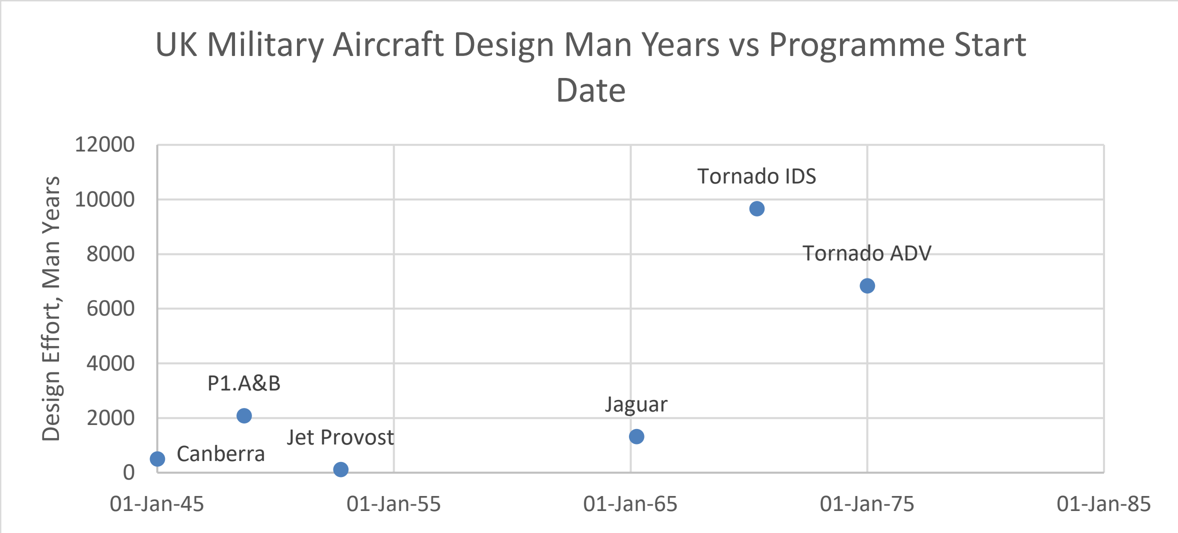

Figure 78 - UK Military Aircraft Design Man Years vs Programme Start Date

Note: Data Source Table 7.26

If we normalise the design effort against the weight of the platform and also adjust the values for multi-national programmes down to the level of an equivalent national programme, we then reach Figure 79. This broadly suggests that the effort per ton has been increasing over time, however, this might be caused by the effect that we observed in Figure 66 where more recent projects had a higher Challenge Rating, which might be expected to require greater design effort. Figure 80 confirms that as the Challenge Rating of projects increases the required design effort per ton also increases.

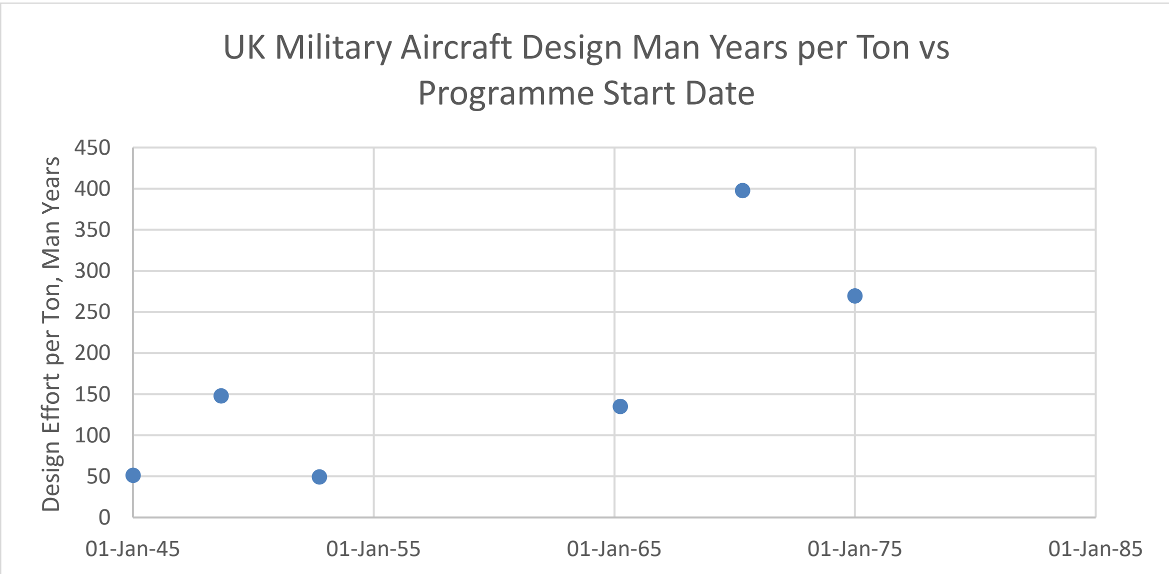

Figure 79 - UK Military Aircraft Design Man Years per Ton vs Programme Start Date

Note: Data Source Table 7.26

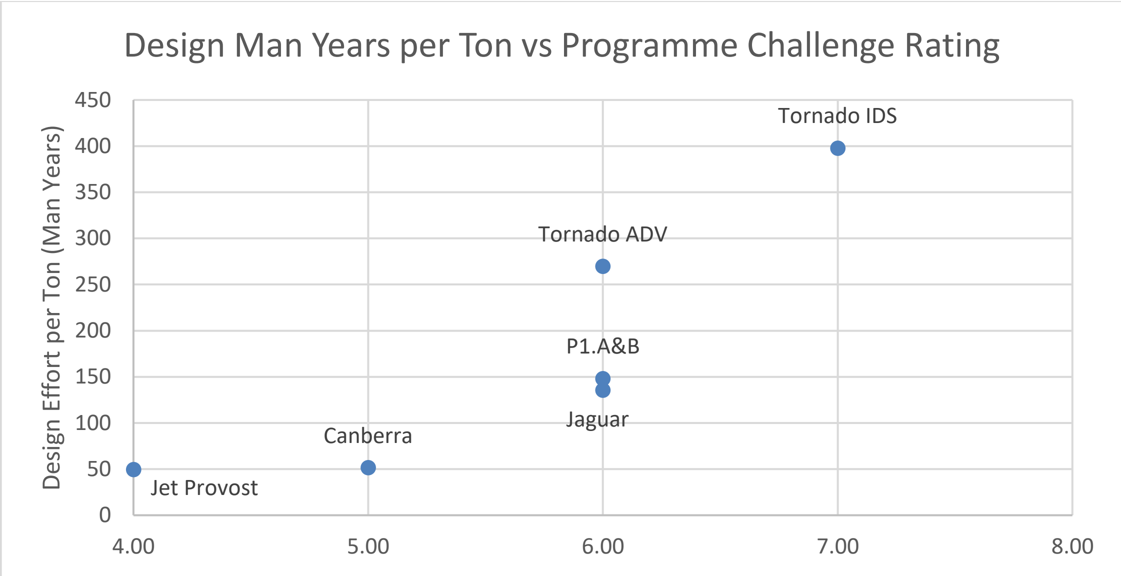

Figure 80 - Design Man Years per Ton vs Programme Challenge Rating

Note: Data Source Table 7.26

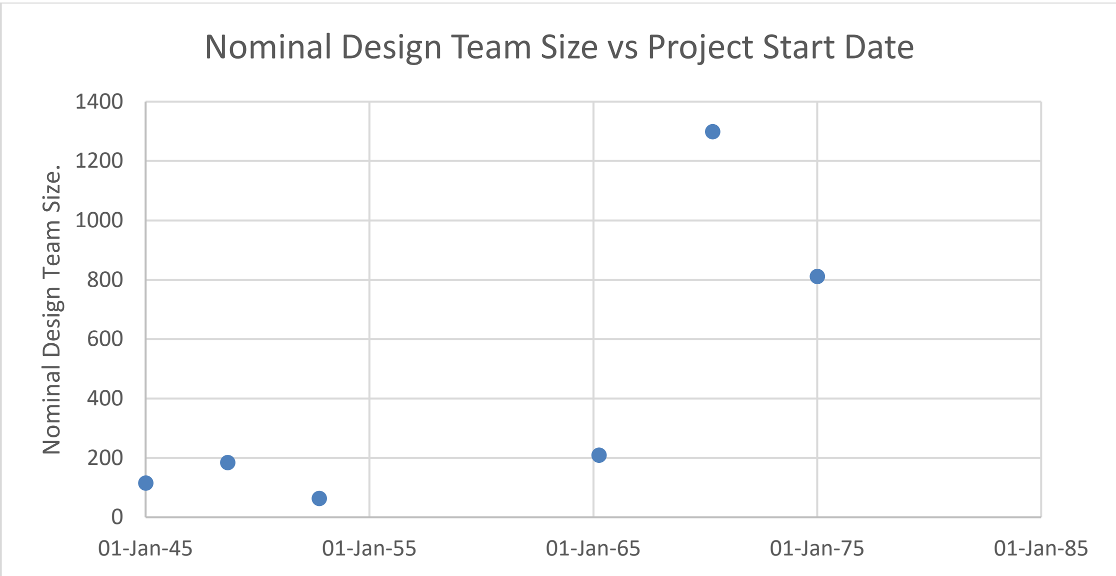

Figure 81 - UK Military Aircraft Nominal Design Team Size vs Project Start Date

Note: Nominal Design Team Size calculated by dividing the Man Years of Effort by the Programme duration. Data Source Table 7.26

Figure 81 illustrates the increase in the nominal[147] size of the design team across the subject projects. As expected, as the design effort required increases, the size of the team increases to deliver the required effort within the timescales desired by the customer.

[147] Assuming each design team member books 1,500 hours per annum to the development programme.

With the data that has currently been released it is not possible to assess what factors have driven the increased size of the design teams. Data does exist that identifies the elements into which the design of these platforms was broken down, and so further work should allow determination of how far this increase was due to specific work elements becoming more complex (and hence labour intensive) and how far the increase was due to an increase in the number of work elements as additional systems and technologies are embodied into a design. Such a task is likely to be rather time consuming, as ensuring, for example, that the scope of work elements was consistent across different programmes, from potentially different manufacturers, is likely to be rather involved. However, aerospace archives exist that do contain much of the necessary material to conduct such further analysis.

Another consideration that must be investigated is that, on occasion, advances in enabling technologies may have significant effects on the direct and indirect manpower required. For example, the introduction of Computer Aided Design systems eliminated the need for the Tracing Department and the Mould Loft and meant that less design effort should be required to deliver a given level of design maturity.

7.6.5 - Defence Systems Case Study Conclusions

From the study of Diesel Electric submarines and some of the aircraft, we have observed that the use of a Reference Class to support forecasting the likely production cost of a future System can be a very effective solution (addressing Question 2 from Table 1). However, unless two systems share similar technologies and development contexts it may not be clear in advance whether they will belong to the same Reference Class. Whilst this appears not to be a significant issue for many civil engineering projects, it is likely to be a challenge for areas where the rate of technological advance is high. This is a particular challenge for new Defence Systems, where there is often a desire for cutting edge technological innovation to be introduced to provide military advantage. Thus, although we are clear on how to compare the costs of different Defence Systems (Table 1, Subtopic 2.1) challenges still exist in comparing quality and normalising for contextual changes (Subtopics 2.2 and 2.3).

Having devised the concept of Challenge Ratings, extending work by NASA, we have also discovered solid evidence that indicates that (perhaps unsurprisingly) civil and military aircraft projects that have a more demanding technology and design Challenge Rating will take longer than those which are less demanding. We observed that for a given size of aircraft, those with a higher Challenge Rating will tend to require more design effort (per ton) than those which are simpler. However, the very limited data set does not allow for unequivocal conclusions to be drawn.

Whilst it is difficult to assemble data that covers all aspects of a Defence System project, based on the author’s experience of aerospace archives, it appears likely that further detailed work could provide insights into the drivers of the increasing size of aircraft design teams. This would provide insights into the context in which different Systems were delivered (Table 1, Subtopic 1.2).

We also observed that assessing the war fighting quality of Defence Systems there were similarities with the Formula 1 case study, although wars occur less often than Grands Prix and are statistically less well controlled for different variables. We also observed that assessing the deterrent quality of a defence system is similar to assessing the quality of an independent school education, in that it is perceptions that matter, rather than hard data. This suggests that Subtopic 2.2 (Comparing the quality of different Systems) may provide long-term challenges as understanding the deterrent effect that a given System would produce in the minds of potential adversaries is non-trivial.

As outlined in the introduction to this case study, we observed that the influences and mechanisms that provide the environment within which Defence System development takes place is similar to that identified for Formula 1. As in the Formula 1 case there are two main sets of influences and mechanisms, firstly there are those such as the geopolitical demands (and hence User Requirements) and also general technological advances that have the effect of defining the solution space within which a Defence System design solution must fit. Secondly there are the contractors whose influences and mechanisms interact with the solution space to produce the best solution that they can. As with the Education case study, the stakeholders appear to have accepted that there is a necessary simplification in how the quality of the product is assessed, as, short of actual war fighting, there are no good approaches to assessing the quality of the product across the uncertain future that will represent its operating context. The assumption is therefore generally made that technical performance parameters such as range and payload are surrogates for quality and so different Systems can be compared using a 'top trumps' approach.

Additionally, the Critical Realist assumption that there are structures and mechanisms that cannot be observed is a useful counterpoint to the Positivist, including authors such as Webb (1990) and Shermon (2011) and (in the author's experience) a significant number of practitioners, who, when presented with a model that does not predict as well as they would like may follow the mantra 'if you add a few more variables, you can do a better job at predictions' (Gelman (2016)). The concept that only some of the relevant effects are observable and that the analysis is designed to find the best explanation for what has been observed may enable reasonable outcomes to be reached without the dangers of over fitted models being produced. The trap of conflating correlation and causation may also be better avoided as the focus is on providing a reasonable explanation, rather than simply following the observed correlation.

Chapter 8 - Overall Conclusions

To round off this research we shall now draw out the relevant points from the areas that have been covered. Initially we shall examine the connections that have been found between the three detailed case studies and the conclusions that can be drawn from these. This is followed by more general conclusions from this work, a review of the progress made in answering the research questions and their subtopics and finally a review of the future work that will be necessary to continue to advance knowledge in this area.

8.1 - Connections between Education, Formula 1 and Defence Systems

Considering the qualitative findings of the three case studies, we may construct Table 8.1 which summarises the context of the three cases studies in a common structure.

Outcomes / Outputs / Roles	Education	Formula 1	Defence Systems
Technical Outputs	Educational experiences available	F1 car(s) plus support necessary to race it effectively.	Defence System and support necessary to use it effectively.
Aspirational Outcome	All pupils can be shown to have been offered the best education possible[148].	Win World Drivers' Championship and World Constructors' Championship.	System widely agreed to have clear and sustained advantage over competitor systems, thus deterring conflict. System performance ensures overwhelming victory when conflict occurs.
Minimum Acceptable Outcome	Perceived quality of education offered to pupils is at least as good as similar schools[149].	Team manages sufficient podium finishes to attract sufficient sponsors to fund following season.	System is perceived as being sufficiently competitive to deter casual aggression. In conflict local advantage may be gained when used carefully.
Output Quality Measurability	Not Directly Measurable	Directly Measurable	Not Measurable (until wartime).

[148] Although management might wish to add some caveats along the lines of 'subject to available resources'.
[149] Although management might also suggest that the minimum quality needs to be sufficient to attract sufficient pupils for the school to remain viable.

Direct Quality Measures	Although various measures will capture some of the output quality picture, parental perceptions capture the whole picture.	Grand Prix qualifying times, Grand Prix race results, season results.	For deterrence, perceptions of potential adversaries. When used in combat, combat exchange ratios etc.
Surrogate Quality Measures	Volume and Cost of relevant inputs	Not required.	Comparison of known / inferred performance measures.
Input Measures	Facilities spend and teacher numbers and pay.	Cost of sales and employee numbers and cost.	Development manpower, Development durations, Development costs, Production costs.

Table 8.1 - Comparison of Education, Formula 1 and Defence Case Studies

From this table we observe that, although the details of the enterprises differ, there are significant similarities from an output quality standpoint. Neither the quality of Independent School education (see Burgess (2016) and Gorard (2019)) nor effectiveness of Defence System deterrence (see Mattiss quoted in UK Ministry of Defence (2020)) can be directly observed, as both depend on the perceptions of relevant stakeholders. In each case some elements of the overall quality can be measured, for Education these will include performance in National Examinations and numbers of pupils progressing to top universities and for Defence deterrence this will include numbers of Defence Systems and their reported performance. However, all the relevant aspects are only brought together in the perceptions of the relevant stakeholders (parents and potential adversaries, respectively).

In the case of Independent School Fees the effect generated by perception will directly inform the willingness of stakeholders to fund this education and hence the number of pupils applying to a given school. In Defence the inability to directly gather opinions from potential adversaries is more challenging, as the ineffectiveness of a deterrence approach may only become clear when a war starts. If a school wished to increase its perceived quality, then there is a clear incentive to increase spending (if this can be afforded) on additional teaching staff and facilities, as these appear to be equated with a perception of quality. If this is undertaken it would produce an effect that is identical to Bowen’s Revenue Theory of Cost (Kimball (2014, p.889)), but the mechanism and motivations behind the effect would be different, as the perception of improved quality would be directed at prospective parents, rather than the peers of those managing the school. Although detailed analysis of decision making within independent schools might provide insights about which explanation of the observed effects is more likely to be true, it is unlikely that this will ever be demonstrated with any certainty. If a government wishes to improve the perceived quality (and hence deterrent value) of its armed forces it can be expected to increase spending on the military equivalent of extra teachers and facilities. It appears likely that the development and fielding of Defence Systems with improved performance characteristics is seen to be an effective approach to communicating the improved quality. This effect would be a clear illustration of when and why governments might wish to increase Defence spending, but does not provide a route to understanding how much spending (and on what Defence Systems) will provide the most effective deterrence, (Subtopic 3.2).

Formula 1 gives (during the racing season) regular feedback on the performance of the car / driver combination, relative to other car / driver combinations, additionally, it is known in advance what format races will take and (generally) what rules will apply. This means that the solution space to which car designs must conform and how performance will be measured is relatively well defined in advance. Designing Defence Systems is more challenging, as it can be challenging to predict the contexts in which it might be used across a (say) 30 year life and also to predict which potential adversaries it might need to deter over this period. This clearly provides a challenge in designing a product that is sufficiently effective against currently identified potential adversary Systems, but is also sufficiently well rounded that it is likely to be effective against future threats. Just as Formula 1 teams can introduce mid-season design changes, Defence Systems can be upgraded whilst in service, but in each case such a strategy will only be effective if the original design can accommodate the necessary changes (see Green (2023) for example).

In considering quantitative analysis of these case studies we found that there key themes emerged in each study, namely:

> The challenge of knowing, in advance[150] which data points will be sufficiently similar to form part of the same Reference Class[151].
>
> The increasing cost and volume of inputs (with staff being consistently important)[152],
>
> Challenges in objectively measuring the quality of the outputs generated in a timely fashion[153].

This was not unexpected for the Independent Schools data, as these reflect the challenges identified by Kimball and Luke (2018) in making such comparisons for the cost of US college education, as summarised in Table 7.2. The occurrence of the same factors in two

[150] And sometimes even when all of the data is available.
[151] Assuming that Flyvberg's assurances (Flyvberg (2006, p.9)) that 'most projects are both non-routine locally and use well-known technologies' are correct, this is most likely to be a particular challenge in areas with high levels of technological innovation. For example, it is likely that the Reference Classes created in the Independent Schools case study could be applied to future analysis with little risk, whereas it would not be implicit whether a new fighter jet would be in the same Reference Class as its immediate predecessors.
[152] As one would expect from the stagnant sector ((Baumol (2012, p.xx)), as indicated in Figure 12.
[153] Even in the case of Formula 1, when the Grands Prix provide a comparison of performance, there is still debate about which of the different measures best represents output quality.

case studies unconnected to education is suggestive of there being common issues across a range of complex, manpower intensive activities. Although some of the other potential case studies examined in Chapter 4 may be subject to one of these issues, none of the others are subject to all three.

We now consider each of these themes, starting with the assessments of which data points appeared to be members of the same Reference Class. This theme is of particular relevance to the forecasting of the cost of future Defence Systems, as most approaches rely on the identification of similar historic projects and hence is central to Question 2 (How should we compare Defence Systems over time).

The Formula 1 case study represented the simplest case. Even though some of the teams were never competitive, they are, by definition, members of the class of Formula 1 teams. On this basis there was no clear reason to exclude any of them from the data set. This data set was satisfactory for examining the performance of different teams but further work to segregate Teams into different Reference Classes would have been required to forecast the future budget requirements of a specific team.

For the Independent Schools data we discovered that (as one might expect) the type of education provided (i.e. Boarding vs Day) influenced the level of fees and hence which data points should be compared. What was unexpected was that the context (Boarding Only, dual Boarding and Day and Day Only) was also an important determinant of the relevant Reference Class for a given data point. We also noted that schools which had chosen to convert to state control appeared to belong to a different class to those which had chosen to remain independent. This segregation of data was based upon differences observed in the visualisation of the different data classes, rather than any preconceived ideas, as there was no indication before the analysis was performed that the different References Classes would be formed by the different contexts. As the variables that identify which Class an Independent School will belong to are clearly defined and there are a limited number of combinations of these variables, future work in this area should be able to build directly on the different Reference Classes identified here.

Unfortunately the same cannot necessarily be said for Defence Systems. The aircraft production cost data from Hartley (2023) (see Figure 64 and Figure 65) and the submarine production cost data from the relevant Naval Estimates and Defence Estimates (see Figure 59 to Figure 63) allow us to forecast the likely future production cost for an item of a given weight, assuming that very similar technologies and techniques are employed in a similar context. However, lacking the discrete categories from the Independent Schools data, challenges will always exist in determining how similar ‘similar’ needs to be to ensure membership of the same Reference Class. This issue is even more challenging for determining the likely development costs of a Defence System, where there is little public domain data and hence it is challenging to determine what characteristics might be used to group different programmes into a finite set of categories. (This challenge is predicated on our rejection of the assertions of authors such as Webb (1990) and Augustine (1997) that (say) all tactical military aircraft necessarily form a single Reference Class).

The second theme concerned the increasing cost and volume of inputs devoted to delivering the outputs. Gaining an understanding of this is relevant both to Subtopic 1.2 (the context of past costs) and 2.3 (understanding how contextual changes have affected the outcomes of past projects). We observed that this appeared to be a significant driver of the recorded costs for each of the case studies. Over the period 1974 – 2022 the Independent Schools data shows an improvement in pupil to teacher ratio of 0.75% per annum, which has had the effect of reducing this value from 12.68 to 8.90. Over the period 1977 to 2007 the average pay of teachers, relative to average UK wages increased by 0.2% per annum[154], which increased the teaching premium from 54% to 63%[155]. We also need to consider the spending on facilities, which we will value on a per pupil basis, uplifted by UK average wages to produce an increase in real terms. There is significant annual variation in these values, presumably due to the size and infrequent nature of such large investments, and so we shall take average values over five years. This approach produces an average reported value for 1974 - 1978 of £281 with the similar result for 2009 – 2013 of £2,024. This represents a growth in real terms of just under 6% per annum. It is presumed that schools and parents both assume that this represents a real-terms growth in the output quality.

As we have seen, Formula 1 experienced a much more significant growth in the number of employees, probably enabled by the increase in sponsorship revenues and other finance available in the sport. Judde et al (2013) reported that sponsorship of Formula 1 accelerated in 1978 when television coverage which had previously been 'fragmented' instead 'proliferated globally'. This led to 'an arms race that saw the budget of the world champion team increase from US$5 million in 1980 to US$40 million in 1990' and 'US$300 million in 2000'. This growth in resources has been mirrored in the increase in employees of Formula 1 Teams. Table 8.2[156] to Table 8.4 shows the increase in employees for McLaren, Williams and Tyrrell (and its successors, including most recently Mercedes) as these are the teams with the longest series of company data.

Team	Period	Employee Numbers – annual growth rate
McLaren	1982 – 2021	7.1%
Tyrrell / Mercedes	1983 – 2020	9.2%

Table 8.2 - Formula 1 Teams growth in employees 1982 - 2021

Team	Period	Employee Numbers - annual growth rate
McLaren	1982 – 1990	15.2%
Tyrrell / Mercedes	1983 – 1990	12.0%

Table 8.3 - Formula 1 Teams growth in employees 1982 – 1990

[154] Bolton (2008) and Department for Education (2023) compared with Clark (2022). (This value does not include increases in employment costs, such as National Insurance and pension contributions).
[155] It should be noted that teachers had received a significant pay rise in 1974, which increased the premium from 32% to 54%.
[156] These are Table 7.17 to Table 7.19 reproduced here for reader convenience.

Team	Period	Employee Numbers - annual growth rate
McLaren	1990 – 2021	5.1%
Tyrrell / Mercedes	1990 – 2020	8.6%
Williams	1990 – 2019	5.3%

Table 8.4 - Formula 1 Teams growth in employees 1990 - 2021

As we have previously seen in the discussion following Table 7.17 to Table 7.19 (reproduced above as Table 8.2 to Table 8.4) where abundant resources are available, then an employee growth rate in excess of 10% may observed, Even in more difficult circumstances and average growth rate of 5% per annum may be required to remain competitive.

Although we have visibility of employment costs, rather than wages, we can compare the changes in the average employee cost in these three teams to changes in average UK wages. Table 8.5 shows that the growth in the cost of employees, when compared with average UK wages, has grown slowly since 1990, indeed the fractions of a percent reported by McLaren and Williams may well represent increases in National Insurance and other overheads that accrue to the employment of staff.

Team	Annual increase up to 1990	Annual increase since 1990
McLaren	6.1%	0.4%
Tyrrell / Mercedes	4.0%	1.25%
Williams	n/k	0.2%

Table 8.5[157] - Formula 1 Teams cost per employee growth above average UK wage increases

We thus observe that despite having significantly increased the number of employees Formula 1 teams do not appear to have had (at least since 1990) to offer significant wage increases (above the UK average) to support an annual staff increase of at least 5% per annum. As Formula 1 teams employ a very small fraction of the UK workforce it is unlikely that increasing workforces here will have any wider impact, however, growth in the workforce of significantly larger industries could have crowding out effects on the rest of the economy.

Whilst the Formula 1 teams are performing a standard annual task of building and racing two cars and produce annual accounts that shed light on the business processes behind this activity, Figure 81 illustrates that the situation is far less clear for Defence Systems. There are few data points and there is significant scatter. However, if one takes the trend as running from Canberra in 1945 to Tornado IDS in 1970 an annual increase in technical effort of 10.2% is reached, whereas, if one takes the trend as running from Canberra in 1945 to Tornado ADV in 1975, then an annual increase of 6.7% is arrived at. It is interesting to note that these results are broadly similar to those observed in Formula 1, even though one might expect the driving factors in each case to be different. In each case growth rate should be related to the available budget, but also subject to the ability to demonstrate the specific value added by the additional employees, although the evidence required to justify additional staff may vary from sector to sector. There is no clear reason to expect similar results to be

[157] This is Table 7.20 reproduced here for convenience.

achieved in these different contexts. This result may be purely coincidental, as there are few Defence Systems data points and those that are present come from projects with a variety of Challenge Ratings, however the result may indicate some fundamental truth about technology intensive industries. Further work to establish how far the additional staff are adding more depth to existing technical areas and how far they represent the establishment of different technical areas is an essential next step in understanding the import of this observed effect.

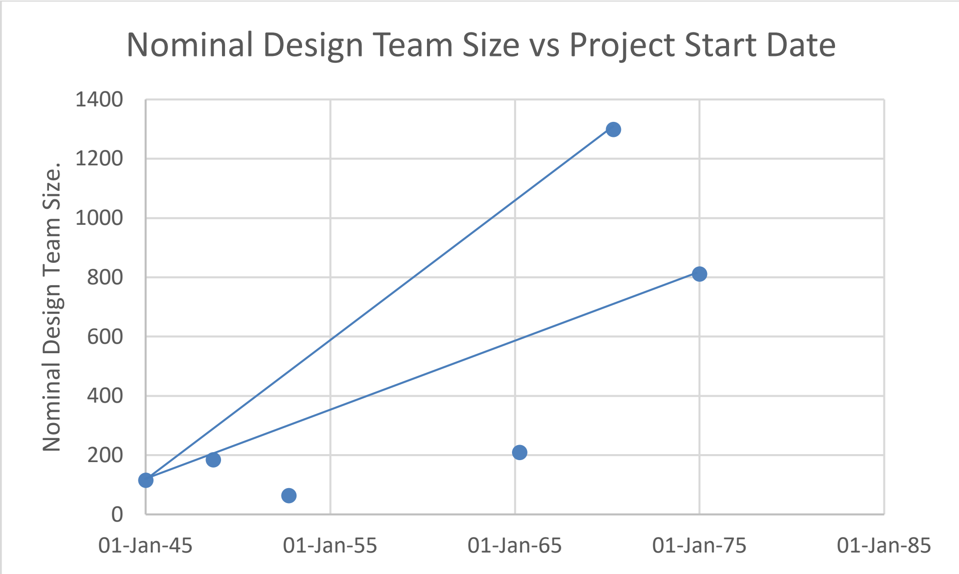

Figure 82 - UK Military Aircraft Trends in Design Team Size

Notes: This is Figure 81 reproduced, with notional upper and lower growth rate lines added. The upper line represents an annual growth of 10.2%, the lower an annual growth of 6.7%.

Considering the final theme, the difficulties in measuring quality (which relates to Subtopics 1.3, 2.2 and possibly 3.1), as we discussed earlier (see Table 8.1) direct measurement of the output quality of Independent School education and the deterrent value of Military Systems is difficult, as it mainly rests on the perception of the relevant stakeholders. Thus, except in extreme situations, decision making in these areas is likely to focus on maximising the factors that are generally agreed to contribute to perceived output quality, rather than seeking new ways to deliver quality, as a failure of a new approach to be perceived as effective could provoke an existential threat for the subject organisation. However, this safe approach may be undermined by exogenous changes, for example, the recent decisions by Oxford and Cambridge Universities to limit the number of students from independent schools may cause parents to perceive that the value of such an education has been diminished, even though the quality of the education has not changed.

In the next section we shall consider what these case study based observations and conclusions, and also the work on Challenge Rating, have added to the theory and understanding of this area. Additionally we shall consider what insights that give into the likely future cost of Defence Systems and hence how one might update forecasting practice.

8.2 - Overall Conclusions on Defence System Cost Drivers and Forecasting

Initially it is worthwhile recalling the two Critical Realist points that appear in the Forward:

- One should accept that there may well be important aspects of the process of developing and producing Defence Systems (and other goods and services) which can be inferred to exist, but whose operation cannot be directly observed.
- One should not always expect direct proof of causation and should be satisfied with the most plausible explanation that fits the known observations.

By bearing these in mind it has been possible to discover additional insights and avoid some of the pitfalls that have been identified in the work of some authors in this field (e.g. Webb (1990), Augustine (1997)). A key example of the type of fallacy that the failure to disconnect correlation and causation can produce is the assumption that as smaller aircraft developed in the past have been cheaper than large, complex aircraft, then if a candidate future large, complex aircraft can be made smaller, then it will become cheaper[158]. As we have observed technology and design Challenge Ratings drive the overall development (and probably manufacturing) costs of the System, therefore historic small, simple, low Challenge Rating aircraft have been cheaper than large, complex, high Challenge Rating offerings. However, attempting to shoe horn a given level of performance into a smaller size is likely to increase the design and perhaps technology Challenge Ratings and hence actually increase the project duration and hence costs.

The insight that there are structures and mechanisms at work which cannot be directly observed and measured and which produce effects which can only sometimes be directly observed is hugely important as it provides a broader and more flexible epistemology, within which to structure and expand our understanding of the cost of Defence Systems. We shall now consider the detailed conclusions about forecasting the cost of Defence Systems, before returning to an overall summary of the achievements made in this research and the future work that is required.

We shall now consider the contribution to knowledge that this work has provided, by revisiting the specific questions and sub topics in Table 3.1.

Question 1 – What did Defence Systems cost in the past?

Sub topics – 1.1 What was the scope and amount of these costs?
1.2 In what context were they incurred?
1.3 What was delivered for these costs?

Initially we discovered that many authors in this field fail to quote sources and / or specific values for the data under consideration and although there may be security and commercial considerations behind these omissions, it is a frustrating situation for other researchers. This work starts to remedy this shortfall by citing sources and / or actually making specific data values available where ever possible. For example, Chapter 6 includes a tabulation of the data that underpins Augustine's Law XVI (Augustine (1997)), Chapter 7 contains new data on the design effort involved in the development of six different UK tactical military aircraft and the

[158] Wilson, D. (2018) Personal Communication, 6 December reported this argument as having been present in the early phases of the Typhoon aircraft requirements definition.

two Annexes cover auction prices of Stradivarius violins and the annual fees for the BA course at the University of Pennsylvania. The location of the Independent School fee data and of the Aircraft Development programme data are given in the data statement on page ii.

Although these steps represent an improvement on the status quo, there is still more work to do to understand the specific scope of costs being quoted, to clarify the context within which they were incurred and to find a consistent approach to describing what was produced. Further detailed investigation of changes in numbers and roles of employees in Formula 1 teams, and a similar examination of Defence company records, to establish the effort expended on different technical areas for different projects would provide an incredibly useful resource. This would allow a better understanding of how far resource increases are being driven by a requirement for individual technology and design areas to be addressed in greater depth and how far the increases are driven by an increase in the number of technology and design areas that need to be addressed. This would need to be coupled with an assessment of the impact of advances in enabling technologies, such as the introduction of Computer Aided Design, which could, in theory, reduce requirements for project manpower.

Whilst the introduction of some technologies to the working environment result in measurable productivity gains, in many other cases it is unclear whether the hoped for improvements have been realised. Comparing the manpower allocated to Defence System projects before and after the introduction of processes such as Computer Aided Design and new Project Management techniques and could provide useful evidence to economists and industrial historians about their effectiveness.

Apart from a consideration of the broad economic situation, which is needed to compare the cost of Defence Systems expenditure over time this work has not had the space to investigate changes in the economic context of the production of Defence Systems. The changes in economic context that are likely to drive available resources are considered in the education case study (changes in UK income levels (see Figure 16) and the Formula 1 case study (see Figure 56), but changes in Defence budgets and government willingness to pay are not considered. Similarly, the structural changes in the UK aerospace and shipbuilding industries have not been directly addressed. For example, as late as 1977 there were two UK based manufacturers of military aircraft (Hawker Siddeley Aviation and the British Aircraft Corporation) which were, in that year nationalised to form British Aerospace, which was, in turn denationalised between 1981 and 1985 (RAF Museum (no date)). Similarly, the maritime sector (Newcastle University (2017)) contained 11 companies that built warships of various sizes before nationalisation in 1977, of which three re-emerged at privatisation in the mid-1980s, all of which are now part of BAE Systems. Whilst probably not influencing the duration or direct labour hours in a project, these structural changes in the Industries might well influence labour rates and the level of overhead charges. Additionally, company specific factors, such as the other work being undertaken at the same time could also affect costs.

Question 2 – How should we compare Defence Systems over time?

Sub topics – 2.1 How can we compare their cost?

2.2 How can we compare their quality?
2.3 How have contextual changes affected outcomes?

Considering this second Question, we observe that Chapter 6 provides a thorough summary of approaches to comparing costs over time. Although no new theory is developed, the analysis specifically identifies that the historic costs of Defence System projects should be uplifted in line with the fraction of GDP that their historic spend represented. This recommendation, together with the discussion of weaknesses with the other approaches should assist researchers and practitioners in selecting and justifying the most appropriate approach. The illustration that different ways of uplifting the cost of Systems produced in the 1910s to current conditions can vary the result achieved by an order of magnitude highlights that having a clearly articulated and justified approach to uplifting historic costs must form an essential part of any regression based upon historic cost data.

In attempting to compare the quality of Defence Systems over time, we recall that Figure 13 illustrated that we currently have no direct way of measuring the impact of changes in the quality of inputs to Defence Systems projects, and short of going to war, all measurements of the output quality are also based on the perceptions of relevant stakeholders. However, it is not clear that stakeholder perceptions in this area would be reliable.

Henry Ford is widely quoted as saying ‘If I had asked people what they wanted, they would have said faster horses’. Although it is disputed whether Ford ever uttered these words (e.g. Vlaskovits (2011)) it conveys a similar sentiment to that offered by Steve Jobs "It's really hard to design products by focus groups. A lot of times, people don't know what they want until you show it to them." (Valentino-DeVries (2011)). These two quotations both speak to the challenge of valuing the quality of past Systems and delivering improved output quality in the future when those making the judgement are not necessarily well placed to imagine what improved quality might look like. Additionally, in such situations, in the absence of direct measures of output quality, those who are less visionary than Ford and Jobs may find it convenient to follow whatever general agreement there is about which of the direct measures are perceived to drive output quality. Any decisions that increase these direct measures will therefore be accepted, by most stakeholders, as contributing to an improvement in output quality.

Thus, it will always be safe for the management of an Independent School to spend any spare resources on improving the pupil to teacher ratio or constructing additional facilities, as the stakeholders who need to be impressed (mainly parents of current pupils and of potential pupils) will tend to agree that this is money well spent. Such behaviours, or course, tend to reinforce the effects noted by Bowen’s ‘Revenue Theory of Cost’ (Kimball (2014, p.889)), as discussed above, because most stakeholders will tend to agree that this will contribute to the ‘Educational Excellence’ and increased quality that the school is seeking.

A similar process may well apply in Defence. There is some literature that discusses the quality of Defence Systems (e.g. Olsson (2020)) and this observes that there are two types of values used in such assessments, there are proxy values such as age and weight and actual performance measures such as protection, mobility and firepower. Olsson's performance measures based option is very similar to the approach taken by Hove and Lillekvelland (2016) in searching for a correlation between performance measures and platform cost. Olsson's approach is heavily context specific and a simple ratio of performance measure to some baseline may not capture the subtleties of performance and quality[159]. For example, the protection, mobility and firepower of a fleet of tanks may well be a good way measure quality, if they are to be pitted against a peer or near peer force in a classic tank battle and a 10% performance advantage may well reflect the average outcome across a number of encounters. However, if a fighter aircraft has a superior radar (say with a 10% greater range) and so is able to detect an adversary aircraft and launch a missile, before the enemy is aware of its presence, then the effective performance difference is likely to be closer to 100:0 rather than the expected 100:90.

Despite the difficulties with this 'top trumps' approach to valuing actual Defence System quality, it has obvious appeal and can also be used to generate modelling with incredible granularity, although its validity is dependent on having made the correct assumptions about future usage of the equipment. On this basis, the only opportunity for a radical departure from the constant evolution of current platforms and the incremental addition of novel technologies, is if the quality of current equipment is tested in war fighting and found to be lacking, in that specific context. For example, the early experience of the British Army in Afghanistan, attempting to mount effective patrols, when hampered by Improvised Explosive Devices led to the rapid development and deployment of the fleet of Protected Patrol Vehicles (see Allen (2020)). Such radical changes occur very infrequently in the Air and Naval environments, and the most accessible examples for future work would probably be Naval Systems between 1850 and 1950 which saw the introduction of steam power, armour, turbine propulsion, submarines and aircraft.

As there appears to be no realistic prospect of meaningful quality measures for Defence Systems, future work is required to understand the geopolitical and technological context of past systems and how this has affected the resources devoted to their delivery. In terms of the input quality demanded, the level of technological and design advancement demanded by the customer can be assessed, but it is not immediately apparent what factors led to these demands being made. The assessment of the novel Challenge Rating concept, as evidenced in past aerospace projects, has demonstrated that these two broad measures, that can also be estimated for future project options, without detailed design work having been executed, provide a reliable estimate of the programme duration. However, this success needs to be coupled with an improved understanding of the drivers of project team manpower requirements. Specifically an understanding is required of the rate at which existing technology areas / design areas require additional resources (for a given Challenge Rating) and also the rate at which novel technology and design areas are added to the project team. An investigation of Formula 1 teams' personnel requirements over time is likely to provide a straight forward initial view of this, which could then guide research on the Defence side.

---

[159] And naturally will not capture the other aspects of fighting power.

This might also provide insights into the impact of enabling technologies such as Computer Aided Design.

Question 3 – How should we expect prices to change over time?

Sub topics – 3.1 How will the cost of systems change over time?
3.2 How will customer willingness to pay change?

As we have previously seen, when designing a new Defence System there is a solution space, defined by available resources, customer aspirations and the geo-political situation into which a new system must fit. There is also a space defined by the available resources, available technologies and contractor competency that defines what system can actually be delivered in the available time. The dialogue between customer and contractor should aim to find an amicable point where the proposed system meets customer aspirations, whilst not imposing too onerous risks (in terms of high Challenge Ratings) on the contractor and the project.

As we have previously seen, in addition to any contextual factors (e.g. relating to industry structure) the cost of the systems will be mainly driven by the duration of the programme (as driven by the Challenge Rating) and by the annual cost of the project team, mainly driven by 'what we had last time', plus additional depth required in technology areas plus additional technology areas required (as driven by customer requirements and Challenge Ratings). An understanding of how different Challenge Ratings of future programme options would drive different programme durations and requirements for team size would allow early estimation of the likely costs of future project options.

In general, the question of how much the customer is willing to pay will be determined by the overall prosperity of the United Kingdom and hence the revenues available to the government and also how threatened the nation feels and hence what proportion of revenues should be spent on Defence. Where resources are insufficient difficult choices are sometimes made, such as when New Zealand disbanded its Air Combat Forces in 2001, however the UK currently aspires not only to own and operate tactical military aircraft, but also to maintain the industries necessary to design and build future platforms. The value in having a sovereign industry with these capabilities is recognised in terms of the operational sovereignty benefits that it produces, but assigning an economic value to this capability is challenging. This current work and future related work can help answer the question of what future Defence Systems are likely to cost. Fortunately the potentially difficult questions of how much the UK ought to spend on Defence are outside the scope of this study area..

8.3 - Augustine and the Starship Enterprise Revisited

As this work started with a consideration of the work of both Augustine and also Kirkpatrick and Pugh, it is worth returning to these topics before sketching out the key future work required.

In nominal terms Augustine observed (see Section 2.1) that the cost of tactical aircraft increases by a factor of four per decade (i.e. a fraction under 15% per annum), however the difficulties in disentangling cost increases from other structural changes makes it very difficult to determine the relative impact of pure cost increases, as against changes in the number of aircraft procured and general inflation in the economy. Using the cost data in Table 6.1 and other historical data to determine build numbers it might be possible to generate an estimate of the real cost increase, as Kirkpatrick and Pugh attempted.

Kirkpatrick and Pugh (see Section 2.2) observed a real cost increase of 8.3% per annum. This value is tantalisingly close to one of the values in Table 8.4, which represented the manpower growth in one of the Formula 1 teams, however this might also be a coincidence. Interestingly this figure is also bounded by the two trend lines in Figure 83.

It is therefore possible that these authors held the key to understanding the growth in the cost of Defence Systems, but were unable to assemble the necessary data to fully investigate the driving factors.

Future work is therefore required to examine the growth in labour inputs to Defence Systems and, if possible, the productivity changes of these inputs.

8.4 - Key Future Work Required

The research reported in this thesis indicates that there are several potentially fruitful areas for future research that would build on the work reported here and contribute to understanding of the drivers of the cost of Defence Systems.

Firstly understanding the changes in the scope and depth of Formula 1 and Defence Systems project teams will provide insights into whether the growth in team size is driven by the need to address traditional design areas in greater depth and / or to address additional design areas created by the introduction of new technologies and / or design approaches. This will complement the research reported earlier in this thesis that investigated how the duration of Defence System and Civil Aircraft projects had varied over time.

Secondly, it is likely that there will be future innovations that will be touted as reducing the labour effort necessary to design and manufacture Defence Systems. If evidence could be gathered examining the effectiveness of previous innovations of this ilk, such as Computer Aided Design and Computational Fluid Dynamics, this would give an initial idea of what the outcome is likely to be.

In addition to these two factors that affect individual projects, there have been significant structural changes in UK Defence industries since 1945. BAE Systems is now the sole UK based fixed wing aircraft manufacturer of any size, with familiar names such as Bristol, De Havilland, Folland, Handley Page, Hawker Siddeley and Vickers all having disappeared through merger or bankruptcy. This rationalisation may have had positive effects on the cost of military aircraft, through reducing spare capacity and concentrating research, development and manufacturing expertise and also freeing up Defence workers to labour in more economically useful areas of the economy. It may also have produced negative effects through the elimination of competition removing the incentive to control costs. Gaining an understanding of the impact of these structural changes would improve our ability to comprehend the range of factors impacting these activities.

Whilst these further investigations may resolve the question of what drives the cost of the current types of Defence Systems, a broader approach would be required to examine issues such as whether a paradigm shift downwards in cost might be achieved in the future. The perceived effectiveness of unmanned systems and artificial intelligence could offer alternative approaches, but historical work is required to examine whether, when technology has changed in Defence, the overall cost has actually ever reduced. Of course, depending on the social and geopolitical situation the future perceived affordability of Defence, against future spending on Welfare and Healthcare may also be a challenge.

Annex A - Stradivarius auction price data

Sources: Data gathered from Tarisio.com (accessed 8 August 2017)

Additional data from Hill and Hill (1902)

These results are consistent with Graddy (2011), but their scope is different as that author also includes private sales and only includes violins where data on two or more sales is available.

Name	Date Made	Auction House	Sale Date	Nominal $	Nominal £	Nominal Franc
Bonualot, Lady Margaret	1694	Puttick & Simpson	05/12/1907		410	
Bucher	1683	Southeby's	05/04/1984		101,200	
Cathedrale	1707	Sotheby's	22/11/1984		396,000	
Dancla	1703	Christie's	01/05/1907		590	
Dancla	1703	Christie's	23/06/1959		8,190	
Ex Joachim Kortschak	1698	Christie's	18/11/1998		529,500	
Ex-Fuchs		Christie's	31/03/1989		156,200	
Ex-Vogelweith	1711	Sotheby's	12/11/1987		165,000	
Falmouth	1692	Puttick & Simpson	28/05/1853		110	
Falmouth	1692	Puttick & Simpson	01/12/1936		1,550	
Falmouth	1692	Christie's	02/06/1982		102,600	
Falmouth	1692	Christie's	26/06/1987		192,500	
Goding / Jupiter / Janze	1722	Christie & Manson	18/02/1857		200	
Hammer	1707	Christie's	16/05/2006	3,544,000		
Hrimaly	1712	Sotheby Parke Bernet	29/06/1984	165,000		
Innes, Loder	1729	Christie's	12/12/1935		1,365	
Innes, Loder	1729	Christie's	09/12/1969		14,500	
Innes, Loder	1729	Sotheby's	22/06/1988		214,500	
Jules Falk	1719	Sotheby's	03/04/1985		286,000	
Kreutzer	1727	Christie's	01/04/1998		947,500	
La Pucelle / La Pucello	1709 or 1710 ?	Hotel Drouot	14/02/1878			22,100
Lady Blunt	1721	Sotheby's	03/06/1971		84,500	
Lady Blunt	1721	Tarisio	22/06/2011		9,808,000	
Lady Tennant	1699	Christie's	22/04/2005	2,032,000		
Le Marien	1714	Tajan	03/02/1998		619,778	
Leopold Auer	1690	Christie's	31/10/1984	308,000		
Lyall	1702	Sotheby Parke Bernet	18/01/1984	231,000		
Marie Hall	1709	Sotheby's	31/03/1988		473,000	

Name	Date Made	Auction House	Sale Date	Nominal $	Nominal £	Nominal Franc
Mendelssohn	1720	Christie's	21/11/1990		902,000	
Mercury	1688	Puttick & Simpson	30/06/1893		500	
Mercury	1688	Puttick & Simpson	19/06/1907		750	
Molitor	1697	Christie's	31/03/1989		209,000	
Molitor	1697	Tarisio	15/10/2010	3,600,000		
Muir-Mackenzie	1694	Puttick & Simpson	02/12/1920		1,700	
Muir-Mackenzie	1730[160]	Sotheby Parke Bernet	13/06/1983	275,000		
Nachez, Hill	1686	Sotheby Parke Bernet	24/11/1979	95,000		
Nachez, Hill	1686	Christie's	29/04/1872		175	
Penny	1700	Christie's	04/04/2008	1,273,000		
Piatti	1717	Sotheby's	19/03/1986		170,500	
Rosenheim	1686	Southeby's	22/11/1984		165,000	
Schreiber	1712	Christie's	18/03/1992		352,000	
Sighicelli	1694	Millon & Associes	01/03/1989		107,131	[161]
Soames	1684	Southeby's	24/05/1973		21,000	
Soames	1684	Glendining & Co	27/03/1907		500	
Solomon, ex-Lambert.	1729	Christie's	02/04/2007	2,728,000		
Taft ex Emil Heerman	1700	Christie's	05/05/2000	1,326,000		
Viotti	1712	Hotel Bouillon	??/??/1824			3,816

[160] It is unclear whether there are one or two violins named Muir-Mackenzie
[161] Actually priced in Francs, but sources report hammer price in GB pounds.

Annex B - University of Pennsylvania data

Sources: Years 1828 - 1899 University of Pennsylvania Catalogues. Available on line, see University of Pennsylvania (2019a).

Years 1900 – 2016 University of Pennsylvania Archives and Record Center compilation (Lloyd and Heavens (2016)).

Years 2017 – 2018 University of Pennsylvania website, as archived at the Internet Archive.

Academic Year Starting September	Annual BA Fees $	Academic Year Starting September	Annual BA Fees $
1828	62	1924	275
1829	62	1925	275
1830	75	1926	400
1831	75	1927	400
1832	75	1928	400
1833	75	1929	400
1834	75	1930	400
1835	75	1931	410
1836	n/k	1932	410
1837	75	1933	410.5
1838	n/k	1934	411
1839	75	1935	411
1840	75	1936	411
1841	75	1937	420
1842	75	1938	420
1843	75	1939	420
1844	75	1940	420
1845	75	1941	420
1846	75	1942	420
1847	75	1943	420
1848	75	1944	420
1849	75	1945	420
1850	75	1946	495
1851	75	1947	570
1852	75	1948	620
1853	75	1949	620
1854	75	1950	625

Academic Year Starting September	Annual BA Fees $	Academic Year Starting September	Annual BA Fees $
1855	90	1951	625
1856	90	1952	785
1857	90	1953	785
1858	90	1954	785
1859	90	1955	935
1860	90	1956	935
1861	90	1957	1,150
1862	90	1958	1,200
1863	105	1959	1,400
1864	105	1960	1,400
1865	105	1961	1,600
1866	105	1962	1,630
1867	105	1963	1,630
1868	105	1964	1,750
1869	105	1965	1,750
1870	105	1966	1,950
1871	105	1967	1,950
1872	150	1968	2,150
1873	150	1969	2,350
1874	150	1970	2,550
1875	150	1971	2,750
1876	150	1972	3,000
1877	150	1973	3,165
1878	150	1974	3,450
1879	150	1975	3,790
1880	150	1976	4,125
1881	150	1977	4,450
1882	150	1978	4,825
1883	155	1979	5,270
1884	155	1980	6,000
1885	155	1981	6,900
1886	155	1982	8,000
1887	155	1983	8,880
1888	155	1984	9,600
1889	155	1985	10,400
1890	150	1986	11,200
1891	160	1987	11,976

Academic Year Starting September	Annual BA Fees $	Academic Year Starting September	Annual BA Fees $
1892	160	1988	12,749
1893	160	1989	13,700
1894	160	1990	14,890
1895	160	1991	15,894
1896	160	1992	16,838
1897	160	1993	17,838
1898	160	1994	18,856
1899	150	1995	19,898
1900	150	1996	21,130
1901	150	1997	22,250
1902	150	1998	23,254
1903	150	1999	24,230
1904	150	2000	25,170
1905	160	2001	26,630
1906	160	2002	27,988
1907	160	2003	29,318
1908	160	2004	30,716
1909	160	2005	32,364
1910	160	2006	34,166
1911	160	2007	35,916
1912	160	2008	37,526
1913	160	2009	38,970
1914	160	2010	40,514
1915	160	2011	42,098
1916	160	2012	43,738
1917	210	2013	45,890
1918	210	2014	47,668
1919	215	2015	49,536
1920	270	2016	51,464
1921	270	2017	53,534
1922	275	2018	55,584
1923	275		

References

AIAA (The American Institute of Aeronautics and Astronautics) (2015). *Aerospace Today…and Tomorrow – An Executive Symposium*. 4 June. Available at https://www.aiaa.org/att2015/ (Accessed 11 October 2015).

Alexander, C. (2017). 'Parametric cost and schedule modeling for early technology development'. Presented at *NASA Cost and Schedule Symposium at NASA Headquarters on August 30.* Available at https://apps.dtic.mil/sti/pdfs/AD1128246.pdf (Accessed 30 September 2023).

Allen, C. (2020). *Foxhound: Protected Patrol Vehicle.* 16 July. Available at https://www.keymilitary.com/article/foxhound-protected-patrol-vehicle#: (Accessed 3 February 2024).

Arena, M. V. (2006). *Why has the cost of Navy ships risen?: A macroscopic examination of the trends in US Naval ship costs over the past several decades*. Rand Corporation, MG-484.

Arena, M.V., Younossi, O., Brancato, K., Blickstein, I. and Grammich, C.A. (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 Corporation.

Augustine, N.R., (1997). *Augustine's Laws*. Reston, USA: AIAA.

Augustine, N.R., (2015) 'Augustine's Laws and Major System Development Programs'. *Defense Acquisition Research Journal,* 22, (1), pp. 2 – 63.

Aversa, P., Furnari, S. and Haefliger, S. (2015). 'Business model configurations and performance: A qualitative comparative analysis in Formula One racing, 2005–2013'. *Industrial and Corporate Change*, 24(3), pp. 655-676.

Awojobi, O. and Jenkins, G.P. (2016). 'Managing the cost overrun risks of hydroelectric dams: An application of reference class forecasting techniques'. *Renewable and Sustainable Energy Reviews*, 63, pp. 19-32.

Bannock, G., Baxter, R.E. and E. Davis (2003). *The Penguin Dictionary of Economics.* London: Penguin Books.

Batselier, J. and Vanhoucke, M. (2016). Practical application and empirical evaluation of reference class forecasting for project management. *Project Management Journal*, 47(5), pp. 36-51.

Baumol, W.J. and W.G. Bowen (1966). *Performing Arts : The Economic Dilemma.* New York: The Twentieth Century Fund (Reprint, 1978). Available at: https://archivesofthecentury.org/myportfolio/performing-arts-the-economic-dilemma/ (Accessed 26 April 2019).

Baumol, W.J. (2012), *The Cost Disease: Why Computers Get Cheaper and Health Care Doesn't,* Yale University Press.

BBC (2007). *MoD confirms £3.8bn carrier order*, 25 July. Available at http://news.bbc.co.uk/1/hi/scotland/6914788.stm (Accessed 7 August 2023).

Begg, D., Fisher, S. and Dornbush, R. (1994). *Economics.* Maidenhead: McGraw-Hill.

Bell, A., Smith, J., Sabel, C.E. and Jones, K. (2016). 'Formula for success: multilevel modelling of Formula One driver and constructor performance, 1950–2014'. *Journal of Quantitative Analysis in Sports*, 12(2), pp.99-112.

Beveridge, W.H. (1939). *Prices and Wages in England from the Twelfth to the Nineteenth Century* London: Longmans, Green & Co.

Bhaskar, R., 1978. On the possibility of social scientific knowledge and the limits of naturalism. *Journal for the Theory of Social Behaviour*, 8(1), pp.1-28.

Bolkcom, C. (2009). *F-35 lightning II Joint Strike Fighter (JSF) Program: Background, Status, and Issues*. Washington: Congressional Research Service, Library of Congress.

Bolton, P. (2008). 'Teachers' pay statistics'. *House of Commons Library*, SN/SG/1877. Available at https://dera.ioe.ac.uk//22821/ (Accessed 27 April 2021).

Bongers, A. and Torres, J.L. (2014). 'Technological change in U.S. jet fighter aircraft' *Research Policy*, 43, pp. 1570 – 1581.

Boyns, T. (2021). 'Organizational change, budgetary control and success and failure in Formula 1: Rubery Owen and British Racing Motors, 1947–1977'. *Management & Organizational History*, 16:3-4, 204-227.

Brauer, J., Hartley, K. and Markowski, S. (2021). 'Rethinking Augustine's Law: armament costs and evolving military technology'. In Chatterji, M. and Gangopadhyay, P. (eds) *New frontiers in conflict management and peace economics: With a focus on human security* (pp. 1-14). Emerald Publishing Limited.

Brooks, F. P. Jr (1975). 'The Mythical Man-Month'. *Essays on Software Engineering.* Reading, USA: Addison-Wesley Publishing Company.

Burgess, S (2016). *Human Capital and Education: The State of the Art in the Economics of Education*. IZA Discussion Paper, No. 9885. Bonn: Institute for the Study of Labor.

Business Insider (2015). *Here's exactly how much the price of a New York City subway ride has changed over the last 100 years.* Available at https://www.businessinsider.com/how-much-the-price-of-the-new-york-city-subway-has-changed-2015-3?r=US&IR=T (Accessed 25 October 2023).

Bygstad, B., & Munkvold, B. E. (2011). 'In search of mechanisms. Conducting a critical realist analysis'. *Proceedings of the 32nd International Conference on Information Systems, Shanghai, China.*

Carvalho, A. (2022). *Using Historical Data to Identify the Best Driver in Formula 1 History*. Doctoral dissertation. National College of Ireland, Dublin.

Clark, G. (2011). *Average Earning and Retail Prices, UK, 1209 – 2010*. Available at: https://www.measuringworth.com/datasets/ukearncpi/earnstudynew.pdf (Accessed 01 August 2017).

Clark, G. (2022) *What Were the British Earnings and Prices Then (New Series).* Available at: http://www.measuringworth.com/ukearncpi/ (Accessed 13 April 2022).

Coad, J. and Guillery, P. (2003). *The Portsmouth Block Mills. The Start of a Revolution.* English Heritage, Swindon. Available at https://historicengland.org.uk/research/results/reports/7165/THEPORTSMOUTHBLOCKMILLS-THESTARTOFAREVOLUTION (Accessed 24 November 2023).

Cole, S.K., Reeves, J.D., Williams-Byrd, J.A., Greenberg, M., Comstock, D., Olds, J.R., Wallace, J., DePasquale, D. and Schaffer, M. (2013). *Technology estimating: A process to determine the cost and schedule of space technology research and development*. NASA Technical Publication-2013-218145.

Companies House (2023). Companies House Website. https://www.gov.uk/government/organisations/companies-house (Accessed 21 July 2023).

Cudhay, B.J. (1995). *Under the Sidewalks of New York: The Story of the Greatest Subway System in the World*. Fordham University Press.

Curtis, P. (2012) 'How the Ofsted chief got his maths wrong on Sats' *Guardian Politics Blog*, 15 March. Available at: https://www.theguardian.com/politics/reality-check-with-polly-curtis/2012/mar/15/ofsted-chief-maths-wrong (Accessed 30 September 2019).

Dallas News (2011). *Only Six Bucks? Super Bowl ticket prices through the years.* Available at http://www.dallasnews.com/sports/super-bowl/the-scene/20110130-only-six-bucks-super-bowl-ticket-prices-through-the-years.ece (Accessed 13 March 2019).

Davies, N., Eager, A., Maier, M. and Penfold, L. (2011). *Intergenerational Equipment Cost Escalation*. London: Ministry of Defence, Defence Economic Research Paper. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/280041/18_december_2012.pdf. (Accessed 22 January 2016).

Davis, E, (1985). *They Moved the Millions.* Livingston Enterprise. Available at: https://www.nycsubway.org/wiki/They_Moved_The_Millions (Accessed 2 November 2023).

Deitchman, S.J. (1979). New technology and military power: general purpose military forces for the 1980s and beyond. *Boulder Colorado: Westview Press*.

Department for Education (2023). *School Workforce in England.* Available at: https://explore-education-statistics.service.gov.uk/find-statistics/school-workforce-in-england (Accessed 19 July 2023).

Department for Transport (2020). *Updating the Evidence Behind the Optimism Bias Uplifts for Transport Appraisals*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/983759/updating-the-evidence-behind-the-optimism-bias-uplifts-for-transport-appraisals.pdf (Accessed 26 June 2023).

Drew, J (2015). *First operational F-35 squadron declared ready for combat.* Flight Global 31 July. Available at: https://www.flightglobal.com/first-operational-f-35-squadron-declared-ready-for-combat/117812.article (Accessed 7 August 2023).

Dunnett, M.(2018) 'Lessons for higher education from private – and quasi-private - schools'. *Higher Education Policy Institute blog*, 22 January. Available at: https://www.hepi.ac.uk/2018/01/22/lessons-higher-education-private-quasi-private-schools/ (Accessed: 11 April 2022).

Economist (2010) 'Briefing - Defence spending in a time of austerity' *Economist.* 26 August. Available at: http://www.economist.com/node/16886851 (Accessed 10 October 2015).

Eldredge, N. and Gould S. J. (1972). 'Punctuated equilibria: an alternative to phyletic gradualism' In Schopf T.J.M. (ed), *Models in Paleobiology*. San Francisco: Freeman Cooper.

Eichenberger, R. and Stadelmann, D., 2009. Who is the best Formula 1 driver? An economic approach to evaluating talent. *Economic Analysis and Policy*, 39(3), pp. 289 - 406.

Eloranta, J. (2023). *Military Spending Patterns in History.* Available at https://eh.net/encyclopedia/military-spending-patterns-in-history/ (Accessed 24 January 2024).

Eskew, H.L. (2000). 'Aircraft Cost Growth and Development Program Length: Some Augustinian Propositions Revisited'. *Acquisition Review Quarterly,* Summer 2000, pp.209 – 220.

Eton College (2019). *Annual Report and Consolidated Financial Statements 2018 – 2019.* Available at https://register-of-charities.charitycommission.gov.uk/charity-search/-/charity-details/4010044/accounts-and-annual-returns (Accessed 3 May 2021).

Fleetwood, W., (1745). *Chronicon Preciosum; or, An account of English money, corn,&c.* London: T Osborne. Available at: https://books.google.co.uk/books?hl=en&lr=&id=TC45AAAAMAAJ&oi=fnd&pg=PA1&dq=Fleetwood,+W.,+(1745).+Chronicon+Preciosum%3B+&ots=K0NFbp-SS_&sig=cKGyVxFeZn_2QZqMwSyo1FDDZxk&redir_esc=y#v=onepage&q=Fleetwood%2C%20W.%2C%20(1745).%20Chronicon%20Preciosum%3B&f=false (Accessed 30 September 2023).

Flyvberg, B. (2006). 'From Nobel Prize to Project Management: Getting Risks Right'. *Project Management Journal*, 37(3), pp. 5 – 15.

Formula 1 Website. https://www.formula1.com/en/racing/2023/Austria/Circuit.html for example. (Accessed 9 June 2023).

Franck, R.E. (1992). *Cost-performance choices in post-Cold War weapon systems*. Air University Press. Available at https://apps.dtic.mil/sti/pdfs/ADA271762.pdf (Accessed 2 February 2024).

Frank, R,H. (1999). *Luxury Fever – Why Money Fails to Satisfy in an Era of Excess.* New York: The Free Press.

Friedman, S.M. (1996). *A Brief History of the University of Pensylvania.* Available at: https://archives.upenn.edu/exhibits/penn-history/brief-history (Accessed 1 January 2019).

Gansler, J.S. and Lucyshyn, W. (2013). *Cost as a Military Requirement*. University of Maryland, Center For Public Policy and Private Enterprise.

GAO (2011) 'Trends in Nunn-McCurdy Cost Breaches for Major Defense Acquisition Programs. GAO11-295R'. Government Accountability Office, Washington D.C. Available at www.**gao**.gov/new.items/d**11295r**.pdf (Accessed 26 January 2016).

Gelman, A.(2016) 'If you add a few more variables, you can do a better job at predictions'. *Columbia University StatModelling Blog* , 19 April. Available at: https://statmodeling.stat.columbia.edu/2016/04/19/28548/ (Accessed: 24 March 2024).

Gilbert, S.W. (2013). *Applying the Hedonic Method.* National Institute of Standards and Technology, Technical Note 1811. Available at: https://nvlpubs.nist.gov/nistpubs/TechnicalNotes/NIST.TN.1811.pdf (Accessed 23 March 2018).

Gorard, S. (2019). *School choice in an established market*. Aldershot: Ashgate.

Gough, C. (2000). Science and the Stradivarius. *Phys World* **13** (4) 27

Graddy, K. and Margolis, P.E. (2011). Fiddling with value: violins as an investment? *Economic Inquiry*, *49*(4), pp.1083-1097.

Green, F., Anders, J., Henderson, M., & Henseke, G. (2017). *Who chooses private schooling in Britain and why?* University College London, LLAKES Research Paper 62. Available at: https://discovery.ucl.ac.uk/id/eprint/10043039/1/Green%2C%20Anders%2C%20Henderson%20%26%20Henseke.pdf (Accessed 15 February 2023).

Green, J. (2023). 'Toto Wolff says Mercedes to make 'fundamental' design changes for 2024 Formula 1 car'. Sky Sports. 23 June. Available at https://www.skysports.com/f1/news/12433/12907559/toto-wolff-says-mercedes-to-make-fundamental-design-changes-for-2024-formula-1-car (Accessed 3 February 2024).

Guardian (2014) 'Edinburgh's tram system, opens £375m over budget and three years late' *Guardian*, 1 June. Available at http://www.theguardian.com/uk-news/2014/jun/01/edinburgh-tram-system-opens (Accessed 20 Jan 2016).

Hartley, K. and Braddon, D., 2014. ‘Collaborative projects and the number of partner nations’. *Defence and Peace Economics*, 25(6), pp.535-548.

Hartley, K., (2020). ‘Rising costs: Augustine revisited’. *Defence and Peace Economics*, 31(4), pp.434-442.

Hartley, K. (2023). ‘Costs and prices of UK military aircraft in war and peace’. *Defence and Peace Economics*, 34(4), pp.512-526.

Henderson, B. (2015). *The Experience Curve – Reviewed (Part I) The Concept.* Boston Consulting Group. Available at: https://www.bcgperspectives.com/content/Classics/strategy_supply_chain_management_experience_curve_reviewed_the_concept/ (Accessed 12 October 2015).

Henry, N., Pinch, S. and Russell, S. (1996). ‘In pole position? Untraded interdependencies, new industrial spaces and the British motor sport industry’. *Area*, 28(1), pp.25-36.

Henry, N., Angus, T. and Jenkins, M. (2021). ‘Motorsport Valley revisited: Cluster evolution, strategic cluster coupling and resilience’. *European Urban and Regional Studies*, 28(4), pp.466-486.

Herbsleb, J.D. and Grinter, R.E. (1999). ‘Splitting the organization and integrating the code: Conway’s law revisited. In *Proceedings of the 21st international conference in Software Engineering* (pp.85 – 95). ACM.

Hess, R.W. and Romanoff, H.P (1987). *Aircraft Airframe Cost Estimating Relationships*. RAND Corporation, R-3255-AF.

Hill, W.H. and Hill, A.F. (1902) *Antonio Stradivari, His Life and Work (1644-1737)*. United Kingdom: Hill.

Hirsch, F (1976). *Social Limits to Growth.* Cambridge, Massachusetts: Harvard University Press, (Third Printing, 1978). Available at: https://archive.org/details/sociallimitstogr00fred (Accessed 26 April 2019).

HM Treasury (2013). *HMT Green Book: Supplementary Guidance.* Available at https://www.gov.uk/government/publications/green-book-supplementary-guidance-optimism-bias (Accessed 10 October 2016).

HM Treasury (2014). *GDP deflators: user guide*. Available at: https://www.gov.uk/government/publications/gross-domestic-product-gdp-deflators-user-guide/gdp-deflators-user-guide (Accessed: 23 March 2018).

House of Commons Library (2009). *Statistical literacy guide. How to adjust for inflation*. Available at http://researchbriefings.files.parliament.uk/documents/SN04962/SN04962.pdf (Accessed 9 March 2018).

Hove, K. and Lillekvelland, T. (2016). 'Investment cost escalation – an overview of the literature and revised estimates'. *Defence and Peace Economics*, 27:2, pp. 208-230.

Independent Schools Council (2001). *ISC Census and Annual Report 2001.* Available at: https://www.isc.co.uk/media/2451/2001_annualcensus_isc.pdf (Accessed: 11 April 2022).

Independent Schools Council (2019). *ISC Census and Annual Report 2019*. Available at: https://www.isc.co.uk/media/5479/isc_census_2019_report.pdf (Accessed: 11 April 2022).

Jefferis, T.J., (2014) 'Cost Analysis for Staff Officers'. [Lecture]. Shrivenham, 11 November.

Jenkins, S.P., Dean, H. and Platt, L. (2016a). The income distribution in the UK. *Social Advantage and Disadvantage, ed. by H. Dean, and L. Platt*, pp.135-160.

Jenkins, M., Pasternak, K. and West, R. (2016). *Performance at the limit: Business lessons from Formula 1 motor racing*. Cambridge University Press.

Jencks, C. and D. Riesman (1968). *The Academic Revolution.* Chicago: University of Chicago Press.

Johnstone, B.M., (2020). 'Augustine's laws: Are we really headed for the $800 billion-dollar fighter'. In *ICEAA Professional Development & Training Workshop, San Antonio, Texas*.

Jones, D. and Woodhill, N. (2009). 'Measuring Defence Inflation'. *Fourteenth Government Statistical Service Methodology Conference*. Available at: http://www.ons.gov.uk/ons/about-ons/get-involved/events/events/fourteenth-gss-methodology-conference--30-june-2009/developing-a-measure-of-defence-inflation-paper.pdf (Accessed 22 January 2016).

Judde, C., Booth, R. and Brooks, R. (2013). 'Second place is first of the losers: An analysis of competitive balance in Formula One'. *Journal of Sports Economics*, 14(4), pp.411-439.

Kahneman, D., 2011. *Thinking, fast and slow*. London: Macmillan.

Kaldor, M. (1982). *The Baroque Arsenal*. London: Andre Deutsch.

Kimball, B.A. (2014) 'The Rising Cost of Higher Education: Charles Eliot's "Free Money" Strategy and the Beginning of Howard Bowen's "Revenue Theory of Cost," 1869—1979'. *The Journal of Higher Education*, 85:6, pp. 886-912.

Kimball, B.A. and Luke, J.B. (2018). 'Historical Dimensions of the "Cost Disease" in US Higher Education, 1870s–2010s'. *Social Science History*, 42(1), pp.29-55.

Kirkpatrick, D. (2003). 'A UK perspective on defence equipment acquisition'. *Singapore: Institute of Defence and Strategic Studies Singapore*.

Kirkpatrick, D. and Pugh, P. (1983). 'Towards the Starship Enterprise – are the current trends in defence unit costs inexorable?' *Aerospace*, May 1983, pp. 16-23.

Kozloski, R (2013). *The Cost of Defence Oversight.* USNI Blog. Available at: http://blog.usni.org/2013/08/28/the-cost-of-defense-oversight. (Accessed 11 July 2023 via Internet Archive).

Kronemer, A. and Henneberger, J.E. (1993). 'Productivity in aircraft manufacturing'. *Monthly Labor Review*, 116, pp.24-33.

Kvalvik, S. and Johansen, P. (2008). *Enhetskostnadsvekst På Forsvarsinvesteringer (EKV-I) [Unit CostGrowth of Defence Investments]*. Forsvarets Forskningsinstitutt (FFI) Rapport 2008/01129.

Large, J.P., Campbell, H.G. and Gates D. (1976). *Parametric Equations for Estimating Aircraft Airframe Costs*. RAND Corporation, R-1694-1-PA&E.

Lawson, J., & Silver, H. (1973). *A Social History of English Education.* London: Metheun.

Lee, W.E. (2016). *Waging war: Conflict, culture, and innovation in world history*. New York: Oxford University Press.

Lehtonen, J. and Anteroinen, J. (2016). 'The Capability Factors as Explanatory Variables of Equipment Unit Cost Growth: A Methodological Proposal'. *Defence and Peace Economics*, 27:2, pp.280-298.

Levi, L. (1885). *Wages and Earnings of the Working Classes: Report to Sir Arthur Bass, MP*. London: J. Murray.

Lloyd, M.F. and Heavens, N.G. (2016). 'Tuition and Mandated Fees, Room and Board, and Other Educational Costs at Penn'. Available at: https://archives.upenn.edu/exhibits/penn-history/tuition/ (Accessed 3 October 2023).

Machin, S. and Vignoles, A. (2005). *What's the Good of Education? The Economics of Education in the UK*. Princeton University Press.

Maddison, A. (2007). *The world economy. Volume 1: A millennial perspective and Volume 2: Historical statistics*. Washington: OECD Publishing.

Marino, A., Aversa, P., Mesquita, L. and Anand, J. (2015). 'Driving performance via exploration in changing environments: Evidence from Formula One racing. *Organization Science*, 26(4), pp.1079-1100.

Markowski, S., Brauer, J. and Hartley, K. (2023). 'Augustine investments and weapons systems'. *Defence and Peace Economics*, 34(3), pp.293-307.

Mastromarco, C. and Runkel, M. (2009). 'Rule changes and competitive balance in Formula One motor racing'. *Applied Economics*, 41(23), pp.3003-3014.

McDonnell Douglas F/A-18 Hornet (2015). *Wikipedia*. Available at: https://en.wikipedia.org/wiki/McDonnell_Douglas_F/A-18_Hornet (Accessed 13 October 2015).

Mill, J.S. (1828). *Speech on Perfectibility* [Speech]. Spoken at the Debating Society. Available at http://www.utilitarian.org/texts/perfectibility.html (Accessed 26 September 2015).

Ministry of Defence (1981). *Statement on the Defence Estimates 1981.1* (Session 1980/81 Cmnd. 8212-II).

Ministry of Defence (2021). *Carrier Strike Group hits important milestone*. 4 January. Available at https://www.gov.uk/government/news/carrier-strike-group-hits-important-milestone (Accessed 7 August 2023).

Ministry of Defence (2022). *UK Defence Doctrine (Joint Doctrine Publication 0-01).* Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1118720/UK_Defence_Doctrine_Ed6.pdf (Accessed 27 October 2023).

Ministry of Housing, Communities and Local Government (2017). *Dwelling Stock Estimates:2017, England.* Available at:https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/710382/Dwelling_Stock_Estimates_2017_England.pdf (Accessed 22 September 2019).

Mott MacDonald (2002). *Review of Large Public Procurement*. Available at: http://www.parliament.vic.gov.au/images/stories/committees/paec/2010-11_Budget_Estimates/Extra_bits/Mott_McDonald_Flyvberg_Blake_Dawson_Waldron_studies.pdf (Accessed 14 October 2016).

MTFCA (Model T Ford Club of America) (2023). *Encyclopedia.* Available at https://www.mtfca.com/encyclopedia/ (Accessed 28 November 2023).

Mustangsmustangs.com (2015). *P-51 Mustang Production.* Available at: http://www.mustangsmustangs.com/p-51/production (Accessed 13 October 2015).

NASA (2010). *Technology Readiness Levels Demystified*. Aug 20. Available at https://www.nasa.gov/topics/aeronautics/features/trl_demystified.html (Accessed 10 August 2023).

Nationwide (2023). *UK House Prices since 1952.* Available at https://www.nationwidehousepriceindex.co.uk/download/uk-house-prices-since-1952 (Accessed 26 September 2023).

Newcastle University (2017). *British Shipbuilders Corporation.* Available at https://co-curate.ncl.ac.uk/british-shipbuilders-corporation/ (Accessed 3 February 2024).

Nordhaus, W.D. (1996). ‘Do real-output and real-wage measures capture reality? The history of lighting suggests not’. In Bresnahan, T.F. and Gordon, R.J. (eds) *The economics of new goods* (pp. 27-70). University of Chicago Press. Available at https://www.nber.org/system/files/chapters/c6064/c6064.pdf (Accessed 11 September 2023).

Nordlund, P. Åkerstström, J. Öström, B. and Löfstedtdt, H. (2011).Kostnadsutveckling FörFörsvarsmateriel[Defence Materiel Cost Development]. Stockholm: FOI, Swedish Defence Research Agency.

Novick, D. (1970). *Are Cost Overruns a Military-Industry-Complex Specialty?* Rand Corporation, Santa Monica, California. P-4311.

O'Donoghue, J., Golding, L., & Allen, G. (2004). Consumer price inflation since 1750. *Economic Trends*, (604) pp. 38-45.

Obama, B. (2013). *Remarks on the Economy* [Speech]. Galesburg, Illinois, 24 July. Referenced at: https://quoteinvestigator.com/2016/05/03/expense/ (Accessed 21 July 2020).

ONS (Office for National Statistics) (2007). *The ONS Productivity Handbook – A Statistical Overview and Guide.* Basingstoke, UK: Palgrave Macmillan, 2007. Available at: https://ons.gov.uk/file?uni=/economy/economicoutputandproductivity/productivitymeasures/methodologies/productivityhandbook/handbook/onsproductivityhandbooktcm77187914.pdf (Accessed 26 March 2018).

ONS (Office for National Statistics) (2011). *History of and differences between the Consumer Prices Index and Retail Prices Index.* Available at: https://www.ons.gov.uk/ons/rel/cpi/consumer-price-indices/history-of-and-differences-between-the-consumer-prices-index-and-retail-prices-index/history-of-and-differences-between-the-consumer-price-index-and-retail-price-index---article.pdf (Accessed 15 July 2017).

ONS (Office for National Statistics) (2014a). *Consumer Prices Indices Technical Manual.* Available at: http://webarchive.nationalarchives.gov.uk/20160109133536/http://www.ons.gov.uk/ons/rel/cpi/consumer-price-indices---technical-manual/2014/index.html (Accessed 22 March 2018).

ONS (Office for National Statistics) (2014b). *UK National Accounts – a short guide, September 2014*. Available at: https://www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/national-accounts/articles/2011-present/uk-national-accounts---a-short-sguide--2014.pdf (Accessed 26 March 2018).

ONS (Office for National Statistics) (2015). *A Short Guide to the UK National Accounts, September 2015*. Available at: http://webarchive.nationalarchives.gov.uk/20160105160709/http:/www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/national-accounts/articles/2011-present/a-short-guide-to-the-uk-national-accounts.pdf (Accessed 23 March 2018).

ONS (Office for National Statistics) (2017). *Consumer price indices, a brief guide: 2017.* Available at: https://www.ons.gov.uk/economy/inflationandpriceindices/articles/consumerpriceindicesabriefguide/2017 (Accessed 9 March 2018).

ONS (Office for National Statistics) (2019). *RPI Average Prices 1914 – 2004*. Available at https://www.ons.gov.uk/file?uri=/economy/inflationandpriceindices/methodologies/consumerpricesindexcpiandretailpricesindexrpibasketofgoodsandservices/rpiaverageprices19142004tcm77168515tcm77420253.xls (Accessed 18 June 2019)

Office of the Secretary of Defense, Cost Assessment and Program Evaluation (2020). 'DOD Cost Estimating Guide v1.0'. *US Department of Defense*. Available at https://www.cape.osd.mil/files/Reports/DoD_CostEstimatingGuidev1.0_Dec2020.pdf (Accessed 1 September 2023).

Officer, L. & Williamson, S. (2006). Better Measurements of Worth. *Challenge*, 49:4, pp. 86-110.

Officer, L. and Williamson, S. (2010). *Measures of Worth*, Measuring Worth. Available at: http://www.measuringworth.com/worthmeasures.php (Accessed 8 October 2015).

Officer, L. & Williamson, S. (2012). *Measures of Worth*. Measuring Worth. Available at: www.measuringworth.com/worthmeasures.php (Accessed 1 August 2017).

Olechowski, A.L., Eppinger, S.D., Joglekar, N. and Tomaschek, K. (2020). 'Technology readiness levels: Shortcomings and improvement opportunities'. *Systems Engineering*, 23(4), pp.395-408.

Olsson, P. (2022). Measuring quality of military equipment. *Defence and Peace Economics*, *33*(1), pp.93-107.

Pagano, U. (2007). 'Positional Goods and asymmetric development' in Yotopoulos, P.A. and D. Romano (eds) *The Asymmetries of Globalization.* Routledge, Abindgdon, pp 28 – 47.

Patil, A., Jain, N., Agrahari, R., Hossari, M., Orlandi, F. and Dev, S. (2023). 'A Data-Driven Analysis of Formula 1 Car Races Outcome'. In *Artificial Intelligence and Cognitive Science: 30th Irish Conference, AICS 2022, Munster, Ireland, December 8–9, 2022, Revised Selected Papers* (pp. 134-146). Cham, Switzerland: Springer.

Peck, R. (2001). 'Athenian Naval Finance in the Classical Period'. *Trireme Trust*. Available at: http://www.triremetrust.org.uk/tt-thesis.htm#ch3 (Accessed 20 July 2017).

Phelps Brown, E.H. & Hopkins, S. (1955). Seven Centuries of Building Wages. *Economica*, 22(87), pp. 195-206.

Phelps Brown, E. H. and Hopkins, S. (1956). Seven Centuries of the Prices of Consumables. *Economica*, 23, pp.194-95.

Phillips, A.J. (2014). 'Uncovering Formula One driver performances from 1950 to 2013 by adjusting for team and competition effects'. *Journal of Quantitative Analysis in Sports*, 10(2), pp.261-278.

Philips, A.J. (2019). ‘A new F1 metrics model’. *F1metrics blog*, 10 September. Available at: https://f1metrics.wordpress.com/2019/09/10/a-new-f1metrics-model/ (Accessed 9 June 2023).

Pugh, P. (1986). *The Cost of Sea Power: The Influence of Money on Naval Affairs from 1815 to the Present Day*. London: Conway Maritime Press.

Pugh, P. G. (2007). *Source book of defence equipment costs*. Self Published, ISBN: 978-0-9556258-0-0

RAF Museum (no date). *Records of the British Aviation Industry in the RAF Museum.* Available at https://www.rafmuseum.org.uk/documents/Research/Company-Papers/Guide-to-Aircraft-Industry-Records.pdf (Accessed 3 February 2024).

Ridley, M. (2010). *The Rational Optimist*. London: Fourth Estate.

Rippy, D. (2014). The first hundred years of the Consumer Price Index: a methodological and political history. *Monthly Lab. Rev.*, 137, pp.1 - 42.

Rockerbie, D.W. and Easton, S.T. (2022). ‘Race to the podium: separating and conjoining the car and driver in F1 racing’. *Applied Economics*, 54, pp.6272-6285.

Royal Commission on the Revenues and Management of Certain Colleges and Schools (1864). London: The Stationery Office. Available at: https://books.google.co.uk/books/download/Appendix.pdf?id=H-Q-AAAAYAAJ&output=pdf&sig=ACfU3U2-IJTgcOlzkgVmMGfDyFtx89--cQ (Accessed 11 April 2022).

Sadin, S., Povinelli, F. and Rosen, R. (1989), ‘The NASA Technology Push Towards Future Space Mission Systems’. *Acta Astronaut*, 20, pp. 73-77.

Sauser, B., Verma, D., Ramirez-Marquez, J. and Gove, R. (2006). ‘From TRL to SRL: The concept of systems readiness levels’. In *Conference on Systems Engineering Research*, 5(2). Los Angeles, USA: Stevens Institute of Technology.

Schneider, M. (2007). ‘The Nature, History and Significance of the Concept of Positional Goods’, *History of Economics Review*, 45:1, pp. 60-81, DOI: 10.1080/18386318.2007.11681237

Shermon, D. (2011). Historical Trend Analysis Analysed. *Journal of Cost Analysis and Parametrics*, 4(1), pp.52-62.

Siemens (2023). ‘How Siemens Xcelerator helps breaking rising development costs in Aerospace & Defense through strategic digitalization’. *Siemens Digital Industries Software blog*, 18 May. Available at https://blogs.sw.siemens.com/nx-design/how-siemens-xcelerator-helps-breaking-rising-development-costs-in-aerospace-defense-through-strategic-digitalization/ Accessed 4 September 2023.

Smallwood D. (2012), 'Augustine's Law Revisited'. *Sound and Vibration,* March 2012 pp.4–5.

Smedley, R., Wise, C., Enkhbayar, D., Price, G., Cheng, R. and Yang, G. (2020). 'The fastest driver in Formula 1', *Amazon Web Services*, 20 August 2020. Available at: https://aws.amazon.com/blogs/machine-learning/the-fastest-driver-in-formula-1/#:~:text=Topping%20our%20list%20of%20rankings,Max%20Verstappen%2C%20and%20Fernando%20Alonso. (Accessed 30 May 2023).

Smithsonian (2015). *SPAD XIII 'Smith IV'.* Smithsonian National Air and Space Museum Website. Available at: http://airandspace.si.edu/collections/artifact.cfm?object=nasm_A19200001000 (Accessed 13 October 2015).

Spurgeon, B. (2009). 'The Machine or the Man: Which Makes a Team Win?'. *New York Times*, 27 August. Available at: https://www.nytimes.com/2009/08/27/sports/autoracing/27iht-PRIX.html (Accessed 31 May 2023).

Stein, H. (1989), 'The Washington Economist'. *AEI Economist*, June 1989, p.1

Thames TV (2015). *Interview with Bernie Eccleston*. Available at: https://www.youtube.com/watch?v=zl68OQPunZs (Accessed: 13 June 2023).

Thorold Rogers, J. E. (1887). *A History of Agriculture and Prices in England: (1259-1792)* (Vol. 6). Oxford: Clarendon Press.

Thorold Rogers, J. E. (1890). *Six centuries of work and wages: the history of English labour*. London: Swan Sonnenschein.

Ticketmaster (2023). *NFL Superbowl LVIII Tickets.* Available at https://www.ticketmaster.com/superbowl (Accessed 25 October 2023)

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), pp.1124-1131.

United Nations, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development & World Bank. (2009). *System of National Accounts 2008*. New York: United Nations.

University of Pennsylvania (1884). *University of Pennsylvania Catalogue and Announcements 1883 – 84.* University of Pennsylvania. Available at: https://archives.upenn.edu/digitized-resources/docs-pubs/catalogues/catalogue-1883-84 (Accessed 1 January 2019).

University of Pennsylvania (2019a). *University Catalogues.* Available at: https://archives.upenn.edu/digitized-resources/docs-pubs/catalogues (Accessed 18 February 2019).

University of Pennsylvania (2019b). *Annual Financial Report 2018 – 19.* Available at: https://www.finance.upenn.edu/wp-content/uploads/Annual-Financial-Report-FY19.pdf (Accessed: 19 April 2022).

Valentino-DeVries, J. (2011). 'Steve Jobs's Best Quotes' *The Wall Street Journal,* 24 August. Available at https://www.wsj.com/articles/BL-DGB-23002 (Accessed 13 November 2023).

Vandermeulen, A. (1968). 'A Remission from Baumol's Disease: Ways to Publish More Articles'. *Southern Economic Journal,* Vol 35, No.2 pp.189 – 191.

Venn, J. (1888). The logic of chance. Third edition. London and New York: Macmillan. Available at https://www.gutenberg.org/ebooks/57359 (Accessed 11 March 2024).

Vignoles, A., & Machin, S. (2018). *What's the Good of Education?*: *The Economics of Education in the UK*. Princeton: Princeton University Press.

Vlaskovits, P. (2011). 'Henry Ford, Innovation and that "Faster Horse" Quote'. *Harvard Business Review blog.* 29 August. Available at https://hbr.org/2011/08/henry-ford-never-said-the-fast (Accessed: 13 November 2023).

Walczak, R. and Majchrzak, T. (2018). 'Implementation of the reference class forecasting method for projects implemented in a chemical industry company'. *Acta Oeconomica Pragensia*, 1, pp.25-33.

Webb, D. (1990). 'Cost Complexity Forecasting: Historical Trends of Major Systems'. *Journal Of Parametrics*, 10(4) pp.67 – 95.

Whitaker's Almanack (1980). London: J Whitaker & Sons

Wright, T.P. (1936). 'Factors Affecting the Cost of Airplanes'. *Journal of Aeronautical Sciences,* 3(4) pp.122 – 128.

Zayed, Y. and Loft, P. (2019). 'Agriculture Historical Statistics'. *House of Commons Library,* Briefing Paper 3339. Available at: https://researchbriefings.files.parliament.uk/documents/SN03339/SN03339.pdf (Accessed 12 October 2022).