Can stimulating demand drive costs down? World War II as a natural experiment

Lafond, Francois and Greenwald, Diana and Farmer, J. Doyne

Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Mathematical Institute, University of Oxford, Isabella Stewart Gardner Museum, Boston, Santa Fe Institute

1 June 2020

Can stimulating demand drive costs down? World War II as a natural experiment*

François Lafond[1,2], Diana Greenwald[3], and J. Doyne Farmer[1,2,4]

[1]*Institute for New Economic Thinking at the Oxford Martin School*
[2]*Mathematical Institute, University of Oxford*
[3]*Isabella Stewart Gardner Museum, Boston*
[4]*Santa Fe Institute*

June 1, 2020

Abstract

For many products, increases in cumulative production are associated with decreasing unit costs. However, a serious problem of reverse causality (lower prices leading to increasing demand) makes it difficult to use this relationship for policy. We study World War II, during which the demand for military products was largely exogenous, and the correlation between production, cumulative production and an exogenous time trend was limited. Our results indicate that decreases in cost can be attributed roughly equally to the growth of experience and to an exogenous time trend.

Keywords: innovation policy; learning curve; natural experiment; World War II.
JEL codes: 031;038;N62.

1 Introduction

An old and well-established literature demonstrates empirically that for many products, unit cost drops as cumulative production increases (Thompson 2007). This relationship has been called several different names, including the *learning curve*, the

*This work was supported by the EU grant FP7-ICT-2013-611272 (GROWTHCOM), Partners for a New Economy, Baillie Gifford, the EU grant 730427 (COP-21 RIPPLES), the Rebuilding Macroeconomics project (which is supported by the Economic and Social Research Council), and the Institute for New Economic Thinking at the Oxford Martin School. We would like to thank Kieran Marray for excellent research assistance and Nick Chater, Giovanni Dosi, Cameron Hepburn, David Hendry, Christopher Magee, John Muellbauer, Bill Nordhaus, Matteo Richiardi, our INET colleagues as well as audiences at the CCS 2016, EHS 2018, EEA-ESEM 2019, Singapore NUS, Oxford, MIT, LSE, OECD, and CONCORDi 2019 for comments. The usual disclaimer applies. Contacts: francois.lafond@inet.ox.ac.uk, dsgreenwald@gmail.com, doyne.farmer@inet.ox.ac.uk.

experience curve and *Wright's law*. Wright's law is named for Theodore Wright, who conjectured that the cost of manufacture of a given model of airplane by a given plant decreases as a power function of the cumulative number of planes produced (Wright 1936).

This observed relationship between costs and production has often been understood as causal and used to set policy. Large-scale public support for a particular industry is often justified by the argument that artificially increasing production will "push it down its learning curve" until it becomes competitive in a global market. Experience curves play a particularly important role in integrated assessment models for climate change mitigation, where they make it possible to model endogenous technological progress (Nordhaus 2014, Witajewski-Baltvilks et al. 2015).

Nonetheless, there are three major related issues with using experience curves as a justification for policy. First, there is the possibility of reverse causality. Experience curves, at least when used to understand the effects of policy, postulate a causal relation from cumulative production to costs. However, as long as demand is elastic and costs are going down, lower prices can induce higher demand, and thus a rise in production and experience. A simple regression of costs on experience therefore does not identify the causal effect of experience on cost, but also picks up the effect of demand elasticity.

A second issue is that experience curves are plagued by an omitted variables bias. Productivity improvements take place for a variety of reasons. One of these is learning-by-doing, which refers to the improvements in the manufacturing process that are stimulated by production experience. We may think of experience as a proxy for all relevant product-specific cumulated efforts, and – absent the reverse causality issue mentioned above – identify their causal effect on cost decrease. However, productivity improvements can also occur independently of product-specific or industry-specific experience or innovation, such as knowledge spillovers from other industries or the general improvement in a country's economic institutions.

Attempts to address this second problem of omitted variable bias lead to a third problem. The simplest option to control for technological progress that is *not* related to experience is to include a time trend, which is intended to capture all productivity improvements that are due to "exogenous" factors. In practice, however, adding a time trend to a regression introduces a serious estimation problem. In most empirical examples production grows exponentially, so that experience also grows exponentially. Because experience is the sum of past production, fluctuations are damped. This means that the log of experience is similar to a deterministic linear time trend, and multicollinearity makes it impossible to distinguish the relative effects of experience from those of the time trend. While multicollinearity is well understood, and does not by itself bias the coefficient of interest, standard errors become very large and coefficients are unstable under changes in specification, yielding ambiguous results.

Military production during World War II provides a unique natural experiment that helps us solve all three of these issues. Demand was largely exogenous because it was driven by battlefield needs. Because battlefield requirements quickly ramped up, then eventually plateaued and decreased, production, experience and time are *not* strongly correlated. This allows us to estimate their relative effects more precisely.

During World War II the U.S. Armed Forces transformed from the world's eighteenth largest military to by far the world's largest (Herman 2012). The U.S. produced war equipment not only for themselves, but for all the allies as well, manufacturing roughly two thirds of the total equipment produced by both sides combined. This is an enormous exogenous demand shock. While the relative allocations to specific types of war equipment may have been somewhat cost sensitive, the overall build up was clearly driven more by needs on the battlefront than by the price of weapons (Hall 2009, Ramey 2011). Moreover, capacity was often stretched to a maximum, further limiting endogeneity concerns that may have arisen if fast learning producers were able to obtain more future contracts.

The massive build-up in military production was then followed by a dramatic slowdown. This makes the time series for production far from exponential and alleviates the problem of multicollinearity. Fig. 1 illustrates a good example. Production of Ford's Armored Car, model M-20 GBK, began in May 1943 and increased to a peak of 400 units per month by September 1943; it dropped dramatically by spring 1944, rising later but never to the previous level. In contrast, costs came down more or less continuously, and a plot of the logarithm of cost against the logarithm of experience yields an approximately straight line. The key point here is that costs continued to drop even as production *decreased*, lending evidence to the importance of cumulative production (experience) rather than production itself as a key variable (Asher (1956), p.87).

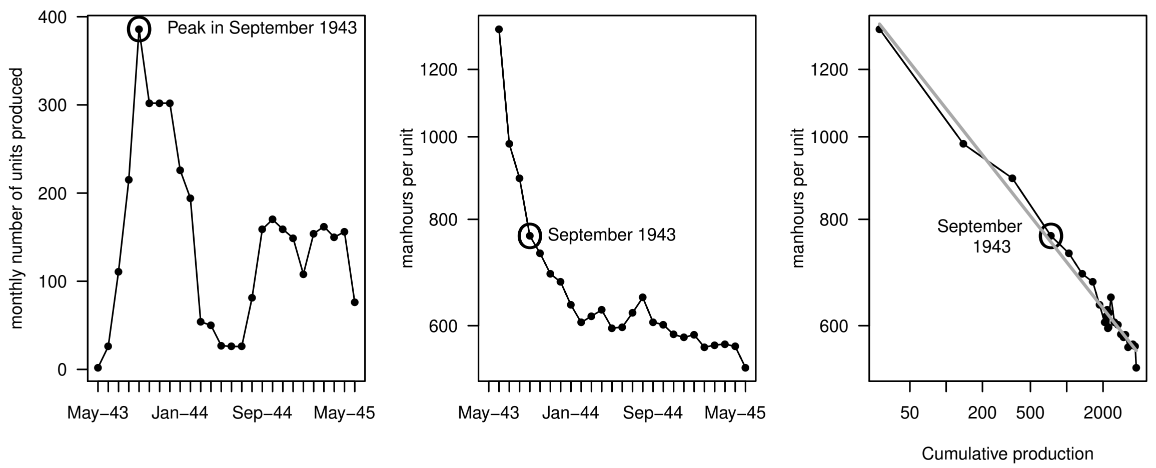

Figure 1: Production, manhours per unit and experience curve (with linear fit) for Ford's Armored Car M-20 GBK.

To take advantage of this unique demand shock, we have gathered novel and unprecedentedly large amounts of data on U.S. military production during the war. First we collected 152 monthly time series on labor productivity for ships and aircraft from historical studies (Searle 1945, Alchian 1963), and for ground transport vehicles from Ford Motor Company archives. Second, we constructed a new dataset for the overall unit cost decline for more than 500 military products by combining information in the *Official Munitions Production of the United States* (OMPUS) Handbook (Civilian Production Administration 1947) with information in the *United States Munitions*

Handbook (USMH), which we discovered in the U.S. National Archives. Finally, we analyze aggregate data on indices of contract prices broken down into ten different categories. These three datasets differ in terms of the main variable available (labor productivity vs unit cost), number of observations (complete time series or not), level of aggregation (from plant to War Department level), and type of product. However, in combination they provide an unprecedentedly comprehensive picture of industrial military production during World War II.

To further bolster the value of this data, we have applied new corrections for estimating cumulative experience prior to the beginning of production observed in our data. Generally, earlier studies have simply applied a correction based on the assumption of exponential growth. Because exponential growth would be a poor assumption in our case, and because our datasets cover highly heterogenous products, we developed an alternative method of estimating prior experience using available historical statistics.

Our results make it clear that costs depend on both experience and an exogenous time trend, in roughly equal amounts. Thus, experience *does* cause cost improvements, lending strong support for arguments in favor of state-sponsored demand to accelerate the development of critical technologies.

Unfortunately, our data does not allow us to evaluate deeper channels of causality. For instance, a higher productivity that follows a rise in experience can be due to direct learning-by-doing, or to any productivity-enhancing factor that tends to scale with experience, such as research and development. We refrain from commenting on these specific channels, and simply conclude that endogenous effects (as captured by experience at producing a specific product) and exogenous effects (as captured by a time trend or time dummies) explain about as much of the average rate of cost decrease. As a result, a large demand shock is likely to accelerate declines in costs of production.

An important caveat is that the relationship between experience and cost is very heterogeneous: while for some products, costs decline sharply if experience increases, other products are less sensitive to experience effects. As a result, a policy stimulus is only likely to be effective for a particular product when there is a favorable historical relationship.

The paper is organized as follows: Section 2 reviews the literature on experience curves, in particular during World War II, and the broader conclusions and criticisms of the relationship between experience and cost. Section 3 explains why the World War II environment provides us with a unique identification strategy. Section 4 introduces the datasets used in this paper. Section 5 explains our identification strategy and provides our empirical framework. Section 6 presents the results for each of three datasets. Section 7 provides an overview of the results obtained in all datasets, and the last section concludes. An extensive appendix provides more detailed information about the data collection and cleaning, corrections for prior experience, and several robustness checks.

2 Literature review

There is large literature documenting the relation between increasing experience, defined as cumulative production, and decreasing costs[1]. The literature has its roots in Wright's study of the aircraft industry but has now gone far beyond that, with studies ranging in aggregation level from individual plants to total global output of individual products.[2]

It is clear that experience is positively correlated with cost declines, but the relative importance of the different channels through which this happens is an unsettled debate. Because Wright's law was originally observed at the plant level, one likely explanation is simply learning-by-doing: workers and managers learn while they are on the job, improve the manufacturing process and thereby lower costs. In fact, there is an established literature in psychology about the "power law of practice" for tasks such as cigar rolling or reading inverted text, with performance often measured as time taken to perform the task (see e.g. Newell & Rosenbloom (1981) and Heathcote et al. (2000) for a review and an alternative functional form). Interestingly, at this level and in experimental conditions, the direction of causality is clear.

Military products have played an important role in the development of the concept of learning-by-doing in economics. According to Hirsch (1956), much of the early impetus to study the evolution of costs came from the U.S. Air Force, which was interested in cost decreases in airplane production. Theodore Wright himself was head of U.S. aircraft production during WWII and is reputed to have used his own law to forecast aircraft production costs. The Air Force commissioned a number of reports, particularly from RAND in the late 1940's and 50's. These early reports suggested strong learning-by-doing effects in aircraft production (Asher 1956, Alchian 1963). Using WWII U.S. shipbuilding data on the well-known Liberty Ships, Rapping (1965) attributed the large decrease in manhours per unit to organizational and individual learning from production experience.

Since the publication of these studies in the mid-twentieth century, scholars have been challenging and complicating the simple learning-by-doing interpretation of the effects of cumulative production on costs. These follow-up studies have focused on several areas: capital intensity, quality of management, knowledge spillovers, and research and development.

A higher capital intensity has been identified as a particularly important driver of reduction in labor needed per unit of production (Scott-Kemmis & Bell 2010, David 1974). Thompson (2001) collected better data on capital intensity in Liberty Ships production, and found that learning-by-doing was largely overestimated, with labor productivity improving mostly thanks to higher capital stock. Mishina (1999) stud-

[1] Preston & Keachie (1964) established a distinction between dynamic cost functions, which are sequential observations of costs over time and can follow Wright's law, and traditional microeconomic theoretical static cost functions, which are pictures of alternatives, representing what cost would be if a firm decides to produce a certain level of output. Here we focus on the literature on the effects of experience, but of course our paper relates to a vast literature on scale elasticities and on the returns to R&D.

[2] Wright (1936), Rapping (1965), Sheshinski (1967), David (1974), Yelle (1979), Argote & Epple (1990), Thompson (2001, 2007), Nagy et al. (2013), Lafond et al. (2018).

ied the Boeing Seattle plant producing the B-17 Flying Fortress aircraft in detail and was able to control for capital intensity. He found that productivity stopped growing when production was scaled down, and rejected cumulative production as a source of cost reduction, a finding that does not generalize in our dataset. Bahk & Gort (1993) worked with plant-level data, measuring experience as cumulative production but adding a variable for capital vintages. One of their conclusions is that when the capital vintages variable is included, the effect of calendar time disappears – or even reverses –suggesting that "industry-wide" learning is actually capital embodied.

Looking at the production of ships during WWII, Searle (1945) found that the rates of cost reduction varied from one yard to another more than they did between different types of vessels within a given yard. This lends support to the idea that the quality of management within an organization has a significant effect on costs of production, as Argote & Epple (1990) have argued. In a classic paper about the cotton spinning industry, Lazonick & Brush (1985) found evidence that management-worker relations impacted work intensity and productivity improvements.

Another well studied source of cost improvement is experience in production of another model or by another firm. Thornton & Thompson (2001) extended Thompson's (2001) dataset with 57 different types of ships and studied a variety of experience spillovers, both between and across yards. They found that such spillovers could be quite important, though this is sensitive to the indicator used. Intergenerational externalities (between different generation of a product) have been identified by Irwin & Klenow (1994) in 2 out of 7 DRAM models, and by Benkard (2000) for two generations of Lockheed airliners. Irwin & Klenow (1994) found strong international interfirm exteralities in the DRAM industry, and Argote et al. (1990) found only mild interyard spillovers. Levitt et al. (2013) estimated the effects of experience on labor productivity and defect rates in an automotive plant, and documented in exceptional detail how learning by the workers of one shift was transmitted to workers of another shift.

More recent studies have focused on the effects of research and development on cost reduction. Sinclair et al. (2000) studied detailed product-level learning curves in the chemical industry, and found that the products for which there was the tightest connection between cost improvement and production experience were also those for which specific (process) R&D projects had been developed. They argue that because cumulative production conditions future returns to R&D, often cumulative production will be associated to R&D and thus to cost decrease. Funk & Magee (2014) have focused on the pre-commercial period of production. They showed that significant pre-commercial technological progress occurs. This suggests that technological progress, at least in this period, is not mostly driven by production experience. While production experience may then contribute to technological progress, it is possible that it remains mostly driven by other factors, such as in-house deliberate R&D but also developments in fundamental research outside of a product's direct knowledge base. Looking at firm-level, non quality-adjusted unit costs in India, Dosi et al. (2017) found that R&D intensive sectors tended to exhibit a "negative" learning rate, pointing towards R& D as a source of improvement in quality, leading to higher costs rather than cost savings. Another explanation for heterogeneous learning rates is design complexity: if engineers perform improvements by trial-and-error on parts of a more complex entity, more complex design require more coordinated improvement, which is less likely (McNerney et al. 2011).

Despite detailed datasets and their focus on different channels that affect productivity, many studies tend to suffer from an issue that plagues experience curves: because experience is a cumulative variable and it tends to grow exponentially, the inclusion of a time trend to correct for omitted variables results in serious multicollinearity that prevents precise estimation of relative effects. To give some examples from energy economics, where experience curves are well studied, Papineau (2006) considered a panel of renewable energy technologies in different countries and found that introducing a time trend resulted in a large decrease of the experience coefficient. Similar dramatic effects were observed and emphasized by Söderholm & Sundqvist (2007), although not in all specifications. In particular, for specifications that included R&D, they found that the time trend tends to pick up the effect of R&D and leave the experience coefficient relatively stable. Other studies where including a time trend led to unstable results include Benkard (2000) and Bahk & Gort (1993). In their regressions of defect rates on experience in an automotive plant, Levitt et al. (2013) report opposite signs for the time trend and experience when using a small dataset (weekly data), but not with a larger (daily) dataset. This may be expected, as more data points compensate for the collinearity between the regressors.

In a simulation study, Thompson (2012) showed that if demand is elastic and cost depends *only* on time and autocorrelated disturbances, experience curve regressions will find a (spurious) effect of experience. Nordhaus (2014) introduced a simple model, which we will describe in detail below, that clarifies how the issues of simultaneity and exogenous technological progress interact. One approach to correct for these problems has been to use instrumental variables (Söderholm & Sundqvist 2007).[3]

Our paper takes a different approach to solving these recurring problems in the literature. It exploits a unique natural experiment, to provide comprehensive evidence on the relationship between unit costs and experience in a context where demand is exogenous and the time trajectory of production is far from exponential, ameliorating problems of collinearity. However, it is beyond the scope of this paper to engage in the debates described above about which of the many channels of productivity improvement is most important. As noted by Levitt et al. (2013), the relationship between productivity and cumulative production may be incidental to a deeper causal mechanism, involving direct accumulation of knowledge at the firm level, and possibly other sources of increasing efficiency when considering sector or product level experience curves.

Instead, here we aim to provide a simpler distinction between causal factors that are well proxied by product-specific accumulated experience, and causal factors that are independent from this but are the same across all products. In the next section, we explain in more detail why the World War II context allows us to use a unique identification strategy to better isolate and estimate learning coefficients.

[3]See also Witajewski-Baltvilks et al. (2015), who found explicit conditions under which using experience as a proxy for other ultimate causal factors would still result in the estimation of coefficients that can be used in Integrated Assessment Models.

3 A unique historical context

The demand shock to military equipment production in the U.S. caused by WWII was enormous. President Roosevelt made this clear in his “Arsenal of Democracy” fireside chat on December 29, 1940, when he said,

> I want to make it clear that it is the purpose of the nation to build now with all possible speed every machine, every arsenal, every factory that we need to manufacture our defense material (Roosevelt 1940).

When war broke out in Europe in September 1939, the U.S. Army had only 189,839 men (Herman 2012). By the time of Pearl Harbor it was more than 1.4 million men strong (The National World War II Museum 2015). Defense spending increased almost 30 times between 1940 and 1942. The U.S. supplied the Soviet Union and the United Kingdom with war materiel, particularly the UK after enacting the Lend-Lease policy Act in 1941. Qualitative evidence suggests that the period between the autumn of 1939 and the American entry into World War II was important for preparing the United States for war production. This ramp-up continued throughout the war. Between 1943 and the end of the war in 1945, the U.S. government spent four times as much money on war production as it had spent in 1942 (Koistinen 2004).

However, these huge expenditures were not allocated to clear and consistent munitions orders. Instead, “munitions orders kept changing . . . as the project size of the American Armed Forces increased as the American military strategy evolved” (Carew 2009). As documented in the Army’s own history of its role in economic mobilization in World War II, “automatic supply gave way in large measure to supply on the basis of specific requisitions from theatre commanders. This permitted procurement and issue of supplies more closely tailored to specific theatre needs as indicated by operation experience and changes in strategic plans” (Smith 1959).

Support for battlefield victory at all costs continued throughout the war. From the highest office to the internal operations of the responsible bureaucracies for military production and procurement; battlefield victory - rather than cost savings - was the priority. In August 1943, the War Department Procurement Review Board was appointed to evaluate the composition, essentiality and balance of procurement programs. It noted that “*a war cannot be run like an industry; the criterion is not low costs but victory.*” (Smith 1959, p. 159). In 1940 the War Production Board set forth 12 criteria for the placement of contracts including speed of delivery, quality and price. In contrast, the War Production Board statement of March 1942 contained three ordered criteria: speed of delivery, conserving of superior facilities for the most difficult items of production, and placement of contracts with firms needing the least amount of additional machinery and equipment (Smith 1959, p. 263). This clearly shows that companies were operating at maximum capacity and that using available capacity, rather than finding the cheapest supplier, was the key driver of procurement decisions. There were no other major changes to the criteria until the defeat of Germany.

The procedures for managing costs and awarding contracts evolved during the course of the war. For products whose structure and costs of production were well known, *Fixed-Price* contracts were the government’s first preference. In situations where firms were creating new products with unknown costs the government instead used *Cost-Plus-a-Fixed-Fee (CPFF)* contracts. In 1941 contracts began to include renegotiation clauses which meant that at the conclusion of the contract a renegotiation would take place between firms and the government to assess actual costs (Smith 1959). This practice was codified with the passage of the 1942 Renegotiation Act and linked legislation that made renegotiation a condition of all contracts or subcontracts greater than \$100,000 (later revised up to \$500,000). Producers were allowed to maintain a fluidly defined "fair profit" in renegotiation, which was typically between 5 and 13 percent. Renegotiation ultimately saved the government approximately \$3 billion on 118,000 contracts (Koistinen 2004). Furthermore, it provided an incentive to conduct the detailed cost accounting that generated the data central to this paper.

While this was the general structure of contracting, procurement generally became more complex as a function of the complexity, novelty, and expense of the product being procured. A clear example of this is the well-documented contracting process for aircraft during World War II (Holley 1964). For the government to reach deals with aircraft and airplane engine makers, there was a longer initial negotiation process and then more comprehensive renegotiation than for less complex products. This allowed both parties to adjust agreed-upon terms as they learned more about production costs and predicted efficiencies in production - including those predicted by Wrights law, which the government frequently used as an analytical tool in negotiation. In short, while there are commonalities across contracting for World War II, the process was not completely homogenous.

The universal inclusion of renegotiation clauses meant that contractors had limited incentive to decrease their costs of production, as they would only be compensated for final demonstrated costs. As has been already noted in the World War II learning literature (Bajari & Tadelis 2001), there is little incentive in a cost-plus contract to supply a product more cost-efficiently. That said, while the contracts themselves do not provide clear incentives for learning, the process of defense procurement does: a high-performing supplier could expect more orders from the government (Rogerson 1994). This possibility of repeat contracts as an incentive for learning and driving costs down may limit our exogeneity argument - at least when seen at the firm-level. However, the combination of the vast scale of the war effort and the intense pressure to ramp production up quickly made controlling costs difficult and required manufacturers to operate at near maximum capacity. The vast scale of the effort was so large that the majority of manufacturing firms in the United States were engaged in war-related production, and, together with the accelerated contracting process, this limited the ability of the government to select the most efficient producers when awarding contracts.

At the sectoral level an argument can be made that cost played a role in determining which equipment was produced. The U.S. government intended to produce the most deadly weaponry at the lowest cost[4], so that the *relative* demand for products that were substitutes was not independent of price. For instance, to the extent that two types of planes could be substituted for one-another, the cheaper one may have been favored. Nonetheless, this was balanced by the fact that military strategy requires a

[4]Rohlfs et al. (2016) use evidence of substitution between military personnel and weaponry to estimate the statistical value of life.

diverse portfolio of different types of military equipment deployed in unison. The war could not be won by planes alone, but also required naval vessels, tanks, troop carriers, bullets, and a host of different types of equipment. Even within a category such as planes, military history suggests product complementarity, rather than substitutability, for instance between bombers and fighters. In any case, our data is quite comprehensive, covering individual products (which may have been substitutes) and aggregates (that are clearly not substitutes).

4 Data

Another key specific advantage of the historical context is that because many products were bought by the government, prices were recorded at a fairly detailed level and are now part of the public record[5]. We were not able to access time series of prices or unit cost at a very detailed level, but we obtained datasets that are either at a very detailed level, or are long time series, or are detailed level time series but only for labour productivity.

We thus used a number of different sources to create three datasets, as summarized in Table 1. The datasets are ordered from the most fine-grained to the most aggregated. The first dataset comprises time series of production and labor productivity for 152 types of aircraft (or plant-aircraft pairs), ships, and motor vehicles, 23 of which have never been published before. The second dataset consists of time series of production for 523 products from all categories of war materiel, paired with "early" and "late" unit costs. This dataset is new. The third dataset comprises time series for indices of contract prices for several branches of the War Department. While this last dataset was easy to collect, we are not aware of any study that has matched the relevant tables to construct experience curves from it.

We now describe each dataset more precisely, leaving the full details of data collection to Appendix A.

4.1 Labor productivity in aircraft, ships and motor vehicles

Our first dataset, which we will refer to as *Labor productivity*, includes time series of labor productivity from three different sources (see Table 2). The first source is the well-known time series of labor productivity for ships, extracted from Searle (1945). It has been already widely extended and analyzed in Rapping (1965), Thompson (2001, 2007), and Argote et al. (1990).

The second source is an extension of the data used in the study of aircraft by Alchian (1963), extracted from the source quoted in his paper, referred to as the *Source Book* (Army Air Forces 1947) .

The third source is entirely original. We have collected a novel set of time series from Ford's archive. This covers 23 products, mostly motor vehicles. Appendix

[5]The lack of availability of prices at the most granular level is a well-known problem for estimating productivity (Klette & Griliches 1996, Foster et al. 2008).

Dataset	Sources	N	Time span	Cost data	Aggregation
Labor Productivity	*Source Book*, Searle (1945), and Ford archives	152	01/1940 to 11/1945, $T \in [2, 64]$	Manhours per unit	Plant or product
OMPUS-USMH	USMH and OMPUS	523	08/1942 to 08/1945, $T = 2$	"Early" and "Late" "Standard Dollar Weight" per unit	Product
Contract Prices	Crawford & Cook (1952)	10	01/1942 to 08/1945, $T = 44$	Index of contract prices	War (sub) departments

Table 1: Summary of the three datasets. N is the number of time series and T is the number of observations in each time series. Time span refers to cost data. In each dataset, we have production data starting in January or July 1940.

A.1 further discusses data collection and transcription issues for each of these three sources.

Source	Type of products	N	Aggregation level
Searle (1945)	Ships	5	Product
Source book	Aircrafts	124	Plant
Ford's archives	Mostly motor vehicles	23	Product

Table 2: Sources of the *Labour productivity* data.

4.2 Total unit costs at the product level

Our second dataset, which we will refer to as *OMPUS-USMH* or *OMPUS* for short, was created by combining two archival publications. The first publication, *Official Munitions Production of the United States, by Months, July 1, 1940 - August 31, 1945 (OMPUS) Handbook* (Civilian Production Administration 1947), provides information about the monthly output of particular categories of munitions and specific products. The second publication is the *United States Munitions Handbook* (USMH), which provides the costs of production, in terms of "Standard Dollar Weights" at two moments–one at the beginning of the war and one towards the end– for many of the products recorded in the OMPUS. While the OMPUS is available online, the formerly classified USMH was located in the National Archives by one of us (D.G.).

Matching OMPUS to the USMH was relatively easy because each product listed in the USMH includes the page and column to which it should be matched in the OMPUS. One of the nice features of these data is that they are nearly comprehensive. The OMPUS publication was intended to record the "munitions figures of *all of the* *procurement services*" (our emphasis), and while the USMH does not report cost values for the majority of products, the coverage appears reasonably representative.

The main issue we had with the dataset concerned the dates at which the costs were recorded; the dates in the USMH are the same for every product category, but some products started production after the "early" date, or ended production before the "late" date – in this case we assumed, guided by the USMH explanatory notes, that the early/late dates correspond to the start/end of production. Appendix A.2 provides a more detailed discussion of our interpretation of early and late costs and Standard Dollar Weights.

While the OMPUS provides production data, it does not provide data on experience, meaning cumulative production starting from a level of initial experience. It is impossible to know what previous manufacturing experience was for each product listed in the OMPUS. While other studies of experience curves have not corrected for prior experience or have applied a correction based on the assumption of exponential growth, we opted for an approach that corrects at the level of semi-aggregated categories, for which we could obtain rough estimates or introduce one ourselves. For each product i, we estimated prior experience as a proportion of total war production experience. More precisely, we estimated initial experience as a category-specific factor ζ_i times the total amount of known production of product i during the war $\sum_{\tau=1}^{T} Q_{i,\tau}$. Denoting experience $Z_{i,t}$, we have

$$Z_{i,t} = \zeta_i \sum_{\tau=1}^{T} Q_{i,\tau} + \sum_{\tau=1}^{t} Q_{i,\tau}, \tag{1}$$

where T is the total number of months from January 1940 to August 1945. The first term represents estimated initial experience, and the second term is the usual experience computed as observed cumulative production up to time t.

Because these products are too detailed, we attempted to estimate the factors ζ_i at a more aggregated level, using the 81 most detailed levels of the table of contents of the OMPUS. The ζ_i and the extensive discussion of how we estimated them are provided in Appendix B. To give one example, consider fighter aircraft. We have 12 models of fighters in the USMH-OMPUS data, and we cannot know initial experience in each model. However, from other sources we were able to estimate that around 300,000 planes were produced during WWII, and 60,000 before. Therefore, we applied a value of $\zeta \approx 60,000/300,000 = 0.20$ to all aircraft models. For example, 680 units of the P–61 Black Widow model (OMPUS ref. 15/3) were produced during the war, so constructed the experience variable as if $0.20 \times 680 \approx 136$ models had been produced before.

4.3 Contract prices at the war department level

Our last dataset, which we will refer to as *Contracts*, contains contract price indices and is taken from the statistics on procurement in Crawford & Cook (1952). This data allows us to construct experience curves for the period January 1942 - August 1945 at several different levels of aggregation. The data is for the eight different branches of the War department. The highest level of aggregation is the entire War department.

It is then subdivided into Air Forces (AAF) and the Service Forces (ASF). The latter is then subdivided into seven technical services, which "were the operating agencies of the War Department in all its supply activities" (Smith (1959), p.114).

The data on production, while expressed in millions of dollars, can be taken as indicative of production *volume* because it has been generated from a large sample of physical quantities, evaluated at 1945 prices, rather than at the prices in force during each month. The indices of contract price changes were computed from price changes to contracts for individual products, representing about half of the total value of War procurement. As for the OMPUS data, we estimated prior experience as a category-specific factor times the total level of the wartime production. The detailed justification for the correction factors applied to this data is also given in Appendix B.

5 Empirical framework

Our objective is to estimate the effect of experience on cost. As discussed above, simply regressing cost on experience, as is done in learning curve studies, suffers from several related issues, all of which are absent in the context of WWII military production.

To restate the issues, because demand is generally elastic, the relationship between price (or cost) and production (or experience) embodies both demand elasticity and the causal effect of experience on cost. However, while demand relates price/cost and production, the effect of experience is between cost and *cumulative* production. It is common that production grows exponentially, so cumulative production grows exponentially too. Because of this, if we use production and cumulative production interchangeably, we can think of the issue as a problem of simultaneity between cost/price and production/experience (Nordhaus 2014), even though the mechanisms are in principle from experience to cost and from price to production.

Second, it is easy to argue that cost reductions cannot possibly come only from the growth of product-specific experience, whatever this proxies. The simplest possible way to acknowledge omitted variables is simply to allow for an exogenous exponential time trend, to capture overall economy-wide trends. However, as discussed in the previous paragraph, experience often grows exponentially and in a smooth manner (Lafond et al. 2018), making it hard to distinguish the effect of experience from the effect of an "exogenous" exponential time trend. Technically speaking this is just an issue of imperfect multicollinearity. In the specifications in first-differences, where the exogenous time trend is simply the intercept of the regression, this is just an issue of limited variance of the regressor. While in principle this does not bias the coefficients, including both experience and a time trend usually makes standard errors extremely large and gives unstable results, as discussed in the literature review (Section 2). It is therefore a serious issue.

5.1 Nordhaus's model

Nordhaus (2014) introduced a very simple model that illuminates the issues above and helps clarify why demand exogeneity makes the identification of the effect of experience on costs possible.

Let Q_t denote production at time t and let Z_t denote cumulative production. Assume the unit cost function c_t depends on both experience Z_t and an exogenous time trend, with the functional form

$$c_t = c_0 Z_t^{-b} e^{-at}. \tag{2}$$

The parameters a and b are expected to be positive, since we generally observe technological progress. Assume that production is equal to demand and price is equal to unit cost. Furthermore assume that demand has constant elasticity $\epsilon > 0$ and that there is an exogenous and growing demand e^{dt}. This gives[6]

$$Q_t = D_t = D_0 c_t^{-\epsilon} e^{dt}. \tag{3}$$

By letting Δ represent the time difference operator $\Delta x = x_t - x_{t-1}$, Eqs. 2 and 3 can be written as

$$\Delta \log c = -a - b\Delta \log Z \tag{4}$$

and

$$\Delta \log Q = -\epsilon \Delta \log c + d. \tag{5}$$

Now suppose that production grows exponentially. Then experience also grows exponentially at the same rate, that is

$$\Delta \log Q \approx \Delta \log Z.$$

Using this in Eq. 5, the solution of the system Eqs. 4-5 is

$$\Delta \log c = \frac{-a - bd}{1 - b\epsilon}, \tag{6}$$

$$\Delta \log Z = \Delta \log Q = \frac{a\epsilon + d}{1 - b\epsilon}. \tag{7}$$

Nordhaus's (2014) critique is that experience curve studies typically assume that the logarithm of experience is the single explanatory factor for the logarithm of costs, i.e. they assume that

$$\Delta \log c = \beta \Delta \log Z, \tag{8}$$

and perform a regression of the logarithm of costs against the logarithm of experience in order to measure the empirical parameter β. Inserting Eqs. 6 and 7 into Eq. 8, we can write the parameter β as the ratio

$$\beta = \frac{\Delta \log c}{\Delta \log Z} = \frac{-a - bd}{a\epsilon + d}. \tag{9}$$

It is clear that in general the empirically estimated parameter β not only depends on the experience parameter $-b$, but it also depends on the exponential trend parameter a for costs, the exponential trend parameter d for exogenous demand growth, and the

[6] Nordhaus (2014) also includes population growth in the exogenous demand term.

demand elasticity ϵ. The estimated β can be interpreted as being purely the effect of experience only when there is no exogenous time trend, i.e. if $a = 0$ then $\beta = -b$.

Let us now consider the case in which demand is completely exogenous ($\epsilon = 0$), but for simplicity still growing exponentially. In this case Eqs. 6 and 7 become

$$\Delta \log c = -a - bd, \tag{10}$$

and

$$\Delta \log Z \approx \Delta \log Q = d, \tag{11}$$

so that putting Eq. 11 into 10 gives

$$\Delta \log c = -a - b\Delta \log Z, \tag{12}$$

which is the same as Eq. 4. If demand is completely exogenous, we can estimate the supply equation directly[7]. A simple first-difference regression model for the logarithm of cost against the logarithm of experience with an additive constant is able to separate the effect of the exogenous trend parameter a and the "learning" parameter b.

While illustrative of the reverse causality issue, this model is not stochastic and all variables grow exponentially, so it cannot do justice to the other important issue, multicollinearity. If production grows exponentially with fluctuations, since computing experience from production involves integrating and therefore smoothing, Z will increase exponentially but with much lower fluctuations, so the term $\Delta \log Z_t$ will be approximately equal to the constant d, for all t[8].

Fortunately, we will find empirically that during WWII demand exogeneity led to production patterns that are *not* exponential, making the correlations between the regressors (time, production and experience) much lower than in other studies using recent data.

5.2 Main specification

It is customary in the literature to motivate the main specification for learning curve regressions by starting from the standard Cobb-Douglas production function. This does not assume exponentially increasing production, it allows for economies of scale and it can be used with data on labor productivity[9]. With constant technical parame-

[7] Pozzi & Schivardi (2016) used survey estimates of demand elasticity coupled with data on firm-level prices to separate the effects of demand and productivity on firm growth.

[8] To give a concrete example, if production follows a geometric random walk with drift r and noise σ, then the change in the log of cumulative production has a variance approximately equal to $\sigma^2 \tanh(r/2)$ (Lafond et al. 2018), which is much smaller than σ^2. For instance, if we consider the Air Force time series from the *Contract prices* data, taking the period where we have monthly data and the growth is roughly exponential (January 1942 to January 1944), we have $\sigma = 0.047$ and $g = 0.068$. The measured variance of the change in the log of *cumulative* production is an order of magnitude less than σ, 0.0063, fairly close to the prediction from the formula above, 0.0087.

[9] In this model, optimal requirements for labor and capital grow at the same rate. This may not be an appropriate assumption, especially during WWII where there was an important build-up of the capital stock (Gordon 1969). However, a more general framework allowing for factor-biased technological change would require more data for estimation, so we limit ourselves to the Cobb-Douglas case here.

ters and a time-varying Hicks-neutral productivity factor A_t that depends on experience and an exogenous time trend, as in Rapping (1965), we have

$$Q_t = A_t K_t^{\theta_k} L_t^{\theta_l},$$
$$A_t = Z_t^b e^{at}, \tag{13}$$

where K_t and L_t are capital and labor in period t and θ_k and θ_l are constants, with an additional term for an exponential exogenous productivity improvement as in Nordhaus (2014).

If firms minimize their total cost subject to the constraint of fixed *exogenous* demand $Q_t = \bar{Q}_t$, and factor prices are constant, the optimal labor demand per unit of production can be written

$$\log \frac{L_t^*}{\bar{Q}_t} \equiv \log l_t = B_L - (a/s)t - (b/s)\log Z_t + (1/s - 1)\log \bar{Q}_t, \tag{14}$$

where B_L is a constant and where $s = \theta_k + \theta_l$ represents economies of scale. Since we have data on the unit labor requirements $l_t = L_t^*/\bar{Q}_t$ in our *Labor productivity* dataset, we can estimate this equation. In the two other datasets, we have data on total unit costs $c_t = C_t/\bar{Q}_t$. Inserting optimal factor requirements in the total cost function $C_t = wL_t^* + rK_t^*$ gives

$$\log \frac{C_t}{\bar{Q}_t} \equiv \log c_t = B_C - (a/s)t - (b/s)\log Z_t + (1/s - 1)\log \bar{Q}_t, \tag{15}$$

where B_C is another constant. Apart from the constants, the two equations (14 and 15) are the same. They become exactly the same if we estimate a model in first differences, that is

$$\Delta \log c = \Delta \log l = -(a/s) - (b/s)\log Z_t + (1/s - 1)\log \bar{Q}_t. \tag{16}$$

Our favored specification is to estimate Eq. 16 in a pooled panel. However, each dataset presents specific patterns of missing data, and thus different opportunities for robustness checks and alternative specifications - typically fixed effects for the levels (Eqs. 14 and 15), and separate regressions for each time series. Thus we present our detailed econometric specifications separately for each dataset.

6 Empirical Results

We now present results from performing regressions on the three data sets here. A point of prior optimism comes from examining the correlations between the independent variables. As shown in Table 6, they range from about 50% to 90%. Although this might seem high, they should be compared with a systematically higher correlations in other contexts, such as the chemical industry numbers reported by Lieberman (1984), which are all greater than 90%.

6.1 *Labor productivity* in aircraft, ships and motor vehicles

A typical model specification (Lieberman 1984) for analyzing a panel of experience curve time series is the fixed effects model

$$\log l_{it} = \kappa_i + \alpha t + \beta \log Z_{it} + \gamma \log Q_{it} + \eta_{it}. \tag{17}$$

Comparing to Eq.14, $\kappa_i = B_L$, $\alpha = -a/s$, $\beta = -b/s$ and $\gamma = 1/s - 1$. The results of this model applied to the *Labor productivity* dataset are reported[10] in Table 3. Including only experience we find that $\beta \approx -0.33$, which implies a drop of cost of 20% for each doubling of experience. This number coincides with the original observation by Wright and is seen in many other studies of products such as airplanes. The estimated coefficient for the exogenous time trend is not very statistically significant, but its value is relatively high: given that the data is monthly, a negative coefficient of 0.004-0.006 implies an annual rate of exogenous technological progress of about 5 to 7 percent. Finally, including production indicates very mild and not very significant economies of scale. Throughout the paper we will find no convincing evidence of non-constant returns to scale [11].

Table 3: Panel regression results for *Labor Productivity*

	Fixed Effects			First Differences		
Experience	-0.326***	-0.304***	-0.276***	-0.253***	-0.217***	-0.203***
	(0.017)	(0.020)	(0.024)	(0.020)	(0.022)	(0.025)
Time		-0.004	-0.006		-0.022***	-0.024***
		(0.003)	(0.003)		(0.004)	(0.005)
Production			-0.036*			-0.010
			(0.015)			(0.007)
N	3034	3034	2981	2830	2830	2740
R^2	0.75	0.75	0.75		0.13	0.13

A typical feature of experience curves regressions is that they have highly autocorrelated residuals (Thompson 2012). We ran the test of Wooldridge (2002) for AR(1) residuals in short panels and find in all three cases very strong evidence against the null of no first-order autocorrelation. This suggests estimating a model with autocorrelated disturbances, or a model with lags of the dependent variable, which we do in Appendix D.1. But since we are not interested in the intercepts, and since from a theoretical point of view we expect that it is the change in experience that induces a

[10] Throughout the paper, we report significance at the 5, 1 and 0.1% levels. Standard errors are clustered by product. For the fixed effect models we report the "within" R^2; for first-difference models with no intercept (no "time" variable), we do not report the R^2 because it cannot be interpreted as the explained proportion of variance.

[11] The number of observations is lower because we removed months during which production was null.

change in cost, we favor a first-differences specification[12], i.e.

$$\Delta \log l_{it} = \alpha + \beta \Delta \log Z_{it} + \gamma \Delta \log Q_{it} + \eta_{it}. \tag{18}$$

The role of experience is fairly consistent with the fixed effects regression but it indicates a larger role for the time trend. The coefficient β now ranges from -0.25 when only experience is included to -0.20 when all independent variables are included and remains solidly statistically significant. However, the coefficient α for the exogenous time trend now has a value of -0.024 when all variables are included, corresponding to a much larger annual improvement rate of about 29%, and is now strongly statistically significant. As we will discuss in Section 7, this represents a third of the rate of cost decrease, with the other two thirds explained by experience. When including production, the coefficient γ is now very small and statistically insignificant.

Finally, we present the results of independent regressions for each individual product or plant, removing all time series with less than 10 available cost values (so 118 of 152 time series remain). Since the coefficient for production is so small in the aggregate regression we exclude the production variable here in the interest of parsimony. The equation estimated is

$$\Delta \log l_{it} = \alpha_i + \beta_i \Delta \log Z_{it} + \eta_{it}. \tag{19}$$

Fig. 2 compares the estimated coefficient $\hat{\beta}$ for the regression Eq.19 when the exogenous time trend is included to the case where it is not, that is, we restrict $\alpha_i = 0$ and Eq. 19 becomes a regression through the origin. We observe a great deal of heterogeneity in experience rates, but the results are broadly consistent with the aggregate regression. The variation is a mixture of intrinsic differences in real progress rates and statistical fluctuations due to the fact that the individual regressions are based on short time series. With some exceptions, most of the data is clustered along the identity line, demonstrating that the coefficient estimated for experience is not overly sensitive to the restriction $\alpha = 0$. This stands in sharp contrast with the results reported by Nordhaus (2014). Fig. 2 in Nordhaus (2014), which is analogous to our Fig. 2, shows coefficients that are often extremely far from the "reasonable" range $\beta \in (0, -0.5)$, and which change by more than an order of magnitude when a time trend is included. We believe that the main difference is that in his data many of the production time series (and thus experience time series) are close to following an exponential trend, creating extreme problems of multicollinearity.

In summary, for the *Labor productivity* dataset we observe a strong and clear relationship between declining labor costs and experience, with exponents $\beta \approx -0.2$. We also show a significant role for the exogenous time trend, whose overall effect is comparable to that of experience (Section 7 will provide an estimate of the relative importance of each variable for all the datasets).

[12]Unfortunately, our panel is strongly unbalanced and contains missing observations, limiting our options for unit root and cointegration tests, but see Appendix D.1. Note also that since the dataset contains missing values, we lose more observations than the number of products (152) when taking first differences.

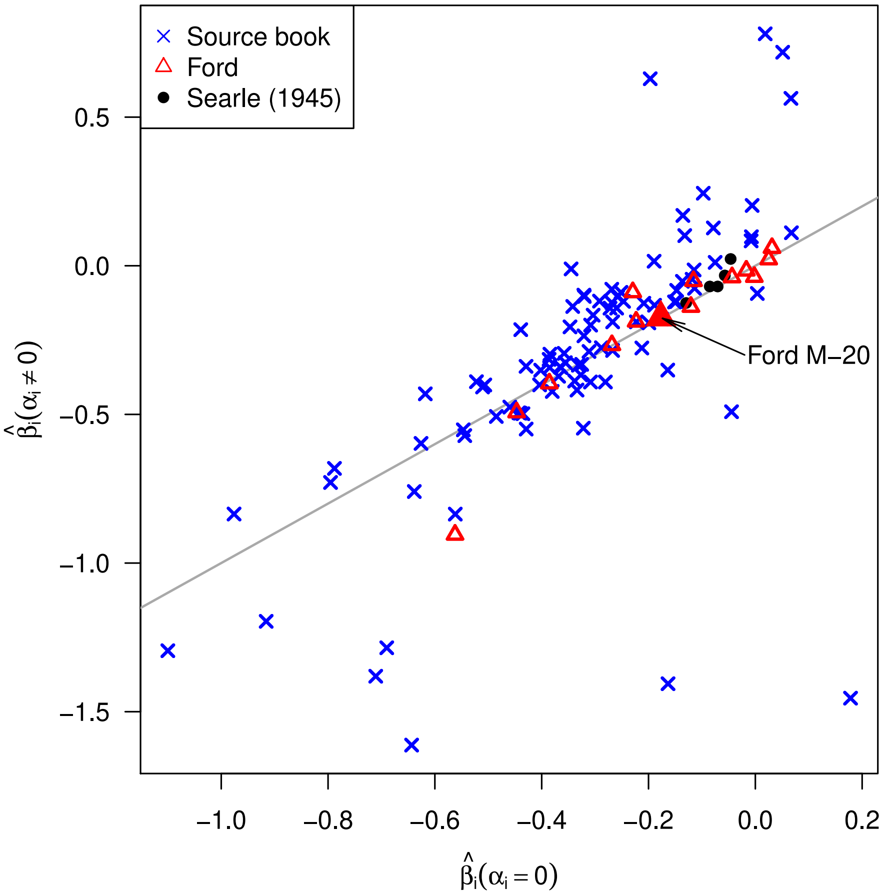

Figure 2: The coefficient β_i for experience when the time trend is included is plotted against the same coefficient when the time trend is not included, using Eq. 19. The grey line is the identity. Three values of β_i which were greater than 2 in absolute value have been excluded. The Ford Armored Car discussed in the introduction (Fig. 1) is shown as a slightly larger red triangle.

6.2 *OMPUS-USMH*: unit costs at the product level

We now turn to the *OMPUS* dataset, which gives total unit costs instead of labor productivity and contains a more complete sampling of military products. A limitation is that we observe the cost for each product at only two different dates, and these dates differ across products. Nonetheless, because we assume that the parameters are the same for all products, we can exploit cross-sectional heterogeneity. We do this in two ways.

First, we are still able to estimate a fixed-effects model, but we also interpret it as a "differences" estimator. As is well-known, when $T = 2$, the first-differences and the fixed-effects estimators are the same. Appendix C shows that this result holds when the differences are heterogenous, in the sense that the span of time between the two observations is different from one product to another. We perform inference and report R^2 from the fixed-effects.

Table 4 shows the results of the regression, first including only experience, and

Table 4: Panel regression results for *OMPUS-USMH*

	Fixed-Effects/Heterog. Differences			Growth rates cross section		
Experience	-0.098***	-0.055**	-0.058**	-0.086**	-0.079	-0.098
	(0.015)	(0.017)	(0.019)	(0.031)	(0.040)	(0.052)
Time		-0.004***	-0.005***		-0.002	-0.002
		(0.001)	(0.001)		(0.003)	(0.003)
Production			0.008			0.024
			(0.009)			(0.044)
N	1046	1046	1002	523	523	482
R^2	0.13	0.17	0.19		0.06	0.06

then including an additional time trend and production. The coefficients for experience are statistically significant, but are smaller than for the *Labor productivity* time series, ranging from -10% when the other variables are not included to about -5.5% when time is included. The coefficients for the exogenous time trend are also significant, and about 0.4% per month or 5% per year (about 60% of the rate of cost decreases). As before the coefficient for production is small and statistically insignificant.

Second, since with two observations we can compute an average growth rate, we perform a cross sectional regression of the growth rates (columns 4-6, see Appendix C, Eq. 25 for details). The results are similar, although with a sensibly higher coefficient for experience and lower statistical significance.

The upshot is that the results for the *OMPUS-USMH* dataset are broadly consistent with those of the *Labor productivity* dataset, albeit with smaller experience exponents.

6.3 *Contract prices* at the war department level

Fig. 3 shows costs against experience in our last dataset on contract price indices at the War Department level. “Total” (on the right) represents the entire War Department, which is divided into Army Air Forces (AAF) and Army Service Forces (ASF); the latter is subdivided into the other 7 services. Apart from Quartermaster, which saw its price *increase* over the period,[13] the other services have a relatively similar slope.

In Fig. 4 we show the experience coefficient obtained from separate regressions for each individual time series, with and without the exogenous time trend. We make four observations that will be useful as a background to motivate the panel specification and understand its result.

[13] Quartermaster contains common goods such as clothing, which were produced extensively prior to the war.

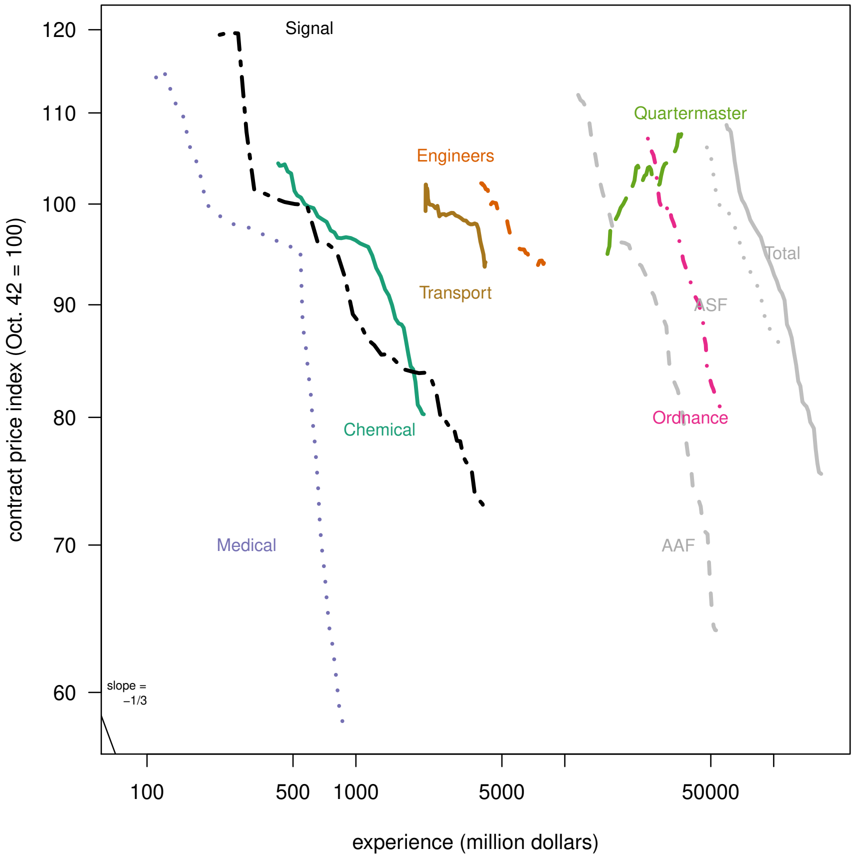

Figure 3: Experience curves for the war departments and component agencies.

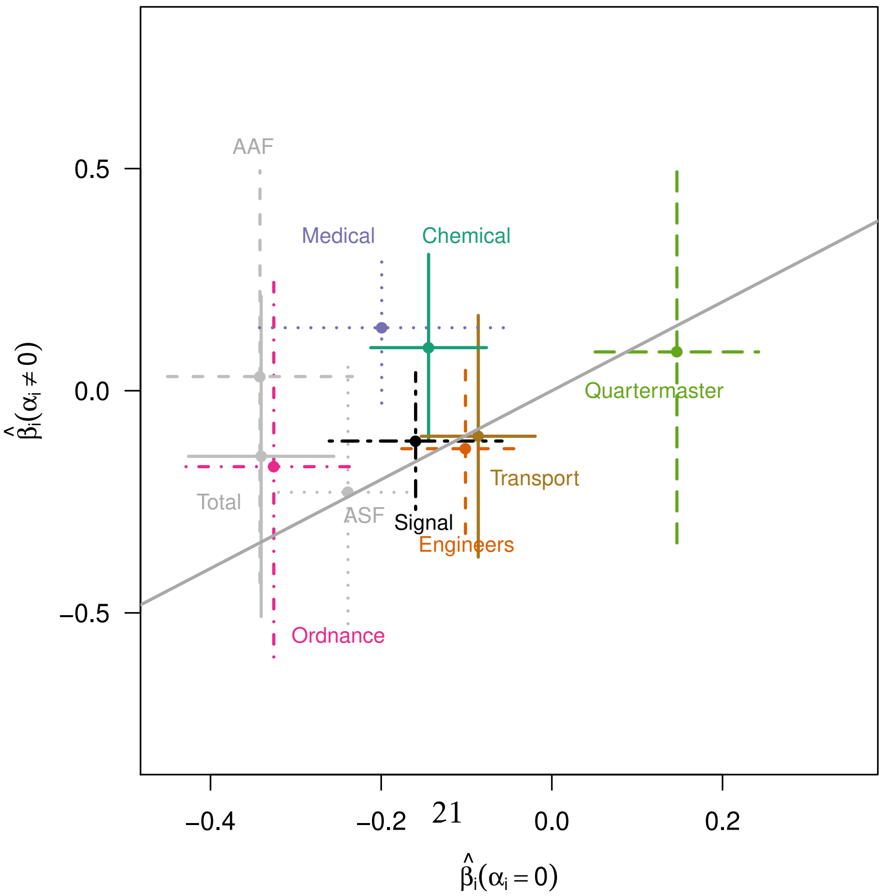

Figure 4: Estimated coefficient of the first difference regression of the log of contract prices on the log of experience, including or excluding an exogenous time trend. The lines show plus or

First, in most cases, we see that standard errors are higher when the time trend is included, again reflecting the multicollinearity problem. Second, for some sub-services (Ordnance, Signal, Transport and Engineers), the coefficient is only slightly affected by the inclusion of the time trend, which in this case confirms that the lack of correlation allows us to pin down the relative effects of the exogenous trend and experience. However, for some other sub-services (Medical, Chemical), the estimated coefficient changes so much that its sign flips. Third, when the time trend is included, all subservices have confidence intervals overlapping zero, indicating a lack of statistical significance. Fourth, the confidence intervals of all subservices largely overlap, suggesting that a panel regression assuming a unique slope could give sensible results.

Because Quartermaster has such an obviously different slope in Fig 3 (on the top right), we remove it from the dataset and present panel regression results for 7 time series (AAF and 6 sub-services of ASF).

Table 5: Panel regression results for *Contracts*

	Fixed Effects			First Differences		
Experience	-0.206**	-0.150*	-0.165**	-0.189***	-0.106	-0.118*
	(0.037)	(0.051)	(0.044)	(0.024)	(0.045)	(0.040)
Time		-0.003	-0.003		-0.004	-0.004
		(0.003)	(0.002)		(0.003)	(0.003)
Production			0.040*			0.006*
			(0.016)			(0.002)
N	308	308	308	301	301	301
R^2	0.77	0.78	0.81		0.05	0.06

Table 5 shows the results. In line with previous estimates, we find an economically significant effect for both experience and the time trend, and a weak effect for production[14].

7 Discussion

In this section we take stock of the results obtained from the three datasets, and briefly summarize our robustness checks.

7.1 Summary of the results

To provide an overview of our results we have gathered some key statistics in Table 6. The first three columns show the average growth rates of the variables, computed as the average one period change in their logarithm, pooling all observations for which

[14]Note that statistical significance levels are based on hypothesis testing using the Student distribution with $G-1$ degrees of freedom, where G is the number of clusters, here equals to 7.

both $\Delta \log c$ and $\Delta \log Z$ are available.[15] The next three columns show the Pearson correlations on pairwise complete observations between the "within" transformation of the regressors. The next two columns report the coefficients for the main regression specification, which includes experience (coefficient β) and time (coefficient α) and is estimated in first-differences (fixed-effects or equivalently "heterogenous differences" for the *OMPUS*, see Appendix C). The last column shows the ratio of α to the average growth rate of cost, which is the share of cost decrease accounted for by the exogenous trend[16].

	Growth rates			Correlations			Coefficients		Share exo.
	c	Z	Q	Q,t	Q,Z	Z,t	β	α	$\alpha/g.r.(c)$
Labor Productivity	-6.6	20.1	4.4	46.9	78.2	80.8	-21.7	-2.2	33.3
OMPUS-USMH	-0.8	7.4	5.2	72.3	70.5	59.2	-5.5	-0.4	60.4
Contract Prices	-0.8	3.5	3.0	62.9	56.5	85.7	-10.6	-0.4	54.3

Table 6: Descriptive statistics and main results. All values are multiplied by 100. Growth rates are monthly.

As is evident from the first column of the table, costs in the *Labor productivity* dataset, which are unit labor requirements, decline more than seven times faster than either of the two unit costs. One explanation is simply that improvement rates are heterogenous and the *Labor productivity* data comes overwhelmingly from the rapidly improving aircraft industry, while the other datasets on unit costs have a much wider and balanced coverage. Another explanation could be that capital deepening during the period improved labor productivity more than overall productivity.

The three columns giving correlations between the independent variables demonstrate that multicollinearity is a limited issue in our data. Although many of these correlations are high, they are still much lower than those reported in most previous studies (see for instance Table 2 in Lieberman (1984)).

The following two columns provide an overview of our regression results. In every case these results are consistent, and highlight the substantial role played by both experience and an exogenous time trend. Given the larger rate at which cost declines in the *Labor productivity* dataset, it is not surprising that both effects are substantially larger.

In the last column we divide the exogenous trend rate α by the overall rate of cost decline (the figures in the first column of the table). This ratio is roughly a third for the *Labor Productivity* dataset but 60% for the *OMPUS-USMH* and a half for the *Contract Prices* datasets.

The overall conclusion is that we see a substantial effect for both experience and an exogenous time trend, in roughly equal proportions. Once we control for time and

[15] For the *OMPUS*, only two observations per product are available, at the starting and ending points t_0 and t_1, so the growth rates shown are averages over all products of the monthly growth rates $(\log x_{t_1} - \log x_{t_0})/(t_1 - t_0)$. In the *Labor productivity* dataset, in contrast to other datasets, the growth rate for experience is much larger than the growth rate of production, partly because in this dataset we did not correct for initial experience.

[16] For the *OMPUS*, this is computed based on the fixed effect equation, see Appendix C.

experience, the effect of monthly output is negligible. When multicollinearity is not as severe as in other settings, and enough data are pooled together, the estimates are relatively stable across specifications.

7.2 Robustness of the results

We discuss the robustness of the results to changes in econometric methodology, assumptions on initial experience, and omitted variables. The details of the robustness checks, when possible, are in Appendix D.

Time series analysis (Appendix D.1). Overall, we tend to find similar results to those reported in the text if we a) include time fixed effects instead of an exponential trend, b) replace experience by its lagged value, c) test for times series properties and estimate the models suggested by these tests.

Heterogenous coefficients (Appendix D.2). We allowed for fixed effects in the first difference regression, thereby allowing time trends to differ. This had no effect on the *Labor Productivity* dataset, but for the *Contracts* data the effect of experience disappears completely, although standard errors are large enough that we cannot reject a value such as the one estimated in the main text. Estimating Swamy's (1970) or Pesaran & Smith's (1995) heterogenous coefficient models gives the same outcome, with robust results for *Labor productivity* but not for *Contracts*.

Instrumenting by lagged values (Appendix D.3). Our dependent variable is a total (either manhours or cost), divided by output. Thus output appears on both sides of the equation. We instrumented output and experience by their lagged value, and found that this does not change the results.

Alternative assumptions on initial experience (Appendix D.4). We re-estimated our models using different assumptions about initial experience. Overall, it is clear that changing the levels of initial experience changes the results, often in the direction of attributing a larger role to the exogenous time trend. Nonetheless, under a very wide range of possible corrections, the signs remain negative. Without implausible changes to our estimates of initial experience, the share of the exogenous time trend cannot exceed about 70%.

Increasing quality and design changes. Technological progress may increase the quality of a product rather than reduce its cost. This is a limited issue here, as many qualitative accounts of war production describe it as mass production of fixed designs, although not without tensions[17]. For instance, Best (2018) describes how mass production techniques were implemented for producing aircraft, which then were sent

[17]Design changes requirements were cited by Ford as one of the reason for the delays in reaching full speed production of the B-24 Liberator plane at the famous purpose-built large-scale Willow Run plant in Detroit (Baime 2014). At General Motors, Keizer had to implement ”flexible mass production” techniques (Herman 2012). Mishina (1999) reports 8 versions of the B-17 “Flying Fortress” since 1934,

to Air Force depot and modification centers for fitting the weapons, which are more susceptible to design changes. These overall limited design changes suggest that our study may not suffer too much from missing data on quality, but also somewhat limit the external validity of our results, as many experience curves relevant to current-day policies include the production of successive designs.

Research and development. We were not able to find monthly, product-level Research and Development (R&D) data. R&D, and especially federal R&D, was important during the war and grew very fast. For instance, the modification centers mentioned above were themselves linked to the Office for Scientific Research and Development and to the National Advisory Committee for Aeronautics (ex NASA), the major institutions organizing war-related R&D. If we consider that cumulative R&D capital lowers costs, our estimates of *both* α and β are very likely to be biased upwards, because one would expect the coefficients of the auxiliary regression of R&D on time and experience to be positive. However, military R&D was often oriented towards *product* innovation. Thus, although it may have decreased the cost per quality-adjusted unit, the non-adjusted unit cost may have been affected *upwards*. As a result, we do not believe that our estimates are driven by an omitted variable bias. However, because we are silent on the role R&D or on ultimate causal factors more generally, our study is unable to provide a more fine-grained policy recommendation.

Comparing *USMH* and *Contracts* (Appendix D.5). In principle, the raw data for these two datasets largely overlaps, so they should give similar results if the *USMH* data can be aggregated at the same level as the *Contracts* data. We used our own concordance table to test whether estimating the time and experience parameters at the war department level in each dataset would give similar results. We found only moderate agreement, but the standard errors are large, so we cannot exclude that this is due to low sample size.

Controlling for inflation (Appendix D.6). An obvious factor in determining cost is the price of inputs, which we are unable to observe for each product. We used the producer price index for all commodities as a control, as a single regressor or interacted with individual dummies in an attempt to acknowledge that different products have different input mixes. Our main results do not change.

External learning (Appendix D.7). We attempted to estimate spillover effects by adding the growth of aggregate production experience (total monthly war spending) as a regressor. We did this for both production and cumulative production. The effects of these variables are not consistent and rarely significant, and leave our estimates for the effects of own experience unchanged.

but note that the design was frozen at model B-17E, after only 134 of a total of 6981 B-17 had been built. Changes between the B-17E and the B-17 F&G were small changes intended to accommodate the changing fitted weapons, which themselves featured continuous and important upgrading.

Depreciation of experience (Appendix D.8). The literature has emphasized that part of past experience may become irrelevant. We constructed experience variables using a retention factor δ, that is $Z_t = \delta Z_{t-1} + Q_t$. We found that allowing for $\delta < 1$ tends to increase the share of cost decrease due to exogenous factors. However, even when this improves the fit (mostly for *Labor Productivity*), the improvement is not important and the share of exogenous progress remains relatively close to our main estimates.

8 Conclusion

Experience curves have been widely applied and remain an important tool for predicting technological progress conditional on deployment. This makes them essential in many applications where modeling endogenous technological progress is necessary, such as economic models of climate change, where one needs to model the balance between the benefits of investing now against the benefits of waiting for exogenously improving backstop technologies.

A major issue in using experience curve models for policy purposes is that the estimation of causal partial effects is plagued not only by problems of endogeneity, as is widely acknowledged, but also by the fact that experience is the cumulative of a variable that grows close to exponentially, and therefore is itself very close to a deterministic exponential trend. This makes it very hard to distinguish “endogenous” technological progress as proxied by experience, and exogenous technological progress as proxied by a deterministic trend.

In this paper we have shown that military production during WWII offers a context that solves both issues. Because the demand for weapons was driven by battlefield needs, which grew and then shrank, not only was it exogenous, but production, experience and time were relatively uncorrelated. As a result, we found that, in contrast to other studies, estimating the partial effect of experience on cost was relatively robust to the inclusion of a time trend. Despite remaining issues of data quality, we conclude that both experience and other effects captured by an exogenous time trend were quantitatively important during the War.

A point that deserves some emphasis is that the effect of experience on cost differs among products. As a result, when assessing a policy intervention for a given product, one must first determine whether there is historical evidence that increasing experience is associated with decreasing cost. With this caveat about heterogeneity in mind, our study broadly confirms that investing in a specific product in order to stimulate experience will cause production costs to decrease. However, our study also finds that a large share of technological progress cannot be attributed to product-specific experience. This suggests that to make targeted investments in a particular technology fully effective, policies should also aim at fostering the broader innovation system.

References

Alchian, A. (1963), 'Reliability of progress curves in airframe production', *Econometrica* **31**(4), 679–693.

Arellano, M. & Bond, S. (1991), 'Some tests of specification for panel data: Monte carlo evidence and an application to employment equations', *The Review of Economic Studies* **58**(2), 277–297.

Argote, L., Beckman, S. L. & Epple, D. (1990), 'The persistence and transfer of learning in industrial settings', *Management Science* **36**(2), 140–154.

Argote, L. & Epple, D. (1990), 'Learning curves in manufacturing', *Science* **247**(4945), 920–924.

Army Air Forces (1947), *The source book of World War II basic data. Airframe industry, Vol. I, direct man-hours - progress curves*.

Asher, H. (1956), Cost-quantity relationships in the airframe industry, Technical report, Santa Monica, CA: RAND Corporation.
URL: *https://www.rand.org/pubs/reports/R291.html*

Ayres, L. P. (1919), *The War with Germany: A Statistical Summary*, Government Printing Office.

Bahk, B.-H. & Gort, M. (1993), 'Decomposing learning by doing in new plants', *Journal of Political Economy* **101**(4), 561–583.

Baime, A. J. (2014), *The arsenal of democracy: FDR, Detroit, and an epic quest to arm an America at war*, Houghton Mifflin Harcourt.

Bajari, P. & Tadelis, S. (2001), 'Incentives versus transaction costs: A theory of procurement contracts', *RAND Journal of Economics* pp. 387–407.

Baltagi, B. H. & Wu, P. X. (1999), 'Unequally spaced panel data regressions with AR(1) disturbances', *Econometric Theory* **15**(6), 814–823.

Benkard, C. L. (2000), 'Learning and forgetting: The dynamics of aircraft production', *The American Economic Review* **90**(4), 1034.

Best, M. H. (2018), *How Growth Really Happens: The Making of Economic Miracles through Production, Governance, and Skills*, Princeton University Press.

Breitung, J. (2001), The local power of some unit root tests for panel data, *in* 'Nonstationary panels, panel cointegration, and dynamic panels', Emerald Group Publishing Limited, pp. 161–177.

Broadberry, S. & Harrison, M., eds (2005), *The Economics of World War I*, Cambridge University Press.

Cain, L. P. (2006), *Historical Statistics of the United States, Millennial Edition*, chapter Domestic sales, imports, and exports of passenger cars and trucks: 1931-1996.

Carew, M. G. (2009), *Becoming the Arsenal: The American Industrial Mobilization for World War II, 1938-1942*, University Press of America.

Carter, S. B., Garntner, S. S., Haines, M. R., Olmstead, A. L., Sutch, R., Wright, G., Atack, J. & Bateman, F. (2006), *Historical Statistics of the United States Millennial Edition Online, Table Dd366-436: Physical output of selected manufactured products: 1860-1997*, Cambridge University Press.

Choi, I. (2001), 'Unit root tests for panel data', *Journal of International Money and Finance* **20**(2), 249–272.

Civilian Production Administration (1947), *Official Munitions Production of the United States*.

Coates Jr, J. B. (1959), 'The history of the us army medical department in world war ii', *Bulletin of the Medical Library Association* **47**(3), 264.

Coker, K. R. & Stokes, C. E. (1991), *A Concise History of the U.S. Army Signal Corps*, Army Signal Center and Fort Gordon GA.

Connors, J. (2010), *Engines of Pratt and Whitney - A Technical History*, American Institute of Aeronautics and Astronautics.

Crawford, R. H. & Cook, Lindsey, F. (1952), *US Army in World Wart II: Statistics - procurement*.
URL: http://www.alternatewars.com/BBOW/Stats/USA_in_WW2_Stats-Procure_9-APR-52.PDF

David, P. A. (1974), 'The "Horndal effect" in Lowell, 1834–1856: A short-run learning curve for integrated cotton textile mills', *Explorations in Economic History* **10**(2), 131–150.

Dosi, G., Grazzi, M. & Mathew, N. (2017), 'The cost-quantity relations and the diverse patterns of learning by doing: Evidence from India', *Research Policy* **46**(10), 1873–1886.

Evans, P. (1982), 'The effects of general price controls in the united states during world war ii', *Journal of Political Economy* pp. 944–966.

Foster, L., Haltiwanger, J. & Syverson, C. (2008), 'Reallocation, firm turnover, and efficiency: selection on productivity or profitability?', *American Economic Review* **98**(1), 394–425.

Funk, J. L. & Magee, C. L. (2014), 'Rapid improvements with no commercial production: How do the improvements occur?', *Research Policy* .

George, J. L. (n.d.), *History of Warships from Ancient Times to the Twenty-First Century*, Naval Institute Press.

Gordon, R. J. (1969), '$45 billion of us private investment has been mislaid', *American Economic Review* **59**(3), 221–238.

Hadri, K. (2000), 'Testing for stationarity in heterogeneous panel data', *The Econometrics Journal* **3**(2), 148–161.

Hall, R. E. (2009), 'By how much does GDP rise if the government buys more output?', *Brookings Papers on Economic Activity* (2).

Harris, R. D. & Tzavalis, E. (1999), 'Inference for unit roots in dynamic panels where the time dimension is fixed', *Journal of Econometrics* **91**(2), 201–226.

Heathcote, A., Brown, S. & Mewhort, D. (2000), 'The power law repealed: The case for an exponential law of practice', *Psychonomic Bulletin & Review* **7**(2), 185–207.

Hecks, K. (1990), *Bombing 1939 - 1945: The Air Offensive Against Land Targets in World War II*, Edmundsbury Press.

Hendry, D. F. (1995), *Dynamic econometrics*, Oxford University Press on Demand.

Herman, A. (2012), *Freedom's Forge: How American Business Produced Victory in World War II*, Random House.

Hirsch, W. Z. (1956), 'Firm progress ratios', *Econometrica* pp. 136–143.

Holley, I. B. (1964), *Buying Aircraft: Materiel Procurement for the Army Air Forces*, Office of the Chief of Military History, Department of the Army.

Hyde, C. K. (2013), *Arsenal of Democracy: The American Automobile Industry in World War II*, Wayne State University Press.

Im, K. S., Pesaran, M. H. & Shin, Y. (2003), 'Testing for unit roots in heterogeneous panels', *Journal of Econometrics* **115**(1), 53–74.

Irwin, D. A. & Klenow, P. J. (1994), 'Learning-by-doing spillovers in the semiconductor industry', *Journal of political Economy* **102**(6), 1200–1227.

Killblane, R. (2012), '70 years of the transportation corps', *Professional bulletin of the the United States army sustainement* **44**(4).
URL: https://alu.army.mil/alog/issues/julaug12/70_Years_transportation_Corps.html

Klette, T. J. & Griliches, Z. (1996), 'The inconsistency of common scale estimators when output prices are unobserved and endogenous', *Journal of Applied Econometrics* **11**(4), 343–361.

Koistinen, P. A. (2004), *Arsenal of World War II: The Political Economy of American Warfare, 1940 - 45*, University Press of Kansas.

Lafond, F., Bailey, A. G., Bakker, J. D., Rebois, D., Zadourian, R., McSharry, P. & Farmer, J. D. (2018), 'How well do experience curves predict technological progress? a method for making distributional forecasts', *Technological Forecasting and Social Change* **128**, 104 – 117.

Lazonick, W. & Brush, T. (1985), 'The "horndal effect" in early us manufacturing', *Explorations in Economic History* **22**(1), 53–96.

LeMay, C. E. & Yenne, B. (1988), *Superfortress: The Story of the B-29 and American Air Power*, McGraw-Hill Companies.

Levin, A., Lin, C.-F. & Chu, C.-S. J. (2002), 'Unit root tests in panel data: asymptotic and finite-sample properties', *Journal of Econometrics* **108**(1), 1–24.

Levitt, S. D., List, J. A. & Syverson, C. (2013), 'Toward an understanding of learning by doing: Evidence from an automobile assembly plant', *Journal of political Economy* **121**(4), 643–681.

Lieberman, M. B. (1984), 'The learning curve and pricing in the chemical processing industries', *The RAND Journal of Economics* **15**(2), 213–228.

Lorell, M. A. (2003), The u.s. combat aircraft industry, 1909-2000: Structure, competition, innovation, Technical report, RAND Corporation.

Mauroni, A. (2015), 'The U.S. army chemical corps: Past, present and future'.
URL: https://armyhistory.org/the-u-s-army-chemical-corps-past-present-and-future

McNerney, J., Farmer, J. D., Redner, S. & Trancik, J. E. (2011), 'Role of design complexity in technology improvement', *Proceedings of the National Academy of Sciences* **108**(22), 9008–9013.

Mishina, K. (1999), Learning by new experiences: Revisiting the flying fortress learning curve, *in* 'Learning by doing in markets, firms, and countries', University of Chicago Press, pp. 145–184.

Nagy, B., Farmer, J. D., Bui, Q. M. & Trancik, J. E. (2013), 'Statistical basis for predicting technological progress', *PloS one* **8**(2), e52669.

Newell, A. & Rosenbloom, P. S. (1981), 'Mechanisms of skill acquisition and the law of practice', *Cognitive skills and their acquisition* **1**(1981), 1–55.

Nordhaus, W. D. (2014), 'The perils of the learning model for modeling endogenous technological change', *The Energy Journal* **35**(1), 1–13.

Papineau, M. (2006), 'An economic perspective on experience curves and dynamic economies in renewable energy technologies', *Energy Policy* **34**(4), 422–432.

Pattillo, D. M. (1998), *A history in the making: 80 turbulent years in the American general aviation industry*, McGraw Hill Professional.

Pedroni, P. (1999), 'Critical values for cointegration tests in heterogeneous panels with multiple regressors', *Oxford Bulletin of Economics and Statistics* **61**(S1), 653–670.

Pesaran, M. H. & Smith, R. (1995), 'Estimating long-run relationships from dynamic heterogeneous panels', *Journal of Econometrics* **68**(1), 79–113.

Pozzi, A. & Schivardi, F. (2016), 'Demand or productivity: What determines firm growth?', *RAND Journal of Economics* **47**(3), 608–630.

Preston, L. E. & Keachie, E. C. (1964), 'Cost functions and progress functions: An integration', *American Economic Review* **54**(2), 100–107.

Ramey, V. A. (2011), 'Identifying government spending shocks: It's all in the timing', *Quarterly Journal of Economics* **126**(1), 1–50.

Rapping, L. (1965), 'Learning and world war ii production functions', *Review of Economics and Statistics* pp. 81–86.

Reed, W. R. (2015), 'On the practice of lagging variables to avoid simultaneity', *Oxford Bulletin of Economics and Statistics* **77**(6), 897–905.

Risch, E. (2014), *The Quartermaster Corps: Organization, Supply and Services, 2 vol.*, Washington DC: Center for Military History.

Rogerson, W. P. (1994), 'Economic incentives and the defense procurement process', *Journal of Economic Perspectives* pp. 65–90.

Rohlfs, C., Sullivan, R. & Kniesner, T. J. (2016), 'Reducing risks in wartime through capital-labor substitution: Evidence from world war II', *Journal of Risk and Uncertainty* **52**(2), 163–190.

Roosevelt, F. D. (1940), 'Fireside chat 16. on the 'arsenal of democracy', *Miller center of public affairs. University of Virginia. Presidential Speech Archive* **29**.

Ross, S. H. (2003), *Strategic Bombing by The United States in World War II: The Myths and the Facts*, McFarland & Company.

Rostker, B. (2013), *Providing for the Casualties of War: The American Experience Through World War II*, RAND Corporation.

Rubis, K. (2012), 'The history of ordnance in america', *Professional bulletin of the United States army sustainment* .
URL: https://alu.army.mil/alog/issues/MayJune12/history_ordance.html

Scott-Kemmis, D. & Bell, M. (2010), 'The mythology of learning-by-doing in world war ii airframe and ship production', *International Journal of Technological Learning, Innovation and Development* **3**(1), 1–35.

Searle, A. D. (1945), 'Productivity of labor and industry', *Monthly Labor Review* **61**(6), 1132–1147.

Sheshinski, E. (1967), 'Tests of the" learning by doing" hypothesis', *Review of Economics and Statistics* pp. 568–578.

Sinclair, G., Klepper, S. & Cohen, W. (2000), 'What's experience got to do with it? sources of cost reduction in a large specialty chemicals producer', *Management Science* **46**(1), 28–45.

Smith, R. E. (1959), *The army and economic mobilization*, Government Printing Office, Washington, D.C.

Söderholm, P. & Sundqvist, T. (2007), 'Empirical challenges in the use of learning curves for assessing the economic prospects of renewable energy technologies', *Renewable Energy* **32**(15), 2559–2578.

Swamy, P. A. (1970), 'Efficient inference in a random coefficient regression model', *Econometrica* pp. 311–323.

The National World War II Museum (2015), *By the Numbers: The US Military*, The National World War II Museum, http://www.nationalww2museum.org/learn/education/for-students/ww2-history/ww2-by-the-numbers/us-military.html.

Thompson, P. (2001), 'How much did the liberty shipbuilders learn? new evidence for an old case study', *Journal of Political Economy* **109**(1), 103–137.

Thompson, P. (2007), 'How much did the liberty shipbuilders forget?', *Management Science* **53**(6), 908–918.

Thompson, P. (2012), 'The relationship between unit cost and cumulative quantity and the evidence for organizational learning-by-doing', *Journal of Economic Perspectives* **26**(3), 203–224.

Thornton, R. A. & Thompson, P. (2001), 'Learning from experience and learning from others: An exploration of learning and spillovers in wartime shipbuilding', *American Economic Review* **91**(5), 1350–1368.

U.S. Bureau of Economic Analysis (2017), Personal consumption expenditures excluding food and energy (chain-type price index), Technical report. [DPCCRG3A086NBEA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DPCCRG3A086NBEA.

War Production Board (1945), *Wartime Production Achievements and the Reconversion Outlook*.

White, G. (1995), *Allied Aircraft Piston Engines of World War II*, Graham White.

Witajewski-Baltvilks, J., Verdolini, E. & Tavoni, M. (2015), 'Bending the learning curve', *Energy Economics* **52**, S86–S99.

Wooldridge, J. M. (2002), *Econometric analysis of cross section and panel data*, MIT Press.

Wright, T. P. (1936), 'Factors affecting the cost of airplanes', *Journal of the Aeronautical Sciences* **3**(4), 122—128.

Yelle, L. E. (1979), 'The learning curve: Historical review and comprehensive survey', *Decision Sciences* **10**(2), 302–328.

Appendix

A Data sources, data collection and data cleaning

A.1 The *Labor Productivity* dataset

The *Labor Productivity* data we assembled comes from three different sources, which we describe here. Fig. 5 shows the times series (only the slopes are comparable, except for the Source Book data which is all expressed in manhours per pound for all aircraft.)

Searle 1945 We extracted the data from Tables 1, 2, 3, 6 and 7. Victory and Cargo vessels are indices constructed from multiple models. A summary of the products we have information about is in the table below.

Product	Start date	End date	T	Total prod.
Liberty Ships	Dec-41	Dec-44	37	2458
Victory Ships	Feb-44	Dec-44	11	199
Cargo Vessels	Apr-43	Dec-44	21	160
Tankers Vessels	Jun-43	Dec-44	19	308
Destroyer Escort	Apr-43	Nov-44	20	351

Table 7: Ships data extracted from Searle (1945)

The Source Book. The report that provided Alchian with the data used in his study is available in the form of a digitized PDF[18]. To create this dataset, we transcribed Tables 3 and 4 in this report, *Source Book of World War II Basic Data - Airframe Industry, Volume 1: Direct Man-Hours - Progress Curves (Dayton, OH: Army Air Forces, Air Materiel Command, January 1950)* (Army Air Forces 1947) The *Source Book* presented significant transcription challenges. Some of the digits were illegible or had been clearly switched during the transcription process with digits similar in appearance, like an 8 for a 0. Luckily, Tables 3 and 4 represent the same data – man-hours per airframe pound and cumulative production – about the same models of airplanes. Table 3 organizes this data principally by manufacturer and plant, while Table 4 organizes it by airplane model. Therefore, the two tables could be compared to one another to corroborate interpretations of certain entries that were hard to read or seemed clearly wrong in one table or another. In the rare cases that it was impossible to transcribe the data faithfully after consulting both tables, we dropped the observations. In one case, the Consolidated Vultee San Diego B-24 August 1943 value for man-hours was changed from a clear 0.07 to a much more plausible 0.77 (the series read $\ldots 0.88, 0.84, 0.84, 0.07, 0.67, 0.65, 0.65\ldots$).

The file gives value of cumulative production, from which we deduce production as the difference in cumulative production. In one case, the B-17 from Seattle, cumulative production was available at the earliest date, January 1940. The series read 45,

[18] https://apps.dtic.mil/dtic/tr/fulltext/u2/a800199.pd

49, 54. Clearly 45 was not the monthly production in January (see Mishina (1999)), so we just assumed that production in January was the same as in February, 4 units.

Finally, there is a page from Table 3 missing in the available PDF. Unfortunately, because of the organization of the PDF, it is impossible to know what data about which manufacturers was on this missing page.

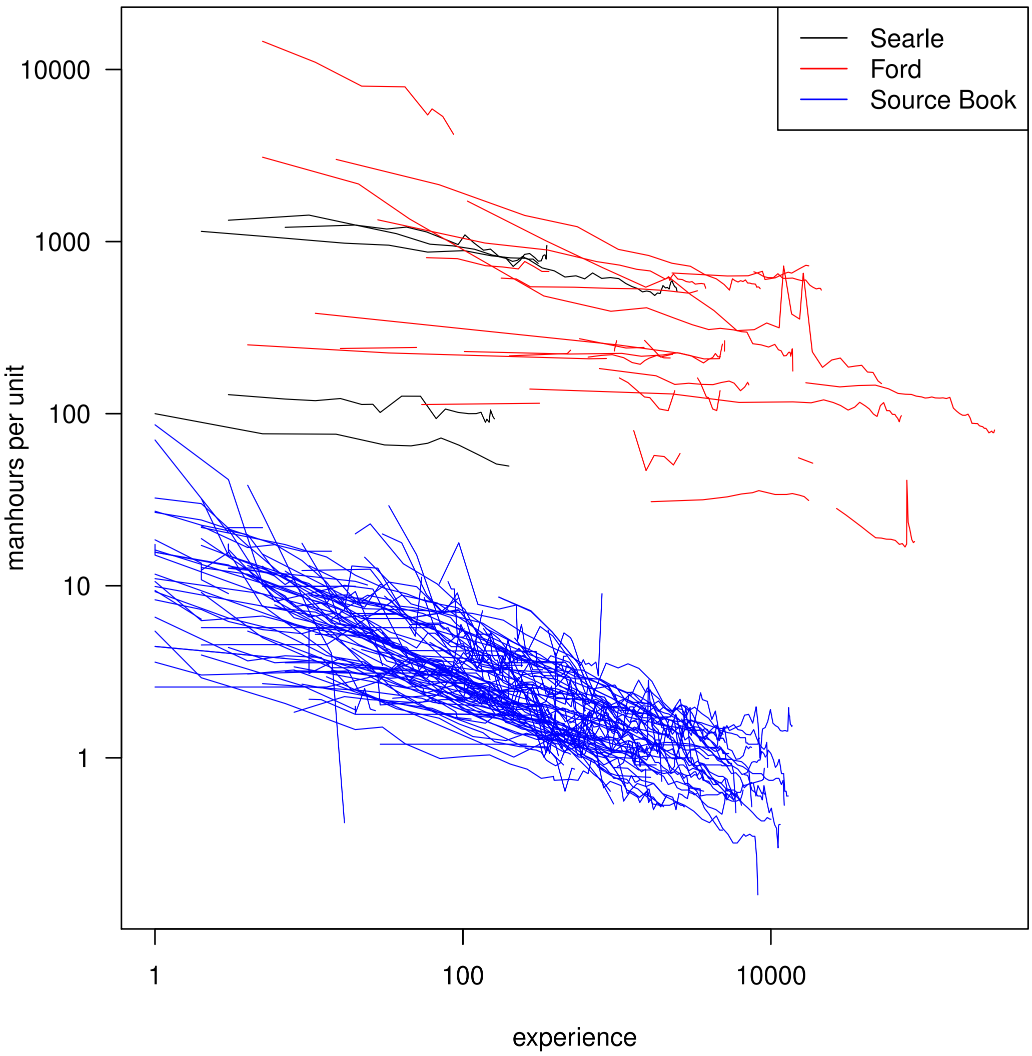

Figure 5: Time series of labor productivity from all included sources.

Table 8 shows some examples of aircraft-plant pairs from this source. We have chosen the examples to show that some plants produced multiple aircrafts, some aircraft were produced by multiple plants, and the total number of units produced and time span of production varied widely.

Ford Archival Records. The Ford Motor Company's archives are held in a dedicated research center, The Benson Ford Research Center in Dearborn, MI. In the Charles La Croix papers, there is a copy of a publication called Record of War Effort: Contributions of the Ford Motor Company in the Development of Production for Victory

Product	Start date	End date	T	Total prod.
B-17, Douglas, Long Beach	Oct-42	Jul-45	34	3666
B-17, Lockheed, Burbank	Jun-42	Jul-45	38	2750
C-69, Lockheed, Burbank	May-44	Sep-45	17	16
A-20, Douglas, Tulsa	Oct-44	Jun-45	9	1085
SB2A, Brewster, Johnsville and L.I.C.	Jan-43	May-44	17	302
R-5, Sikorsky, Bridgeport	Jan-45	Aug-45	8	11

Table 8: Examples of data extracted from the "Source Book"

(Detroit, MI: Ford Motor Company, n.d. 2 vol.). Volume 2 describes the monthly man-hour-per-unit requirements for what appears to be all of the products Ford produced throughout World War II. As this seems to have been an internal publication, there are no sources for this data beyond the implication that they are from extensive internal auditing and record keeping during the war. Table 9 summarizes the data extracted for Ford's archives.

A.2 The *OMPUS-USMH* dataset

This section presents detailed information about the two principal components of an original data set we assembled for this research: the *Official Munitions Production of the United States* and *United States Munitions Handbook*.

The PDF of the *Official Munitions Production of the United States* (OMPUS) is easily available[19], and contains data on production volume. This document is a Special Release, edited in 1947, covering each month from July 1940 ("the beginning of the war program") through August 1945 ("the last month of actual fighting against Japan"). Footnotes often make references to data being only a partial coverage, often because data from some component agencies was not available. In a few cases (experimental aircraft), we also have data from January to June 1940. This source was easy to read. We omitted Canadian data, and products with only one or two values of production. For ships, we took the value in displacement tons instead of units. For a few products, some of the data was available as aggregate for typically 6 months or a year. In these cases we attributed to each month a *pro rata* value.

The *United States Munitions Handbook* (USMH) is a formerly classified publication that was located in the Policy Documentation File (Record Group 179, Stack Group 570) by one of us (D.G.) on a research trip to the National Archives in College Park, MD in April 2015. The transcription from photographs of the document did not present any significant challenges. We note here that for airplanes, the cost data often appears to refer to particular plants, whereas the reference OMPUS production is product-level.

The USMH contains data on "early" and "late" cost for many products. These products are named, and a reference to the OMPUS is provided in the form a page and column number. This provided an uncontroversial match for the vast majority of products, although in some cases the USMH seems to refer to more models than the

[19] http://cgsc.contentdm.oclc.org/cdm/ref/collection/p4013coll8/id/3332

Product	Start date	End date	T	Total prod.
Universal Carrier GAU	Mar-43	May-45	27	13893
Cargo Truck - OTBA	Jul-43	Oct-44	16	2218
CG-13A - Glider (42Places) - GBG	Jan-44	Dec-44	12	87
MX Engine Assembly	Sep-44	Jul-45	11	2378
Aircraft Generators – GAL P-1 and R-1	Dec-42	Jul-45	32	87390
British Engine – GAE	Jun-42	May-43	12	17593
British Axel – GAE	Jun-42	May-43	12	17639
Bomb Service GTBB	Apr-43	Sep-43	6	50
Bomb Service GTBC	Sep-43	Oct-44	14	4701
Cargo Truck - GTB (Less Winch)	Jun-42	Mar-43	10	5007
Stake Truck – G8T	Mar-43	Mar-44	13	7198
Cargo Truck - G8T	Sep-42	May-45	33	70420
Tank Engine GAA	Jul-42	Aug-45	38	21478
Turbo Supercharger - B2 and B22	Aug-42	Oct-44	27	52281
Bomb Service GTBS	Nov-42	Jul-44	21	4679
MX Field Assembly	Sep-44	Aug-45	12	2579
Jeep GP and GPW	Feb-41	Jul-45	54	283664
Tractor Truck G8T	Dec-43	Jan-44	2	314
Armored Car M-8 - GAK	Mar-43	May-45	27	8524
Cargo Truck - GTB (With Winch)	Jul-42	Mar-43	9	995
Tank Engine GAF	Nov-42	Sep-44	23	3908
Tank Engine GAN	Aug-43	Dec-44	17	380
Armored Car M-20 GBK	May-43	Jun-45	26	3773

Table 9: Data extracted from Ford's archives

OMPUS. In a few cases, the match using the page and column number was erroneous, and we used names instead. In a few other cases, product detail was higher in the USMH than in the OMPUS, so that different cost changes were attributed to the same production time series. We did not transcribe these cases.

One specific issue with the matched OMPUS-USMH data is that the USMH does not provide a very clear definition of the cost data ("Standard Dollar Weight"). The Foreword to the USMH states: "The cost figures shown for the separate items are the standard costs which were used in computation of the War Production Board (WPB) index of war production and the Production Statement. They are included in this report to provide the reader with a proper perspective on the magnitude and relative significance of the items involved. Both an "early" and a "late" cost are shown wherever possible, comparison of the two costs oftentimes provides a clue to the tremendous advances in manufacturing techniques which took place in some munitions areas, enabling costs to be cut sharply even during a period of generally rising prices." In short, we think the Standard Dollar Weight represents a nominal dollar amount reported to understand product-level inflation over the course of the war. However, it could be that this is already deflated, so we chose not to deflate the data further in our main text analysis. We note that production in World War II was conducted in a climate of price controls and the rationing of materials and labor (Evans 1982). Despite this inflation was 5% in 1941, 10% in 1942 and averaged 6% annually for the remainder of the war (U.S. Bureau of Economic Analysis 2017).

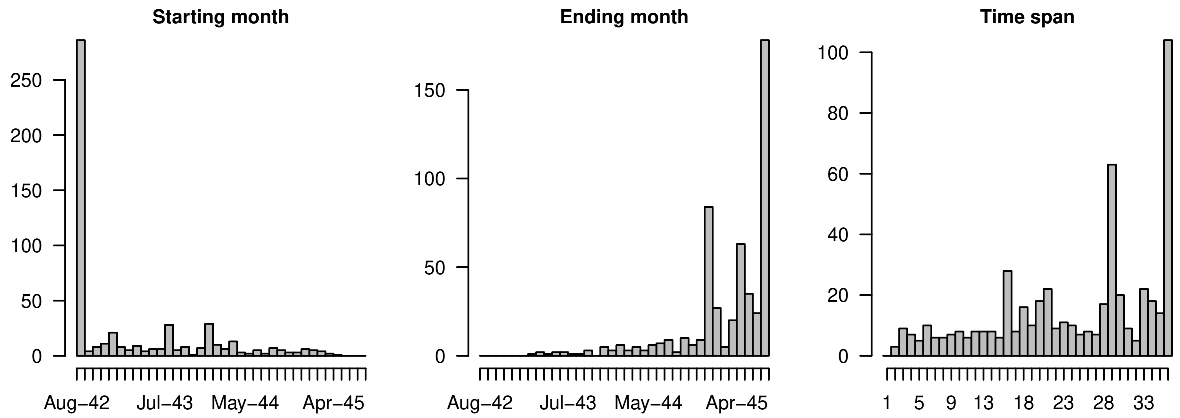

Figure 6: Distribution of the "early" month, "late" month, and number of months in between.

A second issue with the dataset is that we had to modify the dates with which the "early" and "late" costs are associated. While the dates of the early and late costs are reported for each product in the USMH, these dates are always the same in each product category. However, in the explanatory notes (e.g. USMH p.104, explanatory notes for Ordnance and Automotive Vehicles), an important clarification is that "the cost of the item as of the last month of production has been shown as the final cost, while for items produced after 1942, the dollar value shown for the earliest month of production has been listed as the original cost". Therefore, every time we found that a product was not yet in production at the date of early cost, we corrected the date of the early cost as the month in which production started. Every time we found a product for which production had stopped before the date of the late cost, we corrected the date of the late cost as the month in which production stopped[20]. Fig. 6 shows the distribution of the corrected Early and Late dates, as well their difference. For a large number of products we have a start date in August 1942 and an end date in August 1945, 36 months later.

Fig. 7 shows the USMH-OMPUS data (only the slopes are comparable, and "unit" costs may refer to units, pairs, thousands, or millions of units depending on the product).

[20]For 46 Ordnance items (OMPUS ref. 177/2 to 188/7) the Early date reported was the implausible April 1945, but the explanatory notes report that "The August 1942 cost was used as the original production cost for both Army and Navy items" so we edited this subset accordingly. For 341 products, the late date was September 1945 but our production data stops in August 1945, so we assumed that the Late cost was for August. For 2 products, production starts on the month of the "Late" date, so we removed them.

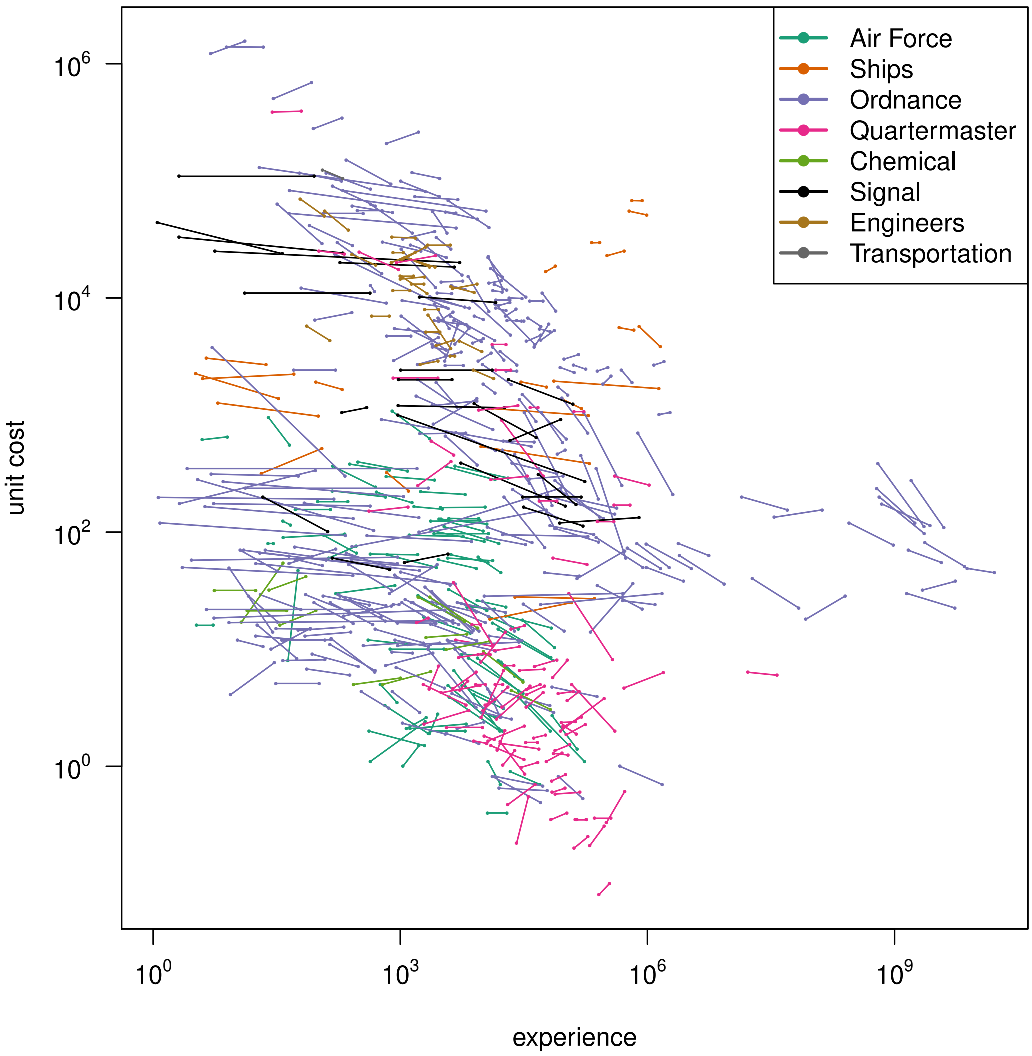

Figure 7: Experience curves constructed by matching USMH and OMPUS.

A.3 The *Contract Prices* dataset

We collected the *Contract Prices* data from Crawford & Cook (1952) (tables PR-2, PR-3 and PR-20), who compiled the estimated value of procurement from various sources, using a sample of products. "For many groups the sample covered more than 90 percent of the total values".

The production index was computed as follows: "Quantities of the sample items delivered each month were multiplied by a weighted unit cost for the item to derive the dollar value of the sample. The unit cost figure for each item was based on the contract or purchase price plus allowances for overhead, the cost of Government-furnished equipment and materiel and any other costs incurred in connection with the item by the War Department." Most importantly, these time series represent physical volume, not the product of physical volume and prices. The relevant excerpt from Crawford & Cook (1952) is footnote *a* on p. 20 "Data were computed from physical quantities delivered and standard dollar weights which for most items were unit costs as of 1945. The figures therefore reflect physical volume rather than cost to the Government; they do not take into consideration price changes or contract renegotiations." The explanatory notes on p.86 further state that "The series was designed to show relative magnitudes and trends in the physical volume of procurement deliveries" Fig. 8 shows the data on production, clearly exhibiting a plateauing in 1943-44 and decrease in 1945.

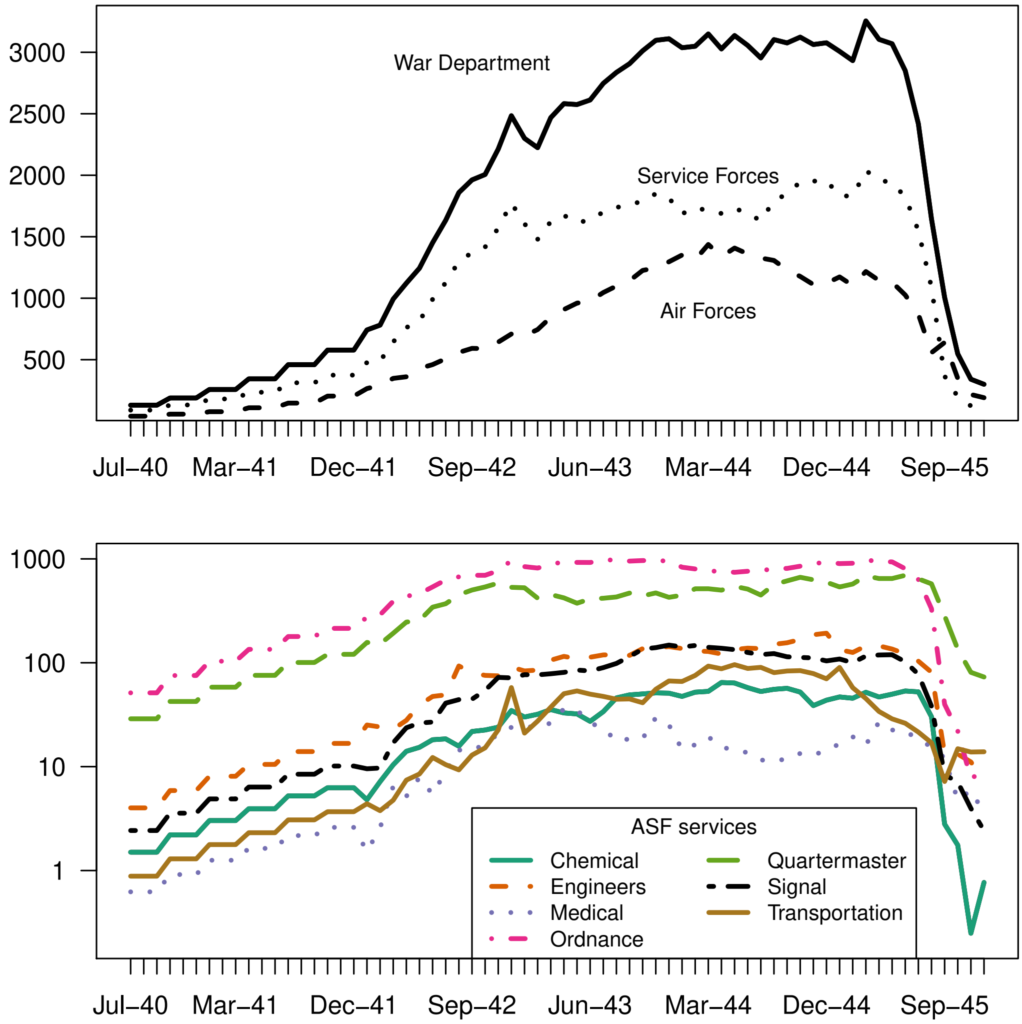

Figure 8: Estimated monthly rates of output between July 1st 1940 and July 31st 1945, total and by main category, in millions of standard dollar weights. Source: War Production Board (1945), p.105.

The monthly data starts in January 1942 for all series. However, for the War department, the AAF and the aggregated ASF, quarterly data on production was available for the previous 6 quarters (1940Q3-1941Q4). Because this information is useful for constructing experience, we used it as follows. For these three series, we constructed monthly data for July 1940-December 1941 by attributing equally to each month the quarterly production data. For the subservices, we computed the share of each subservice in the ASF total for the first 6 months of 1942, and used these shares to calculate monthly production for the period July 1940-December 1941. Note that since price data starts in January 1942, these assumptions are only useful for plotting Fig. 8, and do not change our regression results. Similarly, the production data was not available for the last four months for AAF. We computed the ratio of AAF to ASF in the previous 6 months (March to August 1945), and used this to estimate the values for AAF, and thus for the Total War department as well, for September-December 1945.

The indices of contract price changes were computed as follows: "The items included in these indices cover approximately 50 percent of the total value of War Department procurement. They were selected to be representative of all principal kinds of items purchased. The basic data employed were the contract prices for each company supplying the item on the selected list. All successive prices in additional contracts or revisions of existing contracts were recorded after necessary adjustments were made for specification changes. The price data for all companies supplying a given item were used to compute an index for that item after appropriate weights had been assigned on the basis of relative importance in terms of physical volume".

The explanatory notes also mention that "Individual item indices were combined into group indices and, in turn, into technical service indices. These composites were combined into a master Army Service Forces (ASF) Index, and a similar composite index on Army Air Forces (AAF) items was added to the Army Service Forces composite to provide a War Department index. The indexes do not cover any items produced in government-owned contractor operated plants, or, with the exception of the AAF index, items procured through cost-plus-a-fixed-fee contracts since such purchases were to a degree noncompetitive and the contract terms were often such as to cause the prices to be, incomparable with those of procurement through ordinary commercial channels." Fig. 9 shows the price indices for contract for various wartime agencies, indicating an important decrease for almost every department, the exception being the Quartermaster.

B Estimating prior experience

Estimating relevant prior experience for each of the product categories included in our datasets was one of the greatest challenges we faced in writing this paper. In order to create these estimates, we made a couple of assumptions about prior experience. First, we assumed that Americans gained most of their experience in producing military-specific products (guns, munitions, etc.) during World War I. Therefore, we were able to use the extensive statistical and primary sources available about WWI output to create these estimates. Second, where a technology had both military and civilian applications, we aggregated the WWI military output and rarer estimates of civilian output, where possible. For civilian production, two sources were essential: *the Historical Statistics of the United States* and history of science and technology books about specific products such as radios and airplanes.

The OMPUS data is far more granular than the contract price data. Therefore, after estimating initial experience for products included in the OMPUS data (Section B.1),

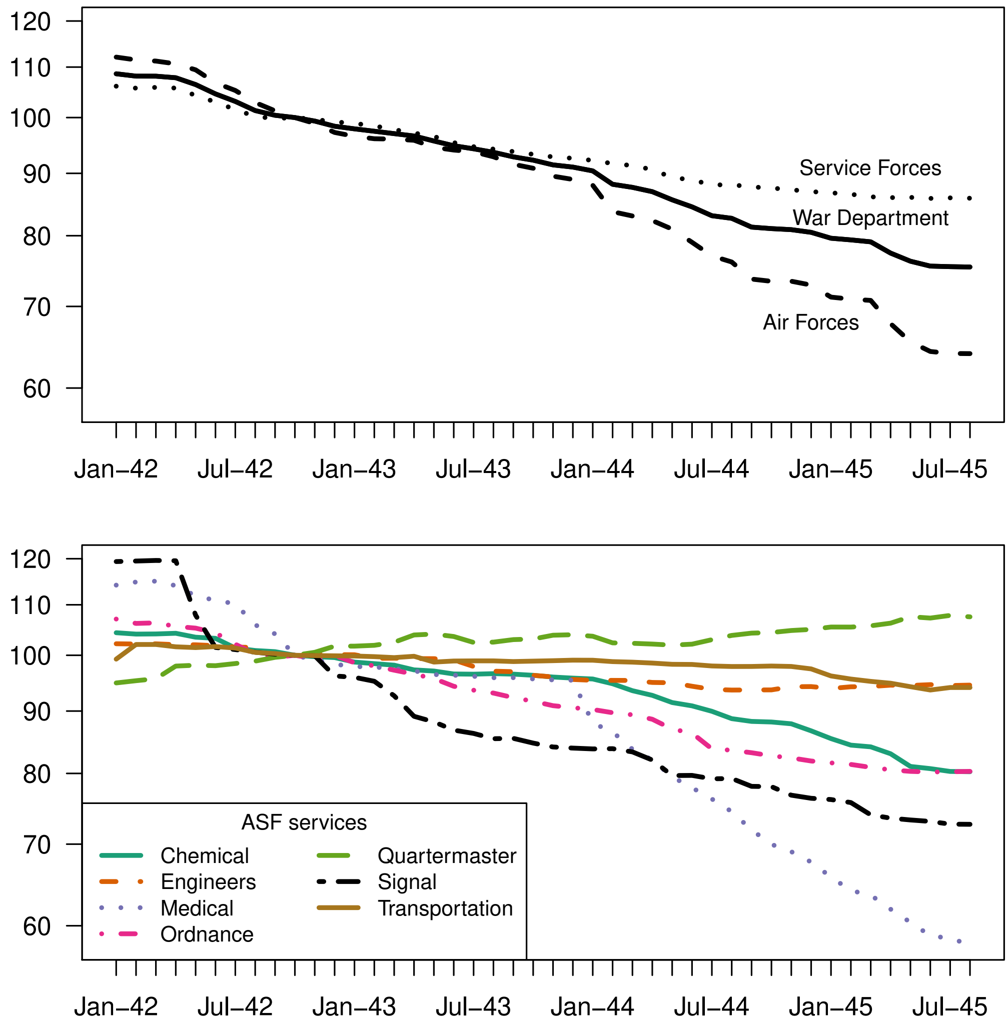

Figure 9: Quarterly Index of Contract Price Changes: 1942 - 1945 for War Department and Component Agencies (October 1942 = 100). Source: Smith (1959), p. 412.

we present a crosswalk that aggregates the OMPUS-level estimates and matched them to the categories in the contract price data (Section B.2).

Main categ.	Subcateg.	ζ	N	War Dep.
Aircraft	Bomber	0.21	15	Air Force
Aircraft	Fighter	0.21	12	Air Force
Aircraft	Reconnaissance (inc. Photographic)	0.21	1	Air Force
Aircraft	Transport	0.21	8	Air Force
Aircraft	Trainer	0.21	4	Air Force
Aircraft	Communication	0.21	3	Air Force
Aircraft	Special Purpose Aircraft	0.21	0	Air Force
Aircraft	Gliders	0.21	0	Air Force
Aircraft	Airships, Barrage Balloons, and Special Devices	0.21	0	Air Force
Aircraft	Aircraft Engines	0.21	21	Air Force

Aircraft	Aircraft Propellers	0.21	12	Air Force
Ships	Combatant	1.21	7	Ships
Ships	Landing vessels	0.05	5	Ships
Ships	Patrol	0.75	3	Ships
Ships	Mine Craft	0.81	1	Ships
Ships	Transports	0.05	0	Ships
Ships	Dry Cargo	0.05	0	Ships
Ships	Tankers	0.05	0	Ships
Ships	Tender and Repair Vessels	0.05	1	Ships
Ships	District Craft	0.05	1	Ships
Ships	Other	0.05	3	Ships
Ships	Maritime Commission Nonmilitary Vessels Delivered to the Armed Forces	0.05	0	Ships
Ships	Army Tugs and Barges	0.05	0	Ships
Ordnance	Field Artillery	0.39	12	Ordnance
Ordnance	Spare Canon, Tubes, and Recoil Mechanisms for Field, Tank, and Self-Propelled Artillery	0.40	0	Ordnance
Ordnance	Tank Guns and Howitzers	0.40	0	Ordnance
Ordnance	Self-Propelled Guns and Howitzers	0.40	6	Ordnance
Ordnance	Aircraft and Army Antiaircraft Guns	0.40	8	Ordnance
Ordnance	Army Rocket Launchers	0.40	6	Ordnance
Ordnance	Mortars	0.40	4	Ordnance
Ordnance	Naval Surface Fire (Guns and Small Arms)	0.40	4	Ordnance
Ordnance	Naval Antiaircraft and Dual-purpose (Guns and Small Arms)	0.40	6	Ordnance
Ordnance	Naval Rocket Launchers	0.40	0	Ordnance
Ordnance	Small Arms	0.40	18	Ordnance
Ordnance	Misc. Army Weapons and Ordnance Mat.	0.40	10	Quartermaster
Ordnance	Misc. Navy Weapons and Ordnance Mat.	0.40	0	Ordnance
Ordnance	Fire Control (excl. radar)	0.40	14	Ordnance
Ordnance	Artillery and Tank Gun	0.35	61	Ordnance
Ordnance	Aircraft (Ammunition)	0.35	7	Ordnance
Ordnance	Army Antiaircraft	0.35	6	Ordnance
Ordnance	Mortar Shells	0.35	13	Chemical
Ordnance	Army Rockets	0.35	6	Ordnance
Ordnance	Army Practice and Drill (All Types)	0.35	3	Ordnance
Ordnance	Naval Surface Fire (Ammunition and Bombs)	0.35	8	Ordnance
Ordnance	Naval Antiaircraft and Dual-purpose (Ammunition and Bombs)	0.35	4	Ordnance
Ordnance	Navy Rockets	0.35	0	Ordnance
Ordnance	Small Arms Ammunition	0.35	15	Ordnance
Ordnance	Land Mines	0.10	0	Ordnance
Ordnance	Grenades	0.10	0	Ordnance
Ordnance	Pyrotechnics	0.10	0	Chemical
Ordnance	Explosives	0.10	0	Ordnance
Ordnance	Propellants : Smokeless Powder	0.10	0	Ordnance
Ordnance	Torpedos	0.10	1	Ordnance
Ordnance	Naval Mines	0.10	1	Ordnance
Ordnance	Depth Charges	0.10	1	Ordnance
Ordnance	Aircraft Bombs	0.00	31	Ordnance
Ordnance	Combat Vehicles (Tanks)	0.01	6	Ordnance
Ordnance	Motor Carriages for Self-propelled Guns	1.00	1	Ordnance
Ordnance	Heavy-heavy Trucks	1.99	17	Ordnance
Ordnance	Light-heavy Trucks	1.99	10	Ordnance

Ordnance	Medium trucks	1.99	5	Ordnance
Ordnance	Light Trucks	1.99	3	Ordnance
Ordnance	Trailers, Semitrailers, and Motorcycles	1.99	2	Ordnance
Ordnance	Remanufactured Automotive Vehicles	1.99	0	Ordnance
Ordnance	Tractors	1.99	8	Ordnance
Comm.	Army (Radio)	0.04	5	Signal
Comm.	Navy (Radio)	0.04	0	Signal
Comm.	Ship and Ship-and-Shore (Radio)	0.04	0	Signal
Comm.	Ground (Radio)	0.04	7	Signal
Comm.	Army (Radar)	0.00	4	Signal
Comm.	Navy (Radar)	0.00	0	Signal
Comm.	Ship and Ship-and-Shore (Radar)	0.00	0	Signal
Comm.	Ground (Radar)	0.00	3	Signal
Comm.	Underwater Sound Equipment	0.00	0	Signal
Comm.	Wire Communication and Misc. Equipment	0.10	4	Signal
Other	Petroleum Products: Aviation Gasoline	0.20	0	Air Force
Other	Machinery	1.00	27	Engineers
Other	Railroad Equipment	1.00	1	Transportation
Other	Clothing	1.00	50	Quartermaster
Other	Medical Supplies and Subsistence Rations	0.10	0	Medical
Other	Misc. Equipment and Supplies	0.50	28	Quartermaster

Table 10: Estimated prior experience for *OMPUS-USMH* data. Main and Subcategory are the main section and finest available section of the OMPUS table of contents (ToC); Product is the most fine grained subsection of the OMPUS ToC available; N is the number of products; War Department is from our hand-made crosswalk. Horizontal lines delineate the higher-level ToC categories discussed in the text. Note that product types for which we have no data matched with USMH ($N = 0$) are also reported.

B.1 Estimates of prior experience for the *OMPUS-USMH* data

This appendix provides an explanation for how we arrived at an estimate of prior experience for each category of product in Table 10. The total wartime production for each category was taken from the summary table "Production of Selected Munitions Items" in War Production Board (1945). As discussed in Section 4.1, the OMPUS dataset disaggregates many products into their component parts. The aggregate table in War Production Board (1945) sums these components into larger product categories and then industry-level categories. We have used this table in lieu of summing the OMPUS ourselves to avoid mismatching components of the same finished product. This appendix explains how we gathered numbers about prior production and wartime production to calculate the prior experience factor ζ_is presented in table 10.

Aircraft. Aircraft were not a novelty in World War II, but the scale and methods of manufacture changed significantly during the war. Furthermore, significant changes were made to their design. Much of this change was linked to improvements in engines and propellers, which are a separate category in the OMPUS dataset and are discussed below. However, the United States did have prior experience in manufacturing aircraft – major firms like Boeing and Curtiss (now Curtiss-Wright) were both founded in 1916. Therefore, we estimated this prior experience by finding the number of individual civil and military airplanes produced in the United States before 1940. Consulting Pattillo (1998) and Lorell (2003), we were able to determine that 62,401 aircraft were produced in the United States before World War II. Many of these aircraft were produced for World War I and for the postal service, which used planes to transport mail over long distances. According to the War Production Board (1945), 296,429 aircraft were produced during World War II. We thus applied a value of $\zeta_i = 0.211$ to all types of aircraft.

Aircraft Engines and Propellers. As mentioned, the OMPUS often disaggregates products into their components. Therefore, it provides cost and production information not only for airframes and completed airplanes, but also separate information for airplane engines and propellers. If we assume that each aircraft built prior and during the war used two engines and two propellers, the estimate of ζ_i remains the same as for aircrafts, $\zeta_i = 0.21$. We note that it may be an overestimate, because unlike for the construction of airframes and other components, there were design changes made to aircraft engines during the war that made prior manufacturing experience less relevant than it was to other areas of aviation. For example, automotive firms with no prior aviation manufacturing experience were asked to adapt the Rolls Royce Merlin piston engine for mass production (Hyde 2013). Other firms, notably Pratt & Whitney, had extensive experience manufacturing piston engines (the type of engine used in the majority of WWII planes) for aircraft. However, the most commonly used types of engines in World War II like the Pratt & Whitney R-2800, were only designed in 1937 and flown for the first time in 1940. (Connors 2010). Other commonly used engines, like the Wright R-3350 used for the famous B-29 bomber, were developed around the same time (LeMay & Yenne 1988). Therefore, while engines were not totally novel at the outbreak of war, they were not fully mature products; furthermore, new designs and changes for mass production were common (White 1995).

Ships. There were two principal categories of ships produced in WWII: transport vessels created for the Maritime Commission and warships made for the Navy.

The Liberty Ships created for the Maritime Commission have been much studied (Thornton & Thompson 2001, Thompson 2001, 2007, 2012). The War Production Board (1945) table states that 53 million deadweight tons of cargo ships were manufactured in WWII. As Thompson documented, there was little pre-war experience in the manufacture of transport ships. 2.7 million deadweight tons of cargo ships were made during the First World War (Ayres 1919). We used $\zeta_i = 2.7/53 \approx 0.05$ for all ships except combatant ships.

We used several different sources for the warships. The first was George (n.d.), which showed that the U.S. had 297 warships at the end of World War 1. The second were the naval treaties agreed upon by the Great Powers during the 1920s and 30s [21]. In addition, we used Roosevelt's 1938 "Message to Congress Making Recommendations for Defense" [22] to work out the number of larger ships built after the treaties

[21] These were the Washington Treaty and the First and Second Treaties of London

[22] https://www.mtholyoke.edu/acad/intrel/interwar/fdr11.htm

lapsed, which showed the numbers of larger ships produced between World War I and World War II. The third was the Dictionary of American Naval Fighting Ships [23], which showed the number of smaller ships and submarines the US produced from the end of World War 1 to the end of 1942. The fourth was the US Navy's Naval Heritage and History Command's record of the size of the US Navy over time [24], which showed the number of each class of ship produced over the course of World War II. Combatant ships are battleships, carriers, escort carriers, cruisers, destroyers, frigates (or 'escort destroyers' as they were called at that time), and submarines. From the sources, we calculated that 889 combatant ships were produced before the war, and 733 during the war. Hence, we used $\zeta_i = 889/733 \approx 1.21$ for combatant ships. From the final source, we calculated the number of landing vessels, patrol boats, and mine craft produced before and after the war. There were 121 landing vessels produced before the war, and 2426 produced during the war. Hence, we used $\zeta_i = 121/2426 \approx 0.05$ for landing vessels. There were 515 patrol boats produced before the war, and 689 produced during the war. Hence, we used $\zeta_i = 515/689 \approx 0.75$ for patrol boats. There were 263 minelayers and minesweepers produced before the war, and 323 produced during the war. Hence, we used $\zeta_i = 263/323 \approx 0.81$ for mine warfare.

Guns and Small Arms. While the production of planes continued for civilian consumption during peacetime in the interwar period, the production of weaponry like guns and small arms slowed significantly between the wars. As stated in Herman (2012), the U.S. army shrunk to being only the 18th largest army in the world before World War II. We exploit this fact to use weapons produced during World War I as our proxy for prior experience manufacturing guns, small arms, ammunition and bombs.

Drawing on information from Broadberry & Harrison (2005) and primary source material from Ayres (1919), we were able to estimate the production of a variety of artillery and guns during World War I. For example, there were 3,077 complete units of artillery equipment manufactured, 226,557 machine guns, 3.43 million rifles and 1.7 million pistols and revolvers. The respective numbers for each of these categories produced during World War II were 7,803 artillery units ($\zeta \approx 0.39$), 2.68 million machine guns ($\zeta \approx 0.08$), 6.5 million rifles ($\zeta \approx 0.53$) and 2.74 million pistols and revolvers ($\zeta \approx 0.621$). Based these ratios, and for simplicity, we assume $\zeta_i = 0.4$ for all items in this category.

Ammunition and General Purpose Bombs. The numbers of ammunition and general purpose bombs produced in World War I were available from the same sources. In World War I 20.3 million artillery rounds were produced and 3.5 billion rounds of ammunition for rifles, revolvers and other small arms. In World War II these numbers were 33.5 million ($\zeta \approx 0.61$) and 41.5 billion ($\zeta \approx 0.08$) respectively. We used $\zeta \approx 0.35$ in these categories.

It is slightly harder to match numbers for conventional bombs. However, we know that 132 million pounds of "high explosives" – an essential component for all bombs –

[23] https://www.hazegray.org/danfs/

[24] https://www.history.navy.mil/research/histories/ship-histories/us-ship-force-levels.html

were produced during World War I. While it is hard to do a clear match of this explosive component to the reported weight of bombs in the War Production Board (1945) table, the closest category reported – "Aircraft bombs (Army and Navy), General Purpose and Demolition" – states that 7.1 billion pounds of bombs were produced, suggesting $\zeta \approx 0.02$. Acknowledging that this seems very low, our choice of the prior experience coefficient for this category is 0.1.

Aircraft bombs, Not General Purpose. The secondary literature generally agrees that there was almost no production of incendiary or fragmentation bombs during World War I, and no testing of this materiel between the wars. Therefore, we can assume prior experience of almost 0 for these models of explosives (Ross 2003, Hecks 1990). According to the War Production Board (1945) table, there were 2.26 million incendiary, fragmentation and armor-piercing bombs produced. We used $\zeta = 0.001$.

Combat and Motor Vehicles. This category unites products that were similar to products that U.S. manufacturers were already producing, such as jeeps and trucks, with others, like tanks, for which they had almost no prior experience. We were able to find disaggregated numbers for many of these products. For example, we know that prior to 1940 a cumulative number of 4.89 million trucks were registered in the United States (Cain 2006). During the war, 2.45 million trucks were manufactured, both for use on the front and for the armament effort at home, suggesting a $\zeta \approx 1.99$. In contrast, prior experience in manufacturing tanks was very low. Only 799 tanks were produced during World War I and there was no military demand for further production in the interwar period (Ayres 1919). During World War II, 86,333 tanks were produced. Since apart from tanks, all products in this category are similar to trucks, we use a category-level $\zeta = 1.99$, except for one product in the "Motor Carriages for Self-propelled Guns" category, which is a light tank chassis, for which we used $\zeta = 1$.

Communications and Electronics. Similar to Combat and Motor Vehicles, this category groups together products that manufacturers had a wide range of experience producing. In particular, there is a clear distinction between radios and radar. Radios were extensively manufactured prior to World War II, primarily for civilian and commercial use. Approximately 86,400 radio sets had been produced by U.S. manufacturers by the end of 1940 (Cain 2006). In stark contrast, only 22 radars had ever been made globally prior to 1940. Only the British "Chain Home" system was operational before 1940, with the first stations opening in 1938. Therefore, we apply prior experience corrections at the product level. Furthermore the table in *War Production Achievements* that we use for aggregate production numbers does not report communications and electronics output in terms of individual units, but rather in dollar values. Therefore, exceptionally for this category, we have summed cumulative production stated in Crawford & Cook (1952). According to this aggregation, 940,852 ground radio sets were manufactured, including vehicular radios, plus 1.25 million air radios were manufactured ($\zeta \approx 0.04$). Just over 66,000 radar sets were completed for ground and airborne use, suggesting truly negligible prior experience. (This number excludes the transponders and fuses that were attached to American materiel for friend-or-foe recognition.). We applied the above-mentioned experience factors for radar and radios, the same as radar for underwater sound equipment, and $\zeta = 0.1$ for products in the "Miscellaneous wire communication" category, mostly wires and cables.

Other Supplies. This is a broad category that unites products used to outfit, house, feed and provide medical treatment for soldiers, as well as machinery and construction equipment. Prior experience varied greatly for these products,

In textiles and household-like goods, such as clothing, tents and cannisters, U.S. manufacturers had extensive prior experience. A priori, WWII military production of clothing compared to previous clothing produced should be small, suggesting a very high ζ. It was surprisingly difficult to find estimates of prior production denoted in units (rather than dollars) for these categories. Therefore, we have to use very rough estimates of prior production from Carter et al. (2006) to estimate prior experience in this category of products. This series allows us to estimate numbers for manufactured apparel, specifically men's and boys suits and coats, back to 1927. In total, 178,496,000 of these items of clothing were produced from 1927 to 1940. This gives a lower bound to be compared with the 428,316,000 items of clothing (not including socks) manufactured for soldiers and sailors during the War, suggesting $\zeta = 0.41$. Since this is clearly lower bound, we assume $\zeta = 1$.

For aviation gasoline, we used the same estimate as for aircraft, $\zeta = 0.2$.

The Machinery category includes mostly construction equipments, such cranes, showels, road rollers, and road scrapers, and railroad equipment includes one model of locomotive. Assuming that prior experience was probably lower than for trucks and automobile (2), but higher than for most other products, we chose $\zeta = 1$.

The miscellaneous equipment categories includes everything from sleeping bags to airplane hangers, through insecticide and steel drums. Overall they tend to be items for which there existed significant prior experience, and we chose $\zeta = 0.5$.

Medical supplies and medicines–like morphine, penicillin, sulfa drugs and plasma–were mass produced for the first time during World War II (Rostker 2013). Therefore, we estimate prior experience for this sub-category to be very low–1% of WWII output. The total output of these products during World War II was 6 billion ampules. We assumed an experience correction factor of 10% because the category including Medical Supplies also includes subsistence rations. Note that we have no product for this category in the OMPUS data,but we will use this estimated prior experience in the Contract Prices dataset,using a procedure which we now explain in details.

B.2 Estimates of prior experience for the *Contract Prices* data

To obtain estimates of prior experience at the level of War departments/Army Service Forces, we take advantage of the fact that we have already justified prior experience coefficients at a lower aggregation level in the previous section. We manually construct a concordance table between each sub-category of the OMPUS Table of Content (ToC) and the War department services (see Table 10). We rely on Crawford & Cook (1952), the source of the War department data, where for each War department there is also a finer grained decomposition for the quantity of individual goods in each War

Department. We compared the items in these finer grained data to the OMPUS categories, and assigned an OMPUS category to a War Department if the goods that War Depatment procured matched the OMPUS category. In the cases where multiple department procured the same good, we assigned it to the department that procured the most of the good. We also supplemented this by consulting extensive histories of the divisions from the Army and military historians (Coates Jr 1959, Coker & Stokes 1991, Crawford & Cook 1952, Killblane 2012, Mauroni 2015, Risch 2014, Rubis 2012). For example, we assigned the 'mortar shells' category to 'Chemical', as Crawford and Cook list the 'Chemical' department as procuring the majority of mortar shells.[25]

War department	ζ	Z_0	Method
Total War Department	0.4472	52845	Sum of sub-departments
AAF Total	0.2096	9421	Average of corresponding OMPUS categ.
ASF Total	0.5931	43423	Sum of sub-services
ASF Chemical	0.2	353	Our assumption
ASF Engineers	0.75	3774	Our assumption
ASF Medical	0.1	81	Average of corresponding OMPUS categ.
ASF Ordnance	0.6158	22437	Average of corresponding OMPUS categ.
ASF Quartermaster	0.6333	14561	Average of corresponding OMPUS categ.
ASF Signal	0.0259	106	Average of corresponding OMPUS categ.
ASF Transportation	1	2112	Average of corresponding OMPUS categ.

Table 11: Estimated prior experience for the *Contract Prices* data

We computed the prior experience coefficient of a War Department as the average of the prior experience coefficient of its associated OMPUS ToC sub-categories. We thus obtained prior experience coefficients for the AAF and for 5 of the ASF subcategories. For the categories ASF Chemical we put 0.2, and for Engineers 0.75.

To get an estimate of prior experience for the aggregate services ASF, we sum up the estimated prior experience of the corresponding subservices. The bottom part of Table 11 reports the estimated values of prior experience Z_0 for the ASF subservices. We sum up these values to obtain Z_0 for ASF Total. The table reports the corresponding value of ζ for information only, we do not use it to estimate Z_0. We proceed similarly for the Total War department, which is the sum of ASF and AAF.

B.3 Estimates of prior experience for the *Labor Productivity* data

The *Labor Productivity* productivity data is mostly at the plant (Source Book) or company (Ford) level. There is an issue with correcting this data for prior experience

[25] Sometimes our mappings are unintuitive due to quirks in how the US procured goods. For example the $M2$ mortars the US used were originally designed to only fire smoke shells as the US peace lobby opposed the use of high explosives and chemical shells after WWI. Hence the Chemical department dealt with the ammunition. But during WWII they adapted them to fire high-explosive ammunition. Thus the Chemical department handled all types of ammunition, even though it was mainly high-explosives and thus seems more likely to be handled by Ordnance.

because if a plant enters in production late, it will not produce a lot and since the factor ζ is applied to total plant-level production, plants that arrive late and benefit most from past experience actually get a lower estimated prior experience. We decided to apply no correction to the data presented in the main text. See Appendix D.4 for the results when we apply a correction for initial experience.

C Estimators for the *OMPUS-USMH* dataset

In the *OMPUS-USMH* dataset, there are two observations per product, but the dates and span of these observations differ across products. Thus, we cannot estimate a *first*-differences model. In this appendix we discuss the "heterogenous-differences" (HD) model, in which we regress the differences in log cost on the differences in log experience and the time span between the two observations (and where no constant is allowed). We show that it results in the same regression coefficients as the fixed-effects (FE) model used in section 6.2, but that it is different from the cross section regression of the product's average monthly growth rates. We also discuss how the share of cost decrease due to the exogenous time trend is computed is Section 7.1.

Equivalence between the point estimates of the HD and FE estimators We define the Heterogenous Differences (HD) estimator as follows. For each individual i, there are two periods t_{0i} and t_{1i}, and we denote the span of time between the two as

$$t_{1i} - t_{0i} = \tau_i.$$

We define the HD operator Δ^{τ_i} on a variable V as the difference of the two available observations

$$\Delta^{\tau_i} V = V(t_{1i}) - V(t_{0i}).$$

Obviously,

$$\Delta^{\tau_i} t = t_{1i} - t_{0i} = \tau_i,$$

and

$$\Delta^{\tau_i}\text{constant} = \text{constant} - \text{constant} = 0,$$

so that applying the operator Δ^{τ_i} to our main equation

$$\log c_{i,t} = \text{constant} + \alpha t + \beta \log Z_{i,t} + \epsilon_{i,t}$$

gives

$$\Delta^{\tau_i} \log c_i = \alpha \tau_i + \beta \Delta^{\tau_i} \log Z_i + \Delta^{\tau_i} \epsilon_i \tag{20}$$

Note that just like in first differencing, applying the Δ^{τ_i} operator leads to the loss of one observation. Since there are two observations per individual to start with, there is now only one observation per individual, so we have removed the subscript t. Note also that there is no intercept in Eq. 20.

We define the HD estimator as the OLS estimator of Equation 20,

$$\hat{\beta}_{HD} = (X^T X)^{-1} X y, \tag{21}$$

where $y = \Delta^{\tau_i} \log c_i$ and, denoting τ_i as x_{i1} and $\Delta^{\tau_i} \log Z_i$ as x_{i2},

$$\underset{n\times 2}{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix}$$

To show that it is equivalent to the Fixed Effects estimator, we observe that when there are only two observations, the "within" transformation is almost equivalent to applying the Δ^{τ_i} operator. The within transformation consists in substracting the group-specific mean from each observation, that is, for a variable V with two observations in t_1 and t_0, the within transformation operator $\mathcal{W}$ gives

$$\mathcal{W}(V(t_{1i})) = V(t_{1i}) - \frac{V(t_{0i}) + V(t_{1i})}{2} = \frac{V(t_{1i}) - V(t_{0i})}{2} = \frac{\Delta^{\tau_i} V_i}{2}$$

and similarly

$$\mathcal{W}(V(t_{0i}) = -\frac{\Delta^{\tau_i} V_i}{2},$$

so the within and HD transformations are very similar. But in contrast to the HD transformation, the within transformation does not reduce the number of observations, so a direct comparison of the matrix of regressors is not possible. However, we can write the within estimator as

$$\hat{\beta}_{FE} = (\tilde{X}^T \tilde{X}) \tilde{X} \tilde{y}, \tag{22}$$

where

$$\underset{2n\times 2}{\tilde{X}} = \frac{1}{2} \begin{bmatrix} x_{11} & x_{12} \\ -x_{11} & -x_{12} \\ x_{21} & x_{22} \\ -x_{21} & -x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \\ -x_{n1} & -x_{n2} \end{bmatrix}.$$

Now, if we compute the entries of $X^T X$ and $\tilde{X}^T \tilde{X}$, we find $\tilde{X}^T \tilde{X} = \frac{1}{2} X^T X$ and thus

$$(\tilde{X}^T \tilde{X})^{-1} = \left(\frac{1}{2} X^T X\right)^{-1} = 2(X^T X)^{-1}. \tag{23}$$

Similarly, writing down explicitly the entries of $\tilde{X}\tilde{y}$ and simplifying shows that

$$\tilde{X}\tilde{y} = \frac{1}{2} X y \tag{24}$$

Putting Eqs. 24 and 23 into 22, and comparing with Eq. 21, we see that

$$\hat{\beta}_{FE} = \left(2(X^T X)^{-1}\right)\left(\frac{1}{2} X y\right) = (X^T X)^{-1} X y = \hat{\beta}_{HD}.$$

■

Non-equivalence between the HD and the growth rates cross-section estimator. When faced with heterogenously spaced data with two observations per individual, another option would simply be to calculate average growth rates, and perform a cross-sectional regression, that is

$$\frac{\Delta^{\tau_i} \log c_i}{\tau_i} = \alpha + \beta \frac{\Delta^{\tau_i} \log Z_i}{\tau_i} + \text{noise}. \tag{25}$$

As can readily be seen by comparing the matrix of regressors, the coefficients estimated from Eq. 25 and the HD/FE estimator Eq. 20 are in general different, so Table 4 (columns 4-6) also reports the estimates based Eq. 25.

Share of cost decrease explained by the exogenous time trend To do a growth decomposition for the *OMPUS-USMH*, we take expectations of Eq. 20

$$E[\Delta^{\tau_i} \log c_i] = \alpha E[\tau_i] + \beta E[\Delta^{\tau_i} \log Z_i],$$

so that the share of the change in cost that is explained by the exogenous time trend is

$$\text{share exo} = \frac{\alpha \frac{1}{n} \sum_i \tau_i}{\frac{1}{n} \sum_i \Delta^{\tau_i} \log c_i} \tag{26}$$

D Robustness checks

D.1 Time series analysis

We can perform time series analysis only in the *Labor productivity* and *Contracts* datasets, and they have different structures (unbalanced and $N > T$ for the first, but balanced and $T > N$ for the second).

Thus we first present results that relate to the time aspect of the models and that can be computed on all three datasets: two-way fixed effects, and using the lag (instead of contemporaneous) experience as regressor. We then proceed to discuss time series properties in the *Labor productivity* and *Contracts* datasets in turn.

We omit the specifications with production as a regressor.

Two-ways fixed effects In the main specification, we constrain the effect of the time variable to be an exponential trend. Instead, we can estimate the fixed effect results when both individual and time dummies are included, that is

$$\log c_{it} = \kappa_i + \theta_t + \beta \log Z_{it} + \eta_{it}. \tag{27}$$

This allows us to control for economy-wide (by contrast to product-specific) effects on costs that are not necessarily growing exponentially in time. Table 12 reports the results[26], showing that an estimate of the effect of experience similar to that obtained with the FE allowing only for individual dummies and an exogenous linear time trend.

[26]The models are estimated by performing a single transformation (removing individual means) and adding time dummies, and the R^2 are the R^2 of this regression.

Table 12: Panel regression results for Two-way fixed effects

	Labor Productivity	USMH	Contracts
Experience	-0.300***	-0.059**	-0.167*
	(0.018)	(0.020)	(0.058)
Observations	3034	1046	308
R^2	0.789	0.250	0.803

First lag of experience We use the first lagged value of experience instead of contemporaneous experience as a regressor. The results, presented in Table 13, do not change much as compared to the baseline results, except for the FD estimator in the Labor Productivity dataset where the experience coefficient drops by a half and the exogenous time trend instead increases. The coefficient for experience in the *OMPUS-USMH* is also somewhat weaker than in the baseline results.

Table 13: Panel regression results for Experience lagged 1 period

	Labor Productivity		USMH	Contracts	
	FE	FD	FE	FE	FD
Experience(t-1)	-0.236***	-0.109***	-0.033**	-0.149*	-0.120*
	(0.019)	(0.020)	(0.013)	(0.051)	(0.045)
Time	-0.008**	-0.036***	-0.005***	-0.003	-0.004
	(0.003)	(0.004)	(0.001)	(0.003)	(0.003)
N	2912	2719	1046	308	301
R^2	0.717	0.046	0.159	0.786	0.060

Labour productivity We first test the null of no first-order autocorrelation in the Fixed Effects results using experience and time as regressors (Second column of Table 3), using Wooldridge's (2002) (section 10.5.4) test, and strongly reject it ($p < 0.001$).

A possibility is that the variables exhibit unit roots, however because our panel is unbalanced our options for testing are limited. We use Fisher-type tests (Choi 2001), which consists in applying a standard test (Augmented Dickey-Fuller or Phillips-Perron) to each time series, aggregating the p-values, and testing the null hypothesis that all panels contain unit roots, against the alternative that at least one panel is stationary. The Fisher-type tests are based on the $T \to \infty$ asymptotic, with finite or infinite N depending on the statistics. All four statistics derived by Choi (2001) deliver a near zero p-value for both the log of experience and the log of manhours per unit, so we reject that all series contain a unit root.

Since there is autocorrelation but it is not as strong as to suggest a first-difference model, we follow two separate directions. First, we simply estimate a Fixed Effects model with autocorrelated errors (Baltagi & Wu 1999), with two different estimators for the autocorrelation parameter (Durbin Watson or the autocorrelation of residuals, both computed on the within transformed data). These two methods of computing

Table 14: Time Series models for *Labor Productivity*

	AR1		Lagged Dep.Var.		
	DW	Corr.	OLS	OLS	Arellano-Bond
Experience	-0.169***	-0.349***	-0.219***	-0.029**	-0.026
	(0.019)	(0.016)	(0.029)	(0.009)	(0.100)
Time	0.004*	0.008***	-0.002*	-0.003***	-0.003
	(0.002)	(0.002)	(0.001)	(0.001)	(0.010)
Manhours(t-1)			0.771***	0.772***	0.744***
			(0.022)	(0.022)	(0.211)
Experience(t-1)			0.159***		
			(0.022)		
Observations	2882	2882	2830	2830	2660
AR(1)	0.83	0.65			
Experience, long-run			-0.260	-0.126	-0.101

autocorrelations result in noticeably different results, yet in both cases the sign of the exogenous time trend is reversed and the coefficient of experience remains important and strongly significant (first two columns of Table 14).

The second approach is to consider that residual autocorrelation is caused by misspecification, whereby the lagged values of the regressor and/or the regressand are missing. The more complete model

$$\log c_{it} = \kappa_i + a \log c_{i,t-1} + \beta \log Z_{i,t} + \beta_2 \log Z_{i,t-1} + \eta_{it}.$$

nests a large class of dynamic linear models (Hendry 1995). The results of estimating this using OLS are in column 3, showing coefficients of opposite signs for experience and its lagged value. Removing the lag of experience (col 4), the autoregressive parameter remains the same, and the estimated long-run effect[27] decreases by half. Finally (col. 5), although we have a fairly "long" panel whereby the Nickell bias is unlikely to be large, we estimate the same equation using the two step Arellano & Bond's (1991) estimator with all possible instruments and robust (Windmeijer) standard errors, and find similar results.

Contracts. We start again with Wooldridge's test for AR(1) residuals and as for the *Labor productivity* data, we strongly reject the null of no autocorrelation. However, in contrast to the *Labor productivity* data, all the unit root tests we performed (Choi 2001, Im et al. 2003, Levin et al. 2002, Hadri 2000, Breitung 2001, Harris & Tzavalis 1999) systematically suggested that both the log of contract prices and the log of production experience have unit roots.

[27] This is estimated as as the sum of the coefficient(s) of experience divided by one minus the autoregressive parameter.

Thus we tested for co-integration, using Pedroni's (1999) test statistics. None of the 7 test statistics suggested rejection of the null of no co-integration. This is not too surprising, as we do not see a compelling reason for the existence of a strong relationship between the *levels* of cost and experience, so that a departure from this long term level relationship would imply an error-correction behavior and a return to this trend[28]. Overall, these tests suggest that costs and experience are related in difference, that is, an change in experience is associated with a change in costs.

D.2 Coefficient heterogeneity

In the main text, we have reported results with either both α and β common across all products, or separate regressions for each product. Here we investigate the results when we constrain only one of the two parameters to be the same across products and allow the other one differ.

Time trend heterogeneity only. We can investigate heterogenous time trends by estimating a fixed effect regression on the first differenced values, that is

$$\Delta \log c_{it} = \kappa_i + \beta \Delta \log Z_{it} + \eta_{it}. \tag{28}$$

The results in Table 15 show results that are robust in the case of the Labor Productivity dataset, but evaporate entirely in the Contracts data. This is not too surprising given what we had reported using individual-level regressions in Figs. 2 and 4.

Table 15: Fixed effects on the first differences

	Labor Productivity	Contracts
Experience	-0.210***	0.004
	(0.020)	(0.083)
Observations	2830	301
R^2	0.105	0.000

Experience effect heterogeneity. We can estimate heterogeneous slopes for experience using Swamy's (1970)'s random coefficients model and the Mean Group estimator of Pesaran & Smith (1995). We estimate these two models on the first differenced variables.

The results are in Table 16. The point estimates in the *Labor Productivity* dataset suggest a stronger effect of experience than in our main specification, while the reverse is true in the *Contracts* dataset. In all cases, however, the standard errors are such that the distributions for the coefficients presented here and the coefficients estimated in the main text overlap significantly.

[28] Note also that experience cannot decrease, which would imply additional restrictions on the error-correction model.

Table 16: Heterogenous coefficients models (Swamy and Mean group)

	Labor Productivity		Contracts	
	Swamy	MG	Swamy	MG
Experience	-0.272**	-0.362***	-0.031	-0.035
	(0.087)	(0.082)	(0.061)	(0.047)
Constant	-0.013	-0.015	-0.007*	-0.008*
	(0.016)	(0.015)	(0.004)	(0.003)
Observations	2817	2817	301	301

D.3 Instrumenting by lagged values

Because unit and labor costs are total costs divided by output, output appears on both sides of the equations. Table 17 reports instrumental variable estimates for the main specification (first-difference) using the first lag of (log) production as instrument for (log) production, and first lag of (log) experience as instrument for (log) experience. We cannot perform this robustness check for the *USMH* data due its structure.

Table 17: Instrumental variable estimates

	Labor productivity		Contracts	
Experience	-0.209***	-0.219***	-0.123***	-0.124***
	(0.021)	(0.018)	(0.033)	(0.029)
Time	-0.021***	-0.021***	-0.004	-0.004**
	(0.005)	(0.004)	(0.002)	(0.001)
Production	-0.025		-0.007	
	(0.014)		(0.083)	
N	2578	2719	301	301
R^2 within	0.70	0.70	0.76	0.77
R^2 overall	0.00	0.00	0.12	0.13

The results are very similar to those reported in the main text. We also performed these regressions using the fixed effect, rather than first-difference, estimator, again finding results very similar to those reported in the main text.

Using lagged regressors is not a fully convincing IV strategy. Reed (2015) shows that it is effective when the lagged regressors are themselves not regressors in the true data generating process, and when the lagged regressors are sufficiently correlated with the (instrumented) regressors.

We have considered other instrumental variable approaches. One approach is to use demand side instruments, for instance battle-related variables, as one may have thought that higher losses of materiel led to an increase in production, and was not correlated with weapons costs decreases. However, as we argued from historical analysis, production operated at maximum capacity, guided by long-term (yearly) targets, and thus was *not* driven by battlefield losses. A second approach would have been to use supply-side instruments, such as the provision of raw materials; because these were in very short supply, they constrained production but their supply may have been argued to be unrelated to cost decreases (but note that there is evidence of induced technical change during the war to save on raw materials). Here we faced the issue that it is virtually impossible to construct product-level instruments.

D.4 Initial experience

Contracts To evaluate the robustness of the results to a different evaluation of the prior experience coefficient, we re-construct experience using values of ζ multiplied by a factor f, with $f = 0 \ldots 5$. Fig. 10 reports the results, showing that indeed the results would change noticeably if we misestimated prior experience by an important factor. The third panel, which shows the share of the decrease in cost attributed to the exogenous time trend, shows that if the true ζs were all 20 times lower ($f = 0.05$), the exogenous time trend would explain all cost decrease, and the "learning" parameter would be close to zero. For even smaller values of f, the sign of $\hat{\beta}$ would even change. On the other hand, if we misestimated all the prior experience coefficients by a factor of 2 ($f = 1/2$ or $f = 2$), say, our main conclusion would not be fundamentally affected.

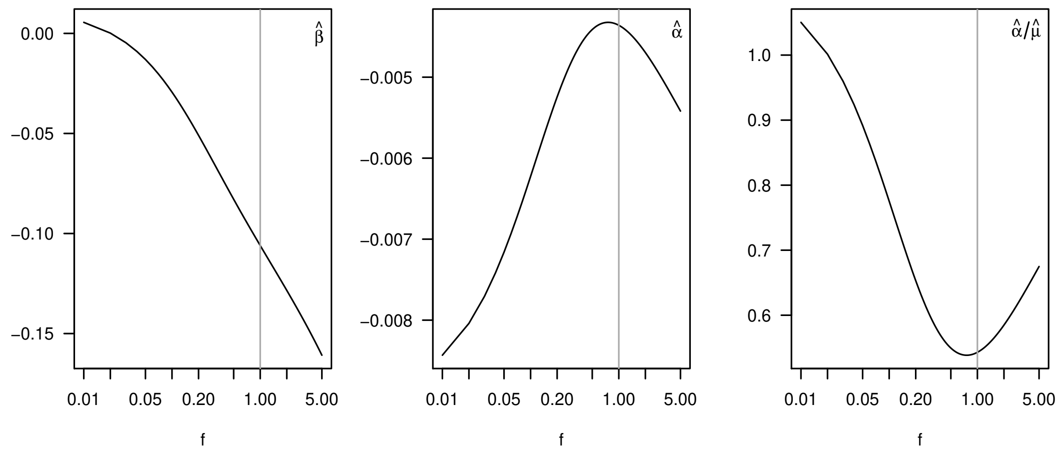

Figure 10: Estimated coefficient of the first difference regression of the log of contract prices on the log of experience and an exogenous time trend, for different values of a factor f that multiplies our baseline vector of estimates of prior experience ζ. The rightmost panel shows $\hat{\alpha}$ divided by the average cost decrease $\hat{\mu} = -0.008$ (as reported in Table 6).

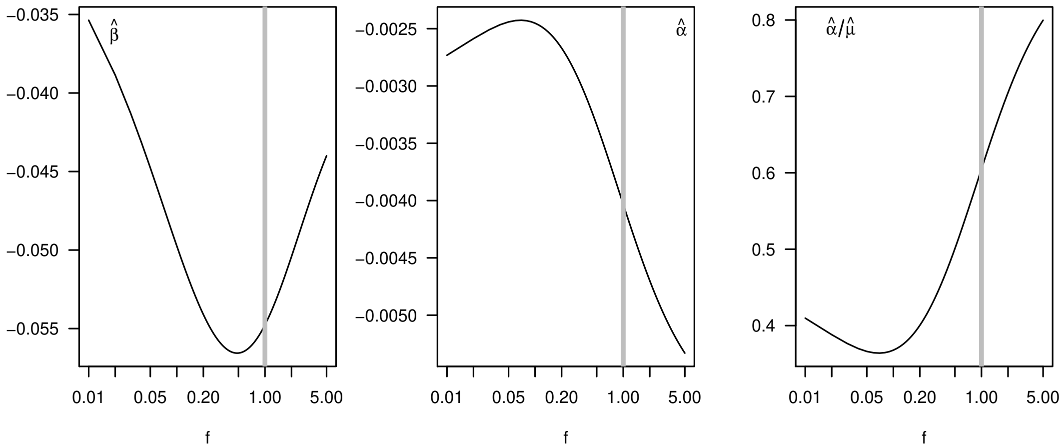

Figure 11: Estimated coefficient of the fixed-effect regression of the log of USMH unit costs on the log of experience and an exogenous time trend, for different values of a factor f that multiplies our baseline vector of estimates of prior experience ζ. The rightmost panel shows $\hat{\alpha}$ divided by the average cost decrease $\hat{\mu} = -0.008$ (as reported in Table 6).

OMPUS-USMH. Fig. 11 shows the robustness of the results to a change of the estimates of initial experience by a factor f, as for the *Contracts* data. Again, the results do change sensibly, but overall the results are fairly robust: it would take a very different, implausible change to the estimates of initial experience to alter our conclusion that experience and the exogenous time trend both explain an important share of the cost trend.

Labor Productivity. In the main text, we did not use data corrected for prior experience (see Section B.3 for a discussion). If we apply the corrections suggested by the discussion in Section B.1 for more aggregated categories, we would apply $\zeta = 0.2$ for aircraft, $\zeta = 0.05$ for ships, and, say, $\zeta = 1$ for Ford. In this subsection we apply this correction, and take it as a baseline on which we apply a factor f as above (for $f = 0.01$, the coefficients correpond almost exactly to the coefficients reported in Table 3, where $f = 0$). Again we observe some change in the results, but the overall qualitative conclusion remains.

We also note that increasing prior experience tends to worsen the problem of collinearity. The estimated effect of experience on individual time series is less robust to the inclusion of a time trend. In Fig. 2, where there is no prior experience correction ($f = 0$, i.e. $\zeta = 0$), the points lie fairly well along the unit line and the correlation between $\hat{\beta}_i(\alpha_i = 0)$ and $\hat{\beta}_i(\alpha_i \neq 0)$ is 0.37. In the baseline correction ($f = 1$), this correlation falls to 0.16.

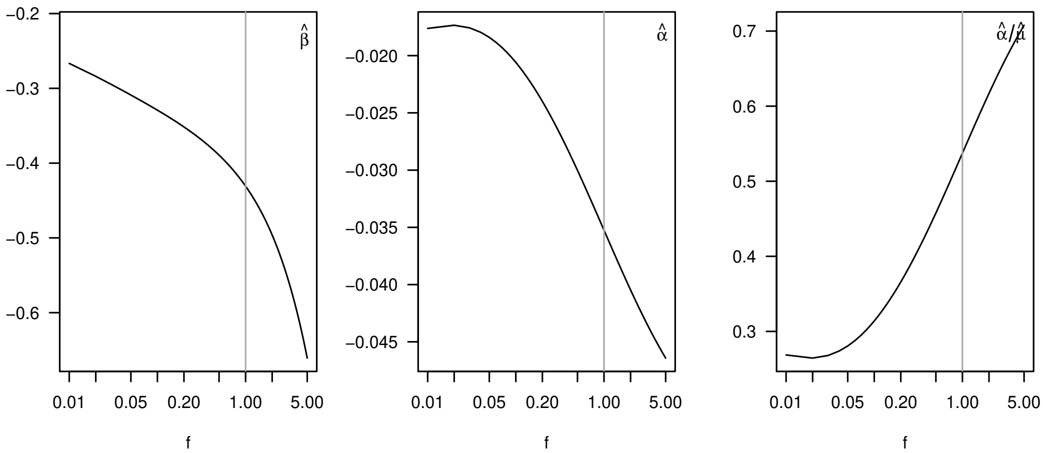

Figure 12: Estimated coefficient of the first difference regression of the log of manhours on the log of experience and an exogenous time trend, for different values of a factor f that multiplies the vector of estimates of prior experience ζ described in this appendix. The rightmost panel shows $\hat{\alpha}$ divided by the average cost decrease $\hat{\mu} = -0.066$ (as reported in Table 6).

The robustness checks described here do not account for the fact that we may have mis-estimated prior experience coefficients much more in some categories than in others. We do not report specific robustness checks for this, but during the process of revising the estimates of the individual ζs, we have re-estimated our main specification several times and while the results somewhat change, as above the main result is not fundamentally affected, with both experience and the exogenous time trend explaining important shares of cost decrease.

D.5 Comparing the datasets

The *OMPUS-USMH* and *Contracts* datasets contain, in principle, overlapping information. Many detailed products in the *OMPUS-USMH* form part of the basis for the price indices in the *Contracts* datasets. To give estimates of prior experience, in Section B.3, we have built a concordance table between the OMPUS Table of Content (ToC) and the War department services (Table 10). Here we exploit this concordance table to compare the estimated coefficients in Fig. 13.

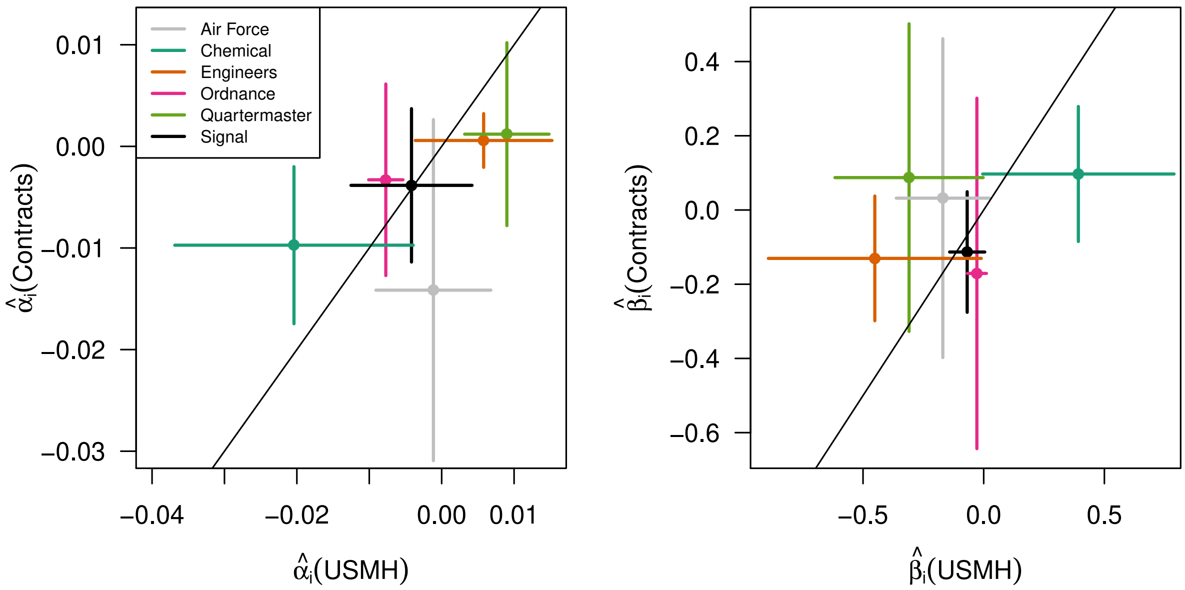

Figure 13: Comparison of the *OMPUS-USMH* and *Contracts* sector-level results.

The estimates of the exogenous technological progress $\hat{\alpha}$ first (left panel) are fairly similar, with a correlation of 0.57. In contrast, the estimates of the effect of experience $\hat{\beta}$ can be quite different, though the correlation remains around 0.33. The estimates for Quartermaster, in particular, are very different. However, the standard errors are very large, and often overlap the identity line, suggesting that the two datasets do not necessarily provide significantly different estimates, and legitimizing our approach of pooling the different war departments.

D.6 Controlling for inflation

During the war, the prices of inputs, including wages, tended to increase, although moderately because of price controls. These input price changes bias our estimates of the effect of experience, which are likely to be higher than what we measure under the assumption of constant input prices.

We cannot control for the price of inputs precisely, due to the lack of available data at a granular level, so we have to resort to an aggregate price index. Of course, even within each dataset the products are quite heterogeneous in terms of their input mix. To control for input prices in this context, we also show a specification which allows each product to have a separate coefficient for the effect of the price index, that is, we interact the price index with the individual dummies (for *OMPUS-USMH*, we used war departments instead of individual products as the basis for interaction terms; for *Labor productivity*, 5 interacted dummies are removed because of perfect multicollinearity).

We used the Producer Price Index for All Commodities (PPIACO), available from FRED. Table 18 reports the results, showing that our main results do not change substantially.

Table 18: Adding PPI as dependent variable

	Labor Productivity		USMH		Contracts	
Experience	-0.219***	-0.222***	-0.061***	-0.043*	-0.110*	-0.104
	(0.022)	(0.022)	(0.017)	(0.018)	(0.042)	(0.054)
Time	-0.024***	-0.023***	0.003	-0.003	-0.005	-0.005
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
PPI	0.899		-3.901*		0.277*	
	(0.593)		(1.752)		(0.095)	
PPI Interacted	No	Indiv.	No	War Dep.	No	Indiv.
N	2830	2830	1046	1046	301	301
R^2	0.126	0.177	0.174	0.224	0.057	0.072

D.7 External learning

A large literature has looked at experience spillovers explicitly, attempting to estimate cross-plant or cross-product spillovers by regressing costs of product i on experience producing i and experience producing j. Here we attempt to capture spillovers at the larger level of the war economy, by constructing an aggregate time series of ”War Effort”. A negative effect on cost would indicate spillovers, whereas a positive effect would suggest that aggregate production negatively affects individual products productivity, perhaps due to scarce inputs, which was the case in the war economy. We take the quantity index for the whole War Departments from the *Contract Prices* dataset, that is, the solid black line in Fig. 8. Cumulative War effort is the cumulative, using the estimated prior experience from Table 11.

Table 19: Adding Total War effort as dependent variable

	Labor Productivity		USMH		Contracts	
Experience	-0.218***	-0.218***	-0.051**	-0.058**	-0.106	-0.117*
	(0.022)	(0.022)	(0.017)	(0.018)	(0.043)	(0.044)
Time	-0.021***	-0.021***	-0.001	-0.005***	-0.004	-0.007
	(0.004)	(0.004)	(0.002)	(0.001)	(0.003)	(0.003)
War Effort	0.012		0.049*		-0.000	
	(0.010)		(0.020)		(0.008)	
Cumul. War Effort		0.022		-0.033		0.129
		(0.014)		(0.041)		(0.066)
N	2830	2830	1046	1046	301	301
R^2	0.126	0.126	0.174	0.167	0.047	0.050

The results are in Table 19. The estimated effects of the “War effort” variables are inconsistent across datasets, but the effects of individual products' experience do not change as compared to our main estimates.

D.8 Depreciation of experience

An important issue in the literature is whether applying a depreciation to experience improves the fit. Usually, one specifies a perpetual inventory method formula for experience and attempts to estimate the depreciation rate. For instance, Levitt et al use non-linear least squares.

A specific problem we have here is that assuming depreciation should logically imply that we decrease the estimate of previous experience. Unfortunately, we were able to give estimates of previous experience but not of how it unfolded over time - thus we cannot easily apply a depreciation factor to it.

Here we simply omit this issue, and take the same estimates of initial experience as in the main text[29]. We then cumulate production using a depreciation factor, as follows

$$Z_t = \delta Z_{t-1} + q_t$$

Instead of estimating δ, we fit the model for a range of values of $\delta \in (0.8, 1)$ and provide the R^2 of the regression (to show what Nonlinear Least Squares would estimate), the estimated coefficients for time and experience, and the implied share of exogenous progress.

[29] Assuming a lower initial experience, e.g. depreciated initial experience = initial experience /6, does not change the patterns in Fig. 14.

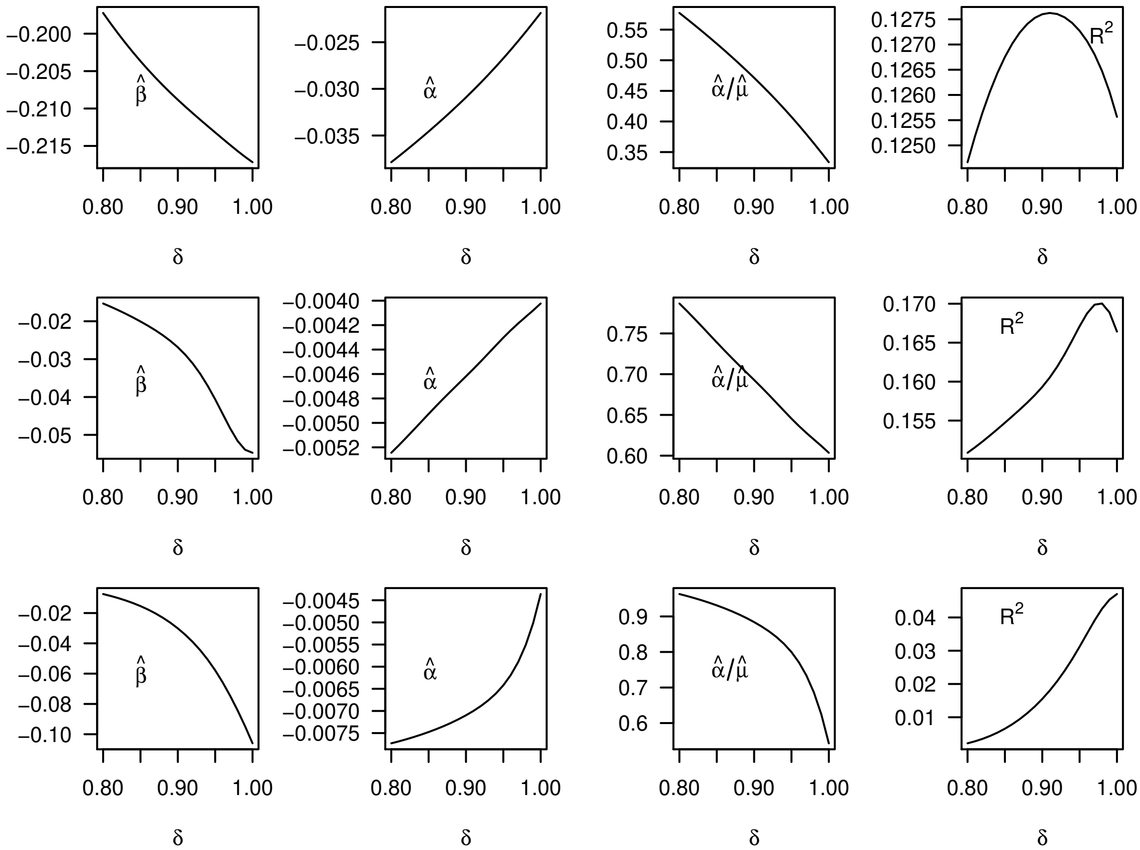

Figure 14: Change of the main result as a function of the depreciation of experience parameter δ. Top: *Labor productivity*; Center: *OMPUS-USMH*; Bottom: *Contracts*

The results are in Fig. 14. The robust pattern that emerges is that allowing $\delta < 1$ would always make $\hat{\beta}$ less negative and $\hat{\alpha}$ more negative, and the estimate of the share of exogenous progress larger. However, the best fit models would imply no or only a moderate increase of the estimated share of exogenous progress. For instance, in the "worst" case, *Labor productivity*, $\hat{\delta} = 0.91$ implies a substantial annual depreciation of $0.91^{12} = 0.32$, but the share of exogenous progress rises only from 0.33 to 0.46.