LEARNING BY NECESSITY:

GOVERNMENT DEMAND, CAPACITY CONSTRAINTS, AND PRODUCTIVITY GROWTH

Ethan Ilzetzki*

London School of Economics

April 2, 2024

Abstract

This paper studies how firms adapt to demand shocks when facing capacity constraints. I show that increases in government purchases raise total factor productivity in quantity units at the production-line level. Productivity gains are concentrated in plants facing tighter capacity constraints, a phenomenon I call "learning by necessity". Evidence is based on newly digitized archival data on US World War II aircraft production. Shifts in demand across aircraft with different strategic roles provide an instrument for aircraft demand. I show that plants adapted to surging demand by improving production methods, outsourcing, and combating absenteeism, primarily when facing tighter capacity constraints.

*Contact: e.ilzetzki@lse.ac.uk. I thank the editor, Emi Nakamura, and the annonymous referees for useful comments and editorial guidance. I thank Lillabelle Aikins, Mary Akinrogbe, Mun Fai Chan, Hugh Hanley, Gonzalo Huertas, Balázs Marko, Tiago Paúl, Anna Pilipentseva, Hugo Reichardt, Laura Richardson, and Martin Souchier for outstanding research assistance; and Tom McAnear, Tab Lewis and the rest of the National Archives staff in College Park, Debbie Seracini at the San Diego Air and Space Museum Archives, Archie Difante and Tammy Horton at the Air Force Historical Research Agency, and Randy Sowell at the Harry S. Truman Library for their help in locating archival materials. I also thank Ufuk Akcigit, Borağan Aruoba, Francesco Caselli, Gabriel Chodorow-Reich (discussant), Jeremiah Dittmar, László Dózsa (discussant), John Fernald, Alex Field, Luca Fornaro, Andy Garin, Alessandro Gavazza, Michaela Giorcelli, Refet Gürkaynak, Josh Hausman, Kilian Huber, Rustam Jamilov (discussant), Xavier Jaravel, Karel Mertens, Mary O'Mahony (discussant), Steve Pischke, Valerie Ramey, Maarten de Ridder, Hugh Rockoff, Barbara Rossi, Mark Schaffer, Jón Steinsson, Claudia Steinwender (discussant), Johannes Wieland, Mark Wilson, Alex Whalley, Noam Yuchtman, and seminar participants at Maryland, Toulouse, Ben Gurion University, the Bonn Macrohistory Lab, INSEAD, the AEA meetings, Nottingham, the IMF, Duke, Johns Hopkins (SAIS), Rutgers, U Chicago, the NBER Summer Institute, CREI, Queen Mary's Belfast, Bank of Finland, the CEPR Joint Conference "New Avenues for Monetary Policy", DG-ECFIN, the Bank of England, CBI Netherlands, Salento Macro Meetings 2022, IDC Hertzlia, Hebrew University, Riksbank, IIES Stockholm, ZEW Public Finance Conference, Banco de España-World Bank Workshop: "How to Improve Public Spending", and IFW-CEPR Conference on Geoeconmics for their useful comments. This project received financial support from UKRI (ERC replacement) grant EP/X025543/1 and from the Centre for Macroeconomics.

How do firms satisfy increased demand for their products when facing tight capacity constraints? The conventional answer is that they can't because demand has no effect on firms' productivity. An alternative view posits that firms can increase productivity when facing demand shocks and that high demand induces innovation that circumvents capacity constraints. This is a common interpretation of the performance of the US economy during the Second World War: Although the US was close to full employment by the time Pearl Harbor was attacked, munitions production nevertheless surged at declining production costs. This observation also spurred post-war research on *learning by doing*: a term that encompasses the many ways plants improve productivity with experience. This work has been extremely influential, fostering a literature on learning by doing (see Thompson 2012, 2010 for reviews) and endogenous growth (Romer 1986, Lucas 1988, 1993).

However, existing empirical work on this topic has some limitations. It has mostly skirted the identification challenge of differentiating whether increased demand, production, and experience lead to higher productivity, or vice versa. I show that this is not merely a theoretical concern but rather that traditional learning-by-doing (LBD) regressions show substantial pre-trends and clear indications of reverse causation. Further, while the focus has often been labor productivity, Thompson (2001) shows the importance of capital in evaluating learning by doing, and Basu *et al.* (2006) demonstrate the importance of capital utilization in measuring productivity over the business cycle. Thompson (2010) notes that the the concept of the "experience curve" is vague: whether productivity gains are passive or driven by active responses to higher demand. Finally, most learning-by-doing studies give static estimates of the experience curve. This is an important drawback: I show almost no contemporaneous effect on productivity, with effects peaking only after a year.

In this paper, I address some of these shortcomings. I utilize detailed archival data on US aircraft production during World War II, including previously untapped measures of physical capital and capital utilization, essential for evaluating total factor productivity (TFP). I address the identification challenge with an instrumental variable. I use the national output of broad aircraft types in each month as a ("leave one out") instrument for aircraft demand in each production line in that month. True, procurement was channeled to plants the military and government expected most likely to deliver aircraft rapidly, *within* broad aircraft types (e.g. which plant should deliver fighter aircraft). However, as outlined in Section 2.3, the allocation of national procurement *across* these broad aircraft types (e.g. the decision of whether to buy more fighter or bomber aircraft) was driven by external factors such as military strategy, combat losses, and battlefield circumstances.

Using a two-way-fixed effects Local Projections Instrumental Variables estimator, I plot the dynamic response of productivity to aircraft demand. I observe a negligible immediate impact on productivity. Instead, I find a delayed response of 0.4% growth in quantity-based, and capital-utilization adjusted, Total Factor Productivity (TFPQ) within 12-months of a 1% demand shock.

I then investigate the role of capacity constraints on the learning curve. I illustrate in a simple model why firms facing capital and labor adjustment costs and convex utilization costs are induced to adopt new production techniques when demand surges. Critically, convex utilization costs mean that technology adoption incentives are greater under tighter capacity constraints, as plants operate on steeper portions of their cost curves. Hence productivity responds more to demand (learning curves are steeper) when plants are already operating at high utilization. I refer to this phenomenon, where increased demand confronts limited production capacity and leads to productivity growth as *Learning by Necessity*. Indeed, I show that plants with higher capacity utilization rates see 80% higher productivity growth in the year following a demand shock. I measure capacity constraints using several separate indicators, all based on new archival data: capital utilization derived from shift-utilization data; labor utilization (weekly hours per worker); high-wage labor markets; and the War Manpower Commission's classification of labor markets by tightness.

Finally, I document several active measures taken by plants to increase productivity. First, production methods in the aircraft industry changed dramatically during the war. The most prominent improvement was the move from job-shop (custom and nearly handmade) production methods to production line methods (standardized products, interchangeable parts, smaller tolerances). Utilizing newly collected data from historical news sources and firms' annual reports, I present suggestive evidence that plants that gained high experience were more likely to adopt new production methods, but only if they were high utilization plants. Second, the airframe industry moved from mostly in-house production to greater reliance on outsourcing and subcontracting, and I find greater such reliance in capacity-constrained plants facing demands shocks. Third, management made concerted efforts to improve working conditions and worker morale, to reduce absenteeism and turnover. I use newly digitized archival data on absenteeism to show that plants with higher labor utilization lost *fewer* labor hours to absenteeism in response to demand shocks.

Previous research has documented learning by doing in aircraft (Wright 1936, Middleton 1945, Asher 1956, Alchian 1963, Rapping 1965) and shipbuilding (Searle 1945 Thompson 2001) industries. These estimates were based on correlations, lacking a causal interpretation. Recent studies have proposed instruments for experience: Benkard (2000) uses lags of global GDP and oil prices as instruments and Levitt *et al.* (2013) use the experience of one production line as an instrument for another in the same plant. Both studies are for a single plant rather than an entire industry and the latter measures production defects rather than labor productivity. Neither controls for capital or its utilization, nor do they provide dynamic estimates that control for lagged demand, which I show to be important in uncovering the causal impact of demand on productivity. Most importantly, this paper is the first to document how learning by doing interacts with capacity constraints: learning by necessity.

An extant macroeconomic literature has estimated returns to scale in industry-level production functions (Hall 1990, Burnside 1996, Basu & Fernald 1997). Existing learning by doing estimates typically ignore returns to scale. Estimates of returns to scale typically assume that demand cannot affect productivity. I provide a framework that nests the two and separate the effect of demand on productivity from static returns to scale. A literature in international trade emphasizes the importance of market size on productivity and innovation (Acemoglu & Linn 2004; Finkelstein 2004 De Loecker, 2007, 2011; Atkin *et al.* 2017 Melitz & Redding 2023). But these focus on the long-run, not business cycle frequency, and don't speak to the importance of capacity utilization.

The notion that demand may affect productivity is implicit in the endogenous growth literature and more explicit in the literature on induced innovation (Romer 1987, Newell *et al.* 1999, Popp 2002). Recent work has further posited that cyclical demand could spur productivity through similar channels (Benigno & Fornaro 2018, Moran & Queralto 2018, Anzoategui *et al.* 2019, and Jordà *et al.* 2020). In an early contribution, Hickman (1957) posited that high utilization could lead to capital investment incentives, what he called "the acceleration princple". Arthur (1989) outlined a theoretical non-linear relationship between technology adoption and demand. In a case study of a single aircraft plant in World War II, Mishina (1999) documents high turnover rates and therefore views experience as a less plausible explanation for productivity growth in the plant. Instead he suggests a phenomenon of "learning by stretching," a precursor to the concept of "learning by necessity" of this paper.

There is a voluminous literature studying the effects of government purchases on the economy, and military spending has been used to identify government spending shocks (Barro 1979, Ramey & Shapiro 1998, Barro & Redlick 2011, Ramey 2011a,b, 2016, 2019, Nakamura & Steinsson 2014, Chodorow-Reich 2019, Auerbach *et al.* 2020). Unlike the extant literature, this article doesn't focus on the aggregate effects of public expenditures on GDP, private consumption, or unemployment (the fiscal multiplier), but rather on its effects on productivity and its dependence on capacity utilization. Antolin-Diaz & Surico (2022) show that the effects of aggregate US military spending are long-lived and that it stimulates innovation and private investment, consistent with the mechanisms studied here. Brunet (forthcoming) uses Wold War II procurement data to study the effects of government spending on output and employment using state-level variation. This paper also speaks to the debate on the dependence of fiscal multipliers on the degree of slack in the economy (Auerbach & Gorodnichenko 2012, 2013, Owyang *et al.* 2013, and Ramey & Zubairy 2018), and to Boehm & Pandalai-Nayar's (2022) finding that supply curves are convex. A large literature has studied the longer-run impact of World War II public spending (Rhode 2000, Fishback & Cullen

2013, Jaworski 2017, Hanlon & Jaworski 2021). Rockoff (2012) and Fishback & Jaworski (2016) give broader reviews of the literature on the impacts of World War II on the post-war economy.

Finally, the paper relates to a literature on capacity utilization, its response to demand shocks, and as a confounding factor in productivity measurement (Burnside & Eichenbaum 1996, Basu *et al.* 2006). This paper shows that TFPQ grows in response to demand shocks (and is procyclical) even controlling for increased utilization, with real productivity gains, not merely reflecting mismeasurement. Additionally, plants with high rates of utilization see relatively higher productivity growth when faced with rising demand, indicating a richer interaction between the business cycle, capacity utilization, and productivity than previously documented.

Admittedly, this paper speaks most directly to the effects of government aircraft purchases during the Second World War. The results suggest that high demand could spur productivity growth in other settings. However, there are some aspects of the war economy that may not translate neatly to a peacetime setting, e.g. workers' patriotism and price controls. Also, the aircraft industry may have been ripe for mass production on the eve of the war, making it particularly poised to "learn by necessity". I discuss concerns of external validity in Section 3.4 and Appendix D. While acknowledging these concerns, I note that it is also possible to overstate the uniqueness of the period. Aircraft firms were exempt from price caps; wages were frequently re-negotiated; and worker strikes and absenteeism were at historical highs, indicating that mundane motivations persisted alongside patriotism.

The remainder of the paper is organized as follows. Section 1 describes the data and the historical setting. Section 2 lays out the empirical strategy with the main results shown in Section 3. Section 4 gives a history and empirical evidence of the actions taken by plants to increase productivity. Section 5 concludes.

1 Data, Institutional Setting, and Historical Context

World War II brought the largest cyclical increase in public consumption in US history. Figure 1a shows government consumption as a share of GDP in the US from 1929 to today. The Second World War stands out as the single largest shock to government purchases. Government consumption and gross investment rose from 9% of GDP at the war's onset to 44% of GDP in 1945, declining again to 16% by 1948. The precise unemployment rate at the onset of World War II is debated, but it is generally agreed that the US economy was approaching full employment by the time the US officially entered the war in late 1941 (Figure 1b, Gordon & Krenn 2010).

The analysis that follows narrows in on aircraft production, which was the single largest procurement item in the military budget and became the largest industry during the war (War Production Board 1945 charts 3 and 11). Figure 1c shows that aircraft procurement peaked at 4% of GDP.

In May 1940, after the fall of France, President Roosevelt set an ambitious objective of producing 50,000 planes during the war (Fireside chat, May 26 1940). Economists Robert Nathan and Simon Kuznetz estimated that the US didn't have the productive capacity to meet this aim. Yet the US aircraft industry produced twice this number of aircraft in 1944 alone (War Production Board 1945 p. 10).

The aircraft industry was young: the average firm was founded in 1927 and the average plant in 1934. Table 1 gives summary statistics for the industry. In total, 38 firms operated 61 plants and produced 109 different aircraft models, with 141 plant-by-model combinations. For simplicity, I refer to plant-by-model combinations as "production lines," although some plants ran several production lines for the same model. The median firm was a single plant producing a single aircraft model. However, there was considerable variation: the 90th percentile firm operated three plants and the 90th percentile plant produced four models. Firm and plant sizes also varied significantly with the 75th percentile firm selling a total of $1.2 billion in aircraft, nearly 50 times more than the 25th percentile firm; and the 75th percentile plant employing 15,000 workers, almost 10 times the 25th percentile plant.

The industry was less concentrated than before or after the war: The Herfindahl-Hirschman Index of aircraft sales (by dollar value) declined slightly from 0.14 in 1939 to 0.10 in 1945, but rose again to 0.11 in 1947 (Reichardt, 1975). Douglass Aircraft, the leading firm, produced only 13% of all aircraft, by sales, a modest proportion by modern aircraft industry standards. There was just one acquisition (Vega by Lockhead) and one merger (Consolidated with Vultee) during the war, and only three small firms exited at the war's end. In contrast, the industry's post-war history has been one of consolidation and concentration: By the time of the Boeing-McDonnell Douglas merger in the late 90s, the industry's Herfindahl index was estimated at around 0.5 (Stock, 1999).

Procurement was under the purview of the Army Air Force and the Navy, in coordination with the War Production Board, which dictated the overarching war production strategy. Most contracts were Cost Plus Fixed Fee, covering suppliers' (audited) costs plus a pre-negotiated payment per aircraft delivered. Concerns over war profiteering led to a legal cap on markups (to 4% by the end of the war) and some contracts were renegotiated ex-post. As a consequence, most aircraft manufacturers' profit margins were lower than they were before or after the war (Smith 1991 pp. 248-293; Wilson 2018, chapter 4).

Aircraft firms, their subcontractors, and their suppliers were exempt from wartime price controls. While wages were were regulated and frozen at their March 1942 levels, they were frequently re-negotiated, leading to a 20% increase in the aircraft industry by 1945 (Smith 1991 pp. 399-403). Previously, most aircraft were made to order based on detailed specifications of the procuring agency, but this became untenable given the new production targets. The military therefore agreed to purchase standardized aircraft models, which were then modified in army or navy modification centers. Standardized aircraft aides productivity analysis, as it ensures consistent specifications across aircraft of the same model and mark.

Productivity data come from the Aeronautical Monthly Progress Reports, collected by the Army Air Force headquarters at Wright Field (later published in USAAF 1952). The military meticulously tracked war production, with all aircraft manufacturers submitting monthly reports. Productivity and production data from this source (Table 3) have been used in previous research, but previous researchers overlooked a second volume of reports, including detailed data on floor space, worker hours, and shift utilization (Tables 5 and 6).[1] To my knowledge, I am the first to have digitized these additional data. Reporting requirements and forms were the same for all plants and were extremely detailed. Figure A.1 in the appendix shows one of the standardized forms.[2]

Productivity measurement starts with the raw variable "Unit Man Hours: Entire Plane," which reports the worker-hours of the last plane delivered in the calendar month. This includes only manufacturing workers: overhead is reported separately. The measure includes hours worked in sub-assemblies, giving a consistent comparison when producers outsourced parts of the production.[3] The variable gives hours per aircraft at the product level, addressing the multi-product plant problem. There are benefits to measuring productivity at the aircraft level, but the last aircraft may be unrepresentative of the plant's average productivity. For comparison, I computed monthly labor productivity by dividing total aircraft deliveries by payroll hours for manufacturing workers, as is commonly done. The two measures are highly correlated but the comparison underscores the advantage of direct aircraft-level measurement. The aircraft-level data incorporates hours across all production months (USAAF 1952 p. 37): important, because production typically exceeded a calendar month (45 to 90 days in the case of Consolidated Vultee bombers, based on data from Consolidated Vultee archives, San Diego Air and Space Museum, Box 17). In contrast, dividing the number of aircraft by hours worked in the current month creates a mismatch between delivery

[1]Data reporting began in 1941, with 60% coverage prior before January 1943, 100% thereafter. However, this was the initial production date for most production lines.

[2]The military also gave plants a 150 page document with minute detail on how to uniformly report production, productivity, capacity utilization, and other data. (ATSC Regulation No. 15-36-3, Air Force Historical Research Agency, Maxwell Field, Reel A2050, starting on slide 850.) Consolidated Vultee Archives, San Diego Air and Space Museum, Box 34 documents how the second largest producer (by revenues) adopted these procedures internally.

[3]USAAF (1952) p. 37 states that these are "direct hours charged to a model... obtained from shop or worked orders and do not refer to payroll hours... They refer to hours expended on the airframe manufacturing process which includes machining, processing, fabricating, assembling, and installing all integral parts of the airframe structure, and rework prior to acceptance." Outsourced production hours are "the estimated direct man-hours it would require to perform within the facility that part of the airframe manufacturing process being produced outside the plant or plants of the reporting facility." The output per hour variable can then be seen as the number of hours worked to produce the portion of the aircraft that was produced in house. On one hand, this introduces some measurement error because the reporting plant is estimating the number of hours it would have taken to produce in-house the portion of production that was outsourced. On the other hand, this has the advantage that we no longer have to concern ourselves with differences in capital per worker between the main facility and feeder plants.

time and production time and severely misstates productivity at the beginning or end of a production batch. The running variable of monthly aircraft production is also from USAAF (1952), Table 3 with Civilian Production Administration (1947) Table 1, pp. 32-55, used to bring coverage from 60% to 100% prior to 1943.

The literature estimating production functions rarely observes plants' physical capital stock. Instead, the nominal (dollar value of the) capital stock is estimated by accumulating past (nominal) investment expenditure, or taken from accounting statements. In many cases, structures are largest component of capital expenditure and such nominal estimates of the capital stock confound differences in land prices and construction costs with real differences in the capital stock. In contrast, USAAF (1952), Table 5, gives a rare proxy for plants' physical capital stock: plant-level quarterly observations of the floor space actively used for manufacturing, measured in square-feet. This measure of physical capital is more comparable across plants and time. Further, the measure includes only floor area actively used for production and therefore incorporates capital utilization to some extent. It excludes office space and other non-production facilities.[4]

Plants also recorded all investments in plant and equipment exceeding $25,000, giving a measure of capital deepening.[5] Structures were the largest component (60%) of capital investment in the airframe industry during the war and we will see in Section 3 that investment in structures and equipment are both highly correlated with future physical floor space, indicating that capital expenditures only translate in to productive capital with a substantial lag.

Figure 2 shows time series indexes of aggregate aircraft production, hours worked, and floor space, from 1942 to 1945. It displays the number of aircraft (top panel) and total aircraft weight (bottom panel); the latter was used by by contemporary researchers to adjust for larger aircraft's greater production complexity. The figures give clear initial evidence of the great increase in productivity during the war. While hours worked and capital grew in tandem by a factor of close to 2.5 to 1944, aircraft production increased by a factor of 3.5, suggesting TFP growth of 35%, under a homogeneous of degree one production function. When measured in units of aircraft weight the growth is even more dramatic at approximately 250%.[6]

The data in USAAF (1952) (Table 6) also give a rare account of capital and labor utilization that hasn't been used in previous research. It includes details on the number of work shifts per day, the number of hours in each shift, and the number of monthly worker-hours active in each one of

[4] I interpolate quarterly floor space to monthly and allocate capital across production lines in the plant to equate the capital to labor ratio across all production lines within a plant in each month, as would optimally occur with standard constant returns to scale production functions. This assumes that the wage rate and rental rate of capital are the same across production lines, which is reasonable given that it was often the same workers shifting across production lines.

[5] "War Manufacturing Facilities Authorized by State and County,"War Production Board Program and Statistics Bureau, June 15, 1945. RG 179, box 984, NARA College Park

[6] This contrasts with Field's 2008; 2018; 2002 evaluation that TFP declined for the US economy as a whole in the war due to mis-allocation *across* industries. Be this as it may, Appendix Figure A.2 shows that productivity dispersion–often used to measure misallocation–across WWII plants declined over the course of the war, *within* the industry.

the shifts in each month. I use these to calculate shift utilization to capture capital utilization, as was done during the war and as suggested more recently by Basu *et al.* (2006). Scheduled working hours in the most active shift, always the Monday morning shift, are used to gauge production potential, with full capacity measured as the number of weekly work hours that would result if the plant operated 24×7 hours a week at this potential. Capital utilization is the ratio between *actual* monthly work hours and full capacity.[7] Additionally, I measure labor utilization as average weekly hours per worker, taken from the same table.[8]

Figure 3 shows the evolution of capital and labor utilization in the median airframe plant. Capital utilization was high and rising in the first year of direct US involvement in the war, peaking at 52%, nearly 90 hours a week. This is perhaps an unremarkable capital workweek by 21st century standards, but was well above typical pre- and post-war utilization rates of around 35% (60 hours per week). The year 1943 sees a surge in aggregate productivity (Figure 2), but a rapid *decline* in capital utilization for the remainder of the war. This suggests that the observed productivity surge was not merely high utilization masquerading as TFP. Instead, it appears that productivity growth substituted for high utilization rates, allowing plants to decrease utilization. The bottom panel of the figure reveals a similar trend in labor utilization, with the average production worker in the median plant working nearly 50 hours a week in 1942; this declines to roughly 45 hours a week by the end of the war.

2 Empirical Strategy

This section describes the paper's empirical strategy, beginning with a conceptual framework that motivates the estimation of "learning by necessity". I then compare this empirical strategy with the existing literature on learning by doing. Finally, I address the identification of demand shifts.

2.1 Conceptual Framework

This section outlines a simple theory of "learning by necessity": how high demand, relative to existing production capacity, induces productivity growth. It is used to frame the empirical analysis. For a more detailed treatment of a dynamic version of the model, refer to Appendix B.

[7] Wartime reports and the data suggest that the use of second shifts, night shifts, and Saturday shifts were the main source of variation in capacity utilization both over time and across plants.

[8] Shift utilization is imperfectly correlated with hours per worker (with a coefficient of 0.23). Shift utilization may seem like it measures labor utilization but it is better thought of a measure of capital utilization. For example, the Martin plant in Omaha had very high average weekly hours per worker (51.3) in early 1942, because many of its workers worked 7 days a week. However, it had very low capital utilization (37%) because the plant mostly worked 9-to-5, with very few workers in a limited evening shift and no night shift. In contrast, workers in the Douglas plant in Santa Monica worked 40 hours per week, but the plant had a high capital utilization (65%) rate because the plant spread its 15,000 workers nearly evenly over 3 shifts a day (operating 6 days a week).

Consider a plant p receiving orders to produce $Y_{mp,t}$ units of an aircraft model m in month t using capital $K_{mp,t}$. Capital is fixed so that $K_{mp,t} = K_{mp}$, but the plant can choose its rate of capital utilization $U_{mp,t}$: utilized capital in period t is $U_{mp,t}K_{mp}$. In the appendix, I extend the model to allow for costly capital adjustment, one period in advance. The empirical analysis incorporates labor as as a second factor of production and the appendix model shows that that the plant chooses to move the utilization of the two factors in tandem, so that no insights are lost in the single-factor model. When using a technology $z_{mp,t}$, the firm produces $F\left(U_{mp,t}K_{mp,t}|z_{mp,t}\right) = z_{mp,t}\left(U_{mp,t}K_{mp}\right)^{1-\alpha}$ units of the final good, with $0 < \alpha < 1$. Utilization incurs costs $\delta\left(U\right)$ per unit of capital, where $\delta\left(.\right)$ is increasing and convex and satisfies $\delta\left(0\right) = 0$. The utilization cost function represents maintenance and depreciation costs that increase with utilization.

In each period, the plant can operate a traditional technology z^T or upgrade to a modern technology $z^M > z^T$ at a monthly fixed cost A_{mp}. This cost could be a financial fixed cost, the cost of exerting managerial effort, or any other costly action that enhances productivity. I outline specific actions undertaken to enhance productivity in World War II aircraft production in Section 4 and Appendix E. Each firm draws the adoption cost from a uniform probability distribution $G\left(A_{mp}\right)$ with support $A_{mp} \in [0, \bar{A}]$. For simplicity, I also assume that technology is entirely reversible in each period, allowing for a static technology choice. It may seem peculiar that the firm cannot adjust factors of production but choose technology freely, but these assumptions are both relaxed in the dynamic model in the appendix. There, the firm makes a one-off and irreversible choice of technology but can adjust factors of production in each period at a cost. In the simple model presented here, fixed factors of production are necessary for meaningful factor utilization choices, and the flexible technology choice helps clarify the concept of "learning by necessity".

The plant chooses technology and utilization to minimize costs, period by period,

$$\min_{U_{mp,t}, z_{mp,t} \in \{z^T, z^M\}} \delta\left(U_{mp,t}\right) K_{mp} + A_{mp}\mathbb{1}\left(z_{mp,t} = z^M\right),$$

subject to satisfying demand,

$$z_{mp,t}\left(U_{mp,t}K_{mp}\right)^{1-\alpha} \geq Y_{mp,t}. \tag{1}$$

If plants receive a fixed payment per aircraft, cost minimization is equivalent to profit maximization. A cost-plus-fixed fee contract gives weaker incentives to minimize costs (see Section 3.4), but doesn't eliminate cost-savings incentives entirely, because future procurement contracts depend on plants' relative performance. For simplicity, I maintain the cost-minimization assumption, but note that incentives may be more subtle and complex, as evaluated in the literature on optimal procurement (McCall 1970, Laffont & Tirole 1988,Laffont & Tirole 1993, Bajari & Tadelis 2001; see Appendix D.

The problem boils down to a discrete choice of technology, whereby the firm chooses the modern technology if

$$C_{mp,t} = K_{mp}\delta\left(\frac{1}{K_{mp}}\left(\frac{Y_{mp,t}}{z^T}\right)^{\frac{1}{1-\alpha}}\right) - K_{mp}\delta\left(\frac{1}{K_{mp}}\left(\frac{Y_{mp,t}}{z^M}\right)^{\frac{1}{1-\alpha}}\right) \geq A_{mp},$$

where $C_{mp,t}$ are cost savings from choosing the modern technology. The two arguments of the $\delta\left(.\right)$ functions are the required utilization rates from (1) when choosing technologies z^T and z^M, respectively. Log-linearizing cost savings in month t around its value in period $t-1$ gives[9]

$$\Delta C_{mp,t} \cong K_{mp}U_{mp,t-1}\left[\delta'\left(U_{mp,t-1}\right) - \left(\frac{z^T}{z^M}\right)^{\frac{1}{1-\alpha}}\delta'\left(U_{mp,t-1}\left(\frac{z^T}{z^M}\right)^{\frac{1}{1-\alpha}}\right)\right]\Delta\log Y_{mp,t}. \tag{2}$$

The term in brackets is positive if $\delta''(.) > 0$, therefore cost savings are increasing in demand. Intuitively, high demand increases the marginal cost of utilization and more so on the steeper end of the cost curve, where the plant finds itself if it uses the traditional technology. The plant adopts the modern technology if $C_{mp,t} > A_{mp}$, which occurs with frequency $G\left(C_{mp,t}\right)$. Therefore,

$$E\log z_{mp,t} = G\left(C_{mp,t}\right)\log z^M + \left(1 - G\left(C_{mp,t}\right)\right)\log z^T. \tag{3}$$

A Log-linearized version of this equation implies

$$E\Delta\log z_{mp,t} \cong \frac{1}{\bar{A}}\log\left(\frac{z^M}{z^T}\right)\Delta C_{mp,t}. \tag{4}$$

Combining (2) with (4) gives

$$E\Delta\log z_{mp,t} \cong \Upsilon\left(U_{mp,t-1}\right)\Delta\log Y_{mp,t}, \tag{5}$$

where

$$\Upsilon\left(U_{mp,t-1}\right) \equiv \frac{K_{mp}U_{mp,t-1}}{\bar{A}}\log\left(\frac{z^M}{z^T}\right)\left[\delta'\left(U_{mp,t-1}\right) - \left(\frac{z^T}{z^M}\right)^{\frac{1}{1-\alpha}}\delta'\left(U_{mp,t-1}\left(\frac{z^T}{z^M}\right)^{\frac{1}{1-\alpha}}\right)\right].$$

A Taylor expansion of $\Upsilon\left(U_{mp,t-1}\right)$ around its value at the median plant, $\bar{U}_{t-1}$, is:

$$\Upsilon\left(U_{mp,t-1}\right) \cong \Upsilon\left(\bar{U}_{t-1}\right) + \Upsilon'\left(\bar{U}_{t-1}\right)\left[U_{mp,t-1} - \bar{U}_{t-1}\right].$$

I show in Appendix C that $\Upsilon\left(\bar{U}_{t-1}\right) > 0$ always, and that $\Upsilon'\left(\bar{U}_{t-1}\right) > 0$ if (but not only if) $\delta'''(.) \geq$

[9] The two-dimensional linearization strategy used here draws on Boehm & Pandalai-Nayar (2022).

0.[10] Combining this last equation with (5) motivates an estimating equation of the form

$$\Delta \log z_{mp,t} = \beta_1 \Delta \log \left(Y_{mp,t}\right) + \beta_2 \left[U_{mp,t-1} - \bar{U}_{t-1}\right] \Delta \log \left(Y_{mp,t}\right),$$

where $\beta_1 = \Upsilon\left(\bar{U}_{t-1}\right)$ and $\beta_2 = \Upsilon'\left(\bar{U}_{t-1}\right)$.

For the practical task of estimation, I modify this equation in a few ways. First, I include fixed effects and lags of the explanatory variable. I discuss their importance for causal inference below. Second, with lags of the dependent variable, we can use the level of $\log Y_{mp,t}$ rather than its first difference, which is useful because the instrument, discussed shortly, is more predictive of the monthly level of demand than its month-on-month growth. Third, we measure utilization $U_{p,t-1} - \bar{U}_{t-1}$ at the plant level at the war's onset, rather than with a single lag, because utilization early in the war is less likely to be endogenous to current productivity growth. Fourth, we transform the continuous variable $U_{p,0}$ into a binary dummy variable that takes on the value of 1 if the plant was above the sample median of capital utilization. This eases interpretation of the coefficient, which becomes a comparison between plants that had high and low capital utilization. The robustness exercises in the following section include a regression with the continuous value of utilization. Finally, the specification is dynamic and allows for lags between the demand shock and the time of technology adoption, for gradual technology adoption, or for gradual effects of technology adoption on TFP. This is done through a local projections specification as follows:

$$\Delta_h \log z_{mp,t+h} = \alpha_{mp} + \alpha_t + \beta_h^{LBD} \log Y_{mp,t} + \beta_h^{LBN} \mathbb{1}\left(U_{p,0} > \bar{U}_0\right) \log Y_{mp,t} + \text{controls} + \varepsilon_{mp,t}^h, \quad (6)$$

where α_t and α_{mp} are month and production line (plant-by-model) fixed effects. The operator Δ_h gives the growth rate of a variable from month $t-1$ to $t+h$, so that $\Delta_h \log z_{mp,t+h} \equiv \log z_{mp,t+h} - \log z_{mp,t-1}$. The variable $\mathbb{1}\left(U_{mp,0} > \bar{U}_0\right)$ equals 1 for plants with above-median utilization at the beginning of the war and zero otherwise. All specifications include six lags of the the independent variable $Y_{mp,t}$ and some include additional controls. The direct effect of initial capacity utilization $U_{p,0}$ or $\mathbb{1}\left(U_{p,0} > \bar{U}_0\right)$ is omitted as it is absorbed by production line fixed effects α_{mp}.

There are two coefficients of interest. First, β_h^{LBD} is the traditional "learning by doing" coefficient. It measures productivity growth in a plant following a 1% increase in demand in period t. (6) is dynamic and controls for lags of the explanatory variable. Controlling for lags gives a nearly perfect correlation between current production used here and "experience", the explanatory variable in previous studies.

Second, β_h^{LBN} is the "learning by necessity" coefficient. It quantifies the differential impact of

[10] This condition holds for a quadratic cost function, for example. It also holds for cost functions that ensure that utilization is bounded, e.g. costs go to infinity as we approach full utilization. Further, the condition $\delta'''(.)$ is sufficient, but not necessary.

demand on productivity in high-utilization plants compared to those with lower utilization. In contrast, the traditional learning literature imposes $\beta_h^{LBN} = 0$.

The raw archival data report labor productivity, which I convert to TFP using a more general production function than the one outlined above:

$$Y_{mp,t} = F\left(U_{mp,t}K_{mp,t}, H_{mp,t}L_{mp,t}|z_{mp,t}\right) = z_{mp,t}\left[\left(U_{mp,t}K_{mp,t}\right)^{1-\alpha}\left(H_{mp,t}L_{mp,t}\right)^{\alpha}\right]^{\gamma},$$

where $L_{mp,t}$ is the number of production workers and $H_{mp,t}$ is hours per worker. The parameter γ allows for economies of scale, with $\gamma > 1$ representing increasing, $\gamma < 1$ decreasing, and $\gamma = 1$ constant returns to scale. With $y_{mp,t} \equiv \frac{Y_{mp,t}}{H_{mp,t}L_{mp,t}}$ denoting labor productivity, we can write:

$$\Delta_h \log z_{mp,t+h} = \Delta_h \log y_{mp,t+h} - (1-\alpha)\left(\Delta_h \log k_{mp,t+h} + \Delta_h \log U_{mp,t+h}\right) - (\gamma - 1)\Delta_h \log S_{mp,t+h}, \tag{7}$$

where $k_{mp,t} \equiv \frac{U_{mp,t}K_{mp,t}}{H_{mp,t}L_{mp,t}}$ is (utilized) capital per hour worked and $S_{mp,t} \equiv \left(U_{mp,t}K_{mp,t}\right)^{1-\alpha}\left(H_{mp,t}L_{mp,t}\right)^{\alpha}$ is production scale.[11]

2.2 Conventional Learning by Doing Estimates

The post-war learning-by-doing literature reports correlations between cumulative output and output per worker as reflecting a "learning curve". But these correlations aren't necessarily informative of demand's causal impact, because demand, experience, and productivity are all jointly determined. Reverse causation isn't merely a theoretical possibility: It is also very likely. Further, in the parlance of modern econometrics, estimated learning curves suffer from substantial pre-trends. This is illustrated in Figure 4, which shows regression coefficients in a standard learning-by-doing regression with pre- and post-trends. (Log) labor productivity (aircraft per hour) are regressed on experience (log cumulative production) and month and production line fixed effects. The existing literature reports the coefficient at $h = 0$. Horizons $h < 0$ show the correlation between current experience and *past* productivity. The regressions show strong pre-trends meaning that production lines accumulating more experience were already more productive in the preceding twelve months. Higher cumulative production at time zero is likely the result of previously high productivity. Horizons $h > 0$ show the correlation between current experience and future productivity. If anything, productivity *declines* in the months after a plant gains experience.

Mishina (1999) (pp. 148, 153) speaks to the challenges of estimating learning curves. He notes that cumulative output follows an upward trend by definition, so that any trend decline in unit

[11] I include factor utilization in "production scale," following Basu & Fernald (1997) and Basu *et al.* (2006) Results are robust to defining $F\left(U_{mp,t}K_{mp,t}, H_{mp,t}L_{mp,t}|z_{mp,t}\right) = z_{mp,t}\left(U_{mp,t}K_{mp,t}^{\gamma}\right)^{1-\alpha}\left(H_{mp,t}L_{mp,t}^{\gamma}\right)^{\alpha}$, which allows economies of scale for both utilized and unutilized capacity.

costs, or trend increase in productivity, will be attributed to "learning". This issue may even be present when using cumulative instruments, e.g. macro variables accumulated over time, as in Benkard (2000). Modern time series econometric methods, which control for lags of the running variable, are a step towards addressing this problem. However, even with two way fixed effects and lags of the running variable, it is plausible that the military diverted demand to production lines with high (anticipated) productivity. I therefore propose a an instrument for aircraft demand.

2.3 Identification Strategy

I instrument the monthly output of each production line with the aggregate output of all other production lines producing the same broad aircraft type in that month. This approach draws on historical evidence that demand for broad aircraft types (e.g. bombers vs. fighter planes) was determined by strategic considerations, not relative productivity in their manufacture. This contrasts with demand for specific aircraft models within a broad category (e.g. B-24 vs. B-17 bombers) or across plants (Douglas vs. Boeing), where demand may well have been affected by plants' relative (expected) productive capacity. I divide aircraft into six broad types: bombers, communications, fighters, trainers, transport, and other specialized aircraft. The instrument $I_{mp,t}$ for demand $Y_{mp,t}$ of aircraft model m in plant p in month t is given by $I_{mp,t} = \sum_{\pi \neq p} \sum_{\mu \in \mathbb{M}_m} Y_{\mu\pi,t}$, where $\mathbb{M}_m$ is the set of aircraft models of the broad type that includes model m.

Instrument relevance requires a correlation between production lines of the same broad aircraft type. Relevance is borne out in F statistics reported in the figures of the following section. The exclusion restriction requires that the national demand for a broad aircraft type affects the subsequent productivity growth in the production line in question only through the correlated demand directed to that production line.

The source of variation captured by the instrument is illustrated in Figure 5, which shows the number of total aircraft delivered for four aircraft types. The four faced different demand fluctuations, with known historical interpretations. Early war production was for lend-lease assistance to US allies in Europe. This primarily took the form of fighter aircraft (e.g. for the Battle of Britain), leading to a boom in fighter production in 1940-1941. Fighters were also used as escorts for US merchant ships during this period. US direct involvement in the war began in December 1941. US military strategy following Pearl Harbor anticipated a heavy reliance on aerial bombing, causing an inflection in bomber aircraft in 1942 and a surge in demand the following year.[12]

Demand for transport aircraft took off only later, supporting the island-hopping operations in

[12]The President's program of January 1942 required that "Offensive planes [be] stressed, and the war department immediately asked that the previous goal of 1,000 heavy bombers a month be increased to 2,000 at the earliest possible date," US Civilian Production Administration (1947b), chapter 46, p. 74.

the Pacific and the invasion of Italy in 1943.[13] Demand for fighter aircraft rose again in mid-1943, when it became apparent that both bomber and transport aircraft benefited from fighter escorts.[14] Trainer aircraft were naturally needed in greater quantities in the early war years than later.

A threat to identification arises if these relative demand shifts were due to differential expected productivity growth across broad aircraft types. But the historical literature indicates that strategic considerations were paramount in determining procurement schedules for broad categories of munitions. In September 1943, a report by the War Manpower Commission[15] notes that (p. 2)

> The primary purpose of the periodical overhauling of aircraft schedules is to shift emphasis from one model to another in the light of combat experience and military needs.

War Production Board (1945) p. 11 explains:

> In 1944, our war production had to meet front-line needs, constantly changing with the shifting locales of warfare, the weaknesses and strengths demonstrated in combat, and our inventiveness as well as the enemy's. Less emphasis was placed on increasing quantities of everything required to equip an army, a navy, and an air force, and more on those specific items needed to replace battle losses and to equip particular forces for particular operations.

The same document (p. 13) narrows in on aircraft production:

> The complex causation of program changes is illustrated by the aircraft program. Each quarterly aircraft schedule represented a cut under its predecessor. In part this reflected lower than anticipated combat losses... [In 1944, t]he demand for four-engine long-range heavy bombers, transport vessels and heavy artillery ammunition rose dramatically during the year, while the need for training planes, patrol vessels, mine craft, and radio equipment fell off in varying degrees.

In summary, procurement of broad categories of aircraft was driven by strategic needs, not aircraft plants' expected productivity. Of course, procurement agencies carefully monitored plant-level productivity and purchased aircraft within these broad categories from plants they viewed as most able to deliver. But this source of variation is discarded, rather than captured by, the

[13]See Air Force Historical Research Agency, Reel 1009, p. 1608 "Airborne Missions in the Mediterranean" on the use of C-47 transport aircraft for glider and paratrooper landings in operations Husky, Landbroke, and Fustan in Sicily. On transport aircraft in the North Burma campaign, see Taylor, Joe G., 1957, Air Supply in the Burma Campaign, USAF Historical Studies No. 75, USAF Historical Division, Maxwell Airforce Base, reel K1009.

[14]Major Lesher, Lee A. (1988). "The Evolution of the Long-Range Escort Doctrine in World War II" United States Air Command and Staff College. An important inflection point was a failed strategic bombing mission on Schweinfurt, Germany in August 1943, that exposed the need for fighter escorts in bombing campaigns.

[15]War Manpower Commission, "Manpower Problems in the Airframe Industry" Sep 18, 1943, RG221, 111, Box 1 National Archives College Park.

instrument. Further, technological improvements and new varieties of aircraft may have moved demand across aircraft models within broad categories (from "heavy" B-17 to "very heavy" B-29 bombers, for example), but not across the broad categories we consider (B-17 bombers to P-39 fighter aircraft), as they were hardly good substitutes in military operations.

3 Learning by Doing and Learning by Necessity

This section summarizes the main results. We begin by restricting $\beta_h^{LBN} = 0$ in (6) to consider the average response of productivity to demand, as in traditional learning-by-doing regressions. We then turn to an unrestricted version of (6), which allows an interaction between demand and capacity utilization: *learning by necessity*.

Impulse responses are based on two-stage least squares. The second stage is estimated using local projections (Jordà 2005), as in (6). In the first stage, (log) aircraft demand and its interaction with initial capacity utilization are instrumented with the (log of the) leave-one-out instrument $I_{mp,t}$ and its interaction with the utilization variable.

3.1 Learning by Doing

The learning-by-doing local-projections impulse responses are shown in Figure 6. These are estimates of (6), imposing $\beta^{LBN} = 0$. Panel 6a gives the response of labor productivity: aircraft per hours worked. Shaded areas in this and subsequent figures give 90% and 95% confidence bands.[16] Regressions include time and plant-by-model fixed effects and are in growth rates relative to productivity at time $t-1$. Hence they reflect the relative cumulative growth in labor productivity at each horizon in a production line receiving 1% higher demand, as predicted by the instrument described in the previous section. The specification controls for six lags of the explanatory variable (log aircraft produced), and dummy variables equaling one if the production line produced more than 25% or 50% of total aircraft of its broad type (e.g. bombers) in that month.[17] Labor productivity increases by around 0.4% per each percent increase in demand, within the first 12 months. Estimates become very noisy beyond the reported horizon.[18]

[16]The instrument is strong, by standard criteria, with an F-statistic of 25 in the 12-month horizon regression. We can reject a bias due to weak instruments greater than 10% according to a Montiel Olea & Pflueger's (2013) test. F-statistics for subsequent regressions are reported in the figure notes. An Anderson-Rubin test gives a p-statistic < 0.01 at the 12-month horizon in this and all subsequent LBD and LBN regressions.

[17]Plants that are dominant in a specific month are non-compilers in the first stage, because production in less significant remaining plants isn't very predictive of output in these dominant ones. These are uncommon occurrences: The monthly median observation produces 4% of its broad type of aircraft that month, and the 90$^{\text{th}}$ percentile observation produces 25%.

[18]Figure A.3 in the appendix shows the OLS version of the baseline IV regression. OLS estimates could be biased upwards or downwards, particularly when looking at the response to demand "shocks", i.e. controlling for past production. On one hand, the War Production Board may have directed demand to plants it expected to deliver aircraft

The production drive was associated with facility expansions; we control for this by calculating TFP as the residual from a constant-returns-to-scale Cobb-Douglass production function with a capital share of 20%, as in (7) with $\alpha = 0.80$ and $\gamma = 1$. The capital share was chosen to match the average ratio of capital costs to the sum of capital and labor costs in aircraft plants during the war.[19] All results are robust to using a capital share of $\frac{1}{3}$, as is common in the macro literature, or to simply controlling for the capital to labor ratio, as we will shortly see. Figure 6b shows a TFP response of similar magnitude to that of labor productivity.

Figure A.4 in the appendix shows the pre-trends of labor productivity and TFP before the shock to demand. There are signs of a slight pre-trend in labor productivity in the run up to the shock, but this is eliminated when considering TFP.

Figure 6c shows the response of production to the 1% increase in demand. The initial shock to aircraft demand leads to a persistent surge in production. The responses in Figure 6 should therefore be considered the response of labor productivity and TFP to an increase in demand with a half-life of slightly over a year.

Capital $K_{mp,t}$ measures *active* floor space. This already accounts to some extent for capital utilization, but TFP in Figure 6b also adjusts for capital utilization $U_{mp,t}$, measured as outlined in Section 1. The impulse response reflects an increase in TFP above and beyond cyclical increases in productivity arising from higher rates of utilization as in Basu *et al.* (2006). Figure A.5 in the appendix shows similar results without accounting for capital utilization.

Although structures reflected more than 60% aircraft plants' capital stock during the war, structures alone don't produce airplanes and Thompson (2001) has shown that capital deepening explains a large portion of shipyards' productivity growth during the war. Fortunately, the War Production Board recorded every investment in plant expansion exceeding $25,000 in this period, whether publicly or privately financed, and these investments are separated into "structure" and "equipment".[20] TFP's response to a demand shock is nearly identical when controlling for each plant's cumulative investment in equipment, as I report in Table A1 in the appendix. This is because investment in equipment is highly correlated with investment in structures (see Figure A.6

at higher productivity, which would lead to an upward bias in OLS estimates. However, it is clear from histories of the war production effort that the War Production Board was more concerned about a plant's ability to deliver a large quantity of aircraft than plants' cost/productivity. This objective, together with the War Manpower Commission's goal of directing demand to lower-pressure labor markets, may have in fact shifted demand to lower productivity plants, leading to a downward bias in OLS.

[19] Source: Aircraft firm balance sheets from Mergent Archives, for the sample of available firms (Curtiss, McDonnell, Nash, Northrop, and Republic). Labour costs are the sum of payroll and benefits. The cost of capital was calculated as depreciation plus the value of property, plant, and equipment times the interest rate. The government offered aircraft plants funding at 4% and this is taken as the interest rate, but doubling or tripling this interest rate to account for a risk premium changes the calculation very little, because depreciation was an order of magnitude larger. Hall (1990) and Basu & Fernald (1997) show that calibrating production function coefficients in this way is robust in the presence of markups.

[20] War Production Board, *War Manufacturing Facilities Authorized by State and County*, RG179, 221.1, Box 986, NARA, College Park.

in the appendix).[21]

Labor productivity and TFP are measured in physical units (TFPQ) so that responses reflect an increase in aircraft produced rather than changes in prices or markups. Model fixed effects reflect narrowly defined models, which controls for (major) product quality changes. Plant-by-model fixed effects also control for any (persistent) quality differences across plants producing the same model. Given the enormous increase in the size and quality of aircraft over the war, estimates shown here are likely lower bounds to quality-adjusted demand-induced productivity growth.[22]

Recent research has warned of potential bias in two-way fixed effects regressions, particularly if treatment effects are heterogeneous. An estimator of de Chaisemartin & D'Haultfoeuille (2020) corrects for this bias, but requires a set of groups whose treatment status remains constant throughout the sample. Instead, I apply a modified version of Goodman-Bacon's (2021) recommendation to compare production lines that were treated early with those that were never treated. When interacting the instrument with a dummy variable equalling one in first half of the sample, results are unchanged (albeit with a weaker instrument, see Figure A.7 in the appendix).

It is difficult to compare the results reported here to the existing literature for two reasons. First, the impulse responses shown here are dynamic, in contrast to the static responses shown in Thompson (2001), for example. Second, to allow a causal interpretation, responses here are to a *shock* to demand, rather than cumulative experience, as in the existing literature. Nevertheless, the 12-month response are of similar magnitude to the impact responses reported in Benkard (2000) and Thompson (2001), and to the naïve contemporaneous learning elasticity reported in Figure 4.

3.2 Learning by Necessity

Turning to learning by necessity, I estimate an unrestricted version of (6). Aircraft demand $Y_{mp,t}$ and its interaction with an indicator variable measuring whether a plant initially had high capital utilization are jointly instrumented by the "leave one out" instrument and its interaction with the indicator variable. Figure 7 plots the local projections impulse responses: the estimated β_h^{LBN} coefficients. This represents the response of productivity to a one percent increase in demand in plants with initially high capital utilization relative to those with lower utilization. High-pressure plants show larger increases in both labor productivity (top panel) and TFP (bottom panel). The magnitudes are substantial with both labor productivity and TFP growing by $\beta_{12}^{LBN} = 0.28$ percentage points more in plants that were initially more constrained at a 12-month horizon. This is on top of

[21] The figure also illustrates the importance of a physical measure of capital. Capital investments in structures correlate with floor space only with a 9-month lag, after controlling for 2-way fixed effects, so investment data may give incorrect measures of TFP, particularly at high frequency.

[22] Results might overstate productivity growth if demand pressures caused plants to cut corners and produce lower quality aircraft. However, I show in Appendix D that there are few indications of systemic demand-induced quality problems.

the $\beta_{12}^{LBD} = 0.23$ percent productivity growth seen in plants with lower utilization (Table A2 in the appendix), themselves operating at utilization rates well above the pre- and post-war norms.

Demand shocks are identified through the instrument, but capital utilization isn't randomly assigned. Plant by model fixed effects absorb productivity differences in plants with differing initial rates of capacity utilization. The remaining concern is that the interaction between capacity utilization and demand shocks is endogenous. Put simply, the concern is that learning by doing is stronger in high utilization plants because of a confounding factor that happens to be correlated with initial capacity utilization. Capacity utilization is endogenous, of course, and was indeed an important consideration in procurement decisions (Fairchild & Grossman 1959 chapter VI). It is reassuring that initially high- and low-utilization plants were similar on most dimensions (Table A3 in the appendix). However, one correlate does stand out: high utilization plants were older on average. This is because older plants were known entities at the onset of the war and they were the first to receive contracts before they'd had a chance to expand their capacity. However, Figure A.8 in the appendix shows that results are, if anything, stronger when controlling for plant age and its interaction with aircraft demand.

Investigating high pressure on labor, as opposed to capital, I use three metrics to evaluate labor shortages. The first is labor utilization, measured at the plant level as average hours per worker in a plant. The second is the wage rate in the plant's labor market excluding plants in the aviation industry, used to capture local labor market tightness. The third is the War Manpower Commission's classification of the tightest labor markets.[23] Table A4 in the appendix shows that these various metrics of capital and labor shortages are correlated but the correlations aren't perfect. Figure A.9 in the appendix shows similar learning by necessity estimates when considering labor rather than capital utilization.

3.3 Robustness

So far we have assumed that production exhibits constant returns to scale, with $\gamma = 1$ in (7). Increasing returns may certainly have played a role, although our data excludes overhead labor and floor space: the fixed costs that are a major source of increasing returns. Nonetheless, two diagnostics indicate that the rise in productivity goes beyond scale effects. Figures 8a and 8b repeat the learning by doing and learning by necessity regressions, but now with controls for the growth of each factor of production from horizon $t-1$ to $t+h$. The regression controls for the growth in (logs of) the capital to labor ratio, hours worked, floor space, and the capital utilization

[23]The War Manpower Commission classified each labor market in the US into four categories in each quarter, with 1 representing the tightest labor markets and 4 representing markets with labor surpluses (unemployment). Nearly half of the production lines in this study were in counties of the first category and an additional 30% were in the second. The dummy in question takes on a value of one of the plant was in a county classified in the first category.

ratio, as in (7). Results are similar to the baseline specification, shown in a dashed line in the figure. Controlling for each factor separately allows for greater flexibility in functional forms, but results are similar when controlling for a single variable measuring scale $S_{mp,t}$ as in (7) and defined in Section 2.1. The estimated coefficient on scale in this case indicates slightly increasing returns to scale, with a value that fluctuates around $\gamma = 1.1$, consistent with Basu *et al.* (2006).

Figures 8c and 8d take a different tack. I run multiple regressions, where the outcome variable TFP is a residual from the production function, using (7). In each regression, I impose a different value of γ, the parameter governing returns to scale. For a wide range of assumed returns-to-scale, we see productivity growth beyond what is explained by economies of scale. In fact, the estimated response of TFP to demand shocks increases the greater are the assumed economies of scale. This is because factors of production themselves *decline* following the demand shock, increasing the required growth in TFP needed to explain explain the increased production, if there are greater scale economies. (The responses of factors of production can be seen in Figure A.10 in the Appendix.) It is perhaps puzzling that plants decreased production inputs in face of high demand, but recall that all responses are relative to other plants. Responses merely suggest that plants receiving demand shocks expanded capacity at no greater pace than other plants (themselves scaling up as part of the nationwide wartime expansion).

Aircraft demand was persistent and productivity may have responded to cumulative, not only current, changes in demand, especially at longer horizons. Figure 9 addresses this issue in two ways. First, I estimate a multiplier-type impulse response, estimating a modified version of (6):

$$\Delta_h \log z_{mp,t+h} = \alpha_{mp} + \alpha_t + \beta_h^{LBD} \log \left(\sum_{\tau=t}^{t+h} Y_{mp,\tau} \right) + \beta_h^{LBN} \mathbb{1} \left(U_{p,0} > \bar{U}_0 \right) \log \left(\sum_{\tau=t}^{t+h} Y_{mp,\tau} \right) + \text{controls} + \varepsilon_{mp,t}^h.$$

The coefficient β^{LBD} now gives the change in productivity from time $t-1$ to $t+h$ resulting from a 1% increase in *cumulative* production over the same period. The results in the figure are from a two-stage-least squares regression, where the (log of) $I_{mp,t}$ now instruments for (log) cumulative production $\left(\sum_{\tau=t}^{t+h} Y_{mp,\tau}\right)$.[24] Similarly, β_h^{LBN} estimates how much larger this response is for plants with initially high capacity utilization. Figures 9a and 9b show multipliers of similar magnitudes to the previous specifications, because demand for aircraft in "treated" plants remained consistently 1% higher for the first 12 months than in the "control" group, with the gap narrowing afterwards. The log-log specification can be interpreted as the percent increase in productivity per one percent increase in accumulated experience, as in traditional learning-curve estimates.

[24] I follow Ramey & Zubairy (2018) and use the instrument at time t rather than the cumulative instrument. The cumulative production of broad aircraft types over longer horizons is more likely to be endogenous to productivity in plants producing those broad types than at higher frequency.

Figures 9c and 9d take a different approach: a difference-in-differences local projections regression proposed by Dube *et al.* (2023). This includes leads of the explanatory variable, in addition to lags, in a regression of the form:

$$\Delta_h \log z_{mp,t+h} = \alpha_{mp} + \alpha_t + \beta_h^{LBD} \log Y_{mp,t} + \beta_h^{LBN} \mathbb{1}\left(U_{p,0} > \bar{U}_0\right) \log Y_{mp,t} + \sum_{\ell=t-L,\ell\neq t}^{t+h} \gamma_\ell \log Y_{mp,\ell} + \text{controls} + \varepsilon_{mp,t}^h.$$

This specification includes separate controls for aircraft production in each period between the demand shock and the estimated productivity response. According to Dube *et al.* (2023), responses can now be interpreted as productivity growth following a 1% shock to demand, holding constant any future increases in demand following the initial shock. The figures show that results are attenuated but similar to the baseline specification.

Productivity spillovers across plants are plausible and these could bias estimates of the response of demand to productivity. I use the the national demand for a broad aircraft type (excluding the plant in question) as an instrument for demand for the aircraft in a particular plant. With productivity spillovers, the plant in question might benefit not only from demand directed to that plant, but also from demand-induced productivity in other plants of that same broad type. This could lead to an over-estimate of the effects of demand on productivity. The concern can be assuaged by controlling for the mediating factor of (average) productivity growth from month $t-1$ to $t+h$ in peer production lines. Given the instrument used, the most relevant peer group is other production lines producing the same broad aircraft types. Figure A.11 in the appendix shows that results are barely affected by this control. The figure also reports regressions that control for the average productivity growth in other production lines that relied on the same motor manufacturer, or the production volume therein. These control for potential productivity spillovers through supply chains, with little change in results. Importantly, these results in no way reject the possibility that there were productivity spillovers across plants. They merely suggest that spillovers were not induced by the identified demand shocks.

Additional robustness exercises are shown in Tables A1 and A2 in the appendix. Beyond the robustness checks already reported, we can see robust results when controlling for cumulative production; cumulative investment in equipment; weighting observations by the production line's cumulative wartime production to date, using a continuous measure of initial capital utilization; or a time varying measure of (lagged) capital utilization.

3.4 External Validity

Is learning by necessity a peculiarity of the Second World War production drive? Appendix D discusses the historical context and its external validity. Wartime price and wage controls suppressed inflationary pressures that might emerge in a peacetime setting. However, the aircraft industry was exempt from price controls and airplane prices *declined* dramatically, making a prize freeze unnecessary. Cost-plus-fixed-fee contracts provide weaker cost-cutting incentives than do fixed-fee contracts that are the default in modern procurement (McCall 1970, Bajari & Tadelis 2001; see Appendix D. With a fixed-fee contract, cost savings contribute to contractor profits, but these are passed through to the buyer under cost-plus-fixed-fee. The industry later faced caps on profit margins, further eroding incentives to cut costs. Now cost savings reduced profits, which were a fixed percentage of costs. However, firms still had an incentive to contain costs to expand quantity produced and to secure future procurement contracts.

Separately, government-induced demand may be different from other demand surges. Firms with market power may have weak incentives to reduce costs when facing high market demand because increased production partly cannibalizes existing profits. In contrast, the government, a monopsonistic buyer of military materiel, has greater power to dictate quantities produced and negotiate contracts that incentivize productivity growth.

Wage controls were frequently renegotiated and wages increased by 20% in aircraft-producing counties during the war. Table A4 in the appendix shows that wages were correlated with reported labor shortages, indicating that the price mechanism was still at play, at least to some extent.

Standardized products are arguably a necessary precondition for mass production. All plants in our dataset delivered standardized aircraft, but differed in the extent to which they adopted mass production techniques. Standardized production was new to this industry, but was commonplace throughout the 20$^{\text{th}}$ century, as in the pre- and post-war automobile industry, the post-war aircraft industry, and "just in time" manufacturing later in the century. Standardization was certainly catalyzed by the large wartime demand surge. However, the aircraft industry may have been at a developmental stage that made it poised for this transition and had the distant cousin of the automotive industry to learn from. It is difficult to assess whether the findings reported here are applicable to industries that are already applying production techniques on the knowledge frontier, are already producing standardized products, or have not yet matured to the point of standardization.

Although the wartime aviation industry may have been poised for a transition to mass production, there is no indication that learning curves were steeper in this setting. Estimates presented in this study are comparable those found in the peacetime aircraft industry (Benkard, 2000), wartime liberty ship building (Thompson, 2001), and truck manufacturing (Lafond *et al.*, 2022), although these all show static rather than dynamic estimates and employ different identifying strategies. These studies don't investigate "learning by necessity" and it is difficult to infer whether this phenomenon depends on the industry's developmental stage.

I was unable to locate data on aircraft faults, but the historical narrative gives little to suggest that there were systemic quality problems in airframe production or that aircraft manufacturers "cut corners" to achieve production targets. Military modification centers, serving as the final checkpoint for aircraft before deployment, were tasked to inspect and repair any faults in aircraft plants' deliveries. If manufacturers traded productivity for quality, these centers would have experienced increased workloads. However, as demonstrated in the appendix (Table A5), there was no correlation between modification center employment and productivity, suggesting that higher productivity didn't come at the expense of quality.

Patriotism may have motivated workers during the war and it is hard to evaluate whether "learning by necessity" requires levels of worker motivation above those typically seen in peacetime. It is difficult to adjudicate this question in our setting, but it is also easy to understate the extent to which more mundane considerations persisted in wartime. Wartime histories summarized in Appendix D show that worker absence, turnover, and strikes–all potentially inimical to productivity growth–were at historical highs around the time aircraft production peaked.

With these caveats in mind, we now inspect some concrete actions taken by airframe manufactures to investigate mechanisms through which productivity increased.

4 Mechanisms: What Plants Did to Increase Productivity

How, then, do capacity-constrained plants increase production in face of surging demand? A voluminous historical literature has studied the productivity "miracle" of the wartime production drive. Here, I focus here on three explanations widely acknowledged in wartime and historical analyses (War Production Board 1945, US Civilian Production Administration 1947a, Nelson 1950, Janeway 1951, Jones & Angly 1951, Herman 2012, Klein 2013). I focus on active decisions, more aligned with the concept of learning by necessity, than on passive learning. Appendix E gives more detailed historical case studies of these practices.

The first significant change was the move from "job shop" production methods to "line" production methods. Craven & Cate (1955) write that the "most conspicuous improvement [in the aircraft industry] was the switch from handwork methods to those of mass production" (p. 385). Mass production methods, long established in the automotive industry, were met with skepticism in the aircraft industry. Klein (2013) p. 71 claims that at the beginning of the war, "Nobody had yet found a way to bring mass-production techniques to airplane building, and prospects for doing so did not look promising". Nonetheless, the enormous demand pressures of the war induced technological adoption.

To evaluate this claim empirically, we assembled a new data set based on newspaper searches for terms related to production technique upgrades. Search terms included the the aircraft firm's name (with plant location verified in the body of the article) and terms indicating modern production technology (MASS and PRODUCTION appearing within 5 words from each other; ASSEMBLY and LINE within 5 words; PRODUCTION and LINE within 5 words; AUTOMOTIVE). A research assistant read each relevant article and a count variable was incremented by one at the earliest mention of a new production technique. For example, an October 1941 Business Week article identified through this procedure states that "The Glenn L. Martin Co. factories in Baltimore, MD have set up a mass-production technique new to aircraft manufacture — a belt-conveyor line... The line has already cut man-hours on these subassemblies in half... to speed bomber production." The "Mass Production" count variable is then increased by one for the Martin Baltimore plant in October 1941.

Sources included the digital archives of main national (business) publications (New York Times, Wall Street Journal, Business Week, Fortune). Local newspapers were searched through the archival platforms Chronicling America and Newspapers.com. Additionally, annual reports for aircraft companies were accessed via Mergent Archives.

By our count, nearly half the aircraft plants adopted new production techniques, with the average plant adopting three new methods (Figure A.12a in the appendix). The higher frequency methods used for the analysis of demand and productivity are less suited to analyze the evolution of methods, which changed at low frequency. Nevertheless, Figure 10 provides indicative evidence linking technology adoption to the volume of production and capacity constraints. It gives a scatter plot of the cumulative number of new production methods adopted in plant p up to month t against the cumulative production of aircraft model m in plant p up to month $t - 12$ (one year earlier). The scatter plot is residualized from time, plant, and aircraft model fixed effects. Notably, there is a statistically significant association between cumulative production ("learning" or "experience") and the subsequent adoption of mass-production methods, but only for plants with high capital utilization.

Outsourcing was a second factor discussed in contemporary reports and indeed the share of outsourced work grew from 10% to 40% of employment over course of the war (Figure A.12b in the appendix). Aircraft plants of the 1930s assembled the entire aircraft in house. However, with the introduction of mass production techniques featuring interchangeable parts produced at narrow tolerances, it became feasible to farm out parts of the production process to feeder plants. Formalizing this argument, Figure 11a shows how the share of outsourced production responded to increased demand in an estimation of (6), with outsourcing as the dependent variable. Plants with high utilization rates outsourced 20 percentage points of their workforce more than low utilization plants, in response to a 1% demand shock. The magnitude is notable considering that the average outsourcing rate was 30%. The effect appears cyclical and transient. Further, while outsourcing was used to increase production volumes, it isn't obvious that it increases productivity. The latter requires the subcontractor to be sufficiently productive and free the "mother plant" to produce the remaining components more efficiently.

Many studies claim that improved labor relations–the third factor I investigate–played a crucial role in driving labor productivity. Labor motivation problems are well documented. The median plant lost 7% of it workforce to absenteeism and 6% to quits in late 1943 (Figures A.12c and A.12d in the appendix, based on new archival data on labor conditions in plants[25]). Demand pressures appear to affect labor: In the half year following a 1% demand shock, plants with low labor utilization saw a 7 percentage point increase in absenteeism. However, Figure 11b shows that absenteeism increased by *less* in high hour-per-worker plants. It estimates (6), with the absence rate as the outcome variable and mean hours per worker over the course of the war as the utilization measure. This counter-intuitive finding–that labor problems increased less in high pressured plants–may suggest that management actions, taken in plants under under duress, were enough to offset these pressures. Appendix E documents specific measures taken by management to improve labor relations when faced with high demand and labor dis-satisfaction.

5 Conclusion

A traditional view of the transmission of government spending posits that increased demand boosts leads firms to soak up under-utilized employment or capital. The neoclassical view focuses increased labor supply. Both theories suggest that cyclical demand does little to expand output at high rates of utilization, nor can they affect productivity. This was also the common view at the onset of the Second World War, where economists warned that the economy could not sustain the planned war production drive, while the military insisted that it must. Using new archival data from this period, we see that plants with rates of capacity utilization met the production challenge through productivity gains. They did so not merely through passive learning, but through active investments in new production methods, improving working conditions, and experimenting with different supply chain management techniques.

The evidence in this paper is based on archival data on airframe production during the Second World War. It is possible that wage and price controls dampened inflationary pressures that might emerge in other settings, but aircraft prices *declined* dramatically during the war, indicating that productivity gains were more than sufficient to counteract inflationary pressures due to high demand. Demand pressures no doubt lead to inflation, but this study suggests a silver lining:

[25] Bureau of Labor Statistics, "Labor Statistics for the Aeronautical Industry," Reel 2237, PDF pp. 2210-2284; and Army Air Force Material Command, "Aircraft Program Progress Report," several volumes, Reel 2237, PDF pp. 2285-2648; both from the archives of the Air Force Historical Research Agency, Maxwell Air Force Base, AL.

Businesses may find ways to enhance productivity when facing exceptional demand. Of course, the findings are based on a particular industry and historical episode, and further research could beneficially examine other periods and industries at different developmental stages.

The case for restrained anti-trust policy in face of learning dynamics (Dasgupta & Stiglitz 1988, Benkard 2000) would appear even stronger with learning by necessity, with its non-linear relationship between demand and productivity. However, the war episode also demonstrates a lesser trade-off between efficiency and market concentration than often presumed. Smaller producers, not only market leaders, gained from robust demand conditions, which appear to have delayed the inevitable march towards market consolidation in this industry.

World wars will hopefully remain a rarity, but there may be lessons from wartime for the age of Covid-19 and wars in Eastern Europe and elsewhere. During the pandemic, some sectors showed substantial excess capacity and shortages were seen in others. Geopolitical risks and sanctions put additional supply constraints on firms worldwide. While such constraints have no doubt contributed to recent inflation, the findings in this paper suggest that private sector firms can at times find ingenious ways to overcome them.

References

ACEMOGLU, DARON, & LINN, JOSHUA. 2004. Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry. *The Quarterly Journal of Economics*, **119**(3), 1049–1090.

ALCHIAN, ARMEN. 1963. Reliability of Progress Curves in Airframe Production. *Econometrica*, **31**(4), 679–693.

ANTOLIN-DIAZ, JUAN, & SURICO, PAOLO. 2022. The Long-Run Effects of Government Spending. *CEPR Discussion Paper*, **17433**.

ANZOATEGUI, DIEGO, COMIN, DIEGO, GERTLER, MARK, & MARTINEZ, JOSEBA. 2019. Endogenous technology adoption and R&D as sources of business cycle persistence. *American Economic Journal: Macroeconomics*, **11**(3), 67–110.

ARTHUR, W. BRIAN. 1989. Competing Technologies, Increasing Returns, and Lock-In by Historical Events. *The Economic Journal*, **99**(394), 116–131.

ASHER, HAROLD. 1956. *Cost-Quantity Relationships in the Airframe Industry*. Santa Monica, CA: Rand Corporation.

ATKIN, DAVID, KHANDELWAL, AMIT K., & OSMAN, ADAM. 2017. Exporting and Firm Performance: Evidence from a Randomized Experiment. *The Quarterly Journal of Economics*, **132**(2), 551–615.

AUERBACH, ALAN J., & GORODNICHENKO, YURIY. 2012. Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy*, **4**(2), 1–27.

AUERBACH, ALAN J, & GORODNICHENKO, YURIY. 2013. Fiscal multipliers in recession and expansion. *Pages 63–98 of: Fiscal policy after the financial crisis*. University of Chicago Press.

AUERBACH, ALAN J, GORODNICHENKO, YURIY, & MURPHY, DANIEL. 2020. Local Fiscal Multipliers and Fiscal Spillovers in the USA. *IMF Economic Review*, **68**(1), 195–229.

BAJARI, PATRICK, & TADELIS, STEVEN. 2001. Incentives versus Transaction Costs: A Theory of Procurement Contracts. *The RAND Journal of Economics*, **32**(3), 387–407.

BARRO, ROBERT J. 1979. On the Determination of the Public Debt. *Journal of Political Economy*, **87**(5), 940–971.

BARRO, ROBERT J., & REDLICK, CHARLES J. 2011. Macroeconomic Effects From Government Purchases and Taxes. *The Quarterly Journal of Economics*, **126**(1), 51–102.

BASU, SUSANTO, & FERNALD, JOHN G. 1997. Returns to Scale in U.S. Production: Estimates and Implications. *Journal of Political Economy*, **105**(2), 249–283.

BASU, SUSANTO, FERNALD, JOHN G, & KIMBALL, MILES S. 2006. Are technology improvements contractionary? *American Economic Review*, **96**(5), 1418–1448.

BENIGNO, GIANLUCA, & FORNARO, LUCA. 2018. Stagnation traps. *The Review of Economic Studies*, **85**(3), 1425–1470.

BENKARD, C. LANIER. 2000. Learning and Forgetting: The Dynamics of Aircraft Production. *American Economic Review*, **90**(4), 1034–1054.

BOEHM, CHRISTOPH E., & PANDALAI-NAYAR, NITYA. 2022. Convex Supply Curves. *American Economic Review*, **112**(12), 3941–69.

BRUNET, GILLIAN. forthcoming. Stimulus on the Home Front: The State-Level Effects of WWII Spending. *Review of Economics and Statistics*.

BURNSIDE, CRAIG. 1996. Production function regressions, returns to scale, and externalities. *Journal of Monetary Economics*, **37**(2), 177 – 201.

BURNSIDE, CRAIG, & EICHENBAUM, MARTIN. 1996. Factor-Hoarding and the Propagation of Business-Cycle Shocks. *American Economic Review*, **86**(5), 1154–1174.

CARTER, SUSAN B. 2006. Labor force, employment, and unemployment: 1890-1990 [Weir], Table Ba470-477. *Pages 93–119 of:* CARTER, SUSAN B., GARTNER, SCOTT SIGMUND, HAINES, MICHAEL R., OLMSTEAD, ALAN L., SUTCH, RICHARD, & WRIGHT, GAVIN (eds), *Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition*, vol. 2. New York, NY: Cambridge University Press.

CHODOROW-REICH, GABRIEL. 2019. Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned? *American Economic Journal: Economic Policy*, **11**(2), 1–34.

CIVILIAN PRODUCTION ADMINISTRATION. 1945. *Alphabetical Listing of Major War Supply Contracts*.

CIVILIAN PRODUCTION ADMINISTRATION. 1947. *Official Munitions Production of the United States.*

CRAVEN, WESLEY F., & CATE, JAMES L. 1955. *Men and Planes*. The Army Air Froces in World War II, vol. 6. Chicago, IL: University of Chicago Press.

DASGUPTA, PARTHA, & STIGLITZ, JOSEPH. 1988. Learning-by-Doing, Market Structure and Industrial and Trade Policies. *Oxford Economic Papers*, **40**(2), 246–268.

DE CHAISEMARTIN, CLÉMENT, & D'HAULTFOEUILLE, XAVIER. 2020. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, **110**(9), 2964–96.

DE LOECKER, JAN. 2007. Do exports generate higher productivity? Evidence from Slovenia. *Journal of International Economics*, **73**(1), 69 – 98.

DE LOECKER, JAN. 2011. Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity. *Econometrica*, **79**(5), 1407–1451.

DUBE, ARINDRAJIT, GIRARDI, DANIELE, JORDÀ, ÒSCAR, & TAYLOR, ALAN M. 2023 (April). *A Local Projections Approach to Difference-in-Differences Event Studies*. Working Paper 31184. National Bureau of Economic Research.

FAIRCHILD, BYRON, & GROSSMAN, JONATHAN. 1959. *The Army and Industrial Manpower*. Washington DC: United States Army, Center of Military History.

FIELD, ALEXANDER J. 2002. *The Economic Consequences of U.S. Mobilization for the Second World War*. New Haven, CT: Yale University Press.

FIELD, ALEXANDER J. 2008. The Impact of the Second World War on US Productivity Growth. *The Economic History Review*, **61**(3), 672–694.

FIELD, ALEXANDER J. 2018. *World War II and the Growth of U.S. Potential Output*. Manuscript.

FINKELSTEIN, AMY. 2004. Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry. *The Quarterly Journal of Economics*, **119**(2), 527–564.

FISHBACK, PRICE, & CULLEN, JOSEPH A. 2013. Second World War spending and local economic activity in US counties, 1939-58 . *Economic History Review*, **66**(4), 975–992.

FISHBACK, PRICE V., & JAWORSKI, TAYLOR. 2016. World War II and US Economic Performance. *Pages 221–241 of: Studies in Economic History*. Studies in Economic History. Springer.

GOODMAN-BACON, ANDREW. 2021. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, **225**(2), 254–277.

GORDON, ROBERT J, & KRENN, ROBERT. 2010. The End of the Great Depression 1939-41: Policy Contributions and Fiscal Multipliers. *NBER Working Paper*, **16380**(September).

HALL, ROBERT. 1990. Invariance Properties of Solow's Productivity Residual. *Pages 63–98 of:* DIAMOND, PETER (ed), *Growth/Productivity/Unemployment: Essays to Celebrate Bob Solow's Birthday*. Cambridge, MA: MIT Press.

HANLON, WALKER, & JAWORSKI, TAYLOR. 2021. Spillover Effects of Intellectual Property Protection in the Interwar Aircraft Industry. *The Economic Journal*, 11. ueab091.

HERMAN, ARTHUR. 2012. *Freedom's Forge: How American Business Produced Victory in World War II*. New York, NY: Random House.

HICKMAN, BERT G. 1957. Capacity, Capacity Utilization, and the Acceleration Principle. *Chap. c5590 of: Problems of Capital Formation: Concepts, Measurement, and Controlling Factors*. NBER.

JANEWAY, ELIOT. 1951. *The Struggle for Survival: A Chronicle of Economic Mobilization In World War II*. New Haven, CT: Yale University Press.

JAWORSKI, TAYLOR. 2017. World War II and the Industrialization of the American South. *The Journal of Economic History*, **77**(4), 1048–1082.

JONES, JESSE H., & ANGLY, EDWARD. 1951. *Fifty Billion Dollars: My Thirteen Years with the RFC (1932-1945)*. New York, NY: Macmillan Company.

JORDÀ, ÒSCAR. 2005. Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, **95**(1), 161–182.

JORDÀ, ÒSCAR, SINGH, SANJAY R., & TAYLOR, ALAN M. 2020. The long-run effects of monetary policy. *CEPR Discussion Paper*, **14338**.

KLEIN, MAURY. 2013. *A Call to Arms: Mobilizing America for World War II*. Bloomsbury Publishing USA.

LAFFONT, JEAN-JACQUES, & TIROLE, JEAN. 1988. The dynamics of incentive contracts. *Econometrica: Journal of the Econometric Society*, 1153–1175.

LAFFONT, JEAN JACQUES, & TIROLE, JEAN. 1993. *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA: MIT Press.

LAFOND, FRANCIOS, GREENWALD, DIANA, & FRAMER, J. DOYNE. 2022. Can Stimulating Demand Drive Costs Down? World War II as a Natural Experiment. *The Journal of Economic History*, **82**(13).

LEVITT, STEVEN D., LIST, JOHN A., & SYVERSON, CHAD. 2013. Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant. *Journal of Political Economy*, **121**(4), 643–681.

LUCAS, ROBERT E. JR. 1988. On the mechanics of economic development. *Journal of Monetary Economics*, **22**(1), 3–42.

LUCAS, ROBERT E JR. 1993. Making a miracle. *Econometrica: Journal of the Econometric Society*, 251–272.

MCCALL, JOHN J. 1970. The simple economics of incentive contracting. *The American Economic Review*, **60**(5), 837–846.

MELITZ, MARC J, & REDDING, STEPHEN J. 2023. Trade and Innovation. *Chap. 6, pages 133–174 of: The Economics of Creative Destruction*. Harvard University Press.

MIDDLETON, KENNETH A. 1945. Wartime Productivity Changes in Airframe Industry. *Monthly Labor Review*, **61**, 215–225.

MISHINA, KAZUHIRO. 1999. Learning by New Experiences: Revisiting the Flying Fortress Learning Curve. *Pages 145 – 184 of: Learning by Doing in Markets, Firms, and Countries*. Cambridge, MA: National Bureau of Economic Research.

MONTIEL OLEA, JOSÉ LUIS, & PFLUEGER, CAROLIN. 2013. A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics*, **31**(3), 358–369.

MORAN, PATRICK, & QUERALTO, ALBERT. 2018. Innovation, productivity, and monetary policy. *Journal of Monetary Economics*, **93**, 24 – 41. Carnegie-Rochester-NYU Conference on Public Policy held at the Stern School of Business at New York University.

NAKAMURA, EMI, & STEINSSON, JÓN. 2014. Fiscal Stimulus in a Monetary Union: Evidence from US Regions. *American Economic Review*, **104**(3), 753–792.

NELSON, DONALD M. 1950. *Arsenal of Democracy: The Story of American War Production*. New York, N.Y.: Harcourt, Brace and Company.

NEWELL, RICHARD, JAFFE, ADAM, & STAVINS, ROBERT. 1999. The Induced Innovation Hypothesis and Energy-Saving Technological Change. *The Quarterly Journal of Economics*, **114**(3), 941–975.

OWYANG, MICHAEL T., RAMEY, VALERIE A., & ZUBAIRY, SARAH. 2013. Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data. *American Economic Review*, **103**(3), 129–34.

POPP, DAVID. 2002. Induced Innovation and Energy Prices. *American Economic Review*, **92**(1), 160–180.

RAMEY, V.A. 2016. Macroeconomic Shocks and Their Propagation. *Chap. 2, pages 71–162 of:* TAYLOR, J. B., & UHLIG, HARALD (eds), *Handbook of Macroeconomics*. Handbook of Macroeconomics, vol. 2. Elsevier.

RAMEY, VALERIE A. 2011a. Can Government Purchases Stimulate the Economy? *Journal of Economic Literature*, **49**(3), 673–85.

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

RAMEY, VALERIE A. 2019. Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research? *Journal of Economic Perspectives*, **33**(2), 89–114.

RAMEY, VALERIE A., & SHAPIRO, MATTHEW D. 1998. Costly capital reallocation and the effects of government spending. *Carnegie-Rochester Conference Series on Public Policy*, **48**(1), 145–194.

RAMEY, VALERIE A, & ZUBAIRY, SARAH. 2018. Government spending multipliers in good times and in bad: evidence from US historical data. *Journal of Political Economy*, **126**(2), 850–901.

RAPPING, LEONARD. 1965. Learning and World War II production functions. *The Review of Economics and Statistics*, 81–86.

REICHARDT, OTTO H. 1975. Industrial Concentration and World War II: A Note on the Aircraft Industry. *The Business History Review*, **49**(4), 498–503.

RHODE, PAUL. 2000. The Impact of World War Two Spending on the California Economy. *Pages 93–119 of:* LOTCHIN, R. (ed), *The Way We Really Were: The Golden State in the Second Great War*, vol. 2. Washington DC: University of Illinois Press.

ROCKOFF, HUGH. 2012. *America's Economic Way of War: War and the US Economy from the Spanish-American War to the Persian Gulf War*. Cambridge, UK: Cambridge University Press.

ROMER, PAUL M. 1986. Increasing Returns and Long-run Growth. *Journal of Political Economy*, **94**(5), 1002–1037.

ROMER, PAUL M. 1987. Crazy Explanations for the Productivity Slowdown. *Pages 163–210 of: NBER Macroeconomics Annual 1987, Volume 2*. NBER Chapters. National Bureau of Economic Research, Inc.

SEARLE, ALLAN D. 1945. Productivity changes in selected wartime shipbuilding programs. *Monthly Labor Review*, **61**, 1132–1147.

SMITH, R. ELBERTON. 1991. *The Army and Economic Mobilization*. Washington DC: United States Army, Center of Military History.

STOCK, ERIC J. 1999. Explaining the Differing U.S. and EU Positions on the Boeing/McDonnell-Douglas Merger: Avoiding Another Near-Miss. *University of Pennsylvania Journal of International Law*, **20**(4), 244â261.

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

THOMPSON, PETER. 2010. Learning by Doing. *Pages 429–476 of:* HALL, BRONWYN H., & ROSENBERG, NATHAN (eds), *Handbook of the Economics of Innovation*. Handbook of the Economics of Innovation, vol. 1. Elsevier.

THOMPSON, PETER. 2012. The Relationship between Unit Cost and Cumulative Quantity and the Evidence for Organizational Learning-by-Doing. *The Journal of Economic Perspectives*, **26**(3), 203–224.

US CIVILIAN PRODUCTION ADMINISTRATION. 1947a. *Industrial Mobilization for War: History of the War Production Board and Predecessor Agencies, 1940-1945*.

US Civilian Production Administration. 1947b. *Industrial Mobilization for War: History of the War Production Board and Predecessor Agencies, 1940-1945*. Vol. 2. Washington, DC: Unpublished manuscript, in RG179, PI-15, Boxes 1-4, National Archives, College Park.

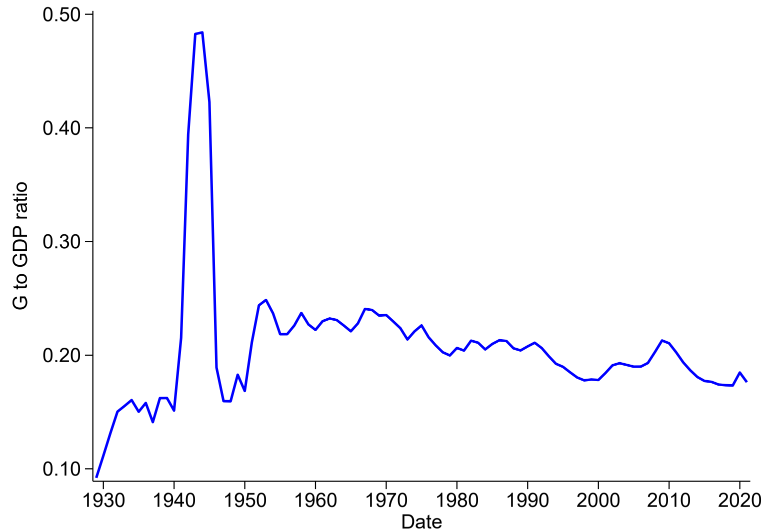
USAAF, United States Army Air Forces Air Materiel Command. 1952. *Source book of World War II basic data. Airframe industry*.

War Production Board. 1945. *War Production in 1944*. Tech. rept. War Production Board.

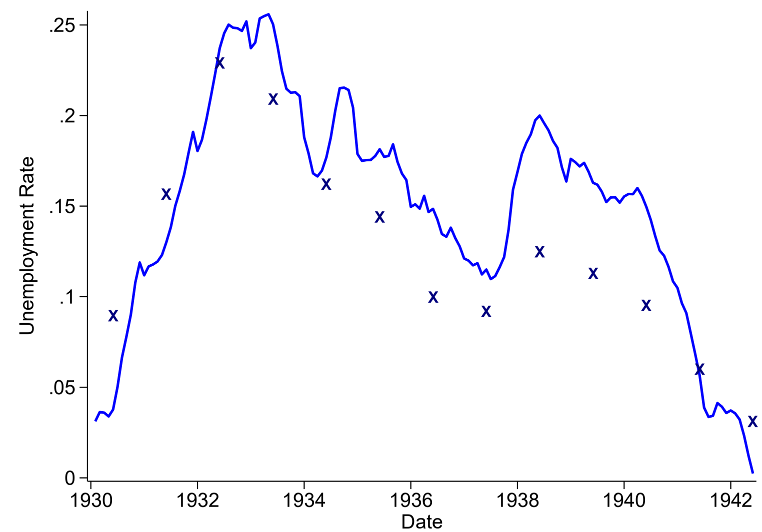
Wilson, Mark R. 2018. *Destructive Creation: American Business and the Winning of World War II*. Philadelphia, PA: University of Pennsylvania Press.

Wright, Theodore P. 1936. Factors affecting the cost of airplanes. *Journal of the aeronautical sciences*, **3**(4), 122–128.

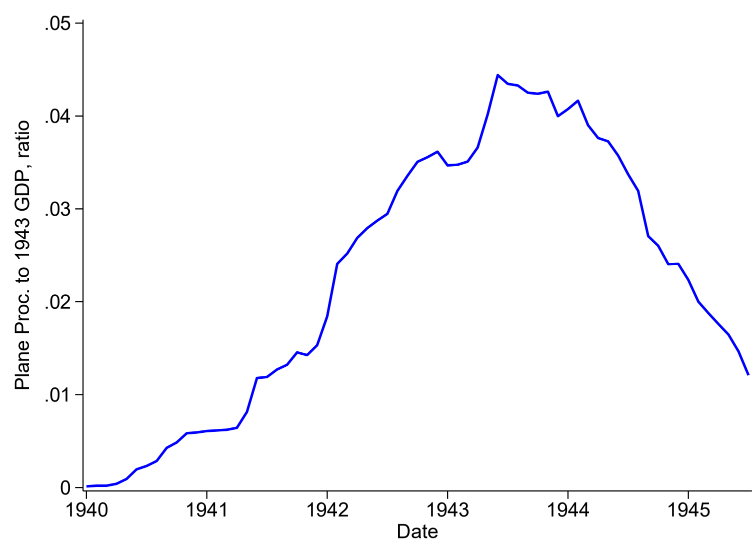
Figure 1: Government Spending, Unemployment, and Aircraft Procurement in the Second World War



(a) Public consumption and gross investment, share of GDP



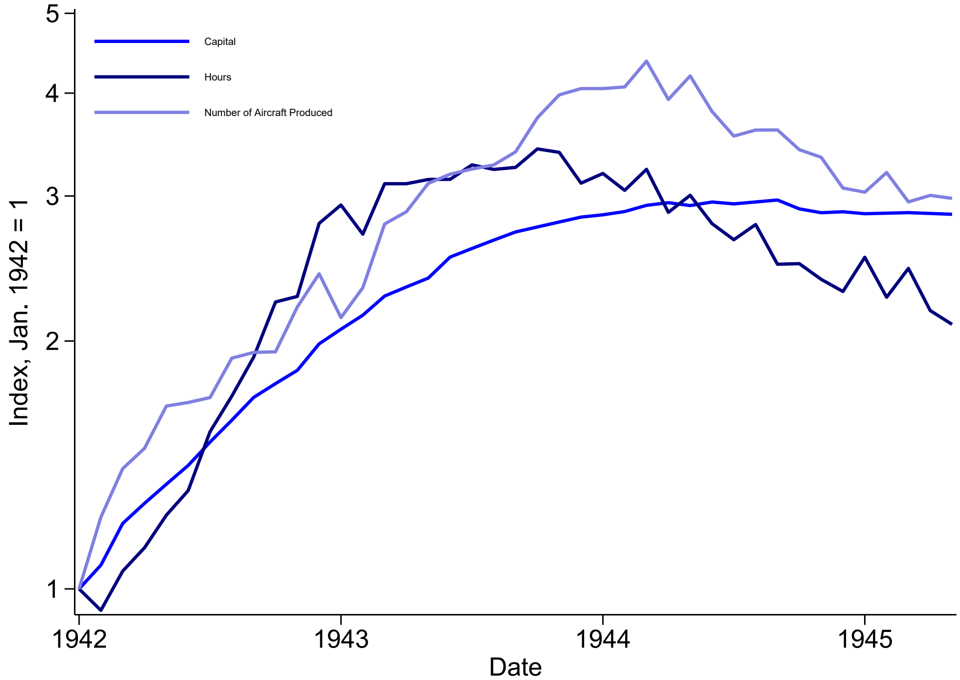
(b) Unemployment



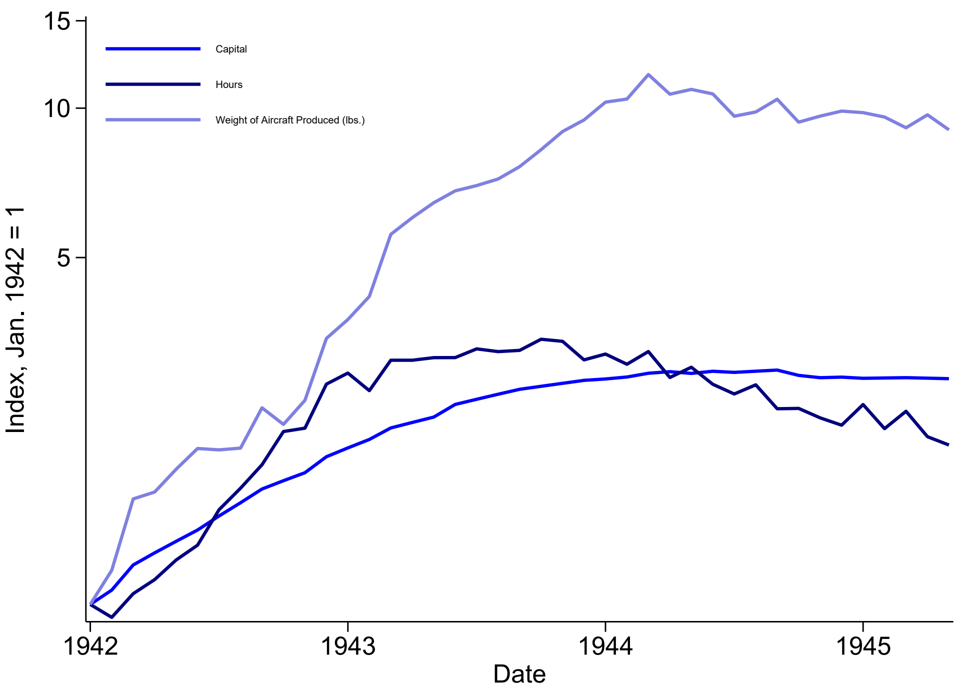
(c) Aircraft Procurement, share of 1943 GDP

Note: Panel (a) shows government consumption expenditure and gross investment as a share of GDP in the US since 1929. Source: Bureau of Economic Analysis, Fed FRED series GDPA and GCEA, retrieved from FRED, Federal Reserve Bank of St. Louis. Panel (b) shows the US unemployment rate from the great depression to the US's formal entry into the Second World War. Monthly series in line: National Bureau of Economic Research, Unemployment Rate for United States [M0892AUSM156SNBR], retrieved from FRED, Federal Reserve Bank of St. Louis. Annual series in Xs: Table Ba470-477 in *Historical Statistics of the United States*, Carter (2006). The latter is considered more authoritative but is available only at annual frequency. Panel (c) shows annualized government aircraft procurement as a share of 1943 GDP. Source: Civilian Production Administration (1945), Bureau of Economic Analysis, Fed FRED series GDPA, and the author.

Figure 2: Capital, Labor, and Output for the US Aircraft Industry in World War II



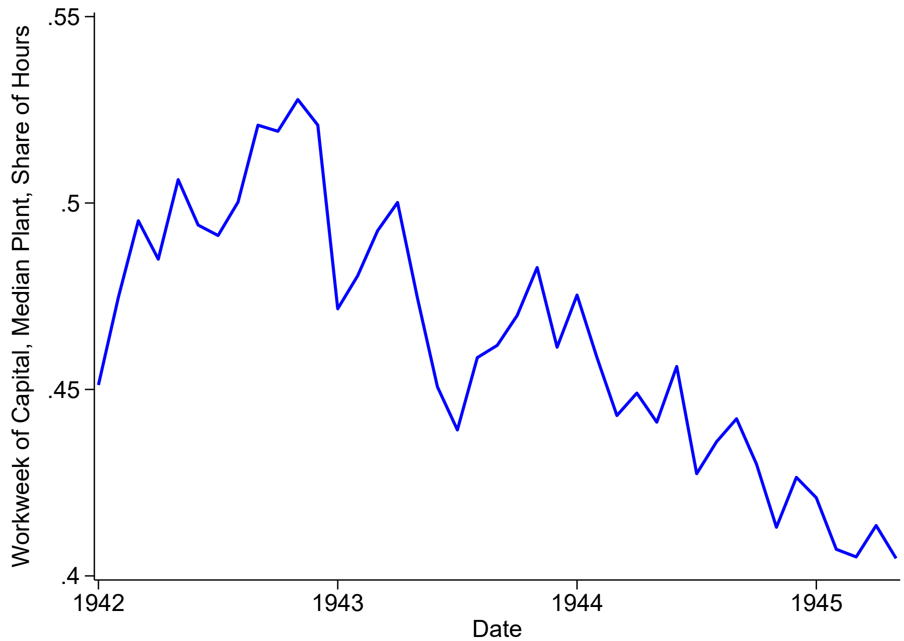
(a) Capital, Labor, and Number of Aircraft Produced, Indexes



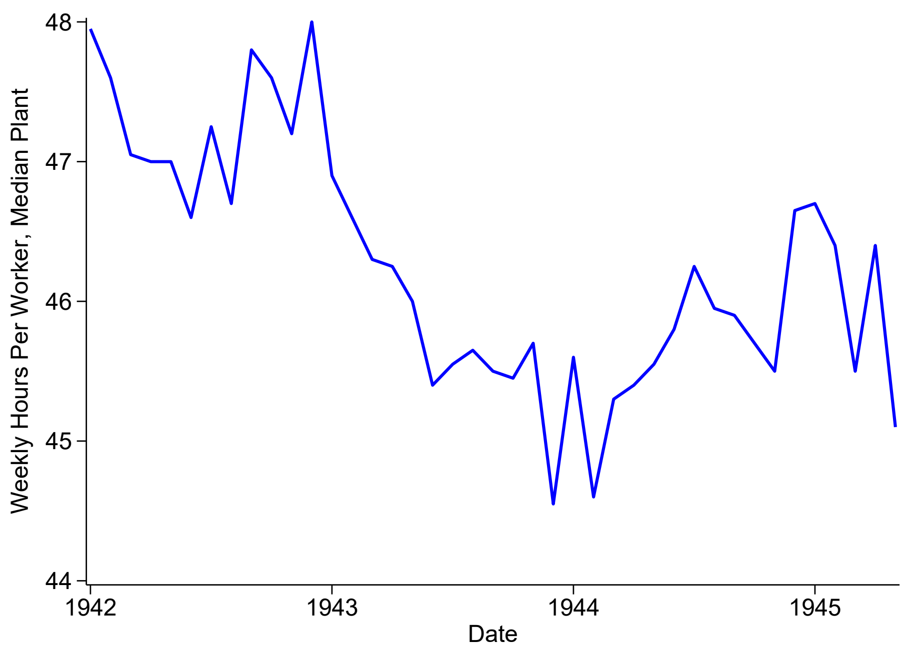
(b) Capital, Labor, and Weight of Aircraft Produced, Indexes

Note: The figures show aggregate inputs to and outputs of production in the airframe industry during World War II. Capital is the aggregate quantity of physical capital used in production, proxied by active floor space in airframe plants. Hours are aggregate hours of workers in direct aircraft manufacturing. Panel (a) measures output as number of aircraft. Panel (b) measures output as aggregate aircraft weight. Values of all variables are normalized to 1 in January 1942. Source: USAAF (1952) Vol. 1 Tables 2 and 3, Vol 2. Table 5, Civilian Production Administration (1947), Table 1, "Airplanes by Plant," pp. 32-55 and the author.

Figure 3: Capital and Labor Utilization in Airframe Plants



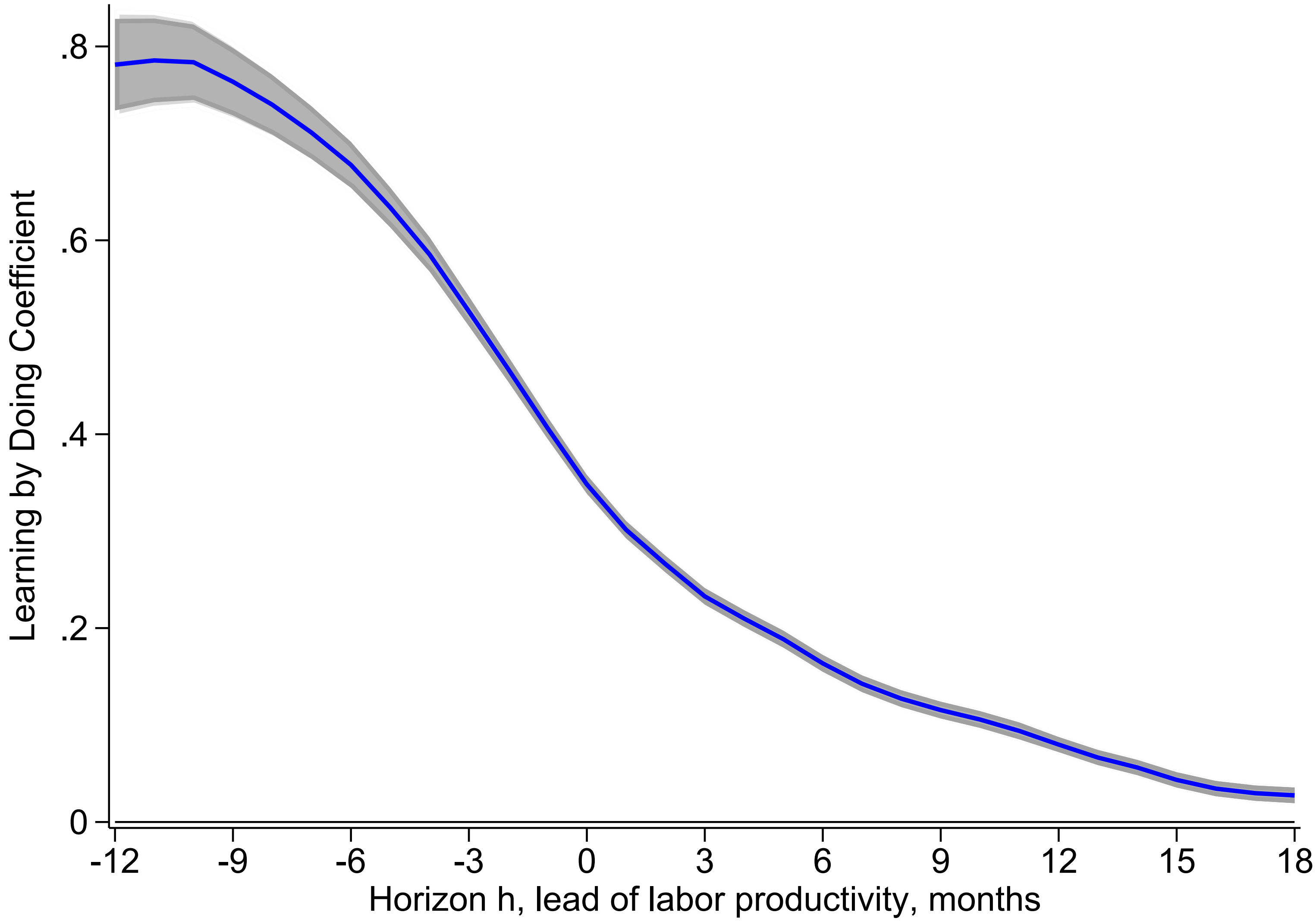
(a) Capital (Shift) Utilization



(b) Hours per Worker

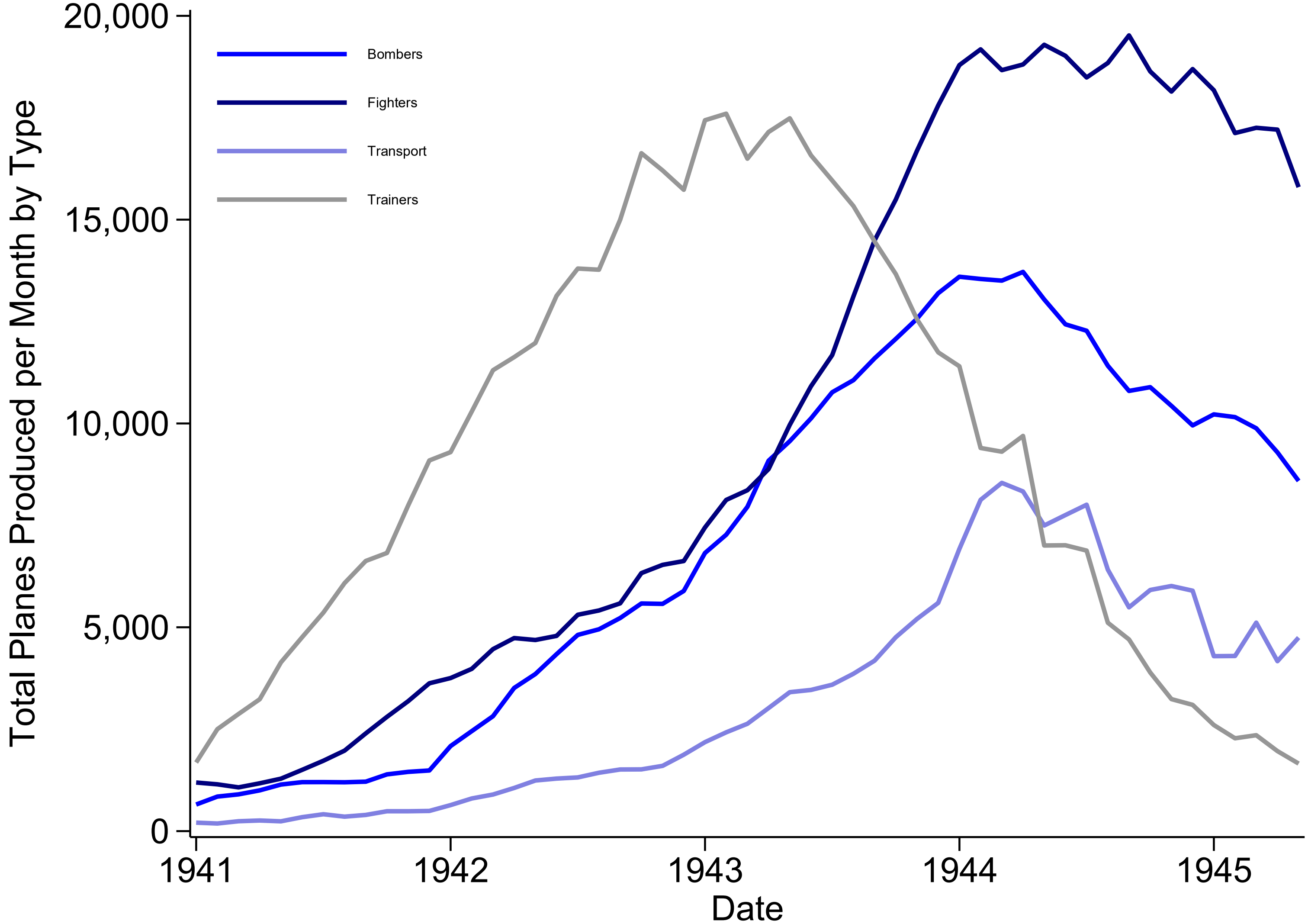
Note: Panel (a) shows shift utilization for the median airframe plant, estimated as described in Section 1, measured as share of hours out of 24×7. Panel (b) shows hours per worker in the median airframe plant. Source: USAAF (1952) Vol. 2, Table 6 and the author.

Figure 4: Dynamic Response and Pre-Trend of Labor Productivity in a Traditional Learning By Doing Regression



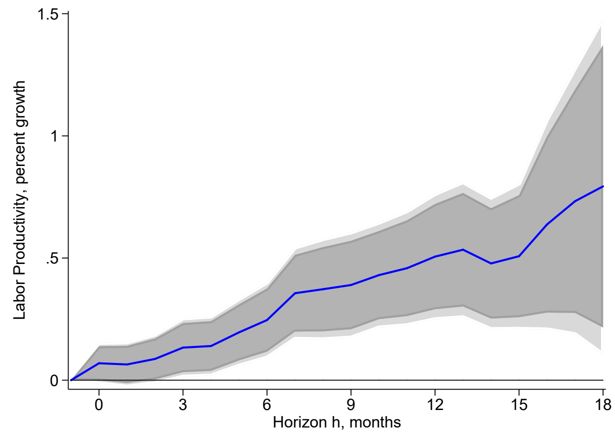
Note: The figure shows the OLS regression coefficient of (log) aircraft per hours worked at month $t+h$ per (log) cumulative aircraft produced at time $t=0$, with the x-axis giving the horizon h, in months. Shaded areas show 95% Newey-West confidence intervals. Traditional LBD regressions report the coefficient of $h=0$, showing a correlation between labor productivity and cumulative production. However, the responses show a substantial pre-trend, strongly indicating that production accumulated due to previous high productivity.

Figure 5: Total Aircraft Production by Broad Aircraft Type

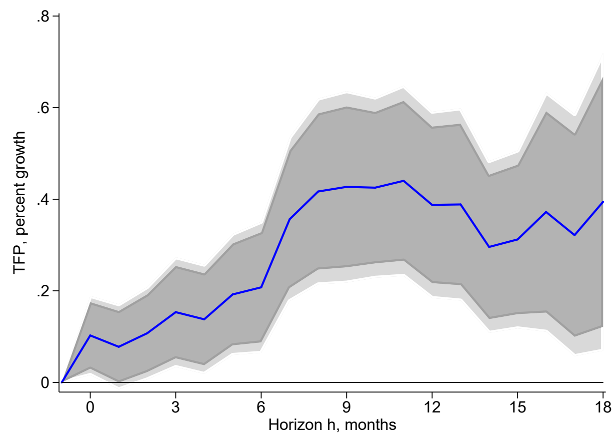


Note: The figure illustrates the instrument described in Section 2. It shows total monthly aircraft production of fighters, bombers, transport aircraft, and trainers. Differential demand for the aircraft types was driven by different strategic needs as the war progressed. The identifying assumption is that different production trajectories across plant types was driven by this differential demand, not other productivity drivers. Fighter aircraft were more prominent in lend-lease acquisitions by US allies in 1941, leading to a boom and bust in their production in 1941-42. Bombers were more central to the US war strategy and saw an inflection point after Pearl Harbor and again in 1943. Transport aircraft become increasingly important later in the war to supply troops when the US had "boots on the ground" in Europe and the Pacific. Trainers were obviously more important earlier in the war. Fighter aircraft saw a resurgence mid-war with increasing realization that bomber and transport aircraft benefited from having fighters as escorts. See historical narrative in Section 2. Source: USAAF (1952) Vol. 1 Table 3, Civilian Production Administration (1947) "Airplanes by Plant," pp. 32-55, and the author.

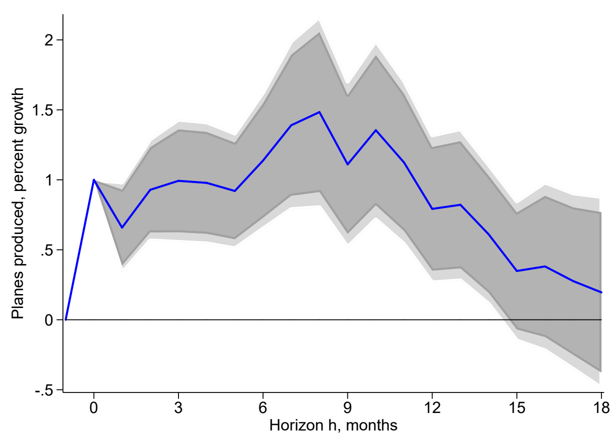
Figure 6: Responses to a 1% Shock to Aircraft Demand



(a) log Aircraft per Hour Worked



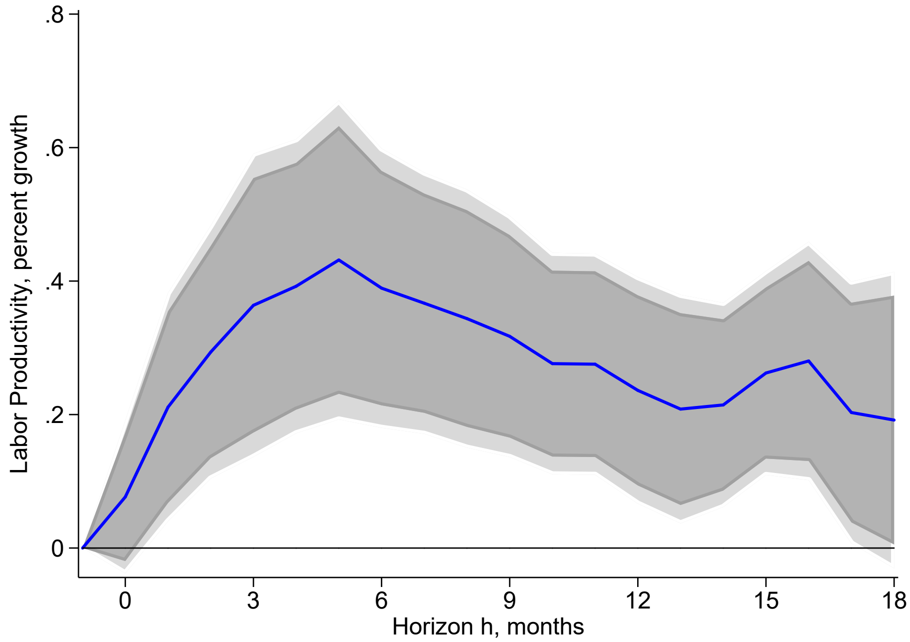
(b) TFP (capital utilization adjusted)



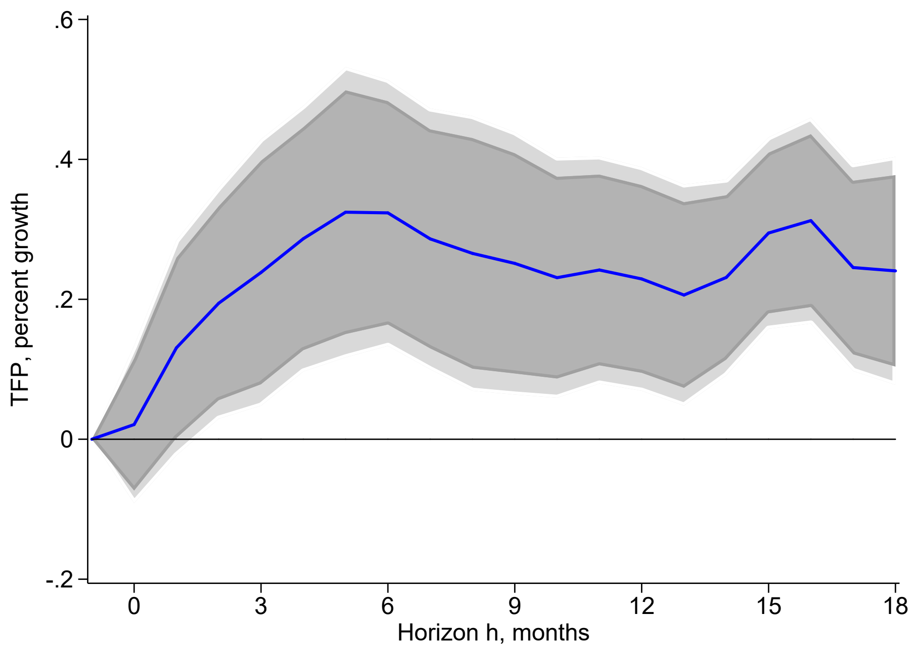
(c) log Aircraft Produced

Note: The figure shows the response of (a) log aircraft per hour worked, (b) TFP (adjusted for capital utilization), and (c) production, to a one percent shock to aircraft demand. Responses are the β_h^{LBD} coefficients of local projections estimates of (6), with $\beta_h^{LBN} = 0$ imposed. Aircraft demand is predicted by the instrument described in Section 2. Shaded areas show 95% Newey-West confidence intervals. First stage F-statistic at 12-month horizon = 24, 30, and 25 in the three panels.

Figure 7: Response of Output per Hour Worked and TFP to a 1% Shock to Aircraft Demand in High Capital Utilization Plants (relative to Low)



(a) Output per Hour Worked

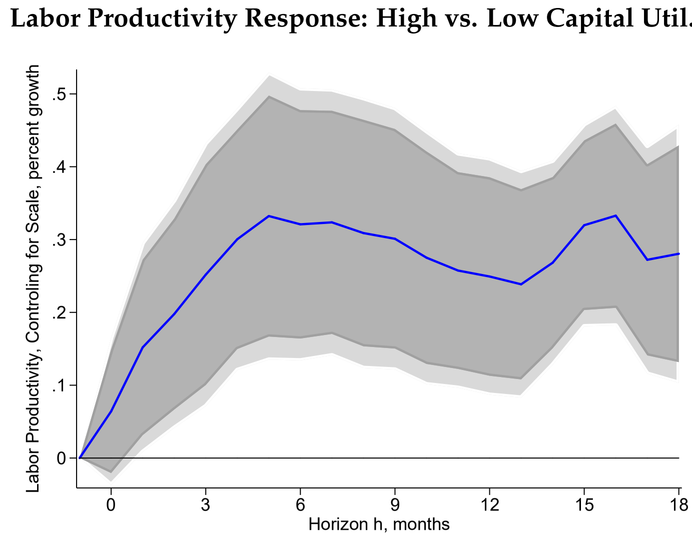


(b) TFP

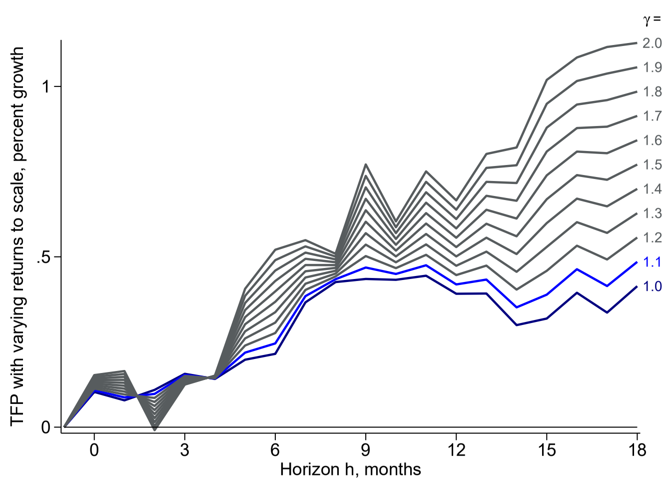
Note: The figure shows responses of (a) log aircraft per hour worked and (b) TFP (adjusted for capital utilization) to a one percent shock to aircraft demand in plants with above median initial capital utilization relative to those with below median utilization. Responses are the β_h^{LBN} coefficients of local projections estimates of (6). Aircraft demand and its interaction with initial capacity utilization are jointly predicted by the instrument described in Section 2 and its interaction with initial capacity utilization. Shaded areas show 90% and 95% Newey-West confidence intervals. First stage F-statistic at 12-month horizon = 14 and 15 in the top and bottom panels, respectively.

Figure 8: TFP or Economies of Scale?

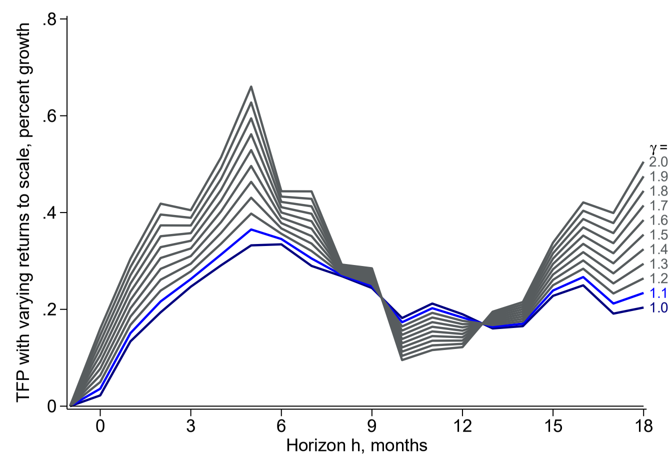
(a) Controlling for Growth of Factors of Production



(b) Controlling for Growth of Factors of Production



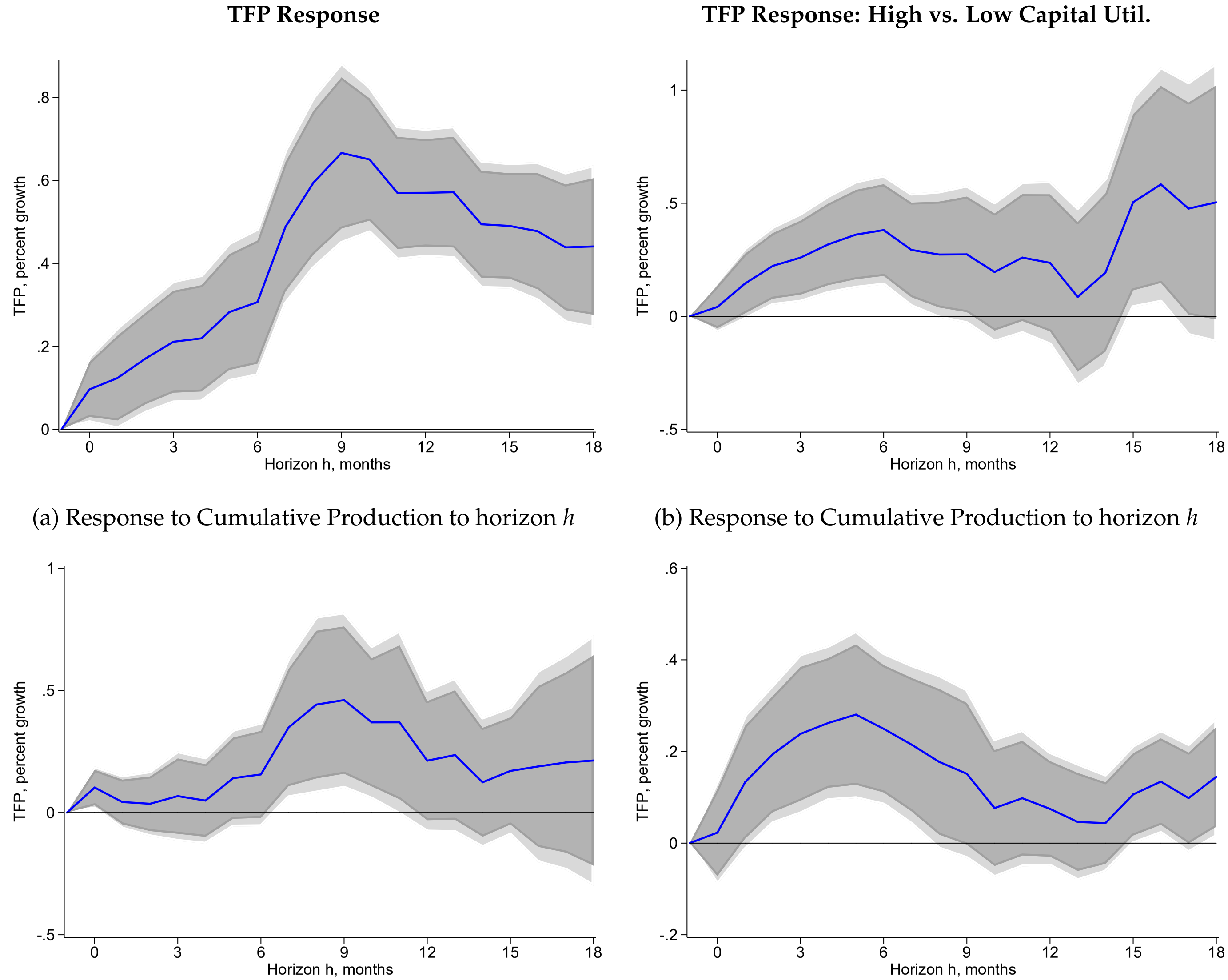
(c) Residualizing with a Range of Scale Parameters



(d) Residualizing with a Range of Scale Parameters

Note: The figure shows the response of productivity to a one percent shock to aircraft demand: local projections estimates of (6). Panels on the left-hand side show responses in the average plant: the β_h^{LBD} coefficients when $\beta_h^{LBN} = 0$ is imposed in (6). Panels on the right-hand side show responses in plants with above median initial capital utilization relative to those with below median utilization: β_h^{LBN} in an unrestricted version of (6). Aircraft demand and its interaction with initial capacity utilization are jointly predicted by the instrument described in Section 2, and its interaction with initial capacity utilization. In the top-row panels, the dependent variable is labor productivity and the regressions control for the growth (log-difference) in hours worked, productive floor space (proxying for capital), and capital utilization from month $t-1$ to month $t+h$, at each horizon h. The bottom row shows responses of TFP as measured by (7): labor productivity residualized for the scale of production. Each line in the bottom-row panels reflects a different value of the scale parameter γ. Shaded areas show 90% and 95% Newey-West confidence intervals. First stage F-statistic at 12-month horizon = 30 and 17 in panels (a) and (b), respectively.

Figure 9: Correcting for Induced Production and Demand



(a) Response to Cumulative Production to horizon h

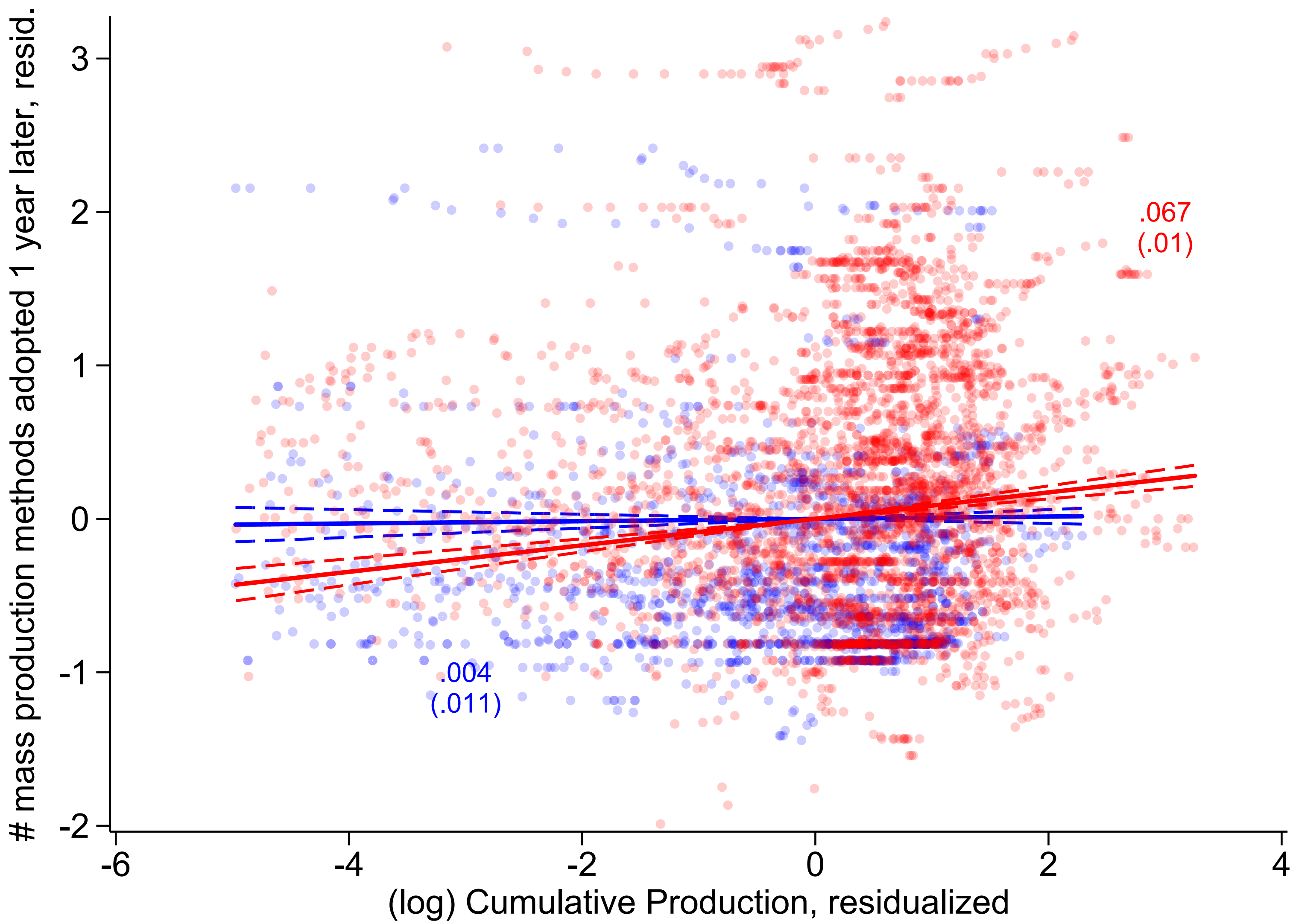
(b) Response to Cumulative Production to horizon h

(c) Controlling for Production in Months $t+1$ to $t+h$

(d) Controlling for Production in Months $t+1$ to $t+h$

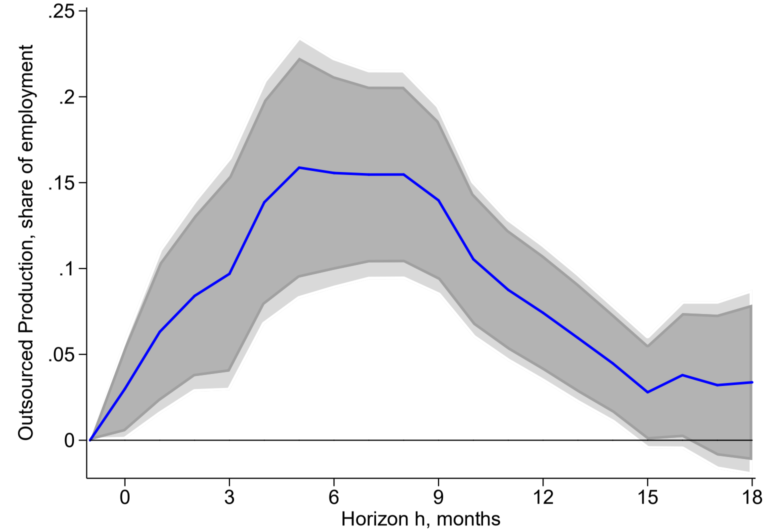
Note: The figure shows the response of TFP (adjusted for capital utilization) to a one percent shock to aircraft demand: local projections estimates of (6). Panels on the left-hand side show responses in the average plant: the β_h^{LBD} coefficients when $\beta_h^{LBN} = 0$ is imposed in (6). Panels on the right-hand side show responses in plants with above median initial capital utilization relative to those with below median utilization: β_h^{LBN} in an unrestricted version of (6). Aircraft demand and its interaction with initial capacity utilization are jointly predicted by the instrument described in Section 2, and its interaction with initial capacity utilization. In the top-row panels, the explanatory variable is the cumulative growth (log difference) of aircraft produced from month $t-1$ to month $t+h$. The estimates represent the percent growth in productivity at horizon h per one percent of cumulative growth in production up to that horizon, as predicted by the instrument at horizon zero. Responses in the bottom row follow Dube *et al.* (2023) and are the response of TFP to a once percent increase in demand at horizon zero, as predicted contemporaneously by the instrument, but include separate controls for production in each month t to $t+h$. The responses reflect the response of TFP to a one percent relative increase in demand at horizon zero, if that plant had experienced no further relative growth in total production to horizon h. Shaded areas show 90% and 95% Newey-West confidence intervals. First stage F-statistic at 12-month horizon = 130, 48, 28, and 14 in panels (a) to (d).

Figure 10: Adoption of Mass-Production Methods by Cumulative Production and Capital Utilization

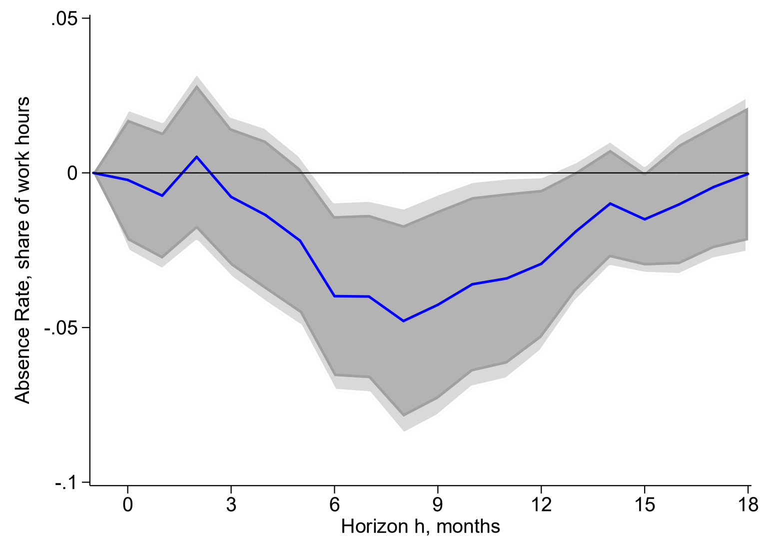


Note: The figure shows the number of mass production methods adopted (with a one-year lag) against the (log of) cumulative production in a production line. Both variables are residualized from monthly, plant, and aircraft model fixed effects. Red dots represent plants that had above median capital utilization at the beginning of the war and blue dots represent plants with below median capital utilization. Fitted regression lines for the two sub-samples are shown in solid lines, with 95% confidence intervals in dashed lines. Coefficients and standard errors for these regression lines are shown. There is a statistically significant association between cumulative production and subsequent adoption of mass-production methods for plants with high capital utilization, but no relationship for plants with low utilization. Source: USAAF (1952), Vol. 1, Table 3 and the author.

Figure 11: Responses to a 1% Shock to Aircraft Demand in High Utilization Plants (relative to Low)



(a) Outsourcing



(b) Absence Rate

Note:The figure shows responses of variables to a one percent shock to aircraft demand in plants with above median initial capital or labor utilization relative to those with below median utilization: the β_h^{LBN} coefficients in (6). Aircraft demand and its interaction with initial capacity utilization are jointly predicted by the instrument described in Section 2, and its interaction with initial capacity utilization. Panel (a) shows the response of the share of hours worked outsourced to feeder plants in high vs. low (initial) capital utilization plants. Panel (b) shows the response of the share of hours lost to absenteeism in high vs. low (initial) hours per worker plants. Shaded areas show 95% Newey-West confidence intervals. First stage F-statistic at 12-month horizon = 13 and 6, in the top and bottom panels, respectively.

Table 1: Summary Statistics

	Mean	Min	P10	P25	Median	P75	P90	Max	Coef. Var.
Panel A: Firm-level statistics									
# of plants	1.6	1	1	1	1	2	3	7	0.85
Models Produced	2.8	1	1	1	1	3	8	12	1.04
Total Sales (USD 1000)	713,575	4,710	14,823	26,896	153,371	1,193,764	2,221,488	3,675,244	1.43
Observations:	38								
Panel B: Plant-level statistics									
	Mean	Min	P10	P25	Median	P75	P90	Max	Coef. Var.
# of models	2.0	1	1	1	1	2	4	8	0.75
Peak production employment	10,170	373	621	1,599	6,977	15,182	24,034	48,128	1.03
Avg. Monthly Production (1000 Lbs.)	992.0	8.0	25.7	76.2	480	1,471	2,404	5,497	1.22
Cum. Investment ($1000)	19,328	276	276	1,447	12,141	31,151	48,658	94,898	1.10
Peak Floor Space (1000 sq. feet)	1,598	72	165	444	1,265	2,443	3,485	6,206	0.85
Observations:	61								
Panel C: Production-line-level statistics									
Peak Employment	7,465	55	481	1,465	4,556	9,818	16,021	125,360	1.63
Avg. Monthly Planes	61.0	0.50	2	11.3	36	83.8	160.6	339.1	1.10
Avg. Monthly Production (1,000s lbs.)	605.8	3.4	13.2	44.9	272.9	919.1	1,906	4,933	1.31
Observations:	141								

Note: Summary statistics of US World War II airframe industry. Sources: USAAF (1952) Vol. 1 Table 1 (production volume and weight) and Vol. 2 Tables 5 (floor space) and 6 (employment); "War Manufacturing Facilities Authorized by State and County," War Production Board Program and Statistics Bureau, June 15, 1945, RG 179, box 984, NARA College Park (investment); Civilian Production Administration (1945) (sales).