Government Purchases and Plant-Level Productivity: Evidence from World War II

Ethan Ilzetzki*

London School of Economics

PRELIMINARY AND INCOMPLETE

First draft: April, 2020

This draft: 23 December, 2020

Abstract

This paper studies the relationship between fiscal policy, productivity, and capacity utilization. It does so in the context of US World War II munition production, where many plants saw productivity growth in the face of substantial capacity constraints. Using archival data on the airframe industry, I show that increases in government purchases raise total factor productivity measured in quantity units (TFPQ) at the plant level. While low capacity utilization plants respond to new government purchases with relative increases in utilization, more constrained plants increase production through TFPQ growth. Increases in TFP are associated with more outsourcing of production. Shifts in military strategy provide an instrument for demand shifts across plants specializing in different aircraft types. The study uses detailed data on production, productivity, and capacity utilization collected by the US War Production Board and Army Air Force.

*I thank Mun Fai Chan, Hugo Reichardt, Laura Richardson, and Martin Souchier for outstanding research assistance; Tab Lewis and the rest of the National Archives staff in College Park, Debbie Seracini at the San Diego Air and Space Museum Archives, and Archie Difante and Tammy Horton at the Air Force Historical Research Agency for their help in locating archival resources. I also thank Jeremiah Dittmar, Josh Hausman, Xavier Jaravel, Johannes Wieland, Noam Yuchtman, and seminar participants at the LSE, Maryland, Toulouse, Ben Gurion University, and INSEAD for their useful comments. I acknowledge financial support from the Centre for Macroeconomics.

1 Introduction

A renaissance in empirical research on fiscal policy in the past decade has shed increasing light on the effects of government purchases on the economy.[1] In recent years, debates have also emerged as to whether the effects of fiscal policy differ depending on the degree of slack in the economy–an important consideration when using fiscal policy as a stabilization tool.[2] This study shows that the transmission of government purchases depends on the degree of capacity utilization at plant level. Using detailed archival plant-level data from the US munitions production drive during the Second World War, I show that airframe plants with low degrees of capacity utilization (more slack) respond to new purchases by relatively increasing capacity utilization. In contrast, plants facing capacity constraints see increases in total factor productivity when satisfying increased demand for their products. In addition to giving more granular evidence of the role of slack in the transmission of fiscal policy, I highlight a new transmission mechanism whereby demand can increase production efficiency, particularly when when producers face capacity constraints.

World War II has some unique advantages for studying this question. The Second World War brought the largest cyclical increase in public spending in American history. Federal public expenditure rose from 8% of GDP at the war's onset to 32% of GDP in 1945, declining again to 15% by 1948. (See Figure 1a.) By the time the US formally entered the war officially in late 1941 and the war production drive took off in earnest, the economy was essentially at full employment (Figure 1b).[3] However, given the large sectoral shift to wartime production, there were large differences in slack across plants and regions at the onset of the production drive. This variation helps shed light on the interaction between public demand and capacity utilization that is the theme of this investigation.

Studying the effects of fiscal policy and capacity utilization on productivity confronts us with substantial data challenges, because productivity and capacity utilization are both notoriously difficult to measure.[4] Given the urgent wartime production needs, the US War Production Board (WPB) and military procurement offices maintained exceptionally detailed production data. These include physical measures of output, direct manufacturing labor hours in production (as opposed to overhead or payroll), physical measures of capital, and direct measures of capital and labor uti-

[1] The literature is too large to cite in full, but Ramey (2011a, 2019) reviews the methodological progress and the evidence using aggregate and sub-national data. Chodorow-Reich (2019) narrows in on evidence based on cross-sectional (regional) and panel data.

[2] Using structural VAR and local projection methods, Auerbach & Gorodnichenko (2012, 2013) argue that multipliers are substantially larger in recessions. In contrast Owyang *et al.* (2013) and Ramey & Zubairy (2018) find little support for higher multipliers in the US in times of slack in the economy, using military spending shocks as the identifying variation.

[3] See Long (1952) for a discussion of the effects of the war on employment.

[4] See Syverson (2011) on productivity and Basu *et al.* (2006) and Fernald (2014) on capacity utilization and its relationship to productivity measurement.

lization. This study builds on a number of merged archival resources on aircraft production from the WPB and the Army Air Force (AAF), some of which haven't been used for empirical analysis since the war.

To study the effects of public spending on plant-level productivity, I utilize a difference in differences strategy and interact the share of national demand for aircraft types with each plantâs degree of specialization in that type of aircraft at the beginning of the war to create an instrument for aircraft demand. The military certainly directed procurement to plants it expected to be more productive.[5] However, the historical narrative outlined in Section 3 strongly suggests that procurement allocation across broad aircraft types (e.g. fighters vs. bombers) at the national level was driven by military strategy (demand) rather than production efficiency concerns. Given that plants' product mixes were stable over the war, this gives variation in procurement coming from strategic military needs rather than plants' productivity. I find that a quantity-based measure of Total Factor Productivity (TFPQ) increases by one third of a percent for each additional percent of aircraft demand in the average plant.

I then use a triple-differences framework to see how the transmission of government spending differs depending on plants' labor and capital constraints (measured at the beginning of the war). I measure capacity constraints using several separate, and imperfectly correlated, indicators: capital utilization based on detailed shift-utilization data, labor utilization (weekly hours per worker), counties that were classified as facing labor shortages according to the War Manpower Commission, and high-wage counties. The first two are direct measures of capital and labor utilization at the plant level, while the latter two reflect the general tightness of local labor markets. All measures tell a similar story: When a plant receives more orders, productivity grows by more in plants facing greater capacity constraints. In contrast, plants with more slack increase production through relative increases in capacity utilization. This indicates a capacity utilization response to demand in plants with production slack, but that more constrained plants met demand for their products through improved production efficiency. Capacity-constrained plants were on average older, but I show that the results are driven by heterogeneity in capacity constraints rather than in this confounding factor.

An earlier literature, discussed below, viewed productivity increases during World War II as reflecting learning by doing (LBD). My findings show that learning occurred primarily when plants faced tight capacity constraints, suggesting that production constraints may have stimulated learning. I will refer to this phenomenon as "learning by necessity", whereby plants increase productivity when facing limits to expansion due to capacity constraints.[6] It was well understood at the time

[5]See Scott-Kemmis & Bell's 2010 critique of endogeneity of procurement in LBD estimation.

[6]Thompson (2010) reserves the term "learning curve" to refer to cases where productivity growth can be linked to knowledge accumulated through experience. He also draws the distinction between "passive learning", with produc-

of the war that excessive capital utilization and long workweeks (labor utilization) were drags on productivity.[7] The War Department and other government agencies recommended a number of measures to resolve these difficulties. These included WMC efforts to relocate labor across regions and Defense Plant Corporation funded public investments. We will see little signs of relative increases in physical capital (in proportion to labor) in response to demand shocks in plants with high capital utilization. The War Department also recommended that constrained plants rely on outsourcing (cf. Fairchild & Grossman 1959). Indeed, we will see that more constrained plants subcontracted a large share of the workload arising from new orders, which may have contributed to their productivity rise. Other sources of productivity improvement are more difficult to measure. These include improved labor relations and spatial reconfiguration of plants (Fairchild & Grossman 1959 pp. 68, 137).[8]

Military spending has long been used to identify the effects of government spending on the economy.[9] Hall (2009) and Ramey (2011b) note that most of the variation in this instrument comes from World War II and the Korean War. A closer look at these large fiscal shocks is therefore important to our understanding of the transmission of fiscal policy. More importantly, this paper builds on the previous literature by using plant-level variation in demand and using shifting strategic military needs as a source of variation across plants over time. Unlike much of the extant literature, I don't focus on the aggregate effects of public expenditures (the "fiscal multiplier"), private consumption, or unemployment but rather on the effects of fiscal policy on productivity and relating the effects of public spending to plant-level capacity utilization or slack.

Research in the immediate post-war period documented learning by doing (LBD) in aircraft (Middleton 1945, Asher 1956, Alchian 1963, Rapping 1965) and shipbuilding (Searle 1945) indus-

tivity growing along an "experience curve" and broader and potentially more proactive causes for productivity growth as production accumulates, along a "progress curve". I will use the term LBD more loosely to refer to any productivity growth associated with accumulated production, regardless of the mechanism.

[7] A November 1943 War Manpower Commission report states that "the tendency for production to level off has been general throughout the war economy and has been attributed by the War Production Board to such factors as 'overall production fatigue' " (*Manpower Problems in the Airframe Industry*). Fairchild & Grossman (1959) reports in Chapter VII that the airframe industry was particularly affected by this problem. See also Yoshpe (1945) (pp. 37), who documents that excessive labor utilization in quartermaster supplying plants led to health and morale problems.

[8] Most of these factors reflect productivity improvements at business cycle frequency, the primary focus of this paper. Analysis is at the plant-by-aircraft model level, so that finding also don't speak to product innovations that may also have been spurred by demand. See Field (2018) in this regard.

[9] Barro (1979), Ramey & Shapiro (1998), Barro & Redlick (2011), Ramey (2011b), and others have used US military spending as instruments for US government consumption and investment shocks. Nakamura & Steinsson (2014) extend this analysis to a panel setting. Auerbach *et al.* (2019) studies the effects of modern military procurement on regional output, but doesn't focus on productivity. Brunet (2017) uses World War II procurement data to study the effects of government spending on output and employment across US states. Fishback & Cullen (2013) use regional data to study the longer term impact of World War II public spending on economic activity and find limited long-run impact. Rhode (2000) studies the effects of wartime spending on the California economy. Jaworski (2017) and Garin (2019) study the effects of wartime public investments on longer-term development. Hanlon & Jaworski (2019) study product improvements in the interwar US aircraft industry, but their focus is on the role of patent protections on innovation. There is of course an extensive literature on other economic implications of the war, including on gender (Goldin & Olivetti 2013), management (Giorcelli & Bianchi 2020), and R&D (Gross & Sampat 2020).

tries. These studies show a positive correlation between the cumulative output of a plant and its output per worker. More recent work suggests that these learning effects are far smaller once one controls for capital, suggesting a far smaller effect of experience on TFP (Thompson 2001). Much of the LBD literature is formulated through the lens of a production learning curve, first observed by Wright (1936), who noted that aircraft manufacturers became more productive with experience. The observation of a learning curve got ingrained in the post-war conventional wisdom and was one of the motivating facts of the endogenous growth literature several decades later (Lucas 1993). Scale effects in production and learning by doing imply that fiscal policy or other sources of demand could increase firms' productivity either cyclically or persistently.[10] The possibility that demand may affect productivity has been a topic of theoretical interest in a more recent literature (e.g. Benigno & Fornaro 2018, Moran & Queralto 2018, Anzoategui *et al.* 2019, Jordà *et al.* 2020).

However, learning curve estimates in the early literature are based on correlations, with the obvious problem that plants with greater productivity growth will have accumulated larger volumes of output over time.[11] As I show in in Section 3, reverse causation isn't merely hypothetical, but can be shown to be an important driver of the correlation between productivity and experience in the data. In contrast, I use shifting strategic demand for aircraft types as a source of variation in production that isn't driven by a specific plant's productivity. Further, I control not only for a physical measure of capital but also for capacity utilization. Most importantly, I investigate heterogeneity in the effects of experience on productivity, shedding some light on the potential mechanisms underlying learning by doing, with a focus on the role of capacity utilization.

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 1994, Basu *et al.* 2006). This paper shows that capacity utilization does indeed react as expected to demand shocks. However TPFQ grows even controlling for this increase in utilization, bringing real productivity gains, not merely reflecting mis-measurement. Additionally, I find that capacity utilization responds more strongly to public demand when plants have spare capacity. But further, plants with high rates of utilization see relatively higher productivity growth when faced with rising de-

[10]See Romer (1994) for a review of the early literature on endogenous growth, including the work of Young (1991), who ties this directly to learning by doing. Jones & Manuelli (2005) summarize a theoretical literature on the effects of fiscal policy on endogenous growth, but this literature typically focuses on the the effects of government size in explaining long-run growth differences across countries. An earlier literature hypothesized that demand could induce innovation: see Romer (1987) for a review; more recent estimates of induced innovation in the context of energy efficiency can be found in Newell *et al.* (1999) and Popp (2002). Hickman (1957) was an early contribution to this literature and linked capacity utilization to incentives for capital investment, dubbed "the acceleration princple". Estimates of the importance of market size on production and productivity can be found in the trade and innovation literatures (Acemoglu & Linn 2004; Finkelstein 2004 De Loecker 2007, 2011; Atkin *et al.* 2017), but these focus on the long-run, as opposed to business cycle frequency and don't speak to importance of capacity utilization. An earlier literature in macro also attempted to estimate returns to scale in aggregate production functions (Hall 1990, Burnside 1996, Basu & Fernald 1997).

[11]See Levitt *et al.* (2013) for a more recent contribution that uses an instrumental variables approach.

mand.

The remainder of the paper is organized as follows. Section 2 describes the data and the historical and institutional setting. Section 3 lays out the empirical strategy with the main results shown in Section 4. Section 5 concludes.

2 Data, Institutional Setting, and Historical Context

World War II led to the largest cyclical increase in public consumption in US history. Figure 1a shows government consumption as a percent of GDP in the US from 1930 to today. There has been a secular increase in public spending as a share of national income since the Great Depression, but the Second World War stands out as the single largest shock to government purchases. The analysis that follows focuses on aircraft purchases, certainly narrowing the analysis to a single sector. However, aircraft was the single largest expenditure item in the military budget and became the largest industry during the war. Figure 1c shows that aircraft procurement rose from one to three percent of pre-war GDP, a share of GDP that is comparable to the entire increase in military spending during the Vietnam War. In May 1940, after the fall of France, President Roosevelt set an ambitious objective of producing 50,000 planes. At the time this was viewed as a nearly impossible task, with Simon Kuznetz estimating that the US didn't have the productive capacity to meet this aim (Wilson 2018 pp. 178.) In fact, the US aircraft industry produced twice this number of aircraft in 1944 alone (War Production Board 1945 pp. 10, Smith 1991). Procurement of aircraft (and other war materiel) increased during 1940-41, but only really took off following the attack on Pearl Harbor in December 1941 and it peaked in 1943. The aircraft industry was a young industry: the median firm was founded in 1928 and the median plant was founded in 1939.

Procurement was under the purview of the relevant military branches, in this case the Army Air Force (AAF) and the Navy. During the war, procurement was also coordinated with the War Production Board (WPB) that provided a strategy for the war production effort. The AAF divided the country into six regional procurement districts, with district commands managing procurement in each region. However, because of the importance and ambition of the aircraft production schedule, procurement of airframe, motors, and propellers was separated from the general Army Supply Program and was managed by a joint agency, the Aircraft Resources Control Office, at Dayton Ohio. This agency dealt directly with the industry and the War Production Board. The AAF base at Wright Field (later Wright-Paterson) monitored aircraft production and aircraft modification to meet the AAF's strategic needs. The majority of contracts were Cost Plus Fixed Fee (CPFF), whereby the suppliers' (audited) costs were reimbursed and augmented with a pre-negotiated lump-sum payment. However, because of concerns of war profiteering, many contracts were renegotiated ex-post and aircraft manufacturers' profitability was lower than it had been before or after the war (Wilson 2018, chapter 4, gives the history of contract re-negotiations). Prior to the war, most aircraft were made to order based on detailed specifications of the procuring agency. This became untenable in an age of mass production and with the need to rapidly out-produce military opponents. Learning from the automotive industry, the WPB demanded that aircraft manufacturers shift to producing standardized aircraft models. These were then modified in army or navy modification centers to the exact specifications of the procuring agency. This aides productivity analysis as one can be more confident that an aircraft of a specific model and mark coming off of a specific production line had the same specifications. The following section outlines in detail how procurement was allocated across plants, which is central to identifying the effects of aircraft demand on plant productivity.

The analysis in this paper draws on a number of archival sources, primarily from the archives of the WPB and the Air Materiel Command of the AAF. While several of the sources have been used in previous research, I have digitized new materials and matched several data sources. Some data, including capacity utilization measures, have not been used in previous research. Here, I briefly outline the data sources for the variables used in the analysis that follows.

The main data are from the Aeronautical Monthly Progress Reports (AMPR), collected by the AAF headquarters at Wright Field. The WPB and AAF carefully monitored the production of war materiel. All aircraft manufacturers were required to provide monthly reports on their production progress. These were collected in monthly issues of the AMPR and were collated in two volumes entitled *Source Book of World War II Basic Data: Airframe Industry*. The data are for assembly of complete aircraft (pre-modification). The data were used to monitor production against plans, to ensure that manufacturers were working as close as possible to full capacity, and to monitor costs for CPFF contracts. Reporting requirements and methodology were uniform across plants and extremely detailed.[12] The WPB and the War Manpower Commission (WMC) were concerned about acute labor and capital shortages and airframe plants were frequently audited.[13]

The AMPR includes monthly plant by aircraft model data for all wartime aircraft manufacturers. 61 plants produced 83 different aircraft models leading to 204 plant-by-model pairs. To ease exposition, I will slightly abuse terminology in what follows and refer to a plant-by-model combination as a "production line", although some plants ran several production lines for the same model. Aircraft models are very narrowly defined in the data and all design changes are noted. I

[12] The AAF provided plants with standardized forms and a 150 page document giving minute detail on how to report production, productivity, capacity utilization, and other data in a very uniform format. The document, ATSC Regulation No. 15-36-3, can be found in the AFHRA archives, Reel A2050, starting on slide 850. See also SDASM archives Box 34 to see how a specific manufacturer (Consolidated Vultee) adopted these procedures internally.

[13] District procurement offices were assigned to monitor these reports and production and were provided with formulae to detect mis-reporting. Wilson (2018) documents (pp. 176) that as many as 60 military and GAO auditors could be on site to monitor production at a single airframe plant. See "AMPR Questionnaire for use in Making In-Plant Audits of Basic Labor Statistics" (AFHRA archives, Reel A2050, starting on slide 1128) and "Basic Labor Statistics–How to Maintain Them", ibid, starting on slide 1179

dropped a small number of plants and production lines that produced fewer than 100 aircraft or operated for less than 6 months, as they don't provide sufficient production line level variation for the subsequent analysis.[14]

The point of departure for productivity measurement is the variable "Unit Man Hours: Entire Plane"Ṗlants were required to report the number of direct worker-hours that entered into the production of the last plane delivered in each calendar month. This includes only workers directly involved in manufacturing; overhead was separately reported as "indirect workers". This is a far more direct measure of labor productivity than commonly available in modern data as it is in hours per physical units and at the product level (thus addressing the multi-product plant problem). While there are clear advantages to measuring productivity at the aircraft level, the last aircraft may be unrepresentative of the plant's average productivity. For sake of comparison, I calculated monthly labor productivity by dividing the number of aircraft delivered by the number of payroll hours for manufacturing workers. This is the standard methodology to calculate labor productivity in physical units. The two measures show very similar time series patterns. However, comparing the two measures highlights the advantage of direct aircraft-level productivity measurement. The typical aircraft took more than a single month to build and the AMPR required the aircraft-level measure to incorporate hours in all production months.[15] In contrast, dividing the number of aircraft by current hours worked creates a mismatch between delivery time and production time and particularly misstates productivity at the beginning or end of a production batch. (The shows many workers producing little output and the latter the opposite.)

Converting output per hour worked to TFP requires a measure of the capital stock used for production. The AMPR provides a direct physical proxy for capital. It gives quarterly observations of the floor space utilized in production at each plant. The measure includes only floor space actively used for production, excluding office space and other facilities, and including any yard space used for production. Structures are a substantial portion of the value of plants' capital stock, but floor space is also a proxy for the amount of equipment used in production. The capital stock is relatively slow-moving in the data and I interpolate quarterly floor space to give a monthly measure of physical capital per plant. Unlike most estimates of the capital stock deriving from investment data and requiring approximations of the initial capital stock, this variable gives the capital stock in physical units that is actively used in a given month.

A quantity-based measure of the capital stock has distinct advantages. Structures were the largest component (60%) of capital investment in the airframe industry during the war. Expenditure on structures confounds variation in land prices and construction costs across regions (ex-

[14]The AMPR begins reporting in 1941, but has only 60% coverage prior to 1943. Coverage is 100% starting in January 1943, which was also the initial production date for a large share of production lines.

[15]Documents from Convair, the largest wartime producer, show that bombers required 45 to 90 days to build, depending on the model (SDASM archives, Box 17).

penditure on structures) with capital accumulation (physical structures). On the other hand, floor space will tend to understate differences in the quality of equipment across plants. This is partially addressed by including time and plant fixed effects, as long as there are no differential trends in capital quality across plants. Given that the dollar value of investment does incorporate some information about the quality of capital used in production, I use investment at the plant level as a control. Investment data was obtained by matching the AMPR with the WPB's *Listing of War Manufacturing Facilities Authorized 1940-45*, which lists every capital investment undertaken in war manufacturing plants exceeding $25,000, whether publicly or privately financed. The document separates investments in structure and equipment and allows a control for time-by-plant improvements in the capital stock and capital vintage, measured by the dollar value of investments in equipment and the time that has elapsed since the most recent investment.

Floor space is given at the plant- rather than production-line level. The AMPR gives the number of worker-hours devoted to each aircraft model in a plant in each month. Using product-level labor inputs, I allocate capital across production lines to equate the capital to labor ratio across all products within a plant. With standard production functions, production efficiency would call for equating the capital-to-labor ratio across production lines. In the baseline specification, TFP is measured by assuming a capital share in production of $\frac{1}{3}$. That is, I subtract one third of the logarithm of the capital-to-labor ratio from log labor productivity. Results are robust to estimating the capital share directly.

While floor space already partly controls for capacity utilization (plants reported actual floor space used rather than floor space available), the data allow an even more careful measurement of capital and labor utilization. The AMPR gives monthly data on the number of workers and hours worked in each work shift. Plants were required to report the number of work shifts per day, the number of daily hours in each shift, and the number of monthly worker-hours active in each one of the shifts each month. This allows a measurement of shift utilization, which was used to assess capital utilization during the war and suggested by Basu *et al.* (2006) as a capital utilization measure. I measure a plant's capacity using the actual number of hours worked in the first shift (always the most active shift) as indicative of full production potential in that month. Full capacity is then measured as the number of monthly worker hours that would result if the plant were active 24 hours a day at full production potential (with the same number of workers per hour as the first shift). Capital utilization is the ratio between actual monthly work hours and full capacity.[16] The AMPR also includes monthly reports of average weekly hours per worker, which I use as a measure of labor utilization.

[16] Wartime reports confirm that the use of second shifts, night shifts, and Saturday shifts were a major source of variation in capacity utilization both over time and across plants. Of course, there is also variation over time within plant in the number of hours employed in the first shift. However, this will already be captured in the capital to labor ratio.

These two variables measure capital (shift) and labor utilization at the plant level. Given the importance of capacity constraints in the thesis forwarded in this paper, I constructed two additional external measures of labor markets tightness to reflect external constraints to labor supply and therefore factor availability. These are both taken from quarterly county-by-industry data on munitions industries included in WPB reports, found in the National Archives. We matched each plant to the average wage in its county of operation to give the local wage, whose variation across counties gives an indication of local labor market tightness. In the baseline specification, we use wages from the non-aircraft industry, to avoid confounding wages with labor productivity itself (although the aircraft industry was often large enough to have a general equilibrium impact on local wages). The WMC classified each county into four categories based on their assessment of the acuteness of labor shortages. These are reported in the OMPI and included as an additional measure of labor market tightness. These variables are all obvious endogenous: I discuss in Section 4 how these are used in a triple-differences specification and potential confounding factors.

The number of monthly aircraft delivered by plant and model is given in the AMPR. The same information is available in Civilian Production Administration's *Official Munitions Production* (OMP). This post-war document recorded all major munitions procured by all military branches during the war at monthly frequency. It gives the number of aircraft "acceptances" received by the military by model from each aircraft plant. This document is slightly more comprehensive than the AMPR, with the latter reaching 100% coverage only in January 1943. I use this source to fill in observations missing from the AMPR and to cross-check the AMPR's data. For those months and production lines where both sources report aircraft deliveries, the two sources correspond closely. Baseline results are given using the AMPR as the primary source for monthly production and the OMP as a secondary source, whenever AMPR data are missing. This approach maximizes coverage, but results are robust to using either of the individual sources.

3 Empirical Strategy

Estimating scale effects in production, learning by doing, and the effects of public spending on plant productivity pose an empirical challenge. Productivity is one reason why certain plants gain larger scale, accumulate more experience, and attract more procurement contracts, so that simple correlations between productivity and scale aren't necessarily informative of demand's causal impact. The post-war LBD literature reported correlations between cumulative output and output per worker as reflecting a "learning curve". Researchers implicitly presumed that wartime procurement reflected a demand shifter that traced the supply curve or production function and thus informed the researcher about the nature of the production process. Of course, procurement wasn't randomly allocated and the government likely purchased more aircraft from those plants it believed could deliver, i.e. ramp up production and productivity, in relatively short order.

Reverse causation is not only possible, but very likely. Certainly experience may cause a plant to increase its productivity over time, but there is also mechanical relationship between output per hour worked and output itself. In a pure accounting sense, there are only two ways a plant can accumulate more output and therefore more experience: greater labor productivity or more labor inputs. Labor productivity has a direct impact on output, but plants are also more likely to increase employment and labor utilization during productive months. Both channels reflect an effect of productivity on output, and ultimately experience. Second, both output and productivity are strongly auto-correlated in the data. This means that both accumulated experience and current productivity will be correlated with past productivity, potentially creating an incorrect impression that experience enhanced productivity.

Reverse causation isn't merely a theoretical possibility. Figure 2 illustrates the empirical challenge in practice. It shows (the natural logarithms of) output and (labor) productivity in four production lines. The left hand panels (a and c) show C-54 aircraft production in two Douglas Aircraft plants; the right hand panels (b and d) show production of two different aircraft models in Convair's San Diego plant. The correlation between production and output per hour worked is nearly perfect. This correlation would survive including time and plant fixed effects as it has a strong time-by-plant component. Labor productivity's growth over time is certainly suggestive of a fixed cost or a learning curve. But the high frequency overlap between productivity and output suggests opposite direction of causation is also likely and that the mere correlation between the two variables overstates the causal effect of experience on productivity.

This can be ascertained more clearly by looking at known productivity shocks. When a firm begins production of an existing aircraft model in a new plant, production and productivity tends to drop in existing plant. Figure 2 shows one such example. In early 1944, Douglas began production on C-54 aircraft in its new Chicago plant (panel c). Productivity declined in the Santa Monica plant (panel a) in early 1944 exactly as production began in Chicago, seen in the lower panel. Similarly, starting a new production line within the same plant slows down production of existing products in the same plant. In mid-1944 Convair began producing the PB4Y (panel d): a variant of the B-24 bomber adapted for the navy. This led to a distinct drop in productivity and output in the existing B-24 production lines (panel b). These correlations clearly reflect productivity shocks that affect output. OLS regressions incorrectly attribute this entire correlation to a deceleration in productivity due to a lower rate of experience accumulation.

Productivity and output are both strongly autocorrelated in the data and this further exacerbates the identification problem. The left-hand panel of Figure A.1 in the appendix shows a scatter plot the (log) cumulative output of aircraft production lines from the beginning to the end of the war in Europe (VE day, May 1945) and (log) output in the single month of May 1945. This correlation indicates that it is very difficult to disentangle scale effects (current production) from learning or experience effects (cumulative output). The right hand panel of the figure now shows (log) cumulative output against the same plant's labor productivity 16 months earlier in early 1944 (a date chosen to maximize observations). The strong correlation between past productivity and cumulative output at the end of the war suggests that past productivity is a confounding factor in OLS LBD regressions. Past productivity causes experience and predicts future productivity.

Table 1 shows OLS estimates of learning and scale effects and illustrates these difficulties in our sample. The simple correlation between (log) cumulative production and (log) aircraft per hours worked is positive and statistically significant (column 1). The following columns introduce combinations of time, plant, aircraft model, and plant-by-model fixed effects, with similar results Strong autocorrelation in production at the plant-by-model level results in similar correlations when using (log) current output as the right hand side variable (column 6) Both current and cumulative output are correlated with productivity when the two are simultaneously included as dependent variables in column 7. This is a common specification in the LBD literature, which uses this specification in attempt to separate LBD from static scale effects. However, including merely a single lag of output per hour worked, cumulative output is no longer correlated with current productivity (column 8). This isn't meant to imply that experience doesn't affect productivity, but rather that it is nearly impossible to disentangle scale effects from learning when production is autocorrelated. In the estimates that follow, I use current demand rather than cumulative output as the explanatory variable and include 6 monthly lags of the explanatory variable. Results are mechanically almost identical when using experience as the explanatory variable: With sufficient number of lags, current output is nearly equal to the residual variation in cumulative output.

To address causal inference, I use an instrument that exploits the fact that plants tend to specialize in specific aircraft types and that demand for broad aircraft types (e.g. bombers vs. fighter planes) was determined primarily due to 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), where procurement was affected both plants' relative expected productive capacity and quality.

I divide aircraft into five broad types: bombers, communications, fighters, trainers, and transport and include a sixth category for other specialized aircraft. Let $S_{p\tau,t-1}$ denote aircraft type τ's share in plant p's *cumulative* production up to month $t-1$.[17] $D_{p\tau t}$ gives total aircraft of type τ delivered in month t from all plants excluding plant p (a "leave one out" specification). The instru-

[17]This gives a dynamic pattern of specialization. Results are nearly indistinguishable when using static shares at the beginning of the war. However, this latter instrument gives smaller F-statistics in the first stage. This is largely because initial shares are often unrepresentative of what happened throughout the war. Results are also similar when using static shares based on cumulative shares of production at war-end.

ment is $I_{mpt} = S_{p\tau,t-1} \otimes D_{p\tau t}$ that interacts the two variables. It gives predicted aircraft demand from plant p based in month t based on its product mix previously in the war. Instrument relevance requires that aircraft manufacturers specializing in particular aircraft receive more orders as overall procurement of that aircraft type increases. This is borne out in F statistics reported in the IV regressions in the following section. The exclusion restriction requires that the differential aggregate procurement trajectories across these broad aircraft types isn't driven by relative expected productivity growth of broad aircraft types.

The first stage of the 2 stage least squares specification is given by

$$D_{mpt} = \gamma I_{mp} + \text{controls} + \text{FE} + \text{lags} + u_{mpt}, \tag{1}$$

where D_{mpt} are (log) aircraft of model m delivered by plant p in month t, I_{mp} is the instrument and the remaining terms include controls, lagged variables, and fixed effects.

Impulse responses of the second stage of the regression are estimated using local projections (Jordà 2005). At each horizon h, the response of productivity $y_{mp,t+h}$ to aircraft demand D_{mpt} is estimated as $\hat{\beta}_h$, arising from the regression

$$y_{mp,t+h} = \beta_h \widehat{D}^h_{mpt} + \sum_{i=1}^{I} \left(\delta^y_i y_{mp,t-i} + \delta^D_i D_{mpt-i} \right) + \alpha_t + \alpha_{mp} + \text{controls} + \varepsilon_{mpt}, \tag{2}$$

where $\hat{D}_{mp,t}$ is predicted aircraft demand from (1) . γ_t and γ_{mp} are time and plant-by-model (production line) fixed effects, respectively. Including these fixed effects give the regression a dynamic difference-in-differences interpretation, whereby we are comparing the differential productivity growth over time across production lines. The items in parenthesis are lagged variables. (2) is estimated in a separate regression at each horizon h, to give a local projections impulse response.[18]

What source of variation does the instrument capture? This is illustrated in Figure 3, which shows the number of aircrafts (of all types) delivered per production line in three categories of plants. These are plants with at least 75% of their production concentrated in bombers, fighters, or transport in January 1943. The threshold is merely for graphical purposes and the instrument itself has more categories and allows for continuous specialization intensities.[19] The figure shows that plants in the three categories saw very demand fluctuations, which have clear historical interpretations.

Early war production was for lend-lease assistance to US allies in Europe. In terms of aviation,

[18]The specification includes both fixed effects and lagged dependent variable. Nickell (1981) warns that this specification may result in biased estimates. Specifications that exclude either fixed effects or lags of the dependent variable lead to weak instruments. However, reduced-form regressions in those two specifications give impulse responses with similar shapes and statistical significance as the ones reported.

[19]The figure looks similar when setting a 50% or 90% threshold.

this primarily came in the form of fighter aircrafts, leading to a boom in fighter procurement in 1940-1941. Fighters were also used as escorts for merchant ships during this period. US direct involvement in the war began in December 1941. US military strategy in the immediate aftermath of the attack on Pearl Harbor anticipated a heavy reliance on aerial bombing (as exemplified by the Battle of Midway in summer 1942), leading to an inflection in the demand for bomber aircrafts in 1942 and surge in demand in 1943. Demand for transport aircrafts took off later in the ground operations phase of the war: transport aircraft supported the island-hopping operations in the Pacific and facilitated the invasion of Italy in 1943.[20] Demand for fighter aircrafts surged again in 1943 as it became apparent that both bomber and transport aircraft benefited from fighter escorts.[21]

The instrument projects this relative time variation across aircraft types differentially across plants based on their mix of aircraft types earlier in the war. The identifying assumption is that shifts in relative procurement across major aircraft types over time shown here were driven by strategic considerations. A threat to identification would arise if these changes in procurement were due to differential expected productivity trends across plants with different specialization.[22]

This assumption has support in the historical literature, where strategic considerations were paramount in determining procurement schedules for broad categories of munitions (see US Civilian Production Administration 1947, Smith 1991, Klein 2013, Holley 1964 for historical overviews of military production schedules and procurement strategies). In September 1943, a report by the WMC on *Manpower Problems in the Airframe Industry* notes that

> 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.

Towards the end of the war, a WPB report looks back and summarizes:

> 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.

[20]See AFHRA Reel 1009, pp. 1608 "Airborne Missions in the Mediterranean" on the use of C-47 transport aircraft for glider and paratrooper landings in operation Husky, Landbroke, and Fustan in Sicily. On the importance of 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.

[21]Major Lesher, Lee A. (1988). "The Evolution of the Long-Range Escort Doctrine in World War II" United States Air Command and Staff College

[22]Given the controls for lagged variables, this would require the AAF to predict differential trends better than an $AR(6)$ process.

The same report 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, the historical narrative supports the notion that procurement of broad categories of aircraft was driven by strategic needs, not productivity of specific aircraft plants. This doesn't mean that the procurement agencies were oblivious to productivity and they certainly attempted to direct demand to those plants that appeared most capable to produce, but this source of variation isn't captured by the instrument. Further, technological improvements and new varieties of aircraft arising from supply side innovation may have moved demand across aircraft models within broad categories (from "heavy" to "very heavy" bombers, for example), but unlikely across the broad categories we consider.

4 The Effects of Public Spending on Productivity

The dynamic response of output per hour worked to a 1% increase in demand is shown in Figure 4. The shaded area in this and subsequent figures give 95% Newey-West confidence bands. The model includes time and plant-by-model fixed effects, so that it reflects a relative increase in demand for a specific model at a specific plant. Predicted demand for aircraft based on the national trajectory of demand for broad aircraft types is used as an instrument for plant-level demand, as described in the previous section. The figure shows that labor productivity increases following a demand shock, by around half a percent within a year. The Cragg & Donald (1993) F-statistic of the first stage is 22, sufficient to reject a bias due to weak instruments greater than 5% with 95% confidence, according to Stock & Yogo (2002).[23]. The response appears persistent within the first year and a half, but it is difficult to ascertain whether the effects are longer-lasting. Estimates become very noisy beyond the reported horizon, as sample size drops substantially as the horizon increases in local projection estimation. Figure A.2 in the appendix shows aircraft production's own response to an increase in aircraft demand. While there is a large variance in the permanence of the demand shock, point estimates appear extremely persistent. It may therefore be useful to view the Figure 4 as reflecting the response of labor productivity to a persistent shock to demand.

[23] In all subsequent regressions, F-stats are reported in the figure notes and a 10% bias can be rejected in all cases and a 5% in most cases.

Nevertheless, this response indicates that by the end of the horizon, half of the monthly increase in plant-level output is met by the demand-induced growth in labor productivity.[24]

Figure 5 now shows the response of TFP in a similar specification. As discussed in Section 2, TFP is measured using a Cobb-Douglas production function with a labor share of $\frac{2}{3}$. Capital, in turn, is measured in physical units. Results are nearly identical when controlling for expenditure on equipment. Production lines receiving more orders show not only an increase in output per hour, but also in TFP. The smaller magnitudes reflect that some of the labor productivity increase was a result of greater use of capital. Nevertheless, we see that a one percent increase in demand increases TFP in quantity units by roughly one third of a percent within the first year, so that one third of the increase demand is partly satisfied through productivity growth by the end of the year. The response of TFP appears to die out by the end of the reported horizon, but estimates admittedly become too inaccurate at longer horizons to fully asses longer-term impacts.

The measure of capital stock used here already partially incorporates capital utilization, as it reflects actual floor space used by rather than available to the plant. However, Figure A.4 in the appendix shows that plants do respond to demand shocks with increased capital utilization, so that it is possible that labor productivity and measured TFP increased as a result. This echoes the concerns of Basu *et al.* (2006), who argue that TFP is mis-measured along the business cycle due to unmeasured capacity utilization. But Figure A.5 in the appendix shows that responses are nearly identical when controlling for capital utilization. Thus the increase in TFP reflects productivity growth that goes beyond measurement error due to increases in capital utilization.

Recall that TFP is measured in physical units (TFPQ) so that responses reflect an increase in aircraft produced rather than changes in prices or markups. Further, (plant-by-) model fixed effects reflect very narrowly defined models, with aircraft models re-coded at every design change. This means that results also largely control for (major) product quality changes. The interaction between plant and model fixed effects also controls for any quality differences in the same model of aircraft across plants. Productivity growth shown in the impulse responses therefore reflect real cost reductions within plant and aircraft model. This precision comes at a cost: we are unable to assess productivity changes manifested through quality growth (or deterioration). On one hand, reported results might understate productivity growth because aircraft quality may have improved and led to innovation in new aircraft types as production experience accumulated. On the other hand, the results might overstate productivity growth if demand pressures caused plants to cut corners and produce lower quality aircraft.

[24]Figure A.3 in the appendix shows that increases in demand at time zero aren't correlated with productivity in the previous months (no pre-trends) in this specification.

Capacity Utilization, Public Purchases, and Productivity

Having documented how demand affects productivity in the average aircraft plant, I now show important heterogeneity in responses to increased demand. The focus will be on the role of capacity utilization as motivated in the introduction.

Four different metrics are used here to measure firms' capacity constraints. These reflect the utilization of both labor and capital inputs and internal and local shortages in factor supplies. Figure 6a shows the evolution of capital utilization, measured through shift utilization, in the median plant over the course of the war. Capital utilization was high early in the war, peaking above 50% in late 1942. To put this in perspective, a plant working at full capacity 6 days a week (plants rarely operated on Sundays) with two 8-hour shifts running from 7am to 11pm would score 43% on this metric. The median plant was operating at a greater capacity than this benchmark throughout most of the wartime period. Beyond the time variation, there was substantial heterogeneity across plants in capacity utilization with the monthly plant observation at the 25 th percentile working at 39% capacity compared to 53% at the 75th percentile. Like other capacity utilization measures, capital utilization is endogenous and may be affected by current and anticipated productivity. I discuss below how the triple difference specification addresses some such concerns and possible remaining confounding factors.

Figure 6b shows the evolution of weekly hours per worker over the war. Workers in the median plant were active nearly 50 hours per week early in the war. Like capital utilization, labor utilization declined over the course of the war. This partly reflects massive increases in employment and capital over the duration of the war and this relieved pressure from each worker and unit of capital. Labor utilization was also very heterogeneous across plants, with workers at the 25 th percentile working at 43 hour per month compared to 54 at 75th.

Two additional indicators measure local labor market tightness. Local wages capture some differences in labor supply. Wartime wage controls limit the variation in the cost of labor across regions, so that relative wages tend to understate differences across counties. This also has an empirical advantage, as wages become less indicative of labor productivity. Relative wages across counties give a picture of labor scarcity, themselves potentially caused by wage controls. Labor shortages, in turn, impose capacity constraints on plants. I use county-level wages excluding the aircraft industry as a measure of labor market tightness. Results are similar including the aircraft industry in measuring local wages, but this helps distance the wage measure from differences in aircraft labor productivity across counties. Of course, the aircraft industry was large enough in some local labor markets to affect wages in other industries.

I include a second indicator of labor market tightness taken from the War Manpower Commission's county-level classification of local labor markets. The WMC classified each county on a scale of 1 to 4, with 4 indicating no local labor shortages and 1 reflecting acute labor shortages. Most aircraft plants were in regions classified as 1 or 2. This classification was used to restrict local hiring in non-essential industries, but the aircraft industry was always classified as high priority and a high WMC classification benefited rather than harmed aircraft plants. Nevertheless, these regulations didn't fully resolve labor market shortages and a higher WMC classification is likely correlated with local labor shortages even for aircraft plants.[25] It is difficult to find an indicator for local capital shortages. The market for capital had more of a national nature and large public investments attempted to address any need for plant expansions.[26]

For each of these indicators, a dummy variable c_p is assigned and equals one if plant p had an above average initial value of capacity constraints.[27] Table A1 in the appendix shows the cross correlations of all four dimensions of heterogeneity. It is hardly surprising that different indicators of capacity constraints are correlated, but correlations are far from perfect and in some cases close to zero. This indicates that results are based on different dimensions of heterogeneity are giving separate indicators of the importance of capacity constraints in the transmission of public demand.

A triple difference in differences gives the differential effects of demand depending on capacity constraints:

$$y_{mpt+h} = \beta_h^{3D} \widehat{[D \times c]}_{mpt} + \omega_h \widehat{D}_{mpt} + \text{lags} + \text{FE} + \varepsilon_{mpt}^{3D}. \tag{3}$$

As in (2), $\widehat{D}_{mpt}$ gives demand for aircraft of model m from plant p in month t, predicted by the instrument in the first stage of the two stage least squares. However, the coefficient of interest is now β_h^{3D}, on the interaction between demand and capacity constraints metric c_p. Demand D_{mpt} and its interaction with capacity constraints c_p are jointly projected on the instrument, the dummy and their interaction in a first stage analogous to (2). The capacity constraints dummy itself is excluded as its variation is absorbed by plant by model fixed effects.

Figure 7 plots local projections–the β_h^{3D} coefficients–for a specification using capital utilization as the heterogeneity dummy c_p. Impulse responses have a triple difference in differences interpretation, showing by how much plants with higher initial capital utilization saw an increase in output per hour worked in response to a one percent increase in demand (predicted by the instrument). Plants with high capital utilization show a substantially larger increase in productivity

[25] I use the earliest observation of wages (or WMC classification) in each county available in the data, typically late 1942, as the labor market tightness metric. The WMC attempted to redirect workers to locations with labor shortages. As long as these efforts left some cross-sectional heterogeneity in labor market tightness, these indicators will be informative of capacity constraints rather than the benefit of WMC support.

[26] Nearly two thirds of all aircraft plant expansions were publicly financed: see Ilzetzki & Reichardt (2020). Plants rented public capital from the government at a fixed rate for the duration of the war and were given the option to purchase capital at its original cost at the end of the war. See Wilson (2018), chapter 6, on the post-war auctioning of public capital and see Garin (2019) on the long run effects of these public investments.

[27] We take the first available observation for each plant, which is typically the first month they delivered aircraft for the war production drive. This initial date differs across plants. Setting the dummy based on the January 1943 value or the average value over the war gives similar results.

in response to an increase in demand than do those with more initial slack. The magnitudes are substantial and approximately two thirds of the response of the full sample shown in Figure 4.

Differences between plants operating at high capacity and those with slack are noticeable in the response of other variables. Figure 8a shows that capacity-constrained plants respond to an increase in demand with a relative decline in capacity utilization. This is better stated conversely: Plants operating at low capital utilization rates respond to demand by *increasing* capital utilization by roughly 4pp relative to capacity constrained plants. The effect is not only statistically significant, but economically very large. It reflects a nearly 10% (relative) increase in capacity utilization compared to a median capacity utilization of 46% (across all plants and months).

It has long been hypothesized that capacity constraints play an important role in plants' responses to demand shocks. Basu *et al.* (2006) argue that measurement of the cyclical properties of TFP are confounded with capacity utilization and may actually reflect demand fluctuations. We previously saw that part of the demand shock is satisfied through increased capital utilization in the average plant (Figure A.4 in the appendix), but that this alone was insufficient to fully account for plants' rise in TFP (Figure A.5 in the appendix). Figure 8a shows, however, that this response isn't uniform across plants. As could be expected, plants with excess capacity see a larger capacity utilization than those operating at full (perhaps unsustainable) capacity. The notion that the impact of fiscal policy depends on the degree of capacity utilization is implicit in the hypothesis that fiscal policy has larger effects in recessions (e.g. Auerbach & Gorodnichenko (2012), 2013). Figure 8a shows evidence of a differential response of capital utilization at the plant-level, depending on the degree of slack in production. However, Figure 7 shows that capacity constrained plants find other ways to respond to demand and do so by increased productivity. This might partly explain why it is difficult to find heterogeneity in fiscal multipliers depending on slack in the economy, in response to military spending shocks (Owyang *et al.* (2013) Ramey & Zubairy (2018)).

How, then, do capacity-constrained plants increase production in face of increased public demand. Figure 8b shows that the increases in labor productivity shown in Figure 7 aren't attributable to increases in capital. It presents the relative response of capital per hour worked in plants with greater capital utilization, in a similar triple differences specification. Impulse responses hover around zero and aren't statistically significant (although error bands are wide).[28]

Figure 8c shows one mechanism that may have contributed to the higher labor productivity in capacity constrained plants. It shows that high utilization plants increased outside production by 10pp in the year following a 1 percent increase in aircraft demand. This is a very large response, given that outsourced production was 30% in the median plant-month. Historical accounts point to

[28]It might seem implausible in any case that plants can increase their capital stock immediately to an increase in demand. However, the measure of physical capital used in these estimates is actual floor space used in production, which implicitly includes an indication of capacity utilization. This means that the capital stock could be installed well in advance and only used when demand increased.

outsourcing as an important tool used by capacity-constrained plants to meet increased demand.[29] For example, Fairchild & Grossman (1959) write:

> The dispersal of subcontracts outside the critical area [of plants failing to meet production schedules] was encouraged, with the result that in September the Boeing Company placed subcontracts for approximately 40 percent of its work and made plans to let out subcontracts for an additional 20 percent. (Page 132)

With labor productivity increasing more in capacity-constrained plants and no difference in capital use, 8d shows that TPF increased by roughly 0.2% more in high utilization plants relative to those with more free capacity. Relative to the literature on learning by doing, I document important heterogeneity in the extent to which increased demand led to productivity gains. This suggests that learning may not be a merely passive process, but depends on plants' incentives and actions, which may differ depending on their production constraints. We see this to some extent in the response of outsourcing in Figure 8c, which shows a very similar trajectory to to the overall increase in TFP shown in Figure 8d.

Figure 9 shows similar heterogeneity using other metrics of capacity constraints. Figure 9a repeats Figure 8d, where capacity utilization is measured through capital (shift) utilization. Figure 9b shows that TFP also increases more in response to demand shocks in plants with higher labor utilization. Figure 9c repeats the exercise comparing plants operating in counties with initially higher wage rates, partly reflecting tighter labor markets. Plants in high wage counties see larger productivity gains from increases in demand. Finally we see similar heterogeneity depending on labor market tightness as classified by the WMC in Figure 9d. Given the categorical nature of the WMC classification, the dummy c_p takes on a value of one for plants with the highest WMC labor market classification of 1 and zero otherwise (in most cases a WMC classification of 2).

All dimensions of capacity constraints are likely endogenous to economic conditions, demand, and plants' productivity. Note, however, that the triple differences specification in (3) absorbs the direct differences in productivity *levels* across plants with different constraints through plant (by model) fixed effects. The government may have directed more procurement to plants with greater perceived capacity, but this is absorbed by the direct control for demand in the specification.[30] Lagged productivity absorbs production-line level productivity trends. The impulse responses therefore reflect remaining *differential* correlation between demand and productivity when comparing plants with high and low capacity constraints. One possible threat to identification would arise

[29] Measures of hours worked per aircraft in the AMPR include outsourced labor hours, so rising labor productivity as as result of outsourcing isn't merely an accounting artefact. However, it is possible that outsourcing is confounded with an increase in capital per worker if the sub-contractor had a higher capital to labor ratio.

[30] Capacity utilization was indeed an important consideration in procurement decisions. See for example Fairchild & Grossman (1959) chapter VI.

if the government directed procurement to constrained plants only when they perceived them to be more productive than usual. However, the instrument ensures that we only rely on variations in demand coming from national demand for broad aircraft types. Additional concerns would have to reflect other ways in which productivity correlates differently with demand along all four dimensions of heterogeneity.

It is nevertheless useful to explore factors correlated with capacity utilization that may give the heterogeneous responses different interpretations. Table A2 in the appendix gives summary statistics along all four dimensions of heterogeneity. The first row shows labor productivity growth rates in the *subsequent* two years–the majority of the war production drive. It shows that the impulse responses donât reflect differences in trend productivity growth across plants facing different capacity. Hence, it is truly the triple differences that is reflected in Figure 7, i.e. it is only conditional on a demand shock that productivity increases relatively in constrained plants. There is also no statistically significant difference in firm age, the number of aircraft produced in January 1943, the unit cost of aircraft, wingspan, and the cumulative amount of public financing received during the war.[31]

One correlate with capacity constraints does stand out. Plants with high capital utilization (and those in WMC local labor markets classified as 1) were older on average. This is in the context of a very young industry: the average high capacity utilization plant was founded in 1932. But low capacity utilization plants were even younger and were founded in 1938 on average. This raises the concern that capacity utilization merely captures plant age. Indeed, young plants were more likely to have lower capital utilization early in the war as many were still ramping up production. It is certainly plausible that young plants respond differently than old ones to increases in demand. However, most stories that might explain such a phenomenon would seem to go the wrong way. If anything, one might expect young plants' productivity to benefit *more* from demand than that of old plants. In any case, Figure A.6 in the appendix shows that differential responses across capital utilization aren't driven by plant age. It runs a similar regression as in (3), but also controls for plant age and an interaction between demand and a dummy equaling one for plants above the median age, allowing for the possibility that the interaction between demand and age rather than capacity drives the triple difference results. While error bands are now wider, the response remains statistically significant and has a similar trajectory over time as the specification without controls.[32]

[31]Plants with low wages received more public financing, but this goes in the wrong direction to explain the higher productivity growth of high-wage plants following a demand shock. Further, TFP estimates already control for capital–public and private–and results are robust to controlling directly for public investments.

[32]With these additional controls, the first stage F statistic is too small to reject weak instruments. The figure shows the reduced form regression of the interaction of the instrument on initial capital utilization.

5 Conclusion

A traditional view of the transmission of fiscal policy posits that increased demand increases output as firms soak up under-utilized employment or capital, either within or outside the firm. The neoclassical view posits that production will increase due to increased labor supply. Both theories would suggest that public demand can do little to expand output when the economy is at very high rates of utilization, as was the case in World War II. Indeed, this was the common view among at the onset of the war, reflected in the "feasibility dispute", where economists warned that the economy could not sustain the planned war production drive, while the military insisted that it must.

This paper contributes to this debate by showing that slack does indeed play an important role in plants' responses to increased public demand. During the Second World War, plants with substantial slack, and in regions with looser labor markets, did indeed respond to the massive increase in public purchases with relative increases in capacity utilization. But plants with higher rates of capacity utilization–in fact rates of utilization that would normally be viewed as "overheating"– met the production challenge through productivity increases.

This evidence is based on archival data on airframe production during the second world war, the largest shock to public spending in US history. Can an episode so distant in history have implications to the modern economy? In the age of Covid-19, it arguably can. The pandemic has affected different sectors of the economy differently, with some showing substantial excess capacity and shortages arising in others. This study suggests that public purchases may stimulate sectors experiencing a demand shortfall through greater utilization of their existing capacity. But it also shows that business that are strained may find ways to enhance productivity in face of increased demand.

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.

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.

ASHER, HAROLD. 1956. *The fiscal revolution in America: policy in pursuit of reality*. 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. 2019. *Local fiscal multipliers and fiscal spillovers in the united states*. Tech. rept. National Bureau of Economic Research.

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.

BRUNET, GILLIAN. 2017. *Stimulus on the Home Front: The State-Level Effects of WWII Spending*.

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

BURNSIDE, CRAIG, & EICHENBAUM, MARTIN. 1994. *Factor hoarding and the propagation of business cycles shocks*. Tech. rept. National Bureau of Economic Research.

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

CRAGG, JOHN G., & DONALD, STEPHEN G. 1993. Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory*, **9**(2), 222–240.

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.

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

FERNALD, JOHN. 2014. A quarterly, utilization-adjusted series on total factor productivity. Federal Reserve Bank of San Francisco.

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, 1939â58. *Economic History Review*, **66**(4), 975–992.

GARIN, ANDY. 2019. *Public Investment and the Spread of "Good-Paying" Manufacturing Jobs: Evidence from World War IIâs Big Plants*. Tech. rept.

GIORCELLI, MICHAELA, & BIANCHI, NICOLA. 2020. *The Dynamics and Spillovers of Management Interventions: Evidence from the Training Within Industry Program*. Tech. rept.

GOLDIN, CLAUDIA, & OLIVETTI, CLAUDIA. 2013. Shocking Labor Supply: A Reassessment of the Role of World War II on Women's Labor Supply. *American Economic Review*, **103**(3), 257–262.

GROSS, DANIEL P., & SAMPAT, BHAVEN N. 2020 (Oct.). *Organizing Crisis Innovation: Lessons from World War II*. NBER Working Papers 27909. National Bureau of Economic Research, Inc.

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

HALL, ROBERT E. 2009. By How Much Does GDP Rise If the Government Buys More Output? *Brookings Papers on Economic Activity*, **40**(2 (Fall)), 183–249.

HANLON, WALKER, & JAWORSKI, TAYLOR. 2019 (Nov.). *Spillover Effects of IP Protection in the Interwar Aircraft Industry*. NBER Working Papers 26490. National Bureau of Economic Research, Inc.

HICKMAN, BERT G. 1957. Capacity, Capacity Utilization, and the Acceleration Principle. *Chap. c5590 of:* ON RESEARCH IN INCOME, CONFERENCE, & WEALTH (eds), *Problems of Capital Formation: Concepts, Measurement, and Controlling Factors*. NBER.

HOLLEY, IRVING BRINTON. 1964. *Buying Aircraft: Materiel Procurement for the Army Air Forces*. Office of the Chief of Military History, Department of the Army.

ILZETZKI, ETHAN, & REICHARDT, HUGO. 2020. Ramping up ventilator production: Lessons from WWII. *VoxEU*, Apr.

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

JONES, LARRY E., & MANUELLI, RODOLFO E. 2005. Neoclassical Models of Endogenous Growth: The Effects of Fiscal Policy, Innovation and Fluctuations. *Chap. 1, pages 13–65 of:* AGHION, PHILIPPE, & DURLAUF, STEVEN (eds), *Handbook of Economic Growth*. Handbook of Economic Growth, vol. 1. Elsevier.

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 (Jan.). *The long-run effects of monetary policy*. CEPR Discussion Papers 14338. C.E.P.R. Discussion Papers.

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

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.

LONG, CLARENCE D. 1952. *The Labor Force in War and Transition: Four Countries*. NBER.

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

MIDDLETON, KENNETH A. 1945. *Wartime Productivity Changes in Airframe Industry*.

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

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.

NICKELL, STEPHEN J. 1981. Biases in Dynamic Models with Fixed Effects. *Econometrica*, **49**(6), 1417–1426.

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, 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.

RHODE, PAUL. 2000. The Impact of World War Two Spending on the California Economy. *Chap. 0, pages 93–119 of:* LOTCHIN, R. (ed), *The Way We Really Were: The Golden State in the Second Great War*, vol. 2. University of Illinois Press.

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.

ROMER, PAUL M. 1994. The Origins of Endogenous Growth. *Journal of Economic Perspectives*, **8**(1), 3–22.

SCOTT-KEMMIS, DON, & BELL, MARTIN. 2010. The mythology of learning-by-doing in World War II airframe and ship production. *International Journal of Technological Learning, Innovation and Development*, **3**(1), 1–35.

SEARLE, 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*. United States Army, Center of Military History.

STOCK, JAMES H., & YOGO, MOTOHIRO. 2002 (Nov.). *Testing for Weak Instruments in Linear IV Regression*. NBER Technical Working Papers 0284. National Bureau of Economic Research, Inc.

SYVERSON, CHAD. 2011. What Determines Productivity? *Journal of Economic Literature*, **49**(2), 326–65.

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. *Chap. 0, 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.

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

WAR PRODUCTION BOARD. 1945. *War Production in 1944*. Tech. rept. War Production Board.

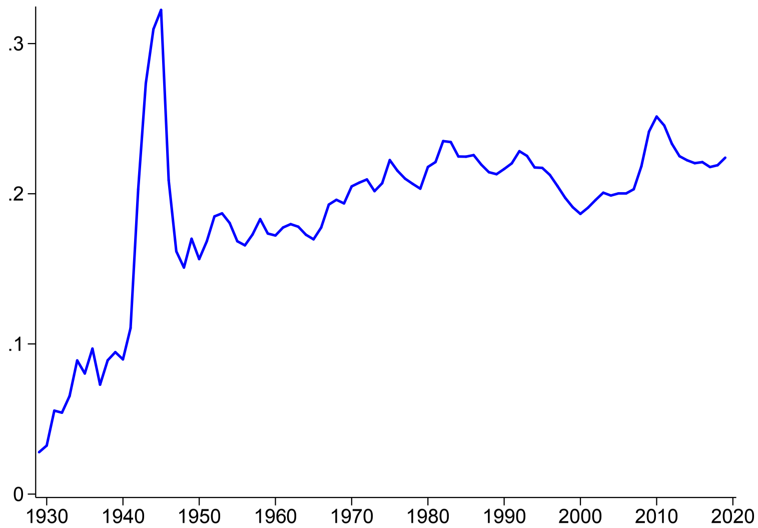
WILSON, MARK R. 2018. *Destructive Creation: American Business and the Winning of World War II*. University of Pennsylvania Press.

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

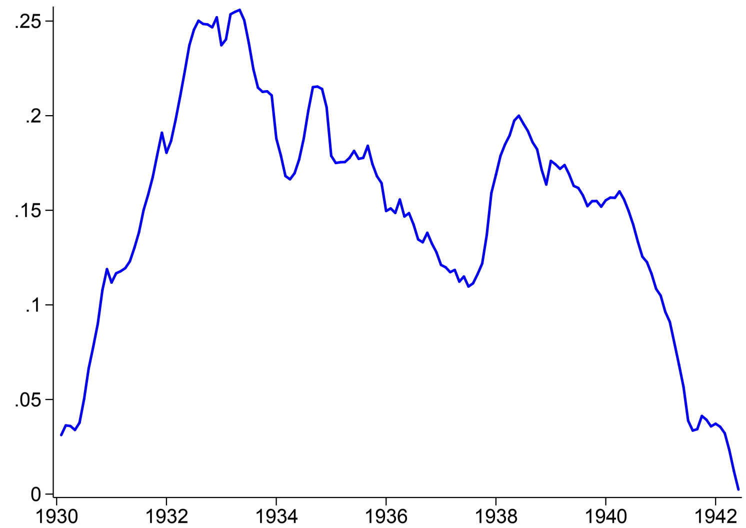
YOSHPE, HARRY B. 1945. *Labor Problems in Quartermaster Procurement: QMC Historical Studies 11*. Office of the Quartermaster General.

YOUNG, ALWYN. 1991. Learning by doing and the dynamic effects of international trade. *The Quarterly Journal of Economics*, **106**(2), 369–405.

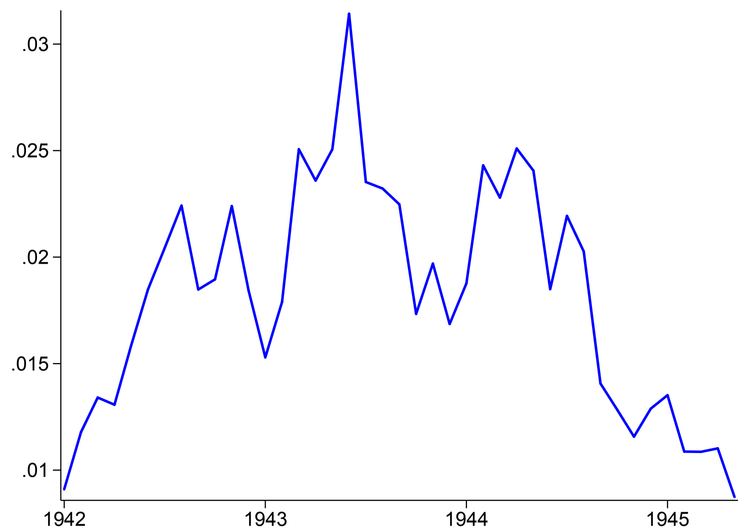
Figure 1: Government Spending, Unemployment, and Aircraft Procurement in the Second World War



(a) Public Expenditure as a Percent of GDP



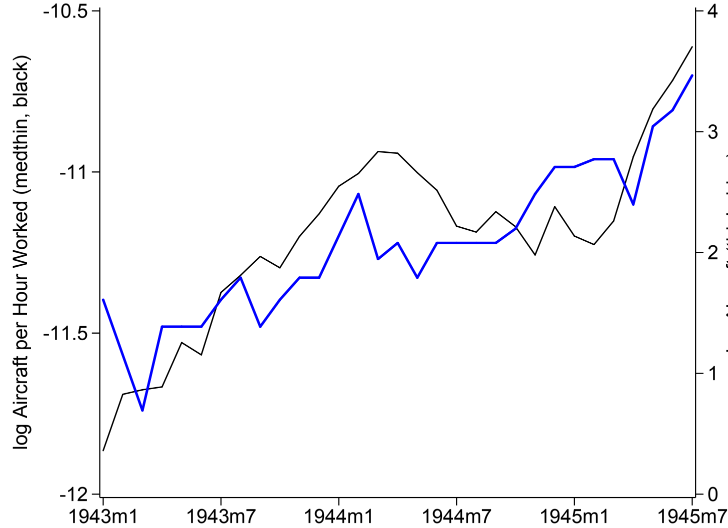
(b) Unemployment



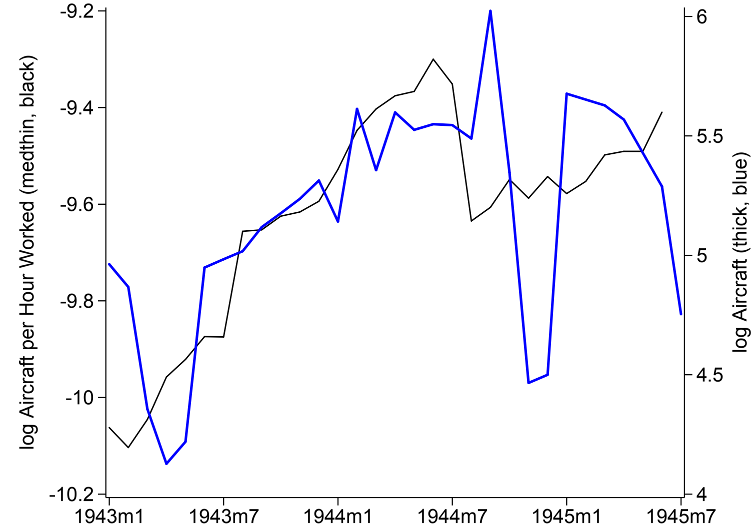
(c) Aircraft Procurement

Note: Panel (a) shows public expenditure as a percent of GDP in the US since 1929. Source: Bureau of Economic Analysis and the author. The World War II public spending surge is the largest in US history by far. Panel (b) shows the US unemployment rate from the great depression to the US formal entry into the Second World War. Source: NBER Macrohistory Database. At the onset of hostilities in Europe, the unemployment rate was still above 15%. By the time the US formally entered the war, the economy was at full employment. Panel (c) shows annualized government aircraft procurement as a percent of pre-war (1939) GDP. Source: War Production Board Listing of Major War Supply Contracts, BEA, and the author. Government spending on aircraft alone exceeded 3% of GDP.

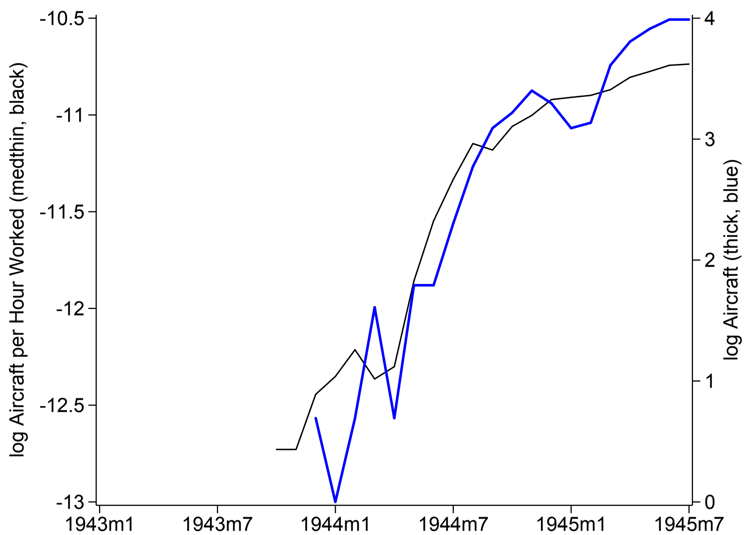
Figure 2: Output per Hour and Monthly Production in Selected Production Lines



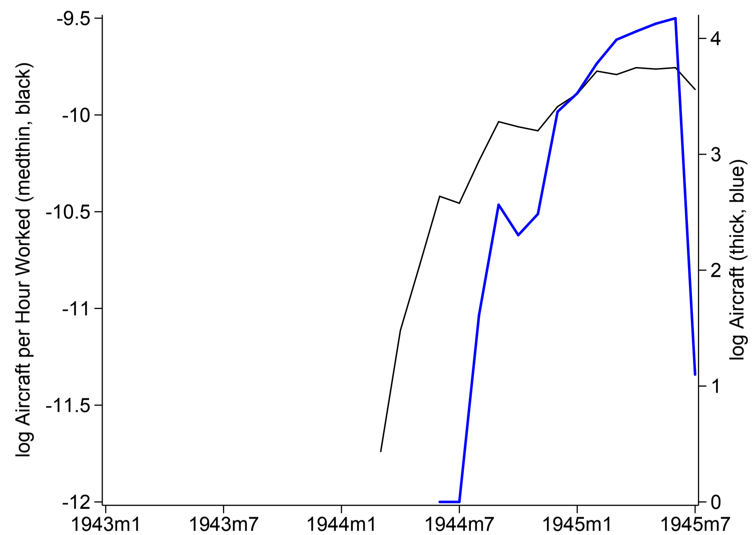
(a) C-54 at Douglas, Santa Monica



(b) B-24 at Consolidated-Vultee, San Diego



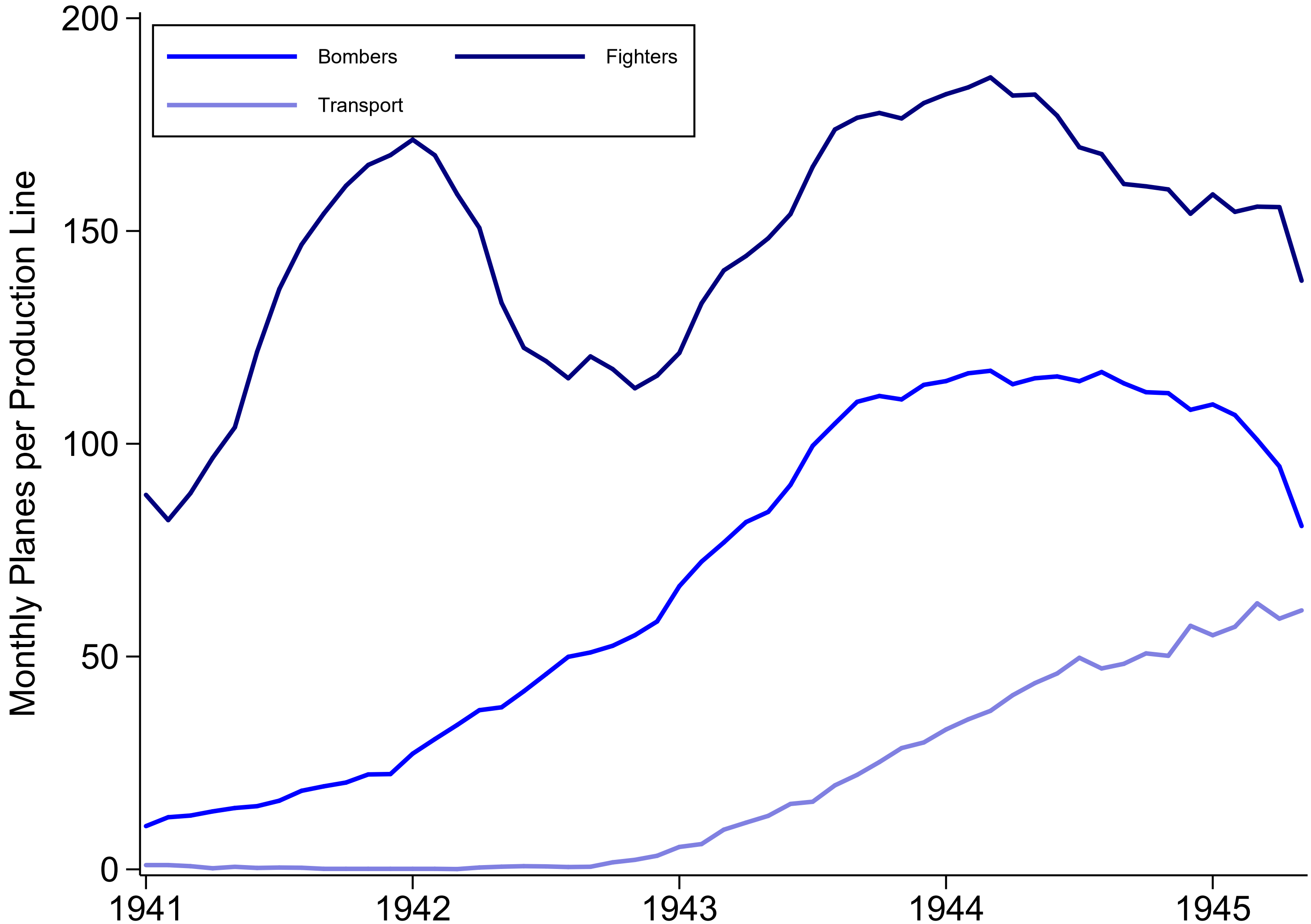
(c) C-54 at Douglas, Chicago



(d) PB4Y at Consolidated-Vultee, San Diego

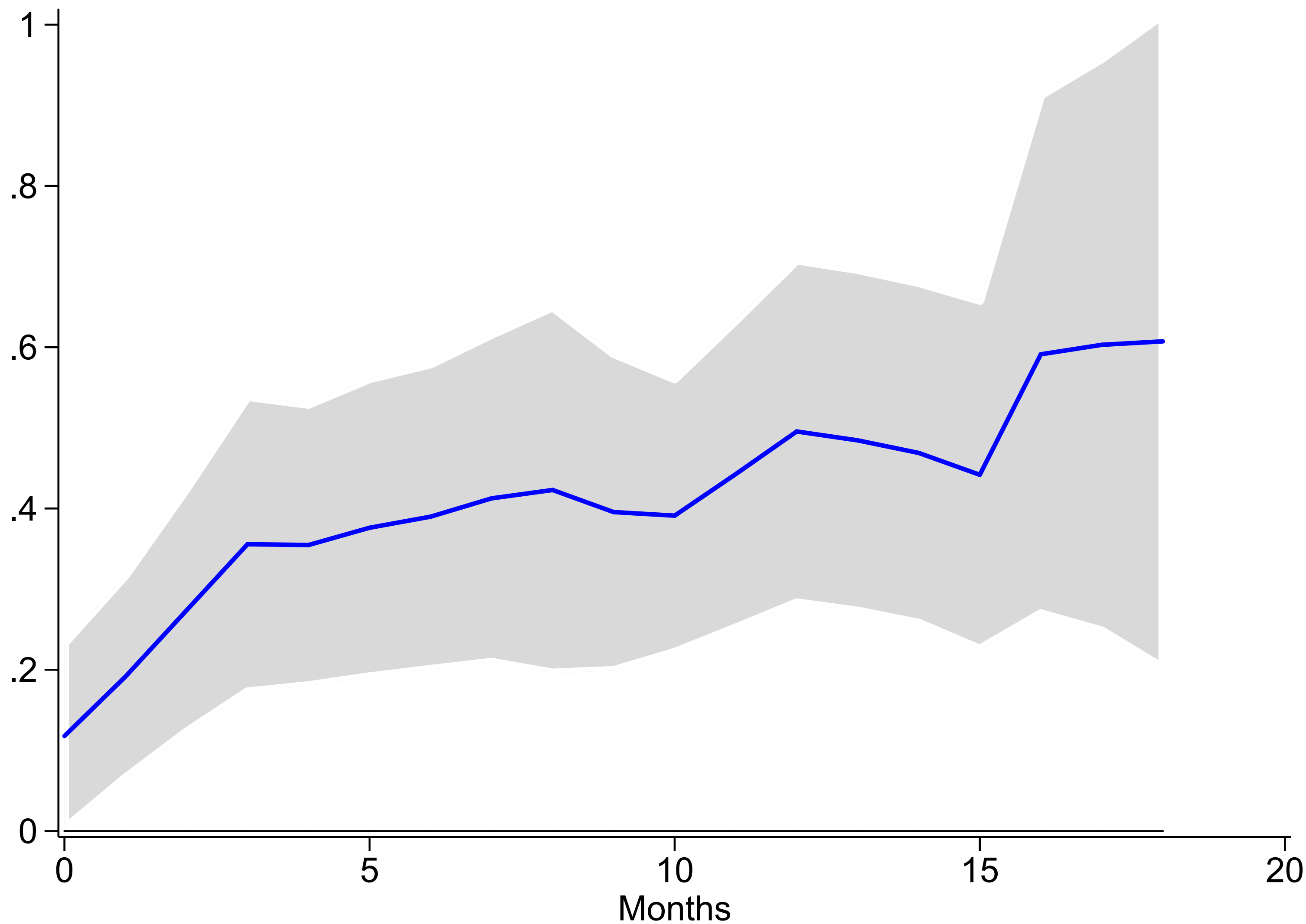
Note: Each panel shows the logarithm of an aircraft model delivered (blue, right hand scale) and the logarithm of aircraft of this type produced per hour worked (black, left hand scale). Panels (a) and (c) show C-54 aircraft in two Douglas Aircraft plants. Panel (a) is the older Santa Monica plant and Panel (c) is the newer Chicago plant, founded in 1942. Panels (b) and (d) show two aircraft models produced in Consolidated-Vultee's San Diego plant. Panel (b) shows B-24 bombers and Panel (d) PB4Y bombers. The figure shows the mechanical coincidence between productivity and output. It also shows that known shocks to productivity–new plants producing the same aircraft and new models produced in the same plant–affect the productivity and therefore output of existing production lines.

Figure 3: Aircraft Production per Plant in Plants Specializing in Bomber and Fighter Aircrafts



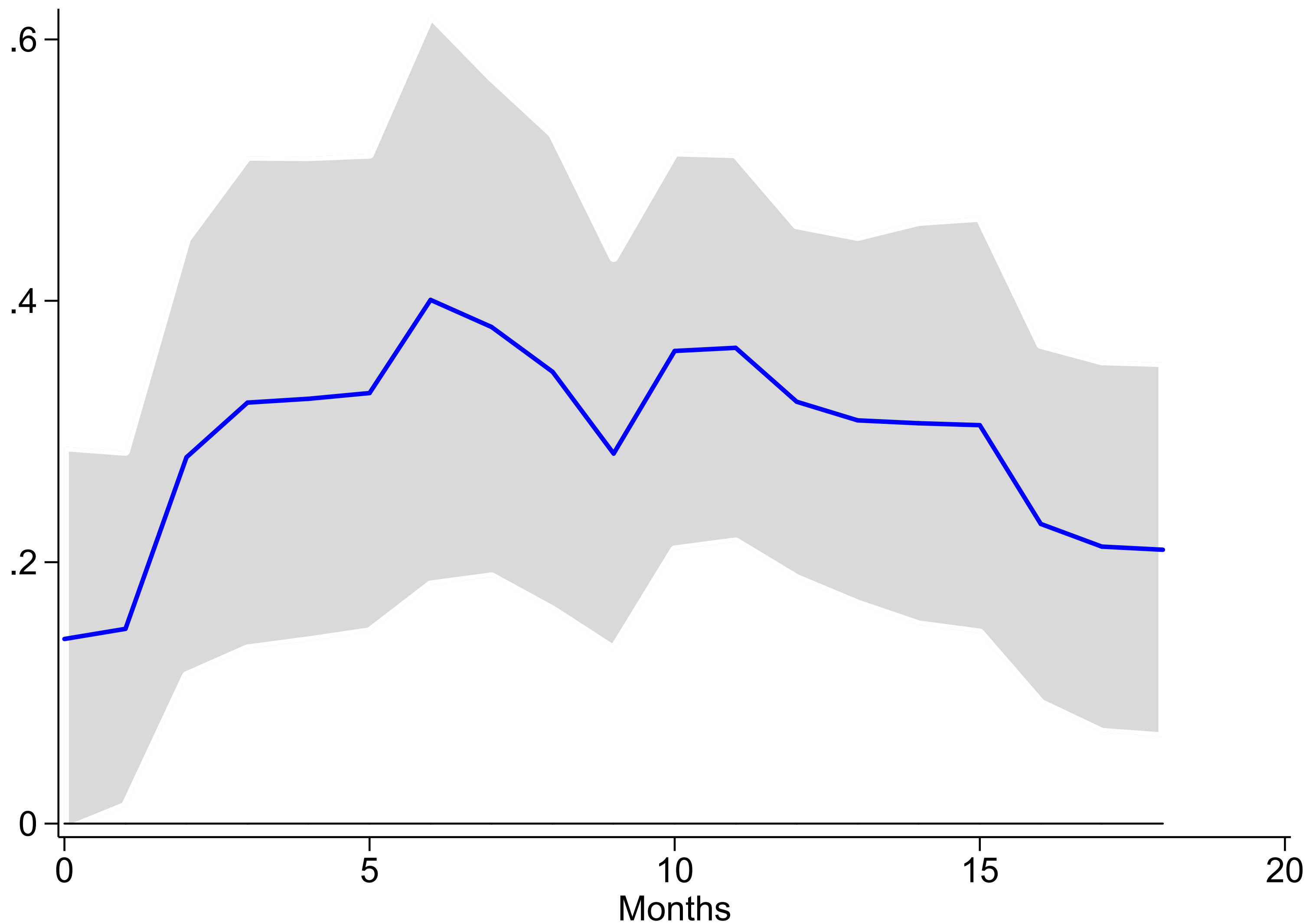
Note: The figure illustrates the instrument described in equation (1) in the text. It shows monthly aircraft production in plants specializing in fighter (top) and bomber (bottom) aircraft types. Specialization is defined as having more than 75% of production concentrated in an aircraft type in January 1943. The two aircraft plant types showed different production trajectories as the war evolved, leading to differential demand from the two plant types. The identifying assumption is that different production trajectories across plant types was driven by this differential demand, not other productivity drivers. Fighter aircrafts were more prominent in lend-lease acquisitions by US allies in 1941, leading to a boom and bust in their production in 1941-42. US entry into the war revived fighter production in 1943. Bombers played a more important part in the later war years and strategic bombing campaigns became more central to US strategy.

Figure 4: Response of Output per Hour Worked to a 1% Shock to Aircraft Demand



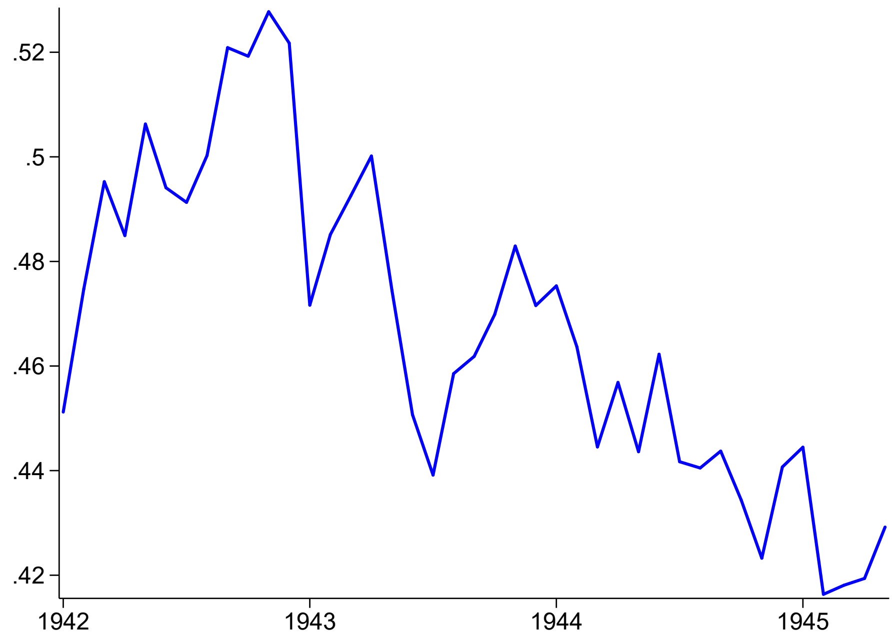
Note: The figure shows the response of (log) aircraft per hour worked to a one percent shock to aircraft demand. The shaded area shows 95% confidence intervals. Estimates are based on local projections, with aircraft demand instrumented with the instrument described in Section 3, and laid out in equations (1) and (2). First stage F-statistic = 22.

Figure 5: Response of TFP to a 1% Shock to Aircraft Demand

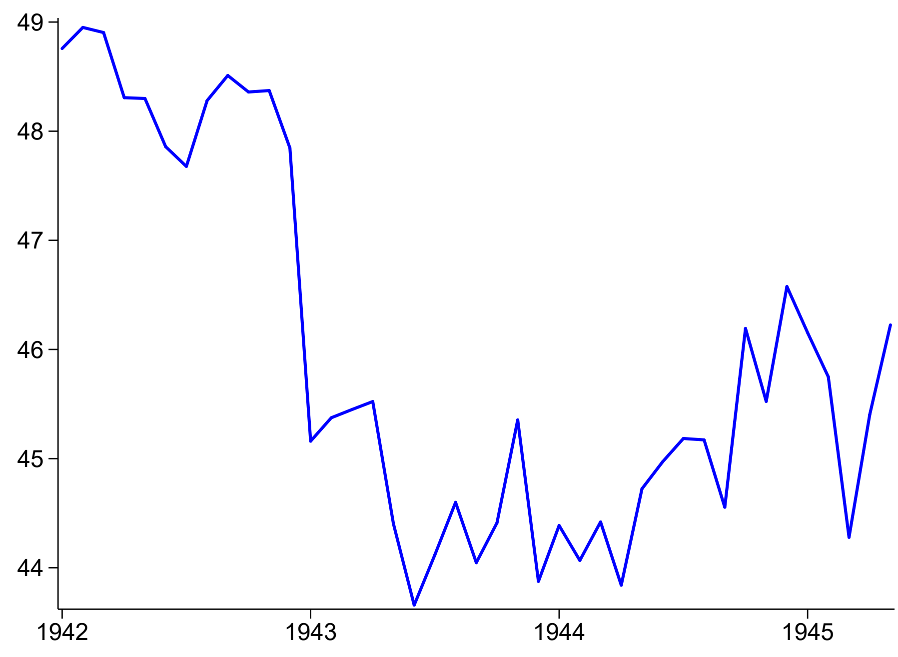


Note: The figure shows the response of TFP to a one percent shock to aircraft demand. The shaded area shows 95% confidence intervals. Estimates are based on local projections, with aircraft demand instrumented with the instrument described in Section 3, and laid out in equations (1) and (2). First stage F-statistic = 33.

Figure 6: Capital and Labor Utilization in Airframe Plants



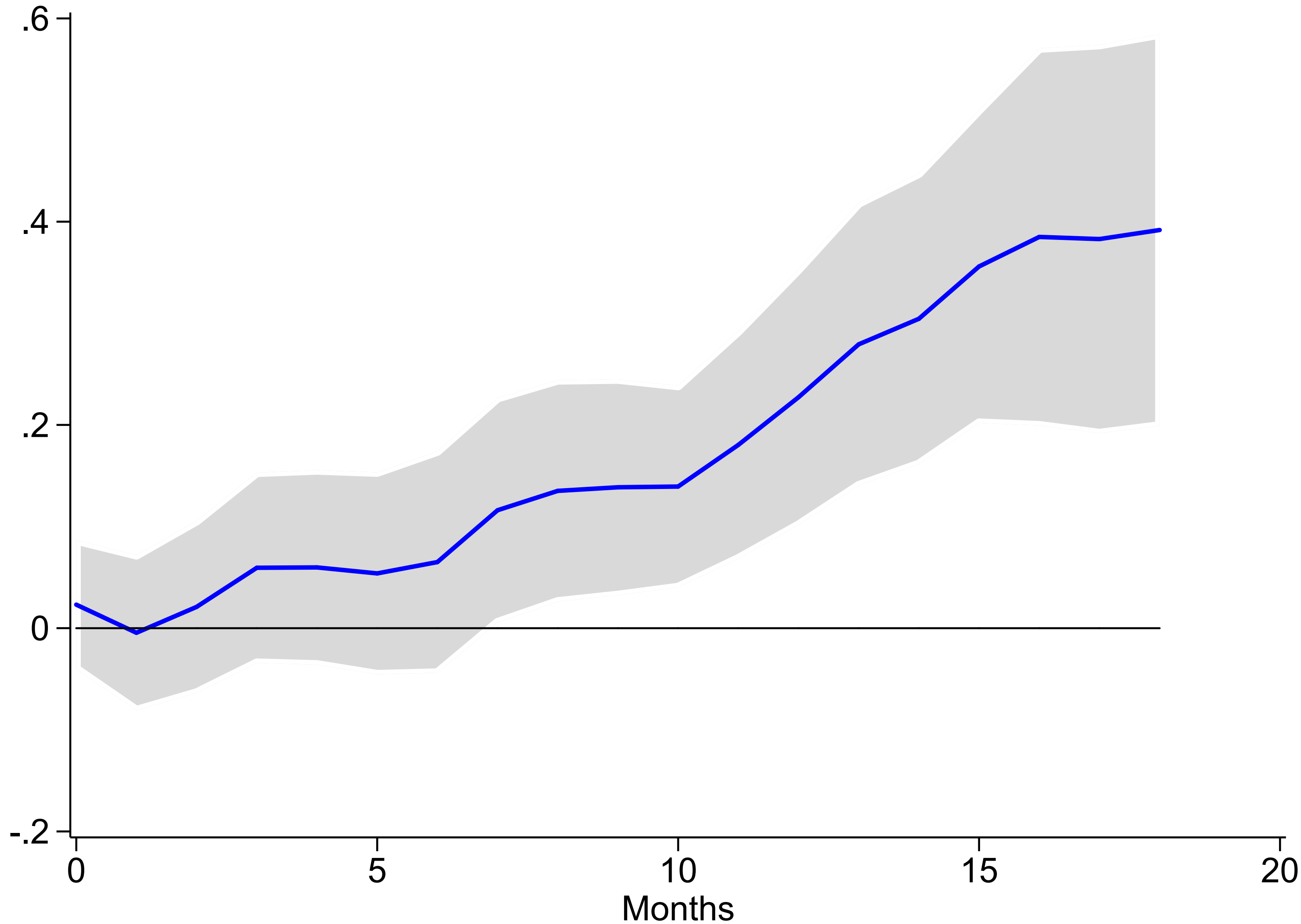
(a) Capital (Shift) Utilization



(b) Hours per Worker

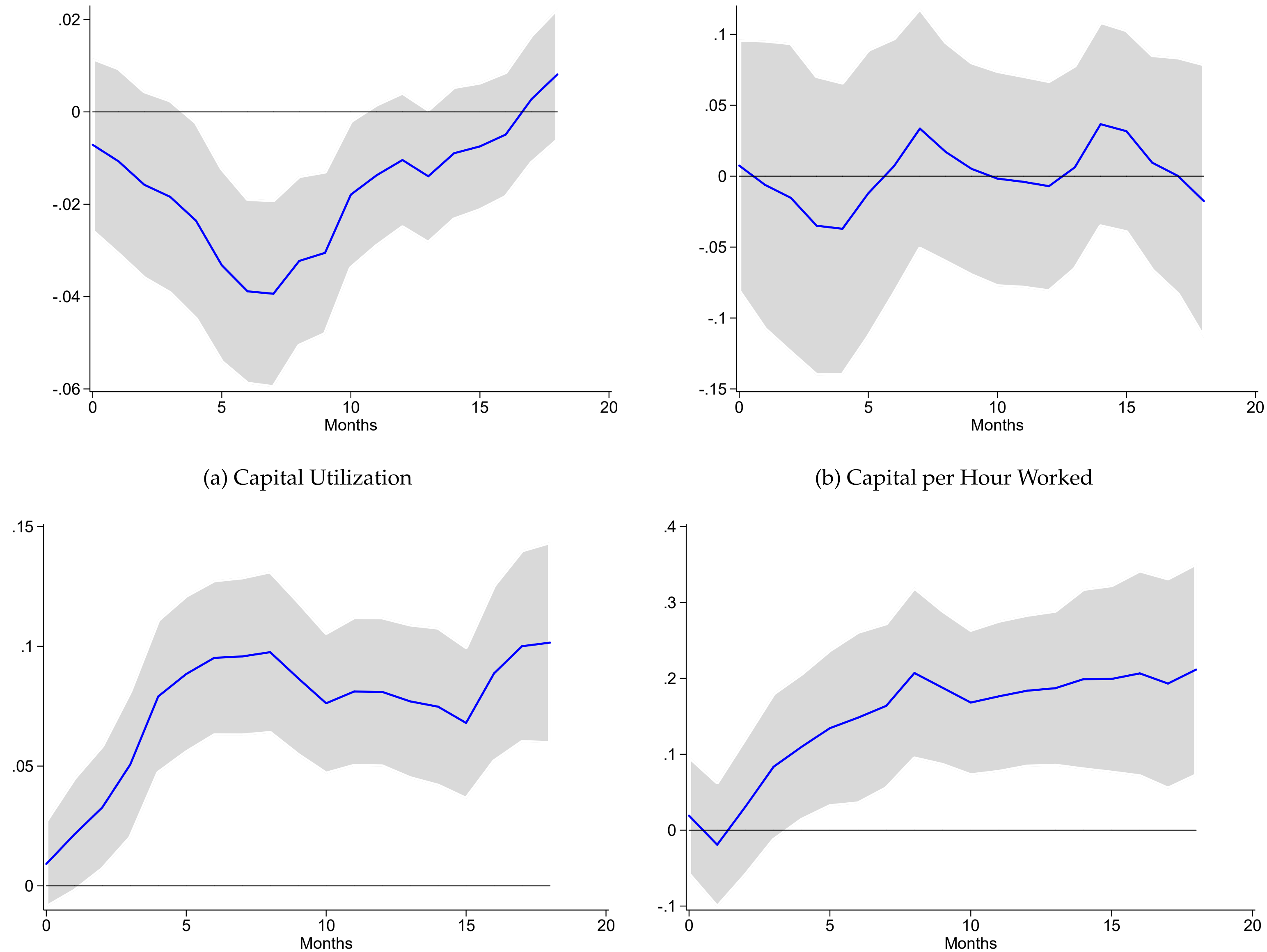
Note: Panel (a) shows shift utilization for the median airframe plant, estimated as described in Section 2. Pane (b) shows hours per worker in the median airframe plant. Source: AMPR and the author.

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



Note: The figure shows the response of (log) aircraft per hour worked to a one percent shock to aircraft demand interacted with a dummy variable equaling one if the plant had above average initial capital utilization. The shaded area shows 95% confidence intervals. Estimates are based on local projections, with aircraft demand instrumented with the instrument described in Section 3, and laid out in equations (1) and (3). First stage F-statistic =15.

Figure 8: Relative responses of to a 1% Shock to Aircraft Demand in High Capital Utilization Plants



(a) Capital Utilization

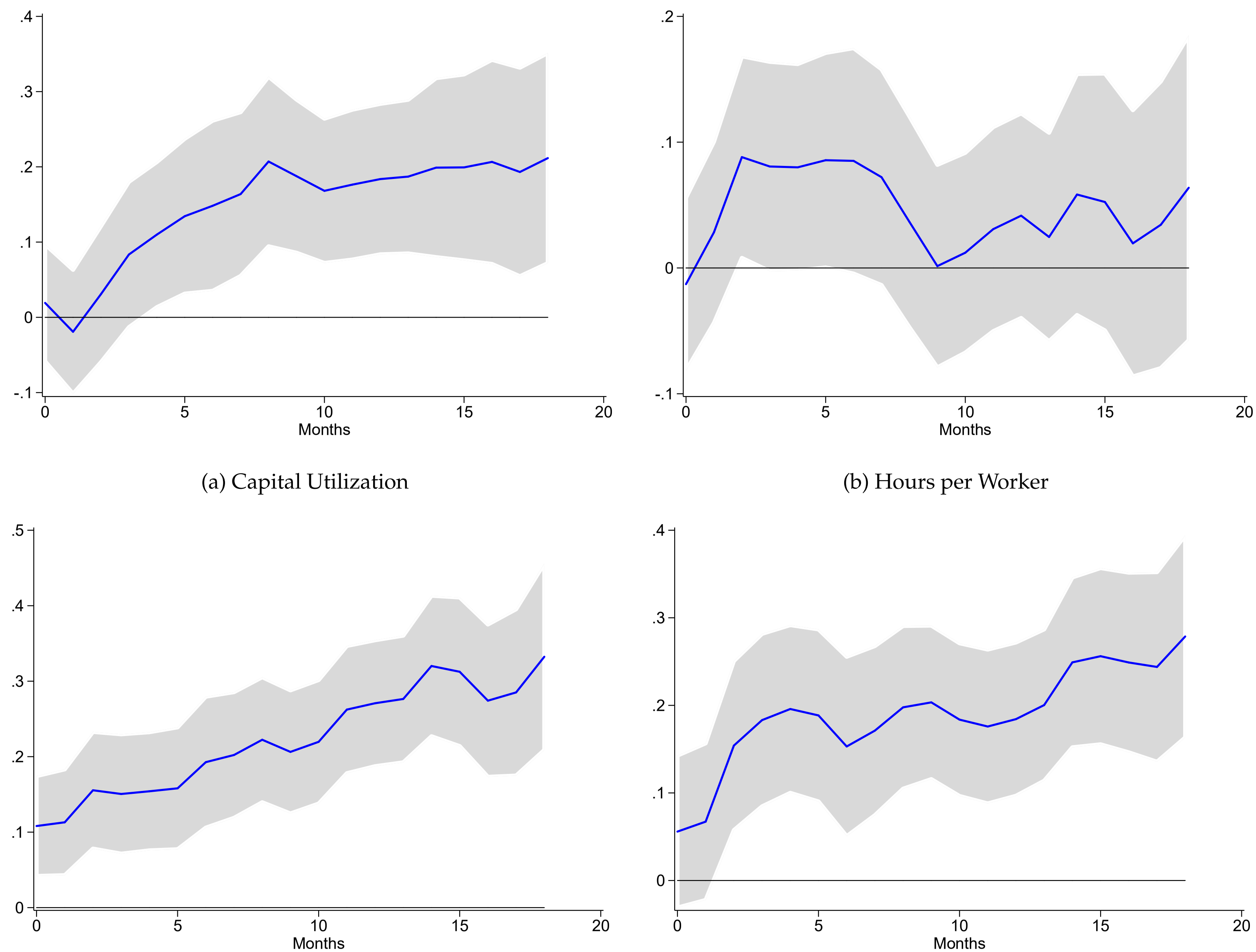
(b) Capital per Hour Worked

(c) Outsourced production (percentage points)

(d) TFP

Note: Each panel shows the response of a variable to a one percent shock to aircraft demand interacted with a dummy variable equaling one if the plant had above average initial capital utilization. The shaded areas show 95% confidence intervals. Estimates are based on local projections, with aircraft demand instrumented with the instrument described in Section 3, and laid out in equations (1) and (3). First stage F-statistics are 18, 25, 26, and 30.

Figure 9: Relative Responses of TFP to a 1% Shock to Aircraft Demand Across Several Dimensions of Heterogeneity



(a) Capital Utilization

(b) Hours per Worker

(c) County-Level Wages

(d) Manpower Classification = 1

Note: Each panel shows the response of TFP to a one percent shock to aircraft demand interacted with a dummy variable equaling one if the plant had the measure of constraints on factors of production indicated in the panel. The shaded areas show 95% confidence intervals. Estimates are based on local projections, with aircraft demand instrumented with the instrument described in Section 3, and laid out in equations (1) and (3). First stage F-statistics are 30, 43, 43, and 42.

Table 1: Correlation Between Productivity and Output

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cumulative output	0.382***	0.406***	0.322***	0.294***	0.326***		0.278***	0.0147
	(0.0105)	(0.0111)	(0.00377)	(0.00540)	(0.00596)		(0.00967)	(0.0113)
Current output						0.268***	0.0574***	0.0426***
						(0.00670)	(0.00930)	(0.00490)
Time FE		X		X	X	X	X	X
Plant FE			X	X				
Plant*Model FE					X	X	X	X
Lagged productivity								X
Observations	2553	2553	2553	2553	2553	2491	2491	1906

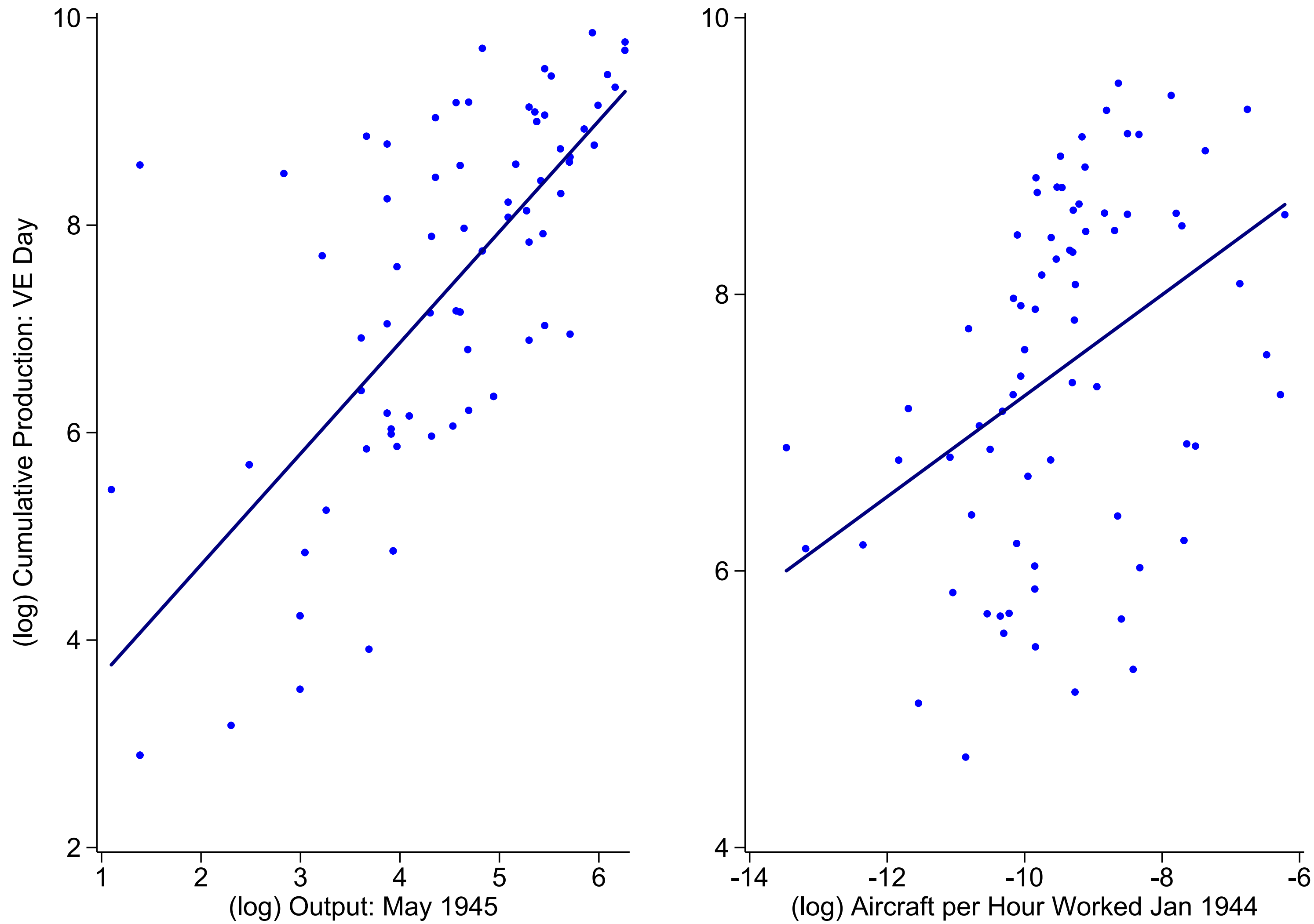
Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The table gives raw correlations (OLS regressions) between (log) aircraft produced per hour worked and (log) production, in a panel of production lines. Columns (1) to (5) are based on regressions of (log) cumulative output on productivity. Column (1) has no controls. The following columns include fixed effects of time (2); plant (3); time and plant (4); time and plant-by-model (5-8). Columns regress current output (6); and current and cumulative output (7) on productivity. Column (8) adds 6 monthly lags of the dependent variable (output per hour) to the regression of both output measures on productivity. Both current and cumulative output are correlated with productivity, even with the most saturated set of fixed effects. The correlation between productivity and cumulative output becomes insignificant when controlling for the lagged dependent variable, suggesting that the correlation between productivity and "experience" is driven by auto-correlation in productivity. Past productivity leads to both large production experience and high current productivity.

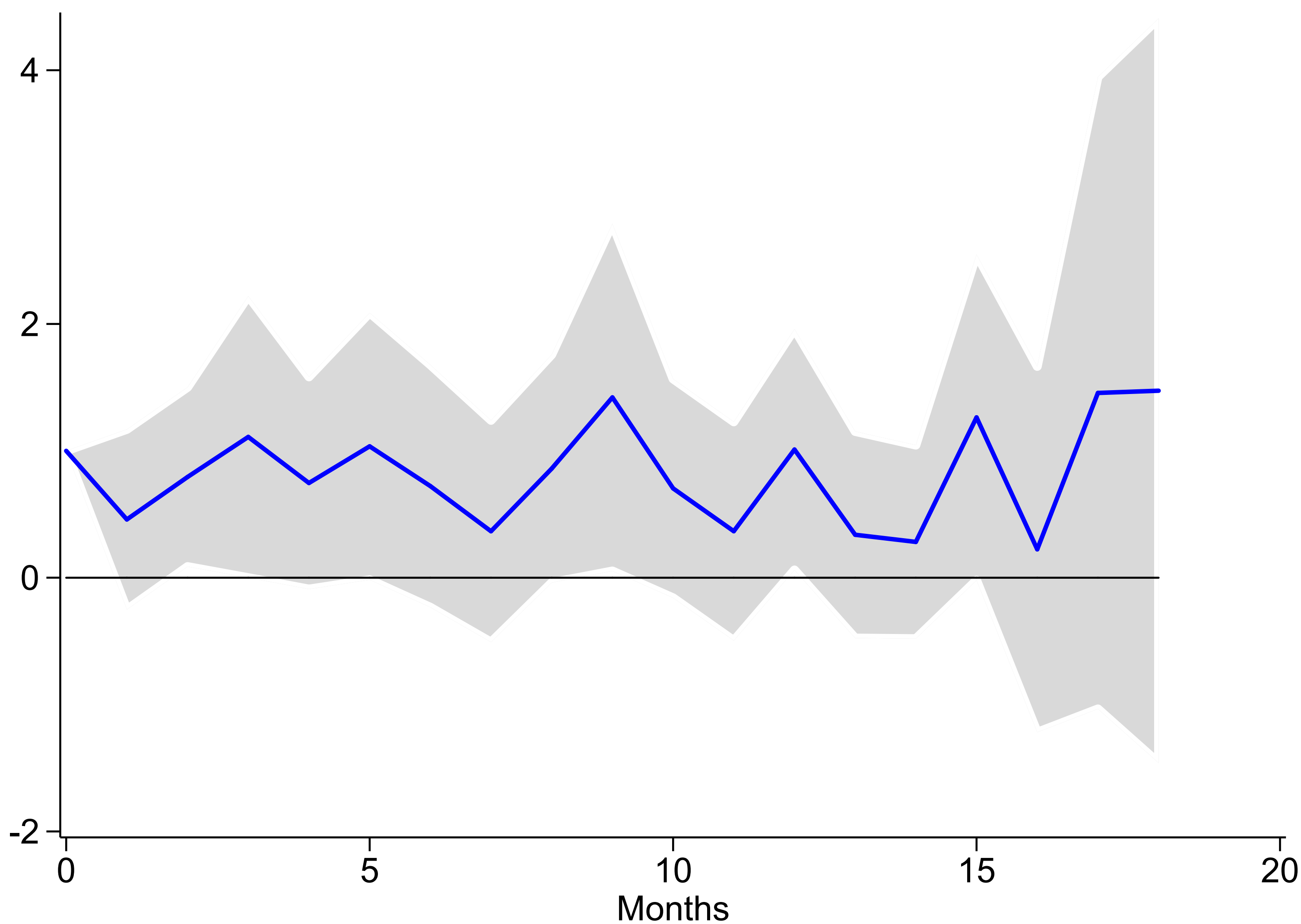
A Appendix Figures & Tables

Figure A.1: Autocorrelation in Ouptut And Productivity



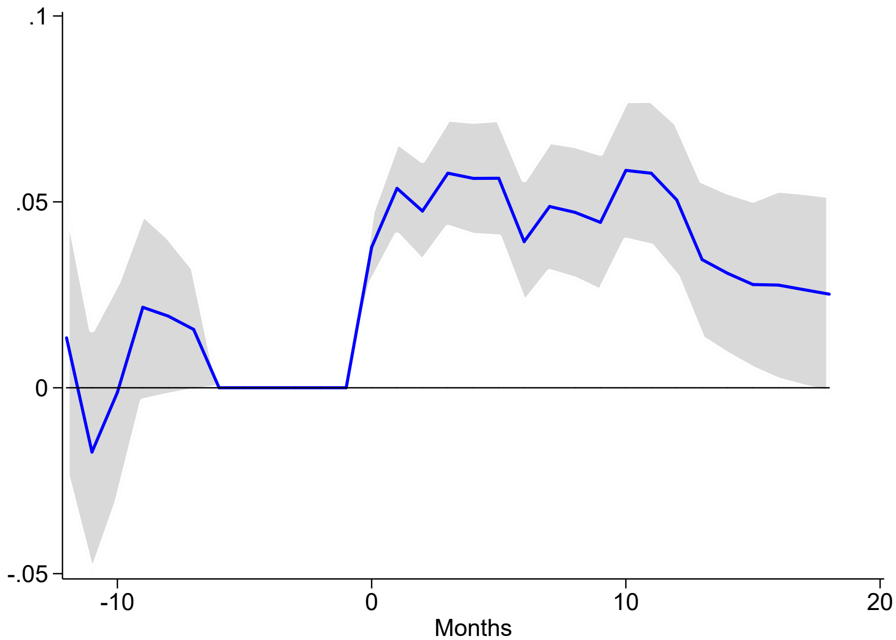
Note: The figure shows scatter plots at the plant-by-aircraft model level. The left hand panel shows (log) cumulative production up to May 1944 against (log) production in the single month of May 1945, representing the allied victory in Europe. The strong correlation demonstrates the strong autocorrelation of production. It is difficult to disentangle the effects current output (scale effects) and cumulative output (experience effects). The right hand plane shows (log) cumulative production up to May 1944 against (log) aircraft per hour worked in January 1944. The strong correlation suggests reverse causation from past productivity to cumulative output. This confounds estimate of the effects of productivity on experience, because productivity itself is strongly autocorrelated.

Figure A.2: Response of Output to a 1% Shock to Aircraft Demand

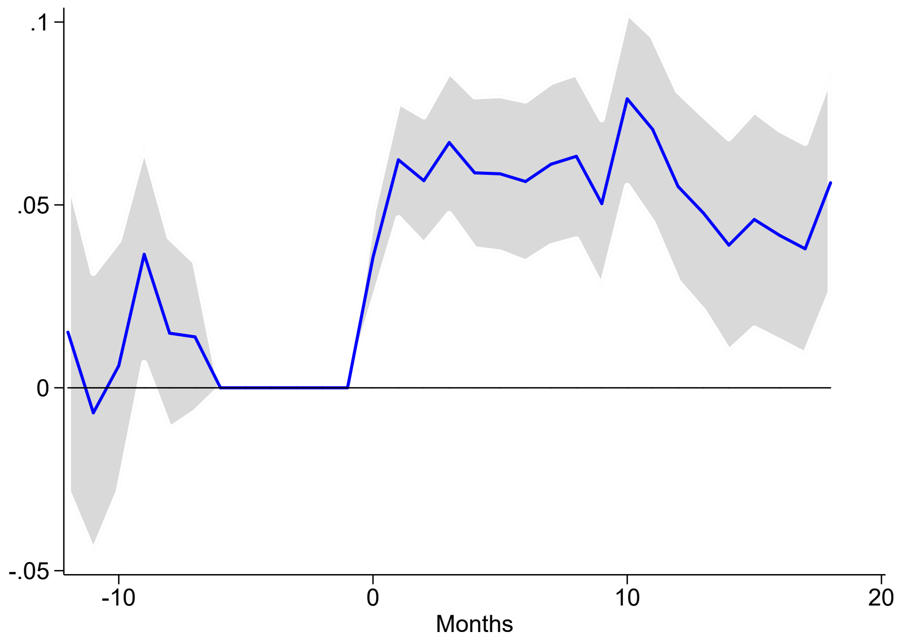


Note: The figure shows the response of TFP to a one percent shock to aircraft demand, controlling for capital utilization. The shaded area shows 95% confidence intervals. Estimates are based on local projections, with aircraft demand instrumented with the instrument described in Section 3, and laid out in equations (1) and (2).

Figure A.3: (No) Pre-trends in Labor Productivity and TFP



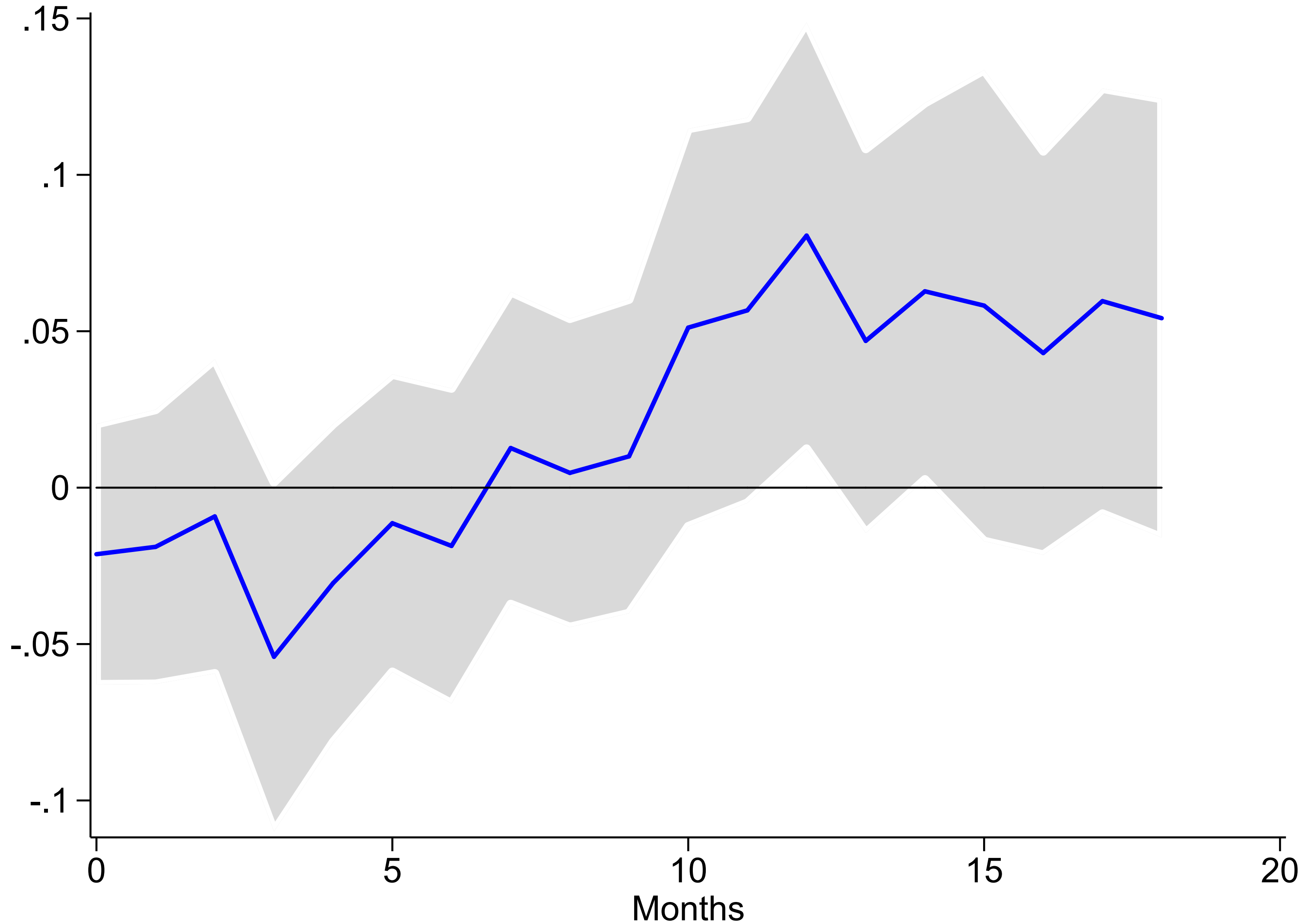
(a) Output per Hour Worked



(b) TFP

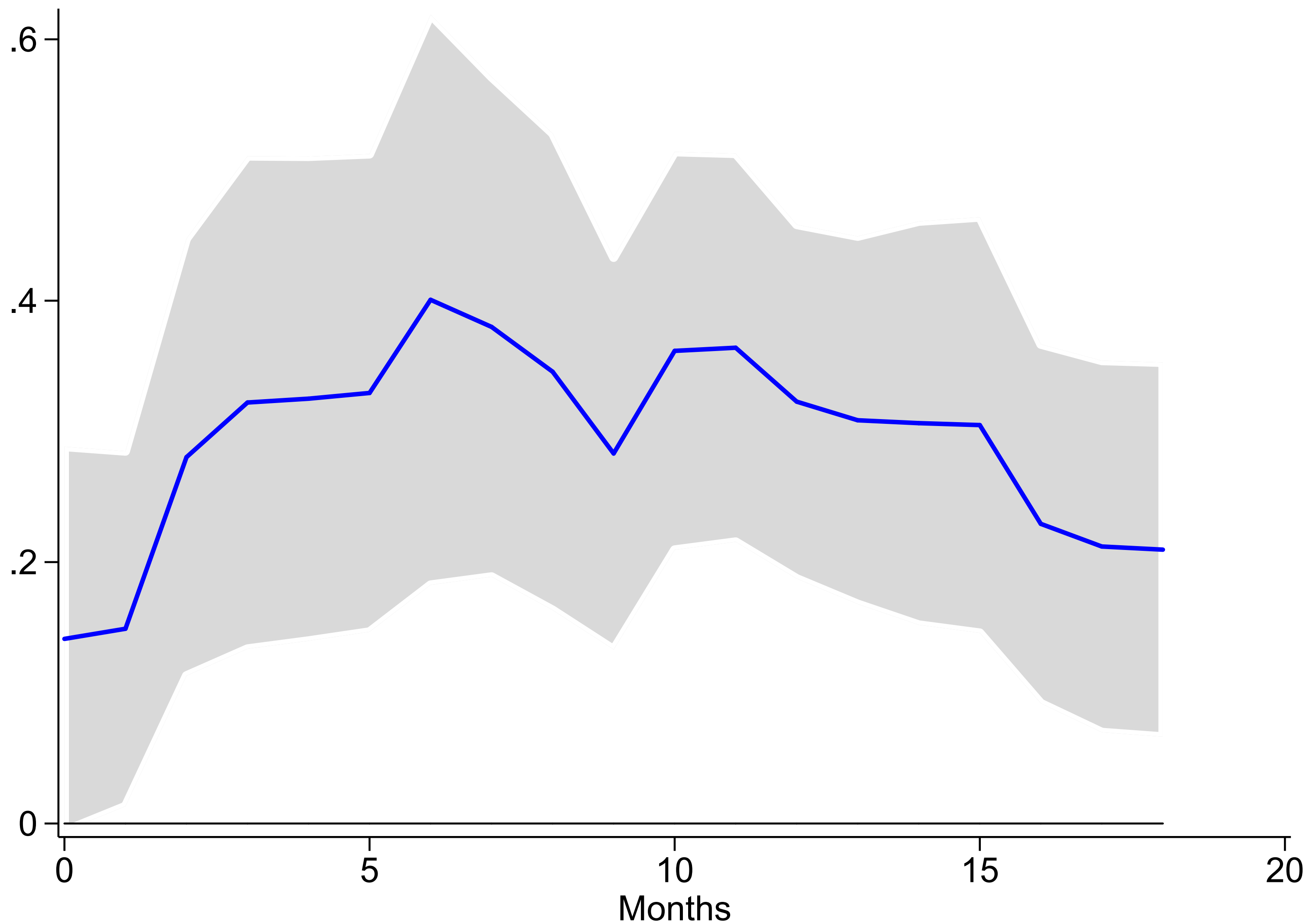
Note: The panels show responses of (a) labor productivity and (b) TFP to 1% shocks to aircraft demand at month zero. The shaded area shows 95% confidence intervals. Estimates are based on OLS local projections, as in (2). Observations in "negative" months are before the shock to demand and show pre-trends showing relative productivity in production lines receiving the shock to receive demand at time zero.

Figure A.4: Response of Capital Utilization to a 1% Shock to Aircraft Demand



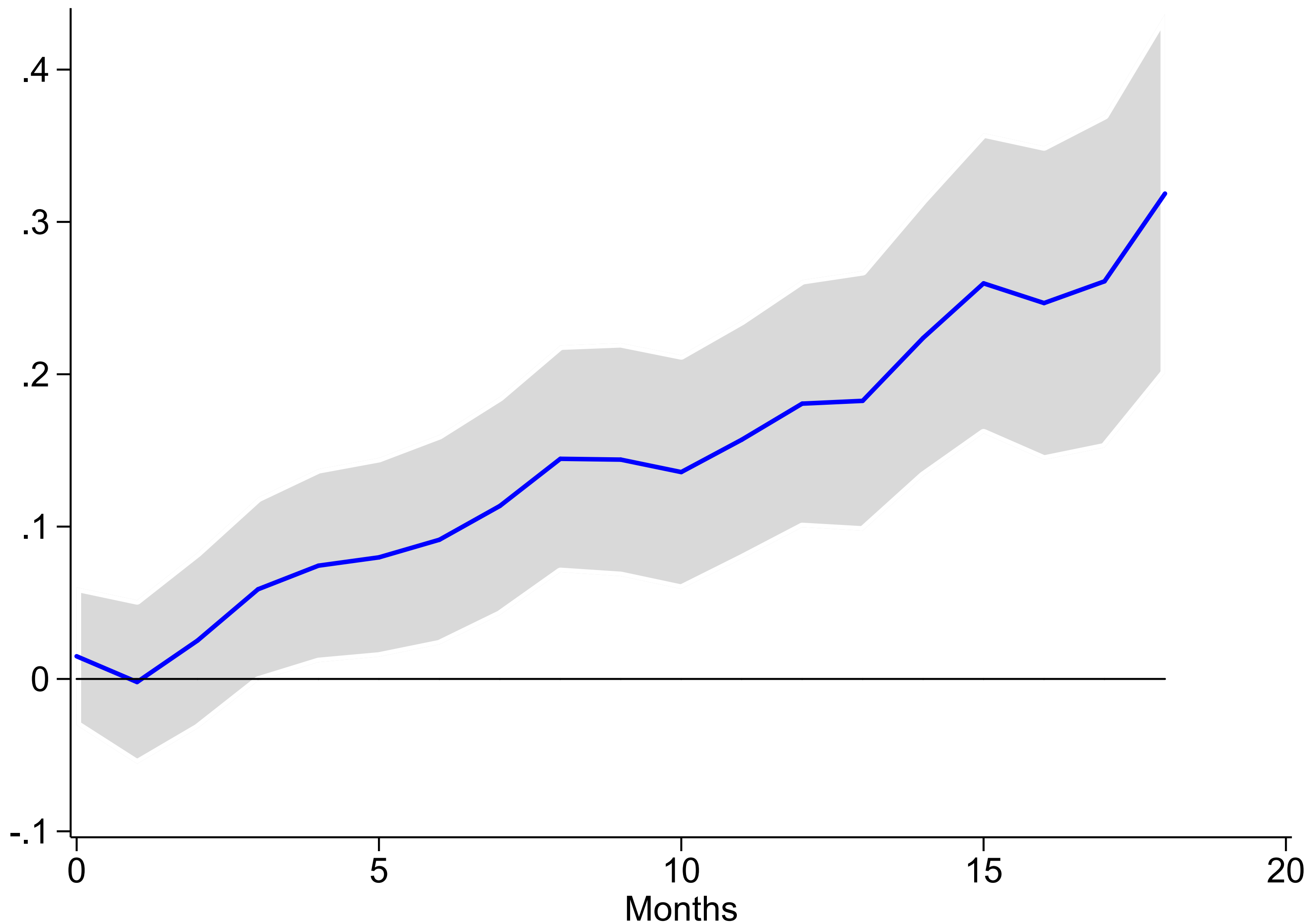
Note: The figure shows the response of capital (shift) utilization to a one percent shock to aircraft demand. The shaded area shows 95% confidence intervals. Estimates are based on local projections, with aircraft demand instrumented with the instrument described in Section 3, and laid out in equations (1) and (2). First stage F-statistic = 13, sufficient to reject a bias due to weak instruments greater than 10% with 95% confidence.

Figure A.5: Response of Capital-Utilization Adjusted TFP to a 1% Shock to Aircraft Demand



Note: The figure shows the response of TFP to a one percent shock to aircraft demand, controlling for capital utilization. The shaded area shows 95% confidence intervals. Estimates are based on local projections, with aircraft demand instrumented with the instrument described in Section 3, and laid out in equations (1) and (2). First stage F-statistic = 34.

Figure A.6: Differential Response of TFP to a 1% Shock to Aircraft Demand: Controlling for Plant Age



Note: The figure shows the response of TFP to a one percent shock to aircraft demand predicted by the instrument described in Section 3 (reduced form) interacted with a dummy variable equaling one if the plant had above average initial capital utilization. The specification includes controls for plant age and the interaction between aircraft demand and a dummy equaling one if the plant was above average age. The shaded area shows 95% confidence intervals. Estimates are based on local projections.

Table A1: Correlation Between Measures of Aircraft Plants' Capacity Constraints

	Capital utilization	Hours per worker	Wages	Labor market priority
Capital utilization	1			
Hours per worker	0.47***	1		
Wages	0.11	0.11	1	
Labor market priority	0.29*	0.10	0.42***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Summary Statistics of Airframe Plants by Capacity Constraint Measures

	Capital Utilization		Hours/Worker		Wages		WMC	
	Low	High	Low	High	Low	High	2-4	1
$\Delta_{\%}$ Output per Worker	127	104	100	115	117	103	111	108
Firm Age (Months)	175	195	183	189	181	193	184	188
Plant Age (Months)	60	139***	123	96	106	108	88	132*
Hours per Pound	4.6	3.1	3.6	3.4	3.7	3.3	4.3	2.8
Airplanes Produced	44	81	85	60	82	59	63	75
Unit Cost (000's $)	113	111	94	130	93	129	92	126
Wing Span (Meters)	21.4	20.1	20.5	20.5	20.9	20.1	21.1	20.1
Public Plant Financing (mln $)	8.2	10.4	7.7	10.9	12.2	7.0*	9.8	9.3

Note: Summary statistics for airframe plants along sample splits reflecting different dimensions of capacity constraints. These are (1) capital utilization as measured by shift utilization, (2) weekly hours per worker, (3) county-level wages, and (4) War Manpower Commission local labor market classification (1 to 4, decreasing in labor shortages). "High" columns give averages for plants above median by the metric in quesion in January 1943. Averages are for January 1943, except for plant financing (cumulative to January 1945), and growth in aircraft per worker (log change from Jan 1943 to Jan 1943). Asterisks reflect statistical significance of the t-test of differences between the two categories: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Sources: AMPR War Production Board *War Manufacturing Facilities Authorized*, and the author.