4-2019

Shocking Technology: What Happens When Firms Make Large IT Investments?

James Bessen
Boston University School of Law

Cesare Righi

SHOCKING TECHNOLOGY: WHAT HAPPENS WHEN FIRMS MAKE LARGE IT INVESTMENTS?

Boston University School of Law
Law & Economics Series Paper No. 19-6

April 2019

James Bessen
Boston University School of Law

Cesare Righi
Boston University School of Management

Shocking Technology:
What happens when firms make large IT investments?

James Bessen and Cesare Righi
Technology & Policy Research Initiative, Boston University School of Law

12 April 2019

Abstract: Many economists see information technology (IT) as central to understanding trends in productivity, labor's share of output, and employment, especially as new "artificial intelligence" (AI) technologies emerge. Yet it has been difficult to measure its effects. This paper takes a first look at the economic impacts of large custom software investment by firms—"IT shocks." Using a novel difference-in-differences methodology, we estimate the productivity of these shocks and the associated effects on revenues and employment and we explore the implications in terms of labor's share and other variables, including heterogeneous relationships by industry, AI use, and time. In our preferred models, IT shocks increase firm productivity by about 5%, followed by increases in revenue of 11% and in employment of 7% on average. However, employment growth following IT shocks is small or negative in mature industries; also, it has been slower in recent years, reducing job reallocation and aggregate productivity growth. Also, labor's share of revenue decreases and operating profits rise following IT shocks.

JEL Codes: D22, J21, O33

Keywords: information technology, artificial intelligence, firm productivity, employment growth, firm growth, labor share, markups

Thanks for helpful comments from Seth Benzell, Richard Freeman, Sina Khoshsokhan, Michael Meurer, Filippo Mezzanotti, Pascual Restrepo, Daniel Rock, Tim Simcoe and participants at the TPRI seminar and the Harvard SBBI seminar. Thanks to Tori Smith for research assistance and special thanks to Ke-wei Huang for sharing data. This work was supported by the Ewing Marion Kauffman Foundation, Google.org, and Schmidt Futures. The contents of this publication are solely the responsibility of the authors.

Introduction

Many economists see information technology (IT) at the crux of several major economic issues including productivity growth, decreasing labor's share of output, and the potentially negative impact of technology on employment. New information technologies, such as Big Data and machine learning, are altering many production activities, enabling computers to interpret X-rays to diagnose disease, select job applicants, handle routine customer support phone calls, and drive cars. Some economists worry about the effects of new IT on wages, employment, and inequality (Brynjolfsson and McAfee 2014), some fearing mass unemployment (Frey and Osborne 2017). Others see new information technology implicit in labor's declining share of output (Karabarbounis and Neimann 2014, Autor et al. 2017 Acemoglu and Restrepo 2018, Calligaris et al. 2018). And many economists see IT as central to ongoing productivity growth but they differ widely as to whether the positive impact of IT has waned (Gordon 2017) or is about to start a resurgence (Brynjolfsson, Rock, and Syverson 2019).

There is a sizeable empirical literature on IT (see Brynjolfsson and Yang 1996, Kretschmer 2012, and Stiroh 2005 for reviews). However, this literature is limited in several ways. First, it focuses mainly on the impact of IT on productivity and little on some of the other issues just noted.[1] Second, much of this literature measures IT as hardware, either computers and peripherals or also other information processing equipment and

[1] An important strand of the literature studies the relationship of IT and complementary investments in organizational capital, human capital, and management including Bartel, Ichniowski, and Shaw 2007; Bessen 2015; Bloom et al. 2012; Bloom et al. 2014; Bloom et al. 2017; Bresnahan, Brynjolfsson, and Hitt 2002; Bresnahan and Greenstein 1996; Brynjolfsson et al. 2008; Caroli and van Reenen 2001; and Crespi, Criscuolo, and Haskel 2017. Akerman 2015 and Gaggl and Wright (2018) explore a variety of outcome variables beyond productivity.

communications equipment.[2] Third, in line with the focus on productivity measurement, this literature treats IT as an input to production rather than as investment in innovation.

These limitations are important because the nature of IT has been changing. IT is now dominated by firm-specific custom software rather than hardware or pre-packaged software. While firm-specific software accounted for 33% of IT investment in 1985, it accounts for 55% today.[3] In 2016, the Bureau of Economic Analysis (BEA) estimates that private investment in proprietary software—both self-developed and custom contracted—was $250.4 billion, almost as much as net capital investment. Not only is this investment large, it is also qualitatively different from routine investments in IT inputs. There are, of course, different types of IT investments. Installing routine word processing software on personal computers is an example of IT as an input factor. But when a firm builds a custom logistics system with proprietary features that rivals do not have, that is different because it is innovation. Innovative investments can, for instance, let the firm earn quasi-rents, increasing markups and decreasing labor's share of output. Investment in custom software is largely investment in innovation (otherwise the software could be purchased) and, as we develop below, this implies different outcomes and different ways of measurement than investments in routine inputs.

This paper takes a first look at the economic impacts of large custom software investment by firms on firm productivity, revenue and employment, and also explores the implications of these effects on quantities such as labor's share of revenue and markups. We do this by focusing on events where firms make major investments in developing new

[2] Some important exceptions include Bloom et al. 2012, Harrigan et al. 2016, Jin and McElheran 2018, and Tambe and Hitt 2012.

[3] This counts custom contracted and own-account (self-developed) software as a share of total gross investment in software, computers, and peripherals.

software, a kind of "technology shock." We identify these events using online resume data to detect large increases in the share of software developers in each firm's workforce. Firm financial data from Compustat allow us to relate these events to changes in firm outcomes in a difference-in-differences model bolstered by a control function to capture unobserved productivity and demand shocks.

We make several contributions to this literature. First, we develop a new way to measure major IT investments and adapt a novel methodology from Bessen et al. (2019) to estimate impacts. Second, we are able to explore not only impacts on productivity, but also associated effects on firm revenue and employment, and study the association of these shocks with labor's share of revenue and profits. Third, we perform this analysis along multiple dimensions to understand heterogeneous impacts. These include whether or not the firm uses AI/Big Data, industry, geographical location, firm age and calendar time of the shocks. Together, this analysis provides a rich picture of what IT does and where it does it with implications for a range of policy issues.

We find that these technology shocks are followed on average by increases in productivity, sales, and—contrary to alarmist predictions—employment. These increases are economically significant (respectively, 5, 11 and 7 percent) and we provide a variety of estimates that support a causal interpretation of our main results. However, the impact of these shocks on employment depends on the industry; it is strong in technologically new industries but weak or negative in mature ones. Moreover, we find that the productivity impact of IT shocks has not declined in recent years, however, employment has not grown as much following these shocks. This means that IT shocks contribute less to *aggregate* productivity growth now than they did in the past as argued by Decker et al. (2018). That is, IT shocks do not contribute to the correlation between productivity and firm size as much as they did in the past. Finally, major IT investments are followed by a decrease in labor's share of firm revenue and an increase in firm operating margins in some industries.

The paper begins by discussing the nature of custom software and the significance of IT shocks. A simple model outlines the relationship of these shocks to a variety of outcome variables. We then describe the construction of our data. The last two sections summarize our empirical findings and conclude.

Background and Theory

IT shocks

A large literature has looked at the productivity contribution of IT both at the aggregate level (see for example, Oliner and Sichel 2000, Jorgenson, Ho, and Stiroh 2005) and at the firm level (see for example, Brynjolfsson and Hitt 1996, 2003). This body of research uses two main methods of obtaining estimates of the productivity contribution of IT: non-parametric growth accounting estimates of Solow residuals or parametric estimates of production functions. In both methods, IT enters as an input factor of production. IT investments in innovation only appear tangentially, as embodied technical change reflected in the quality-adjusted price of IT goods or as part of IT's contribution to disembodied technical change.

However, the large investments that firms are making in developing their own software suggest a number of important differences in the role of IT that affect a broader range of outcomes and also affect how we measure the impact of IT. First, innovations can generate quasi-rents. To the extent that new custom software lowers prices or increases product quality in a way that rival firms cannot imitate, IT investment should lead to quasi-rents that increase the firm's markup of price over cost. Below we will explore whether major IT investments tend to be followed by increases in operating margins and decreases in labor's share of output, both of which might follow from increased markups.

Second, innovation can change the production function. For example, Acemoglu and Restrepo (2018) emphasize how automation of tasks—increasingly done by information technology—can change labor's share of output.

Third, the impacts of innovative IT investment are likely much more heterogeneous than those of routine investments. One reason for heterogeneous outcomes is wide variation in the technological opportunities facing different industries and firms. Another is that firms have disparate capabilities in terms complementary assets needed for innovation such as software talent, managerial capability, and organizations. A significant literature has looked at how such complementary assets affect the returns to IT, however, the connection to the innovative role of IT is not always drawn.[4]

Finally, innovation tends to be risky in general and this is particularly true for IT innovation. Indeed, large IT projects have notoriously high rates of failure.[5] The high uncertainty, combined with some other characteristics of software investment, tend to make much IT investment "lumpy," that is, occurring in discrete episodes of high investment. It is well-established that capital investment tends to be lumpy (Haltiwanger et al. 1999; Doms and Dunne 1998; Nilsen and Schiantarelli 2003). In theory, high uncertainty gives rise to lumpy behavior when the investment is irreversible and when there are indivisibilities or

---

[4] See footnote 1 for some citations to this literature.

[5] Michael Bloch, Sven Blumberg, and Jürgen Laartz, "Delivering large-scale IT projects on time, on budget, and on value," Digital McKinsey, October 2012, a study of 5,400 IT projects > $15 million found failures so bad in 17% of the projects that they threatened the existence of the company; on average costs run 45% higher and 56% less value than planned. Lars Mieritz, "Survey Shows Why Projects Fail," Gartner, June 2012 in survey of 154 organizations, North America and Europe found failure rates of 28% for projects > $1 million; 20% for projects smaller than $350k.

nonconvex adjustment costs (Pindyck 1991, Rothschild 1971). Investments in custom software are typically irreversible; they cannot be resold because they are firm-specific. And large software systems have large, indivisible fixed costs. Also, associated organizational changes may have large adjustment costs.

All of this suggests that substantial innovative IT investment may occur in discrete episodes. Below we define "IT spikes" and find that a large share of IT investment occurs in these episodes. Indeed, custom software development appears to be substantially lumpier than capital expenditure. Using our definition, 47% of the total increase in the employment of software developers occurred during IT spikes, yet these spikes accounted for only 12% of firm-year observations in our data.

These episodes appear to reflect specific technological and market opportunities of the firm. Consider three examples from 2007. After a decade of rapid growth, Ebay invested $89 million in software development staff and consultants to enhance user experience and add new products. That year Danske Bank, Denmark's largest retail bank, also hired a large number of software developers. Danske Bank had developed an effective IT banking platform, but on acquiring a group of banks in Baltic countries, they needed to adapt their platform and integrate existing systems from this group. In 2007 also, the aerospace division of Crane Company identified a market opportunity to sell rugged mobile computers to military suppliers. They hired software developers and also acquired a business unit of an embedded computer firm in order to rapidly put together a product and bring it to market.

These examples show that the occurrence of spikes may often be tied to particular technological opportunities such as Crane Co.'s opportunity to deliver a new kind of computer to the military market or Danske Bank's opportunity to enter new markets or Ebay's opportunity to expand its market with better service. The opportunities were mostly specific to the particular firm. In addition, exploiting these opportunities depended on complementary firm assets such as Ebay's large customer base, Danske Bank's existing platform, and Crane Co.'s technical capabilities including those in the acquired business unit. In short, the occurrence of these IT shocks is substantially idiosyncratic to the firm. Moreover, the IT shock should be viewed not simply as a pure investment in software, but an investment in complementary goods as well.

Because such major investments in innovation often occur in discrete episodes, we can exploit an empirical approach developed by Bessen et al. (2019) that uses these events to conduct difference-in-differences analyses. This approach allows us to isolate the relationship between large IT investments and subsequent performance of a variety of outcome variables. While an association between the timing of these events and changes in firm behavior over time is not, by itself, sufficient to establish causality, we use some standard tools of production function estimation to strengthen the plausibility of our estimates and support a causal interpretation of our empirical models.

A simple model

It is helpful to sketch a simple model and define the outcome variables that we will explore. Assume that firm *i* creates output *Y* at time *t* according to a Cobb-Douglas production function,

(1)

$$Y_{it} = A_{it} \cdot L_{it}^{\beta} \cdot K_{it}^{1-\beta}$$

where *A* is productivity, *L* is employment, and *K* is capital. Assume also that firms have some market power so that price, *p*, can be expressed by an inverse demand function,

(2)

$$p_{it} = d_{it} \cdot Y_{it}^{-\frac{1}{\varepsilon}}, \quad R_{it} = p_{it} Y_{it} = d_{it} \cdot Y_{it}^{1-\frac{1}{\varepsilon}},$$

where ε is the elasticity of demand, *d* is a firm-specific demand shifter, and *R* is revenue.

The firm seeks to maximize profits,

$$\pi_{it} = R_{it} - w_t L_{it} - r_t K_{it}$$

where *w* is the wage and *r* is the capital rental rate. The first order conditions are

(3)

$$\frac{\partial Y_{it}}{\partial L_{it}} = \frac{\beta Y_{it}}{L_{it}} = \frac{w\mu}{p_{it}}, \quad \frac{\partial Y_{it}}{\partial K_{it}} = \frac{(1-\beta) Y_{it}}{K_{it}} = \frac{r\mu}{p_{it}}, \quad \mu \equiv \frac{\varepsilon}{\varepsilon - 1}.$$

These conditions apply in equilibrium, however, firms facing adjustment costs might not reach long-term equilibrium immediately.

Several useful expressions can now be derived. First, since we observe firm revenue, but not actual output, it is helpful to define a log revenue production function

(4)

$$\ln R_{it} = \frac{1}{\mu} [\ln A_{it} + \beta \ln L_{it} + (1-\beta) \ln K_{it}] + \ln d_{it}.$$

We will estimate a version of this equation below, specifying *A* as a function of the IT shock.

Second, unit cost, *c*, can be defined

(5)

$$c_{it} \equiv \frac{wL_{it} + rK_{it}}{Y_{it}} = \frac{c_0}{A_{it}},$$

where the latter expression is derived from the first order conditions and applies in equilibrium.[6] Then, using Euler's theorem and the first order conditions (3), it is straightforward to show that the price markup is,

(6)

$$\frac{p_{it}}{c_{it}} = \mu.$$

One concern is that IT may create market power, increasing the markup. This might occur, for instance, if a firm earned quasi-rents from its innovative software. Below we will want to explore whether markups change, but since we do not observe unit costs, we need to measure closely related variables. Three related quantities are labor's share of revenue, *s,* operating margins, *m*, and the capital labor ratio, respectively,

(7)

$$s \equiv \frac{w_t L_{it}}{p_{it} Y_{it}} = \frac{\beta}{\mu}, \qquad m \equiv \frac{R_{it} - c_{it} Y_{it}}{R_{it}} = 1 - \frac{1}{\mu}, \qquad \frac{K_{it}}{L_{it}} = \frac{1-\beta}{\beta} \cdot \frac{w_t}{r_t}.$$

Labor's share of revenue is also interesting because Acemoglu and Restrepo (2018) propose that automation of tasks can decrease labor's share, decreasing β in our model. On the other hand, Autor et al. (2017) find that the changes in labor's share of output in the aggregate are largely a result of reallocation of employment between firms, not within-firm changes. Our data may provide some indication whether IT shocks cause within-firm decreases in labor's share of output. Another issue has been the extent to which rising markups and firm profits have been caused by lax competition policy as opposed to changes in technology (Barkai 2016, De Loecker and Eeckhout 2017). Bessen (2017) finds some evidence that information

[6] $c_0 = w\left(\frac{\beta r}{w(1-\beta)}\right)^{1-\beta} + r\left(\frac{\beta r}{w(1-\beta)}\right)^{-\beta}$.

technology is associated with higher operating margins. Below we will see whether IT shocks cause measurable changes in operating margins as well.

Also, using (2), (3), and (6), revenue and optimal labor demand can be written,

(8)

$$\ln R_{it} = (\varepsilon - 1)(\ln A_{it} - \ln \mu - \ln c_0) + \ln d_{it}.$$

$$\ln L_{it} = \ln \frac{\beta R_{it}}{w_t \mu} = (\varepsilon - 1)(\ln A_{it} - \ln \mu - \ln c_0) + \ln d_{it} + \ln \frac{\beta}{w_t \mu}.$$

This latter equation is important because it shows that the impact of productivity on employment depends on the firm's demand elasticity with respect to price. This allows us to explore whether the predictions of major job losses associated with artificial intelligence/machine learning appear to be occurring. Bessen (2019) argues that demand elasticity has historically declined in industries subject to heavy productivity-enhancing technological change. As a result, firms in manufacturing industries, which have been subject to extensive automation and rapid technical change for a long period of time, are likely to have lower demand elasticities than firms in service, trade, and finance sectors. This raises the possibility, that we will test below, that industries may differ in response to IT productivity, some growing, others not.[7]

This equation also provides us a way to explore how firm productivity improvements contribute to aggregate productivity growth. That is, aggregate productivity grows when firms improve their productivity and also when more productive firms grow proportionately larger. Decker et al. (2018), analyzing the recent slowdown in job reallocation, finding evidence that positive productivity shocks are occurring just as frequently now as in the past,

[7] Bessen (2019) is concerned with industry-level elasticities of demand; here we are concerned with firm level elasticities, which reflect both industry elasticity as well as the substitutability between firms within the industry. In common models of oligopoly, firm demand elasticities reflect industry demand elasticities.

but that the correlation between firm employment size and productivity is weakening. We will explore below whether IT shocks have changed over the last two decades both regarding their productivity impact and also their impact on firm employment growth.

Data

In this section we describe our sample construction and define the main variables for the empirical analysis.

Sample

Our main sources of data are Compustat and LinkedIn. We retrieve data on firm characteristics such as revenues, capital, total employment, capital investment, market value of equity, wages, industry codes and country of incorporation from Compustat, and use these variables to define operating margins and labor share of revenue.

We convert all variables defined in current dollars to 2009 dollars using deflators from the BEA.[8] We calculate capital as net plant and equipment. We define operating margins as the operating income before depreciation and taxes divided by revenue. To calculate labor's share of revenue, it is necessary to calculate the wage bill. For a fraction of Compustat firms, mostly in the financial sector, staff expense is reported. For the remainder of firms, we impute the wage bill by multiplying the number of firm employees by the industry employee compensation per employee. As long as firm wages move more or less in parallel with industry wages, this imputation will be representative.[9]

[8] We deflate revenues by the industry gross output deflator. We deflate capital and capital investment by the investment deflator. We deflate wages and market value of equity using the deflator of the gross domestic product.

[9] For those observations where we have wage data, the correlation between the actual wage bill and the imputed wage bill is .714.

LinkedIn is our source of data on the composition of firms' workforce. LinkedIn allows users to post their profiles including resumes and they may choose to make their profiles public. Our data were obtained for another project that used Google to search for public profiles on LinkedIn from June to November of 2013 (Ge et al. 2016).[10] Each profile's work experience section reports a series of jobs by date, job title, and employer. For instance, one person might have been a "graphic design intern" at company X from 1998 to 2002, and an "information architect" at company Y from 2002 to 2007.

We limit our sample period to years between 1990 and 2012. LinkedIn data may be less reliable as we go further back in time, and 2012 is the last full year before these data were collected.

We matched firms in Compustat and LinkedIn in a multistep process. First, we used ticker symbols where these were available in the LinkedIn data. Next, for each firm name in Compustat that we could not match at the previous step, we tried to identify all the possible variations in the LinkedIn data, as LinkedIn users may list variations of a company name or provide the name of a subsidiary as their employer. So we cleaned and standardized firm names consistently in the two data sets and used a fuzzy matching algorithm on these names. Then we manually reviewed the fuzzy matches to reduce false positives. This was a burdensome task, so we focused our efforts on large companies as it is more difficult to retrieve additional information for a careful review on smaller organizations. Eventually, we matched 4,262 firms active between 1990 and 2012.[11]

Our match coverage improves over time. The percentage of Compustat firms we can match with LinkedIn firms increase from 25% in 1990 to 54% in 2012. Not surprisingly,

---

[10] We thank Ke-wei Huang for graciously sharing these data with us.

[11] Details on the matching process are available upon request.

matched firms are substantially larger than unmatched firms in terms of sales, employees and capital and other variables related to firm size (see Table A1 in the Appendix). Moreover, the coverage of LinkedIn improves over time, so we have more matches in recent years. As a consequence of the focus on large firms and the increase in coverage over time, the match covers firms that account for 68% of the employees in Compustat in 1990, rising to over 90% of the employees in 2012. Finally, software engineers may be over-represented in LinkedIn, so our matched firms are also more likely to be in IT-related industries. Note that both LinkedIn and Compustat are international, including non-US companies, but both sets are dominated by US firms.

We use resume data from LinkedIn to define our key measure of IT shocks. This measure is based on changes in the IT share of each firm's workforce, that is, changes in the ratio of software developers to total employees. We tally how many LinkedIn profiles report working at a given firm in a given year and to calculate the share of these profiles that are in software development jobs. To do this, we created a list of 1,791 job titles for software development occupations. We included managers such as "information systems project manager;" and we excluded job titles for tech support, maintenance, and basic operations. Identifying software developers in this way, we tabulate the ratio of LinkedIn software developers to LinkedIn total employees for each year for each firm from 1990 through 2012. However, this ratio might not be representative of the total population for employees because the relative usage of LinkedIn by software developers compared to non-IT employees might have changed over time. To correct for changes in coverage, we calculated the total ratio of software occupations to all workers in each year of the Current Population Survey. We use this ratio to weight the firm-year observations so that they correspond to the ratio from the Current Population Survey in aggregate.

We use also use the ratio of IT workers to total employees to calculate the number of non-IT employees, multiplying the number of employees reported in Compustat by one minus the IT share. Throughout the paper, when we refer to labor generally, we mean this measure of non-IT labor. In addition, we use the LinkedIn data to flag companies that use AI or Big Data by identifying a list of job titles associated with these technologies.[12]

For each firm, we keep all the consecutive years with a positive IT share of employees and those for which this is equal to zero but preceded by a year with a positive IT share of employees, so we can compute growth rates. We discard firms for which we can never define a growth rate in IT share of employees and firm-years without data in Compustat, keep only the longest series of observations without gaps for each firm and, if there are ties, we keep the most recent series because the quality of the LinkedIn improves over time.

Table A2 in the Appendix shows summary statistics for firm-years in our matched sample. About 10 percent of the observations are related to firms that use AI or Big Data, the mean growth rate of the IT share of the workforce is 10 percent, mean operating margins are 11 percent, and the mean ratio of employee compensation to revenue is 32 percent.[13] Table A3 shows the distribution by sector.

IT Spikes

Above we proposed that software investment tends to be "lumpy" because it is risky, irreversible, and has large indivisibilities and/or non-convex adjustment costs. In this section

[12] These are "Hadoop" "big data" "quantitative analyst" "data scientist" and "machine learning." We define a time-invariant indicator equal to one for firms that employ at least one person listing these titles in their profile during our sample period.

[13] The last two means are trimmed of the 1% tails.

we define a practical way of identifying these "spikes." We can see evidence of the lumpiness of the IT share growth in Figure 1. This shows a histogram of the year-over-year growth rate in firm IT shares. The smooth line represents a normal distribution. Comparing the histogram to the normal curve shows that the growth rates in the IT share have a heavy upper tail—there are a disproportionate number of events where the growth rate is 40% or above. Moreover, a very large proportion of the firm-years has a growth rate close to zero.

We define a spike as a year when the percentage growth rate in the IT share exceeds 30 percent (for example, the IT share goes from 1.00 percent to 1.35 percent).[14] A way to gauge the lumpiness of IT hiring is to look at the share of hiring that occurs during spike years. Although only 12 percent of the firm-year observations are spike years, these account for 47 percent of the total increase in IT hiring in aggregate over the sample period.

Characteristics of these spikes are explored further in the Appendix. Among the findings: 1) the frequency of spikes grew during the 1990s, but does not exhibit a strong trend since then; nor does the frequency seem to respond consistently to changes in the business cycle (figure A2). 2.) Table A4 shows the frequency of spikes in the sample of matched firms. About one third of the firms do not spike during the sample period. Of those that do spike, about 40 percent spike only once, the remainder spiking more often, up to 7 times. For firms that spike multiple times, our analysis focuses on the spike with the largest growth.[15]

[14] There are, of course, many other ways one could define a spike. We experimented with several of these, including using different thresholds, using an absolute measure of increase rather than a relative measure, using stocks of software developers, and using thresholds relative to a firm average. Generally, results are similar. We report results based on a higher threshold (50%) to define spikes in the robustness checks in the Appendix.

[15] The largest spikes account for 25 percent of the total increase in IT hiring in aggregate over the sample period, although they are only 5 percent of the firm-year observations.

Figure 2 provides further evidence on the lumpiness of spikes. The chart shows the mean and median growth rate of the IT share around the year of the largest spike. On average, there is little growth prior to the spike, a sharp and discrete increase with the spike, and little growth afterwards. The spike typically represents a permanent increase in the firm's IT share of the workforce.[16]

Firms that spike tend to be smaller than firms that do not spike, both before and after the largest spike (see Table A5). Unreported analysis confirms this using only data for the first year in our sample for each firm. This is related to the definition of spike that we use. Large firms, which aggregate multiple business divisions, are less likely to see a 30 percent growth in IT share for the entire company even if individual business units spike. The difference in size may also be related to differences in the distributions of spikers and non-spikers by industry.

We also analyze the relationship between firm size and the occurrence of spikes in a regression framework. We estimate a set of linear probability models in which we regress an indicator equal to one in the year of a spike (multiplied by 100 to facilitate the interpretation of the coefficients as percentage-point changes) against measures of firm size in the previous year and calendar year effects. For these estimates we use all the spikes, allowing firms to have more than one spell. Spells start either in a firm's second year in the sample (because of our definition of spike and because we use lags of the predictors) or in the year after a spike, and end either in a spike year or at the end of the sample period. We model the baseline hazard of a spike with a set of "age" dummies, where age is defined as the number of years since the start of the spell. We use three measures of firm size (all in natural logarithms):

[16] Our econometric analysis below is robust to excluding firms with multiple spikes.

revenues, non-IT employees and market capitalization. The first three columns of table A6 show that they are all negatively correlated with spikes. However, in the last three columns of that table we also show that the growth of revenue, non-IT employees and market capitalization is not significantly correlated with spikes. These results suggest that while smaller firms are more likely to experience IT spikes, they are not growing more than other firms before the spikes, a result that we will confirm below with other methods.

Finally, Figure 3 shows general trends of revenue and employment around the year of the largest spike. The chart shows medians of these variables across all firms that spike, without any controls. Both quantities exhibit secular trends both before and after the spike, with revenue tending to grow faster. A slight pickup, especially in revenue, occurs immediately after the spike. In the analysis below we will look at these trends, but also control for a variety of considerations.

Empirical Strategy

In this section we describe our empirical strategy to estimate the effect of IT spikes on productivity and other firm outcomes. Our first set of models estimates a simple difference-in-differences production function. Then, we complement these estimates with other difference-in-differences analyses.

Productivity

We begin with the most basic analysis, estimation of the productivity impact of IT spikes using a modified revenue production function (4). Define the revenue productivity $a_{it} \equiv \frac{1}{\mu} \ln A_{it} + \ln d_{it} = \alpha_i + \beta_t + \omega_{it} + \gamma \cdot D_{it}$, where α_i is a firm fixed effect that controls for differences across firms that do not change over time, β_t is a year fixed effect that controls for all factors that change over time and are common to all firms, and ω_{it} is an unobserved time-varying firm-specific shock, capturing both productivity and demand shocks. Let D_{it} represent the contribution of firm IT/software development to productivity. In the literature, this variable is typically captured as a perpetual-inventory stock of past investments in IT. While we will not use such a stock, it is helpful to start with the standard estimation problem from the literature. Inserting this expression for productivity in production function, (4),

(9)

$$\ln R_{it} = \alpha_i + \beta_t + \gamma \cdot D_{it} + \delta \ln L_{it} + \theta \ln K_{it} + \epsilon_{it}, \quad \epsilon_{it} \equiv \omega_{it} + \rho_{it}$$

where ρ_{it} represents a standard error term. Our interest, and that of much of the literature, is to estimate the elasticity of IT, γ.

The standard problem with estimating this equation is that the firm may base its decisions about variable inputs on unobserved productivity shocks, ω_{it}, making these shocks an omitted variable. In particular, an unobserved increase in productivity may be positively associated with higher investments in IT, so $E[\epsilon_{it} \cdot D_{it}] \neq 0$, possibly biasing coefficient estimates upwards. This simultaneity problem was identified by Marschak and Andrews (1944) and a large literature has developed on methods to obtain unbiased production function estimates. Leading techniques include control functions (Olley and Pakes 1996; Levinsohn and Petrin 2003; and Ackerberg, Caves, and Fraser 2015) and dynamic panel data estimates (Arellano and Bond 1991, Blundell and Bond 2000). We will mainly use the control function method of Olley and Pakes (1996), augmented with firm and year fixed effects, with some robustness checks using dynamic panel data estimates (which we do not report for the sake of brevity).[17] Our analysis differs from the standard production function estimation in that we measure the contribution of a discrete increase in IT investments to productivity rather than the relationship between changes in IT stock calculated using the perpetual inventory method.

Specifically, let τ_i be the year that firm *i* spikes (if at all). Then we specify $D_{it} \equiv \mathbf{1}(t \geq \tau_i)$ as an indicator variable that equals one for the year of the spike and subsequent years. This approach has several advantages for our purposes. First, this form corresponds more closely with the discrete phenomenon we are studying, providing better measurement. While major investment episodes are captured in stock measures, their contribution to changes in productivity may be diluted by the use of measures that include past expenditures. Second, we can test whether the occurrence of IT spikes is correlated with other, potentially confounding variables.[18] For example, we find that IT spikes are uncorrelated with prior revenue growth, reducing concerns related to trends in the pre-spike period that may confound our estimates (see Appendix). Third, it avoids heroic assumptions about depreciation. IT stocks are calculated on the assumption of a constant rate of "depreciation." However, software does not actually depreciate; instead, it becomes obsolete based, in part, on the developments of rivals and these developments are likely to affect firm revenue.

Using a discrete form for D_{it}, makes equation (9) a difference-in-differences (DID) specification. As we explore below, this also has advantages in estimating the relationship of IT spikes with revenue and employment. It is important to note that, as we emphasized

[17] Because our data do not include firm-level measures of intermediate inputs exclusive of IT-related costs, we cannot use the similar methods of Levinsohn and Petrin (2003) and Ackerberg, Caves, and Fraser (2015).

[18] Comparable inquiries on investment stocks are not possible without a clear model of how and when investment occurs.

above, firms sometimes make these large investments in response to unique technological and market opportunities that may increase the gains to investments in IT. That is, in the context of this DID specification, the simultaneity problem emerges as a selection bias: the error term may be correlated with the *occurrence* of an IT spike, biasing estimates. We can still use the Olley-Pakes control function approach to correct for this selection bias. Olley and Pakes (1996) take advantage of the notion that a firm facing a positive shock will increase capital investment; that is, investment is correlated with the unobserved productivity shock. Following their approach, we make an assumption about timing: firms will adjust labor, including software development labor, in the same year as the shock; they will also invest in capital, but the effective capital stock adjusts more slowly. Under some general conditions (see Pakes 1994), investment can be written as an increasing function of revenue productivity and capital, $I_{it} = h(a_{it}, K_{it})$ in the region where investment is positive. If productivity is the only scalar unobservable and if the investment demand function is strictly monotonically increasing, then this function can be inverted, $a_{it} = h^{-1}(I_{it}, K_{it})$. This is a "control function" that reflects the disturbances to productivity that are observed by the firm but not measured by the econometrician. We approximate this control function using a third order polynomial in log investment and log capital, $f(\ln I_{it}, \ln K_{it})$, such that $a_{it} = f(\ln I_{it}, \ln K_{it}) + \eta_{it}$. Then we can estimate

(10)

$$\ln R_{it} = \alpha_i + \beta_t + \gamma \cdot D_{it} + \delta \ln L_{it} + f(\ln I_{it}, \ln K_{it}) + \epsilon_{it}.$$

This equation is the first stage of an Olley-Pakes estimation procedure and can be estimated with or without the fixed effects. The control function should capture the contribution of the unobserved shocks, reducing the concerns related to possible bias in γ.[19] Equation (10) introduces a slight wrinkle to the typical Olley-Pakes estimation in that productivity here is only *partially* unobserved. That is, we include the IT spike term, $\gamma \cdot D_{it}$, explicitly in the regression even though it is implicitly captured by the control function as well. We can show this does not bias the estimates of γ, although it does potentially create multicollinearity.[20] But the precision of our estimates below, as well as formal tests for multicollinearity, show that this is not a problem for our analysis. We also can estimate γ in a two-step procedure where we first estimated (10) without the $\gamma \cdot D_{it}$ term, and then regress the first stage productivity estimates against $\gamma \cdot D_{it}$ and a set of fixed effects.

In addition, we estimate a more general specification, one that allows us to estimate a variable effect before and after the spike:

(11)

$$\ln R_{it} = \alpha_i + \beta_t + \sum_{k \neq -1} \gamma_k \cdot \mathbf{1}(k = t - \tau_i) + \delta \ln L_{it} + \theta \ln K_{it} + \epsilon_{it}.$$

In this equation, we substitute the post-spike indicator with a set of year-relative-to-spike dummies, omitting the indicator for the year before the spike. This specification allows us to measure how the effect of IT spikes changes over time and to examine the trend in productivity before the spikes both graphically and formally. While the lack of significant differences in pre-spike trends in productivity between spikers and other firms does not rule

[19] Under Olley and Pakes's assumptions, it should also produce unbiased estimates of δ. Ackerberg, Caves, and Fraser (2015) argue for alternative assumptions and an alternative estimation procedure, however, this is not our focus.

[20] If the control function *exactly* captured productivity, then the regression routine would have to drop the $\gamma \cdot D_{it}$ term. But as long as η_{it} has non-negligible variation, we can show that the estimates of γ are unbiased. Calculations available from authors.

out all the endogeneity concerns described above, failing to find substantial differences in the trends would reduce the concerns related to pre-spike confounding factors.

Revenue and Employment

When a firm experiences a positive productivity shock, two effects determine the revenue it produces: 1) it is able to produce more with given quantities of labor and capital and 2) it reduces prices in response to lower costs; lower prices increase optimal revenue, so the firm also scales up the amount of labor and capital it employs. Equations (4) and (9) capture the first effect, accounting for revenue relative to given levels of labor and capital. Equation (8a) describes the combined response of revenue to both greater productivity and greater demand. We can derive a difference-in-differences version of this equation using the equivalent procedure as with the productivity equation, that is, let $(\varepsilon - 1)\ln A_{it} + \ln d_{it} = \alpha_i + \beta_t + \omega_{it} + \gamma \cdot D_{it}$ and let $\epsilon_{it} \equiv \omega_{it} + \rho_{it}$. Then we have a general difference-in-differences specification,

(12)

$$\ln R_{it} = \alpha_i + \beta_t + \gamma \cdot D_{it} + \epsilon_{it}.$$

In this DID specification we do not include labor and capital as predictors of revenue, as our interest here is in estimating the total effect of the IT shock on revenue and not the effect on productivity. As above, we rely on firm fixed effects and year fixed effects to capture time-invariant unobserved heterogeneity across firms and all time-varying factors that are common to all firms. Nevertheless, this DID specification has the same potential concerns about endogeneity related to unobservable productivity and demand shocks as does equation (9). However, if we accept that $E[\epsilon_{it} \cdot D_{it}] \approx 0$ for equation (9), then the comparable simultaneity biases in the estimates of γ in (12) should be small as well.

Equation (8b) provides a comparable expression for firm non-IT labor, $\ln L_{it}$. The equivalent DID specification for this outcome variable takes the same form as (12) with labor as the outcome variable. And estimates of γ from this equation, too, should be likely unbiased, assuming the occurrence of the spike is uncorrelated with the error term.

As with the production function, we also estimate models that allow for the effect of the IT spike to vary by year relative to the spike, and test for the existence of pre-spike trends. The maintained assumption for these models is that, conditional on firm and year effects, the trends in revenues and labor of the spikers and the other firms would be parallel in absence of the IT spike. We will provide evidence supporting this assumption at least for the pre-spike period below.

Results

In this section we present the results of our empirical analysis. We begin showing the estimates for the production function. Then, we provide the estimates of the effects of IT shocks on revenues and employment. We also estimate heterogeneous effects of the IT shocks and finally explore the implications for operating margins, labor share of revenues and capital labor ratio.

Estimates for the production function

Table 1 presents our estimates of the production function based on equation (9). Instead of the firm fixed effects, the specification in column 1 includes a single dummy variable for whether the firm is in the treatment group (a spiker) or not, which captures the "selection effect" into the IT spikes. The estimates show that after the IT spike there is a 9%

increase in productivity, and that spikers are less productive than other firms before the spike.

The model in column (2) drops the spiker dummy and includes a full set of firm fixed effects. The coefficient of the post-spike dummy drops substantially, but it is still statistically significant at 1% level and large in magnitude, implying an increase in productivity after the spike by almost 5%.

Columns (3) and (4) report the estimates for the specifications based on the Olley and Pakes models. Model (3) does not include the firm and year fixed effects and it is estimated applying the full procedure following Olley and Pakes (1996), which allows us to provide an estimate for the coefficient of capital. Here the control function is a third order polynomial in log investment and log capital and the estimates also take into account the probability of firm exit.

For simplicity, we estimate only the first stage of the Olley and Pakes (1996) model in specification (4), which also includes fixed effects for firms and years, as well as the polynomial of log capital and log investment, so we do not provide the coefficient of capital.[21] The coefficients of the IT spikes are similar to those in column (2). Because the specifications in columns (3) and (4) might give rise to multicollinearity, we measure the variance inflation factor for the first stage of the column (3) specification. The value for γ is 1.01, suggesting that this estimate is not subject to serious multicollinearity. Thus, unobserved shocks do not seem to bias the estimates of productivity gains accompanying IT spikes. Therefore, for the analysis below, we assume $E[\epsilon_{it} \cdot D_{it}] \approx 0$.

[21] The coefficient of the IT spike and its standard error are not affected by this choice, as they are estimated in this first stage.

In Table A4 in the Appendix, we report on a number of robustness checks of these productivity regressions, estimating variations of the model in column (2) of Table 1. First, we estimate a model that substitutes the year effects with year-by-4-digit-SIC-code. These effects should capture all the time-varying shocks that are common to firms in an industry. Second, we allow the coefficients of capital and labor to be different for each SIC 2-digit industry. Third, we estimate a model discarding the control group and exploiting only the timing of the spikes for the subsample of spikers for identification. Fourth, we re-define the spikes using a higher threshold in the growth of the IT share of employment (50% instead of 30%). Fifth, we drop all firms that spike more than once during the sample period. Finally, we drop firm-years characterized by very high or low revenue growth (those in the top and bottom 1% of the distribution of revenue growth). In general, the estimated coefficients are similar across these variations: they are all positive and statistically significant at 1%, and although the point estimates vary, all imply an increase in productivity after the IT spike.

In addition to these robustness checks, we conducted several alternative specifications not reported here:[22] we added the log of advertising and marketing expenditure and the log of research and development expenses in the control function polynomial; we performed a two-step procedure, first estimating a standard Olley-Pakes estimation without the IT term and then regressing the resulting productivity estimates on the IT term and fixed effects; we performed a similar two-step procedure using dynamic panel data methods in the first step (Arellano and Bond 1991). In all of these, the estimated coefficients were similar to those reported in the table. Finally, we also considered a model where an IT spike might change the long term markup. While the equations just estimated assume a constant markup,

[22] And available from the authors on request.

below we find that IT spikes tend to be followed by higher markups. In step one, we estimate the inverse of the markup as a function of the IT shock then insert the prediction of this variable into equation (9). Bootstrapping the errors, this gives us a somewhat larger estimate of the productivity shock of .066 (.017).

In Figure 4 we provide a graphical representation of the trends in productivity before and after the IT spikes. The figure plots the coefficients of a set of year-relative-to-spike dummies from our estimates of equation (10) in an 11-year window around the spike. Spikers and other firms are on very similar productivity trends before the spike. After the spikes, there is a 5% increase in productivity that remains about constant at least for the first five years. This trend provides further support for a causal interpretation of our estimates of equation (9). Also, this time pattern is distinctly different from the result of Brynjolfsson and Hitt (2003), who find that productivity growth continues to increase up 5 or 7 years after investment in computers. We find that productivity responds quickly with little further adjustment after the first year.[23]

Estimates for revenues and employment

Table 2 shows the difference-in-differences results for revenues and non-IT employment. The model in column (1) is a simplified version of equation (12) in which we

[23] The estimation includes year-relative-to-spike dummies for the entire sample period (from year -22 to 21). However, few observations are available far from the spike time, leading to large standard errors. For this reason, we only display the most central coefficients. We also tested the equality of the pre-spike trends with an F-test of the null hypothesis that the pre-spike coefficients (including those not shown in the figure) are jointly zero. This test cannot reject the null hypothesis (p-value 0.46). We also estimated this general specification using only the subsample of spikers, discarding the control group. Following Borusyak and Jaravel (2017), we omit also the dummy for the earliest year before the spike. The results are very similar: the coefficients for the pre-spike dummies in the 5 years before the spike are very close to zero and their pattern is very similar to the one reported for the full sample, the F-test cannot reject the null hypothesis for the pre-spike coefficients (p-value 0.52), and there is a marked increase in productivity after the spike. These results provide support to the idea that the increase in productivity of the spikers is driving the results, rather than a possible decrease in productivity of the firms that do not spike.

use the spiker dummy defined above instead of the firm fixed effects. The estimates in column (1) imply that after the IT spike there is very large increase in firm revenues. The coefficient of the "selection" indicator captures the difference in firm size before the IT spike, and confirms that the occurrence of IT spikes is more likely for smaller firms.

The model in column (2) is our preferred specification, in which we use firm fixed effects. The coefficient for the post-spike dummy in this model is much lower than in column (1). Nevertheless, we observe an increase in revenues by 11% after the spikes.

Columns (3) and (4) estimate similar models, in which our outcome variable is the natural logarithm of non-IT employment. The results are analogous to those of models (1) and (2): spikers are smaller on average in terms of employment (column (3)), and there is a 7% increase in employment after the spikes (column (4)).

IT shocks are followed by strong revenue growth on average and, far from destroying jobs, employment grows substantially on average. Note however that employment grows more slowly than revenue. This could be because IT might also reduce labor's share of output. We will explore that possibility below.

Table A5 in the Appendix reports the results of robustness checks similar to those we performed for the production function estimates. The results are similar to those reported in the main text.

Figures 5 and 6 plot the coefficients of more general models that allow us to check the pre-spike trends and to observe the dynamics of the effects after the IT shocks. Firms affected by the IT shocks and those that are not have very similar trends in revenues and employment before the IT shocks, and then the spikers experience an increase in these outcomes. This provides additional support for a causal interpretation of the results.[24]

Heterogeneity of the effects

Table 3 and Table 4 explore the heterogeneity of the responses of productivity, revenue, and labor to IT shocks. To explore heterogeneous responses to different values of categorical variable z, we estimate specifications based on

(13)

$$\ln Y_{it} = \alpha_i + \beta_t + \sum_j \gamma_z \cdot \mathbf{1}(t \geq \tau_i \,\&\, z = j) + \pi X_{it} + \epsilon_{it},$$

where Y_{it} is one of our outcomes, and X_{it} contains the natural logarithms of capital and labor for the production function estimates.

Some people argue that artificial intelligence technologies will have a different effect on employment because these technologies are more about replacing human tasks (Ford 2015). Panel A measures the impacts by whether the firm employs AI or Big Data software developers at any point in time in our sample period. We find that the productivity impact of an IT spike is about the same between AI-users and others, even if the coefficient for the latter group of firms is estimated more precisely. This finding complements the deeper look taken by Tambe (2014) who finds that firm returns to investments in Hadoop, a key

[24] Also for these models we tested the equality of the pre-spike trends with an F-test of the null hypothesis that the pre-spike coefficients (including those not shown in the figures) are jointly zero. The test for the trend in revenues cannot reject the null hypothesis (p-value 0.46). The test for the trend in non-IT labor rejects the null hypothesis at 5%. However, this rejection is essentially driven by few coefficients for time periods more than 10 years before the spike, where we have relatively few observations and the results may be driven by outliers. So we do not view this rejection as problematic for a causal interpretation. As we did for the production function, we also estimate these models on the subsample of spikers. Results are very similar to those reported here.

machine learning technology, are higher, but only for firms in geographically concentrated Hadoop clusters (Silicon Valley) that also have made other large IT investments. In contrast to the productivity results, the AI-using firms exhibit much stronger revenue and employment growth.[25] This does not necessarily mean that AI/Big Data causes more rapid growth; it could simply be that more rapidly growing companies are more likely to employ AI. But these estimates are hard to reconcile with the idea that AI is particularly job-destroying in the firms using it.

Panel B looks at differences between industries that use software as part of their products and those that do not.[26] Those that use software as part of their products exhibit higher growth in all three outcome measures on average.[27] But those that develop software for internal use nevertheless show strong growth in each outcome variable.[28]

Panel C looks at changes in response over time. This is important because aggregate productivity grows when firms increase their productivity *and* when more productive firms grow faster. Decker et al. (2018) argue that the pace of job reallocation has declined since 2000, contributing to slowing aggregate productivity growth. They find slowing job reallocation is driven by declining firm responsiveness to productivity shocks. We find supportive evidence regarding IT shocks. The productivity gains from IT shocks are roughly

---

[25] A formal test of the equality of these coefficients rejects the null hypothesis of equal coefficients at least at 5% level.

[26] The former include NAICS 5112, software publishers, 5181, Internet service providers and web search portals, 5182, Data Processing, Hosting, and Related Services, 5191 Other information services, 5415 Computer Systems Design and Related Services, 3341 Computer and peripheral equipment manufacturing, 3342 Communications Equipment Manufacturing, 3344 Semiconductor and Other Electronic Component Manufacturing, and 3345 Navigational, measuring, electromedical, and control instruments manufacturing.

[27] There is significant overlap between firms that produce software-based products and firms that use AI.

[28] Formal tests of the equality of these coefficients cannot reject the null hypothesis at conventional significance levels for the productivity regressions, while it rejects at 10% and 1% respectively for the revenue and labor specifications.

the same before and after 2002. However, the response of employment and revenue declined sharply, and these differences are statistically significant at least at 5% level.

Panel D explores whether "US does IT better," as proposed by Bloom et al. (2012), who argue that better managerial practices at US-based firms generate greater returns to IT investments. While we cannot interpret differences between US and non-US firms in our data as entirely driven by managerial practices, we find that the productivity coefficient is significantly greater for former group (difference statistically significant at 1% level). Interestingly, the coefficient for labor is not much different, either economically or statistically. This might also reflect different managerial capabilities.

Finally, Panel E explores the differential effect of IT for new firms (those publicly listed for 5 years or fewer). Consistent with the view that startups are better able to utilize new technology, we find much larger gains in productivity, revenue, and employment after new firms make major investments in IT.

Table 4 compares several industry sectors. All sectors show an increase in productivity following an IT spike, in many cases highly significant, except for the small "Other" category.[29] But in the other outcome variables, there is significant heterogeneity. In particular, manufacturing industries exhibit much smaller revenue growth than do tertiary sector industries and they exhibit minimal or even negative labor growth compared to strong employment growth for trade, FIRE (Finance, Insurance and Real Estate), and services. Equation (8) provides an explanation: firms in manufacturing industries may have lower price elasticities of demand. Bessen (2019) suggests that demand elasticity in manufacturing industries was initially high but declined as these industries were progressively automated;

[29] The "Other" sector represents only 227 firms, of which 155 experience an IT spike.

tertiary sector industries may have higher demand elasticity in part because they have experienced far less productivity-improving technical change.

Labor's share of revenue, operating profits and capital labor ratio

Economists have attributed changes in a number of measures to information technology, including labor's share of output, operating margins and the capital labor ratio. These possible effects of IT are important because they relate to economic inequality, firm market power, and other economic trends.

In this section, we explore possible associations between some of these variables and IT shocks. While our above analysis used control functions to correct for selection bias, we do not attempt to correct for selection bias in the analysis of these variables. Nevertheless, the associations between IT spikes and subsequent changes in labor's share and margins provide some insights about trends in these outcome variables. What we can do is to establish a sort of Granger causality that certain changes in these outcome variables tend to follow IT shocks.

Also, we do not present a complete model of these effects. Instead, we identify four explanations for the decline in labor's share of output from the literature and use our empirical analysis to distinguish between them by looking at the responses of the variables in equation (7):

1. Autor et al. (2017) find that most of the decline in labor's share of revenue occurs from the relatively greater growth of firms with low labor share. They find little within-firm change in labor's share. In their model, IT may play a role in heightening competition via reduced communication costs. We test whether IT spikes are followed by a decline in labor's share of revenue at the affected firms.
2. Karabarbounis and Nieman (2014), analyzing aggregate data, contend that falling IT prices induced firms to replace labor with capital. In their model, labor and

capital have an elasticity of substitution that is greater than one so that the decline in the relative price of IT should cause a decrease in labor's share and an increase in the capital-labor ratio. While their analysis is at an aggregate level, if falling IT prices are driving major IT investments, then we should expect to see these investments also followed by a decline in labor's share and an increase in the capital-labor ratio. We test these below.

3. Acemoglu and Restrepo (2018) propose that automation replaces human labor with machines on particular tasks, causing a decline in labor's share of output.[30] In this case, the capital-labor ratio should also rise.
4. IT innovations might generate quasi-rents, causing a rise in markups and a fall in labor's share of income. Bessen (2017) finds a link between industry IT investment and firm operating margins. Calligaris et al. (2018), using cross-country data, find a link between digitalization and markups. We test whether IT spikes are followed by a rise in operating margins, which are related to markups.

We can understand these explanations in terms of equation (7). A decline in labor's share of revenue can follow from a decline in the output elasticity of labor, β, or from a rise in markups, μ. But a decline in β should also correspond to an increase in the capital-labor ratio, all else equal. And a rise in markups should correspond to an increase in long-term equilibrium operating margins. We can distinguish between these explanations depending on how these variables respond following an IT shock.

We begin with Table 5, looking at estimates in the spirit of a difference-in-differences analysis for wages and labor's share of revenue. In this table and the next, the outcome variable is trimmed of the 1 percent tails to remove influential outliers (mostly firms with very small revenues) when the outcome is labor's share of revenue or operating margin. The first two columns use reported wage data. Only a fraction of Compustat firms report "staff expense," mostly in the financial sector. There is no significant change in either

[30] They also note that technology might increase the number of tasks in a way to offset this tendency.

wages or labor's share for these firms associated with an IT spike, but the sample size may be too small to identify an association. Columns 3 and 4 impute wages using BEA industry estimates for employee compensation for those firms that do not report staff expense. To the extent that firm wages move in tandem with industry wages, this imputation should be representative and, as above, the imputed wage bill correlates with the actual wage bill where we have data. Here, we see no effect on wages but a significant decline in labor's share of revenue. Since wages are little changed, this decline represents the relative growth rates of revenue and labor, which diverge, as we have seen above. We thus find that there *is* a significant association at the firm level on labor's share following major IT investments. This finding does not, of course, necessarily contradict the finding of Autor et al. (2017) that most of the aggregate decline comes from between-firm changes.

Columns 5 through 7 explore related variables to distinguish between the various explanations. Operating margins increase significantly, both in the economic sense and in the statistical sense. Moreover, the magnitude of the implied change in $\frac{1}{\mu}$, -1.6%, closely matches the decline in labor's share of revenue, -1.4%. Quasi-rents might be larger in industries where IT is part of the product if, for instance, these industries also experience network effects. The 6th column excludes these industries. The result is still significant and similar in magnitude. The capital-labor ratio shows a small increase after an IT spike that is not statistically significant. These estimates suggest that, overall, explanation #4, a rise in rents, fits the data best.

Table 6 looks at these differences across industries, suggesting some important industry heterogeneity. Column 1 repeats the regression on labor's share. Manufacturing, transport, utilities, and other industries all show significant declines in labor's share of revenue; services show an increase. To the extent that manufacturing, transport and utilities are arguably industries subject to automation, this finding lends support to the Acemoglu-Restrepo hypothesis.[31]

Column 2 looks at operating margins where the increase is substantial in nondurable manufacturing and especially in services. Because services includes the software industry, Column 3 excludes IT-producing industries; now the services coefficient is smaller and not statistically significant. IT producing industries might, indeed, have higher quasi-rents.

Column 4 looks at the capital-labor ratio. There are significant increases in transport/utilities and "other" industries. In these two sectors, declining share of labor might be a result of a replacement of labor by capital, either from factor substitution (Karabarbounis and Nieman) or from biased technical change (Acemoglu and Restrepo). But that does not seem to be the case for manufacturing generally or for tertiary sector industries.

In summary, IT spikes are associated with declining labor share and that is closely linked to rising markups overall. There are some industry differences, suggesting that in some industries the decline in labor's share could be the result of rising markups while in others it could be from labor-displacing automation.

Conclusion

This paper firmly takes the view that much investment in information technology is innovation. We measure this innovative activity using the share of software developers in a firm's workforce. And we show how major changes in firm software development can be

[31] Bessen et al. (2019) find that automation occurs in tertiary sector industries as well as manufacturing and transport. On the other hand, in Acemoglu and Restrepo's model, the labor-displacing effect of automation can be offset if technology creates new tasks for workers to perform. Arguably, this might happen more frequently in tertiary sector industries.

used to analyze various impacts and associations of information technology using difference-in-differences.

This analysis helps answer several questions but also raises several more. We find that these IT investments boost firm productivity and this effect has not diminished after 2002, contrary to the view of Robert Gordon (2017). Nor has the frequency of spikes declined. However, we also find that firms have grown less in response to this improved productivity in recent years. That means that the contribution of IT to aggregate productivity growth is likely smaller—these newly more productive firms are not increasing their share of aggregate employment as much. This finding parallels the evidence of Decker et al. (2018) that the response to productivity shocks has declined in recent years. The question, of course, is why.

We also find that, on average, major IT investments increase firm employment, even at firms that use AI, contrary to a common view. New information technologies are not creating unemployment overall. Nevertheless, even though firm employment rises on average, some jobs are lost (and others are created), perhaps an increased number. We see some evidence that jobs are lost in manufacturing, transport, and utilities while job growth following IT shocks is robust in trade, services, and finance. This follows the pattern proposed by Bessen (2019) where technologically mature industries will tend to have lower elasticity of demand and, hence, a weaker or negative employment response. But this pattern means that workers may need to transfer from one industry to another, perhaps changing occupations and skills. The policy challenge here is to facilitate these transitions, reducing the burden on workers and facilitating more rapid adoption of new technologies. An important question, then, is to understand which occupations are most affected, what new skills are most needed and how workers can be best helped to make these transitions.

We also find that labor's share of revenue declines following IT shocks on average because firms earn greater markups following the investment. This might be because they earn quasi-rents on their innovations. In any case, more research is needed to understand the mechanism and how it relates to aggregate trends in the labor share as in Autor et al. (2017).

Finally, this paper has focused on episodes where firms sharply increase the share of software developers in their workforces. This is a useful measurement device, but much innovative IT occurs outside of these episodes, including at very large firms that spend heavily on IT—at large firms, spikes at individual business units are obscured in aggregate numbers. In addition, we have only explored direct impacts at the firms making the investments. These investments surely affect rival firms, firms in downstream and upstream industries, and consumers.[32] So while our results highlight some important impacts of IT, more research is needed to understand the full social impact. Nevertheless, we can conclude that understanding IT as innovation is critical to appreciating its role.

[32] Autor and Salomons (2018) address some of these issues with productivity growth.

References

Acemoglu, D. and Restrepo, P. (2018). Modeling automation. AEA Papers and Proceedings, 108:48–53.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. "Identification properties of recent production function estimators." Econometrica 83, no. 6 (2015): 2411-2451.

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad. "The skill complementarity of broadband internet." The Quarterly Journal of Economics 130, no. 4 (2015): 1781-1824.

Arellano, Manuel, and Stephen Bond. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." The review of economic studies 58, no. 2 (1991): 277-297.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. The Fall of the Labor Share and the Rise of Superstar Firms. No. 23396. National Bureau of Economic Research, Inc, 2017.

Autor, David, and Anna Salomons, "Is automation labor-displacing? Productivity growth, employment, and the labor share," Brookings Papers on Economic Activity, 2018 forthcoming.

Barkai, Simcha. "Declining labor and capital shares." Stigler Center for the Study of the Economy and the State New Working Paper Series 2 (2016).

Bartel, Ann, Casey Ichniowski, and Kathryn Shaw. "How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills." The Quarterly Journal of Economics 122, no. 4 (2007): 1721-1758.

Bessen, James. Learning by doing: the real connection between innovation, wages, and wealth. Yale University Press, 2015.

Bessen, James E. "Information technology and industry concentration." Boston Univ. School of Law, Law and Economics Research Paper 17-41 (2017).

Bessen, James E. "Automation and jobs: When technology boosts employment." (2019).

Bessen, James E., Maarten Goos, Anna Salomons, and Wiljan Van den Berge. "Automatic Reaction-What Happens to Workers at Firms that Automate?." Boston Univ. School of Law, Law and Economics Research Paper (2019).

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron S. Jarmin, Megha Patnaik, Itay Saporta Eksten, and John Van Reenen. "IT and Management in America." (2014).

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron S. Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen. "What drives differences in management?" No. w23300. National Bureau of Economic Research, 2017.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. "Americans do IT better: US multinationals and the productivity miracle." The American Economic Review 102, no. 1 (2012): 167-201.

Borusyak, Kirill, and Xavier Jaravel. "Revisiting event study designs." Working paper (2017).

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt. "Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence." The Quarterly Journal of Economics 117, no. 1 (2002): 339-376.

Bresnahan, Timothy, Shane Greenstein, David Brownstone, and Ken Flamm. "Technical progress and co-invention in computing and in the uses of computers." Brookings Papers on Economic Activity. Microeconomics 1996 (1996): 1-83.

Brynjolfsson, Erik, and Lorin Hitt. "Paradox lost? Firm-level evidence on the returns to information systems spending." Management Science 42, no. 4 (1996): 541-558.

Brynjolfsson, Erik, and Lorin M. Hitt. "Computing productivity: Firm-level evidence." Review of economics and statistics 85, no. 4 (2003): 793-808.

Brynjolfsson, Erik, and Andrew McAfee. The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company, 2014.

Brynjolfsson, Erik, Andrew McAfee, Michael Sorell, and Feng Zhu. "Scale without mass: business process replication and industry dynamics." (2008).

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. "Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics." In The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, 2018.

Brynjolfsson, Erik, and Shinkyu Yang. "Information technology and productivity: a review of the literature." In Advances in computers, vol. 43, pp. 179-214. Elsevier, 1996.

Calligaris, Sara, Chiara Criscuolo, and Luca Marcolin. Mark-ups in the digital era. No. 2018/10. OECD Publishing, 2018.

Caroli, Eve, and John Van Reenen. "Skill-biased organizational change? Evidence from a panel of British and French establishments." The Quarterly Journal of Economics 116, no. 4 (2001): 1449-1492.

Correia, Sergio. "REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects." (2015).

Crespi, Gustavo, Chiara Criscuolo and Jonathan Haskel "Information Technology, Organisational Change and Productivity Growth: Evidence from UK Firms" CEP Discussion Paper 783 (2017).

De Loecker, Jan, and Jan Eeckhout. The Rise of Market Power and the Macroeconomic Implications. No. w23687. National Bureau of Economic Research, 2017.

Decker, Ryan A., John C. Haltiwanger, Ron S. Jarmin, and Javier Miranda. Changing business dynamism and productivity: Shocks vs. responsiveness. No. w24236. National Bureau of Economic Research, 2018.

Doms, M. E. and Dunne, T. (1998). Capital Adjustment Patterns in Manufacturing Plants. Review of Economic Dynamics, 1(2):409–429.

Frey, C. B. and Osborne, M. (2017). The Future of Employment: How Susceptible are Jobs to Computerisation? Technological Forecasting and Social Change, 114:254–280.

Gaggl, Paul, and Greg C. Wright. "A short-run view of what computers do: Evidence from a uk tax incentive." American Economic Journal: Applied Economics 9, no. 3 (2017): 262-94.

Ge, Chunmian, Ke ‐ Wei Huang, and Ivan PL Png. "Engineer/scientist careers: Patents, online profiles, and misclassification bias." Strategic Management Journal 37, no. 1 (2016): 232-253.

Gordon, Robert J. The rise and fall of American growth: The US standard of living since the civil war. Vol. 70. Princeton University Press, 2017.

Haltiwanger, J., Cooper, R., and Power, L. (1999). Machine Replacement and the Business Cycle: Lumps and Bumps. American Economic Review, 89(4):921–946.

Harrigan, James, Ariell Reshef, and Farid Toubal. The march of the techies: Technology, trade, and job polarization in France, 1994-2007. No. w22110. National Bureau of Economic Research, 2016.

Jin, Wang, and Kristina McElheran. "Economies before Scale: Productivity and Survival of Young Plants in the Age of Cloud Computing." 2017.

Jorgenson, Dale W., Mun S. Ho, and Kevin J. Stiroh. "productivity, Volume 3: information technology and the American growth Resurgence." MIT Press Books 3 (2005).

Karabarbounis, Loukas, and Brent Neiman. "The global decline of the labor share." The Quarterly Journal of Economics 129, no. 1 (2013): 61-103.

Kretschmer, Tobias. (2012), "Information and Communication Technologies and Productivity Growth: A Survey of the Literature", OECD Digital Economy Papers, No. 195, OECD Publishing.

Levinsohn, James, and Amil Petrin. "Estimating production functions using inputs to control for unobservables." The review of economic studies 70, no. 2 (2003): 317-341.

Marschak, Jacob, and William H. Andrews. "Random simultaneous equations and the theory of production." Econometrica, Journal of the Econometric Society (1944): 143-205.

Nilsen, Ø. A., & Schiantarelli, F. (2003). Zeros and lumps in investment: Empirical evidence on irreversibilities and nonconvexities. Review of Economics and Statistics, 85(4), 1021-1037.

Oliner, Stephen D., and Daniel E. Sichel. "The resurgence of growth in the late 1990s: is information technology the story?" Journal of economic perspectives 14, no. 4 (2000): 3-22.

Olly, Stephen, and Ariel Pakes. "The dynamics of productivity in the telecommunications equipment industry." Econometrica 64, no. 6 (1996): 1263-1297.

Pindyck, R. (1991). Irreversibility, Uncertainty, and Investment. Journal of Economic Literature 29(3):1110–48.

Rothschild, M. (1971). On the cost of adjustment. The Quarterly Journal of Economics, 85(4):605–622.

Stiroh, Kevin. "Reassessing the Impact of IT in the Production Function: A Meta-Analysis and Sensitivity Tests." In Contributions in Memory of Zvi Griliches, pp. 529-561. Annales D'Économie et de Statistique, 79-80, July-December 2005, 2010.

Tambe, Prasanna. "Big data investment, skills, and firm value." Management Science 60, no. 6 (2014): 1452-1469.

Tambe, Prasanna, and Lorin M. Hitt. "The productivity of information technology investments: New evidence from IT labor data." Information Systems Research 23.3-part-1 (2012): 599-617.

Yasar, Mahmut, Rafal Raciborski, and Brian Poi. "Production function estimation in Stata using the Olley and Pakes method." The Stata Journal 8, no. 2 (2008): 221-231.

Tables and Figures

Table 1: production function estimates

Outcome	Log(revenue)			
Specification	OLS	OLS	Olley-Pakes	Olley-Pakes
Model	Selection	Firm FE	Olley-Pakes	Olley-Pakes, FE
	(1)	(2)	(3)	(4)
Post spike	0.091**	0.048**	0.052**	0.046**
	(0.021)	(0.010)	(0.019)	(0.011)
Spiker	-0.214**			
	(0.028)			
Log(non-IT labor)	0.647**	0.647**	0.627**	0.632**
	(0.012)	(0.027)	(0.011)	(0.027)
Log(capital)	0.295**	0.204**	0.136**	
	(0.009)	(0.018)	(0.023)	
Year effects	✓	✓		✓
Firm effects		✓		✓
Control function				✓
Observations	47,448	47,416	45,438	44,177
R-squared	0.871	0.976	n.a.	0.975
Firms	4,088	4,056	3,872	3,800

The unit of observation is firm-year. Outcome is the natural logarithm of revenue in all models. Post spike is an indicator variable equal to one starting in the year of largest spike in IT employment. Spiker is an indicator variable equal to one for firms that spike during the sample period. All models include the natural logarithms of non-IT employees and capital. Model (3) estimated using the logarithm of capital investments as proxy variable (Olley and Pakes 1996). Models (1), (2) and (4) estimated with the Stata package developed by Correia (2016). Model (3) estimated with the Stata package developed by Yasar et al. (2008). Model (4) includes a third order polynomial in log investment and log capital as a control function for capital. We only estimate the first stage of the Olley-Pakes algorithm for this model, so we do not report the coefficient of capital. Robust standard errors clustered by firm in parentheses. ** $p<0.01$, * $p<0.05$

Table 2: IT spikes, sales and non-IT employment

Specification	OLS			
Outcome	Log(revenue)		Log(non-IT employees)	
Model	Selection	Firm FE	Selection	Firm FE
	(1)	(2)	(3)	(4)
Post spike	0.628**	0.109**	0.596**	0.070**
	(0.056)	(0.018)	(0.053)	(0.017)
Spiker	-1.239**		-1.125**	
	(0.088)		(0.082)	
Firm effects		✓		✓
Year effects	✓	✓	✓	✓
Observations	50,205	50,202	49,991	49,967
R-squared	0.051	0.939	0.049	0.942
Firms	4203	4200	4242	4218

The unit of observation is firm-year. Outcome is the natural logarithm of revenue in models (1) and (2), and the natural logarithm of non-IT employees in models (3) and (4). Post spike is an indicator variable equal to one starting in the year of largest spike in IT employment. Spiker is an indicator variable equal to one for firms that spike during the sample period. All models include year effects. Models (2) and (4) also include firm fixed effects. All models estimated with the Stata package developed by Correia (2016). Robust standard errors clustered by firm in parentheses. ** p<0.01, * p<0.05

Table 3: Heterogeneous “effects” of IT spikes

Specification	OLS		
Model	Production function	Revenue	Non-IT employees
Outcome	Log(revenue)	Log(revenue)	Log(non-IT employees)
	(1)	(2)	(3)
Panel A: AI/Big Data			
Post spike not AI	0.048**	0.098**	0.056**
	(0.011)	(0.018)	(0.018)
Post spike AI	0.045	0.247**	0.237**
	(0.031)	(0.068)	(0.067)
Panel B: IT producing			
Post spike not IT	0.043**	0.089**	0.042*
	(0.012)	(0.020)	(0.020)
Post spike IT	0.064**	0.179**	0.170**
	(0.021)	(0.042)	(0.040)
Panel C: Time period			
Post spike pre-2002	0.044**	0.154**	0.128**
	(0.014)	(0.027)	(0.027)
Post spike post-2002	0.052**	0.060*	0.005
	(0.016)	(0.027)	(0.025)
Panel D: US based			
Post spike not US	-0.011	0.055	0.055
	(0.023)	(0.041)	(0.043)
Post spike US	0.056**	0.117**	0.072**
	(0.011)	(0.019)	(0.019)
Panel E: New firms			
Post spike old firm	0.019	0.007	-0.016
	(0.011)	(0.021)	(0.022)
Post spike new firm	0.121**	0.366**	0.291**
	(0.022)	(0.032)	(0.027)

The unit of observation is firm-year. Outcome is the natural logarithm of revenue columns (1) and (2), and the natural logarithm of non-IT employees in column (3). Post spike is an indicator variable equal to one starting in the year of largest spike in IT employment, and it is interacted with different variables in each model. All models include year and firm fixed effects. Models in column (1) also include the natural logarithms of non-IT employees and capital. All models estimated with the Stata package developed by Correia (2016). Robust standard errors clustered by firm in parentheses. ** $p<0.01$, * $p<0.05$

Table 4: heterogeneity by industry

Specification	OLS		
Model	Production function	Revenue	Non-IT employees
Outcome	Log(revenue)	Log(revenue)	Log(non-IT employees)
	(1)	(2)	(3)
Post spike nondurable manufacturing	0.050	-0.019	-0.090*
	(0.031)	(0.048)	(0.044)
Post spike durable manufacturing	0.049**	0.060	0.024
	(0.015)	(0.033)	(0.032)
Post spike transport and utilities	0.135**	0.142*	-0.028
	(0.026)	(0.056)	(0.059)
Post spike trade	0.045	0.175**	0.209**
	(0.038)	(0.068)	(0.053)
Post spike finance	0.072**	0.229**	0.146**
	(0.025)	(0.040)	(0.051)
Post spike service	0.033	0.141**	0.172**
	(0.026)	(0.046)	(0.043)
Post spike others	-0.144**	0.099	0.129
	(0.039)	(0.074)	(0.075)

The unit of observation is firm-year. Outcome is the natural logarithm of revenue columns (1) and (2), and the natural logarithm of non-IT employees in column (3). Post spike is an indicator variable equal to one starting in the year of largest spike in IT employment, and it is interacted with industry indicators in each model. All models include year and firm fixed effects. Models in column (1) also include the natural logarithms of non-IT employees and capital. All models estimated with the Stata package developed by Correia (2016). Robust standard errors clustered by firm in parentheses. ** p<0.01, * p<0.05

Table 5: Relationship of IT spikes with wages, labor share, operating margin and capital to labor ratio

Specification	OLS						
Outcome	Log(wages)	Labor share of revenue	Log(wages)	Labor share of revenue	Operating margin	Operating margin	Log(Capital/ non-IT employees)
Sample	Reporting firms	Reporting firms	All firms	All firms	All firms	No IT producers	All firms
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post spike	0.003	-0.006	-0.001	-0.014**	0.016**	0.012*	0.007
	(0.020)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.013)
Observations	3,502	3,806	49,624	49,062	49,159	37,033	48,577
R-squared	0.885	0.801	0.919	0.754	0.644	0.672	0.922
Firms	339	346	4190	4144	4114	3035	4116

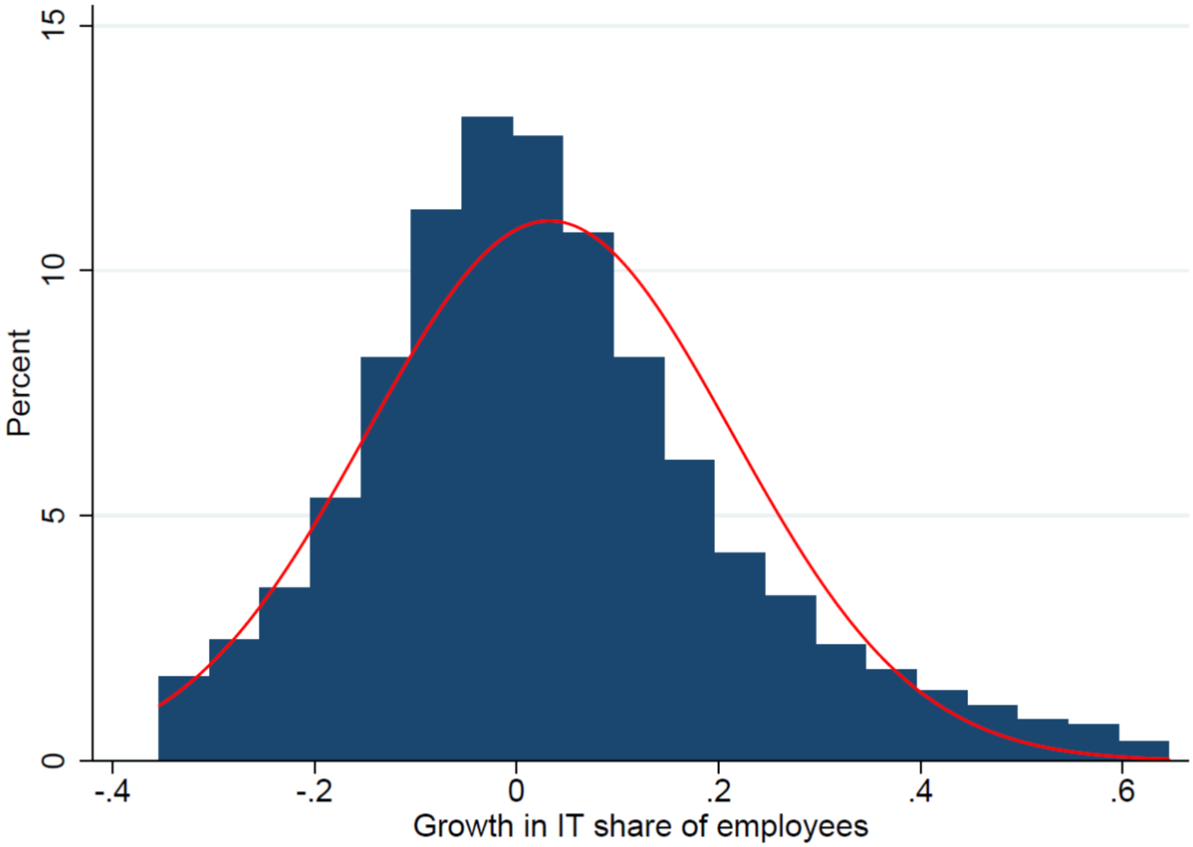
The unit of observation is firm-year. The sample for models (1) and (2) does not include firm-years that do not report wages. The sample for model (6) excludes IT producers. Outcome is the natural logarithm of wages in models (1) and (3), the labor share of revenue in models (2) and (4), the operating margin in models (5) and (6), and the natural logarithm of the ratio between capital and number of non-IT employees in model (7). Models (2), (4), (5) and (6) do not include firm-years in the top 1% and bottom 1% of the distribution of the outcome variable. Post spike is an indicator variable equal to one starting in the year of largest spike in IT employment. All models include year and firm fixed effects, and are estimated with the Stata package developed by Correia (2016). Robust standard errors clustered by firm in parentheses. ** $p<0.01$, * $p<0.05$

Table 6: Relationship of IT spikes with labor share, operating margin and capital to labor ratio by industry

Specification	OLS			
Outcome	Labor share of revenue	Operating margin	Operating margin	Log(Capital/ non-IT employees)
Sample	All firms	All firms	No IT producers	All firms
	(1)	(2)	(3)	(4)
Post spike nondurable manufacturing	-0.069**	0.033*	0.038**	0.029
	(0.012)	(0.014)	(0.014)	(0.028)
Post spike durable manufacturing	-0.029**	0.007	0.015	-0.035
	(0.005)	(0.008)	(0.009)	(0.020)
Post spike transport and utilities	-0.033**	-0.006	-0.000	0.198**
	(0.009)	(0.010)	(0.010)	(0.042)
Post spike trade	-0.002	0.003	0.007	0.002
	(0.007)	(0.007)	(0.007)	(0.037)
Post spike finance	0.001	-0.014	-0.009	0.045
	(0.008)	(0.012)	(0.012)	(0.037)
Post spike service	0.055**	0.058**	0.012	-0.177**
	(0.012)	(0.013)	(0.013)	(0.034)
Post spike others	-0.050**	0.009	0.016	0.351**
	(0.016)	(0.019)	(0.019)	(0.070)
Observations	49,062	49,159	37,033	48,577
R-squared	0.757	0.645	0.673	0.923
Firms	4,144	4,114	3,035	4,116

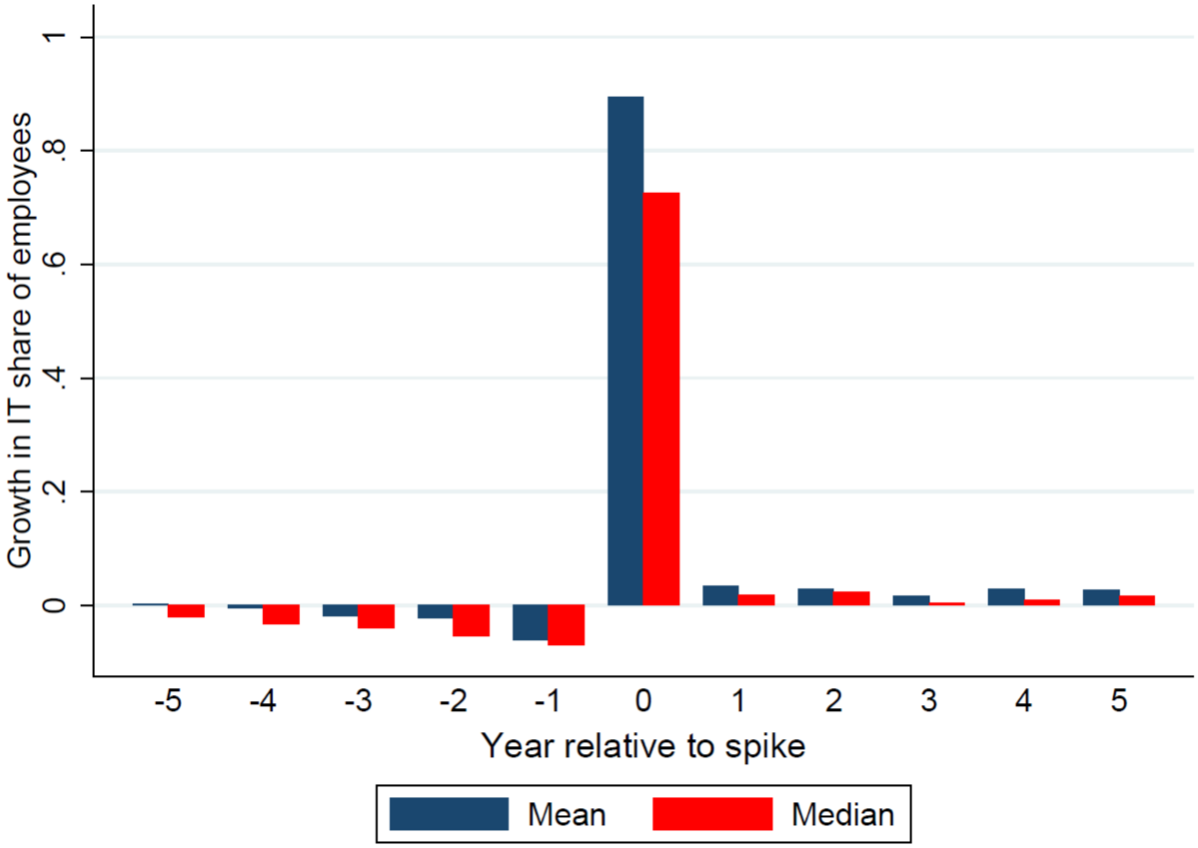
The unit of observation is firm-year. The sample for model (3) excludes IT producers. Outcome is the labor share of revenue in model (1), the operating margin in models (2) and (3) and the natural logarithm of the ratio between capital and number of non-IT employees in model (4). Models do not include firm-years in the top 1% and bottom 1% of the distribution of the outcome variable. Post spike is an indicator variable equal to one starting in the year of largest spike in IT employment, and it is interacted with industry indicators in all models. All models include year and firm fixed effects, and are estimated with the Stata package developed by Correia (2016). Robust standard errors clustered by firm in parentheses. ** p<0.01, * p<0.05

Figure 1: Distribution of growth in IT share of employment



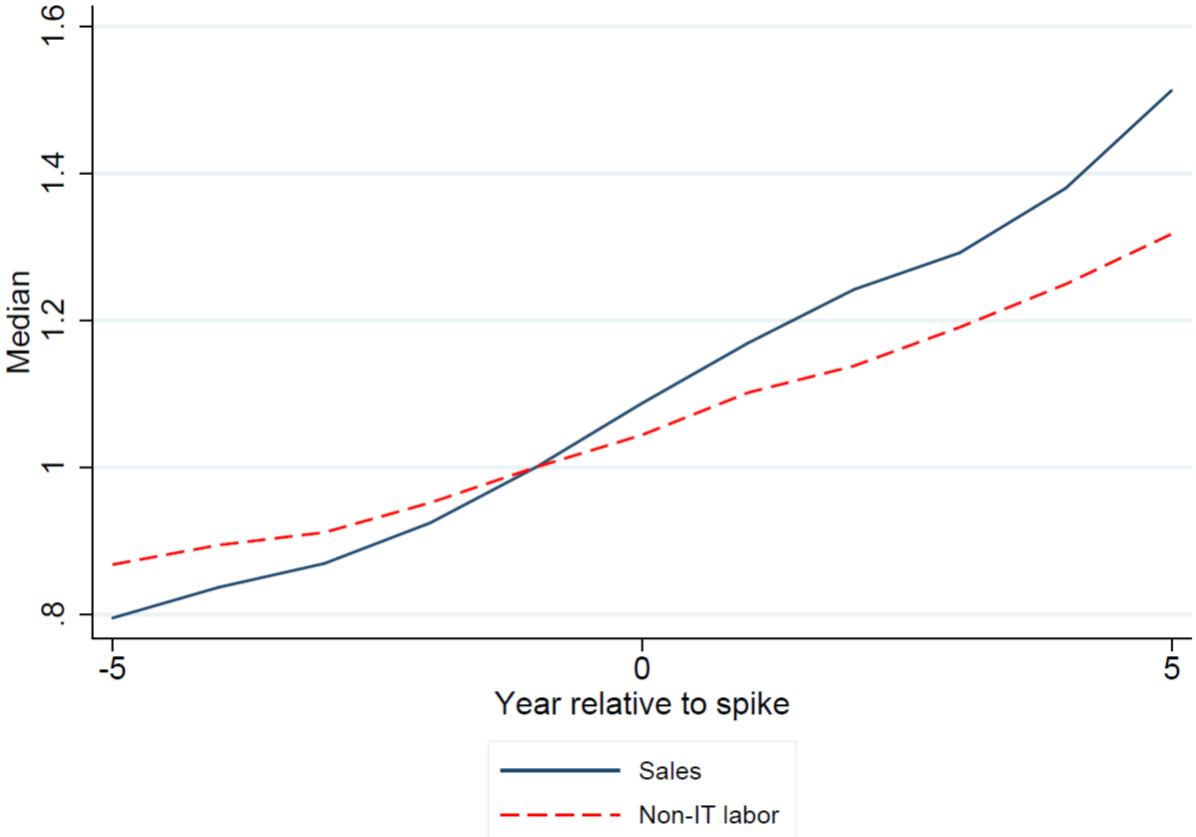
The figure reports the percentage of firm-years in each 5 percentage-point bin of the distribution of the growth in IT share of employees (blue histogram) and compares it with a normal distribution (red line).

Figure 2: Growth in IT share of employees around highest spikes



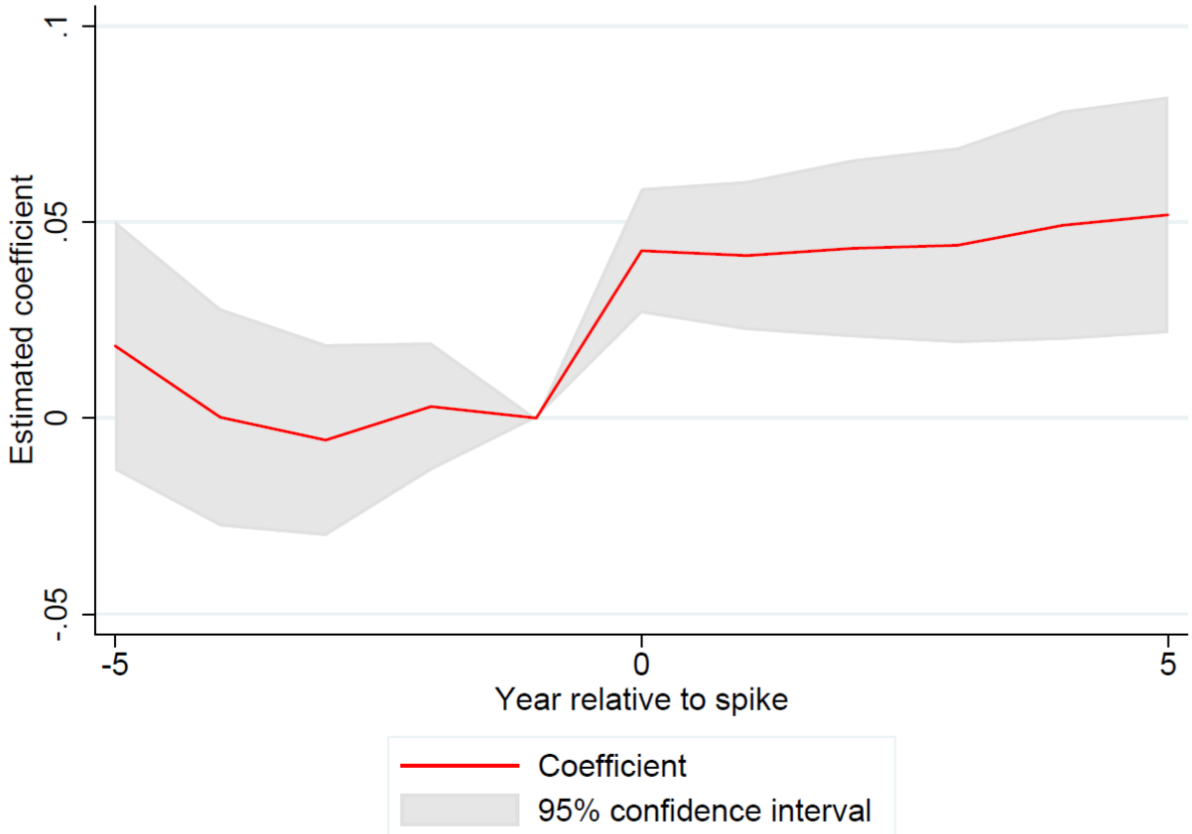
The figure reports the mean and the median of the growth in IT share of employees in an 11-year window around the year of the highest spike for firms that spikes at least once.

Figure 3: Growth in revenue and non-IT employees around highest spikes



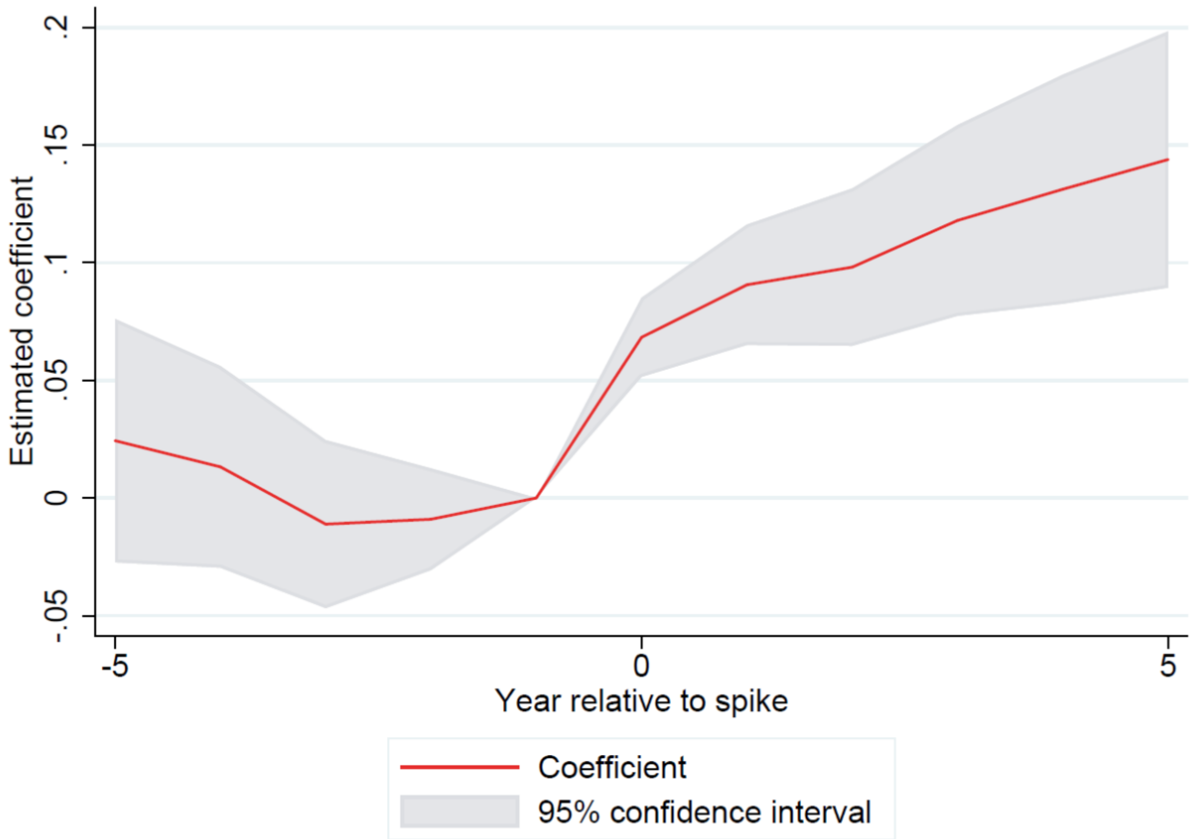
The figure reports the medians of revenues and non-IT employees in an 11-year window around the year of the highest spike for firms that spikes at least once. Revenues and non-IT employees in each year are normalized dividing their value in the current year by the value in the year before highest spike.

Figure 4: trend in revenues around spike, production function estimates



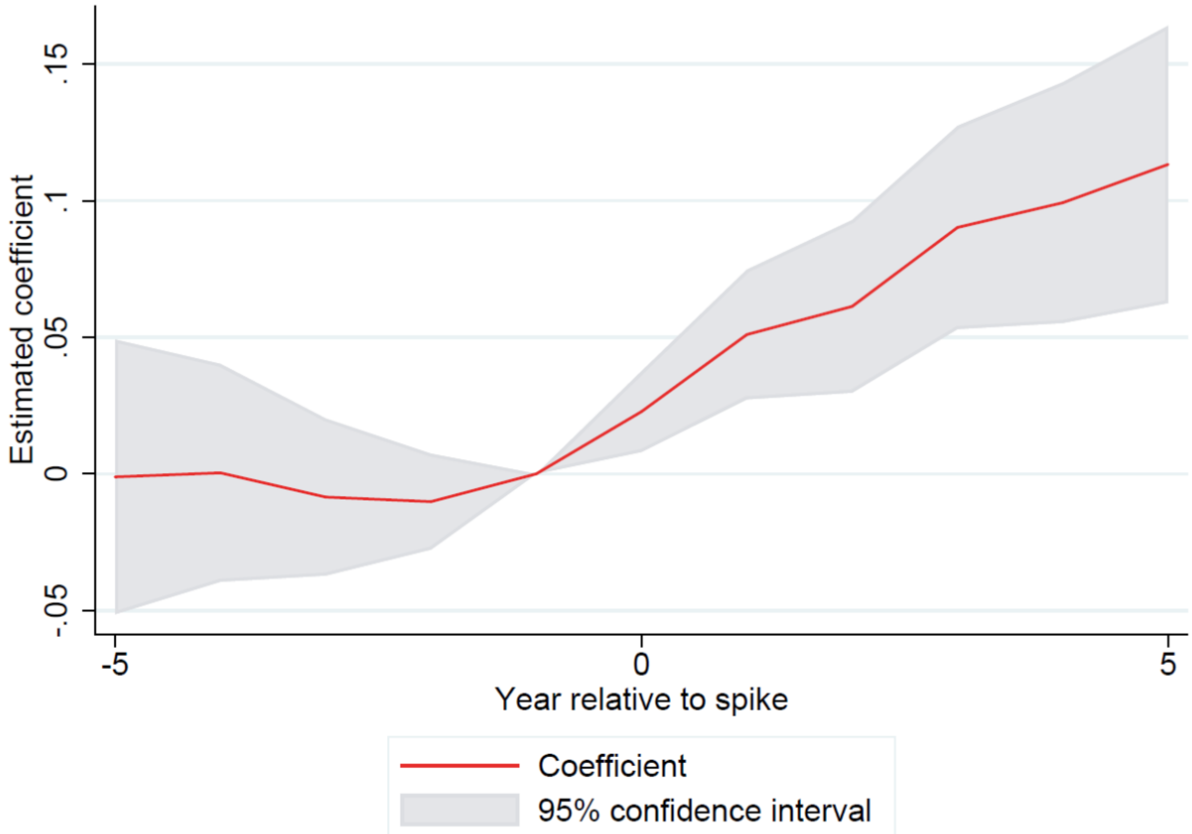
This figure plots the coefficients (solid line) and the 95% confidence intervals (shaded area) of a set of year-relative-to-spike from an OLS regressions similar to model (2) in table 1. The omitted category is the year before the spike. While the regression includes the full sets of coefficients, we report here only those in an 11-year window around the year of the spike. The unit of observation is a firm-year. The sample includes both spikers and firms that do not spike. The outcome variable is the natural logarithm of revenues. The model includes year and firm fixed effects, as well as the natural logarithms of non-IT employees and capital. Robust standard errors are clustered by firm.

Figure 5: trend in revenues around spike, difference-in-differences estimates



This figure plots the coefficients (solid line) and the 95% confidence intervals (shaded area) of a set of year-relative-to-spike from an OLS regressions similar to model (2) in table 2. The omitted category is the year before the spike. While the regression includes the full sets of coefficients, we report here only those in an 11-year window around the year of the spike. The unit of observation is a firm-year. The sample includes both spikers and firms that do not spike. The outcome variable is the natural logarithm of revenues. The model includes year and firm fixed effects. Robust standard errors are clustered by firm.

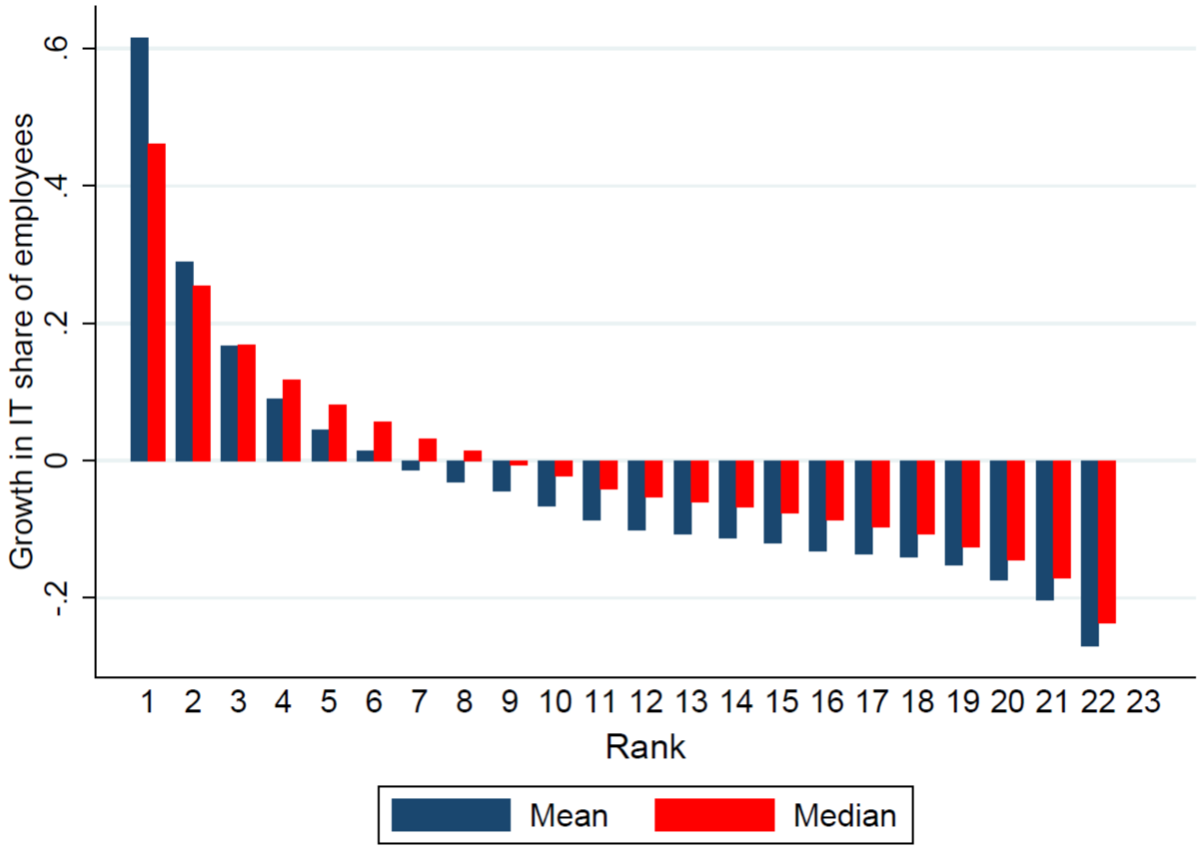
Figure 6: trend in non-IT employees around spike, difference-in-differences



This figure plots the coefficients (solid line) and the 95% confidence intervals (shaded area) of a set of year-relative-to-spike from an OLS regressions similar to model (4) in table 2. The omitted category is the year before the spike. While the regression includes the full sets of coefficients, we report here only those in an 11-year window around the year of the spike. The unit of observation is a firm-year. The sample includes both spikers and firms that do not spike. The outcome variable is the natural logarithm of non-IT employees. The model includes year and firm fixed effects. Robust standard errors are clustered by firm.

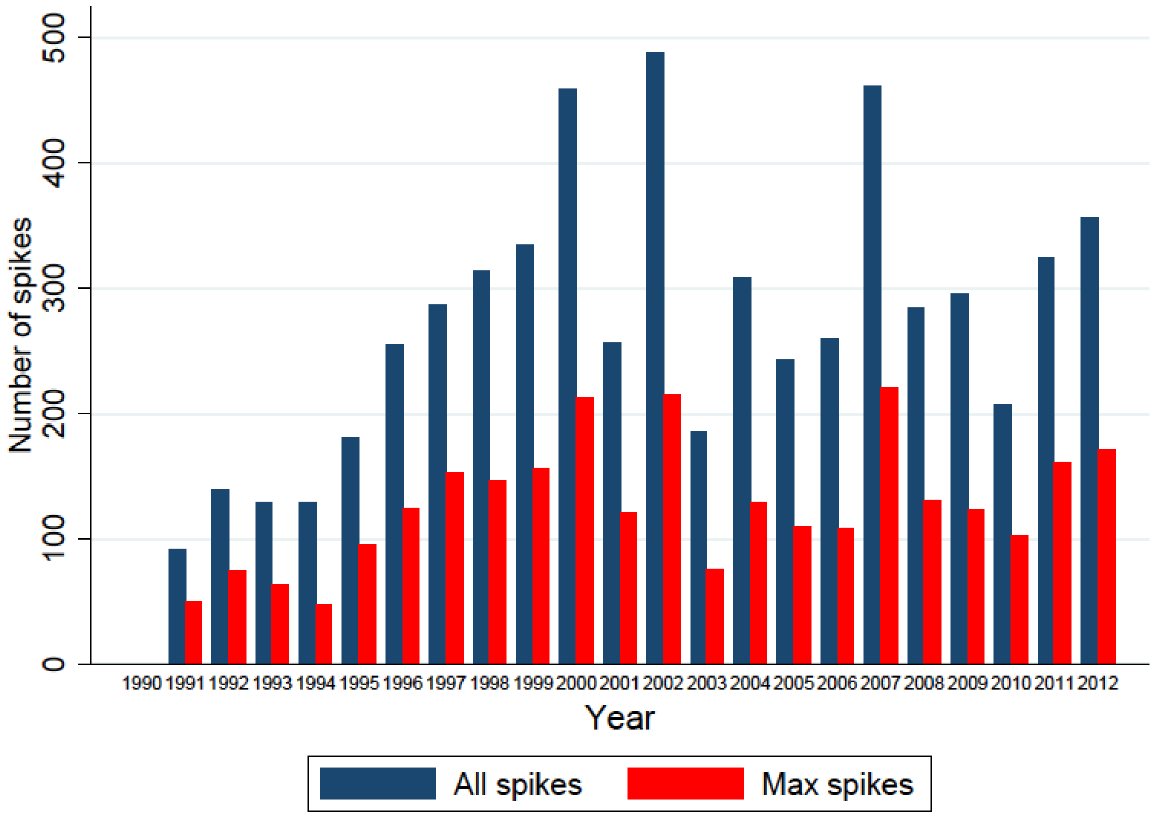
Appendix

Figure A1: Distribution of growth in IT share of employees by rank within firm



The figure plots the mean and the median of the growth in IT share of employees by rank within firm.

Figure A2: Frequency of spikes by year



This figure plots the number of spikes occurring in each sample year and the number of spikes limited to the largest spike per firm.

Table A1: Comparison of means for matched and unmatched firms

Variable	Unmatched	Matched	T-statistic	Norm. Diff.
	(1)	(2)	(3)	(4)
Sales (mill $2009)	642.64	4,626.56	72.27	0.34
Employees (total, 1000s)	2.50	15.11	70.55	0.34
Capital (mill $2009)	312.31	1,888.88	58.38	0.27
Wages (mill $2009)	68.82	74.16	17.07	0.09
IT producing industry	0.16	0.21	26.06	0.13
US firms	0.91	0.88	19.98	-0.10
Capital investment (mill $2009)	52.64	339.75	50.76	0.25
Market value of equity (mill $2009)	742.29	5,944.15	72.30	0.33
First year in Compustat	1,989.60	1,986.15	50.73	-0.25
Year	1,999.69	2,002.34	85.28	0.43
Observations	97,563	64,086		

Unit of observation is firm-year. Unmatched are firm-years in Compustat we cannot match to firm-years in LinkedIn. Matched are those we can match. Columns (1) and (2) report means by group. Column (3) reports the t-statistics from a test of the difference between the means in the first two columns. Column (4) reports the normalized difference in average covariates between groups.

Table A2: descriptive statistics

Variable	N	Mean	SD	Min	Max
Post spike	51,382	0.4	0.5	0.0	1.0
Spiker	51,382	0.7	0.4	0.0	1.0
Growth in IT share of employees	47,120	0.1	0.3	-1.0	4.7
Revenue	50,205	5,423.7	17,541.3	0.0	440,944.5
Non-IT employees	50,108	17.2	53.9	0.0	2,182.0
Capital	50,134	2,161.0	8,328.5	0.0	247,286.0
Capital/Non-IT employees	48,809	206.4	1947.7	0.0	229,661.4
Wages	49,655	74.7	74.4	0.0	13,773.5
Capital investment	48,012	386.2	1,663.4	-401.6	49,105.4
Market value of equity	51,290	6,974.4	23,668.1	0.0	744,989.4
AI/Big Data	51,382	0.1	0.3	0.0	1.0
It producing	51,382	0.2	0.4	0.00	1.0
First year in Compustat	51,382	1,985.4	16.3	1,950.0	2,011.0
Year	51,382	2,002.9	6.0	1,990.0	2,012.0
Labor share of revenue*	49,100	0.3	0.3	0.0	2.7
Operating margin*	49,200	0.1	0.3	-3.6	0.6

Unit of observation is firm-year. All the dollar amounts are in millions of 2009 dollars. Employment data in thousands of employees. *Trimmed of 1% top and bottom tails.

Table A3: sample by industry

Industry	N	%
Durable manufacturing	14,518	28.30%
Service	9,847	19.20%
Finance	7,629	14.80%
Nondurable manufacturing	7,088	13.80%
Transport and utilities	5,245	10.20%
Trade	4,417	8.60%
Other	2,638	5.10%
Total	51,382	100.00%

Table A4: Frequency of spikes per firm

Number of spikes per firm	Number of firms	Percent of firms
0	1,468	34.44
1	1,107	25.97
2	768	18.02
3	513	12.04
4	271	6.36
5	93	2.18
6	34	0.80
7	8	0.19
Total	4,262	100.00

Unit of observation is a firm.

Table A5 comparison of firm-years pre- and post-spike

	Pre spike (1)	Post spike (2)	T-stat (3)	Norm. Diff. (4)
Revenue	6,460.05	4,140.15	14.76	-0.13
Non-IT employees	19.72	14.19	11.45	-0.10
Capital	2,404.33	1,858.98	7.29	-0.07
Capital/Non-IT employees	193.65	222.02	1.60	0.01
Wages	74.04	75.48	2.14	0.02
AI/Big Data	0.15	0.08	25.86	-0.23
IT producing	0.26	0.22	9.30	-0.08
Labor share of revenue*	0.32	0.31	1.62	-0.01
Operating margin*	0.11	0.11	2.37	0.02
Observations	28,536	22,846		

Unit of observation is firm-year. Pre-spike observations include firm-years for firms that do not spike and those before the year of the highest spike for the spikers. Columns (1) and (2) report means by group. Column (3) reports the t-statistics from a test of the difference between the means in the first two columns. Column (4) reports the normalized difference in average covariates between groups. *Trimmed of 1% top and bottom tails.

Table A6: Predictors of spikes

Outcome	1[spike] X 100					
Specification	OLS					
	Revenue (1)	Non-IT labor (2)	Market cap (3)	Change in revenue (4)	Changed in non-IT labor (5)	Change in market cap (6)
Lagged revenue	-1.11**					
	(0.08)					
Lagged non-IT employees		-1.07**				
		(0.08)				
Lagged market cap			-1.04**			
			(0.07)			
Lagged change in revenue				0.14		
				(0.54)		
Lagged changed in non-IT employees					-0.25	
					(0.60)	
Lagged change in market cap						0.40
						(0.29)
Observations	46,002	45,838	47,058	41,802	41,401	42,780
R-squared	0.03	0.03	0.03	0.03	0.03	0.03
Mean outcome	12.72	12.63	12.71	12.23	12.11	12.21
Firms	4200	4231	4260	3973	3984	4033

Unit of observation is firm-year. The outcome is an indicator variable equal to one in the year of a spike, multiplied by 100. Firms at risk of a spike from their first year in the sample, or the first year after a spike (i.e. firms may have multiple spells). All models include year effects and year-since-spell-start dummies. Robust standard errors in parentheses clustered by firm. ** $p<0.01$, * $p<0.05$

Table A4: robustness of production function estimates

Outcome	Log(revenue)					
Specification	OLS					
Model	industry-by-year	L & K by industry	spikers	higher spikes	no multi-spikers	no growth-outliers
	(1)	(2)	(3)	(4)	(5)	(6)
Post spike	0.049**	0.049**	0.058**	0.034**	0.084**	0.046**
	(0.011)	(0.010)	(0.010)	(0.012)	(0.017)	(0.009)
Observations	45,644	47,416	34,778	47,416	23,743	46,508
R-squared	0.980	0.977	0.970	0.976	0.982	0.980
Firms	3981	4056	2674	4056	2438	4013

The unit of observation is firm-year. Sample for model (3) contains only firms with an IT spike. Sample for model (5) excludes firms with more than one IT spike. Sample for model (6) excludes firms with at least one year in the top 1% or bottom 1% of the distribution of revenues for firm-years in the sample. Outcome is the natural logarithm of revenue in all models. Post spike is an indicator variable equal to one starting in the year of largest spike in IT employment. In model (4) this variable is computed using a higher threshold. All models include the natural logarithms of non-IT employees and capital, and firm fixed effects. Model (2) allows the coefficients of non-IT employees and capital to differ by industry (SIC 2-digit). Model (1) includes industry-by-year effects (SIC 4-digit). Models (2)-(6) include year effects. All models estimated with the Stata package developed by Correia (2016). Robust standard errors clustered by firm in parentheses. ** p<0.01, * p<0.05

Table A5: IT spikes, sales and non-IT employment, robustness checks

Panel A

Outcome	Log(revenue)				
Specification	OLS				
Model	industry-by-year	spikers	higher spikes	no multi-spikers	no growth-outliers
	(1)	(2)	(3)	(4)	(5)
Post spike	0.098**	0.118**	0.088**	0.170**	0.110**
	(0.019)	(0.017)	(0.021)	(0.030)	(0.017)
Observations	48,447	36,859	50,202	25,045	49,238
R-squared	0.953	0.924	0.939	0.953	0.943
Firms	4,129	2,757	4,200	2,533	4,158

Panel B

Outcome	Log(non IT-employees)				
Specification	OLS				
Model	industry-by-year	spikers	higher spikes	no multi-spikers	no growth-outliers
	(1)	(2)	(3)	(4)	(5)
Post spike	0.065**	0.068**	0.060**	0.106**	0.074**
	(0.018)	(0.016)	(0.021)	(0.028)	(0.017)
Observations	48,160	36,599	49,967	24,993	49,041
R-squared	0.955	0.930	0.942	0.956	0.946
Firms	4,146	2,775	4,218	2,539	4,198

The unit of observation is firm-year. Sample for column (2) contains only firms with an IT spike. Sample for column (4) excludes firms with more than one IT spike. Sample for column (5) excludes firms with at least one year in the top 1% or bottom 1% of the distribution of revenues or non-IT employees for firm-years in the sample. Outcome in Panel A is the natural logarithm of revenue in all models. Outcome in Panel B is the natural logarithm of non-IT employees in all models. Post spike is an indicator variable equal to one starting in the year of largest spike in IT employment. In column (3) this variable is computed using a higher threshold. Models in column (1) includes industry-by-year effects (SIC 4-digit). Models (2)-(6) include year effects. All models estimated with the Stata package developed by Correia (2016). Robust standard errors clustered by firm in parentheses. ** p<0.01, * p<0.05