5-2020

Information Technology and Firm Employment

James Bessen

BU School of Law

Information Technology and Firm Employment

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

May 2020

James Bessen
Boston University School of Law

Cesare Righi
Boston University School of Management

Information Technology

and Firm Employment

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

May 2020

Abstract
Do firms displace labor with new information technologies such as "artificial intelligence"? It is challenging to distinguish the effects of technology adoption from unobserved productivity and demand shocks. We take a first look at the economic impacts of large custom software investment —"IT spikes"—using a novel methodology to obtain consistent estimates. Following these events, firm employment increases by about 7% and revenues by about 11%. Rather than displace labor, IT spikes increase revenues and markups, implying decreased labor share of output. Moreover, growth is greater for firms that use AI, IT-producing firms, newer firms, and those in the trade, service, and financial sectors.

JEL Codes: D22, J21, O33

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

This paper is a major revision of a working paper titled "Shocking Technology: What happens when firms make large IT investments?" Thanks for helpful comments from Seth Benzell, Josh Feng, Richard Freeman, Sina Khoshsokhan, Kristina McElheran, Michael Meurer, Filippo Mezzanotti, Pascual Restrepo, Daniel Rock, Tim Simcoe, Ken Simons, Gabriel Unger and participants at the TPRI seminar, the Harvard SBBI seminar, the Harvard Growth Lab seminar, and the NBER productivity lunch. 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

What happens to employment when firms adopt major new technologies? Concerns about job losses have been heightened as new digital technologies ranging from robotics to machine learning and other forms of artificial intelligence are being adopted across many sectors of the economy. Following Frey and Osborne (2017), over 18 recent studies predict job losses from new automation technologies, including some predictions of massive job losses (Winick 2018). These studies are based on presumed occupational exposure to automation. Beginning with Autor, Levy, and Murnane (2003), a large literature has explored how occupational characteristics make certain jobs more "automatable." However, just because certain jobs may be more automatable does not mean that firms will adopt a new technology or that its adoption will displace workers in these jobs.

If workers in automatable occupations are being replaced by new technology, we should see the clearest evidence of this in studies of firm or establishment data.[1] However, firm-level measures of technology adoption are difficult to obtain for large representative samples of firms.[2] Following Bessen et al. (2019), a number of studies have sought new sources of data that identify discrete adoption events in broad populations of firms.[3] For reasons we discuss below, major new technologies are often adopted in "lumpy" investment episodes. These discrete events help isolate the impact of technology adoption from other

[1] A number of studies have looked at the impact of robots at a more aggregate level, including Acemoglu and Restrepo (2020), Graetz and Michaels (2018) and Dauth et al. (2017). Of course, the effect of new technology on firm employment can be different from the aggregate effect, but firm level analysis should nevertheless most clearly reveal the displacement of workers.

[2] Some studies have explored specific natural economic experiments, but these measure effects only for limited samples. Akerman et al. (2015) study broadband rollout in Norway; Gaggl and Wright (2018) use a tax incentive in the UK that affected small businesses.

[3] Acemoglu, Lelarge, and Restrepo (2020); Aghion et al. (2020); Bessen et al. (2020); Bonfiglioli et al. (2020); Domini et al. (2019); Humlum (2019); Koch et al. (2019).

changes using difference-in-differences designs. However, these designs do not fully control for the endogeneity of adoption. For instance, adoption might be correlated with unobserved productivity or demand shocks that also affect labor demand, thus biasing estimates.

This paper proposes to understand the impact of technology adoption by considering the structural relationship between labor demand and the production function. There are two benefits from modeling labor demand this way. First, the production function literature has well-established techniques for estimating unobserved productivity and demand shocks using control functions. We incorporate these estimates in our difference-in-differences labor demand equation, thus controlling for the most likely sources of endogeneity bias. Second, our labor demand equation helps distinguish different margins along which a major new technology affects labor demand. Specifically, a new technology can replace workers with machines (displacement effect), it can increase productivity, and it can increase markups, especially if the new technology is proprietary. Each of these margins affects employment and also other outcomes such as the labor share of output. While displacement decreases employment, all else equal, growth in productivity or markups can increase employment.

We apply this empirical approach to study the firm-level impact of major investments in proprietary information technology among firms listed publicly in the US. We identify adoption events by major increases in firm employment of software developers ("IT spikes") and estimate the effect of these events on labor and revenue with a difference-in-differences design. We find these events are followed by increases in firm (non-IT) employment of about 7% and increases in revenue of about 11%, after controlling for time-invariant unobserved heterogeneity among firms as well as unobserved productivity and demand shocks. These results are robust to controlling for shocks common to all firms in an industry, using only the subsamples of firms that experience an IT spike for the analysis or dropping firms that experience multiple IT spikes, alternative definitions of an IT spike, excluding firm-years in the tails of the distribution of labor or revenue growth, and controlling for factors that vary over time such as the number of mergers and acquisitions of a company and expenses in adverting and research and development.

Moreover, an event-study analysis shows that, before the IT spikes, the trends in labor and revenue for firms "treated" by the IT spikes are similar to those of "control" firms, and there is a gradual increase in both outcomes starting in the year of the IT spike. This suggests that our difference-in-differences estimates of the positive effects of the IT spikes are not due to pre-trends.

Exploring the sources of increased employment, we find no significant displacement of workers by capital, a small effect of productivity on employment, and most of the increase appears related to increases in markups. The latter might not be surprising considering that these investments in own-developed software are proprietary and thus might generate rents.

Finally, we explore the heterogeneous impacts of these adoption events. We find that job growth is greater in firms that use AI, IT producing firms, in firms headquartered in the US, in young firms, at the beginning of our sample period, and in firms in the trade, financial, and service sectors.

An emerging literature uses technology adoption events to study impacts on employment. Previous studies have measured these events by increases in automation expenditures (J. E. Bessen et al. 2019), in the use of robots (Bonfiglioli et al. (2020); Humlum (2019); Koch et al. (2019)), in the use of automated equipment (Acemoglu, Lelarge, and

Restrepo (2020); Domini et al. (2019)) and in motive electric power (Aghion et al. 2020). The challenge for this literature has been to control for unobserved factors correlated with adoption that might contemporaneously affect employment. Several of the prior studies find that firms that automate tend to have growing employment or revenues prior to the time of adoption, suggesting that unobserved productivity or demand factors could be related to the firm's decision to adopt new technology. Koch et al. (2019) use propensity score matching to control for unobserved trends; Aghion et al. (2020) use a Bartik-style instrumental variable estimation under some assumptions about factor input rigidities. Our paper contributes to this literature by using spikes in employment of software developers to measure large custom software investment events, and proposing a novel methodology to purge the effects of unobserved productivity and demand shocks by using control function estimates adapted from the production function literature. Moreover, our paper contributes by providing a picture of the differing impacts of IT spikes for firms with different characteristics.

Another contribution of our paper is that our labor demand equation separates the effects of productivity, displacement, and markups on labor demand. A large literature has looked at the productivity contribution of IT both at the aggregate level (see for example, (Oliner and Sichel 2000; Jorgenson et al. 2005) and at the firm level (see for example, (Brynjolfsson and Hitt 1996; 2003). However, these studies do not distinguish major custom software from routine IT investments nor do they estimate labor demand impacts of IT. We measure both IT spikes and also routine IT activity, measured as IT labor, following Tambe and Hitt (2012).[4] In addition, although we find that routine IT investment is correlated with

[4] See also Bloom et al. (2012) and Harrigan et al. (2016).

higher output, IT spikes are not correlated with greater productivity, suggesting that the impact of major IT investments on employment does not occur through this channel.

Other researchers see new information technology implicit in labor's declining share of output (Karabarbounis and Neiman 2014; D. Autor et al. 2020; Acemoglu and Restrepo 2018; Calligaris, Criscuolo, and Marcolin 2018). This could be because declining prices of IT hardware cause firms to substitute capital for labor (Karabarbounis and Neiman 2014) or because automation transfers tasks from humans to machines (Acemoglu and Restrepo 2018) or because IT raises markups and margins (J. Bessen 2020b; Calligaris, Criscuolo, and Marcolin 2018) or all of the above. Our analysis finds that IT spikes are not significantly associated with rising capital-labor ratios, suggesting that their impact on labor's share may be from higher margins rather than from technological displacement or capital-labor substitution.

We begin by motivating our use of discrete events to measure large custom software investments and outlining our empirical approach. Then we describe our data, report our results and conclude.

Information Technology Spikes

The nature of information technology (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.[5] In 2016, the Bureau of Economic Analysis estimates that private investment in

[5] This counts custom contracted and own-account (self-developed) software as a share of total gross investment in software, computers, and peripherals. In addition, there has been a shift of hardware processing to the cloud (Jin and McElheran 2017)

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. For instance, routine word processing software running on personal computers might well enhance productivity. But when a firm builds a custom logistics system with proprietary features that rivals do not have, that is different. Such systems may allow firms to earn quasi-rents, raising markups and affecting labor demand.

Another difference is that large, custom IT projects tend to be risky and have notoriously high rates of failure.[6] The high uncertainty, combined with 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, Cooper, and Power 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 nonconvex adjustment costs (Pindyck 1991; Rothschild 1971). Investments in custom software are typically irreversible; they cannot be resold because they are firm-specific. Moreover, large software systems have large, indivisible fixed costs. In addition, associated organizational changes may have large adjustment costs.

All of this suggests that much custom IT investment may occur in discrete episodes. Below we define "IT spikes" and find that a large share of IT investment occurs in these

[6] 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.

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.

Empirical Model

Unobserved productivity and demand shocks

The standard staggered difference-in-differences equation for estimating the effect of technology adoptions on employment is

(1)

$$l_{it} = \alpha_i + \delta_t + \gamma \cdot D_{it} + \omega_{it} + \varepsilon_{it}, \quad D_{it} \equiv \mathbf{1}(i \in T \ \& \ t \geq \tau_i)$$

where l_{it} is log employment for firm *i* at time *t*, α_i and δ_t are, respectively, firm and time period fixed effects, T represents the group of firms that adopt the technology, τ_i is the time firm *i* adopted, ω_{it} represents unobserved productivity or demand shocks known to the firm but unobservable to the econometrician, and ε_{it} represents optimization errors, random events such as weather, and other unforeseen errors. Under a parallel trends assumption, the coefficient γ measures the average treatment effect of technology adoption.

A challenge to estimating this equation is that the unobserved shock might be correlated with the likelihood of technology adoption, that is, $\text{corr}[D_{it}, \ \omega_{it}] \neq 0$. For example, a positive demand shock might make firms more likely to adopt and the demand shock might increase employment independently of the adoption, biasing upwards the estimate of γ.

This problem is very similar to the problem of simultaneity bias in production function estimation discussed by Marschak and Andrews (1944) and the subject of a large empirical literature. Using lowercase to designate logs, a Cobb-Douglas revenue production function can be written

(2)

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \varepsilon_{it}$$

where *y* is revenue, *l* is labor, and *k* is capital. If the firm chooses labor with knowledge of ω_{it}, OLS estimates will be biased.

A major innovation of this paper is based on the recognition that a) there are well-established techniques in the production function literature for obtaining estimates of ω_{it}, and, b) under some simple assumptions, a labor demand function (and product demand equation) similar to equation (1) can be derived from equation (2) incorporating these estimates, $\widehat{\omega}_{it}$. Regressions on this labor demand equation should yield estimates of γ that are consistent with respect to unobserved productivity and demand shocks.

Beginning with Olley and Pakes (1996), researchers have estimated unobserved ω_{it} using control functions. Under some assumptions, Olley and Pakes argue that unobserved productivity can be captured as a nonparametric function of investment and capital, permitting consistent estimates of β_l. The intuition is that while capital is slow to adjust to a productivity shock, investment adjusts immediately and increases with the magnitude of the shock. With a further assumption that ω_{it} evolves as a Markov process, they are able to estimate β_k and back out estimates, $\widehat{\omega}_{it}$. Levinsohn and Petrin (2003) propose using materials instead of investment as a proxy variable. Ackerberg, Caves, and Frazer (2015) allow for more general assumptions regarding the timing of employment in response to shocks.

Below, we use the Ackerberg, Caves, and Frazer (ACF) setup using investment as the proxy variable. We use ACF because it has more general assumptions about timing. We use investment rather than materials as a proxy because in our data materials need to be imputed, possibly introducing biases regarding IT (see below). We use revenue as our outcome variable, assuming a fixed relationship between materials and output. However, we find that our results are robust to alternative specifications.

Labor and product demand

Assuming a fixed product price elasticity of demand, we obtain a labor demand equation from the first-order conditions implied by (2). The production function and price equations are, suppressing firm and time subscripts,

(3)

$$q = \ln A + \beta \cdot l + (1-\beta) \cdot k, \quad p = d \cdot q^{-\frac{1}{\epsilon}}$$

where q is log output, A is unobserved productivity and d is an unobserved demand parameter. Defining

(4)

$$\beta_l \equiv \frac{\beta}{\mu}, \quad \beta_k \equiv \frac{1-\beta}{\mu}, \quad \omega \equiv \frac{\ln A}{\mu} + \ln d\,, \quad \mu \equiv \frac{\epsilon}{\epsilon - 1}$$

recovers equation (2). The associated first order profit maximizing conditions are

(5)

$$\frac{\partial\, Q}{\partial\, L} = \frac{\beta Q}{L} = \frac{w\mu}{p}, \quad \frac{\partial\, Q}{\partial\, K} = \frac{(1-\beta)Q}{K} = \frac{r\mu}{p}$$

where w is the wage and r is the capital rental rate. From these we derive a labor demand equation (see Appendix) where log labor is a function of three variables— ω, β, and ϵ—that might be affected by firm technology choices:

(6)

$$l = f^1(\epsilon) + \epsilon \cdot \omega - (\epsilon - 1)f^2(\beta) + \ln \beta - \ln \frac{w}{r}.$$

where $f^1(\epsilon)$ and $f^2(\beta)$ are provided in the Appendix.

This equation captures three margins along which technology might affect labor demand. First, by raising productivity, ω, technology might increase product demand and hence demand for labor. Second is β, which equals labor's share of output.[7] In Hicks (1963), labor-saving technical change reduces β; in Acemoglu and Restrepo (2018), automation does so as tasks are transferred from humans to machines. Third is ϵ, which reflects firm market power. If proprietary technology allows the firm to earn rents, then average markups, the ratio of revenue to operating costs, should increase. It is straightforward to show (Appendix) that $\mu = \frac{\epsilon}{\epsilon - 1}$ equals the average markup. This means that technology rents would increase μ and decrease ϵ. Below we explore how each of these margins is associated with IT spikes.

While (6) is a complicated expression, we develop an approximation that can be estimated. To do this, we obtain estimates, $\widehat{\omega}_{it}$, from production function regressions. Also, because the first order conditions imply

$$\beta = \frac{1}{1 + \frac{rK}{wL}},$$

we use the log capital-labor ratio, κ, to proxy for β. Finally, our measure of IT spikes, D_{it}, captures changes in ϵ and any other impacts. Using Taylor series expansions, we obtain a first-order approximation,

[7] By the first order condition for labor, $\beta = \frac{wL}{pQ}$.

(7a)

$$l_{it} = \alpha_i + \delta_t + \gamma_1 \widehat{\omega}_{it} + \gamma_2 D_{it} + \gamma_3 \kappa_{it} + \varepsilon_{it}$$

and a second order approximation, including interaction terms,

(7b)

$$l_{it} = \alpha_i + \delta_t + \gamma_1 \widehat{\omega}_{it} + \gamma_2 D_{it} + \gamma_3 \kappa_{it} + \gamma_4 \widehat{\omega}_{it} \cdot D_{it} + \gamma_5 \kappa_{it}^2 + \gamma_6 \kappa_{it} \cdot D_{it} + \varepsilon_{it.}$$

These equations let us estimate the impact of IT spikes on labor demand, controlling for unobserved productivity and demand shocks and also for changing labor share of output. These equations also provide a framework for distinguishing the relative importance of the three margins of impact for IT spikes. We develop similar equations for product demand and estimate these as well.

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, industry codes and country of incorporation from Compustat.

We convert all variables defined in current dollars to 2009 dollars using deflators from the BEA.[8] We calculate capital as net property, plant and equipment. We define operating margins as the operating income before depreciation and taxes divided by revenue.

Some production function estimation methods require measures of materials or intermediates. Where data are available, we calculate materials as operating expenses minus depreciation and staff expense (employee compensation). However, staff expense is only reported for a fraction of Compustat firms, mostly in the financial sector. For the remainder of firms, we impute the wage bill by multiplying the number of firm employees by the industry employee compensation per employee. Unfortunately, this imputation is likely to be problematic because wage levels are likely correlated with the IT share of the workforce, making measurement error correlated with our variable of interest. For the same reason, our production functions use revenue as the outcome variable rather than value added, which is also likely to be mis-measured in a problematic way.

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, Huang, and Png 2016).[9] 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.

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

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.[10]

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, matched firms are substantially larger than unmatched firms in terms of revenue, 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. Because 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

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

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 spikes. 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 of 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.[11]

[11] These are "Hadoop" "big data" "quantitative analyst" "data scientist" "data science" "artificial intelligence" "machine learning" "deep learning" "neural network" and "natural language processing." We define a time-

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. Table A3 shows the distribution by sector.[12]

Evidence on 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, we define a practical way of identifying these "spikes." We can see evidence of the lumpiness of the IT share growth in Figure 1. The solid line represents the distribution of the year-over-year growth rate in firm IT shares. The dashed line represents a normal distribution. Comparing the two distributions 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.

invariant indicator equal to one for firms that employ at least one person listing these titles in their profile during our sample period.

[12] We also collected data from additional sources. These include mergers and acquisitions made by a company from the SDC Platinum dataset, and the identity of CEOs from the ExecuComp database.

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).[13] 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.[14]

To gain a sense of spikes and their proprietary nature, 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 acquired a business unit of an embedded computer firm in order to rapidly put together a product and bring it to market. In each case, the firm responded to an idiosyncratic opportunity, aiming to gain a competitive advantage that might earn rents.

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

[13] We report results based on alternative definitions of IT spike in the Appendix.

[14] Figure A1 shows also shows evidence that the growth in IT share of employees tends to be lumpy: within firm, few years are characterized by very large growth.

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]

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.

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

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

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 logs): 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.

Results

Production functions

Table 1 shows a range of alternative production function estimates, including our preferred specification in column 4. In all regressions, the outcome variable is the log of revenue, and we cluster the standard errors by firm. For reference, the first two columns show a simple OLS regression and an OLS regression with firm and year fixed effects. The next four columns show regressions using various control function methods. These regressions all include adjustments for firm exit. Columns 3 and 4 use the log of investment as the proxy variable. Column 3 uses the Olley-Pakes (1996) method while column 4 uses Ackerberg, Caves, and Frazer (2015). We prefer the ACF method because it is more general and because it allows more consistent assumption about the timing of production decisions. Columns 5 and 6 use the log of materials as a proxy variable, column 5 showing the Levinsohn-Petrin method and column 6 the ACF method. As noted above, we impute materials for most observations and this imputation may be correlated with firm IT use, introducing possible biases.[16]

In the following difference-in-differences models, we will use the $\widehat{\omega}_{it}$ estimated by the model in column (4) to control for productivity and demand shocks.

Employment and revenue estimates

The top panel of Table 2 reports difference-in-differences estimates of the labor demand equations, while the bottom panel reports findings for the comparable revenue equations. Instead of firm fixed effects, the specification in column 1 includes a single dummy variable for whether the firm is a spiker or not, which captures the "selection effect" into the IT spikes. This specification show a very large increase in employment and revenues after the IT spike, and the coefficients of the "selection" indicators confirm that firms that experience the IT spikes are smaller than the other firms in our sample.

Column 2 drops the spiker dummy and adds firm fixed effects to the regression model. The IT spikes are associated with an increase in employment by 7.0%, and an increase in revenue by 10.9%.

[16] Throughout the paper, we estimate all models with firm fixed effects using the Stata package "reghdfe" developed by Correia (2018) and all models with a control function using the Stata package "prodest"(Rovigatti and Mollisi 2018).

The previous regressions do not control for unobserved productivity and demand shocks. Columns 3 and 4 introduce controls as specified in equations (7a) and (7b) respectively. We obtain estimates of $\widehat{\omega}$ in a first stage regression, using the ACF method with investment as the proxy variable, and use these estimates as additional control variable in our difference-in-differences models. To obtain the standard errors for these regressions, we bootstrap (50 repetitions) the two-stage procedure, clustering by firm. Column 4 reports the marginal effects of the interaction equation. We report the corresponding regression coefficients in Table A8. The inclusion of the additional controls has little influence on coefficients of the IT spikes, which are almost unchanged. This implies that these controls are not strongly correlated with the occurrence of spikes.[17] Our interpretation is that the estimates in column 2 are not substantially biased by unobserved productivity and demand shocks.

While we saw in Table 1 that there are differences between the estimates of the production function obtained with various methods, these differences create only modest variation in our estimates of the effects of IT spikes. Table A7 in the Appendix shows estimates for the difference-in-differences models of employment and revenue in column (3) of Table 2 using the $\widehat{\omega}$ estimated with all the control function methods in Table 1. The estimates that use the Olley-Pakes method, the Levinsohn-Petrin method, and the ACF method with materials as proxy variable to produce $\widehat{\omega}$ are somewhat lower, but the coefficient of the IT spike dummy is positive and statistically significant at conventional levels in all models.

[17] Calculating simple correlation coefficients, $\widehat{\omega}$ has a small, but statistically significant negative correlation (-.03) with D, and κ has a small (.03) and statistically insignificant correlation; adding firm and year fixed effects, neither relationship is significant, economically or statistically. For $\widehat{\omega}$, the coefficient is -.002 (.004); for κ, it is .007 (.013).

Under a parallel trends assumption, the coefficient of the post IT spike dummy estimates the impact of IT spikes on employment and revenue. This assumption is more plausible if firms that experience an IT spike and those that do not are on similar trends in employment and revenue before the IT spike. In order to examine the pre-spike trends and the dynamic effects of the IT spikes, we estimate two event-study models, and plot the coefficients for an eleven-year window around the IT spikes in Figures 4 and 5.[18] Both graphs show that the pre-IT-spike trends in employment and revenue are very similar, and there is significant rise in employment and revenue of the spikers after the event.[19]

Revenues appear to grow significantly faster than labor after the spikes, suggesting that labor's share of revenue declines. This will be true as long as wages do not substantially increase. For that portion of the sample reporting staff expense, we found no evidence of rising average pay, implying that labor share of revenue did indeed decline.[20]

Labor's share of revenue can decline for two reasons: labor's share of output (β) declines *or* the ratio of cost to revenue declines, that is, average markups, (μ), increase. The first changes the composition of costs away from labor, the second changes the wedge

[18]To produce the event study graphs, we estimate two regressions based on the equation $\ln Y_{it} = \alpha_i + \beta_t + \sum_{k\neq -1} \gamma_k \cdot \mathbf{1}(k = t - \tau_i) + \epsilon_{it}$ where y_{it} is the log of labor or revenue, α_i and β_t are firm and year fixed effects, and $\gamma_k \cdot \mathbf{1}(k = t - \tau_i)$ is an indicator variable equal to one for firms that have an IT spike k years before/after the spike (the omitted category is the year before the spike). The coefficients γ_k measure the difference in the trend of employment or revenue before (k<0) and after ($k \geq 0$) the IT spike, *ceteris paribus*. Although our regressions include the year-relative-to-spike dummies for all the pre- and post-spike time periods (from year -22 to 21), few observations are available far from the spike time, leading to large standard errors. So we report only the most central coefficients.

[19] We test 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 at conventional levels. 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.

[20] Using a simple difference-in-differences specification similar to the one in column (2) of table 2 with log mean wage as the outcome variable, the treatment coefficient is .003 (.020) in our sub-sample of 3,502 observations.

between costs and revenue. Our estimates suggest that the main impact of IT spikes is to increase markups—a decrease in β would reduce firm employment and increase the capital labor ratio, all else equal. Instead, we find that employment increases and the capital labor ratio does not change following IT spikes.[21] Moreover, although it is a noisy measure, the operating margin is positively correlated with IT spikes.[22] Thus, it appears that the main impact of major proprietary IT investments is to increase firm market power rather than to displace labor.

In Table A9 in the Appendix, we report on a number of robustness checks, estimating variations of the model in column (2) of Table 2. 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 estimate a model discarding the control group and exploiting only the timing of the IT spikes for the subsample of spikers for identification. Third, we re-define the spikes using a higher threshold in the growth of the IT share of employment (50% instead of 30%). Fourth, we re-define the IT spikes requiring that an IT spike is not only the largest in relative terms for a firm, but it also represents an increase in the number of software developers by at least 10 employees (column (4)). Fifth, we drop all firms that spike more than once during the sample period. Sixth, we drop firm-years characterized by very high or low employment or revenue growth (those in the top and bottom 1% of the distributions of employment or revenue growth). Finally, one may be concerned that the IT spikes might occur around

[21] Estimating a model similar to the model in column (2) of table 2 with the log of the capital labor ratio as outcome, we obtain a coefficient of the IT spike equal to .007, with a standard error of .0132 (p-value= 0.592).

[22] In an unreported model, we regress the operating margin (trimmed of the 1% tails) against the post-IT-spike dummy, controlling for firm and year fixed effects. This regression shows that the IT spike is associated with a statistically significant increase of .016 (standard error .005) in operating margin. This regression is sensitive to extreme values and the extent of winsorizing.

mergers and acquisitions or changes in other company characteristics. So we control for the number of M&A events that involved the focal firm in each year and for two time-varying company characteristics such as adverting and research and development expenses.[23] Although the magnitude of the coefficient of the IT spike changes, the coefficient of the IT spike is positive and statistically significant at 1% in all these additional models for non-IT labor, and in all but one of the models for revenue, where the it is statistically significant at 10%.[24]

Heterogeneity of effects

Table 3 and Table 4 explore the heterogeneity of the response of employment and revenue to the IT spikes. To explore heterogeneous responses to different values of categorical variable *z*, we estimate specifications based on

(8)

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

where Y_{it} is either employment or revenue.

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

[23] We take the logs of one plus advertising expenses and research and development expenses to use firm-year observations that contain zeros for these variables.

[24] Another concern is that IT spikes might happen when there is change in the top management of the company, and our regressions are picking up the effect of these changes. However, CEO changes are not correlated with the IT spikes in the subsample of firms for which we have information on the identity of CEOs. Specifically, the correlation between IT spikes and a CEO change in the same year or in one of the previous four years is between -0.0221 and 0.0087 (-0.0149 and 0.0026 for the largest spike, i.e. the one we use in our regressions).

developers at any point in time in our sample period. We find that the AI-users 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. However, 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 both employment and revenue, on average. Nevertheless, those that develop software only for internal use show strong growth.[27]

Panel C looks at changes in response over time. This is important because aggregate productivity grows when firms increase their productivity *and* when firms that are more productive 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. Consistent with this idea, our analysis shows that the response of employment and revenue to the IT spikes declined sharply.[28]

[25] A Wald test rejects the null hypothesis of equality of the coefficients for AI-users and for other firms at 1% level in both columns.

[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] A Wald test rejects the null hypothesis of equality of the coefficients for IT-producers and for other firms at 1% and 10%, respectively, for the labor and revenue models.

[28] Wald tests of the equality of these coefficients reject the null hypothesis at 1% and 5% respectively for the labor and revenue models.

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 US firms grow more both in terms of employment and revenue after the IT spikes, although we cannot reject the null hypothesis of the equality of the coefficients for US and non-US firms at conventional levels of statistical significance.

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 employment and revenue after the IT spikes for new firms.

Table 4 compares several industry sectors. There is significant heterogeneity. In particular, manufacturing industries exhibit much smaller or even negative labor and revenue growth than do tertiary sector industries such as trade, FIRE (Finance, Insurance and Real Estate), and services. A possible explanation is that firms in manufacturing industries may have lower price elasticities of demand. Bessen (2020a) suggests that demand elasticity in manufacturing industries was initially high but declined as these industries were progressively automated; tertiary sector industries may have higher demand elasticity in part because they have experienced far less productivity-improving technical change.

Conclusion

In recent decades, firms have sharply increased their investments in intangibles in general and in proprietary information technology in particular, including investments in artificial intelligence and big data. Furthermore, these new technologies have dramatic new capabilities, raising fears that they may lead to large scale displacement of workers.

We have developed a novel way to measure the adoption of major proprietary IT. In contrast to routine investments, major new information technology tends to occur in discrete episodes marked by substantial hiring of software developers. We use these events to measure the effects of adopting this technology, controlling for time-invariant firm characteristics as well as unobserved productivity and demand shocks. Contrary to common fears about machines replacing humans, we find that major firm investments in custom information technology tend to increase firm employment and revenues. Moreover, this effect is substantially stronger for firms using AI. The evidence shown here does not support the view that these technologies are mainly about taking over tasks performed by humans.

However, IT spikes do seem to decrease labor's share of output. Revenue increases faster than labor, suggesting that firm markups increase. This is a natural result because proprietary IT presumably earns quasi-rents and it is consistent with evidence that IT is related to higher markups and greater industry concentration (Calligaris, Criscuolo, and Marcolin 2018; J. Bessen 2020b). On the other hand, the capital-labor ratio does not rise, implying that this is not biased technical change as in Acemoglu and Restrepo (2018).

Different types of firms respond very differently following IT spikes. Employment grows faster following a spike in firms that use AI, in new firms, and in US-based firms, perhaps because of differences in management quality (Bloom, Sadun, and Van Reenen 2012). Mature industries such as non-durable manufacturing experience employment declines while the trade, services, and financial sectors see strong growth, perhaps evidence of differences in demand elasticity (J. Bessen 2020a). Also, employment growth following spikes seems to have slowed after the midpoint of our sample in 2002, consistent with other evidence on firm growth dynamics (Decker et al. 2018).

Even though IT spikes appear to increase employment, other automation technologies might well reduce employment at the firm level.[29] And even though IT spikes increase firm employment, they might reduce employment in the aggregate economy. These investments surely affect rival firms, firms in downstream and upstream industries, and consumers. Also, our research design does not necessarily capture the effects of proprietary investment at the largest multi-division firms because spikes at individual business units may be obscured in aggregate numbers. Despite the limitations of this study, large custom investments in firm-specific IT appear to be an important aspect of the role of technology in the economy.

[29] Bessen et al. (2019) find that automation expenditures have a different effect on worker separations than computer investments.

References

Acemoglu, Daron, Claire Lelarge, and Pascual Restrepo. 2020. "Competing with Robots: Firm-Level Evidence from France." In *AEA Papers and Proceedings*, 110:383–88.

Acemoglu, Daron, and Pascual Restrepo. 2018. "Modeling Automation." *AEA Papers and Proceedings* 108: 48–53. https://doi.org/10.1257/pandp.20181020.

———. 2020. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy* 128 (6).

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica* 83 (6): 2411–2451.

Aghion, Philippe, Céline Antonin, Simon Bunel, and Xavier Jaravel. 2020. "What Are the Labor and Product Market Effects of Automation? New Evidence from France."

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *The Quarterly Journal of Economics* 135 (2): 645–709.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content Of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118 (4): 1279–1333.

Bessen, James. 2020a. "Automation and Jobs: When Technology Boosts Employment." *Economic Policy* forthcoming.

———. 2020b. "Information Technology and Industry Concentration." *Journal of Law and Economics* forthcoming.

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

Bessen, James, Maarten Goos, Anna Salomons, and Wiljan van den Berge. 2020. "Firm-Level Automation: Evidence from the Netherlands." In *AEA Papers and Proceedings*, 110:389–93.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. "Americans Do IT Better: US Multinationals and the Productivity Miracle." *American Economic Review* 102 (1): 167–201.

Bonfiglioli, Alessandra, Rosario Crinò, Harald Fadinger, and Gino Gancia. 2020. "Robot Imports and Firm-Level Outcomes."

Brynjolfsson, Erik, and Lorin Hitt. 1996. "Paradox Lost? Firm-Level Evidence on the Returns to Information Systems Spending." *Management Science* 42 (4): 541–558.

Brynjolfsson, Erik, and Lorin M Hitt. 2003. "Computing Productivity: Firm-Level Evidence." *Review of Economics and Statistics* 85 (4): 793–808.

Calligaris, Sara, Chiara Criscuolo, and Luca Marcolin. 2018. "Mark-Ups in the Digital Era."

Correia, Sergio. 2018. "REGHDFE: Stata Module to Perform Linear or Instrumental-Variable Regression Absorbing Any Number of High-Dimensional Fixed Effects."

Dauth, Wolfgang, Sebastian Findeisen, Jens Südekum, and Nicole Wößner. 2017. "German Robots – The Impact of Industrial Robots on Workers." Institute for Employment Research, Nuremberg, Germany.

Decker, Ryan A, John C Haltiwanger, Ron S Jarmin, and Javier Miranda. 2018. "Changing Business Dynamism and Productivity: Shocks vs. Responsiveness." National Bureau of Economic Research.

Domini, Giacomo, Marco Grazzi, Daniele Moschella, and Tania Treibich. 2019. "Threats and Opportunities in the Digital Era: Automation Spikes and Employment Dynamics." *LEM Working Paper Series, Scuola Superiore Sant'Anna*.

Doms, Mark E., and Timothy Dunne. 1998. "Capital Adjustment Patterns in Manufacturing Plants." *Review of Economic Dynamics* 1 (2): 409–29. https://doi.org/10.1006/redy.1998.0009.

Frey, Carl Benedikt, and Michael A. Osborne. 2017. "The Future Of Employment: How Susceptible Are Jobs To Computerisation?" *Technological Forecasting and Social Change* 114 (C): 254–80.

Ge, Chunmian, Ke-Wei Huang, and Ivan PL Png. 2016. "Engineer/Scientist Careers: Patents, Online Profiles, and Misclassification Bias." *Strategic Management Journal* 37 (1): 232–253.

Graetz, Georg, and Guy Michaels. 2018. "Robots at Work." *Review of Economics and Statistics* 100 (5): 753–768.

Haltiwanger, John, Russell Cooper, and Laura Power. 1999. "Machine Replacement and the Business Cycle: Lumps and Bumps." *American Economic Review* 89 (4): 921–46.

Harrigan, James, Ariell Reshef, and Farid Toubal. 2016. "The March of the Techies: Technology, Trade, and Job Polarization in France, 1994-2007." National Bureau of Economic Research.

Hicks, John. 1963. *The Theory of Wages*. Springer.

Humlum, Anders. 2019. "Robot Adoption and Labor Market Dynamics." Working Paper.

Jin, Wang, and Kristina McElheran. 2017. "Economies Before Scale: Survival and Performance of Young Plants in the Age of Cloud Computing." *Rotman School of Management Working Paper*, no. 3112901.

Jorgenson, Dale W, Mun S Ho, Kevin J Stiroh, and others. 2005. "Productivity, Volume 3: Information Technology And The American Growth Resurgence." *MIT Press Books* 3.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline of the Labor Share." *The Quarterly Journal of Economics* 129 (1): 61–103.

Koch, Michael, Ilya Manuylov, and Marcel Smolka. 2019. "Robots and Firms."

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies* 70 (2): 317–341.

Marschak, Jacob, and William H Andrews. 1944. "Random Simultaneous Equations and the Theory of Production." *Econometrica, Journal of the Econometric Society*, 143–205.

Nilsen, Oivind Anti, and Fabio Schiantarelli. 2003. "Zeros and Lumps in Investment: Empirical Evidence on Irreversibilities and Nonconvexities." *The Review of Economics and Statistics* 85 (4): 1021–37. https://doi.org/10.1162/003465303772815907.

Oliner, Stephen D, and Daniel E Sichel. 2000. "The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?" *Journal of Economic Perspectives* 14 (4): 3–22.

Olley, G Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64 (6): 263–97.

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

Rothschild, Michael. 1971. "On the Cost of Adjustment." *The Quarterly Journal of Economics* 85 (4): 605–622.

Rovigatti, Gabriele, and Vincenzo Mollisi. 2018. "Theory and Practice of Total-Factor Productivity Estimation: The Control Function Approach Using Stata." *The Stata Journal* 18 (3): 618–662.

Tambe, Prasanna. 2014. "Big Data Investment, Skills, and Firm Value." *Management Science* 60 (6): 1452–1469.

Tambe, Prasanna, and Lorin M Hitt. 2012. "The Productivity of Information Technology Investments: New Evidence from IT Labor Data." *Information Systems Research* 23 (3-part-1): 599–617.

Winick, Erin. 2018. "Every Study We Could Find on What Automation Will Do to Jobs, in One Chart." *Technology Review. Https://Www. Technologyreview. Com/s/610005/Every-Study-We-Could-Find-on-Whatautomation-Will-Do-to-Jobs-in-One-Chart.*

Tables and Figures

Table 1: Production function estimates

Outcome	Log revenue					
Model	OLS (1)	OLS FE (2)	OP (3)	ACF investment (4)	LP (5)	ACF materials (6)
Log IT labor	0.080**	-0.005	0.056**	0.083**	0.059**	0.064**
	(0.011)	(0.010)	(0.011)	(0.000)	(0.005)	(0.000)
Log non-IT labor	0.579**	0.653**	0.575**	0.569**	0.324**	0.331**
	(0.017)	(0.029)	(0.018)	(0.000)	(0.009)	(0.001)
Log capital	0.302**	0.206**	0.135**	0.309**	0.195**	0.165**
	(0.009)	(0.018)	(0.047)	(0.000)	(0.020)	(0.013)
Log materials					0.340**	0.522**
					(0.014)	(0.006)
Observations	47,086	47,003	43,897	43,897	44,143	44,143
Firms	4,083	4,000	3,830	3,830	3,955	3,955

The unit of observation is firm-year. Model (2) has firm and year fixed effects. Model (3) uses the Olley and Pakes (1996) model; Model (4), our preferred specification, uses the Ackerberg, Caves, and Frazer method (2015) with investment as the proxy variable. Models (5) and (6) use imputed log materials as the proxy variable, model (5) with the Levinsohn-Petrin method and model (6) with the Ackerberg, Caves, and Frazer method. Estimates in (3)-(6) include adjustments for firm exit. Robust standard errors clustered by firm in parentheses (for models (3)-(6), based on 50 bootstrap replications). ** p<0.01, * p<0.05

Table 2: Employment and Revenue Equation Estimates

Model	(1) No firm FE	(2) FE	(3) Omega and K/L	(4) Interactions (marginal effects)
Panel A. Outcome: Log non-IT employees				
Post spike	0.596**	0.070**	0.071**	0.077**
	(0.053)	(0.017)	(0.018)	(0.019)
Productivity/demand shock ($\widehat{\omega}$)			-0.211	0.504
			(0.301)	(0.310)
Log capital / non-IT labor (κ)			-0.091**	-0.068*
			(0.033)	(0.033)
Spiker	-1.125**			
	(0.082)			
Observations	49,991	49,967	43,823	43,823
R-squared	0.049	0.942	0.943	0.944
Firms	4,242	4,218	3,756	3,756
Panel B. Outcome: Log revenue				
Post spike	0.628**	0.109**	0.107**	0.109**
	(0.056)	(0.018)	(0.019)	(0.019)
Productivity/demand shock ($\widehat{\omega}$)			0.613*	0.818**
			(0.258)	(0.262)
Log capital / non-IT labor (κ)			0.172**	0.169**
			(0.025)	(0.025)
Spiker	-1.239**			
	(0.088)			
Observations	50,205	50,202	43,823	43,823
R-squared	0.051	0.939	0.940	0.941
Firms	4,203	4,200	3,756	3,756

The unit of observation is firm-year. All models are estimated with OLS and include year effects. Columns 2-4 also include firm fixed effects. First stage estimation of $\widehat{\omega}$ uses the ACF method with investment as the proxy variable. Column 3 reports the coefficient estimates from equation (7a). Column 4 shows the marginal effects from the specification in equation (7b) (coefficients in Appendix Table A8). Robust standard errors clustered by firm in parentheses (for models (3)-(4), based on 50 bootstrap replications). ** p<0.01, * p<0.05

Table 3: Heterogeneous effects of IT spikes

Outcome	(1) Log non-IT employees	(2) Log revenue
Panel A: AI/Big Data		
Post spike not AI	0.052**	0.095**
	(0.018)	(0.018)
Post spike AI	0.268**	0.271**
	(0.064)	(0.065)
Panel B: IT producing		
Post spike not IT	0.042*	0.089**
	(0.020)	(0.020)
Post spike IT	0.170**	0.179**
	(0.040)	(0.042)
Panel C: Time period		
Post spike pre-2002	0.128**	0.154**
	(0.027)	(0.027)
Post spike post-2002	0.005	0.060*
	(0.025)	(0.027)
Panel D: US based		
Post spike not US	0.055	0.055
	(0.043)	(0.041)
Post spike US	0.072**	0.117**
	(0.019)	(0.019)
Panel E: New firms		
Post spike old firm	-0.016	0.007
	(0.022)	(0.021)
Post spike new firm	0.291**	0.366**
	(0.027)	(0.032)

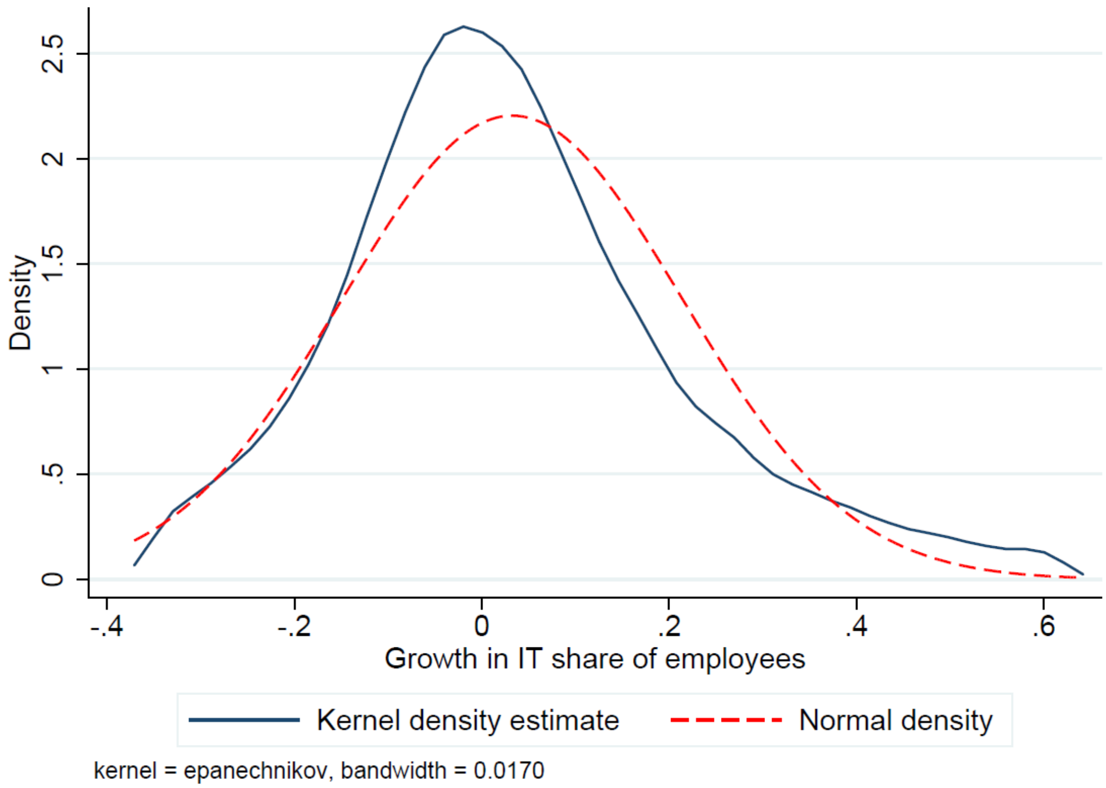
The unit of observation is firm-year. All models are estimated with OLS, and include year and firm fixed effects. Robust standard errors clustered by firm in parentheses. ** p<0.01, * p<0.05

Table 4: Heterogeneity by industry

	(1)	(2)
Outcome	Log non-IT employees	Log revenue
Post spike nondurable manufacturing	-0.090*	-0.019
	(0.044)	(0.048)
Post spike durable manufacturing	0.024	0.060
	(0.032)	(0.033)
Post spike transport and utilities	-0.028	0.142*
	(0.059)	(0.056)
Post spike trade	0.209**	0.175**
	(0.053)	(0.068)
Post spike finance	0.146**	0.229**
	(0.051)	(0.040)
Post spike service	0.172**	0.141**
	(0.043)	(0.046)
Post spike others	0.129	0.099
	(0.075)	(0.074)

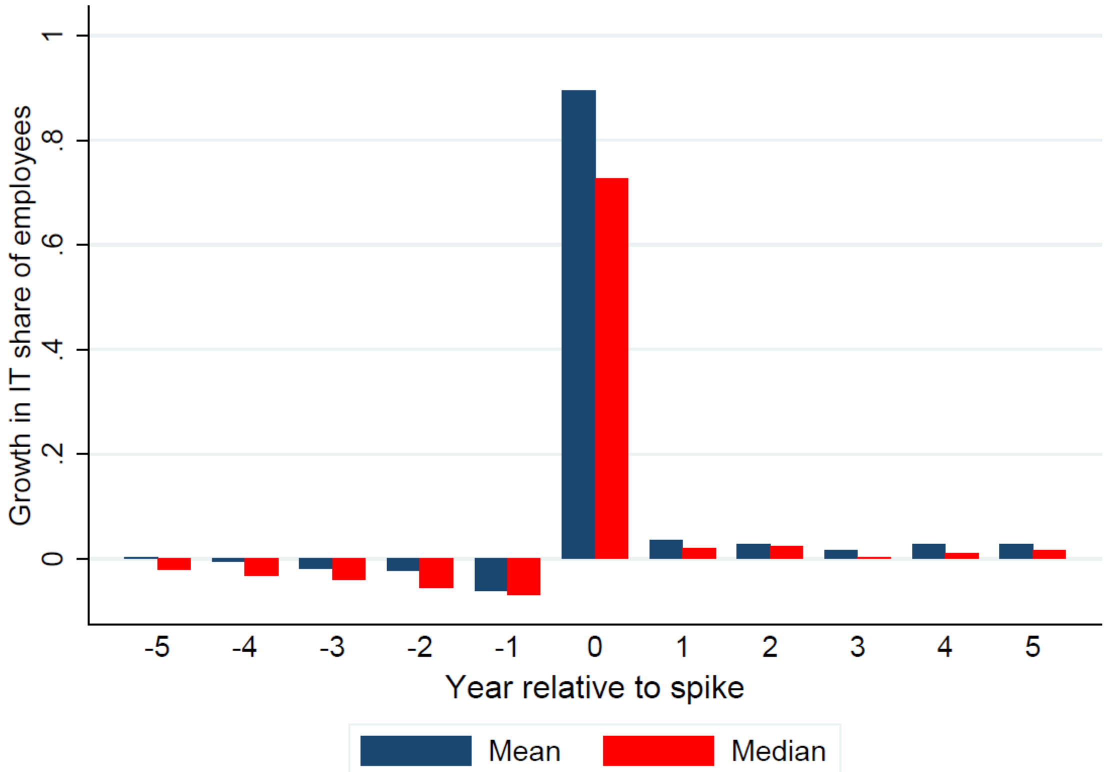
The unit of observation is firm-year. All models are estimated with OLS, and include year and firm fixed effects. Robust standard errors clustered by firm in parentheses. ** $p<0.01$, * $p<0.05$

Figure 1: Distribution of growth in IT share of employment



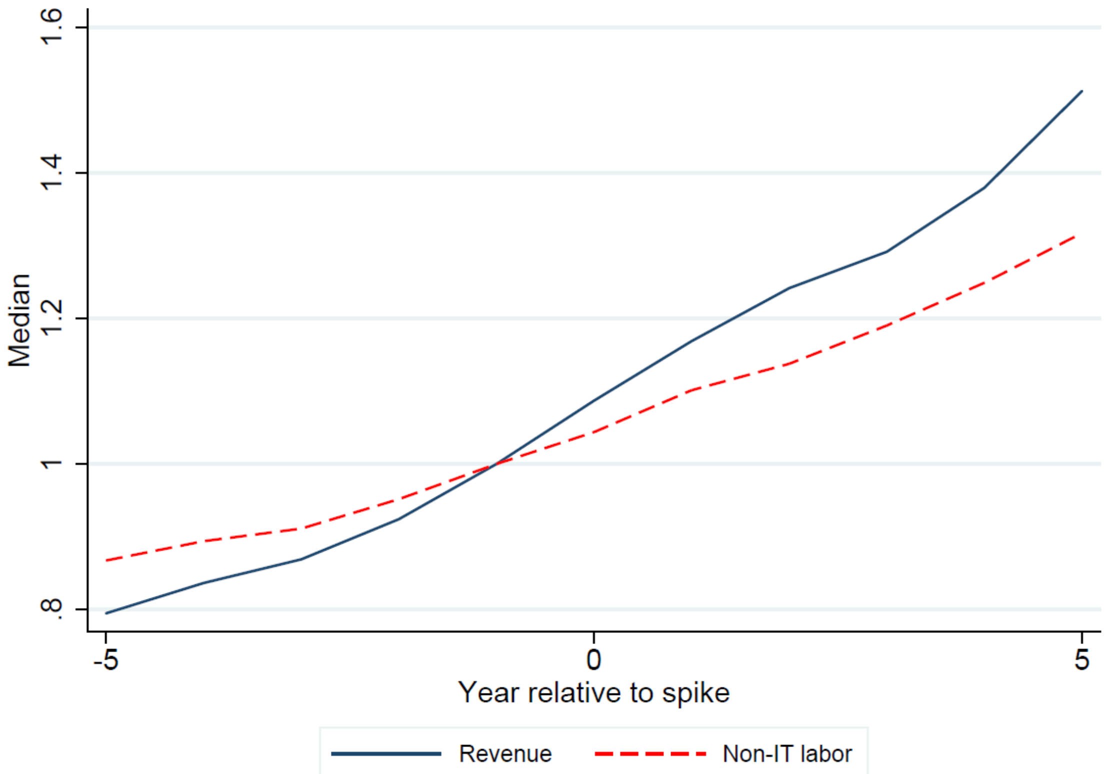
Notes. The figure plots the kernel density estimate of growth in IT share of employees (solid line) and compares it with a normal distribution (dashed line).

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



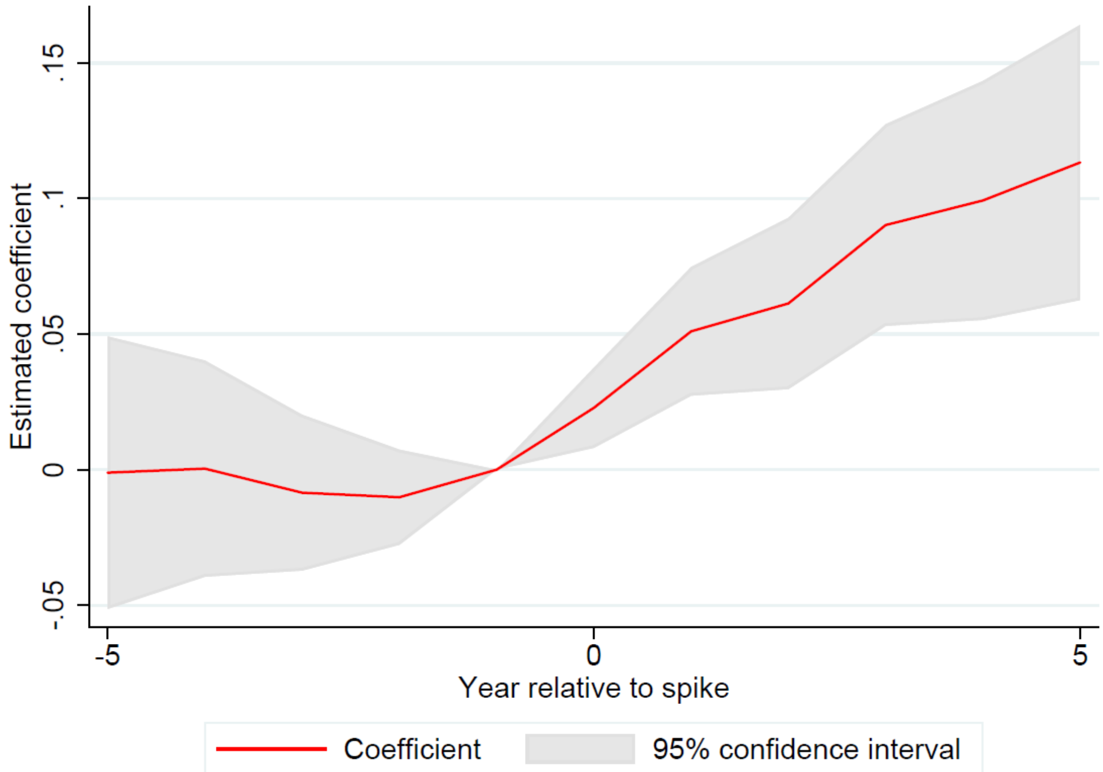
Notes. The figure plots the mean and the median of the growth in IT share of employees in an 11-year window around the year of the highest spike for firms that spikes at least once.

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



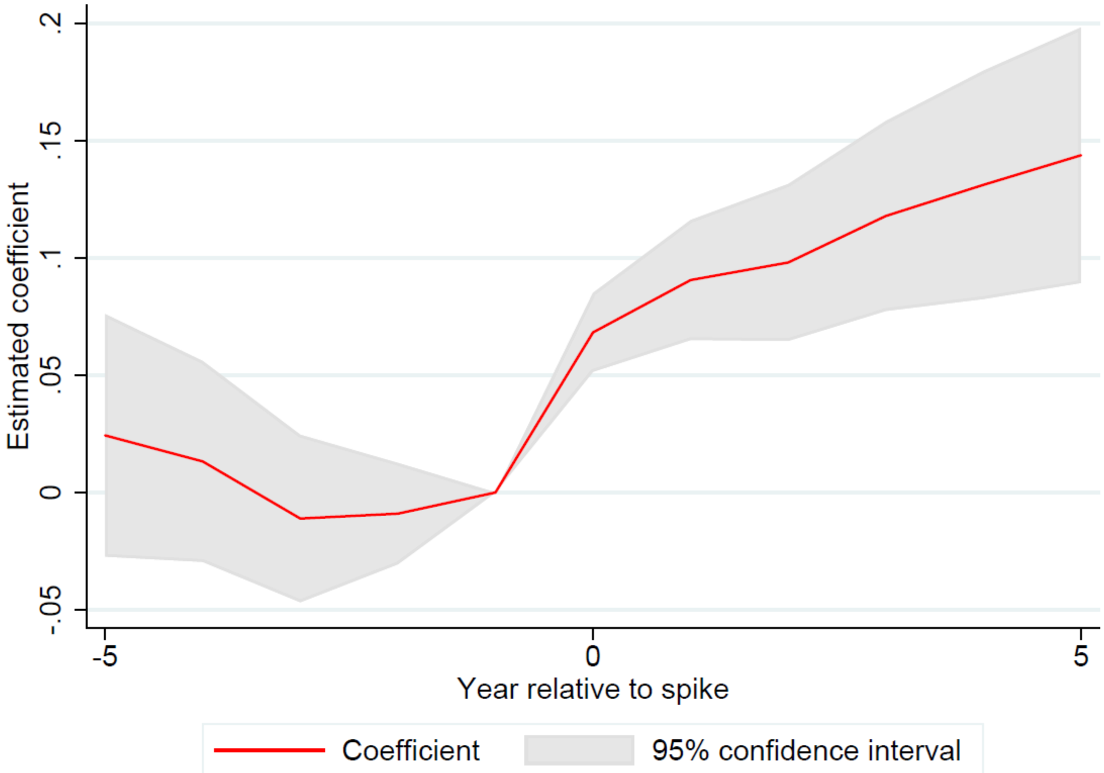
Notes. The figure plots the medians of revenue and non-IT employees in an 11-year window around the year of the highest spike for firms that spikes at least once. Revenues and non-IT employees in each year are normalized dividing their value in the current year by the value in the year before highest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

Figure 4: trend in non-IT employees around spike, difference-in-differences



Notes. This figure plots the coefficients (solid line) and the 95% confidence intervals (shaded area) of a set of year-relative-to-spike indicators 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 outcome variable is the log of non-IT employees. The model includes year and firm fixed effects. Robust standard errors are clustered by firm.

Figure 4: trend in revenues around spike, difference-in-differences estimates

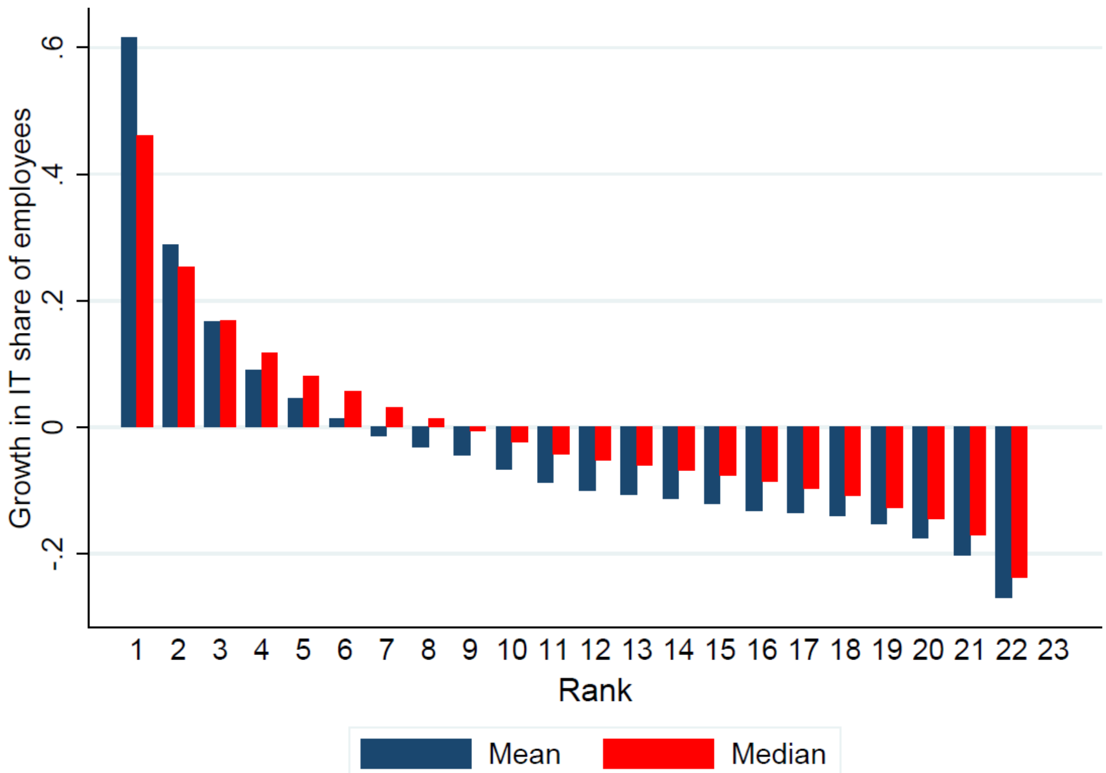


Notes. This figure plots the coefficients (solid line) and the 95% confidence intervals (shaded area) of a set of year-relative-to-spike indicators from an OLS regressions similar to model (2) in table 2. The omitted category is the year before the spike. While the regression includes the full sets of 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 outcome variable is the log of revenues. The model includes year and firm fixed effects. Robust standard errors are clustered by firm.

Appendix

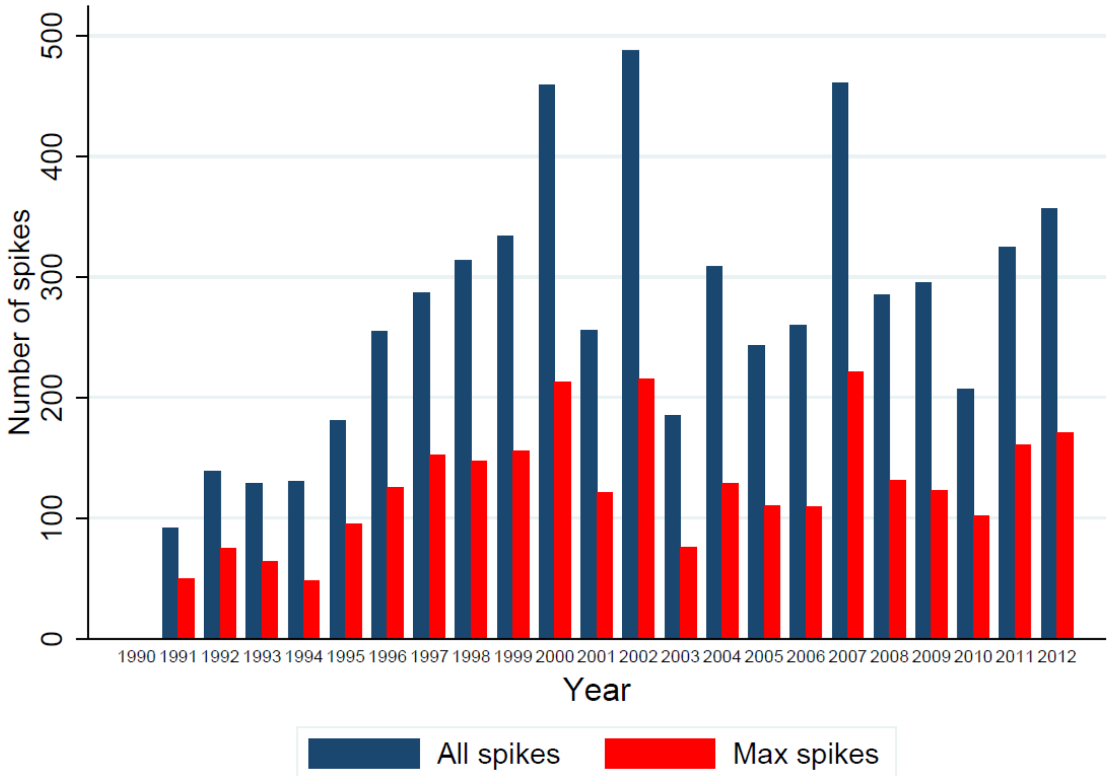
Supplementary Figures and Tables

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



Notes. 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



Notes. 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)
Revenue (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		

Notes. 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
IT employees	50,108	0.3	1.1	0.0	32.7
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
Materials	48,599	2,939.7	12,291.3	0.0	389,757.0
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
US firm	51,382	0.9	0.3	0.0	1.0
M&As	51,382	0.8	1.8	0.0	60.0
Advertising expenses	51,382	57.3	329.5	0.0	9,016.1
R&D Expenses	51,382	123.0	631.4	0.0	14,046.2
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
CEO change	22,477	0.1	0.3	0.0	1.0

Notes. 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%

Notes. Unit of observation is firm-year.

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

Notes. 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	0.18	0.09	29.25	-0.26
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		

Notes. 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

Notes. 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 A7: Sensitivity to alternative production function estimates

Model	(1) OP	(2) ACF/invest	(3) LP	(4) ACF/materials
Panel A. Outcome: Log non-IT employees				
Post spike	0.045**	0.071**	0.046*	0.059**
	(0.012)	(0.018)	(0.018)	(0.019)
Productivity/demand shock ($\hat{\omega}$)	3.020**	0.262	1.191	-0.124
	(0.432)	(0.301)	(0.645)	(0.295)
Log capital / non-IT labor (κ)	-0.503**	-0.065*	-0.013	-0.048
	(0.083)	(0.033)	(0.044)	(0.029)
Observations	43,823	43,823	44,021	44,021
R-squared	0.975	0.943	0.953	0.947
Firms	3,756	3,756	3,833	3,833
Panel B. Outcome: Log revenue				
Post spike	0.085**	0.107**	0.076**	0.096**
	(0.014)	(0.019)	(0.018)	(0.020)
Productivity/demand shock ($\hat{\omega}$)	2.745**	0.613*	2.073**	0.175
	(0.353)	(0.258)	(0.576)	(0.297)
Log capital / non-IT labor (κ)	-0.248**	0.172**	0.185**	0.150**
	(0.076)	(0.025)	(0.025)	(0.026)
Observations	43,823	43,823	44,021	44,021
R-squared	0.962	0.940	0.957	0.943
Firms	3,756	3,756	3,833	3,833

Notes. The unit of observation is firm-year. Showing second stage estimates where first stage uses the indicated method to estimate production function. All models contain firm and year fixed effects. All models estimated with the Stata package developed by Mollisis and Rovignati (2016) in the first stage and OLS in the second. Robust standard errors clustered by firm in parentheses. ** p<0.01, * p<0.05

Table A8: Interaction Regressions

	(1)	(2)
Outcome	Log non-IT employees	Log revenue
Post spike	-1.325*	-0.819
	(0.543)	(0.606)
Productivity/demand shock ($\widehat{\omega}$)	0.195	0.582
	(0.333)	(0.302)
Post spike x shock ($D \cdot \widehat{\omega}$)	0.298**	0.202
	(0.107)	(0.121)
Log capital / non-IT labor (κ)	0.248	0.448**
	(0.128)	(0.091)
Post spike x capital/labor ($D \cdot \kappa$)	-0.006	-0.010
	(0.016)	(0.016)
Capital/labor squared (κ^2)	-0.041**	-0.036**
	(0.015)	(0.011)
Observations	43,823	43,823
R-squared	0.944	0.941
Firms	3,756	3,756

Notes. The unit of observation is firm-year. Models estimated with ACF in the first stage and OLS in the second and include firm and year fixed effects. First stage estimation of $\widehat{\omega}$ uses the ACF method with investment as the proxy variable. Bootstrapped standard errors clustered by firm in parentheses (50 repetitions). ** p<0.01, * p<0.05

Table A9: Employment and Revenue Equation Estimates, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Model	Industry-by-year	Spikers	Higher spikes	At least 10 more IT employees	No multi-spikers	No growth tails	Time-varying controls
Panel A. Outcome: Log non-IT employees							
Post spike	0.065**	0.068**	0.060**	0.069**	0.106**	0.074**	0.074**
	(0.018)	(0.016)	(0.021)	(0.025)	(0.028)	(0.017)	(0.016)
Observations	48,160	36,599	49,967	49,967	24,993	49,041	49,967
R-squared	0.955	0.930	0.942	0.942	0.956	0.946	0.949
Firms	4,146	2,775	4,218	4,218	2,539	4,198	4,218
Panel B. Outcome: Log revenue							
Post spike	0.098**	0.118**	0.088**	0.042	0.170**	0.110**	0.113**
	(0.019)	(0.017)	(0.021)	(0.025)	(0.030)	(0.017)	(0.016)
Observations	48,447	36,859	50,202	50,202	25,045	49,238	50,202
R-squared	0.953	0.924	0.939	0.938	0.953	0.943	0.945
Firms	4,129	2,757	4,200	4,200	2,533	4,158	4,200

Notes. The unit of observation is firm-year. All models are estimated with OLS, and include firm fixed effects. Sample for column (2) contains only firms with an IT spike. Sample for column (5) excludes firms with more than one IT spike. Sample for column (6) excludes firms with at least one year in the top 1% or bottom 1% of the distribution of non-IT employment (panel A) or revenue (panel B) in the sample. In column (3) the post-spike indicator is computed using a 50% threshold. In column (4) the post-spike indicator is computed using a 30% threshold and an increase in the number of IT employees by at least 10 employees. Models in column (1) includes industry-by-year effects (SIC 4-digit). Models (2)-(7) include year effects. Model (7) includes also the annual count of M&As, and the logs of one plus advertising expense and one plus research and development expenses. Robust standard errors clustered by firm in parentheses. ** p<0.01, * p<0.05

Labor demand equation

From the first order condition for labor and the price elasticity of demand,

(A1)

$$L = \frac{\beta Y}{w\mu} = \frac{\beta p^{1-\epsilon} d^{\epsilon}}{w\mu}.$$

We can obtain an expression for *p* using Euler's theorem:

$$Q = \frac{\partial Q}{\partial L} L + \frac{\partial Q}{\partial K} K$$

so that inserting the first order conditions,

(A2)

$$p = \frac{\mu(wL + rK)}{Q} = \mu \cdot c = \frac{\mu}{A} \frac{wL + rK}{(L^{\beta} K^{1-\beta})} = \frac{\mu r}{A} \left(\frac{K}{L}\right)^{\beta} \left(\frac{wL}{rK} + 1\right)$$

where *c* is the average unit cost, implying that $\mu = \frac{p}{c} = \frac{Y}{c \cdot Q}$ is the average markup. From the first order conditions,

(A3)

$$\frac{K}{L} = \frac{1-\beta}{\beta} \frac{w}{r}.$$

Substituting (A2) and (A3) into the log of (A1) and defining $\omega \equiv \frac{\ln A}{\mu} + \ln d$,

(A4)

$$l = f^{1}(\epsilon) + \epsilon \cdot \omega - (\epsilon - 1) f^{2}(\beta) + \ln \beta - \ln \frac{w}{r}$$

$$f^{1}(\epsilon) = -\epsilon \left(\ln r + \ln \frac{\epsilon}{\epsilon - 1}\right)$$

$$f^{2}(\beta) = \beta \ln \frac{1-\beta}{\beta} \frac{w}{r} - \ln(1-\beta).$$