Computer networks and productivity revisited: Does plant size matter? Evidence and implications

Henry R. Hyatt* and Sang V. Nguyen
Center for Economic Studies, U.S. Census Bureau, Washington, DC, USA

Numerous studies have documented a positive association between information technology investments and business- and establishment-level productivity. Most of these studies, however, rely on empirical specifications that over-represent small businesses. In this paper, we revisit one piece of evidence, the Computer Network Use Supplement to the 1999 U.S. Annual Survey of Manufactures, which has previously been used to show that there is a positive relationship between computer networks and productivity in manufacturing plants. We show that this is only true for small- and medium-sized plants, and that for larger plants the relationship is negative. We give critical consideration to alternative methods for weighting these data, and show that employment-weighted estimates indicate the presence of a computer network has, on average, a negative relationship with the productivity of employees.

Keywords: Computer network, productivity, size, employment, information technology, manufacturing, internet

1. Introduction

During the last decade, an increasing number of empirical studies have explored the relationship between information technology (IT) investments and economic performance, and most find that IT investments are a significant source of additional productivity. For example, Jorgensen and Stiroh [13], Jorgenson [12], Stiroh [15], and Triplett and Bosworth [21] use industry-level data to demonstrate that IT plays an important role during the surge of productivity growth in the U.S. economy in the late 1990s. While in aggregate, U.S. labor productivity clearly grew during the 1990s while businesses were investing in IT, the story becomes more complex when studied at the business or establishment level. One crucial limitation, which we address in this study, is that most previous researchers have used empirical specifications that give all firms or plants equal weight. This method treats the relatively infrequent large businesses with the same weight as that of the much more numerous small plants, which only yield consistent estimates of the effects of IT investments on productivity if they affect businesses similarly, regardless of size.

---

*Corresponding author: Henry R. Hyatt, Center for Economic Studies, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233, USA. Tel.: +1 301 763 7884; Fax: +1 301 763 5935; E-mail: Henry.R.Hyatt@Census.gov.

Under this (generally implicit) assumption of homogenous effects by size, IT investments have been associated with productivity increases. In a study of large firms, Brynjolfsson and Hitt [6] find that computer investments yield large productivity returns over a five-year time horizon. Using panel data for the manufacturing industry, Dunne et al. [8] document that increases in computer investment are associated with increases in the plant-level wages and productivity. Shin [17] analyzes the relationship between software usage and productivity on a sample of 525 small to medium size firms, and finds evidence of enhanced productivity in both manufacturing and service firms. Atrostic and Nguyen [2,3] and Atrostic et al. [1] use plant-level data to study the relationship between computer network use and the labor productivity of manufacturing plants, and find that computer networks are associated with higher productivity.

Less attention has been devoted to the role of business size in the effect of IT investments on productivity, and there is almost no empirical evidence on this subject. A recent exception was provided by Tambe and Hitt [20], who explore tabulations that distinguish between the effects of the presence of different numbers of IT workers on the productivity of Fortune 500 vs. non-Fortune 500 firms, the former of which are in general larger. Otherwise, studies contain at most provide a discussion of mechanisms by which IT investments may produce different productivity returns for businesses of different sizes, but without empirical evidence. For example, Brynjolfsson et al. [7] assert that higher IT investments will lower "coordination costs" within and between firms that would lead to smaller, more specialized firms. Similarly, Iansiti et al. [11], discussing their sample of firms that employ between 100 and 500 employees, suggest that in larger firms, successful implementation of an IT infrastructure may be hindered by conflicts between the business units dedicated to IT and other units within a business. The implication of the listed mechanisms is that productivity returns to the implementation of an IT solution may be greater for small entities than large ones. In this study, we consider manufacturing plants operating in the United States in the late 1990s, and we offer evidence that a particular IT investment, a computer network, is associated with higher productivity for smaller business establishments (specifically, manufacturing plants) but not larger ones.

To do so, we exploit a unique survey, the Computer Network Use Supplement (CNUS) to the 1999 Annual Survey of Manufactures (ASM), linked with longitudinal data on plant size and productivity from the 1992 and 1997 Census of Manufactures (CM). The Supplement asked about whether manufacturing plants employed any sort of telecommunications infrastructure in which computers were able to interact with one another, including general connectivity through internet access, or something more specific to the location such as a local area network or intranet. This survey took place when most, but not all, manufacturing plants, had adopted some sort of computer network, and so it allows comparison between the technology adopters and non-adopters.

Atrostic and Nguyen [3] and related studies have used the CNUS linked with the CM to demonstrate that manufacturing plants with computer networks, on average, exhibit 5% higher productivity than those without a network. In this study, we

measure how the relationship between computer network adoption and productivity varies by plant size. We document that the relationship between computer networks is strongly positive only for the more numerous small- to medium-sized establishments in this dataset, but is strongly negative for larger plants. These different computer network productivity differentials have implications for how to interpret the previous finding that a computer network is associated with higher plant-level productivity. In this study, we argue that the Atrostic and Nguyen [3] regression specification, in which all plants are given equal weight, recovers the productivity difference associated with the presence of a computer network for the *typical plant*, but it does not represent the association for the *typical worker*, or the economy as a whole.

We give critical consideration to the implications of different weights that can be applied to plant-level regression specifications. Weighting all plants equally produces regression estimates that represent the relationship between computer networks and productivity that exists for the much more numerous smaller plants. However, large plants, by definition, constitute a share of aggregate labor productivity that is proportionate to their size. In order to recover the impact of computer networks on aggregate labor productivity, we estimate regression models in which plants are weighted by total employment. In order to do so, we construct weights that are appropriate to the ASM, which over-samples larger plants. We also consider the components of this product as alternative weights, the total employment of plants, as well as the ASM weights alone.[1] We find that both sets of employment-weighted estimates indicate a substantially different relationship between computer networks and labor productivity: estimates from the employment-weighted regressions indicate a strongly negative association between the presence of a computer network and productivity rather than the more commonly reported strongly positive one. This finding suggests that researchers should pay careful attention to weighting when estimating the productivity impacts of technology adoption.

We would like to caution readers that, in the 1999 ASM, nearly 90% of sampled plants have a computer network, although after applying the ASM sample weights this frequency drops to 77%. However, of the largest plants, 97% report operating a computer network (because the largest plants nearly all have weights of exactly one by construction, this fraction does not change materially after the application of ASM sample weights). In order to provide an additional empirical reference point in the presence of different levels of adoption of this technology among establishments of different sizes, we estimate regression specifications similar to those in Atrostic and Nguyen [3] separately for three broad size categories. However, this disaggregation of our results does not overcome the fundamental difficulty of conducting our empirical analysis using the 1999 ASM, which is that, by 1999, most manufacturing plants had adopted computer networks. Therefore, the cross-sectional differences in

---

[1] We thank an anonymous referee for this suggestion.

our sample of plants may not represent productivity differentials that are implicit in studies of longer time periods, such as the national accounts estimates in Jorgenson [12] or manufacturing plant-level estimates in Dunne et al. [8]. We also would like to caution readers that we only consider one sort of technology that businesses adopt, that is, computer networks, and that investments in this sort of technology may not be representative of technology investments more generally.

The paper proceeds as follows. In the next section, we describe our data and variables. Section 3 presents the empirical model and estimation method. We present the empirical results in Section 4. Section 5 concludes the paper.

2. Data sources and variable construction

2.1. Data sources

The Computer Network Use Supplement (CNUS) to the 1999 ASM provides the first large-scale picture of the prevalence and use of computer networks in U.S. manufacturing. The CNUS was sent to manufacturing plants in the year 2000, and asked about whether manufacturing plants had some computer network such as an internet, intranet, LAN, EDI, extranet or some other type of computer network. The survey confirmed that computer network adoption was widespread, see Mesenbourg [15]. The ASM is sent to U.S. manufacturing establishments, and the frequency with which firms are selected for the ASM varies by size: the largest establishments (those that employ 250 or more employees) are sampled with certainty, and those who employ fewer people are selected with a probability that is increasing in their size. The Census Bureau's Longitudinal Research Database maintains longitudinally consistent identifiers between successive waves of the ASM, as well as the CM, which is the quinquennial (1992, 1997, etc.) census of all manufacturing plants in the U.S.

We use data from the Supplement to the 1999 ASM data and any linked data from the 1992 and 1997 CM in order to obtain data for our empirical analysis. This link to the CM provides observable characteristics of the plants in previous years. The Supplement contains responses for 39,056 plants of which 30,313 merge with the 1992 and 1997 CM and have all values of the key variables in the estimation of labor productivity, total employment and total value of shipments, greater than zero. Deleting observations with total employment less than five and observations with missing values in 1997 and 1999 further reduces the sample to 29,177. Additional removal of any observations with missing values for any other variable in 1992 (to be used in the two-stage regressions) described below yields 27,007 plants.[2]

---

[2]For more details on deriving a sample from these data sources, see Atrostic and Nguyen [2,3] and Mesenbourg [15]. We deleted plants with 4 or fewer employees because data on these small plants are largely imputed based on administrative records. In our preliminary work, we included plants with less than 5 employees and found similar results.

2.2. Variable construction

Variables in our models are defined as follows. Subscript t takes the values 92, 97 and 99, which imply source data is from the 1992 and 1997 CM, and the 1999 ASM, respectively.

Output (Q_t): A plant's total value of shipments serves as a revenue proxy for output.

Labor (L_t): The labor input is the total number of employees working at the plant.[3]

Capital (K_t): The book value of a plant's gross capital stock (including buildings and machinery assets) serves as a proxy for capital services. While book values are likely to be subject to measurement error, this approach is well established and is used in, for example, McGuckin et al. [14], Greenan and Mairesse [10] and Atrostic and Nguyen [3]. Because this variable is only available in CM years, we use K_{97} (and, for calculating the capital-to-labor ratio, K_{97}/L_{97}) when estimating empirical specifications that otherwise employ data from the 1999 ASM.

Materials (M_t): The materials input is the sum of values of materials, parts, energy and contract work. Expenditures on materials is an important control variable when estimating the relationship between computer networks on productivity because it permits comparison between plants that have an in-house IT department and those who outsource their IT services. Outsourcing plants may have a lower number of individuals that are included in payroll and, therefore, raw labor productivity will increase. However, these same plants will have a higher materials expenditure to labor ratio. This control variable accounts for what might otherwise cause omitted variables bias in our regression specifications.

Labor Productivity (LP_t): Labor productivity is revenue output divided by labor (the number of employees), or Q_t/L_t.

Total Factor Productivity (TFP_{97}): In a robustness check for the validity of our labor productivity measure, we also use a multi-factor productivity measure. This variable was only available with a two year lag (the constraint is that the capital stock is only reported in a CM year), and so it is not our preferred specification. We compute a multi-factor measure of productivity that computes efficiency using cost-shares allocated to capital, labor, and materials assuming a Cobb-Douglas production technology as in our theoretical framework below and follows the standard method used by, for example, Baily et al. [4] and Olley and Pakes [16]. Specifically, we estimate

$$\text{Log}(Q_{97}) = \omega_0 + \omega_1 \log(L_{97}) + \omega_2 \log(K_{97}) + \omega_3 \log(M_{97}) \tag{1}$$

using data from the 1997 CM, where Q_{97} is the quantity of output, Q_{97} is the quantity of output, K_{97} is the book value of capital, and M_{97} is expenditure on materials. This

---

[3]We also used hours worked instead of total employment and obtained similar results. In keeping with Atrostic and Nguyen [3], we use total number of employees as a proxy for labor.

recovers the productivity parameters ω_1, ω_2, and ω_3, which are the estimated returns to labor, capital, and materials, respectively. We then plug in these globally estimated parameters into our data at the plant level to recover total factor productivity through the calculation

$$TFP_{97} = \text{Log}(Q_{97}) - \omega_1\text{log}(L_{97}) - \omega_2\text{log}(K_{97}) - \omega_3\text{log}(M_{97}). \qquad (2)$$

Computer Network (Network$_{99}$): This variable takes on a value of 1 if the plant's response to the CNUS indicated any kind of computer network, and 0 otherwise. The Supplement allowed respondents to report operating an internet, intranet, Local Area Network, Electronic Data Interchange, Extranet, or an "other" computer network.

Multi-unit firm status (Multi$_t$): This variable takes on the value of 1 if the plant is part of a multi-unit firm, and equals 0 otherwise.

Industry (Industry$_{t,j}$): We construct two different sets of dummy variables for 3-digit NAICS industries for 1999 and 2-digit SIC industries in 1992, that is, for each industry j.

Relative labor Productivity (RLP$_{92}$): This variable is the plant's labor productivity divided by its industry-average labor productivity, measured at the 2-digit SIC from the 1992 CM.

Computer Expenditures (Computer$_{92}$). This variable is defined as the computer expenditure of the plant divided by total number of employees in the 1992 CM.

Skill$_t$: We use the number of non-production workers divided by total number of employees in the plant as a proxy for skill.

Size$_{99,s}$: We define six size classes based on total employment in the 1999 ASM, L_t:

- *Size$_{99,1}$* = 1 if $5 \leqslant L_{99} < 50$, *Size$_{99,1}$* = 0, otherwise;
- *Size$_{99,2}$* = 1 if $50 \leqslant L_{99} < 100$, *Size$_{99,2}$* = 0, otherwise;
- *Size$_{99,3}$* = 1 if $100 \leqslant L_{99} < 250$, *Size$_{99,3}$* = 0, otherwise;
- *Size$_{99,4}$* = 1 if $250 \leqslant L_{99} < 500$, *Size$_{99,4}$* = 0, otherwise;
- *Size$_{99,5}$* = 1 if $500 \leqslant L_{99} < 1000$, *Size$_{99,5}$* = 0, otherwise;
- *Size$_{99,6}$* = 1 if $L_{99} \geqslant 1000$, *Size$_{99,6}$* = 0, otherwise.

Small, Medium and Large: In some estimates, we split the dataset into 3 size classes and estimate separate regressions by 1999 ASM categories, as follows:

- *Small*: $5 \leqslant L_{99} < 100$ employees;
- *Medium*: $100 \leqslant L_{99} < 250$ employees;
- *Large*: $L_{99} \geqslant 250$ employees.

We emphasize that it is impossible to offer universally accepted definitions for small, medium and large plants, and that the terms should be viewed as relative. However, our results are not particularly sensitive to the values used to define the categories.

Weight: We utilize three weights in this paper (we also present unweighted specifications implicitly assign a weight of one to all observations). The meaning and proper interpretation of these weights are discussed further in the next section, but we list them here for completeness:

- *ASM*: This is the ASM sample weight, and reflects the sampling rate and the influence of ex-post edits and non-response to ensure that *ASM*-weighted responses reflect the averages associated with the manufacturing sector or its different subcomponents;
- *Employment:* This is the employment count in the 1999 ASM, and is identical to the *Labor (L_t)* variable defined above;
- *ASM*Employment:* We calculate an "employment weight" as the 1999 ASM sample weight (not adjusted for non-response) multiplied by total employment.

3. The empirical model and estimation method

3.1. The empirical model

Following Atrostic and Nguyen [3], we specify a three-factor Cobb-Douglas production function, extended to incorporate plant characteristics as control variables, and estimate log-labor productivity with corresponding marginal effects, as follows:

$$
\begin{aligned}
\text{Log}(Q_{99}/L_{99}) = {} & \beta_0 + \beta_1 Network_{99} + \alpha_1 \log(K_{97}/L_{97}) + \\
& \alpha_3 \log(M_{99}/L_{99}) + \alpha_5 \log(Skill_{99}) + \alpha_6 Multi_{99} + \qquad (3) \\
& \sum \gamma_j Industry_{99,j} + \sum \lambda_s Size_{99,s} + \varepsilon,
\end{aligned}
$$

where the variables are as described above with corresponding marginal effects and ε is an error term. We estimate Eq. (1) using Ordinary Least Squares using SAS Version 9.[4]

Note that the empirical specifications in this paper differ from that in Atrostic and Nguyen [3], who measure size as a continuous variable, using total employment. Since we focus on the productivity differentials for plants of different sizes, we incorporate categorical variables for size. Another advantage of this approach over the use of total employment as a right-hand variable is to avoid potential bias from measurement error.[5]

As Atrostic and Nguyen [3] point out, the estimated coefficient for the *Network99* variable Eq. (1) is most likely subject to endogeneity bias.[6] Thus, following Atrostic

---

[4] SAS was employed because most Census Bureau datasets, including the CM, ASM, and CNUS, are stored as SAS datasets. For our regressions in SAS, we use the GLM procedure. Furthermore, all tabulations were done on a Linux server.

[5] We note, however, that using the size variables does not significantly affect the estimated coefficient of the Network variable.

[6] Readers may be tempted to refer to our two-step estimation procedure as a "causal" estimate. We caution that both regressions are essentially descriptive, and that the two-step estimator only corrects for selection bias implied by the functional form.

and Nguyen [3] we employ a two-step specification in which the presence of a computer network is related to its past characteristics according to the following probit regression:

$$Pr(\textit{Network}_{99}) = \Phi(\eta_0 + \eta_1 \log(\textit{RLP}_{92}) + \eta_2 \log(\textit{Skill}_{92}) +$$
$$\eta_3 \log(\textit{Computer}_{92}) + \eta_4 \textit{Multi}_{99} + \varphi_j \textit{Industry}_{92j}) \tag{4}$$

where variables are as described above with corresponding marginal effects.

We first estimate Eq. (2) and use its parameter estimates to calculate probability of a plant having a computer network, and use this in the estimation of Eq. (1) to account for possible endogeneity bias associated with the $\textit{Network}_{99}$ variable. Specifically, $Pr(\textit{Network}_{99}) = \Omega(-\textit{Networkhat})$, where Ω is the cumulative density function for the standard normal variable and $\textit{Networkhat}$ is the fitted value of $\textit{Network}_{99}$. In short, we estimate the equation

$$\text{Log}(Q_{99}/L_{99}) = \beta_0 + \beta_{TS}Pr(\textit{Network}_{99}) + \alpha_1\log(K_{97}/L_{97}) +$$
$$\alpha_3\log(M_{99}/L_{99}) + \alpha_5\log(\textit{Skill}_{99}) + \alpha_6\textit{Multi}_{99} + \tag{5}$$
$$\sum \gamma_j \textit{Industry}_{99,j} + \sum \lambda_s \textit{Size}_{99,s} + \varepsilon$$

To estimate this two-step procedure and obtain appropriate standard errors, we apply Stata's treatreg procedure with robust standard errors.[7]

3.2. Weights and size categories

We estimate both weighted and un-weighted versions of our Ordinary Least Squares and two-stage estimation strategies described above. Note that an "un-weighted" empirical specification implicitly assigns a uniform weight to all observations. Our weighted specifications use employer size to recover the aggregate relationship between computer networks and productivity. If the computer network differential is similar for both small and large plants, this weight will not alter the point estimate associated with the computer network variable. Because our empirical evidence suggests differences by plant size, any single summary statistic for the average association between computer networks and a given worker's labor productivity will need to weight by total employment. It proved necessary to use ASM sample weights to adjust the total employment weight because plants are sampled for the 1999 ASM with greater probability as they increase in size, with plants of greater than 250 employees being sampled with certainty. Weighting by total employment alone under-represents the employment of smaller plants, and this correction affects the weighted estimates that we report in this paper.

---

[7]At this point, we switched to Stata specifically to have the standard errors for two-step estimation procedure correctly specified.

Note that we estimated two additional sets of empirical specifications, which we also include in this paper. One specification weighted the Atrostic and Nguyen [3] regressions by ASM sample weights. This provides appropriate weights to smaller plants sampled with lower probability in the ASM. However, because the productivity differential is so similar among small plants, estimates that weight by the ASM sample weights are nearly identical to unweighted estimates. Furthermore, we estimated regression models in which plants are weighted by total employment, without an ASM correction. This specification produces empirical estimates of the association between computer networks and productivity that were also negative and sometimes more than twice the magnitude of employment-weighted estimates that employ the ASM correction. These larger negative estimates do not have a straightforward interpretation, but we include them for completeness.

To summarize, in the following section we report results for the two weighting methods that best describe the patterns in the sample under consideration: "unweighted" regressions refer to ASM-based regressions that weight all respondents equally, as in Atrostic and Nguyen [3], and "weighted" or "employment-weighted" results refer to those specification in which the weight is calculated as total employment multiplied by the ASM sample weight.

In order to provide an additional empirical reference point, we estimate regression specifications similar to those in Atrostic and Nguyen [3] separately for plants we categorize as *Small* (5–99 employees), *Medium* (100–249) and *Large* (250+), to obtain the relationship between computer networks and productivity for each size class. These regression specifications may prove informative to readers who, in light of later evidence that we present on the presence of different levels of adoption of this technology among establishments of different sizes, and the heterogeneous relationship between computer networks and productivity, do not find the estimation of an average association between computer networks and productivity meaningful.

4. Results

Table 1 shows average plant productivity and employment by whether or not the plant has a computer network, disaggregated by size and whether the firm that operates the plant is multi-unit or not. Employment characteristics are as follows. The average employment per plant is 237 employees. Among these plants, 45.2 percent are small plants (<100 employees), 29.2 percent are medium plants (100–249 employees), and 25.5 percent are large plants (250+ employees). The average employment counts for small, medium and large plants are 46.8, 161.6, and 661.4 workers, respectively. While small and medium plants account for 74.5 percent of the total number of plants in the sample, they employ only 29 percent of total employment. In contrast, large plants account for only 25 percent of total plants, but employ 71.0 percent of the sample's employment. Note that while our sample includes less than 10 percent of 350,000 establishments in the U.S. manufacturing sector, it accounts

Table 1
Plant size, productivity, and employment

	Productivity			Employment			Observations		
	With network	Without network	All	With network	Without network	All	With network	Without network	All
By Size									
Small	261.3	198.6	249.9	198.6	34.6	46.8	11,245	2,468	13,713
Size$_{99,1}$	266.6	200.7	251.1	28.2	22.8	27.0	5,972	1,834	7,806
Size$_{99,2}$	255.2	192.4	248.5	75.5	68.9	73.0	5,273	634	5,907
Medium	249.7	201.1	247.3	201.1	151.6	161.6	8,409	451	8,860
(*Size*$_{99,3}$)									
Large	255.6	270.5	256.0	270.5	990.7	661.4	7,508	232	7,740
Size$_{99,4}$	244.2	184.2	242.7	350.4	341.6	350.2	4,379	118	4,497
Size$_{99,5}$	251.3	290.3	252.5	677.8	658.9	677.2	2,167	69	2,236
Size$_{99,6}$	316.7	466.6	323.4	1960.9	3201.5	2016.4	962	45	1,007
By multi-unit status									
Single	157.6	148.7	155.9	44.2	44.2	87.8	8,102	1,867	9,969
Multi	298.0	285.0	297.0	234.5	234.5	310.2	19,060	1,284	20,344
All plants	256.1	204.2	250.7	250.4	250.4	237.0	27,162	3,151	30,313

Notes: Employment is the average number of employees employed by a manufacturing plant in a year. Productivity is the total amount of revenue at a plant, measures in thousands of dollars, divided by the plant's employment. See text for exact definitions of different size categories.

for more than 40 percent of all manufacturing employment in the U.S. because the ASM over-samples large plants.

These descriptive statistics indicate that computer networks are associated with higher productivity among smaller plants, and lower productivity among larger plants. Note that there are different degrees of computer network penetration by size. While 10.4 percent of all plants do not have a computer network, 18 percent of *Small* plants, 5 percent of *Medium* plants, and only 3 percent of *Large* plants do not have a computer network. Among plants with fewer than 250 employees, labor productivity is approximately 25% higher among plants with a computer network, indicating a strong, positive relationship between the presence of a computer network and productivity. However, the productivity differential is clearly negative among larger plants. Based on simple averages, the largest plants with a computer network are roughly 25% less productive than those that do not operate a computer network. Regression adjustment reduces the magnitude of the productivity differences, but not their direction.

Our regression results using all 29,261 observations (for OLS) are shown in Table 2. The un-weighted and ASM-weighted estimates, which over-represent small plants, indicate that computer networks are associated with 4.3%–4.5% higher labor productivity. The positive estimates obtained from un-weighted regressions confirm the Atrostic and Nguyen [3] finding that computer networks are associated with on the order of 5% higher labor productivity among plants, and who used an un-weighted empirical specification. Note that un-weighted regressions assign equal weight to every plant and small and medium plants dominate the sample, and weight-

Table 2
OLS regression results: All observations

	Un-weighted	ASM weighted	Employment weighted	ASM*Empl. weighted
Intercept	2.781** (170.33)	2.923** (208.66)	2.857** (133.72)	2.808** (169.11)
$Network_{99}$	0.045** (5.83)	0.043** (7.63)	−0.075** (7.99)	−0.033** (3.97)
$Log(Skill_{99})$	0.044** (13.15)	0.053** (14.71)	0.051** (15.98)	0.053** (16.24)
$Log(K_{97}/L_{97})$	0.089** (40.30)	0.072** (42.94)	0.098** (44.71)	0.093** (42.94)
$Log(M_{99}/L_{99})$	0.491** (197.67)	0.430** (178.64)	0.513** (201.77)	0.492** (195.78)
$Multi_{99}$	0.124** (22.02)	0.143** (24.91)	0.104** (13.94)	0.107** (16.50)
$Size_{99,2}$	−0.047** (7.00)	−0.017** (2.66)	−0.034* (2.14)	−0.016+ (1.72)
$Size_{99,3}$	−0.079** (12.23)	−0.025** (3.52)	−0.060** (4.21)	−0.028** (3.20)
$Size_{99,4}$	−0.091** (11.56)	−0.002 (0.16)	−0.066** (4.58)	−0.016+ (1.66)
$Size_{99,5}$	−0.082** (8.33)	0.031* (2.10)	−0.058** (4.00)	−0.005 (0.56)
$Size_{99,6}$	−0.016 (1.23)	0.117** (5.52)	−0.010 (0.73)	0.098** (9.82)
R^2	0.7375	0.7033	0.7244	0.7211
Number of plants	29,261	29,261	29,261	29,261

Notes: Table contains results of OLS regressions with plant-level log-labor productivity in 1999 as the dependent variable. See the discussion of equation (3) for additional details. t-statistics are in parentheses. **denotes "significant at the 1 percent level", *denotes "significant at the 5 percent level", and + denotes "significant at the 10 percent level". All empirical specifications include dummy variables for 3-digit NAICS industry. See text for exact variable definitions.

ing by the ASM provides even more weight to the smaller plants, which are sampled less frequently.

The simple and composite employment-weighted estimates, by contrast, indicate a negative relationship. However, even here, there are noticeable differences: the unadjusted employment weight indicates that the presence of a computer network is associated with a 7.5% lower labor productivity, while the employment weighted estimate that corrects for ASM sampling indicates that labor productivity 3.3% lower at plants with a computer network. When weighted by employment, the association is significantly negative, indicating that computer networks have a strong negative relationship with the typical worker exposed to them. Note that all estimates of the coefficient on the computer network variable are statistically significant at the 1 percent level, and so all of the specifications, whether weighted or unweighted, are robust to sampling error. The results that follow show that the basic distinction – that the unweighted and ASM weighted estimates show a positive relationship between computer networks and labor productivity, while the employment weighted estimates show a negative one – appears in all the subsequent results.

Our empirical specifications also employ a number of control variables. Those in Table 2 show expected relationships between observable characteristics and labor productivity, however, many of the characteristics are correlated with each other because larger plants often have higher material-to-labor ratios and capital-to-labor ratios, and have a higher share of non-production workers. The sign and magnitude of these other controls are similar across weighting methods. Higher capital-to-labor and material-to-labor ratios are also consistently associated with higher productivity.

Table 3
OLS regression results: All observations

	Un-weighted	ASM weighted	Employment weighted	ASM*Empl. weighted
Intercept	0.268** (20.97)	0.168** (15.40)	0.415** (22.40)	0.289** (21.14)
$Network_{99}$	0.025** (3.42)	0.043** (7.88)	−0.094** (10.24)	−0.039** (5.02)
$Log(Skill_{99})$	0.023** (7.06)	0.032** (9.07)	0.022** (7.13)	0.027** (8.76)
$Multi_{99}$	0.058** (11.11)	0.057** (10.85)	0.051** (7.18)	0.054** (8.94)
$Size_{99,2}$	0.004 (0.64)	0.024** (3.91)	0.008 (0.54)	0.026** (2.98)
$Size_{99,3}$	−0.007 (1.18)	0.022** (3.18)	0.006 (0.48)	0.035** (4.26)
$Size_{99,4}$	−0.024** (3.19)	0.029** (2.74)	−0.012 (0.84)	0.038** (4.27)
$Size_{99,5}$	−0.036** (3.94)	0.025+ (1.74)	−0.025+ (1.81)	0.031** (3.31)
$Size_{99,6}$	−0.026* (2.00)	0.035+ (1.71)	−0.008 (0.59)	0.050** (5.39)
R^2	0.0439	0.0408	0.0829	0.0715
Number of plants	29,254	29,254	29,254	29,254

Notes: Table contains results of OLS regressions with plant-level total factor productivity in 1997 as the dependent variable. See the discussion of equations (2) and (6) for additional details. t-statistics are in parentheses. **denotes "significant at the 1 percent level", *denotes "significant at the 5 percent level", and + denotes "significant at the 10 percent level". All empirical specifications include dummy variables for 3-digit NAICS industry. See text for exact variable definitions.

Those plants that have more non-production workers also exhibit higher productivity. The size category controls are somewhat less consistent: the larger size category dummy variables indicate that a higher size is associated with lower productivity in the unweighted and ASM weighted specifications, but when larger plants are given a weight according to their size, the size categories indicate that larger plants exhibit higher productivity.

The basic result also appears when we employ a multi-factor productivity measure from the 1997 CM, which precedes the 1999 ASM by two years. This follows the calculation method described in Eq. (2) above to estimate total factor productivity. Although we have chosen to focus on labor productivity in this paper, note that the main results of our paper are qualitatively similar if we use an alternative specification that uses a multi-factor measure of productivity,

$$
\begin{aligned}
Log(TFP_{97}) = {} & \beta_0 + \beta_1 Network_{99} + \alpha_5 \log(Skill_{99}) + \alpha_6 Multi_{99} \\
& + \sum \gamma_j Industry_{99,j} + \sum \lambda_s Size_{99,s} + \varepsilon,
\end{aligned}
\tag{6}
$$

and results are shown in Table 3. Readers should exercise caution, because the timing of the dependent variable, total factor productivity, is 1997, while the computer network status refers to a plant's status in 1999. Nevertheless, the results indicate that the overall findings in Table 2 are robust to the particular productivity measure employed, as the point estimates show similar effects for lagged total factor productivity and contemporaneous labor productivity. The estimates that weight large plants with exactly (unweighted) or approximately (ASM weighted) the same weight as small plants show an indicates that the presence of a computer network is associated with 2.5%–4.3% higher total factor productivity, whereas the labor productivity specification indicated 4.3%–4.5%. The employment weighted estimates also agree in sign

and approximately in magnitude: the unadjusted employment weight indicates that the presence of a computer network is associated with 9.4% lower total factor productivity, whereas the labor productivity estimate only suggested a 7.5% penalty, and the final correction of both productivity measure indicates a smaller penalty, of 3.9% for total factor productivity, but only 3.3% for the labor productivity specifications. These results are quite similar despite the lag because productivity is relatively persistent within U.S. manufacturing plants, see Foster, Haltiwanger, and Syverson [9]. Readers should also note that the assumed production function underlying the functional forms in both sets of specifications is a Cobb-Douglas function of capital, labor, and materials.

Results for the two-stage estimation technique include 27,007 observations, and are shown in Table 4.[8] The two-stage regression results in Table 4 are similar in sign to those obtained from OLS regression in Table 2: the unweighted and ASM-weighted estimates both indicate a positive relationship between the presence of a computer network and labor productivity, while the employment-weighted and composite weights indicate that computer networks are associated with lower productivity. This indicates that the direction of the OLS results (see our discussion of Table 6 below for the magnitude) are robust to instrumenting: they are apparent even when we restrict identification to differences in the presence of a computer network could be predicted based on computer expenditures in 1992.

Table 5 shows the regression results by our three broad size categories. It is clear that the effect of computer networks on productivity is positive and significant for plants in our *Small* and *Medium* categories. However, for larger plants in our *Large* category, which have 250 or more employees, computer networks are clearly associated with lower productivity, and this relationship is robust to both control for observable characteristics and our two-stage approach. Specifically, OLS results show that computer networks are associated with a 4.6 percent increase in labor productivity for *Small* plants and 7.6 percent increase for *Medium* plants. However, computer networks are associated with a 6.4 percent decline in labor productivity among the set of largest plants.

The interpretation of the magnitudes of the two-step estimates in Tables 4 and 5 is less straightforward than those of the OLS regressions. This is because the variable $Network_{99}$ in the OLS regression is a dummy variable whose value is either 0 or 1, whereas the $Pr(Network_{99})$ variable in the two-stage regressions is continuous and has value between 0 and 1. Therefore, following Atrostic and Nguyen [3], we compare the effects of computer network on plant productivity at two points in the predicted productivity having a computer network. These comparisons are reported in Table 6. The percentage productivity increase (or decrease) is calculated by multiplying the difference in the predicted probability of having a computer network by

---

[8]The sample reduced to 27,007 observations due to missing values in 1992 for certain variables such as $COMP_{92}$ (computer investment in 1992) and $SKILL_{92}$ (non-production workers/total employment ratio).

Table 4
Two-stage regression results: All observations

	Un-weighted	ASM weight	Employment weight	ASM*Empl. weight
Intercept	2.688** (135.41)	2.890** (48.39)	3.316** (61.31)	3.314** (72.69)
Pr($Network_{99}$)	0.310** (17.86)	0.504** (18.32)	−0.628** (17.96)	−0.616** (25.35)
Log($Skill_{99}$)	0.044** (12.50)	0.051** (6.28)	−0.035** (4.10)	0.036** (5.09)
Log(K_{97}/L_{97})	0.091** (39.95)	0.072** (12.22)	0.098** (15.54)	0.093** (15.73)
Log(M_{99}/L_{99})	0.488** (184.37)	0.406** (38.08)	0.510** (46.54)	0.448** (55.18)
$Multi_{99}$	0.091** (15.04)	0.072** (5.00)	0.118** (11.43)	0.151** (15.17)
$Size_{99,2}$	−0.048** (6.87)	−0.018+ (1.71)	−0.027** (3.27)	−0.009 (0.82)
$Size_{99,3}$	−0.078** (11.51)	−0.022* (2.05)	−0.047** (5.29)	−0.017+ (1.61)
$Size_{99,4}$	−0.087** (10.72)	0.007 (0.57)	−0.050** (4.62)	−0.005 (0.38)
$Size_{99,5}$	−0.082** (8.19)	0.033* (2.20)	−0.047** (3.78)	0.020 (1.48)
$Size_{99,6}$	−0.014 (1.06)	0.123** (6.68)	0.024 (1.37)	0.105** (5.26)
Number of plants	27,007	27,007	27,007	27,007

Notes: Table contains results of OLS regressions with plant-level log-labor productivity in 1999 as the dependent variable, see the discussion of equation (3) for additional details; also the results of two stage regressions with plant-level log-labor productivity in 1999 as the dependent variable, see the discussion of equations (4) and (5) for additional details. t-statistics are in parentheses. **denotes "significant at the 1 percent level", *denotes "significant at the 5 percent level", and + denotes "significant at the 10 percent level". All empirical specifications include dummy variables for 3-digit NAICS industry. See text for exact variable definitions.

Table 5
Regression results by plant size

	Small plants $5 \leqslant L_{99} < 100$		Medium plants $100 \leqslant L_{99} < 250$		Large plants $250 \leqslant L_{99}$	
	OLS	Two-stage	OLS	Two-stage	OLS	Two-stage
Intercept	2.821**	2.788**	2.561**	2.448**	2.853**	3.406**
	(117.46)	(95.35)	(82.21)	(64.87)	(70.20)	(83.31)
$Network_{99}$	0.045**	(–)	0.076**	(–)	−0.063**	(–)
	(4.86)		(4.48)		(2.60)	
Pr($Network_{99}$)	(–)	0.443**	(–)	0.356**	(–)	−0.667**
		(20.18)		(11.37)		(25.43)
Log($Skill_{99}$)	0.053**	0.051**	0.025**	0.023**	0.051**	0.044**
	(9.52)	(8.60)	(4.29)	(3.83)	(8.17)	(7.08)
Log(K_{97}/L_{97})	0.073**	0.075**	0.096**	0.095**	0.099**	0.099**
	(21.52)	(20.43)	(24.21)	(23.84)	(23.03)	(23.053)
Log(M_{99}/L_{99})	0.475**	0.461**	0.512**	0.506**	0.498**	0.500**
	(131.25)	(112.63)	(110.61)	(105.98)	(93.67)	(93.93)
$Multi_{99}$	0.148**	0.108**	0.088**	0.083**	0.111**	0.116**
	(18.52)	(12.08)	(9.51)	(8.66)	(2.60)	(6.94)
R^2	0.7547	–	0.7442	–	0.6924	–
Numbers of plants	12,887	11,117	8,726	8,400	7,648	7,490

Notes: Table contains results of two stage regressions with plant-level log-labor productivity in 1999 as the dependent variable. See the discussion of equations (4) and (5) for additional details. t-statistics are in parentheses. **denotes "significant at the 1 percent level", *denotes "significant at the 5 percent level", and + denotes "significant at the 10 percent level". All empirical specifications include dummy variables for 3-digit NAICS industry. See text for exact variable definitions.

Table 6
Two-stage estimates by percentile

		Percent (%) increase in labor productivity	
	Percentiles (%) of Pr(*Network*$_{99}$)	Unweighted	*ASM*Empl.* Weighted
Small	1% (0.5049) versus 99% (0.9805)	21.06%	
plants	5% (0.6219) versus 99% (0.9805)	15.87%	
	18% (0.7535) versus 99% (0.9805)	**10.06%**	
	25% (0.7885) versus 99% (0.9805)	8.51%	
Medium	1% (0.8386) versus 99% (0.9886)	5.34%	
plants	**5% (0.8861) versus 99% (0.9886)**	**3.65%**	
	10% (0.9042) versus 99% (0.9886)	3.00%	
	25% (0.9365) versus 99% (0.9886)	1.85%	
Large	1% (0.9228) versus 99% (0.9929)	−4.67%	
plants	**3% (0.9379) versus 99% (0.9929)**	**−3.67%**	
	10% (0.9464) versus 99% (0.9929)	−3.10%	
	25% (0.9577) versus 99% (0.9929)	−2.35%	
All	1% (0.6085) versus 99% (0.9941)	11.95%	−23.75%
plants	5% (0.7320) versus 99% (0.9941)	8.12%	−16.14%
	10% (0.7941) versus 99% (0.9941)	**6.20%**	**−12.32%**
	25% (0.8725) versus 99% (0.9941)	3.77%	−7.49%

Notes: The estimated increases in labor productivity are calculated by comparing different points in the distribution of the predicted probabilities of having a computer network (i.e., Pr(*Network*$_{99}$)). For example, the first row compares plants in the 1st and 99th percentiles of the predicted probability of having a computer network. Bolded lines indicate the percentile closest to the sample average for ownership of a computer network. See text for exact variable definitions.

the estimated coefficient of the Pr(*Network)* variable. For example, a *Small* plant at the 99th percentile of the probability distribution (0.9805) enjoys a labor productivity that is higher than that in the plant at the 1st percentile (0.5049) in the predicted probability distribution by 21.06 percent (= (0.9805–0.5049) × 0.443). Since 18 percent of *Small* plants in our working sample do not have a computer network, we compare a plant at the 18th percentile (likely not having a computer network) to a plant at the 99th percentile of the probability distribution (most likely to have a computer network). We find that computer networks are associated with 10.06 percent higher labor productivity in *Small* plants. For *Medium* plants, only 5 percent do not have a computer network. Comparing a *Medium* plant at the 5th percentile to its counterpart at the 99th percentile in the probability distribution, we find that the association between computer networks and labor productivity among medium plants is 3.65 percent. Finally, comparing a *Large* plant at the 3rd percentile and its counterpart at the 99th percentile in the productivity distribution, we find that the productivity difference is −3.67 percent.

5. Discussion and concluding remarks

The empirical phenomenon that establishments with computer networks exhibit higher labor productivity, presented in Atrostic and Nguyen [3], is entirely due to

small- and medium-sized establishments. Larger plants that employ 250 or more employees but lack a computer network exhibit higher productivity than similar large plants with a computer network. Regression specifications that weight all plants equally indicate a positive relationship between the presence of a computer network and productivity, while employment-weighted regressions indicate a negative relationship. This indicates that although the *typical plant* exhibits an increase in labor productivity in the presence of a computer network, the *typical worker* exhibits lower productivity in the presence of a computer network.

This apparent negative employment-weighted effect of computer network suggests that empirical evidence on the effect of computer networks on small- to medium-sized manufacturing establishments has limited ability to explain aggregate trends in economic growth. For example, Stiroh [19] notes that IT-intensive industries have grown at a 1% faster rate than industries that use IT less intensively. Jorgenson [12] furthermore argues that information technology during the 1990s accounts for roughly 0.5% percent of GDP growth and around 1% of GDP growth in later years, for a total of on the order of 10% or so to U.S. GDP. The unweighted point estimate in this study and in Atrostic and Nguyen [3] indicates that the adoption computer networks adds around 5% to manufacturing plants' productivity. This implies that the near-universal (90%) adoption of computer networks by the manufacturing industry in 1999, which at the time had about 15% of U.S. employment would contribute roughly 5%*15% = 0.75% to U.S. GDP, and so could plausibly explain some of these overall gains. However, the micro-level gains from computer networks seem only to represent the effects of computer networks at smaller plants, and when considering aggregate output, these gains are outweighed by a negative relationship among the relatively small number of large plants.

Our results demonstrate the importance of considering sample weights when conducting establishment-level efficiency analyses. Un-weighted regressions give establishments that employ few people as much influence on the parameter estimates as larger plants. While this makes sense when treatment effects do not vary by plant size, we have showed that this is clearly not the case in our dataset. Regression specifications that measure the relationship between computer networks and labor productivity, even those that include control variables for establishment size, show a positive relationship. Assigning establishments weights commensurate with establishment size highlights the negative relationship between computer networks and labor productivity, and, therefore, represents aggregate relationships more clearly. While we do not want to discourage un-weighted regression specifications, we hope that this study highlights the danger of using un-weighted establishment-level estimates to explain economy-wide trends. We furthermore recommend that researchers who have an establishment size variable either run results separately by plant size to test for heterogeneous treatment effects, or compare unweighted estimates with employment-weighted ones as a more general test for model misspecification, see Solon et al. [18].

We would like to point out the greatest limitation of our results is the fact that the negative effect of computer networks on productivity is driven by a relatively small number (3%) of large employers. This limitation is inherent to the measure of IT intensity employed, and may be able to be overcome using different data that permits IT investments to be captured with a continuous rather than a binary variable. Another potential concern is whether large, productive plants may have become smaller due to the presence of a computer network. There is some evidence that manufacturing plants employ fewer people when they utilize IT more intensively. For example, Brynjlofsson et al. [7] show that IT investment leads to a shift in the firm size distribution in favor of smaller firms. Similarly, Bardhan, Whitaker and Mithas [5], using CM data and a supplemental survey, find that information technology investments are positively associated with outsourcing decisions, which naturally may allow establishment sizes to decrease.

We would like this paper to serve as a reference point in how to interpret the computer network productivity differential. An increasing number of studies note positive effects of computer networks on business outcomes, and discussion of differential effects generally considers which point estimate is the "most positive." We have found using data that produces the "standard" result of IT investments on productivity, that weighting by employment actually produces a negative effect. This may be because small establishments outsource their computer network rather than having an in-house IT department, or there could be some ability for small establishments to increase productivity based on the fact that the information managed by the computer network is shared among fewer individuals. We hope that future studies can re-assess this productivity differential to see whether it holds more generally, and, if it does, attempt to provide a mechanism by which it occurs.

6. Disclaimer and acknowledgements

References

[1]   B.K. Atrostic, K. Motohashi and S.V. Nguyen, *Computer Network Use and Firms' Productivity Performance: The United States vs. Japan*, Center for Economic Studies Discussion Paper CES 08-30, 2008.

[2]   B.K. Atrostic and S.V. Nguyen, Computer Input, Computer Networks, and Productivity, in: *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*, E. Berndt and C. Hulten, eds, Conference on Research in Income and Wealth, Studies in Income and Wealth 67. University of Chicago Press for the National Bureau of Economic Research, 2007, pp. 383–410.

[3]   B.K. Atrostic and S.V. Nguyen IT and Productivity in U.S. Manufacturing: Do Computer Networks Matter? *Economic Inquiry* **43** (July 2005), 493–506.

[4]   M.N. Baily, C. Hulten and D. Campbell. Productivity Dynamics in Manufacturing Plants, *Brookings Papers on Economic Activity: Microeconomics* **1992** (1992), 187–267.

[5]   I. Bardhan, J. Whitaker and S. Mithas, Information Technology, Production, Process Outsourcing, and Manufacturing Plant Performance, *Journal of Management Information Systems* **23** (Fall 2006), 13–40.

[6]   E. Brynjolfsson and L.M. Hitt Computing Productivity: Firm-Level Evidence, *The Review of Economics and Statistics* **85** (November 2003), 793–808.

[7]   E. Brynjolfsson, T.W. Malone, V. Gurbaxani and A. Kambil, Does Information Technology Lead to Smaller Firms? *Management Science* **40** (December 1994), 1628–1644.

[8]   T. Dunne, L.S. Foster, J. Haltiwanger and K. Troske, Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment, *Journal of Labor Economics* **22** (April 2004), 397–430.

[9]   L. Foster, J. Haltiwagner and C. Syverson, Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review* **98** (March 2008), 394–425.

[10]   N. Greenan and J. Mairesse, *Computers and Productivity in France: Some Evidence*, Working Paper No. 5836, National Bureau of Economic Research, 1996.

[11]   M. Iansiti, G. Favaloro, J. Utzschneider and G. Richards, *Why IT Matters in Midsized Establishments*, Harvard Business School Working Paper Series 06-013, 2005.

[12]   D.W. Jorgenson, Information Technology and the U.S. Economy, *The American Economic Review* **91** (March 2001), 1–32.

[13]   D.W. Jorgenson and K.J. Stiroh, *Raising the Speed Limit: U.S. Economic Growth in the Information Age*, Brookings Papers on Economic Activity **2000** (2000), 125–211.

[14]   R.H. McGuckin, S.V. Nguyen and A.P. Reznek, On Measuring the Impact of Ownership Change on Labor: Evidence from U.S. Food-Manufacturing Plant-Level Data, in: *Labor Statistics Measurement Issues*, J. Haltiwanger, M.E. Manser and R. Topel, eds, Conference on Research in Income and Wealth, Studies in Income and Wealth 60. University of Chicago Press for the National Bureau of Economic Research, 1998, pp. 207–248.

[15]   T. Mesenbourg, *Measuring Electronic Business*, http://www.census.gov/econ/estats/papers/ebusasa.pdf, 2001 (accessed: February 28, 2012).

[16]   G.S. Olley and A Pakes, The Dynamics of Productivity in the Telecommunications Equipment Industry, *Econometrica* **64** (November 1996), 1263–1297.

[17]   I. Shin, Adoption of Enterprise Application Software and Firm Performance, *Small Business Economics* **26** (April 2006), 241–256.

[18]   G. Solon, S. Haider and J. Wooldridge, *What are We Weighting For?* Working Paper #18859, National Bureau of Economic Research, 2013.

[19]   K.J. Stiroh, Information Technology and the U.S. Productivity Revival: What do the Industry Data Say? *The American Economic Review* **92** (December 2002), 1559–1576.

[20]   P. Tambe and L. Hitt, The Productivity of Information Technology Investments: New Evidence from IT Labor Data, *Information Systems Research* **23** (September 2012), 599–617.

[21]   J. Triplett and B. Bosworth, "Baumol's Disease" has been Cured: IT and Multifactor Productivity in US Service Industries, in: *The New Economy and Beyond: Past, Present and Future*, D. Jansen, ed., Northampton, MA: Edward Elgar Publishing, Inc., 2006, pp. 34–71.