Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Green Cloud? An Empirical Analysis of Cloud Computing and Energy Efficiency

Jiyong Park, Kunsoo Han, Byungtae Lee

To cite this article:
Jiyong Park, Kunsoo Han, Byungtae Lee (2023) Green Cloud? An Empirical Analysis of Cloud Computing and Energy Efficiency. Management Science 69(3):1639-1664. https://doi.org/10.1287/mnsc.2022.4442

informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 69, No. 3, March 2023, pp. 1639–1664
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Green Cloud? An Empirical Analysis of Cloud Computing and Energy Efficiency

Jiyong Park,[a] Kunsoo Han,[b,*] Byungtae Lee[c]

[a] Bryan School of Business and Economics, University of North Carolina at Greensboro, Greensboro, North Carolina 27412; [b] Desautels Faculty of Management, McGill University, Montreal, Quebec H3A 1G5, Canada; [c] College of Business, Korea Advanced Institute of Science and Technology, Seoul 02455, Korea
*Corresponding author

Contact: jiyong.park@uncg.edu, https://orcid.org/0000-0003-4444-7261 (JP); kunsoo.han@mcgill.ca, https://orcid.org/0000-0002-5203-4654 (KH); btlee@kaist.ac.kr (BL)

Received: May 14, 2018
Revised: November 29, 2019; April 17, 2021; February 19, 2022
Accepted: March 19, 2022
Published Online in Articles in Advance: June 7, 2022

https://doi.org/10.1287/mnsc.2022.4442

Abstract. The rapid, widespread adoption of cloud computing over the last decade has sparked debates on its environmental impacts. Given that cloud computing alters the dynamics of energy consumption between service providers and users, a complete understanding of the environmental impacts of cloud computing requires an investigation of its impact on the user side, which can be weighed against its impact on the vendor side. Drawing on production theory and using a stochastic frontier analysis, this study examines the impact of cloud computing on users' energy efficiency. To this end, we develop a novel industry-level measure of cloud computing based on cloud-based information technology (IT) services. Using U.S. economy-wide data from 57 industries during 1997–2017, our findings suggest that cloud-based IT services improve users' energy efficiency. This effect is found to be significant only after 2006, when cloud computing started to be commercialized, and becomes even stronger after 2010. Moreover, we find heterogeneous impacts of cloud computing, depending on the cloud service models, energy types, and internal IT hardware intensity, which jointly assist in teasing out the underlying mechanisms. Although software-as-a-service (SaaS) is significantly associated with both electric and nonelectric energy efficiency improvement across all industries, infrastructure-as-a-service (IaaS) is positively associated only with electric energy efficiency for industries with high IT hardware intensity. To illuminate the mechanisms more clearly, we conduct a firm-level survey analysis, which demonstrates that SaaS confers operational benefits by facilitating energy-efficient production, whereas the primary role of IaaS is to mitigate the energy consumption of internal IT equipment and infrastructure. According to our industry-level analysis, the total user-side energy cost savings from cloud computing in the overall U.S. economy are estimated to be USD 2.8–12.6 billion in 2017 alone, equivalent to a reduction in electricity use by 31.8–143.8 billion kilowatt-hours. This estimate exceeds the total energy expenditure in the cloud service vendor industries and is comparable to the total electricity consumption in U.S. data centers.

History: Accepted by Chris Forman, information systems.
Supplemental Material: Data files and the online appendices are available at https://doi.org/10.1287/mnsc.2022.4442.

Keywords: cloud computing • software-as-a-service • infrastructure-as-a-service • IT outsourcing • energy efficiency • green IT • green IS • sustainability • stochastic frontier analysis

Introduction

Cloud computing—providing information technology (IT) services over a network—has been widely adopted by businesses across industries because of such benefits as flexibility, scalability, always-on availability, and pay-as-you-go pricing (Marston et al. 2011). A cloud service vendor provides users with on-demand access not only to various computing resources in the cloud (e.g., storage, networking, servers) but also to software and applications through the Internet. The former service model is commonly known as infrastructure-as-a-service (IaaS; e.g., Amazon Web Services EC2, Google Compute Engine), whereas the latter is typically called software-as-a-service (SaaS; e.g., Salesforce, Google Workspace, Microsoft Office 365). Gartner (2021) predicts that public cloud spending will exceed 45% of all enterprise IT spending by 2026, up from less than 17% in 2021. According to an IT trend report by the Society for Information Management (2020), cloud computing was the top organizational investment in 2020 for the first time since 2010. The rapid widespread adoption of cloud computing over the last decade has sparked debates on its environmental impacts among practitioners and academics alike. In particular, sustainability has been identified as one of major driving forces that may reshape the relationship between cloud service providers and users; as Gartner (2021, p. 9) puts it, "New sustainability requirements will be mandated over the next few years and the choice of cloud services providers may hinge on the provider's 'green' initiatives."

On one hand, there has been a pervasive view that cloud computing is one of the main culprits of global energy consumption and environmental degradation. For example, the large volume of electricity consumed to support cloud services has often been described as a "dirty secret" of cloud computing (Fortune 2019). According to a study conducted by the Lawrence Berkeley National Laboratory (Shehabi et al. 2016), U.S. data centers consumed approximately 70 billion kilowatt-hours of electricity in 2014, representing 2% of the country's total electricity consumption. These numbers are expected to grow rapidly as the demand for cloud-based IT and data services increases (Mastelic et al. 2014, Jones 2018). In addition to data center operations, data transfer from data centers to users may also consume a significant amount of energy (Baliga et al. 2011).

On the other hand, some practitioners and scholars suggest that cloud computing can lead to net energy savings for the overall economy if we factor in the user-side benefits. For example, organizations can replace an energy-inefficient internal IT infrastructure with cloud-based IT services accessible on demand over a network, which allows them to optimize IT resource utilization and enhance energy efficiency. Capitalizing on economies of scale, cloud computing can save a substantial amount of energy through virtualization, which allows a number of different applications/instances that typically have a large amount of unused capacity in the absence of cloud computing to be consolidated onto a single server (Bose and Luo 2011). A simulation study argues that if all U.S. businesses moved their applications to the cloud (i.e., cloud service providers' servers), they could reduce their computing energy footprint by 87% (Masanet et al. 2013). Moreover, on-demand cloud services provide low-cost access to scalable, high-powered software running on the cloud that can confer greater operational and sustainability benefits in ways to reduce wasted resources, including energy (Battleson et al. 2016), which could not be achieved through in-house development or traditional software licensing.

Despite ongoing debates over the environmental impacts of cloud computing, a comprehensive assessment of the impacts from the perspectives of both service providers (vendors) and users (clients) has been elusive, mainly because of a lack of the necessary data. Mytton (2020) highlights the challenge in evaluating the environmental impacts of cloud computing, as it may alter the energy consumption dynamics between users and service providers. In this regard, a complete understanding of the environmental implications of cloud computing requires an investigation of its impact on *the user side*, which can be weighed against its *vendor-side* effect. Although prior literature has examined how the energy consumption on the vendor side adversely affects the environment and how the energy efficiency of cloud computing facilities can be improved (e.g., hyperscale data centers) (Mastelic et al. 2014, Jones 2018), little attention has been devoted to the broader impacts of cloud computing on the user side (Marston et al. 2011). Drawing on production theory, this study aims to fill the current void by empirically examining the *user-side* effect of cloud computing on energy efficiency improvement, which allows us to assess the economy-wide impact of cloud computing on energy use. Moreover, cloud computing has transformed the way various IT resources (i.e., hardware/infrastructure and software) are procured. Given that hardware and software may have different implications for energy consumption and environmental sustainability because of first- and second-order effects (Horner et al. 2016), we also scrutinize the differential roles of cloud service models provisioning hardware/infrastructure (i.e., IaaS) and software (i.e., SaaS) in shaping energy efficiency, which helps us tease out the underlying mechanisms and thus offer novel theoretical and managerial implications.

For the empirical analysis, we develop a novel industry-level measure of cloud computing that captures purchased cloud-based IT services by combining industry-level product sales data from the U.S. Economic Census and interindustry purchase flows from input-output use tables. This approach allows us to distinguish different types of cloud services (SaaS and IaaS), based on the types of IT services purchased by an industry. In addition, a stochastic frontier analysis is employed to estimate an energy frontier—the minimum level of energy consumption needed to produce a given output with extant production inputs—which is used to measure an industry's energy efficiency as a deviation from the energy frontier. Then, we examine whether and how cloud-based IT services influence users' energy efficiency. We further delve into the differential effects involving two types of cloud computing services.

Using data from 57 U.S. private industries over the period 1997–2017, we find that cloud-based IT services improve client industries' energy efficiency. In contrast, no significant effect has been found for non–cloud-based IT services such as computer systems design, which typically relies on clients' own IT infrastructure. Although cloud-based IT services have no significant impact on energy efficiency prior to 2006, its impact becomes statistically significant after 2006, when the first commercial cloud services (i.e., Amazon Web Services (AWS)) were launched; its impact becomes even larger after 2010, when the prices of cloud services substantially declined due to increased competition.

Moreover, we find heterogeneous impacts of cloud computing on energy efficiency, depending on the cloud service models, energy types, and internal IT hardware intensity, which jointly assist in teasing out the underlying mechanisms. Although SaaS is significantly associated with energy efficiency improvement across all industries, the impact of IaaS varies, depending on the internal IT intensity and energy type. Specifically, the positive contribution of IaaS to energy efficiency is stronger in industries that make more intensive use of IT hardware. In addition, while the main effects of SaaS on energy efficiency are significant for both electric and nonelectric energy efficiency, IaaS only improves electric energy efficiency for industries with high IT hardware intensity.

To illuminate the underlying mechanisms more clearly, we conduct a firm-level analysis using survey data from 187 firms. The results corroborate our industry-level findings and validate the underlying mechanisms at a more granular level. Specifically, although both SaaS and IaaS are positively associated with energy efficiency, the effects of SaaS and IaaS are mainly mediated by operational benefits and a reduction in the energy consumption of internal IT, respectively. Taken together, the findings from industry- and firm-level analyses highlight the intricate way that cloud computing leads to energy efficiency improvement: although SaaS facilitates energy-efficient production by mainly helping to optimize operations and redesign production processes (which consume both electric and nonelectric energy), the primary role of IaaS is to mitigate the adverse impacts of internal IT equipment and infrastructure (which rely on electric energy) on energy efficiency.

According to the back-of-the-envelope calculation based on our industry-level estimates, the total energy cost savings from cloud computing on the user side in the U.S. economy are estimated to be USD 2.8–12.6 billion in 2017 (based on the 95% confidence interval). This estimate is equivalent to a reduction in electricity use by 31.8–143.8 billion kilowatt-hours, which represents approximately 0.9%–3.9% of total electricity use in the United States. The estimated energy savings exceed the total energy consumption by cloud service vendor industries and is comparable to the total electricity consumption by U.S. data centers estimated by the Lawrence Berkeley National Laboratory (70 billion kilowatt-hours in 2014; Shehabi et al. 2016). Our estimates highlight the notion that the cloud computing's energy efficiency benefits on the user side far outweigh its vendor-side adverse effects in the overall economy.

From the business perspective, this research presents the first empirical evidence of the user-side effect of cloud computing on energy efficiency, which is imperative for a comprehensive assessment of the economy-wide impact of cloud computing on energy use. Going beyond the business value of IT, this study advances the body of knowledge in the information systems (IS) literature by shedding new light on the "green" value of cloud computing and uncovering the way cloud computing confers sustainability benefits to business users and the economy as a whole. This study also provides meaningful policy and managerial implications, informing ongoing dialogues on cloud computing and environmental sustainability.

Related Literature and Theoretical Framework

Cloud Computing

Cloud computing is generally defined as an "information technology service model where computing services (both hardware and software; HW and SW hereafter) are delivered on-demand to customers over a network in a self-service fashion, independent of device and location" (Marston et al. 2011, p. 177). The National Institute of Standards and Technology suggests five essential characteristics of cloud computing—on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service—and three service models—*SaaS*, *platform-as-a-service (PaaS)*, and *IaaS* (Mell and Grance 2011). SaaS allows users to access SW and applications run on the cloud, eliminating the need to install and run applications on their own machines. Examples range from enterprise systems such as Salesforce, Workday, and NetSuite to personal applications such as Gmail and Microsoft Office 365. PaaS provides a platform on which applications can be developed without considering the underlying technical layers such as the server uptime, virtualization parameters, and network security (e.g., Microsoft Azure, Salesforce Platform, Google Cloud Platform). IaaS provides users with on-demand access to storage, networking, servers, and other computing resources over a network (e.g., Amazon EC2, Google Compute Engine, and Alibaba Cloud).

On-demand IT services with scalability and the pay-as-you-go pricing model make it attractive for firms to source more IT services from the cloud and use computing resources as a utility (Brynjolfsson et al. 2010). Although a few studies have examined the economic impacts of cloud computing (Wauters et al. 2016, Jin and McElheran 2019) and its operational benefits (Loukis et al. 2019, Rodrigues et al. 2021), its environmental impacts on users have only been discussed conceptually (Iyer and Henderson 2010) with little empirical evidence. By reviewing 236 scholarly journal articles on cloud computing published between 2009 and 2014, Bayramusta and Nasir (2016) highlight that the energy consumption dimension of cloud computing is the least attention-grabbing theme in the literature, accounting for only 4% of publications; even those studies examine the energy use of cloud systems on the vendor side only. Using survey data with a focus on supply chain management, Schniederjans and Hales (2016) present evidence of a positive relationship between the use of cloud computing and perceived environmental performance such as reducing solid waste, resource consumption, and hazardous material consumption; however, the authors neither quantify the impact of cloud computing on actual energy consumption/efficiency in the overall economy nor disentangle the roles of different types of cloud services (e.g., SaaS versus IaaS).

Given that cloud computing alters the energy consumption dynamics between users (clients) and service providers (vendors) (Mytton 2020), there is a critical need for a comprehensive empirical assessment of the environmental impacts of cloud computing, considering both clients and vendors. Hence, the current study aims to fill this important gap in the literature and advance our understanding of whether and how cloud computing confers environmental benefits to its users, based on a novel measure of cloud computing and industry-level data that cover the entire US economy. Doing so will allow us to assess the economy-wide net effect of cloud computing on energy use.

Cloud Computing and Energy Efficiency from the User Perspective

The environmental effects of IT can be classified into first-order effects—direct effects from the physical existence of IT that increase energy consumption—and second-order effects—indirect, energy-saving effects from the ongoing use and application of IT (Hilty et al. 2006, Horner et al. 2016).[1] In the IS literature, mitigating the first-order effects is often termed as "green IT," and facilitating the second-order effects is referred to as "green IS" (Malhotra et al. 2013). Against this backdrop, we posit that cloud computing services can improve users' energy efficiency through two mechanisms: (i) by *mitigating the energy consumption of IT infrastructure and equipment* and (ii) by *facilitating energy-efficient production and operations*.

Mitigating the Energy Consumption of IT Infrastructure and Equipment. Given that the United States is home to three million data centers (Fortune 2019), IT equipment and infrastructure have been blamed for their intensive energy consumption (Murugesan 2008). An annual report by Greenpeace (2017) highlights the importance of the IT sector in global energy demand, stating that "the IT sector was estimated to already consume over 7% of global electricity demand in 2012 … and [will] continue to grow at least 7% annually through 2030, double the average rate of electricity growth globally" (p. 15). The Lawrence Berkeley National Laboratory also estimates the electricity consumption of U.S. data centers as being approximately 2% of the country's total electricity use (Shehabi et al. 2016).

In this regard, IaaS allows its users to optimize IT resource utilization and minimize the cost of their own energy-intensive IT infrastructure by providing on-demand access to computing resources (e.g., processors, networks, servers, storage) over a network (Armbrust et al. 2010, Bose and Luo 2011). A report by Deloitte and the European Commission suggests that the largest economic benefit of cloud computing is direct cost savings such as those from reduced IT infrastructure investment and maintenance, which account for 20%–50% of total IT costs (Wauters et al. 2016). Given that energy-related costs account for 42% of server operations costs—direct power consumption (19%) and cooling infrastructure (23%) (Jing et al. 2013), cloud service users can reduce their energy costs by migrating their own servers and data centers to the cloud without needing to power and cool them. In addition, SaaS can reduce the energy consumption required to run and support SW applications on the user side. For instance, the power consumption needed to run Excel 365 (a cloud-based version of Microsoft Excel) on user devices is estimated as being 15.7% (tablet) to 27.3% (laptop) lower than that needed to run the on-premises Excel 2010 (Williams and Tang 2013).

Moreover, cloud computing services allow users to respond to uncertain demand and meet excess requirements without incurring the costs of purchasing and running internal IT infrastructures, which are highly underused during low-demand periods (Marston et al. 2011). This cloud-driven flexibility and agility can reduce wasted resources and increase the utilization rate of IT resources, leading to energy savings. For example, after making its service available via Facebook in 2008, a U.S.-based online video service, Animoto, experienced a demand surge of 750,000 new users in three days, a phenomenon known as the "Slashdot effect" (Klems et al. 2009), resulting in a dramatic increase in Amazon EC2 use, from 50–100 instances to 3,400 instances (*AWS* 2008). Had Animoto relied on its internal IT infrastructure, it would have needed to purchase many additional servers, incurring huge fixed costs; however, it still would not have been able to scale up quickly enough, resulting in severely disrupted operations and the loss of tremendous sales opportunities. Recently, cloud service providers have introduced autoscaling, which allows users to scale their computational capacity automatically in real time without having to make capacity precommitments, thereby helping them avoid extra costs from overinvestment (Fazli et al. 2018).

Facilitating Energy-Efficient Production and Operations. Cloud-based SW and applications can help improve operational efficiency and reduce resource waste in production, thereby enhancing energy efficiency. Loukis et al. (2019) suggest that the operational benefits of SaaS can drive firm performance. Using a manager survey, Rodrigues et al. (2021) find that the business value of the operational benefits, cost efficiency, and quality of process and products/services mediates the effect of SaaS on firm performance. Indeed, interviews with senior executives responsible for making decisions on cloud computing investment indicate that cloud computing has transformed firms' business processes and has significantly enhanced efficiencies in terms of cost and time to serve customers (Battleson et al. 2016). According to a survey of product development professionals, cloud-based system users reported higher levels of success (e.g., meeting cost targets, being more innovative, and delivering on time) than users of server-based SW or those with no formal systems (Engineering.com 2018).

It is noteworthy that the operational benefits from IT may not be exclusive to cloud computing. For instance, Han and Mithas (2013) demonstrate that IT outsourcing can save non-IT operating costs by increasing operational efficiencies and allowing the reallocation of internal IT resources. Chou and Chang (2008) suggest that enterprise resource planning (ERP) systems can confer benefits through improved coordination and task efficiency. However, by virtue of subscription-based (or pay-as-you-go) pricing, always-on accessibility, scalability, modularity, and its broader community of users, we argue that cloud-based SaaS plays a more prominent role in conferring operational benefits, thereby reducing waste and energy costs and ultimately enhancing energy efficiency to a greater extent than traditional SW and other IT service outsourcing.

First, SaaS provides low-cost access to expensive, high-powered SW running on the cloud, with a great potential for efficiency gains, which might be unobtainable through in-house development or traditional SW licensing. A case in point is Emerson, a climate system manufacturer (heating, ventilation, air conditioning, and refrigeration). Although Emerson's supermarket customers may not be able to afford large upfront investments for standalone analytics SW, they could use Emerson's products equipped with cloud-enabled big data analytics applications that enable remote diagnosis, maintenance, and repairs. As a result, the customers were able to achieve energy cost savings of more than 10% (*KPMG* 2014).

Second, SaaS allows users not only to access their work anywhere and anytime but also to add new applications, services, or SW features on the fly, which helps reduce or remove any unnecessary energy use caused by traveling or occupying fixed spaces. According to an industry survey (*Business Advantage* 2017), higher mobility and access from anywhere was the primary reason for choosing cloud-based computer-aided design (CAD) programs, followed by the ease of software updates, improved collaboration, better scalability of computing power, storage, and memory, and the ability to pay only for what is used.

Third, SaaS provides a cost-effective and standardized way to rapidly deploy and scale up IT solutions, thereby reducing operating costs and improving efficiency. Iyer and Henderson (2012) illustrate an experience related to implementing a cloud-based human resources management solution, Workday: "Fairchild Semiconductors switched from an on-premises provider of enterprise SW to a cloud service provider, helping it simplify and standardize core human resources business processes for more than 10,000 employees and managers globally … Paul Lones of Fairchild Semiconductor … estimates that the cost to implement and run Workday was about 15% less than buying, installing and maintaining the traditional ERP software. In addition, he estimates that implementing an SaaS solution required about 50% to 70% less time" (p. 55). Another case in point is Volkswagen's development of an industrial cloud on AWS. In their global plants, a large volume of data from several hundred thousand machines and plant items are recorded by sensors and analyzed by standardized apps on their cloud platform; these actions are mainly carried out for the predictive maintenance of machines and the reduction of reworking on vehicles. The company expects implementation of the industrial cloud to save approximately 200 million euros up to the end of 2025 (Volkswagen Newsroom 2020).

Finally, SaaS provides a higher quality alternative to on-premises SW with up-to-date features, which leads to a greater potential for enhancing operational efficiency and reducing waste. Choudhary (2007) argues that the subscription-based licensing model of SaaS gives vendors more incentives to invest in product development than traditional perpetual licensing, leading to higher SW quality. In addition, the modularity of cloud services makes SW updates and maintenance easier (Hardy 2018). Also, collective problem solving from a community of users (typically broader than traditional IT outsourcing clients) is another benefit of cloud computing that can lead to higher quality services; as Iyer and Henderson (2012) put it, "Cloud technology allows multiple users to share data and processes owned by a vendor. The vendors can choose to allow their partners to modify and enhance this shared asset, while allowing all users to enjoy the benefits of continuous improvement" (p. 53).

In addition to SaaS, IaaS provides scalability and flexibility, which allows users to free up and reallocate internal resources to more urgent and mission-critical activities, leading to higher efficiency in production and operations. For example, the ride-sharing company, Uber, highlights the importance of its hybrid cloud model in not only ensuring constant uptime but also facilitating product development and deployment processes (*ZDNet* 2019). Table 1 summarizes the theoretical propositions based on the two mechanisms with anecdotal evidence, which will guide our empirical analyses.

Table 1. Theoretical Framework on Cloud Computing and Energy Efficiency

	SaaS		IaaS	
	Mitigating the energy consumption of IT infrastructure and equipment	Facilitating energy-efficient production and operations	Mitigating the energy consumption of IT infrastructure and equipment	Facilitating energy-efficient production and operations
Potential mechanism	(1)	(2)	(3)	(4)
Theoretical proposition	SaaS running on cloud infrastructure can reduce the energy consumption required to run and support software applications.	SaaS can help redesign the production process to enhance operational efficiency and reduce energy waste by leveraging scalable applications delivered over the Internet on an on-demand basis.	IaaS can substitute cloud-based IT services for an energy-inefficient internal IT infrastructure.	Scalable, on-demand IT resources can increase scalability and flexibility, which can free up and reallocate internal resources to more urgent activities, leading to a higher level of overall efficiency in production and operations.
Anecdotal evidence	• On user devices, the power consumption for running Excel 365 (cloud-based version of Excel) was estimated to be 15.7% (tablet) to 27.3% (laptop) lower than that for running the on-premises Excel 2010 (Williams and Tang 2013).	• In the Volkswagen Group, global plants were integrated into the industrial cloud, improving the efficiency of manufacturing and logistics processes through the predictive maintenance of machines and the reduction of reworking on vehicles (Volkswagen Newsroom 2020). • Emerson's climate systems were equipped with cloud-enabled big data analytics applications that offer remote diagnosis and maintenance, allowing their customers to save energy costs by more than 10% (KPMG 2014). • According to a survey of product development professionals, cloud-based system users could meet cost targets and deliver products on time better than users of server-based software (Engineering.com 2018).	• At data centers, 42% of server operation expenses are tied to energy costs, including direct power consumption (19%) and cooling infrastructure (23%) (Jing et al. 2013), which can be saved by cloud migration. • When Animoto made its service available via Facebook in 2008, the company experienced a demand surge of 750,000 new users in three days. The company coped with the surge by immediately scaling up Amazon EC2 usage from 50–100 to 3,400 instances, instead of equipping itself with additional in-house servers that could incur huge fixed costs and be underutilized during low-demand periods (AWS 2008).	• The ride-sharing company, Uber, leverages its hybrid cloud model not only to ensure constant uptime, but also to facilitate product development and deployment. The company developed their cloud infrastructure to be highly automated, which allows the company to add new features rapidly (ZDNet 2019).

Data and Variables

Data

We use the economy-wide panel data of U.S. private industries over the period 1997–2017, obtained from the Multifactor Productivity (MFP) database of the U.S. Bureau of Labor Statistics (BLS). The MFP database provides the annual output, capital stock, labor costs, and intermediate input costs at the three-digit North American Industry Classification System (NAICS) industry level. We exclude one industry that does not use any IT services at measurable levels—water transportation (NAICS 483; see Table A1 in Online Appendix A for a list of 57 industries). Table 2 summarizes the variables used in our analysis (see Table A2 for the correlations).

For capital inputs, we use data on "productive capital stocks" in 2012 constant dollars, which measure the income-producing capacity of the existing stock during a given period (Stiroh 2002). We measure HW capital as the productive capital stock of "computers" and "communication equipment" in the "information capital" category, and SW capital as the productive capital stock of "software" in the "intellectual property capital" category. We calculate IT capital by adding HW and SW capital, and non-IT capital by subtracting IT capital from the total capital. Non-IT capital includes non-IT equipment (i.e., industrial and transportation equipment), structures (including land), and intellectual property products, excluding SW. Furthermore, labor input is measured as labor compensation.

The MFP database provides information on intermediate inputs purchased from both domestic and offshore suppliers in three categories: energy, materials, and purchased services. We measure an industry's energy consumption by the cost of energy input. Other intermediate inputs are calculated by subtracting the energy cost from the costs of the total intermediate inputs, which include materials and purchased services. It is common in the business and economics literature to use energy expenditure to measure energy use in production (e.g., energy intensity is measured by the energy expenditure over the output in Bloom et al. 2010 and Lyubich et al. 2018). In addition, it is worth noting that detailed products may be assigned to different categories, depending on the industry (Strassner et al. 2005). For instance, although a petroleum-derived product is categorized as an energy input in most industries, it is categorized as a material input in the petroleum-refining and chemical-manufacturing industries. Hence, our measure of energy consumption only captures the energy *consumed* by industries to operate and support their production processes. In converting the nominal costs of intermediate inputs into real values (2012 constant dollars), we multiply nominal values from the base year 2012 by the chain-type quantity indices of each intermediate input. Thus, our measure of energy consumption is proportional to the quantities of energy consumed, with energy prices held constant at the 2012 level. To further control for substitution between production factors because of relative price changes, we also consider the price index of each factor (base year 2012). The industry-level energy price index reflects the composition of energy sources.

Measurement of Cloud Computing

Given that cloud computing is considered as part of IT service outsourcing (Choudhary and Vithayathil 2013), we define an industry's use of cloud computing as cloud-based IT services purchased as an intermediate production input. To the best of our knowledge, there is no existing measure of cloud computing at the aggregate level, which makes it difficult to study the economy-wide impacts of cloud computing. Although some studies rely on survey-based subjective measures (Schniederjans and Hales 2016, Loukis et al. 2019), they have a limitation in evaluating the economy-wide implications of cloud computing. Other studies measure cloud computing via overall IT service expenditures (Jin and McElheran 2019), but doing so is likely to overestimate the use of cloud computing, as it captures both cloud- and non–cloud-based IT services. Thus, we develop a novel industry-level measure of cloud-based IT services by combining industry-level product/service sales data (which provide a detailed description of the products/services that each industry sells) with interindustry purchase flows.

First, we define product/service types corresponding to IT service outsourcing, based on the North American Product Classification System (NAPCS). Prior studies have defined an industry's IT outsourcing as the purchased services from two IT service industries (Han et al. 2011, Qu et al. 2011): *data processing, hosting, and related services* (NAICS 5182) and *computer systems design and related services* (NAICS 5415). Using industry-level product/service sales data obtained from the U.S. Economic Census, we define IT services as product/service types that account for more than 1% of sales in the two IT service industries. Table 3 shows the list of product/service types corresponding to IT services.

Second, we distinguish cloud-based from non–cloud-based IT services. Among the IT services reported in Table 3, we consider application service provisioning (ASP) as SaaS, given that "the SaaS model can be viewed as an evolution of the ASP model" (Loukis et al. 2019, p. 38). By contrast, IT infrastructure provisioning services for website hosting, content streaming, data storage, and so on, are considered IaaS (see Table A3 for detailed descriptions of cloud computing services).[2]

Table 2. Summary Statistics (N = 1,197 for 57 Industries During 1997–2017)

Variable	Mean	Standard deviation	First quartile	Third quartile	Description
Output	346,504.90	329,884.40	96,833.41	501,977.50	Sectoral output by industry
IT capital	20,236.90	45,854.51	1,910	17,132	Sum of HW and SW capital
HW capital	11,105.21	37,341.60	936	8,945	Productive capital stock of computers and communication equipment by industry
SW capital	9,131.69	16,004.80	728	7,963	Productive capital stock of software by industry
Non-IT capital	455,264.20	624,607.60	121,529	465,749	Productive capital stock of total capital, excluding IT capital, by industry
R&D capital	29,213.25	79,655.69	789	21,875	Productive capital stock of research and development by industry
Labor	118,892.60	140,598.20	28,457.99	146,069.7	Labor cost by industry
Energy input	12,276.86	23,935.12	1,749.99	11,772.81	Energy input by industry
Electric energy	4,230.65	23,935.12	619.42	3,905.40	Purchased intermediate input from electric power generation, transmission, and distribution industries (NAICS 2211) by industry
Nonelectric energy	8,073.06	7,915.76	727.26	5927.07	Energy input, excluding electric energy, by industry
Other intermediate inputs	151,496.40	157,244	42,967.70	196,421.50	Intermediate inputs, excluding energy, cloud computing, and non–cloud-based IT services, by industry
Cloud-based IT services	361.83	702.01	30.74	346.03	Sum of SaaS and IaaS
SaaS	177.21	370.31	8.07	156.59	Purchased services of application provisioning, defined in Table 3
IaaS	184.62	337.46	19.17	173.53	Purchased services of IT infrastructure provisioning, defined in Table 3
Non–cloud-based IT services	3,225.72	4,624.51	423.03	3,992.92	Purchased IT services other than cloud computing
Price index of capital input	93.29	28.98	77.59	106.48	Price index of capital input (Base year = 2012)
Price index of labor input	87.78	16.54	75.68	100	Price index of labor input (Base year = 2012)
Price index of energy input	78.35	22.44	58.49	99.23	Price index of energy input among intermediate inputs (Base year = 2012)
Price index of material input	87.27	16.30	76.40	100	Price index of material input among intermediate inputs (Base year = 2012)
Price index of purchased services	91.10	11.21	81.85	100	Price index of purchased services among intermediate inputs (Base year = 2012)

Note. All variables, except the price indices, are in millions of constant 2012 USD.

Our measures of SaaS and IaaS are consistent with commonly accepted definitions (Mell and Grance 2011).[3] It is noteworthy that cloud-based IT services have been available even before the advent of major cloud computing services (e.g., AWS). A case in point is ASP, which has been rapidly replaced by SaaS. Although the ASP model provides computer-based services to clients over a network similar to SaaS, ASP provides monolithic enterprise applications and does not typically provide shared services to multiple clients. However, in the SaaS model, a single instance of an SW application and the supporting infrastructure serve multiple users (Ju et al. 2010). Also, Ju et al. (2010) state that "most ASP-supported applications were immense client-server programs with simple HTML Web interfaces, [but] SaaS solutions today are designed specifically for the Web environment, which improves usability and manageability" (p. 385).

Table 3. Product Types of IT Services and Cloud Computing

			Sales percentage (%)					
			Data processing, hosting, and related services (NAICS 5182)			Computer systems design and related services (NAICS 5415)		
Category	Subcategory	Product type	2002	2007	2012	2002	2007	2012
Cloud-based IT services	SaaS	Application service provisioning	10.5	17.4	22.5	0.3	0.7	1.2
	IaaS	Website hosting services	5.2	6.8	11.2	0.2	0.3	0.4
		Video and audio streaming infrastructure provisioning services	0.3	1.3	3.8	0	0	0.1
		Data storage infrastructure provisioning services	2	2.1	1.5	0.4	0.2	0.3
		IT infrastructure collocation services	6.6	1.1	1.4	0	0	0
		Other IT infrastructure provisioning services	1.3	5.1	2.2	0	0	0
Non–cloud-based IT services		Business process management services	24	32	17	0.4	1.2	1.6
		Data management services	11.8	11.7	14.1	0.3	0.5	0
		Computer systems design, development, and integration services	1.1	0.8	0.4	30.6	36.3	25.3
		Custom application design and development services	4.1	2.6	2.8	28.6	25.8	35
		Network design and development services	1.1	0.4	0.2	3.5	3.5	1.3
		IT infrastructure and network management services	8.7	1.8	2	9.7	6	5.4
		Information and document transformation services	5.5	3.3	1.3	0	0	0
		IT technical consulting services	3.3	1.5	2.2	10.5	4.8	10.1
		IT technical support services	8.3	6.2	9.3	7.4	10.1	11.4
		Temporary staffing-IT staff	0.3	0	0	1.5	1	1
Total cloud services (%)			25.9	33.8	42.6	0.9	1.2	2
Total non–cloud-based IT services (%)			68.2	60.3	49.3	92.5	89.2	91.1
Total IT services (%)			94.1	94.1	91.9	93.4	90.4	93.1
Total non-IT services (%)			5.9	5.9	8.1	6.4	9.6	6.9

Notes. Product types are based on the NAPCS. The sales percentages for each product are obtained from the 2002, 2007, and 2012 Economic Census of the U.S. Census Bureau. See Table A3 in Online Appendix A for detailed descriptions of cloud computing products. See Figures A1 and A2 for trends in cloud computing and non–cloud-based IT services by product type.

Finally, we calculate an industry's purchased cloud-based IT services by combining the product/service sales data with the interindustry purchase flows from input-output tables, which show the output produced by one industry and the intermediate inputs purchased by another industry for each pair of industries (Han et al. 2011). Using input-output use tables provided by BLS,[4] we compute industry i's cloud-based IT services in year t by summing up the intermediate inputs purchased from each supplier industry j weighted by the sales share of cloud computing in industry j:

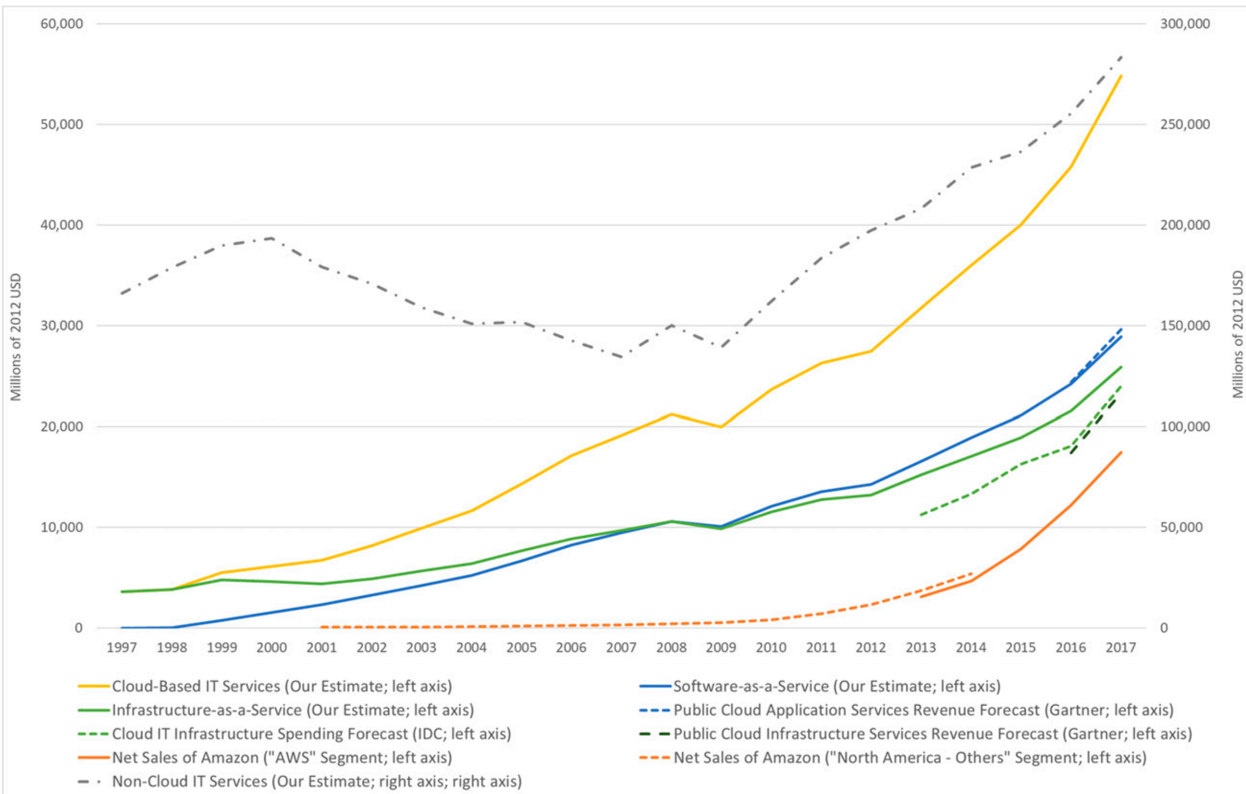
$$\begin{aligned}\text{Cloud}_{it} = \sum_j &(\text{sales share of cloud-based IT services}\\ &\text{in industry } j \text{ in year } t)\\ &\times (\text{intermediate inputs purchased by}\\ &\text{industry } i \text{ from industry } j \text{ in year } t).\end{aligned}$$

Because the Economic Census is conducted every five years, the sales share of cloud-based IT services is linearly interpolated during 1997–2017, based on the 2002, 2007, and 2012 Economic Census data that were available at the time of our analysis.

Figure 1 presents the trends of cloud-based IT services in U.S. industries over the period 1997–2017 (see Figures A1 and A2 in Online Appendix A for trends in IT services by type). This trend illustrates that the use of cloud computing has consistently increased in U.S. industries during the last two decades, and its growth has accelerated since 2010. Cloud-based IT services have increased from USD 3.6 billion in 1997 to USD 20.0 billion in 2009, and to USD 45.8 and 54.8 billion in 2016 and 2017 in the overall U.S. economy.[5] As shown in Figure A3, the industries that rely heavily on cloud services include the administrative and support service industry (NAICS 561), financial institutions (NAICS 521, 522), the broadcasting and telecommunications industry (NAICS 515, 517), and the legal service industry (NACIS 5411).

Our estimate of cloud-based IT services is comparable to the cloud computing market revenue in the United States estimated by Statista (USD 47.3 and 53.2 billion in 2016 and 2017).[6] Moreover, our measures of SaaS and IaaS are comparable to the SaaS and IaaS market revenues in the Unites States estimated by Gartner and International Data Corporation (IDC).[7] For comparison, we also present trends in the net sales of AWS, the global market leader of public cloud services.[8] The net sales of AWS have also increased, especially since 2010, reaching USD 17.5 billion in 2017, which accounts for 31.9% of our estimate. Given that AWS' public cloud market share was estimated at

Figure 1. (Color online) Trends in Cloud-Based IT Services in U.S. Private Industries



Notes. Net sales of Amazon are obtained from its annual reports. Amazon began to report AWS as a separate segment since its 2015 annual report. Until then, AWS sales were included in sales from nonretail activities in the North American segment (North America-Others). In 2013 and 2014, when both AWS and North America-Others statistics are available, AWS sales account for approximately 85% of nonretail sales in North America.

approximately 34%–52% in 2017, our measure of cloud-based IT services appears to represent the total market size of cloud computing in the United States quite well, thereby adding to the validity of the measure.[9]

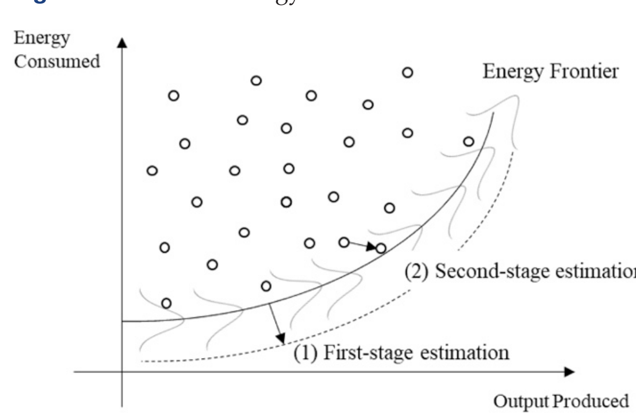
Methodology

Two-Stage Stochastic Frontier Analysis

Prior studies have estimated energy efficiency as a ratio of the optimal-to-actual energy input from the total factor productivity framework (see Filippini and Hunt 2015 for a review on energy efficiency). Thus, we adopt the neoclassical production theory to measure energy efficiency in production. In the production theory, energy inefficiency is defined as a deviation from the best practice, or the energy frontier, which represents the minimum level of energy consumption required to produce a given output using extant production inputs (see Online Appendix C for the theoretical background).

For our empirical analysis, we use a two-stage stochastic frontier analysis (SFA), which has the primary advantage of isolating efficient energy use from the technological change that shapes and shifts the energy frontier. For such a reason, SFA has been widely used to examine the impacts of IT on productivity and various types of efficiency (Lee and Barua 1999, Shao and Lin 2001, Chang and Gurbaxani 2013, Pang et al. 2014). Figure 2 depicts the two-stage stochastic frontier approach where each dot represents an industry-year observation. In the first-stage estimation, we derive an

Figure 2. Stochastic Energy Frontier



energy frontier using SFA. By capturing how far the actual energy consumption is from the given energy frontier, we measure energy inefficiency at the industry-year level, from which we derive energy efficiency. In the second stage, we regress the estimated energy efficiency on the production factors, including cloud-based IT services, to examine the factors' contribution to energy efficiency.

First Stage: Estimating Energy Efficiency

In general, a production technology can be defined as

$$T = \{(X, E, Y) : (X,\ E)\,\text{can produce}\ Y\},$$

where Y is the industry output, and X and E are the production inputs (e.g., capital, labor) and energy input, respectively. The production technology T consists of all feasible input-output vectors for the given technology. Consistent with prior work (Lin and Du 2014), we define the Shephard energy distance function (EDF) (Shephard 1970) as follows:

$$EDF(X, E, Y) = sup\left\{\alpha : \left(X, \frac{E}{\alpha}, Y\right) \in T\right\},$$

where $EDF(X,\ E,\ Y)$ is the ratio of the actual energy consumption to the energy frontier, which reflects the extent to which energy consumption decreases with the extant inputs and output being maintained. If $EDF(X,\ E,\ Y)$ equals one, an industry's technology is on an energy frontier where the industry consumes the minimum level of energy for the given inputs and output; industries on the energy frontier are considered energy efficient, given the extant production technology. When it is greater than one, an industry is considered energy inefficient and can reduce energy proportionally by $EDF - 1$.

Among others, the Cobb-Douglas production function has been used most widely in the IT productivity literature (Brynjolfsson and Hitt 1996, Dewan and Kraemer 2000, Tambe and Hitt 2012). It has been extended to include intermediate inputs as production inputs in addition to capital and labor (Han et al. 2011). In the literature, the "embodied energy" required to produce goods and services (measured using input-output data) is found to play a significant role in productivity (Costanza 1980), and energy has been included in a production function as an input alongside with other intermediate inputs (materials and purchased services) (Bloom et al. 2010, Boyd and Curtis 2014). Following the prior literature, we consider the following extended Cobb-Douglas production function:

$$Y = A \times K^{c_1} L^{c_2} M^{c_3} E^{c_4},$$

where A is a parameter capturing total-factor productivity (TFP), and Y, K, L, M, and E are the output, capital, labor, other intermediate inputs (materials and purchased services), and energy, respectively.

In the energy literature, a cost function–based energy demand model has been adopted to explicitly take the energy price effects into account, which is not possible in a production function (Fisher-Vanden et al. 2004, Hang and Tu 2007, Wurlod and Noailly 2018). From the extended Cobb-Douglas production function, we can obtain the following total-cost function:

$$C(P_K, P_L, P_M, P_E, Y) = \delta\left(\frac{Y}{A}\right)^k P_K^{c_1'} P_L^{c_2'} P_M^{c_3'} P_E^{c_4'},$$

where δ is a constant, P_i is the price for each factor i and its multiplier c_i' is $\frac{c_i}{c_1+c_2+c_3+c_4}$, and k is $\frac{1}{c_1+c_2+c_3+c_4}$. By applying Shephard's lemma, we can derive the cost-minimizing energy demand function to determine the energy frontier as

$$E_{frontier} = \frac{\partial C(P_K, P_L, P_M, P_E, Y)}{\partial P_E} = \delta c_4' \left(\frac{Y}{A}\right)^k P_K^{c_1'} P_L^{c_2'} P_M^{c_3'} P_E^{c_4'-1},$$

where $E_{frontier}$ is the optimal level of energy required to produce the same level of output with the minimum costs.

To capture the energy inefficiency, we incorporate a multiplicative energy distance function with a random error into the energy demand function, E_{it}, for industry i at time t:

$$E_{it} = E_{frontier} \times EDF_{it}(X_{it}, E_{it}, Y_{it}) \times exp(v_{it}), \qquad (1)$$

where $EDF_{it}(X_{it}, E_{it}, Y_{it})$ is defined in the same way as previously, and v_{it} is an idiosyncratic random error. Taking the natural logarithm of the energy demand equation, we can obtain the following stochastic energy frontier model:

$$\begin{aligned}\ln(E_{it}) = \beta + k \ln Y_{it} - k \ln A_{it} + c_1' \ln P_K + c_2' \ln P_L \\ + c_3' \ln P_M + (c_4' - 1) \ln P_E + u_{it} + v_{it},\end{aligned} \qquad (2)$$

where β is a constant and $u_{it} \equiv \ln EDF_{it}\ (X_{it}, E_{it}, Y_{it})$ is a nonnegative random variable associated with time-varying energy inefficiency that follows an exponential distribution.

We model TFP ($\ln A_{it}$) as a function of the IT and research and development (R&D) shares of the total capital (i.e., IT intensity and R&D intensity) and the material and purchased service intensities—the other two categories of intermediate inputs in addition to energy. In the productivity literature, IT intensity has been identified as one of the main determinants of TFP (Gu and Wang 2004, Seo and Lee 2006, Jorgenson et al. 2011). R&D intensity has also been shown to influence TFP (Isaksson 2007). In addition, we further consider intermediate input intensities that are related to TFP growth (Baptist and Hepburn 2013). Specifically, we include material and purchased service intensities because they play a critical role in energy use and carbon emission (hence, the energy frontier) (Allwood et al. 2011).

By defining the TFP as an additive form, we obtain the estimation-friendly energy demand function as a form of the cost frontier (Kumbhakar and Lovell 2000):

$$\begin{aligned} ln(E_{it}) = {} & \beta_i + \beta_1 Y_{it} + \beta_2 \ln(\text{IT share of total capital})_{it} \\ & + \beta_3 \ln(\text{R\&D share of total capital})_{it} \\ & + \beta_4 \ln(\text{Materials share of total cost})_{it} \\ & + \beta_5 \ln(\text{Purchased services share of total cost})_{it} \\ & + \beta_6 \ln P_K + \beta_7 \ln P_L + \beta_8 \ln P_{Materials} \\ & + \beta_9 \ln P_{Purchased\ Services} + \beta_{10} \ln P_E + \tau_t + u_{it} + v_{it}, \end{aligned} \tag{3}$$

where the multipliers of the terms are replaced with estimated coefficients, and the prices of other intermediate inputs are decomposed into those of materials and purchased services. To capture industry-level heterogeneity in the production processes and output elasticity of energy, we use a fixed-effects stochastic frontier model, which Greene (2005a, b) calls the "true" fixed-effects model. This fixed-effects stochastic frontier model allows us to isolate the time-varying inefficiency term after netting out the industry-specific heterogeneity in energy use. We also include year dummies (τ_t) to control for common year-specific shocks (Filippini and Hunt 2015).

From the first-stage estimation based on Equation (3), the energy efficiency (EE), which ranges from zero to one, can be measured as

$$\begin{aligned} EE_{it} &\equiv \frac{1}{EDF_{it}(X_{it}, \mathrm{E}_{it}, Y_{it})} = \frac{\mathbb{E}[E^*_{it} | 0,\ v_{it}, X_{it}, Y_{it}]}{\mathbb{E}[E_{it} | u_{it},\ v_{it}, X_{it}, Y_{it}]} \\ &= exp(-\mathbb{E}[u_{it} | v_{it}]), \end{aligned}$$

where $\mathbb{E}[*]$ is an expected value, and E^*_{it} is the minimum level of energy consumed to produce Y_{it}, given the production inputs (X_{it}) (Battese and Coelli 1988, Lin and Du 2014). If u_{it} is equal to zero (i.e., $EE = 1$), an industry's technology is on the frontier, and the industry consumes the minimum energy level for a given level of inputs and output. Conversely, if u_{it} is greater than zero (i.e., $EE < 1$), it indicates that an industry's production lies above the energy frontier, which means that the industry is energy inefficient.

Second Stage: Estimating Factor Contribution to Energy Efficiency

In the second stage, we model energy efficiency as a function of IT capital and cloud-based IT services. Additionally, we include non-IT capital and other intermediate inputs (including non–cloud-based IT services) to control for capital deepening and production dependence on intermediate inputs, which might influence energy efficiency. Given that energy efficiency ranges from zero to one, we normalize the production factors by dividing each by labor compensation, as in Chang and Gurbaxani (2013).

Specifically, we estimate the following energy efficiency equation:

$$\begin{aligned} EE_{it} = {} & \alpha_i + \alpha_1 EE_{it-1} + \alpha_2 \ln\left(\frac{IT_{it}}{L_{it}}\right) + \alpha_3 \ln\left(\frac{Non-IT_{it}}{L_{it}}\right) \\ & + \alpha_4 \ln\left(\frac{M_{it}}{L_{it}}\right) + \alpha_5 \ln\left(\frac{Cloud_{it}}{L_{it}}\right) \\ & + \alpha_6 \ln\left(\frac{Non-Cloud_{it}}{L_{it}}\right) + \theta_t + \varepsilon_{it}, \end{aligned} \tag{4}$$

where EE_{it} is the energy efficiency estimated in the first stage, and ε_{it} is a random error for industry i at time t. We consider industry fixed effects (α_i) and year dummies (θ_t) to control for industry-level heterogeneity and common shocks (e.g., changes in global energy prices and nationwide environmental policies). We are primarily interested in α_5, which represents the percentage point change in energy efficiency associated with a 1% increase in the intensity of cloud-based IT services.[10]

To account for the potential endogeneity of production factors and the dynamic nature of efficiency, we use a system generalized method of moments (GMM) model to estimate Equation (4) (Arellano and Bond 1991, Blundell and Bond 1998). The system GMM model estimates a system of two equations—the original equation and the first-differenced one—using lagged values in the first differences and levels as internal instrumental variables (IVs), respectively. The system GMM model has been used as an effective way to account for the endogeneity of IT investment in the IT productivity literature (Tambe and Hitt 2012, Chang and Gurbaxani 2013, Chung et al. 2019). At the same time, the system GMM model can also consider the dynamic nature of efficiency by accounting for a lagged efficiency level (EE_{it-1}). Given that successful performance in the past may be a result of superior capabilities or know-how and structural changes with respect to energy utilization, it is reasonable to assume that the current level of energy efficiency is correlated with its past level (Chang and Gurbaxani 2013).

We use the one-step system GMM with robust standard errors. In doing so, we assume that all production factors are endogenous and restrict the number of lags to two to avoid overidentification. In our empirical analyses, we test the validity of the internal instruments of the system GMM by conducting statistical tests for instrument validity (Hansen test of overidentification restrictions) and second-order serial correlation (Arellano-Bond test for AR(2) in the first differences).

Across all estimations, the test results indicate that the instruments are orthogonal to the error terms, and there is no second-order serial correlation, which supports the validity of the system GMM results.

Results

Impacts of Cloud Computing on Energy Efficiency

In the first stage, we estimate the energy frontier model and then measure the energy efficiency (see Online Appendix B for the result of the first-stage estimations). In the second stage, we regress the estimated energy efficiency on the production factor intensities. Table 4 presents the results from the system GMM model. In column 1, our findings demonstrate that cloud-based IT services have significantly contributed to energy efficiency improvement, moving the industries closer to the energy frontier. The results remain unchanged after controlling for non–cloud-based IT services (column 2). Our estimates suggest that doubling the cloud computing services (a 100% increase) leads to a 1.5–2.0-percentage-point increase in energy efficiency. To delve deeper into the distinct roles of different cloud computing types, we separate cloud-based IT services into SaaS and IaaS. The results show that the contribution of cloud computing to energy efficiency is driven mainly by SaaS rather than IaaS (column 3); moreover, the results are virtually identical after further separating IT into HW and SW (column 4).

Time-Split Analysis of Cloud Computing and Energy Efficiency

We examine whether the magnitude of the effect has changed over time as cloud services have advanced and penetrated into the broader economy, based on two notable milestones related to cloud computing: (i) the year 2006 as the beginning of the cloud computing era and (ii) the year 2010 as the start of the rapid growth period with intensified competition.

Although the underlying concepts of cloud computing might trace back to the 1960s, only after 2000 has it emerged as a commercial reality (Qian et al. 2009). In particular, commercial cloud computing services were first introduced by AWS in 2006. AWS delivered the first storage service (Amazon S3) in the spring of

Table 4. Estimation Results of Cloud Computing and Energy Efficiency

	System GMM									
					Time-split analysis					
Dependent variable:	Full sample				1997–2005		2006–2017		2010–2017	
Energy efficiency	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lagged efficiency	0.722***	0.715***	0.713***	0.710***	0.639***	0.610***	0.723***	0.722***	0.748***	0.763***
	(0.051)	(0.052)	(0.048)	(0.048)	(0.097)	(0.091)	(0.046)	(0.039)	(0.071)	(0.056)
IT intensity	−0.009	−0.010**	−0.010**		−0.007		−0.012*		−0.021**	
	(0.006)	(0.004)	(0.004)		(0.010)		(0.007)		(0.010)	
HW intensity				−0.009		−0.024*		−0.010		−0.009
				(0.005)		(0.014)		(0.006)		(0.007)
SW intensity				0.000		0.013		−0.001		−0.006
				(0.004)		(0.012)		(0.005)		(0.007)
Non-IT intensity	0.012*	0.015**	0.016***	0.016***	0.018*	0.021*	0.017**	0.019***	0.023**	0.022***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.011)	(0.007)	(0.006)	(0.010)	(0.007)
Other intermediate inputs intensity	−0.021**	−0.021**	−0.022***	−0.023***	−0.021**	−0.027**	−0.028***	−0.028***	−0.029*	−0.028*
	(0.010)	(0.009)	(0.008)	(0.009)	(0.010)	(0.012)	(0.009)	(0.010)	(0.016)	(0.015)
Cloud-based IT services intensity	0.015***	0.020***			0.016		0.031***		0.038***	
	(0.005)	(0.007)			(0.011)		(0.009)		(0.013)	
SaaS intensity			0.023**	0.019**		0.012		0.028*		0.045**
			(0.009)	(0.009)		(0.012)		(0.014)		(0.020)
IaaS intensity			−0.000	0.001		−0.006		0.006		−0.004
			(0.009)	(0.010)		(0.013)		(0.011)		(0.014)
Non–cloud-based IT services intensity		−0.012	−0.011	−0.010	−0.005	0.003	−0.028***	−0.025***	−0.031***	−0.028***
		(0.009)	(0.008)	(0.008)	(0.013)	(0.011)	(0.010)	(0.008)	(0.011)	(0.010)
Arellano-Bond test for AR(2)	0.833	0.822	0.837	0.833	0.601	0.593	0.398	0.394	0.828	0.830
Instrument validity test	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Observations	1,140	1,140	1,140	1,140	456	456	684	684	456	456

Notes. Robust standard errors are in parentheses. All intensity variables are log-transformed. We report the p values of the test statistics for serial correlation (Arellano-Bond test) and instrument validity (Hansen test of overidentification restrictions), respectively. Year dummy variables are suppressed for brevity.

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

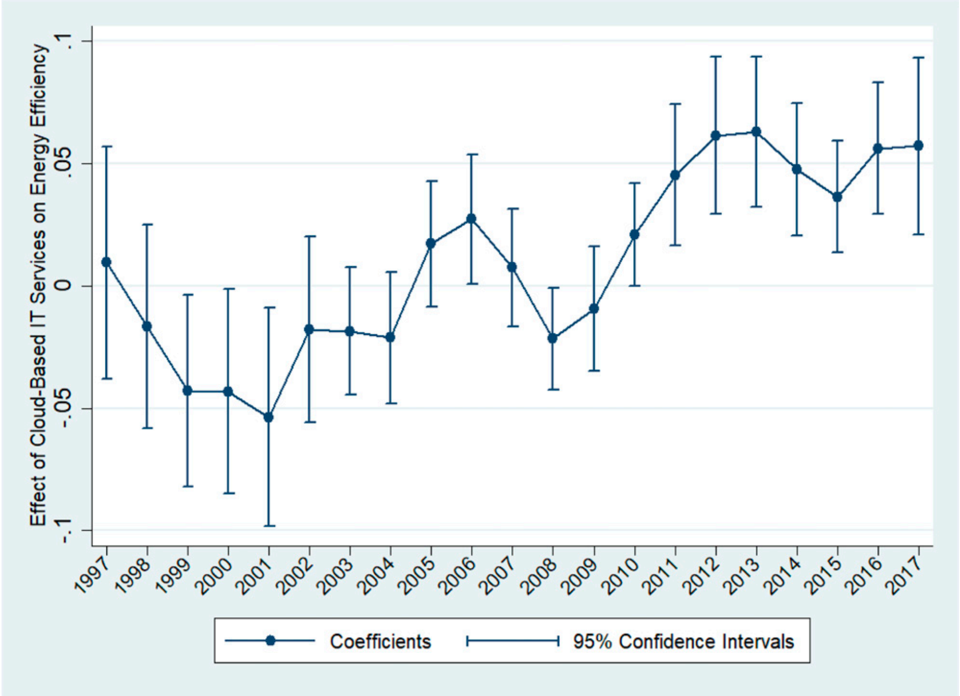
2006 and provided computing capacity (Amazon EC2) in the fall of that year. To further justify the choice of 2006 as the beginning of the cloud computing era, we look into the Google search volume of cloud computing-related terms. From the upper panel of Figure A4 in Online Appendix A, two noteworthy patterns are observed: (i) the terms "Amazon EC2" and "Amazon S3" were first mentioned in 2006 and (ii) the search volumes of "cloud computing" started to soar in 2007. Thus, it is reasonable to assume that cloud computing services began to be commercialized in 2006. This time-split approach is consistent with prior literature; for example, Ewens et al. (2018) separate the period into before and after 2006 when investigating how cloud computing reduces costs for startup firms and thus how it influences venture capital investments.

In Figure A4, the Google search volume of "cloud computing" peaked in 2010, which coincides with the time when the deployment of cloud computing in U.S. industries accelerated (Figure 1). This may be because several other major cloud services have launched since 2010. For instance, Microsoft Azure and Google Cloud, which follow AWS in the public cloud services market, made their services generally available in the spring of 2010 and the fall of 2011, respectively. The rapid deployment of cloud services was driven partially by a sharp decline in prices because of increased competition. Furthermore, advanced features have been continually added to cloud computing services such as autoscaling (Fazli et al. 2018) and predictive scaling powered by machine learning (*AWS* 2018). Because such wide deployment and advanced features of cloud computing may have helped increase its users' energy efficiency, we further compare the effect of cloud computing before and after 2010, as in Jin and McElheran (2019).

The results in columns 5–10 of Table 4 confirm our conjecture. In columns 5 and 6, cloud-based IT services (both SaaS and IaaS) have a positive yet insignificant association with energy efficiency during 1997–2005, whereas their coefficients (particularly SaaS) increase in magnitude and become statistically significant after 2006 (columns 7 and 8). We also find that the effect of cloud computing becomes even more salient after 2010. During 2010–2017, the coefficient of cloud-based IT services is 0.038, meaning that a 100% increase (doubling) in cloud services leads to an *increase* in energy efficiency by 3.8 percentage points.

To investigate long-run cross-sectional effects, we estimate a static model of energy efficiency (i.e., Equation (4) without a lagged term) year by year, an approach used by Dewan and Kraemer (2000). To account for correlations in error terms across years, we estimate them using seemingly unrelated regression (SUR). Figure 3 shows the regression coefficients of cloud computing by year. In keeping with our arguments on the advent and proliferation of cloud computing, the coefficient of cloud services is positive and significant only from 2006 at the 5% level. Although this effect disappears during the

Figure 3. (Color online) Year-by-Year Effects of Cloud-Based IT Services on Energy Efficiency



Note. Coefficients are estimated year by year using SUR.

financial crisis (2007–2009), it becomes significant again from 2010 on.

Taken together, the results of the time-split analysis support our argument that the rapid migration of IT services to the cloud has contributed to energy efficiency improvement in U.S. industries.

Robustness Checks

To lend further credence to our findings, we conduct a series of robustness checks as summarized in Table 5, and detailed descriptions and results are reported in Online Appendix C, D, and E.

First, we test the sensitivity of our results to the modeling choices in the system GMM model. Following recommendations by Cheng and Bang (2021), who critically review the use of the Arellano–Bond GMM in IS research, we re-estimate the system GMM model with different sets of IVs (Table C1) and also replicate the estimations using a difference GMM model (Table C2). In addition, we present the results using alternative panel models, including (i) the fixed-effects model with robust standard errors clustered by industry, (ii) feasible generalized least squares (FGLS), and (iii) ordinary least squares with panel-corrected standard errors (OLS-PCSE) (Table C3). All estimation results of the alternative models are consistent with our main findings from the system GMM model.

Second, we use a difference-in-differences (DID)-style model by exploiting two sources of variation: (i) temporal variation before and after the launch of the first commercial cloud services in 2006 and (ii) cross-sectional variation across industries based on the intensity of cloud-based IT services in the precloud computing era (before 2006). The rationale is that industries that heavily used cloud-based IT services (such as ASP and IT infrastructure provisioning services) in the precloud computing era would have adopted cloud services more aggressively after 2006 when the launch of AWS played a role as an exogeneous shock. Table C4 demonstrates that the industries with a higher proportion of IT outsourcing spending in cloud-based IT services before 2006 experienced a significantly greater increase in energy efficiency in the postcloud computing era (after 2006). As a falsification test, alternative treatments based on the intensity of internal IT capital and general IT outsourcing prior to 2006 yield no significant impact on energy efficiency. These findings support our argument: cloud-based IT services that rapidly migrated to cloud environments after 2006 are indeed one of the main drivers of energy efficiency improvement in U.S. industries. Third, we further address the endogeneity of cloud computing investment by using an industry's prior IT investment and customer (downstream) industries' use of cloud

Table 5. Summary of Robustness Checks

Concern	Test	Location
Sensitivity of system GMM estimations	• Replicate the system GMM estimations with different numbers of instrument variables.	Table C1
	• Estimate the difference GMM model using one-step and two-step estimations.	Table C2
	• Estimate alternative panel models, including the fixed-effects model, FGLS, and OLS-PCSE.	Table C3
Endogeneity of cloud computing investment	• Estimate the DID model by constructing the treatment group with the industries in the top quartile of the percentage of IT outsourcing invested in cloud-based IT services in the precloud computing era (before 2006).	Table C4
	• Estimate the 3SLS model in which an industry's use of cloud services is assumed to be endogenously determined by prior IT investments and downstream customer industries' cloud computing investments.	Table C5
Measurement errors in cloud computing and energy efficiency	• Replicate the system GMM estimations with alternative measures of cloud computing constructed by different interpolation assumptions.	Table C6
	• Replicate the system GMM estimations using a hypothetical measure of cloud computing with simulated measurement errors added.	Figure D1
	• Replicate the system GMM estimations for energy efficiency adjusted by energy use uncorrelated with cloud computing.	Table E1
Model misspecifications	• Replicate the system GMM estimations with an alternative intensity measure by output instead of labor.	Table C7, Columns 1–5
	• Replicate the system GMM estimations after excluding IT service industries.	Table C7, Columns 6–10

Note. All results of the robustness checks can be found in Online Appendices C, D, and E.

services as IVs. Table C5 shows that the results from IV-based three-stage least squares corroborate our main findings.

Fourth, we address potential measurement errors regarding cloud computing. Table C6 shows that the results remain consistent across alternative interpolation methods for measuring cloud-based IT services. In addition, in Online Appendix D, we discuss why a measurement error in cloud services is unlikely to alter our findings. Moreover, in Online Appendix E, we also provide an in-depth theoretical discussion regarding a measurement error in energy efficiency and an additional sensitivity test to demonstrate that the unmeasured portion of energy consumption unrelated to cloud computing is not likely to affect our estimate. Finally, Table C7 confirms that our results are robust to using alternative measures of factor intensity and to excluding a few influential industries that might drive the results.

The aforementioned robustness checks yield results that are consistent with the main findings, thereby strengthening the credence and validity of our findings (see Online Appendices C, D, and E for details).

Investigation of Underlying Mechanisms

As summarized in Table 1, both types of cloud services (IaaS and SaaS) have the potential to contribute to energy efficiency through the two distinct mechanisms. Although it is challenging to disentangle the mechanisms, we attempt to tease them out by empirically investigating (i) how the impact of cloud computing varies by the intensity of the internal HW that encompasses IT infrastructure and equipment and (ii) how this impact varies across different energy types (i.e., electric versus nonelectric energy). Moreover, we conduct a firm-level survey analysis that allows us to directly investigate the underlying mechanisms by measuring factors that could mediate the relationship between cloud computing and energy efficiency.

Interaction Between IT Intensity and Cloud Computing

Given that internal HW constitutes a major portion of IT equipment and infrastructure, we examine how IT intensity (HW intensity in particular) moderates the relationship between cloud computing and energy efficiency. To this end, we first include an additional interaction term involving IT intensity and cloud-based IT services.[11] In column 1 of Table 6, no significant relationship between internal IT intensity and cloud-based IT services is found. Given that SaaS and IaaS play distinct roles in improving energy efficiency as we demonstrated, we interact SaaS and IaaS separately with IT intensity. Interestingly, in column 2, IT intensity positively moderates the effect of IaaS on energy efficiency at the 10% level, but not for SaaS. Finally, we further split IT intensity into HW and SW intensity. In column 3, we find that the interaction of HW intensity with IaaS is positive and significant, but its interaction with SaaS is not significant. In contrast, SW intensity does not play any moderating role in the relationship between cloud-based IT services and energy efficiency. These findings validate our conjecture that IaaS mainly affects energy efficiency by mitigating energy consumption of IT equipment and infrastructure. In column 4, we also interact HW and SW with non–cloud-based IT services but find no significant interaction effects; however, the interaction between HW and IaaS remains significant after controlling for non–cloud-based IT services. In column 5, we use the HW percentage of the total IT capital as an alternative measure of HW intensity and obtain a similar result.

Distinguishing Electric and Nonelectric Energy Efficiency

Given that energy inputs include both electric and nonelectric energy (i.e., fuels), we separate the two types of energy to further illuminate the underlying mechanisms. Specifically, although the first mechanism (i.e., mitigating the energy consumption of IT infrastructure and equipment) is expected to mainly apply to electric energy, the second mechanism (i.e., facilitating energy-efficient production and operations) would concern not only electric energy but also nonelectric energy used in production processes.

We first measure electric energy use as an intermediate input purchased from the industry of electric power generation, transmission, and distribution (NAICS 2211). Then, we calculate nonelectric energy by subtracting electric energy from the total energy input. We estimate electric and nonelectric energy efficiencies separately in the first stage of our energy frontier model, which are subsequently included as dependent variables in the second stage. As reported in Table 7, the results reveal the nuanced roles of different cloud computing types in improving energy efficiency. Whereas the main effect of SaaS on energy efficiency is significant for both types of energy, IaaS only improves electric energy efficiency (but not nonelectric energy efficiency) for industries with high HW intensity.

Taken together, the interaction between internal HW intensity and cloud-based IT services (Table 6) and the breakdown of different energy types (Table 7) indicate that SaaS can facilitate energy-efficient production by helping to optimize operations and redesign production processes that consume electric and nonelectric energy, thereby enhancing the efficiency of both energy types. Conversely, IaaS seems to play a primary role in optimizing the utilization of IT

Table 6. Interactions Between Internal IT Capital and Cloud Computing

	System GMM				
Dependent variable: *Energy efficiency*	(1)	(2)	(3)	(4)	(5)
Lagged efficiency	0.714***	0.715***	0.701***	0.699***	0.700***
	(0.050)	(0.045)	(0.047)	(0.047)	(0.049)
IT intensity	−0.010**	−0.010**			
	(0.004)	(0.004)			
HW intensity			−0.011	−0.011*	−0.008
			(0.007)	(0.006)	(0.006)
SW intensity			0.001	0.004	−0.001
			(0.005)	(0.005)	(0.004)
Non-IT intensity	0.015**	0.015**	0.013**	0.014**	0.013***
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Other intermediate inputs intensity	−0.021**	−0.021**	−0.021***	−0.020**	−0.021***
	(0.009)	(0.0099)	(0.008)	(0.009)	(0.008)
Cloud-based IT services intensity	0.020***				
	(0.007)				
SaaS intensity		0.022***	0.019***	0.021***	0.021***
		(0.008)	(0.007)	(0.008)	(0.008)
IaaS intensity		−0.000	−0.006	−0.008	−0.009
		(0.009)	(0.008)	(0.009)	(0.009)
Non–cloud-based IT services intensity	−0.013	−0.011	−0.009	−0.006	−0.009
	(0.009)	(0.008)	(0.007)	(0.007)	(0.007)
Cloud-based IT services × IT intensity	0.000				
	(0.003)				
SaaS × IT intensity		−0.005			
		(0.004)			
IaaS × IT intensity		0.006*			
		(0.004)			
SaaS × HW intensity			−0.006	−0.006	
			(0.005)	(0.006)	
SaaS × SW intensity			0.001	−0.001	
			(0.004)	(0.004)	
IaaS × HW intensity			0.015***	0.017**	
			(0.006)	(0.007)	
IaaS × SW intensity			−0.010	0.001	
			(0.007)	(0.008)	
Non–cloud-based IT services × HW intensity				−0.005	
				(0.008)	
Non–cloud-based IT services × SW intensity				−0.009	
				(0.010)	
SaaS × HW percentage of IT					−0.018
					(0.018)
IaaS × HW percentage of IT					0.067**
					(0.027)
Arellano-Bond test for AR(2)	0.821	0.835	0.813	0.807	0.821
Instrument validity test	1.000	1.000	1.000	1.000	1.000
Observations	1,140	1,140	1,140	1,140	1,140

Notes. Robust standard errors are in parentheses. All intensity variables are log-transformed. We report the p values of the test statistics for serial correlation (Arellano-Bond test) and instrument validity (Hansen test of overidentification restrictions), respectively. Year dummy variables are suppressed for brevity.

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

equipment and infrastructure—which have been blamed as a major "electricity hog"—and in mitigating the adverse impacts of internal HW on electric energy efficiency.

Firm-Level Survey Analysis

Although our industry-level analysis allows us to estimate the economy-wide effect of cloud-based IT services on energy efficiency, it has a limitation in directly testing the underlying mechanisms through which cloud computing improves energy efficiency. Thus, we conduct a survey analysis by collecting data from business managers to further validate the effects of cloud services on energy efficiency at the firm level. Moreover, the survey method allows us to test the underlying mechanisms more directly by considering factors that could mediate the relationship between cloud computing and energy efficiency.

Table 7. Distinct Effects on Electric and Nonelectric Energy Efficiency

	System GMM							
	Dependent variable: *Electric energy efficiency*				Dependent variable: *Nonelectric energy efficiency*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lagged efficiency	0.721***	0.712***	0.680***	0.688***	0.753***	0.757***	0.749***	0.753***
	(0.038)	(0.038)	(0.044)	(0.042)	(0.049)	(0.046)	(0.045)	(0.046)
IT intensity	−0.004	−0.005			−0.017**	−0.017**		
	(0.004)	(0.005)			(0.008)	(0.007)		
HW intensity			−0.016**	−0.011*			−0.010	−0.005
			(0.007)	(0.006)			(0.011)	(0.009)
SW intensity			0.012**	0.007			−0.008	−0.011
			(0.005)	(0.005)			(0.009)	(0.009)
Non-IT intensity	0.008**	0.010***	0.009**	0.009**	0.013	0.015	0.014*	0.013
	(0.004)	(0.004)	(0.004)	(0.004)	(0.012)	(0.010)	(0.009)	(0.009)
Other intermediate inputs intensity	−0.008	−0.011***	−0.012***	−0.011***	−0.031*	−0.032**	−0.027**	−0.028**
	(0.005)	(0.004)	(0.004)	(0.004)	(0.016)	(0.013)	(0.014)	(0.013)
Cloud-based IT services intensity	0.011*				0.017			
	(0.006)				(0.012)			
SaaS intensity		0.017**	0.010*	0.014**		0.020**	0.018	0.020*
		(0.008)	(0.006)	(0.007)		(0.009)	(0.012)	(0.011)
IaaS intensity		−0.002	−0.008	−0.012		0.001	0.005	0.002
		(0.008)	(0.007)	(0.008)		(0.011)	(0.014)	(0.014)
Non–cloud-based IT services intensity	−0.010	−0.012*	−0.011*	−0.011*	−0.001	−0.004	−0.006	−0.003
	(0.007)	(0.007)	(0.006)	(0.006)	(0.015)	(0.013)	(0.012)	(0.012)
SaaS × HW intensity			−0.014***				0.006	
			(0.005)				(0.006)	
SaaS × SW intensity			0.003				−0.008	
			(0.004)				(0.007)	
IaaS × HW intensity			0.026***				−0.007	
			(0.008)				(0.010)	
IaaS × SW intensity			−0.014				0.016	
			(0.009)				(0.011)	
SaaS × HW percentage of IT				−0.039*				0.029
				(0.020)				(0.028)
IaaS × HW percentage of IT				0.105***				−0.047
				(0.037)				(0.047)
Arellano-Bond test for AR(2)	0.133	0.131	0.116	0.119	0.520	0.523	0.517	0.519
Instrument validity test	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Observations	1,140	1,140	1,140	1,140	1,140	1,140	1,140	1,140

Notes. Robust standard errors are in parentheses. All intensity variables are log-transformed. We report the p values of the test statistics for serial correlation (Arellano-Bond test) and instrument validity (Hansen test of overidentification restrictions), respectively. Year dummy variables are suppressed for brevity.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

In partnership with a professional market research firm in the United States, we collected data from managers at 201 firms, 187 of which are included inthe final analysis after excluding unreliable responses. The sample covers a wide range of organizational sizes and industries. In particular, our sample represents 46 of 57 industries included in the industry-level analysis. No respondents reported "no use of cloud computing services," implying the pervasiveness of cloud computing services across the overall economy. The wide coverage of our sample helps us generalize the results of the survey analysis. In addition, most respondents work in IT-related functions and are top or senior managers, ensuring that our survey participants are knowledgeable about their firm's IT and cloud computing investments, as well as their organizational and environmental performance.

By adapting survey instruments from Khuntia et al. (2018) and Loukis et al. (2019), we measure variables, including the cloud computing expenditure, relative importance of IaaS (as a percentage of the cloud computing budget allocated to IaaS), and performance in terms of energy efficiency. Furthermore, we measure two mediating factors that correspond to the underlying mechanisms we propose in this study: (i) energy reduction in IT equipment and infrastructure and (ii) operational benefits in terms of cost reduction and the improved quality regarding the electronic support of a company's operations and business processes (Loukis et al. 2019). See Online Appendix F for details about the methodology of the survey analysis.

Table 8 reports the estimation results of the causal mediation analysis (Imai et al. 2010). Columns 1–4

Table 8. Estimation Results of Survey Analysis

	Effect of cloud computing on mediator				Effect of cloud computing on energy efficiency					
					Dependent variable: *Energy efficiency*					
	Dependent variable: *Energy reduction in IT equipment/infrastructure*		Dependent variable: *Operational benefits of cloud computing*				Mediator: *Energy reduction in IT equipment/infrastructure*		Mediator: *Operational benefits of cloud computing*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Cloud Computing Expenditure	0.475*** (0.114)	0.023 (0.151)	0.272*** (0.061)	0.167** (0.079)	1.210*** (0.153)	0.500** (0.226)	0.719*** (0.132)	0.478*** (0.177)	0.894*** (0.165)	0.336 (0.211)
Cloud Computing Expenditure × Relative Importance of IaaS		0.617*** (0.126)		0.143*** (0.052)		0.968*** (0.195)		0.375** (0.150)		0.827*** (0.180)
Energy Reduction in IT Equipment/ Infrastructure							1.033*** (0.097)	0.961*** (0.099)		
Operational Benefits of Cloud Computing									1.161*** (0.261)	0.985*** (0.255)
Mediation effect with 95% confidence interval (Imai et al. 2010)										
Percentage of the main effect *(Reflecting the Effect of SaaS)*, mediated by each mediator							40.7% (32.3%, 55.4%)	3.2% (1.6%, 19.9%)	26.0% (21.2%, 35.2%)	32.6% (16.7%, 124.9%)
Percentage of the interaction effect *(Reflecting the Effect of IaaS)*, mediated by each mediator								61.4% (44.4%, 98.2%)		14.4% (10.2%, 22.9%)
Control variable	Firm size									
R^2	0.135	0.237	0.215	0.239	0.338	0.423	0.649	0.660	0.420	0.481
Observations	187	187	187	187	187	187	187	187	187	187

Notes. Robust standard errors are in parentheses. For brevity, we omit firm size dummies.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

show the effect of cloud computing on each mediator, and columns 5–10 present the effects of cloud computing on energy efficiency while accounting for mediating factors. In columns 1 and 3, we find that cloud computing expenditure is positively associated with both energy reduction in IT equipment/infrastructure and operational benefits. To estimate the distinct effects of different types of cloud services, we additionally consider the interaction term involving cloud computing expenditure and the relative importance of IaaS, which represents the effect of IaaS. Given that the relative importance of IaaS is normalized between zero and one as a percentage measure, the coefficient of the main term of cloud computing can be interpreted as the effect of cloud services when IaaS is rarely deployed, which reflects the effect of SaaS. In columns 2 and 4, the results demonstrate that the main term of cloud computing (reflecting the use of SaaS) has a significant effect on operational benefits only. In contrast, the interaction term (reflecting the use of IaaS) is positively associated with both energy reduction in IT equipment/infrastructure and operational benefits, although the former effect is much stronger than the latter.

In columns 5 and 6, we find that cloud computing expenditure is positively associated with energy efficiency for both types of cloud services, as the coefficients of both main and interaction terms are positive and significant. More importantly, columns 7–10 show that these effects are partially mediated. Specifically, we find that 40.7% and 26.0% of the effect of cloud computing services on energy efficiency is mediated by energy reduction in IT and operational benefits, respectively (columns 7 and 9). However, the two mediating factors play distinct roles in different types of cloud services. In column 8, energy reduction in IT equipment and infrastructure mediates 61.4% of the effect of the interaction term (i.e., IaaS effect) on energy efficiency, whereas a very small portion (3.2%) of the effect of the main term (i.e., SaaS effect) is mediated by this factor. Conversely, operational benefits mediate 32.6% and 14.4% of the effects of the main (SaaS effect) and interaction (IaaS effect) terms on energy efficiency, respectively (column

10). The results remain consistent after controlling for general IT expenditure and top management commitment to energy efficiency (see Table F7 in Online Appendix F).

Taken together, the results from the firm-level survey analysis corroborate our industry-level findings and further validate our theoretical expositions on the underlying mechanisms. In sum, non-IaaS cloud services—reflecting SaaS—improve a company's energy efficiency through operational benefits. In contrast, the use of IaaS improves energy efficiency mainly by reducing energy consumption in IT equipment and infrastructure while also contributing, albeit to a lesser extent, to energy efficiency by conferring operational benefits.

Discussion and Conclusions

User-Side Energy Savings vs. Vendor-Side Energy Consumption

To assess the economic significance regarding the impact of cloud computing on energy consumption, we perform a back-of-the-envelope calculation to estimate the amount of energy savings led by the use of cloud computing across U.S. industries. In Equation (1), the actual level of energy consumption (E) is calculated as a hypothetical energy frontier ($E_{frontier}$) divided by energy efficiency (EE). To estimate the contribution of cloud computing to energy consumption, we decompose the energy efficiency into the change in energy consumption because of cloud computing services from the previous year (ΔEE_{cloud}) and the remainder (EE'):

$$E = \frac{\mathbb{E}[E_{frontier}]}{EE} = \frac{\mathbb{E}[E_{frontier}]}{EE' + \Delta EE_{cloud}}.$$

Then, the change in energy consumption because of cloud computing (ΔE_{cloud}), all else being equal, can be calculated as

$$\Delta E_{cloud} = \frac{\mathbb{E}[E_{frontier}]}{EE' + \Delta EE_{cloud}} - \frac{\mathbb{E}[E_{frontier}]}{EE'} = E \times \frac{\Delta EE_{cloud}}{\Delta EE_{cloud} - EE}.$$

According to our energy efficiency equation (Equation (4)), we calculate year-by-year efficiency changes because of cloud computing (ΔEE_{cloud}) by multiplying the coefficient of cloud-based IT services (based on the year-by-year estimates in Figure 3) with the change in the average cloud computing intensity across all industries. From this calculation, the user-side energy cost savings because of cloud computing are estimated to be approximately USD 2.8 to 12.6 billion in 2017 alone, based on the 95% confidence interval. Also, the cumulative energy savings from 2010 to 2017 are estimated to range between USD 13.2 and 52.6 billion.

To weigh these user-side energy savings against vendor-side energy consumption, we estimate the energy consumption by cloud computing vendor industries. Assuming that an industry's energy consumption to produce a product/service is proportional to its sales, we calculate a vendor industry's portion of the energy consumption needed for cloud service provisioning by multiplying its sales share of cloud-based IT services by the industry's total energy costs. According to our estimates, the total user-side energy savings in U.S. industries in 2017 (USD 2.8–12.6 billion) and during 2010–2017 (USD 13.2–52.6 billion) are greater than the total energy costs incurred by cloud computing vendor industries during the same periods (USD 382 million in 2017 and USD 3.5 billion during 2010–2017). It is noteworthy, however, that our data from U.S. industries have some limitations in measuring the energy consumption by cloud service vendors for two reasons. First, our data can only capture the energy consumption within the United States without accounting for the offshore portion of energy consumption (e.g., energy consumption by offshore captive data centers). Second, although we assume that an industry's energy consumption to produce a product/service is proportional to its sales, the energy intensity needed to offer cloud services could differ from that to offer other services or products.

Therefore, we next compare our estimate of energy cost savings from the use of cloud computing to the total electricity consumption by U.S. data centers, estimated by the Lawrence Berkeley National Laboratory (Shehabi et al. 2016). We convert the monetary energy expenditure into electricity use based on the average electricity price in industrial and commercial sectors, obtained from the U.S. Energy Information Administration (EIA). The user-side energy savings from cloud computing in 2017 are equivalent to a reduction in electricity use by 31.8 to 143.8 billion kilowatt-hours, based on the 95% confidence interval, which represents approximately 0.9%–3.9% of the total electricity use in the U.S. economy.[12] This estimate appears to exceed (or at least be comparable to) the total electricity consumption by U.S. data centers in 2014 (70 billion kilowatt-hours) (Shehabi et al. 2016).

Taken together, our analysis demonstrates that the energy consumption by cloud service vendors—an oft-cited culprit for the energy footprint of the IT industry—is more than offset by the reduced energy consumption by cloud computing users. Furthermore, given that firms will use more cloud-based IT services particularly in the postpandemic era and that cloud service vendors have invested in energy efficiency technologies (Masanet et al. 2020) and transitioned to more efficient hyperscale data centers (Jones 2018),[13] we expect that the economy-wide energy saving effects of cloud computing will become stronger over time, thereby helping to usher in an era of sustainable growth.

Theoretical Contributions

Our study makes several important scholarly contributions. First, we contribute to the literature on the value and impacts of IT, wherein an abundance of evidence exists regarding the positive impacts of IT investment and IT outsourcing on economic performance (Melville et al. 2004), technical efficiency (Chang and Gurbaxani 2013), and cost savings (Han and Mithas 2013). However, few empirical works have been conducted to examine the environmental effects of IT in general and cloud computing in particular, despite calls for research on the impacts of IT on environmental performance (Dedrick 2010, Melville 2010, Malhotra et al. 2013, Gholami et al. 2016). By examining the impact of cloud computing on energy efficiency, our study broadens the scope of this research stream into the green value of IT investments, especially related to cloud computing, which constitutes an increasing portion of IT investments. Moreover, our theoretical framework and firm-level survey analysis help address the "how" question by illuminating the mechanisms through which cloud computing drives energy efficiency. Although the first-order effect (increasing energy consumption because of the physical presence of IT) and second-order effect (reducing energy consumption because of the ongoing use and application of IT) of IT have been discussed conceptually (Horner et al. 2016), this study is the first to disentangle these two effects within the context of cloud computing.

Second, we contribute to the nascent literature on cloud computing by taking a perspective involving both clients and vendors. Our approach is similar to the IT outsourcing literature that has considered service venders, clients, and their relationship in reaping the value of IT outsourcing (Levina and Ross 2003, Goo et al. 2009). Given that "corporate users of cloud computing have an active role to play in ensuring that cloud computing ends up delivering on its promise of revolutionizing corporate computing" (Marston et al. 2011, p. 183), there has been a call for research on cloud computing from the corporate users' perspective. However, little empirical work has been conducted regarding the user-side impact of cloud computing on energy use, and our study fills this void. The paucity of empirical research can be attributed to the lack of a proper measure concerning the use of cloud computing. In this regard, our industry-level measure and firm-level survey instrument for the use of cloud computing can be useful for future researchers studying the value and impacts of cloud computing. Furthermore, our proposed approach to categorizing industry-level IT services into distinct types of outsourced IT services (cloud-based IT services versus non–cloud-based IT services), as well as different cloud computing types (SaaS versus IaaS), highlights the importance of considering such differences when studying the impacts of IT outsourcing and cloud computing.

Finally, our study is one of the first attempts in the IS literature to measure energy efficiency based on a panel stochastic frontier model that can account for heterogeneity in energy efficiency across industries and over time while accounting for TFP and factor prices. Because of its solid theoretical foundation, the stochastic frontier approach has been widely applied to various contexts related to IT productivity and technical efficiency in the literature. We believe that our SFA-based two-stage approach can be used to examine diverse aspects of the relationship between IT and energy efficiency at various levels (e.g., plant, firm, and country levels).

Practical Implications

This study provides meaningful practical implications not only for corporate users of cloud computing but also for cloud service vendors and policymakers. From the user's perspective, our findings can assist business managers in gauging the environmental impacts of cloud computing investments and formulating their IT strategies accordingly. Although previous studies have focused on the direct economic benefits of IT outsourcing, such as productivity gains (Han et al. 2011, Chang and Gurbaxani 2012) and cost reductions (Han and Mithas 2013), we suggest that the energy efficiency gains from using cloud computing be taken seriously, especially under the current social and regulatory pressures. In particular, firms can benefit generally from SaaS investments in terms of energy efficiency, although firms heavily relying on HW investments need to invest more in IaaS by migrating their energy-inefficient IT infrastructure into the cloud as the potential efficiency gains from doing so would be much greater than those in less HW-intensive industries.

Our findings also have implications for cloud service vendors that have faced public scrutiny and criticism related to the environmental sustainability of cloud computing. For instance, Greenpeace (2017) annually publishes a "Clean Energy Index" for major IT companies, including Amazon and Google, and its criteria are primarily related to vendor-side practices for reducing energy consumption (e.g., renewable energy commitment, energy efficiency of data centers, and the pollution mitigation of operations). Our findings can help IT managers and related stakeholders deal with societal concerns regarding environmental sustainability by allowing them to evaluate the environmental impacts of cloud computing from a holistic perspective, encompassing both the vendors and clients of cloud computing. Although cloud service vendors need to keep improving the energy efficiency in data centers and related infrastructure that support cloud computing services, they must also devise strategies to leverage the energy saving benefits delivered to their users.

Moreover, our study can inform policymakers by providing novel evidence on the economy-wide energy-saving capacity of cloud computing on the user side. Given that the public discourse to date has focused on the negative environmental impact of cloud computing on the vendor side, our findings regarding its positive user-side environmental impacts can inform the sustainability debate of cloud computing: it is imperative to carefully consider the positive, indirect user-side effects and differentiate them from the negative, direct vendor-side environmental effects of cloud computing. Also, our empirical approach provides guidance for assessing the economy-wide net effect of cloud computing on energy use.

Limitations and Future Research

This study is not without limitations, which point to avenues for future research. First, in addition to what we account for in this paper (i.e., operational benefits from cloud computing, SaaS in particular), there may be unobserved systematic differences in the production processes between firms that invest in cloud computing and those that do not. However, not all unobserved production heterogeneities would confound our estimation results; they would influence the relationship between cloud computing investment and energy efficiency in different ways, depending on their complementarity to cloud computing. On one hand, there could be organizational complements to cloud computing, which tend to be adopted in concert with the use of cloud services (Brynjolfsson and Milgrom 2013), as in the case of traditional IT investments (Bresnahan et al. 2002). For example, IT-related business and management practices that were found to play a significant role in realizing the value of IT investment (Aral et al. 2012, Saunders and Brynjolfsson 2016) could also interplay with cloud computing and energy efficiency, as management practices are often associated with energy efficiency (Martin et al. 2012). Although such complementary factors may lead to heterogeneity in the effectiveness of cloud computing, they would not invalidate the overall contribution of cloud computing to energy efficiency that we found in this study. Given that our results represent the *average* effect of cloud computing on energy efficiency, future research should examine the role of organizational complements to understand organization-level heterogeneity in reaping energy efficiency benefits from cloud computing investment.

On the other hand, some organizational practices and process innovations are not complementary to cloud computing but might coincide with replacing outdated IT systems with cloud-based IT services. For example, digital transformation initiatives might involve not only streamlining business processes and workflows that could potentially improve energy efficiency, but also replacing outdated IT with state-of-the-art cloud services. Although we account for time-invariant heterogeneity across industries using fixed effects and address omitted variable bias using several econometric techniques and a range of robustness tests, we cannot completely rule out the effect of unobserved confounders, which we leave for future research.

Second, although our findings demonstrate the overall contribution of cloud computing and the distinct roles of different types of cloud services in enhancing energy efficiency, we cannot tease out the impacts of different design features and functionalities of cloud computing. In particular, recent technical advances may open up additional opportunities for greater energy efficiency gains through cloud computing. For example, advanced analytics and machine learning applications can be more effective in improving energy efficiency compared with legacy systems. Although such advanced functionalities could be either "rented" via cloud computing or purchased through traditional SW licensing, they are more likely provisioned via cloud computing because it can offer access to advanced service features on demand without large upfront investments for on-premises SW. In this case, firms' improved ability to access such advanced applications would partially explain the greater energy efficiency impact of cloud computing after 2010. Moreover, machine learning and artificial intelligence have empowered cloud computing, which can further enhance the energy efficiency of client organizations. A case in point is AWS' machine learning-powered predictive autoscaling, which provides its users with more opportunities to optimize their cloud computing usage (*AWS* 2018). Drawing on our findings concerning the differential effects of distinct cloud computing types on energy efficiency and its increasing effects over the years, future research can examine how to design and evaluate unique features and functionalities of cloud-based IT services, vis-à-vis traditional on-premises SW, to improve users' energy efficiency.

Third, environmental issues are systematically interconnected, and there have been concerns about the rebound effect (Dimitropoulos 2007), which implies that improved energy efficiency may unexpectedly increase energy consumption in the long term. A case in point involves autoscaling in cloud computing. Although the cloud's autoscaling feature could improve a user's energy efficiency in the short run (as we found in this study), it may also alter the nature of competition and influence entrepreneurial market entry by reducing the operating costs for new entrants (Fazli et al. 2018), which could change the total energy demand by incumbents and entrants in the long run. In addition, given the global distribution of data centers that support cloud services, cloud computing may shift the energy consumption of companies overseas. Thus, future research should consider the rebound effects and global dynamics of energy consumption in relation to cloud computing to advance our understanding of the longer-term, global environmental impact of cloud computing.

Finally, the current structure of NAPCS does not explicitly consider a range of cloud computing services. Although our measure of purchased cloud services based on the product/service types of NAPCS is a reasonable proxy for cloud computing investment, researchers and policymakers must incorporate cloud computing into national accounts (Baer et al. 2020), given the increasing significance of cloud computing in the overall economy.

Concluding Remarks

This study investigated whether and how the use of cloud computing services influences energy efficiency. By using two data sets at the industry and firm levels, we provided what we believe to be the first evidence that the use of cloud computing increases users' energy efficiency by allowing them to use production inputs in a more energy-efficient manner. Moreover, additional analyses revealed that SaaS confers operational benefits to business users by facilitating energy-efficient production across all industries, whereas IaaS helps its users enhance energy efficiency, especially in HW-intensive industries, by mitigating energy consumption of internal IT equipment and infrastructure. Overall, our findings underscore the green value of cloud computing from the business perspective. We hope that our theoretical perspective and empirical approach can provide guidance to IT managers and policymakers regarding how to devise IT strategies and policies to reconcile their economic and environmental goals and can spur future research on the role of IT and IS in environmental sustainability, especially in the era of cloud computing.

Endnotes

[1] Third-order effects refer to the structural and behavioral effects of digital technologies in the long run on transforming economic and industrial structures and altering consumer behaviors, thereby broadly affecting the environment, which we do not consider, as this is beyond the scope of our study.

[2] As shown in Table 3, cloud-based IT services are provided by the data processing and hosting service industry (NAICS 5182; 42.6% of sales), the motion picture, video, and sound recording industries (NAICS 512; 2.1%), the computer systems design service industry (NAICS 5415, 2.0%), other information services (NAICS 519; 0.5%), and software publishers (NAICS 5112; 0.4%). Although business process management and data management services account for a large portion of sales in the data processing and hosting industries (NACIS 5182), we do not consider them as cloud computing because these services cover more comprehensive services, including application and systems design, as well as technical and management consulting.

[3] PaaS is another type of cloud service, although there is no product/service type corresponding to PaaS, according to the current structure of NAPCS. We conjecture that PaaS may be distributed across types of IT infrastructure provisioning services (IaaS), as Armbrust et al. (2010) put it, "the line between 'low-level' infrastructure and a higher-level 'platform' is not crisp. We believe the two are more alike than different" (p. 50).

[4] The raw data of BLS' input-output tables come from the BEA. In this study, we use BLS' input-output tables instead of those from BEA because BLS' input-output tables provide more granular data at the four-digit NAICS level. Thus, to be more precise, we calculate the cloud computing purchases at the four-digit industry level and then aggregate them at the three-digit industry level, which correspond to the production account data (from BLS MFP).

[5] All monetary values based on our measure of cloud computing are based on constant 2012 USD.

[6] Source: https://www.statista.com/forecasts/963837/cloud-services-revenue-in-united-states.

[7] We use the worldwide cloud IT infrastructure spending forecast by IDC (https://www.statista.com/statistics/503686/worldwide-cloud-it-infrastructure-market-spending/) and the worldwide public cloud service revenue forecast for SaaS (Cloud Application Services) and IaaS/PaaS (Cloud Application/System Infrastructure Services) by Gartner (https://www.gartner.com/en/newsroom/press-releases/2017-10-12-gartner-forecasts-worldwide-public-cloud-services-revenue-to-reach-260-billionin-2017), although detailed statistics by country are not available. To infer the cloud market size in the United States, we use Statista's estimations on public cloud revenues by country—the United States accounts for 50.6% of global public cloud revenues, followed by China (8.5%), the United Kingdom (4.4%), and Germany (4.1%) (https://www.statista.com/outlook/16000/100/public-cloud).

[8] Net sales of Amazon are obtained from its annual reports. Amazon began to report AWS as a separate segment since its 2015 annual report. Until then, AWS sales were included in sales from non-retail activities in the North American segment ("North America-Others"). In 2013 and 2014, when both "AWS" and "North America-Others" statistics are available, AWS sales account for approximately 85% of nonretail sales in North America.

[9] For example, see reports by the Synergy Research Group (https://www.srgresearch.com/articles/leading-cloud-providers-continue-run-away-market) and Gartner (https://www.gartner.com/en/newsroom/press-releases/2018-08-01-gartner-says-worldwide-iaas-public-cloud-services-market-grew-30-percent-in-2017).

[10] Because energy efficiency measure is normalized between zero and one, the unit of energy efficiency is a percentage.

[11] For ease of interpretation, we mean-center the variables involving the interaction term so that the main terms can be interpreted as the average effect.

[12] According to EIA, total electricity use in all sectors was 3,723,356 million kilowatt-hours in 2017, and the average electricity price in the industrial and commercial sectors in 2017 was 6.88 and 10.66 cents per kilowatt-hour, respectively.

[13] A report by the Lawrence Berkeley National Laboratory estimates that if 80% of servers in small U.S. data centers were moved over to hyperscale facilities, it would result in a 25% drop in energy use for U.S. data centers (Shehabi et al. 2016). Additionally, some case studies highlight how cloud-based solutions and virtualization could help achieve energy savings in the economy by utilizing a smaller number of cloud servers with high utilization and efficiency, as well as advanced, continuously optimized, and highly efficient cooling systems (Microsoft 2010, Google 2012, Masanet et al. 2013, Williams and Tang 2013).

References

Allwood JM, Ashby MF, Gutowski TG, Worrell E (2011) Material efficiency: A white paper. *Resourse Conservation Recycling* 55(3): 362–381.

Aral S, Brynjolfsson E, Wu L (2012) Three-way complementarities: Performance pay, human resource analytics, and information technology. *Management Sci.* 58(5):913–931.

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econom. Stud.* 58(2):277–297.

Armbrust M, Stoica I, Zaharia M, Fox A, Griffith R, Joseph AD, Katz R, et al. (2010) A view of cloud computing. *Comm. ACM.* 53(4):50.

AWS (2008) Animoto: Scaling through viral growth. Accessed April 3, 2022, https://aws.amazon.com/blogs/aws/animoto-scali.

AWS (2018) Predictive scaling for EC2, powered by machine learning. Accessed April 3, 2022, https://aws.amazon.com/blogs/aws/new-predictive-scaling-for-ec2-powered-by-machine-learning.

Baer A, Lee K, Tebrake J (2020) Accounting for cloud computing in the national accounts. IMF Working Paper No. 20/127, International Monetary Fund, Washington, DC.

Baliga J, Ayre RWA, Hinton K, Tucker RS (2011) Green cloud computing: Balancing energy in processing, storage, and transport. *Proc. IEEE* 99(1):149–167.

Baptist S, Hepburn C (2013) Intermediate inputs and economic productivity. *Philosophical Trans. Roy. Soc. A* 371:20110565.

Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *J. Econometrics* 38(3):387–399.

Battleson DA, West BC, Kim J, Ramesh B, Robinson PS (2016) Achieving dynamic capabilities with cloud computing: An empirical investigation. *Eur. J. Inform. Systems* 25(3):209–230.

Bayramusta M, Nasir VA (2016) A fad or future of IT?: A comprehensive literature review on the cloud computing research. *Internat. J. Inform. Management* 36(4):635–644.

Bloom N, Genakos C, Martin R, Sadun R (2010) Modern management: Good for the environment or just hot air? *Econom. J. (London)* 120(544):551–572.

Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. *J. Econometrics* 87(1):115–143.

Bose R, Luo X (2011) Integrative framework for assessing firms' potential to undertake green IT initiatives via virtualization: A theoretical perspective. *J. Strategic Inform. Systems* 20(1):38–54.

Boyd GA, Curtis EM (2014) Evidence of an 'energy-management gap' in U.S. manufacturing: Spillovers from firm management practices to energy efficiency. *J. Environment. Econom. Management* 68(3):463–479.

Bresnahan TF, Brynjolfsson E, Hitt LM (2002) Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *Quart. J. Econom.* 117(1):339–376.

Brynjolfsson E, Hitt L (1996) Paradox lost? Firm-level evidence on the returns to information systems spending. *Management Sci.* 42(4):541–558.

Brynjolfsson E, Milgrom P (2013) Complementarity in organizations. Gibbons R, Roberts J, eds. *The Handbook of Organizational Economics* (Princeton University Press, Princeton, NJ), 11–55.

Brynjolfsson E, Hofmann P, Jordan J (2010) Cloud computing and electricity: Beyond the utility model. *Comm. ACM* 53(5):32–34.

Business Advantage (2017) *CAD in the Cloud: Market Trends 2017 Report* (Business Advantage).

Chang YB, Gurbaxani V (2012) Information technology outsourcing, knowledge transfer, and firm productivity: An empirical analysis. *Management Inform. Systems Quart.* 36(4):1043–1063.

Chang YB, Gurbaxani V (2013) An empirical analysis of technical efficiency: The role of IT intensity and competition. *Inform. Systems Res.* 24(3):561–578.

Cheng N, Bang Y (2021) A comment on the practice of the Arellano-Bond/Blundell-Bond generalized method of moments estimator in IS research. *Comm. Assoc. Inform. Systems* 48:38.

Chou S-W, Chang Y-C (2008) The implementation factors that influence the ERP (enterprise resource planning) benefits. *Decision Support Systems* 46(1):149–157.

Choudhary V (2007) Comparison of software quality under perpetual licensing and software as a service. *J. Management Inform. Systems* 24(2):141–165.

Choudhary V, Vithayathil J (2013) The impact of cloud computing: Should the IT department be organized as a cost center or a profit center? *J. Management Inform. Systems* 30(2):67–100.

Chung S, Animesh A, Han K, Pinsonneault A (2019) Software patents and firm value: A real options perspective on the role of innovation orientation and environmental uncertainty. *Inform. Systems Res.* 30(3):1073–1097.

Costanza R (1980) Embodied energy and economic valuation. *Science* 210(4475):1219–1224.

Dedrick J (2010) Green IS: Concepts and issues for information systems research. *Comm. Assoc. Inform. Systems* 27(11):173–183.

Dewan S, Kraemer KL (2000) Information technology and productivity: Evidence from country-level data. *Management Sci.* 46(4): 548–562.

Dimitropoulos J (2007) Energy productivity improvements and the rebound effect: An overview of the state of knowledge. *Energy Policy* 35(12):6354–6363.

Engineering.com. (2018) Research report: Identifying the core issues that frustrate product development teams. *Engineering.com*, https://www.engineering.com/story/research-report-the-core-issues-that-frustrate-product-development-teams-pvjlj.

Ewens M, Nanda R, Rhodes-Kropf M (2018) Cost of experimentation and the evolution of venture capital. *J. Financial Econom.* 128(3):422–442.

Fazli A, Sayedi A, Shulman JD (2018) The effects of autoscaling in cloud computing. *Management Sci.* 64(11):5149–5163.

Filippini M, Hunt LC (2015) Measurement of energy efficiency based on economic foundations. *Energy Econom.* 52:S5–S16.

Fisher-Vanden K, Jefferson GH, Liu H, Tao Q (2004) What is driving China's decline in energy intensity? *Resource Energy Econom.* 26(1):77–97.

Fortune (2019) The Internet cloud has a dirty secret. Accessed April 3, 2022, https://fortune.com/2019/09/18/internet-cloud-server-data-center-energy-consumption-renewable-coal.

Gartner (2021) Gartner says four trends are shaping the future of public cloud. Accessed April 3, 2022, https://www.gartner.com/en/newsroom/press-releases/2021-08-02-gartner-says-four-trends-are-shaping-the-future-of-public-cloud.

Gholami R, Watson RT, Hasan H, Molla A, Bjørn-andersen N (2016) Information systems solutions for environmental sustainability: How can we do more? *J. Assoc. Inform. Systems* 17(8):521–536.

Goo J, Kishore R, Rao HR, Nam K (2009) The role of service level agreements in relational management of information technology outsourcing: An empirical study. *MIS Quart.* 33(1):119–145.

Google (2012) *Google Apps: Energy Efficiency in the Cloud* (Google).

Greene W (2005a) Fixed and random effects in stochastic frontier models. *J. Production Anal.* 23(1):7–32.

Greene W (2005b) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *J. Econometrics* 126(2): 269–303.

Greenpeace (2017) *Clicking Clean: Who Is Winning the Race to Build a Green Internet* (Greenpeace, Washington, DC).

Gu W, Wang W (2004) Information technology and productivity growth: Evidence from Canadian Industries. Jorgenson DW, ed. *Economic Growth in Canada and the United States in the Information Age*, Industry Canada Research Monograph Series (Industry Canada, Ottawa), 57–82.

Han K, Kauffman R, Nault B (2011) Returns to information technology outsourcing. *Inform. Systems Res.* 22(4):824–840.

Han K, Mithas S (2013) Information technology outsourcing and non-IT operating costs: An empirical investigation. *Management Inform. Systems Quart.* 37(1):315–331.

Hang L, Tu M (2007) The impacts of energy prices on energy intensity: Evidence from China. *Energy Policy* 35(5):2978–2988.

Hardy Q (2018) How cloud computing is changing management. *Harvard Bus. Rev.*, https://hbr.org/2018/02/how-cloud-computing-is-changing-management.

Hilty LM, Arnfalk P, Erdmann L, Goodman J, Lehmann M, Wäger Pa (2006) The relevance of information and communication technologies for environmental sustainability: A prospective simulation study. *Environ. Modeling Software* 21(11):1618–1629.

Horner NC, Shehabi A, Azevedo IL (2016) Known unknowns: Indirect energy effects of information and communication technology. *Environ. Res. Lett.* 11(10):103001.

Imai K, Keele L, Yamamoto T (2010) Identification, inference and sensitivity analysis for causal mediation effects. *Statist. Sci.* 25(1):51–71.

Isaksson A (2007) Determinants of total factor productivity: A literature review. Working paper, United Nations Industrial Development Organization, Vienna.

Iyer B, Henderson JC (2010) Preparing for the future: Understanding the seven capabilities of cloud computing. *MIS Quart. Executive* 9(2):117–131.

Iyer B, Henderson JC (2012) Business value from clouds: Learning from users. *MIS Quart. Executive* 11(1):52–60.

Jin W, McElheran K (2019) Economies before scale: Survival and performance of young plants in the age of cloud computing. Working paper, Rotman School of Management, University of Toronto, Toronto.

Jing S-Y, Ali S, She K, Zhong Y (2013) State-of-the-art research study for green cloud computing. *J. Supercomput.* 65(1):445–468.

Jones N (2018) How to stop data centres from gobbling up the world's electricity. *Nature* 561(7722):163–166.

Jorgenson DW, Ho MS, Samuels JD (2011) Information technology and U.S. productivity growth: Evidence from a prototype industry production account. *J. Production Anal.* 36(2): 159–175.

Ju J, Wang Y, Fu J, Wu J, Lin Z (2010) Research on key technology in SaaS. *Proc. Internat. Conf. on Intelligent Comput. and Cognitive Inform.* (IEEE, New York), 384–387.

Khuntia J, Saldanha TJV, Mithas S, Sambamurthy V (2018) Information technology and sustainability: Evidence from an emerging economy. *Production Oper. Management* 27(4):756–773.

Klems M, Nimis J, Tai S (2009) Do clouds compute? A framework for estimating the value of cloud computing. *Part of the Lecture Notes in Business Information Processing Book Series,* vol. 22 (Springer, Berlin), 110–123.

KPMG (2014) *2014 Cloud Survey Report: Elevating Business in the Cloud* (KPMG).

Kumbhakar S, Lovell K (2000) *Stochastic Frontier Analysis* (Cambridge University Press, Cambridge, UK).

Lee B, Barua A (1999) An integrated assessment of productivity and efficiency impacts of information technology investments: Old data, new analysis and evidence. *J. Production Anal.* 12(1):21–43.

Levina N, Ross JW (2003) From the vendor's perspective: Exploring the value proposition in information technology outsourcing. *MIS Quart.* 27(3):331–364.

Lin B, Du K (2014) Measuring energy efficiency under heterogeneous technologies using a latent class stochastic frontier approach: An application to Chinese energy economy. *Energy* 76:884–890.

Loukis E, Janssen M, Mintchev I (2019) Determinants of software-as-a-service benefits and impact on firm performance. *Decision Support Systems* 117:38–47.

Lyubich E, Shapiro JS, Walker R (2018) Regulating mismeasured pollution: Implications of firm heterogeneity for environmental policy. *AEA Paper Proc.* 108(2):136–142.

Malhotra A, Melville NP, Watson RT (2013) Spurring impactful research on information systems for environmental sustainability. *Management Inform. Systems Quart.* 37(4): 1265–1274.

Marston S, Li Z, Bandyopadhyay S, Zhang J, Ghalsasi A (2011) Cloud computing—The business perspective. *Decision Support Systems* 51(1):176–189.

Martin R, Muûls M, de Preux LB, Wagner UJ (2012) Anatomy of a paradox: Management practices, organizational structure and energy efficiency. *J. Environ. Econom. Management* 63(2):208–223.

Masanet E, Shehabi A, Lei N, Smith S, Koomey J (2020) Recalibrating global data center energy-use estimates. *Science* 367(6481):984–986.

Masanet E, Shehabi A, Ramakrishnan L, Liang J, Ma X, Walker B, Hendrix V, et al. (2013) *The Energy Efficiency Potential of Cloud-Based Software: A U.S. Case Study* (Lawrence Berkeley National Laboratory, Berkeley, CA).

Mastelic T, Oleksiak A, Claussen H, Brandic I, Pierson J-M, Vasilakos AV (2014) Cloud computing: Survey on energy efficiency. *ACM Comput. Survey* 47(2):1–36.

Mell P, Grance T (2011) *The NIST Definition of Cloud Computing* (National Institute of Standards and Technology, Gaithersburg, MD).

Melville NP (2010) Information systems innovation for environmental sustainability. *Management Inform. Systems Quart.* 34(1):1–21.

Melville NP, Kraemer K, Gurbaxani V (2004) Information technology and organizational performance: An integrative model of IT business value. *Management Inform. Systems Quart.* 28(2):283–322.

Microsoft (2010) *Cloud Computing and Sustainability: The Environmental Benefits of Moving to the Cloud* (Accenture, WSP Environment & Energy, Microsoft).

Murugesan S (2008) Harnessing green IT: Principles and practices. *IT Professional* 10(1):24–33.

Mytton D (2020) Hiding greenhouse gas emissions in the cloud. *Nature Climate Change* 10(8):701.

Pang M-S, Tafti A, Krishnan MS (2014) Information technology and administrative efficiency in U.S. state governments: A stochastic frontier approach. *Management Inform. Systems Quart.* 38(4):1079–1101.

Qian L, Luo Z, Du Y, Guo L (2009) Cloud computing: An overview. *Cloud Computing. CloudCom 2009*, vol. 5931. *Lecture Notes in Computer Science* (Springer, Berlin), 626–631.

Qu WG, Pinsonneault A, Oh W (2011) Influence of industry characteristics on information technology outsourcing. *J. Management Inform. Systems* 27(4):99–128.

Rodrigues J, Ruivo P, Oliveira T (2021) Mediation role of business value and strategy in firm performance of organizations using software-as-a-service enterprise applications. *Inform. Management* 58(1):103289.

Saunders A, Brynjolfsson E (2016) Valuing information technology related intangible assets. *Management Inform. Systems Quart.* 40(1):83–110.

Schniederjans DG, Hales DN (2016) Cloud computing and its impact on economic and environmental performance: A transaction cost economics perspective. *Decision Support Systems* 86: 73–82.

Seo H-J, Lee YS (2006) Contribution of information and communication technology to total factor productivity and externalities effects. *Inform. Tech. Development* 12(2):159–173.

Shao BBM, Lin WT (2001) Measuring the value of information technology in technical efficiency with stochastic production frontiers. *Inform. Software Tech.* 43(7):447–456.

Shehabi A, Smith SJ, Sartor DA, Brown RE, Herrlin M, Koomey JG, Masanet ER, et al. (2016) *United States Data Center Energy Usage Report* (Lawrence Berkeley National Laboratory, Berkeley, CA).

Shephard R (1970) *Theory of Cost and Production Functions* (Princeton University Press, Princeton, NJ).

Society for Information Management (2020) *2020 SIM IT Trends* (Society for Information Management, Mount Laurel, NJ)

Stiroh KJ (2002) Information technology and the U.S. productivity revival: What do the industry data say? *Amer. Econom. Rev.* 92(5): 1559–1576.

Strassner EH, Medeiros GW, Smith GM (2005) *Annual Industry Account: Introducing KLEMS Input Estimates for 1997–2003 Survey of Current Business* (U.S. Department of Commerce, Washington, DC).

Tambe P, Hitt LM (2012) *The productivity of information technology investments: New evidence from IT labor data. Inform. Systems Res.* 23(3-part 1):599–617.

Volkswagen Newsroom (2020) *Volkswagen Steps Up Development of Industrial Cloud* (Volkswagen Newsroom).

Wauters P, Peijl S, Van Der Cilli V, Bolchi M, Janowski P, Moeremans M, Taylor G, et al. (2016) *Measuring the Economic Impact of Cloud Computing in Europe* (European Commission/ Deloitte).

Williams DR, Tang Y (2013) Impact of office productivity cloud computing on energy consumption and greenhouse gas emissions. *Environ. Sci. Tech.* 47(9):4333–4340.

Wurlod J-D, Noailly J (2018) The impact of green innovation on energy intensity: An empirical analysis for 14 industrial sectors in OECD countries. *Energy Econom.* 71:47–61.

ZDNet (2019) Uber vs. Lyft: How the rivals approach cloud, AI, and machine learning. Accessed April 3, 2022, https://www.zdnet.com/article/uber-vs-lyft-how-the-rivals-approach-cloud-ai-machine-learning/.