R&D AND CATCH-UP EFFECT AMONG SOFTWARE-AS-A-SERVICE FIRMS: A STOCHASTIC FRONTIER APPROACH

Chunmian Ge, Department of Information Systems, National University of Singapore, Singapore, chunmian@nus.edu.sg
Ke-wei Huang, Department of Information Systems, National University of Singapore, Singapore, huangkw@comp.nus.edu.sg

Abstract

Since its inception, SaaS market has been one of the fastest growing segments in the software industry. This paper is the first attempt to measure the productivity of pure-SaaS firms by adopting a stochastic frontier approach. Using an annual dataset from 2002 to 2009, we conduct a two-stage analysis. In the first stage, we derive the efficiency scores of each SaaS firm. We found average technical efficiency has been increasing due to catch-up effect in recent years. In the second stage’s analysis, R&D investments are found to be negatively associated with SaaS firm’s technical efficiency. Our results also show that R&D investments significantly contribute to the 1-year growth of technical efficiency of SaaS firms.

Keywords: Software-as-a-service, On-demand computing, Stochastic Frontier analysis, Technical efficiency, R&D, Catchup effect, Spillover.
1 INTRODUCTION

Software-as-a-Service\(^1\) (SaaS) is a relatively new software delivery business model. International Data Corporation (IDC) states that in 2009, worldwide SaaS software generated $13.1 billion in revenue. They forecast that revenue will be $40.5 billion in total by 2014, representing a 25.3% compound annual growth rate. Besides, SaaS revenue will grow at just over five times the growth rate of traditional packaged software\(^2\). To date, the most successful pure-SaaS vendor is Salesforce.com, which delivers customer relationship management (CRM) solutions to business over the Internet. It was founded in 1999 and went public in June 2004. From 2004 to 2009, the revenues of Salesforce.com grew from $176.4 million to $1.3 billion, at an annual growth rate of 49.2%. Since its inception, SaaS market has been one of the fastest growing segments in the software industry and this paper aims at studying the drivers of the sales productivity growth.

The methodology used in this paper is Stochastic Frontier for productivity analysis in economics. Stochastic frontier has been widely adopted in several disciplines (please see Section 2.2 for the details). One main benefit of this approach is that it produces an estimated efficiency score of each SaaS firm in each year. Given these estimated efficiency scores, we can further investigate the dynamics and the drivers of the sales performance differences among pure-SaaS firms. To the best of our knowledge, no other study has used the Stochastic Frontier model to measure the productivity of pure-SaaS firms. This paper is the first attempt to provide additional insights into sources of inefficiency through a panel regression model.

We compile an unbalanced yearly panel dataset of 24 publicly listed pure-SaaS USA firms from 2002 to 2009. Economic value-added (defined as “sales” minus “cost of goods sold” in our analysis) is chosen as the output variable of the production function. Total fixed assets and total number of employees are the input variables of the production function. Our work is then carried out in two stages. In the first stage, we use a stochastic frontier model to construct a production frontier of all pure-SaaS firms and estimate the total factor productivity growth and also each firm’s technical efficiency score. In the second stage, those estimated technical efficiency scores and the growth of efficiency scores are treated as dependent variables and are regressed upon the firm-level explanatory variables to examine the source of technical inefficiency by fixed-effect panel regression models.

Our analysis shows that not only the total factor productivity (TFP) but also the average efficiency score of pure-SaaS firms grows significantly over years. The main explanation is due to the “catching-up” effect. One evidence is that we found that the variance of efficiency scores among pure-SaaS firms reduces over years and those losing firms (less efficient firms) become relatively more efficient over time. To further investigate the drivers behind this phenomenon, we conduct our second-stage regressions and found that: (1) the R&D investment intensity is negatively correlated with technical efficiencies in the same period because less efficient SaaS firms spend more R&D to try to catch up with those more efficient firms. (2) The R&D investment intensity is positively correlated with the 1-year growth rate of technical efficiency score. In other words, firms spend more R&D this year indeed improve more of their relative productivity performance in the following year. We also conduct various robustness checks to confirm the abovementioned findings. Our results also suggest that R&D investment is a more important contributor to the productivity growth than the marketing or advertising expenses.

Understanding the dynamics and the drivers of the sales productivity growth of SaaS firms will not only contribute to the productivity studies of SaaS, but also help SaaS practitioners identify the sources of efficiency, and eventually find out a way to improve their firms’ technical efficiency. For example, if research and development (R&D) investment could be proved to be a main source of technical efficiency for SaaS firms, then implication for placing more investment in R&D could be

---

\(^1\) In this paper, SaaS is defined as a model of software deployment via the Internet whereby the SaaS provider licenses an application to customers as a service based on usage or periodic subscription payments.

\(^2\) http://www.idc.com/research/viewdocsynopsis.jsp?containerId=223628
made. Moreover, R&D investment plays a more important role in contributing to productivity growth according to our study. Therefore, SaaS firms should divide their investment in R&D, marketing, and advertising more properly. These mentioned contributions will also be helpful to the policy decision makers.

The current paper is organized as follows: Section 2 discusses the related literature and presents the hypotheses while Section 3 specifies the empirical model. Section 4 reports the data source and the relevant variables. The main results and related discussions are given in Section 5, with Section 6 concluding the paper.

2 THEORETICAL BACKGROUND

2.1 Production Theory

A production function is a function that specifies the output of a firm, an industry, or an entire economy for all combination of inputs. The inputs used in production process are called factors of production. Typically the inputs consist of capital, labor and others. The economic theory of production places certain technical constraints on the choice of functional form, such as quasi-concavity and monotonicity (Varian 1992). Perhaps the simplest functional form that relates inputs to outputs and is consistent with these constraints is the Cobb-Douglas specification, variants of which have been used since 1896 (Berndt 1991). While this approach is not the only method used for conducting productivity analysis, it is by far the most common functional form used for estimating production functions, calculating the elasticities and marginal products of inputs (Hitt and Brynjolfsson 1996), and remains the standard for studies (Brynjolfsson and Hitt 1996).

The Cobb-Douglas production function with two inputs, capital (K) and labor (L), and one output (Y) can be specified as:

$$ Y_{it} = A K_{it}^{\beta_K} L_{it}^{\beta_L} $$  \hspace{1cm} (1)

Where $Y_{it}$ denotes the output of the $i$-th firm at the $t$-th period. $A$ is a scale factor defined as total factor productivity in the literature. $K_{it}$ and $L_{it}$ represent the capital input and labor input of the $i$-th firm at the $t$-th period.

After taking the logarithms and adding an error term, we had the following estimating equation:

$$ \ln(Y_{it}) = \ln(A) + \beta_K \ln(K_{it}) + \beta_L \ln(L_{it}) + \epsilon_{it} $$  \hspace{1cm} (2)

In this specification, $\beta_K$ and $\beta_L$ represent the output elasticities of capital and labor, which indicate the percentage increase in output associated with a one-percent increase in the corresponding input. For example, the output elasticity of capital, $\beta_K$, represents the percentage increase in output provided by a 1% increase in capital.

Extant literature on IT productivity has examined this logarithm expression by various regression methodologies (Black and Lynch 1996; Dewan and Kraemer 2000; Hitt et al. 2002). This approach is closely related to a literature on the impact of R&D investments on productivity, as well as to a literature on the productivity of information technology investments (Tambe and Hitt 2010). By adopting this approach, Brynjolfsson and Hitt (1996) documented how IS spending had made a substantial and statistically significant contribution to firm output; Kudyba and Diwan (2002) re-examined the productivity paradox with updated data; Aral et al. (2006) found that firms that successfully implement IT, react by investing in more IT; Cheng and Nault (2007) estimated the effects of IT investments made upstream on downstream productivity; Mitt and Nault (2009) studied the indirect impact of IT on the production function at the industry level.
2.2 Stochastic Frontier Approach

The stochastic frontier production function was developed independently by Aigner, Lovell, and Schmidt (1977), and Meeusen and Van den Broeck (1977). Battese and Coelli (1995) defined a stochastic frontier production function for panel data on firms.

\[ Y_{it} = \exp(x_{it}\beta + V_{it} - U_{it}) \]  

(3a)

Where \( Y_{it} \) denotes the output of the \( i \)-th firm (\( i = 1, \ldots, N \)) at the \( t \)-th period (\( t = 1, \ldots, N \)), \( x_{it} \) is a \((1 \times k)\) vector of known functions of inputs of production and other explanatory variables associated with the \( i \)-th firm at the \( t \)-th period. \( \beta \) is a \((k \times 1)\) vector of coefficients to be estimated.

\( V_{it} \) is a random variable that accounts for measurement error and other random factors. It is assumed to be normal distribution and can be positive or negative. \( U_{it} \) is a non-negative random variable. It is assumed to be independently distributed and represents production loss due to firm-specific technical inefficiency. Thus, it is always greater than or equal to zero. More details about the definitions of parameters could be found in Battese and Coelli (1995).

For stochastic frontier production function, the generalized Cobb-Douglas functional form is one of the most frequently used specifications. According to Battese and Coelli (1995), the stochastic frontier production function to be estimated is

\[ \ln(Y_{it}) = \beta_0 + \beta_K \ln(K_{it}) + \beta_L \ln(L_{it}) + V_{it} - U_{it} \]  

(3b)

The main difference between stochastic frontier model and others is that it attributes part of the deviations to technical inefficiency (\( U_{it} \)) and part of the deviations to random noise (\( V_{it} \)). In other words, stochastic frontier approach takes both inefficiency and random noise into account while others do not. It is generally believed that stochastic frontier is a better approach to measure productive efficiency than deterministic frontier (Schmidt 1985).

The technical efficiency of production for the \( i \)-th firm at the \( t \)-th observation is defined by equation (4) below,

\[ TE_{it} = \exp(-U_{it}) \]  

(4)

In our study, stochastic production frontiers will be used to measure the SaaS firm’s productive technical efficiency. Moreover, Kumbhakar et al. (1991) and Battese and Coelli (1995) suggested that determining the factors responsible for inefficiency is an essential component of efficiency analysis. Therefore we will also define an efficiency model later. Before formally introducing the efficiency model, more explanations on technical efficiency are outlined first.

3 HYPOTHESES DEVELOPMENT

3.1 Technical Efficiency and the Catch-Up Effect

Technical efficiency is an important and useful economic measure of firms’ performance. It simply means that firms get the most production from available resources. If firms cannot attain the most production, it is said that they are technically inefficient. Production theory suggests the economic process of transforming different inputs (resources) into outputs. The input-output transformation process can be described by a production frontier, which tells the maximum output that can be achieved given certain inputs. Firms in a certain industry operate either on the frontier, or beneath the frontier (Lin 2009)(see Figure 1). Thus, the difference between the production frontier and a firm’s actual output is referred to as technical inefficiency (line AB). Technical efficiency (especially those derived from Stochastic Frontier) has not been widely used in the study of IT value in the past. It was recently utilized by Lin and Shao (2000) to empirically investigate the business value of IT at the firm level in the MIS literature.
Due to the nascence of SaaS, it is natural to observe differences in technical efficiency among SaaS firms. We define leaders and followers as firms with higher and lower technical efficiency scores, respectively. Therefore, catch-up effects are defined as: the followers gradually improve their technical efficiency to the level of those leaders, and could be measured by technical efficiency (Kumar and Russell 2002).

Firms with relatively low technical efficiency (defined as followers in this paper) typically attempt to spend more resources to catch up with the “leaders” in the same industry. Several authors have documented the existence of the imitation of other followers and the catch-up effects (Bernstein and Nadiri 1988; Jaffe 1986; Romer 1990). However, a follower’s potential for growth weakens as its productivity level converges towards that of the leader (Abramovitz 1986). Assume there is decreasing return of R&D investment and access to the newer generation of technologies, followers will eventually catch up with the leading firms (Eeckhout and Jovanovic 2002). In the stochastic frontier context, this means that all firms will fall on the efficient frontier with 100% efficiency score and zero variance in the ideal case. From the abovementioned literature, we hypothesize that,

\[ \text{H1a: Average pure-SaaS firms’ technical efficiency scores increases over recent years.} \]

\[ \text{H1b: The variance of pure-SaaS firms’ technical efficiency scores decreases over recent years.} \]

3.2 The Drivers of the Catch-up Effect among SaaS Firms

R&D has long been seen as an important source of knowledge generation and productivity improvement (Shell 1966). In research-intensive industries, firms invest in R&D not only to gain immediate profit by selling better products, but also to maintain the level of their R&D technology or Knowledge (Aoki 1991). However, firms undertaking R&D investments are unable to completely appropriate all of the benefits from their R&D projects. The R&D investments by a firm not only reduces its own production cost, but also reduces costs of other firms as a result of spillovers (Bernstein and Nadiri 1988).

Information sources for incoming spillovers are usually situated in the public domain (Cassiman and Veugelers 2002). For example, the large amount of public articles talking about new technological or business model improvements of pure-SaaS firms play as one of the sources for spillovers. Since knowledge is inherently a public good, the existence of technologically related research efforts of other firms may allow a given firm to achieve results with less research effort than otherwise (Jaffe 1986). And that partially explains why that catch-up effect can be created.

We know imitation is much cheaper than innovation (Mansfield 1977). A firm’s own R&D will then be stimulated to capitalize on spillovers, since the firm is more likely to be a follower or imitator with respect to innovation (Cassiman and Veugelers 2006). Therefore, pure-SaaS firms with a lower technical efficiency will tend to place more R&D investments to absorb the external knowledge (spillovers), and then to “catch up” and compete with the more efficient firms.
On the contrary, high productivity firms may have less incentive to invest intensively because the return to further investment is low (Aw et al. 2007). Empirical evidences also confirmed this phenomenon. Cui and Mak (2002) found that in high R&D firms, firms with higher growth rate in assets tend to have lower R&D intensity. Research has been also done and found that the great increase in R&D investments were associated with a stagnant or decreasing productivity growth rate, both at aggregate level (most OECD countries) and firm levels (Jones 2002; Klette and Kortum 2004). The implication of these studies lies in that, firms with a high productivity are likely to reduce their R&D investments since the marginal benefits are now not as attractive as they were.

We therefore believe that less technical efficient pure-SaaS firms will tend to place more R&D investments to catch up with the more efficient ones, while the more technical efficient pure-SaaS firms will tend to place less R&D investments. Based on that, we hypothesize that,

**H2.** R&D investments of SaaS firms are negatively associated with technical efficiency scores in the same period.

### 3.3 R&D Payoff of SaaS Firms

Many researchers have investigated the returns to R&D investments. Extensive empirical studies have been conducted in this domain. Jeffe (1986) studied 432 firms and found that R&D pays off as technological opportunity in an industry increases. Lichtenberg and Siegel (1991) found that in firms that formalize R&D, investments pay off significantly in improved productivity. Graves and Waddock (1994) documented the R&D payoffs in most industries. Ettlie (1998) confirmed R&D intensity was significantly associated with improvements in market share and agility in manufacturing.

The abovementioned evidences suggested that once SaaS firms followers spend relatively larger R&D expense, it is highly possible these R&D investments would have positive payoffs according to literature. If the R&D investment exhibits decreasing return to cumulative R&D investment, SaaS followers can catch up with the leaders in the end. As a result, the technical efficiency scores would increase as time goes by. Therefore we hypothesize that:

**H3.** R&D investments of SaaS firms are positively correlated with the 1-year growth rate of technical efficiency scores.

### 4 EMPIRICAL MODEL SPECIFICATION

For empirical estimations, we perform a two-stage analysis on our data. In the first stage, we adopt Cobb-Douglas production function into the stochastic frontier model. This model is employed to construct a production frontier and estimate the technical efficiency scores of pure-SaaS firms. Based on the estimation we can then calculate the mean, variance and 1-year growth rate of these scores. In the second stage, an efficiency model is specified to examine the source of technical inefficiency for pure-SaaS firms. In this stage, the technical efficiency scores and the 1-year growth of these efficiency scores are separately treated as dependent variables in the efficiency model.

In the first stage, we specify a stochastic frontier production function for the pure-SaaS firms as shown in equation (5):

\[
\ln(Y_{it}) = \beta_0 + \beta_K \ln(K_{it}) + \beta_L \ln(L_{it}) + \theta_t + V_{it} - U_{it}
\]  

(5)

In the second stage, the technical efficiency effects are assumed to be defined by equation (6) and the 1-year growth of technical efficiency are assumed to be defined by equation (7):

**Efficiency Model I:**  
\[
TE_{it} = \delta_0 + \delta_1 (RD_{it}) + \delta_2 (AD_{it}) + \delta_3 (FSIZE_{it}) + \theta_t + \epsilon_{it}
\]  

(6)

**Efficiency Model II:**  
\[
TEG_{it} = \delta_0 + \delta_1 (RD_{it}) + \delta_2 (AD_{it}) + \delta_3 (FSIZE_{it}) + \theta_t + \epsilon_{it}
\]  

(7)

In these two models, we also include advertising expenses as an alternative explanatory variable because marketing expenses and R&D expenses are two critical items on the income statements of enterprise software companies. Firm size is added as a control variable because we observe
diseconomies of scale in the first stage (Huang and Wang 2009). Year dummy variables are included to control for the time trends and unobservable heterogeneity associated with years.

Variables:

- $Y_{it}$ is the economic value added of firm $i$ in year $t$;
- $K_{it}$ is the fix assets of firm $i$ in year $t$;
- $L_{it}$ is the number of employees of firm $i$ in year $t$;
- $\theta_t$ is a set of year dummies indicating 2002-2009;
- $TE_{it}$ is the technical efficiency score obtained from model (5);
- $TEG_{it}$ is the growth of technical efficiency score obtained from model (5);
- $RD_{it}$ is the R&D intensity of firm $i$ in year $t$;
- $AD_{it}$ is the advertising intensity of firm $i$ in year $t$;
- $FSIZE_{it}$ is the firm size of firm $i$ in year $t$;
- $\epsilon_{it}$ is the error term.

5 VARIABLES AND DATA

5.1 Data Source

The list of pure-SaaS firms in our study is obtained from annual industry reports from the Software Equity Group. Later the list is used to retrieve annual financial data from COMPSTAT. It is an unbalanced panel dataset for 24 publicly listed pure-SaaS USA firms from year 2002 to 2009.

5.2 Dependent Variables

The standard output measure used in the literature is economic value added, which is defined as the additional value of the final product over the cost of input materials used to produce it from the previous stage of production (Brynjolfsson and Hitt 1996; Dewan and Min 1997; Kudyba and Diwan 2002). Since the software is unique in that the “input materials from the previous stage” are not well-defined, a simple definition of output is adopted in this paper. Output is operationalized as the total revenue minus the cost-of-goods-sold (COGS) with total revenue deflated by PPI in the software industry and COGS deflated by PPI for intermediate goods (Huang and Wang 2009).

5.3 Independent Variables

The following variables are collected for stochastic frontier model estimation in the first stage: (i) total fix assets (a typical measure of “capital” in the literature), and (ii) number of employees (a typical measure of “labor” in the literature). These two variables are standard inputs in the productivity analysis literature. And all inputs must be measured as real rather than nominal quantities (Lieberman et al. 1990). Accordingly, the capital input must then also be adjusted for inflation. The price deflators used in this study are described in Table 1.

The following variables are collected for efficiency model estimations in the second stage: (i) Research and development (R&D) intensity (a typical measure of R&D investments in the literature), and (ii) Advertising intensity (a typical measure of advertising expense in the literature), and (iii) Firm size as a control variable. For Descriptive statistics, please see Table 2. The correlations are given out in Table 3.

---

3 http://www.softwareequity.com/research_annual_reports.aspx
Variable | Notation | Measurement Construction Process | Deflator |
---|---|---|---|
Output | Y | Total Revenue (revt) minus Cost of Good Sold (cogs), converted to constant 2002 dollars | Producer Price Index for software (NAICS code = 511210) (Bureau of Labor Statistics 2010) |
Labor | L | Total number of employees (emp) | N/A |

Table 1 Data Construction Procedures and Deflators

To be specific, R&D intensity can be measured as firms’ R&D expenses normalized by firms’ annual sales (Dewan and Ren 2009). Advertising intensity can also be measured as advertising expenses divided by firm’s annual sales (Dewan and Ren 2009). Firm size is measured as the natural logarithm of total assets (Finkelstein and Boyd 1998) in this study. All the data are deflated before being carried into model estimations.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value Added</td>
<td>92.201</td>
<td>145.035</td>
<td>4.341</td>
<td>1131.515</td>
</tr>
<tr>
<td>Fix Assets</td>
<td>25.128</td>
<td>57.924</td>
<td>0.498</td>
<td>517.366</td>
</tr>
<tr>
<td>emp</td>
<td>0.623</td>
<td>0.614</td>
<td>0.033</td>
<td>3.969</td>
</tr>
<tr>
<td>RD</td>
<td>0.118</td>
<td>0.068</td>
<td>0.011</td>
<td>0.375</td>
</tr>
<tr>
<td>AD</td>
<td>0.025</td>
<td>0.041</td>
<td>0.000</td>
<td>0.267</td>
</tr>
<tr>
<td>FSIZE</td>
<td>4.308</td>
<td>1.066</td>
<td>1.763</td>
<td>7.508</td>
</tr>
</tbody>
</table>

Table 2 Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Value Added</th>
<th>Fix Assets</th>
<th>emp</th>
<th>RD</th>
<th>AD</th>
<th>FSIZE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value Added</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix Assets</td>
<td>0.945</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>emp</td>
<td>0.935</td>
<td>0.843</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RD</td>
<td>-0.255</td>
<td>-0.239</td>
<td>-0.290</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>0.016</td>
<td>-0.009</td>
<td>-0.001</td>
<td>0.098</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>FSIZE</td>
<td>0.696</td>
<td>0.605</td>
<td>0.773</td>
<td>-0.342</td>
<td>-0.040</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3 Correlations of variables

---

4 Names after variables in the parenthesis are the variable names in Compustat.
6 ANALYSIS

6.1 Technical Efficiency

The task of estimation was carried out by Stata. The dependent variable is economic value added and measured by the logarithm of “Value Added” in the estimation, while the independent variable capital is measured by logarithm of “Fix Assets” and labor by logarithm of “emp”. Year 1 to 8 is the dummies for year 2002-2009. Year 1 stands for year 2002, and year 2 stands for year 2003, and so on. The results of stochastic frontier model estimation are then reported in Table 4.

The significantly positive estimate of the coefficient of capital suggests that the input of capital (fix assets, in our study) has a positive impact on pure-SaaS firms’ output. The coefficient of labor is also significantly positive. However, if we compare the two coefficients, we can find that the output elasticity of labor is much greater than capital’s. The difference between two estimated output elasticities implies that labor factor has played a more important role in the production process of pure-SaaS firms. It seems quite reasonable to us, since software firms rely more on their skilled employees rather than fix assets.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital</td>
<td>0.240704</td>
<td>0.040056</td>
<td>0.000</td>
</tr>
<tr>
<td>Labor</td>
<td>0.680122</td>
<td>0.062581</td>
<td>0.000</td>
</tr>
<tr>
<td>year1</td>
<td>-0.105980</td>
<td>0.156139</td>
<td>0.497</td>
</tr>
<tr>
<td>year2</td>
<td>-0.000830</td>
<td>0.136359</td>
<td>0.995</td>
</tr>
<tr>
<td>year3</td>
<td>-0.131780</td>
<td>0.109535</td>
<td>0.229</td>
</tr>
<tr>
<td>year4</td>
<td>-0.099440</td>
<td>0.097094</td>
<td>0.306</td>
</tr>
<tr>
<td>year5</td>
<td>-0.072760</td>
<td>0.081761</td>
<td>0.373</td>
</tr>
<tr>
<td>year6</td>
<td>-0.006130</td>
<td>0.077906</td>
<td>0.937</td>
</tr>
<tr>
<td>year7</td>
<td>0.025264</td>
<td>0.073795</td>
<td>0.732</td>
</tr>
<tr>
<td>Constant</td>
<td>4.341433</td>
<td>0.176381</td>
<td>0.000</td>
</tr>
<tr>
<td>$\delta_s^2$ ($\delta_s^2=\delta_u^2+\delta_v^2$)</td>
<td>0.159453</td>
<td>0.174175</td>
<td></td>
</tr>
<tr>
<td>$\gamma$ ($\gamma=\delta_u^2/($$\delta_u^2+\delta_v^2$)</td>
<td>0.666044</td>
<td>0.364252</td>
<td></td>
</tr>
<tr>
<td>$\delta_u^2$</td>
<td>0.106202</td>
<td>0.173766</td>
<td></td>
</tr>
<tr>
<td>$\delta_v^2$</td>
<td>0.053250</td>
<td>0.007511</td>
<td></td>
</tr>
</tbody>
</table>

Table 4 Estimation Results of Stochastic Frontier Production Function

Table 4 also shows that the sum of the coefficients is 0.921 and is significantly less than 1, thus pure-SaaS firms exhibit decreasing returns to scale: larger pure-SaaS firms have lower overall productivity. This result is consistent with the findings in Huang and Wang (2009). Interested readers may refer to that paper for the detailed discussions about the economies of scale of SaaS firms.

The average technical efficiency scores of pure-SaaS firms obtained from the stochastic frontier model are presented in Table 5. Their variances and total factor productivity are also listed.

<table>
<thead>
<tr>
<th>Year</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical Efficiency</td>
<td>0.690</td>
<td>0.631</td>
<td>0.688</td>
<td>0.681</td>
<td>0.728</td>
<td>0.756</td>
<td>0.805</td>
<td>0.831</td>
</tr>
<tr>
<td>Variance of Technical Efficiency</td>
<td>0.033</td>
<td>0.038</td>
<td>0.026</td>
<td>0.034</td>
<td>0.027</td>
<td>0.021</td>
<td>0.015</td>
<td>0.011</td>
</tr>
<tr>
<td>TFP Growth Rate</td>
<td>/</td>
<td>0.111</td>
<td>-0.123</td>
<td>0.033</td>
<td>0.027</td>
<td>0.068</td>
<td>0.033</td>
<td>-0.026</td>
</tr>
</tbody>
</table>

Table 5 Average Technical Efficiency Scores and Variance over Year
The average technical efficiency and total factor productivity are both generally increasing over time (See Figure 2). **Hypothesis 1a is supported.** What’s more, we can see from Table 5 that the variance of technical efficiency is generally becoming smaller over time. **Hypothesis 1b is supported.** It tells us that performance differences among pure-SaaS firms are generally getting smaller in the studied years (2002-2009). As we have presented, the creation of catch-up effect helps less efficient pure-SaaS firms to improve their technical efficiency closer and gradually converge to the “leader”. The supported of these two hypotheses can be seen as evidences that the creation of catch-up effect is very important for less efficient pure-SaaS firms.

### 6.2 Sources of Inefficiency

In the second stage, we are interested in the source of technical inefficiency. The results of the efficiency model are reported in Table 6 and Table 7. We estimated 6 models for the efficiency model I (See Table 6, and technical efficiency score is used as the dependent variable). Model 1 is estimated using panel data with fixed effect. Model 2 is estimated using panel data with fixed effect and robust standard errors. Model 3 is estimated using panel data with random effect. Model 4 is estimated like Model 1 but without advertising expenses and firm size. Model 5 is estimated like Model 1 but without R&D investment and firm size. Model 6 is estimated like Model but without firm size.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD</td>
<td>-0.481***</td>
<td>-0.481***</td>
<td>-0.470***</td>
<td>-0.468***</td>
<td>-0.475***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.005)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>-0.227</td>
<td>-0.227</td>
<td>-0.0913</td>
<td>-0.624**</td>
<td>-0.204</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.409)</td>
<td>(0.526)</td>
<td>(0.728)</td>
<td>(0.022)</td>
<td>(0.418)</td>
<td></td>
</tr>
<tr>
<td>FSIZE</td>
<td>-0.00217</td>
<td>-0.00217</td>
<td>0.00703</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.826)</td>
<td>(0.822)</td>
<td>(0.464)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year2</td>
<td>-0.0160</td>
<td>-0.0160</td>
<td>-0.0184</td>
<td>0.0118</td>
<td>-0.0108</td>
<td>-0.0162</td>
</tr>
<tr>
<td></td>
<td>(0.416)</td>
<td>(0.360)</td>
<td>(0.402)</td>
<td>(0.479)</td>
<td>(0.620)</td>
<td>(0.407)</td>
</tr>
<tr>
<td>year3</td>
<td>0.0197</td>
<td>0.0197</td>
<td>0.0130</td>
<td>0.0537***</td>
<td>0.0420*</td>
<td>0.0187</td>
</tr>
<tr>
<td></td>
<td>(0.332)</td>
<td>(0.237)</td>
<td>(0.563)</td>
<td>(0.002)</td>
<td>(0.050)</td>
<td>(0.342)</td>
</tr>
<tr>
<td>year4</td>
<td>0.0501**</td>
<td>0.0501**</td>
<td>0.0384*</td>
<td>0.0781***</td>
<td>0.0791***</td>
<td>0.0482**</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.023)</td>
<td>(0.097)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>year5</td>
<td>0.0953***</td>
<td>0.0953***</td>
<td>0.0787***</td>
<td>0.122***</td>
<td>0.127***</td>
<td>0.0923***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>year6</td>
<td>0.133***</td>
<td>0.133***</td>
<td>0.110***</td>
<td>0.160***</td>
<td>0.171***</td>
<td>0.129***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>year7</td>
<td>0.165***</td>
<td>0.165***</td>
<td>0.143***</td>
<td>0.194***</td>
<td>0.205***</td>
<td>0.161***</td>
</tr>
</tbody>
</table>
**Table 6 Estimation Results of the Efficiency Model I**

We can see from Table 6 that R&D investments do have a negative relationship with pure-SaaS firms’ technical efficiencies in the same period. And the results are quite robust. We can say that, hypothesis 2 is supported. The coefficient of firm size is negative and consistent to our finding above that pure-SaaS firms exhibit decreasing returns to scale, even though the coefficient is not significant. We can also see that the effect of advertising expense is negative, but not significant. The supported of this hypothesis again confirmed that less efficient pure-SaaS firms tend to “catch up” by placing more R&D investments.

We also estimated 6 models for the efficiency model (7) (See Table 7, and 1-year technical efficiency growth rate is used as the dependent variable). The 6 estimated models are the same as we did for model (6).

**Table 7 Estimation Results of the Efficiency Model II**

The results are also quite robust. R&D investments are positively associated with technical efficiency growth in all estimations. Therefore hypothesis 3 is supported. R&D investments pay off significantly in one year. One unit investment in R&D can bring about 0.23 unit increase in technical efficiency. The effect is sufficiently large. For pure-SaaS firms, R&D investments play a critical role in improving technical efficiency. The coefficient of advertising expense is generally positive, but not significant, while the coefficient of firm size is slightly negative and also not significant. Advertising may also have positive effect on technical efficiency improvement. However, we can not confirm that.
CONCLUSION

In this study we have constructed a stochastic frontier model to measure the productivity of 24 pure-SaaS firms in the period from 2002 to 2009. The result shows that capital and labor all played important roles in the production process of pure-SaaS firms, while labor did have much more influence on the outputs of the production. We also observed decreasing returns to scale for pure-SaaS firms. In addition, the technical efficiency (obtained from the stochastic frontier) of pure-SaaS firms is generally increasing in recent years due to catch-up effect. However, the variance of technical efficiency is generally decreasing. In other words, the performance differences between pure-SaaS firms converge in recent years. Later R&D investments are found to be negatively associated with pure-SaaS firm’s technical efficiency. However, R&D investments significantly contribute to the 1-year growth of technical efficiency of pure-SaaS firms. Placing more investments in R&D indeed helps a pure-SaaS firm to improve its technical efficiency.

This paper is the first attempt to measure the productivity of pure-SaaS firms using a stochastic frontier approach. Stochastic frontier approach takes both inefficiency and random noise into account and is a better approach for productivity analysis. It provides additional insights from the perspective of technical efficiency. Our work may shed light on SaaS business model and help improve SaaS firm performance. Practitioners in less efficient pure-SaaS firms may need to place more R&D investments to catch up according to our study, since the return is rather promising. Future research may focus on the difference of productivity dynamics between pure-SaaS firms and traditional software firms. It would be interesting to learn more about the output elasticities of different input factors in these two types of software firms. The source of technical inefficiency may be different. Practical implications could be made through the comparison. Another direction for future research is to investigate the labor mobility among SaaS firms, which could also contribute to the productivity study of SaaS firms.

References


