Asia-Pacific Journal of Accounting & Economics

ISSN: 1608-1625 (Print) 2164-2257 (Online) Journal homepage: www.tandfonline.com/journals/raae20

Studying the production efficiencies of Taiwan's public audit firms under the framework of two-stage models

Yi-Chun Lin & Kuo-Jui Huang

To cite this article: Yi-Chun Lin & Kuo-Jui Huang (11 Jan 2026): Studying the production efficiencies of Taiwan's public audit firms under the framework of two-stage models, Asia-Pacific Journal of Accounting & Economics, DOI: 10.1080/16081625.2025.2612086

To link to this article: https://doi.org/10.1080/16081625.2025.2612086

Published online: 11 Jan 2026.

ASIA-PACIFIC JOURNAL OF ACCOUNTING & ECONOMICS
https://doi.org/10.1080/16081625.2025.2612086

Studying the production efficiencies of Taiwan's public audit firms under the framework of two-stage models

Yi-Chun Lin [a] and Kuo-Jui Huang [b]

[a]Department of Accounting, National Changhua University of Education, Changhua, Taiwan (R.O.C.);
[b]Department of Applied Economics and Management, National Ilan University, Yilan, Taiwan (R.O.C.)

ABSTRACT

We apply a two-stage model to characterize the multistage production process of Taiwan's public audit firms. In the first stage, firms are assumed to hire some ratios of partners and managers (x_1) and fixed assets (x_2) to undertake various cases. In the second stage, firms employ the remaining x_1 and x_2, together with cases and the number of employees, to complete the entrusted cases. We use copula methods to derive the log-likelihood function and find that the smaller firms outperform that of large firms in both stages, and scale and scope economies prevail in either or both stages.

ARTICLE HISTORY
Received 6 October 2024
Accepted 31 August 2025

KEYWORDS
Two-stage model; audit firm; production efficiency; copula methods; scale and scope economies

1. Introduction

Two-stage models are also known as network models. In this context, network data envelopment analysis (network DEA), as first proposed by Färe and Grosskopf (2000), allows the production processes of a firm to be divided into different stages or different departments (sectors) and then measures the production efficiency of different departments. This approach differs from the traditional DEA or stochastic frontier approach (SFA), where a firm is assumed to employ some production factors to produce single or multiple outputs through a specific production process and considers the entire production process as a black box that ignores the interactions among different sectors in the process. Therefore, one can measure a firm's overall efficiency instead of individual sectors. Network models can compensate for traditional methods' deficiencies and help researchers solve classification problems. Specifically, it is sometimes difficult to identify some variables involved in production processes as inputs or outputs. Many empirical researchers use network DEA models to deal with such problems. See Note #[3].

The transparency of the capital market affects national economic development and the international financing activities of firms. Since investors do not engage in corporate operations, auditors must verify financial statements prepared by management to reduce information asymmetry. Audit reports have the essence of public goods since they provide external monitoring functions and facilitate capital market efficiency. The reports offer transparency to participants in financial markets and can be viewed as

CONTACT Kuo-Jui Huang kjhuang@niu.edu.tw Department of Applied Economics and Management, National Ilan UniversityNo. 1, Sec. 1, Shennong Rd., Yilan City 260007, Taiwan (R.O.C.)

Accepted by Hong Hwang upon the recommendation by CHING-I HUANG.

a public good in a capital market. When audit firms conduct auditing jobs, they assign scarce resources to different sectors, including audit personnel with differing specialties, various auditing software and computers, etc. Since audit firms face fierce competition, they must improve their economic efficiency and produce high-quality auditing reports. Economic efficiency is crucial for the sustainability and profitability of the audit firm. Therefore, the government of Taiwan performs operating and financial inspections of all audit firms annually, based on the law of Article 19 of the Taiwan Certified Public Accountant (CPA) Act.

The article authorizes the government to send staff for inspection to safeguard the general public's interests and promote society's good. The competent authority may dispatch personnel to inspect the operations and operations-related financial status of a CPA firm approved to provide attestation services to public companies. A CPA firm may not avoid, impede, or refuse to cooperate with such an inspection. Audit quality can strengthen the transparency and fairness of corporate financial statements through auditing and promote the efficiency of the capital market (Black 2000; Watts and Zimmerman 1983). Audit quality has become the core of supervision by competent authorities worldwide. Referring to the global standard capital market practices, the government of Taiwan also requires corporations to set up audit committees and enforce the regularly rotating audit CPA. In auditing production processes, audit firms hire inputs to transfer into auditing services (outputs) and earn revenue in return. Thus, the efficiency of audit production in the transformation procedure is worth investigating.

Since the primary revenue source of an audit firm comes from their audit fees, there is a close trade-off relationship between audit firms' operation conditions and their audit quality. The British government has enacted laws requiring the Big Four audit firms to issue transparent reports on their operations. The issue of audit firms' organizational efficiency is worth further discussion from the perspective of audit supervision, which is one of the main motives of this study.

Partners and managers in an audit firm are usually responsible for auditing and connecting with clients.[1] These new audit engagements are then transferred to team leaders of the firm, who are in charge of the subsequent jobs, such as planning and executing the engagement signed with clients, to earn audit attestation and non-audit fees.[2] All previous studies assume the above activities to be performed in a single production stage. Note that the number of engagements handled by partners and managers directly influences audit fees, so it should be categorized as input.

The number of audit engagements plays dual roles, i.e. it is an intermediate output in the first stage and an input, similar to the raw materials used by the manufacturing industry, in the second stage. If the number of engagements is considered in traditional single-production stage models, it is difficult to define whether it is an input or output variable. In previous studies on audit production, researchers view the intermediate audit process as a black box, hindering one from observing the entire production process. See, for example, Banker, Chang, and Natarajan (2005) and T.-H. Huang, Chang, and Kuo (2019). This dilemma can be solved using network models described in the previous paragraphs.[3]

This study divides audit firms' production process into two stages under the framework of network models. In the first stage, the audit firm employs specific proportions of two inputs, i.e. the number of partners and managers and total floor area (a proxy variable to capital stock), to undertake clients' audit engagement, and the number of engagements is considered as the sole, intermediate product in this stage. In the second stage, the firm employs four inputs, including the remaining portions of the number of partners and managers and floor area, the total number of team leaders and staff, and the number of engagements, to produce the final output, i.e. audit fees (or audit revenues) and non-audit fees.

We explore the production efficiency of Taiwan's audit firms by the network SFA model, as proposed by T.-H. Huang, Lin, and Chen (2017), T.-H. Huang, Chen, and Lin (2018), and T.-H. Huang, Lin, and Wu (2019). This model is advantageous over the network DEA model in several aspects. First, if there are shared inputs among different production stages, the network SFA model allows one to estimate the ratios of these factors hired in various stages. Researchers must develop economic and econometric models that characterize these multistage production processes and estimate the parameters of interest. Second, and more importantly, the production efficiency of different stages can be linked with a set of environmental variables, which reflect the effects of these variables on the production efficiency of different stages. Researchers may infer valuable operation strategies and policy implications for managers and government authorities based on the parameter estimates.

The rest of this paper is organized as follows. Literature review briefly reviews some relevant papers. Methodology develops our economic and econometric models for regression analysis and establishes three research hypotheses to be tested. Data analysis describes the data and provides sample statistics. Empirical study shows the parameter estimates, discusses the empirical outcomes, and conducts the hypothesis testing raised in Methodology, while the last section concludes this paper.

2. Literature review

2.1. Network models

DEA and SFA are the two most commonly used methods to study decision-making units' (DMUs) efficiency and productivity changes. Since Charnes, Cooper, and Rhodes (1978) and Banker, Charnes, and Cooper (1984) proposed DEA to evaluate DMU's efficiency through linear programming methods under constant and variable returns to scale, respectively, numerous papers have applied the DEA method. Tone (2001, 2002) proposed using slacks-based measures (SBM) of efficiency in DEA to evaluate the efficiency of firms that can validly deal with slacks.[4] Tone and Tsutsui (2010) further proposed the dynamic SBM method. Juo et al. (2016) empirically analyzed the profit efficiency of Taiwan's banks by the SBM method.

Ever since they were proposed by Färe and Grosskopf (2000), network DEA models have received much attention from many scholars, such as Lewis and Sexton (2004), Prieto and Zofio (2007), Kao (2009, 2014), Kao and Hwang (2010), Tone and Tsutsui (2010, 2014), Holod and Lewis (2011), Yang and Liu (2012), Lewis, Mallikarjun, and Sexton (2013), Wang et al. (2014), Fukuyama and Weber (2015), Grosskopf et al. (2015), and Zha et al. (2016), to mention a few.

The DEA model utilizes programming techniques to assess DMU's efficiency and productivity change without specifying the functional form for production or cost functions. However, it has several disadvantages, such as its estimates lack statistical properties and are easily affected by random factors (O'Donnell and Coelli 2005). Another downside is that it is difficult to directly explore the influences of various environmental and management factors on (in)efficiency. Although Simar and Wilson (2007) suggested using bootstrapping to correct the efficiency estimates obtained from the DEA model and the standard error of the second-stage truncated regression analysis, their procedures are difficult to apply straightly to network DEA.

Numerous researchers have employed SFA, pioneered by Aigner, Lovell, and Schmidt (1977) and Meeusen and Van Den Broeck (1977), to investigate a firm's technical efficiency. C. J. Huang, Huang, and Liu (2014) extended the (deterministic) metafrontier production model, developed by Battese, Rao, and O'Donnell (2004), to the stochastic metafrontier framework.[5] Kutlu (2010), Tran and Tsionas (2013, 2015), and Amsler, Prokhorov, and Schmidt (2016, 2017) brought the endogeneity of production factors into the efficiency evaluation model. Notably, the DEA cannot take endogenous inputs into account. From this angle, SFA is preferable to DEA.

Given the shortcomings of the DEA models, this study instead exploits the network SFA model, as developed by T.-H. Huang, Lin, and Chen (2017), to examine the production efficiency of Taiwan's audit firms. To date, only a few papers have applied this model to probe the bank's efficiency in China (T.-H. Huang, Lin, and Chen 2017) and in the United States (T.-H. Huang, Chen, and Lin 2018), and the insurer's efficiency in Taiwan (T.-H. Huang, Lin, and Wu 2019).

2.2. Efficiency of audit firms

Fewer papers address the efficiency of audit firms, likely because most are privately held, rendering their accounting data inaccessible to the public. The lack of input-output variables impedes researchers from conducting empirical studies on the performance of audit firms. A few researchers have extracted data on audit firms from the top 100 audit firms in various countries and applied DEA with these small samples to investigate multiple performances from different perspectives. For example, Banker, Chang, and Natarajan (2005), H. Chang, Choy, Cooper, Parker, et al. (2009, 2009), and Knechel, Rouse, and Schelleman (2009) estimated the efficiency of the top 100 audit firms in the U.S. H. Chang et al. (2011) concerned how information and human capital affect the productivity of 51 audit firms in Taiwan.

B.-G. Chang, Huang, and Kuo (2015) and T.-H. Huang, Chang, and Kuo (2019) carried out cross-country comparisons and studied the differences in efficiency and productivity change of the top 100 audit firms across Taiwan, the United States, and China by the stochastic metafrontier model, as proposed by C. J. Huang, Huang, and Liu (2014). Unlike the papers cited in the previous paragraph, which use DEA models, these two papers use SFA to perform empirical studies. In addition, Dopuch et al. (2003) explored the relationship between the technical efficiency of audits and audit fees. Banker, Chang, and Natarajan (2007) studied the growth of allocative efficiency (revenue efficiency) and technical efficiency after the merger of audit firms. The above studies used the one-stage DEA to assess the efficiency of audit firms, which can neither observe the actual conversion process from inputs to outputs in different production stages nor overcome the problem that evaluation results are susceptible to the effects of data noise. The network SFA model can solve the two issues above; thus, the empirical results are likely more reliable.

3. Methodology

3.1. Theoretical model

This paper attempts to construct an economic model that assumes that the entire production procedure of audit firms consists of two stages. The subscript for individual firms is omitted for simplicity. In the first stage, the firms employ two factors, including a fraction (α_1) of the number of partners and managers (x_1) and a fraction (α_2) of total floor area (x_2), to undertake cases (x_3), which is the single, intermediate output in this stage. The fractions of α_1 and α_2 are unknown parameters between zero and one to be estimated. Their sizes show how much of the corresponding two inputs are used in the sector and reveal the relative importance of the sector in the entire production process. The greater the values of α's, the more critical the sector will be.

In the second stage, the firms utilize the unused portions of the two inputs, i.e., $(1-\alpha_1)x_1$ and $(1-\alpha_2)x_2$, together with total team leaders and staff (x_4) and the number of cases to produce two final goods, i.e. audit fees (y_1) and non-audit fees (y_2). Let $X = (\alpha_1 x_1, \alpha_2 x_2)'$. Following T.-H. Huang, Lin, and Chen (2017), the stochastic production frontier model for the number of cases is formulated as follows:

$$x_3 = g(X)e^{\varepsilon_1}, \tag{1}$$

where $g(\cdot)$ represents the production function, $\varepsilon_1 = v_1 - u_1$ is the composed error, $v_1 \sim N(0, \sigma^2_{v_1})$ is the disturbance term, $u_1 \sim N^+(0, \sigma^2_{u_1})$ is the one-sided error that signifies technical inefficiency, and v_1 and u_1 are assumed to be statistically independent.

The presence of u_1, arising from managerial inabilities, causes the realized output of x_3 to be below the potential output of $g(\cdot)e^{v_1}$. Empirical study will take the natural logarithms on both sides of Equation 1 and let $\ln g(\cdot)$ be the translog function, as commonly used in the literature. The fractional parameters α's act as a bridge between the first and second production processes. However, almost all databases fail to offer such information in firms' production stages. Researchers must rely on an econometric approach to estimate them based on appropriate econometric models and data.

Note that estimating Equation 1 alone cannot identify the extra parameters of α's. We solve this problem using a cost function to describe the production technology of the second stage.[6] In this manner, we can derive the corresponding cost share equations, which offer additional information to identify α's. Assuming that, in the second stage, the firms exploit the leftover portions of x_1 and x_2, i.e., $(1-\alpha_1)x_1$ and $(1-\alpha_2)x_2$, the intermediate output x_3 produced in the first stage, and x_4 to supply y_1 and y_2 in an attempt to minimize costs. Let $PC^*(\cdot)$ be the optimal cost, which can be expressed as:

$$PC^*\left(Y, \frac{W}{b}\right) = \min_{b\tilde{X}} \left[\frac{W'}{b}\left(b\tilde{X}\right) \middle| H\left(b\tilde{X}, Y\right) = 0 \right]$$

$$= \frac{1}{b} PC(Y, W), \tag{2}$$

where $\tilde{X} = [(1-\alpha_1)x_1, (1-\alpha_2)x_2, x_3, x_4]' = [\tilde{\alpha}_1 x_1, \tilde{\alpha}_2 x_2, \tilde{\alpha}_3 x_3, \tilde{\alpha}_4 x_4]$, with $\tilde{\alpha}_3 = \tilde{\alpha}_4 = 1$, W is a 4×1 input price vector, $H(\cdot)$ is the production transformation function, and parameter b $(0 < b \leq 1)$ denotes the degrees of technical (in)efficiency in the second sector

(Atkinson and Cornwell 1993, 1994). The closer the value is to 1, the higher the firm's technical efficiency will be, and vice versa. Equation 2 says that the firm can optimize its cost by minimizing the expenditure of $W'\tilde{X}$ subject to producing a given level of outputs (Y), i.e. $H\left(b\tilde{X}, Y\right) = 0$, where the firm may suffer from input-oriented technical inefficiency equaling b.

We choose not to show the entire derivation for brevity. Readers are suggested to refer to T.-H. Huang, Lin, and Chen (2017). Using Shephard's lemma, we can prove the share equation of the k^{th} input to be:

$$S_k(Y, W) \equiv \frac{\partial \ln PC^*}{\partial \ln W_k} = \frac{\partial PC^*}{\partial W_k}\frac{W_k}{PC^*} = \frac{\tilde{\alpha}_k x_k W_k}{PC^*}, k = 1, \ldots, 4. \tag{3}$$

We can then relate the actual expenditure (E) to the share equations:

$$E = \sum\nolimits_k W_k x_k = \sum\nolimits_k W_k \frac{PC^* S_k}{\tilde{\alpha}_k W_k} = PC^* \sum\nolimits_k S_k \tilde{\alpha}_k^{-1}. \tag{4}$$

Taking the natural logarithm on both sides of Equation 4, we yield:

$$\ln E = \ln PC^*\left(Y, \frac{W}{b}\right) + \ln \sum\nolimits_k S_k \tilde{\alpha}_k^{-1}$$

$$= \ln PC(Y, W) - \ln b + \ln \sum\nolimits_k S_k \tilde{\alpha}_k^{-1}$$

$$= \ln PC(Y, W) + \ln \sum\nolimits_k S_k \tilde{\alpha}_k^{-1} + u_2, \tag{5}$$

where $u_2 \equiv -\ln b \sim N^+\left(0, \sigma_{u_2}^2\right)$ is a nonnegative random variable, signifying the increase in actual expenditure incurred by technical inefficiency in the second sector.

We can write the actual share equation of the k^{th} input as:

$$\frac{W_k x_k}{E} = \frac{W_k}{E}\frac{S_k PC^*}{\tilde{\alpha}_k W_k} = \frac{S_k \tilde{\alpha}_k^{-1}}{\sum_k S_k \tilde{\alpha}_k^{-1}}, k = 1, \ldots, 4. \tag{6}$$

Equation 6contains four share equations because the firm is assumed to hire four inputs to proceed with production activities in the second stage. Equation 5 and Equation 6 connect the realized expenditure and shares with the optimal cost (PC^*) and shares (S_k). Notably, the share equations of Equation 6 contain fractional parameters α's, and therefore, we can use them to help identify α's. After appending an error term to each of the two equations to capture random shocks that are uncontrollable by the firm, we can estimate them to obtain all parameters of interest, including α_1 and α_2, using a simultaneous equation model. Specifically, we add $v_2 \sim N(0, \sigma_{v_2}^2)$ to Equation 5, and random disturbances $\varepsilon_i \sim N(0, \sigma_{\varepsilon_i}^2)$, $i = 3, \ldots, 6$, to Equation 6 since there are four inputs. Again, $u_2$andv_2 are assumed to be statistically independent and $\varepsilon_2 = u_2 + v_2$. In Empirical study, $\ln PC(Y, W)$ will be specified as the translog form. Since the sum of the four shares equals unity, we can include only three of them in the simultaneous regression model to avoid the singularity of the covariance matrix of ε_i.

In summary, the above model should be simultaneously estimated for the five equations of Equation 1, Equation 5, and three share equations of Equation 6 by maximum likelihood

(ML), such that all coefficients in Equation 1, Equation 5, as well as α_1 and α_2, can be identified. These estimates have the desirable statistical properties. The inclusion of the share equation of Equation 6 is vital. Excluding this equation, we cannot identify all parameter values in the model, including α_1 and α_2. The subsequent calculations for technical efficiency, scale economies, and scope economies of the two production stages, comparisons among different groups, and hypothesis testing cannot be accomplished. However, these calculated results help gain insights into managerial abilities and operational strategies for the references of managers and government authorities.

3.2. Econometric model

The main difficulty of simultaneously estimating Equation 1, Equation 5, and Equation 6 comes from the fact that both Equation 1 and Equation 5 have composed error terms that are known as having skewed normal distributions, leading the joint probability density function (PDF) of the composed errors to be unknown. Lai and Huang (2013) suggested using a Gaussian copula to derive the joint PDF of ε_1 and ε_2. In this study, we must derive the joint PDF of five random variables, i.e. $\varepsilon_1, \ldots, \varepsilon_5$. Note that ε_i, $i = 3, 4, 5$, are univariate random variables assumed to be jointly normally distributed in this study.

According to Sklar (1959), the joint cumulative distribution functions (CDF) $F(\cdot)$ of the five random variables of $\varepsilon_1, \ldots, \varepsilon_5$ can be related to a copula distribution function $C(\cdot)$, whose components consist of marginal CDF, $F_i(\varepsilon_i)$, $i = 1, \ldots, 5$. Taking the partial derivatives of $C(\cdot)$ with respect to $\varepsilon_1, \ldots, \varepsilon_5$, we obtain the joint PDF:

$$f(\varepsilon_1, \ldots, \varepsilon_5) = c[F_1(\varepsilon_1), \ldots, F_5(\varepsilon_5); \rho] \times \prod_{i=1}^{5} f_i(\varepsilon_i), \tag{7}$$

where $c(\cdot)$ represents the copula PDF, $f_i(\cdot)$ is the marginal PDF of ε_i, and ρ is a vector, including the dependence parameters among all marginal CDFs, $F_i(\varepsilon_i)$, $i = 1, \ldots, 5$.

The literature has many different copula functions, such as multivariate Student's t copula, Archimedean copula, Clayton n-copula, and Gumble n-copula. Each copula function has its dependency parameters, and economists most widely use the Gaussian copula. Following Lai and Huang (2013), Amsler, Prokhorov, and Schmidt (2016), T.-H. Huang, Lin, and Chen (2017), T.-H. Huang, Chen, and Lin (2018), and T.-H. Huang, Lin, and Wu (2019), we also select the Gaussian copula to derive the joint PDF of Equation 7 since it is tractable and is comprehensive. Further extensions to other copula functions should be able to follow the same procedure with a similar calculation. The derivation procedure is not displayed here. Readers are suggested to refer to the papers above.

The log-likelihood function of the simultaneous regression of Equation 1, Equation 5, and Equation 6 is shown below:

$$\ln L(\theta) = \sum\nolimits_{j=1}^{N} \ln f\left(\varepsilon_{1j}, \ldots, \varepsilon_{5j}\right)$$

$$= \frac{-N}{2}\ln|\Omega| - \frac{1}{2}\sum\nolimits_{j=1}^{N} \xi_j'\left(\Omega^{-1} - I_5\right)\xi_j + \sum\nolimits_{i=1}^{5}\sum\nolimits_{j=1}^{N} \ln f_i\left(\varepsilon_{ij}\right), \tag{8}$$

where θ represents all parameters to be estimated in the simultaneous regression model, j is the j^{th} sample point, $\xi_j = \left\{\Phi^{-1}\left[F_{1j}\left(\varepsilon_{1j}\right)\right], \ldots, \Phi^{-1}\left[F_{5j}\left(\varepsilon_{5j}\right)\right]\right\}'$ is a 5×1 vector, $\Phi^{-1}(\cdot)$ is the inverse of the CDF of the standard normal distribution, Ω is a 5×5 correlation coefficient matrix, whose main diagonal elements are all equal to one and the off-diagonal element signifies the correlation coefficients of the components in ξ_j, and I_5 is a 5×5 identity matrix. If the assumed copula function is correct, the ML estimates of Equation 8 are consistent, asymptotically efficient, and asymptotically normally distributed.

The first two terms on the right-hand side of Equation 8 describe the correlations between the five equations, and the last term corresponds to the log-likelihood function obtained from a separate regression. The separate regression may give consistent estimates and suitable standard errors under correctly specified marginal densities, but the standard errors are less efficient than those yielded by maximizing Equation 8, if the copula function is correct.[7]

Differing from the previous works, we allow the inefficiency terms u_1 and u_2 in Equation 1 and Equation 5 to be influenced by a set of environmental variables, namely

$$u_i = \gamma_i' q_i + \omega_i \geq 0, i = 1,\ \ 2. \tag{9}$$

If we assume $\omega_i \sim N(0, \sigma_{\omega_i}^2)$, then $-\gamma_i' q_i \leq \omega_i < \infty$ is a truncated normal random variable.[8] Notation q_i denotes an array of exogenous environmental variables, and γ_i is the corresponding parameter to be estimated. Plugging Equation 9 into Equation 1 and Equation 5, we obtain

$$\ln x_3 = \ln g(\alpha X) + v_1 - u_1 = \ln g(\alpha X) - \gamma_1' q_1 + \varepsilon_1, \tag{10}$$

$$\ln E = \ln PC(Y, W) + \ln \sum\nolimits_i S_i \tilde{\alpha}_i^{-1} + \gamma_2' q_2 + \varepsilon_2, \tag{11}$$

where $\varepsilon_1 = v_1 - \omega_1$ and $\varepsilon_2 = v_2 + \omega_2$. The PDF of ε_1 can be shown to be:

$$f_1(\varepsilon_1) = \frac{1}{\sigma_1 \Phi\left(\frac{\gamma_1' q_1}{\sigma_{\omega_1}}\right)} \Phi(B_1) \phi\left(\frac{\varepsilon_1}{\sigma_1}\right), \tag{12}$$

where $\sigma_1^2 = \sigma_{\omega_1}^2 + \sigma_{v_1}^2$, $B_1 = \frac{\gamma_1' q_1 \sigma_1^2 - \varepsilon_1 \sigma_{\omega_1}^2}{\sigma_{v_1} \sigma_{\omega_1} \sigma_1}$, and $\phi(\cdot)$ is the PDF of the standard normal variable. The PDF of ε_2, $f_2(\varepsilon_2)$, can be easily derived by simply replacing the subscript in Equation 12 with '2' and defining $B_2 = \frac{\gamma_2' q_2 \sigma_2^2 + \varepsilon_2 \sigma_{\omega_2}^2}{\sigma_{v_2} \sigma_{\omega_2} \sigma_2}$.

Equation 7 and Equation 8 have the CDF of ε_i, $F_i(\varepsilon_i)$, $i = 1, 2$, acquired by integrating the PDFs of $f_i(\varepsilon_i)$, $i = 1, 2$. However, Equation 12 shows that the skewed normal distribution of $f_1(\varepsilon_1)$ and $f_2(\varepsilon_2)$ enclose the CDFs of the standard normal distribution, which have no closed forms and hence are unable to be integrated, making the estimation of Equation 8 infeasible. Tsay et al. (2013) developed an approximate integral method to deduce $F_i(Q_i)$, $i = 1, 2$, which have closed forms without associating the inefficiency terms with a set of environmental variables.[9] Note that Q_i is a given value. This article allows the inefficiency term to be affected by several environmental variables so that $F_1(Q_1)$ and $F_2(Q_2)$ must be re-derived through a tedious process. Re-express $F_1(Q_1)$ as:

$$F_1(Q_1) = \int_{-\infty}^{Q_1} f_1(\varepsilon_1)d\varepsilon_1 = \frac{1}{\sigma_1 \Phi\left(\frac{\gamma_1' q_1}{\sigma_{\omega_1}}\right)} \int_{-\infty}^{Q_1} \Phi(B_1)\phi\left(\frac{\varepsilon_1}{\sigma_1}\right)d\varepsilon_1$$

$$= \frac{1}{\sigma_1 \Phi\left(\frac{\gamma_1' q_1}{\sigma_{\omega_1}}\right)} \int_{-\infty}^{Q_1}\left[\int_{-\infty}^{B_1}\phi(\varepsilon)d\varepsilon\right]\phi\left(\frac{\varepsilon_1}{\sigma_1}\right)d\varepsilon_1$$

$$= \frac{I(Q_1)}{\sigma_1 \Phi\left(\frac{\gamma_1' q_1}{\sigma_{\omega_1}}\right)}, \tag{13}$$

where $I(Q_1) = {}_{-\infty}^{Q_1}\left[{}_{-\infty}^{B_1}\phi(\varepsilon)d\varepsilon\right]\phi\left(\frac{\varepsilon_1}{\sigma_1}\right)d\varepsilon_1$ and the CDF of $\Phi(B_1)$ has no closed form. By extending Tsay et al. (2013), the approximate integral of $I(Q_1)$, $I^{app}(Q_1)$, can be derived. Plugging $I^{app}(Q_1)$ into Equation 13, we get the approximated $F_1(Q_1)$, $F_1^{app}(Q_1)$. By the same procedure, we can derive $I^{app}(Q_2)$ and $F_2^{app}(Q_2)$ sequentially. Since $I^{app}(Q_1)$ and $I^{app}(Q_2)$ are similar, Appendix 1 only shows the functional forms of $I^{app}(Q_1)$. Substituting $F_1^{app}(Q_1)$ and $F_2^{app}(Q_2)$ for $F_1(Q_1)$ and $F_2(Q_2)$ in Equation 8, we can then estimate the parameters by ML.

Those parameter estimates can be used to calculate such measures as returns to scale (RTS), scope economies (SC), and various efficiency scores. Based on Equation 10, the returns to scale measure, RTS1, of the production function is defined as:

$$\text{RTS1} = \sum_{i=1}^{2} \frac{\partial \ln x_3}{\partial \ln x_i}. \tag{14}$$

RTS1 > (<) 1 means that a proportionate increase in inputs increases output more (less) than proportionately. The production exhibits increasing (decreasing) returns to scale. Expanding the firm's production scale lowers (raises) its long-run average cost. This firm enjoys scale economies (diseconomies). Based on Equation 11, the returns to scale measure, RTS2, of the cost function is defined as:

$$\text{RTS2} = \sum_{j=1}^{2} \frac{\partial \ln E}{\partial \ln y_j}. \tag{15}$$

RTS2 < (>) 1 means that a proportionate increase in outputs increases expenditures less (more) than proportionately. The production exhibits increasing (decreasing) returns to scale. Expanding the firm's production scale lowers (raises) its long-run average cost. This firm is experiencing scale economies (diseconomies). Economies of scale result from higher production levels permitting specialization and labor division. Each worker is getting better at a specific task. Conversely, diseconomies of scale arise from coordination problems that often occur in huge firms.

We can use the coefficient estimates of the cost function to calculate the measure of scope economies (SC):

$$\text{SC} = \frac{PC(0, Y_2, W) + PC(Y_1, 0, W) - PC(Y_1, Y_2, W)}{PC(Y_1, Y_2, W)}, \tag{16}$$

where $PC(Y_1, Y_2, W)$ is the optimal cost function of Equation 2. The first two terms of the numerator in Equation 16 represent the sum of costs of specialized production on Y_2 and Y_1, respectively. If the numerator exceeds zero, then economies of scope prevail. Joint production of the two goods can help firms save production costs. Conversely, if the numerator is negative, then diseconomies of scope exist. Specialized production of individual goods can help firms save production costs.[10]

To calculate the technical efficiency of the production function, it is necessary first to obtain the conditional PDF of $\omega_1|\varepsilon_1$:

$$f(\omega_1|\varepsilon_1) = \frac{1}{\Phi\left(\frac{\gamma_1' q_1 + \mu_{s_1}}{\sigma_{s_1}}\right)\sqrt{2\pi}\sigma_{s_1}} e^{-\frac{1}{2}\left(\frac{\omega_1 - \mu_{s_1}}{\sigma_{s_1}}\right)^2}, \tag{17}$$

where $\mu_{s_1} = -\varepsilon_1\left(\frac{\sigma^2_{\omega_1}}{\sigma^2_1}\right)$ and $\sigma^2_{s_1} = \frac{\sigma^2_{v_1}\sigma^2_{\omega_1}}{\sigma^2_1}$. The conditional PDF of Equation 17 is, in effect, the normal distribution $N\left(\mu_{s_1}, \sigma^2_{s_1}\right)$ truncated at $-\gamma_1' q_1$ from below. The conditional expectation of $\omega_1|\varepsilon_1$ can be shown to be:

$$E(\omega_1|\varepsilon_1) = \mu_{s_1} + \sigma_{s_1}\frac{\phi\left(\frac{\gamma_1' q_1 + \mu_{s_1}}{\sigma_{s_1}}\right)}{\Phi\left(\frac{\gamma_1' q_1 + \mu_{s_1}}{\sigma_{s_1}}\right)}. \tag{18}$$

The estimate of Equation 18, $\hat{E}(\omega_1|\varepsilon_1)$, is regarded as the estimated value of ω_1, $\hat{\omega}_1$. Here, symbols with '^' represent estimates. After summing $\hat{\omega}_1$ over $\hat{\gamma}_1' q_1$, we get the estimated value of technical inefficiency, $\hat{u}_1$. Finally, the technical efficiency score (TE) is computed as:

$$\mathrm{TE} = \exp(-\hat{u}_1) = \exp\left(-\hat{\gamma}_1' q_1 - \hat{\omega}_1\right) = \exp\left(-\hat{\gamma}_1' q_1\right)\exp(-\hat{\omega}_1). \tag{19}$$

TE ranges between zero and unity. A large value of TE score close to unity indicates that the actual output level is nearer the production frontier, and vice versa. As for the cost function, the cost efficiency (CE) score can be obtained by a similar approach with minor modifications, i.e.

$$\mathrm{CE} = \exp(-\hat{u}_2) = \exp\left(-\hat{\gamma}_2' q_2 - \hat{\omega}_2\right) = \exp\left(-\hat{\gamma}_2' q_2\right)\exp(-\hat{\omega}_2), \tag{20}$$

where $\hat{\omega}_2 = \hat{E}(\omega_2|\varepsilon_2)$ is similar to Equation 18 with partial modifications. CE also lies between zero and unity. A large value of CE score close to unity indicates that the actual production cost is nearer the minimal cost frontier, and vice versa.

3.3. Research hypotheses

Audit firms should undertake the following two strategies to achieve full technical efficiency and high audit quality. They first provide training and development projects to enhance auditors' skills and acquire the latest auditing standards and technologies. They second utilize advanced audit software and newly developed data analytics tools to improve the effectiveness and efficiency of audit procedures. Because audit reports are public goods, implementing robust audit quality control and maintaining ethical standards is essential.

Given that we have split the production process of audit firms into two stages, it is interesting to discuss whether the technical efficiencies of different sizes and sectors are the same. Large audit firms may possess economies of scale and scope and brand effect, which aids in attracting excellent auditors. Large audit firms are expected to be superior to small and medium-sized ones in technical and cost efficiency. In addition, large audit firms have more CPA licenses than small and medium-sized ones, and such professional licenses reflect the professional abilities of accountants. Some empirical studies found that the number of CPA licenses is an important driver impacting audit firms' productivity (H.-H. Chang et al. 2015).

When audit firms accept clients' audit work contracts, the audit engagement is verified and planned by supervisors and team leaders based on audit risk assessment. Auditors assess clients' inherent and internal control risks, schedule audits, and evaluate audit risks. Large audit firms have accumulated richer and more diversified audit experience than smaller ones, implying that large firms' audit reports may have higher audit quality for different industries.

This so-called learning-by-doing effect prompts large firms' resource scheduling and coordinating capability among sectors. They tend to possess more excellent auditing skills when executing client engagements and are expected to be more productive in the first production stage than smaller firms.

We propose the first hypothesis (H1) as follows:

H1: Large audit firms have higher technical efficiency than small and medium-sized firms in the first production stage.

Organizational structure influences a firm's daily operations, performance, and span of control (Gerloff 1985). The design of the organizational structure must respond to the external competitive environment. Therefore, the environments inevitably influence the design of the organizational structure (Duncan 1972). According to the audit firm service report of Taiwan's Financial Supervisory Commission in 2019, the total number of certified public accountants in Taiwan was 1050, 1111, 1134, and 1140 between 2016 and 2019. The numbers of one-person audit firms (the ratio of one-person audit firms to total firms) in the four years are 806 (76%), 835 (75%), 883 (77%), and 895 (79%), respectively.[11] The associated firms are 244, 276, 251, and 245, respectively. Research has observed that most of Taiwan's audit firms are small-sized ones. The same report also displays that the average profit rates for one-person and associated audit firms declined yearly, showing that the degree of competition in the audit market is likely getting fierce over time.

Here, the term 'span of control' describes the number of people a supervisor can lead, command, and supervise. It should be considered when designing a firm's organizational structure. Due to limited personal abilities and other environmental factors, a supervisor must have a limited span of control, such that each employee is fully managed and supervised without subordinates' dissatisfaction, which is an essential topic in studying organizational management. Small and medium-sized firms are disadvantageous due to their need for scale economies in the increasingly competitive audit market. The largest four (henceforth, the Big Four) audit firms possess higher brand recognition by capital market investors and government agencies. Conversely, small and medium-sized audit firms are frequently concerned about their less audit experience and ability to offer reliable reports.

Large audit firms employ a significant workforce and operate on a sizeable scale relative to small and medium-sized ones. The latter's organizational structures are usually simple, with accountants as the top leaders supervising a few staff. This flattening organization may operate more efficiently than a large, multi-level one. After the engagement, partners and managers will work with team leaders and staff in the second production stage to conduct the audit. As small and medium-sized audit firms make up more than 75% of all firms in Taiwan, they face more intense competition than large ones. Lin and Huang (2023) have shown that small firms outperform large firms because the former have higher managerial abilities and adopt superior technology to the latter. Small and medium-sized audit firms likely have higher technical efficiency than large ones in the second production stage to be viable and sustainable.

We propose the second hypothesis (H2) as follows:

H2: Small and medium-sized firms have higher technical efficiency than large ones in the second production stage.

Audit firms are known as knowledge-intensive. Auditors provide audit and non-audit services to mitigate the audited firms' information asymmetry in the capital and financial markets. Their services include corporate financial and tax attestation, financial statement review and assurance, and consulting services. In the production process of audit firms, dedicating specialized and intensive work efforts is an essential requirement for accountants in delivering services to their clients. Audit firms rely heavily on educated talents and direct their professional and complex knowledge into the production process to offer professional advice to clients (Greenwood et al. 2005; Morris and Empson 1998). Scott (1998) referred to audit firms as the intellectual industry.

As clients face the challenges of uncertain business environments, professional and knowledge-intensive institutes, such as audit firms, are better to have two characteristics (Starbuck 1992). First, they must integrate complex knowledge into intangible outputs – professional services. Second, they should provide customized and competent advice to solve client problems. In other words, clients expect accountants to present 'exceptional and valuable professional advice' that can solve their business problems, such as shareholders' fundraising, corporate bond issuance, tax-efficient planning, and capital expenditure efficiency. Thus, accountants must have professional knowledge of various disciplines and the characteristics of multidisciplinary practice (MDP) (Greenwood et al. 2005).

According to previous empirical studies, when knowledge-intensive professional organizations gradually build up their abilities to use multidisciplinary knowledge, they will benefit from the following advantages as reputation stickiness (Berardino 2000; Trebilcock and Csorgo 2001), scope economies (Melancon 1998), cross-selling (Nayyar 1990), and diversified services (Greenwood et al. 2005; Nayyar 1990), generated from customer trust. Accountants are recommended to supply professional services from MDP, diversifying their products (services) to expand the market power, retain clients, and improve the operation efficiency of the firms. Specifically, the presence of scope economies lowers a firm's production cost through resource sharing among different products. The subsequent input savings may enhance the operation efficiency of the firm.

For the production and cost efficiencies in the first and second stages, we propose the third hypothesis (H3A and H3B) as follows.

H3A: Audit firms with highly diversified products have higher technical efficiency in the first production stage.

H3B: Audit firms with highly diversified products have higher cost efficiency in the second production stage.

4. Data analysis

The data used are mainly from the Taiwan Economic Journal (TEJ) database spanning 2010–2014 because, after 2015, the databank does not provide the variable of the total floor area of audit firms. We shall define this variable as one of the inputs. We deleted some observations with incomplete data and obtained 3410 pooling data.[12] In the next section, we classify sample firms into two forms, i.e. small firms that have the number of employees below 30 and large firms that have the number of employees above 31. Since the former contains too few observations, we do not split our sample firms into the Big Four and Non-Big Four ones.

The scale distribution of audit firms in Taiwan differs from that in European and American countries, where the Big Four audit firms account for more than 70% market share. In addition to the Big Four audit firms, the audit market in Taiwan is full of many small and medium-sized audit firms. The group of large firms embraces 310 observations, and the small firms have 3100 observations.

4.1. Variable definitions

Table 1 shows the definitions of all variables in this study. There are four inputs, i.e. the number of partners and managers (x_1), the total floor area (the proxy for capital stock) (x_2), the number of cases (x_3), and the number of employees (x_4). The number of cases plays a dual role, as described in Subsection 3.1. It is the intermediate output in the first production stage and one of the inputs in the second production stage. The corresponding factor prices are denoted by p_1, p_2, p_3, and p_4, respectively, and the sum of the expenditures on the four factors is equal to the total cost (cost). There are two output variables: audit fees (y_1) and non-audit fees (y_2).

We use a production frontier to represent the technology in the first sector, in which the firm utilizes $\alpha_1 x_1$ and $\alpha_2 x_2$ to produce the single output, i.e. the number of cases. In the second sector, the firm is assumed to hire the remaining proportions of x_1 and x_2, i.e. $(1-\alpha_1)x_1$ and $(1-\alpha_2)x_2$, x_3, and x_4 to supply two outputs. Here, we use the cost frontier to characterize the production technology under multiple outputs and inputs. Moreover, the corresponding input share equations can be formulated and give the necessary information to estimate α_1 and α_2.

Two sets of environmental variables are considered. One is for the production frontier, and the other is for the cost frontier. For the production frontier, four environmental variables are being considered, i.e. the proportion of female partners (female), the

Table 1. Variable definitions.

Variable	Definition (units)
Inputs	
Partners and managers (x_1)	Number of partners plus managers (persons)
Fixed assets (x_2)	Floor area (pin = 3.3 square meters)
Cases (x_3)	The number of cases undertaken per year
Employees (x_4)	The number of team leaders and staff (persons)
Input prices	
Price of x_1 (p_1)	Partner and manager income/x_1 (NT dollars)
Price of x_2 (p_2)	Expenditure on rent/x_2 (NT dollars)
Price of x_3 (p_3)	Expenditure on fulfilling entrusted cases/x_3 (NT dollars)
Wages of employees (p_4)	Salaries of all staff/x_4 (NT dollars)
Outputs	
y_1	Audit fees (NT dollars)
y_2	Non-audit fees (NT dollars)
Total costs (cost)	$p_1 \times x_1 + p_2 \times x_2 + p_3 \times x_3 + p_4 \times x_4$ (NT dollars)
Environmental variables	
First stage	
female	The ratio of the number of female partners to all partners
MA	The ratio of the number of partners and managers with MA or Ph. D. degree to all partners and managers
Share	Market share of individual firm's audit fees
growth	GDP growth rate per annum (%)
Second stage	
Fe_worker	The ratio of female workers to total employees
PhD	The ratio of employees with master's or Ph. D. degree to total employees
HHI	Herfindahl-Hirschman index calculated based on audit fees
cpi_rate	Inflation rate calculated using CPI
divy	Product diversification index, according to Laeven and Levine (2007), is defined as $\text{divy} = 1 - \frac{\lvert y_1 - y_2 \rvert}{y_1 + y_2}$. It takes values between zero and one, with higher values indicating greater diversification.

Note: Notation 'NT dollars' means New Taiwan dollars.

proportion of partners and managers with master's or doctoral degrees (MA), the market share of the firm (Share), and the GDP growth rate per annum (growth). While females are cautious in their work, they may not be able to meet the needs of corporate business development due to family or physiological factors. Thus, the effect of this variable on the production efficiency of the number of cases is indeterminant. Partners and managers with master's or above degrees may indicate they have gained more professional knowledge to undertake more cases. This positive effect on production efficiency may be partially offset by their charging high service fees, leaving the impact of MA on efficiency indeterminant. A positive coefficient of Share means that firms with high market share are less efficient, which supports the 'quiet life hypothesis,' while a negative coefficient of Share rejects the hypothesis. The higher the macroeconomic growth rate, the more cases enterprises entrust, improving audit firms' production efficiency. [13] On the other hand, large increases in client demands may induce audit firms to over-expansion and neglect to strengthen their managerial abilities, which hurts technical efficiency. In summary, it is ambiguous for researchers to judge whether the above four environmental variables positively or negatively affect efficiency.

The cost frontier of the second stage has five environmental variables, i.e. the proportion of female workers (Fe_worker), the proportion of employees with master's or doctoral degrees (PhD), the Herfindahl-Hirschman index (HHI), the inflation rate (cpi_rate), and

Table 2. Descriptive statistics.

Variable	Mean	Standard deviation
cost	2.0978D + 07	1.9671D + 08
x_1	6.9980	51.1486
x_2	493.4677	6,204.7691
x_3	365.5628	1,746.0680
x_4	19.4282	123.5139
p_1	622,066.4082	644,249.97
p_2	4,040.1410	3,975.2946
p_3	26,051.2525	204,885.8950
p_4	357,479.2575	156,865.4735
y_1	1.8404D + 07	1.6491D + 08
y_2	1.1042D + 07	9.7680D + 07
female	0.1351	0.1342
MA	0.0937	0.1217
Share	0.00139	0.01326
growth	4.5541	2.8572
Fe_worker	0.7479	0.1628
PhD	0.1104	0.1389
HHI	1,213.5982	17.0450
cpi_rate	1.2751	0.3969
divy	0.5243	0.2985
Sample size	3,410	

Note: 'D' signifies scientific notation, such that $2.0978D + 07 = 2.0978 \times 10^7$.

product diversity index (divy). The effects of the first four variables on efficiency are analogous to the previous paragraph and hence ignored. The definition of the product diversification index, i.e. divy, is reported in Table 1. Firms with a high divy index can spread risks and benefit from resource-sharing advantages, prompting cost efficiency.

4.2. Sample statistics

Table 2 displays the descriptive statistics of all variables. Most have high standard deviations relative to their sample mean values, indicating significant heterogeneity among sample firms. Recall that we separate the audit firms into small firms, hiring the number of employees less than or equal to 30, and large firms, hiring the number of employees above 30. Table 3 reports the descriptive statistics of all variables for the two types of firms, which diverge substantially in most of these variables in terms of sample means and standard deviations, and the differences of some variables exceed 20 times. One may infer that the two groups may have their characteristics and advantages in the two production stages, which is worth further inspection by testing the hypotheses raised in Methodology.

5. Empirical study

5.1. Coefficient estimates

We exploit the ML approach to simultaneously estimate the log-likelihood function of Equation 8, which contains a total of five equations, i.e. Equation 10, Equation 11, and Equation 6.[14] Table 4 exhibits the coefficient estimates of the two production stages. We see that all coefficient estimates attain at least a 5% significance level, arising mainly from the use of simultaneous estimation of the five equations, which reduces the standard

Table 3. Descriptive statistics of the large and small firms.

	Large firms		Small firms	
Variable	Mean	Std dev	Mean	Std dev
cost	3.05506D + 08	9.97257D + 08	6,505,397.9448	5,460,052.3962
x_1	54.4226	162.3400	2.2555	1.6197
x_2	4,087.9968	20,256.3837	134.0148	125.0770
x_3	2,337.51290	5,387.6312	168.3677	177.8122
x_4	150.0774	386.3106	6.3632	4.8258
p_1	909,275.8574	1,060,596.1559	593,345.4633	579,086.1185
p_2	4,822.8051	3,466.3389	3,961.8746	4,014.7614
p_3	71,241.9421	580,461.9702	21,532.1836	111,161.7519
p_4	438,783.03414	127,498.9789	349,348.8798	157,225.9598
y_1	1.61382D + 08	5.26587D + 08	4,106,216.8110	4,235,463.8231
y_2	9.69971D + 07	3.11485D + 08	2,446,530.7923	2,945,760.1201
female	0.1314	0.0590	0.1354	0.1395
MA	0.0654	0.0427	0.0966	0.1266
Share	0.0128	0.0424	0.00025	0.00023
growth	4.6901	2.9486	4.5405	2.8481
Fe_worker	0.7558	0.0901	0.7471	0.1683
PhD	0.0999	0.0906	0.1115	0.1428
HHI	1,212.7377	17.5245	1,213.6843	16.9969
cpi_rate	1.2660	0.3952	1.2760	0.3972
divy	0.5679	0.2818	0.5199	0.2998
Sample size	310		3,100	

Note: 'D' signifies scientific notation, such that 2.0978D + 07 = 2.0978×10^7.

errors of the estimates. The estimates of fractional parameters of α_1 and α_2 are equal to 0.0795 and 0.1360E-05, respectively, indicating that the sample firms consume 7.95% of partners and managers (x_1) and 0.000136% of fixed assets (total floor area, x_2) in the first production stage to undertake cases. These estimates tell that x_1 is the primary input and that x_2 plays quite a minor role, conforming to the current audit practices in Taiwan.[15] Moreover, the second stage is the primary production procedure since both fractional parameters are estimated to be less than 8% in the first stage.

Although the firm office is a crucial input in the audit production process, audit firms attract customers mainly by providing professional and high-quality audits, which makes the quality of financial statements trustable and credible to different users (Ball 2009; Barton 2005; Francis and Wilson 1988; Frendy and Hu 2014). Many enterprises in Taiwan appoint accountants based on the accountants' professional qualifications and the firms' reputation. The first production stage mainly involves partners and managers undertaking cases requiring little capital, i.e. x_2. Most proportions of the two factors are used in the second production stage. In practice, after partners and managers bring the cases entrusted to clients into the firm, the internal audit begins, including back-office audit assignments and field audit planning. The group leaders and their members are in charge of the front-line audit work proceeded at their clients' locations, and supervisors review the audit working paper to ensure the audit quality. From this perspective, the second stage is the primary production process and source of profit for firms.

To show that the simultaneous estimation of five equations improves the parameter estimates' efficiency and test the validity of the chosen Gaussian copula, we alternatively estimate the same five equations under the assumption that they are independent. Note that the copula methods are of no use in this case. We do not show the coefficient estimates to save space. This time, 16 out of the 45 coefficient estimates are insignificantly estimated. Only one coefficient estimate of the nine environmental variables attains

Table 4. Coefficient estimates of the two stages.

First stage: Production frontier			Second stage: Cost frontier		
Variable	Coefficient estimate	Standard error	Variable	Coefficient estimate	Standard error
α_1	.0795***	.6324E-03	Intercept	10.0647***	.3559E-02
α_2	.1360E-05***	.5445E-17	$\ln(p_2/p_1)$	.2840***	.3464E-03
Intercept	6.0319***	.1769E-02	$\ln(p_3/p_1)$	.4756***	.4702E-03
$\ln x_1$	.6027***	.4232E-02	$\ln(p_4/p_1)$	.0392***	.6566E-03
$\ln x_2$	.3443***	.3067E-03	$0.5[\ln(p_2/p_1)]^2$	.0394***	.1238E-03
$0.5(\ln x_1)^2$	.2637***	.3515E-02	$0.5[\ln(p_3/p_1)]^2$	.1347***	.1595E-03
$0.5(\ln x_2)^2$	.0429***	.6963E-04	$0.5[\ln(p_4/p_1)]^2$	.1302***	.8719E-03
$\ln x_1 \ln x_2$	−.0454***	.9235E-04	$\ln(p_2/p_1)\ln(p_3/p_1)$	−.0177***	.1056E-03
			$\ln(p_2/p_1)\ln(p_4/p_1)$	−.0164***	.3782E-03
Ω_{12}	.6375***	.2101E-02	$\ln(p_3/p_1)\ln(p_4/p_1)$	−.0746***	.4222E-03
Ω_{13}	−.2494***	.4883E-02	$\ln y_1$	−.9410***	.1607E-03
Ω_{14}	.2321***	.6543E-02	$\ln y_2$	−.4085***	.9823E-04
Ω_{23}	−.0935***	.5840E-02	$0.5(\ln y_1)^2$	.0997***	.3262E-04
Ω_{24}	−.0860***	.6636E-02	$0.5(\ln y_2)^2$	.0466***	.2203E-04
Ω_{34}	−.2501***	.0110	$\ln y_1 \ln y_2$	−.7513E-02***	.7185E-05
Ω_{15}	.0857***	.6531E-02	$\ln(p_2/p_1)\ln y_1$	−.6318E-03***	.2858E-04
Ω_{25}	.2739***	.4819E-02	$\ln(p_2/p_1)\ln y_2$	−.1870E-02***	.3258E-04
Ω_{35}	−.0908***	.0158	$\ln(p_3/p_1)\ln y_1$	.0152***	.2669E-04
Ω_{45}	−.6723***	.9125E-02	$\ln(p_3/p_1)\ln y_2$	−.2514E-02***	.3441E-04
			$\ln(p_4/p_1)\ln y_1$	.1859E-03**	.8683E-04
σ_3	.0512***	.1637E-03	$\ln(p_4/p_1)\ln y_2$	−.3941E-03***	.1039E-03
σ_4	.1109***	.1133E-02			
σ_5	.1188***	.1724E-02			
σ_{ω_1}	.3618***	.2647E-02	σ_{ω_2}	.5757***	.9360E-03
σ_{v_1}	1.2269***	.6259E-02	σ_{v_2}	.1551***	.2884E-03
Environments					
female	.2687***	.6231E-02	Fe_worker	−.0494***	.2961E-03
MA	.0968***	.2822E-02	PhD	−.0784***	.3220E-02
Share	23.3635***	.3571	HHI	−.2444E-03***	.1213E-05
growth	.4147E-03***	.2984E-04	cpi_rate	.0482***	.2417E-03
			divy	−.2636***	.1560E-02
Log-likelihood	18,296.3				
Sample size	3,410				

Note: 1. ***, **, and * represent significance at the 1%, 5%, and 10%, respectively.
2. σ_3, σ_4, and σ_5 are estimated standard deviations of the error terms in three share equations.
3. Ω_{ij}, $i, j = 1, \ldots, 5$, $i < j$, is the estimated off-diagonal element of the correlation coefficient matrix of Ω in Equation 8.

statistical significance. The estimates of proportional parameters α_1 and α_2 are equal to 0.00117 and 0.00141, respectively, and both are insignificant. These findings highlight the advantages of a simultaneous estimation procedure.

According to Lai and Huang (2013), we apply Hausman's test to test the null hypothesis that the simultaneously estimated coefficients are equal to the coefficients under the independent assumption among the five equations. The chi-squared statistic equals 28,601.60, and the degree of freedom is 45, which rejects this null hypothesis and supports the appropriateness of the Gaussian copula.[16]

The coefficient estimates of all four environmental variables in the first production process are positive, implying that the increases in these variables adversely impact the sample firms' technical efficiency. This finding is attributed to the fact that, by referring to Subsection 4.1, the adverse effects of these environments on efficiency outweigh the positive effects. In other words, sample firms should lower the values of these variables to improve technical efficiency. In the second production stage, an increasing proportion of female workers (Fe_worker), improving employee quality (PhD), and strengthening product diversification (divy) stimulate the cost efficiency of the sample firms. These findings are consistent with previous empirical studies, e.g. Greenwood et al. (2005). The negative coefficient estimate of HHI leads us to reject the 'quiet life hypothesis.' A rising inflation rate corresponds to economic recovery or expansion, and firm executives should avoid overexpansion and strengthen managerial abilities to upgrade cost efficiency.

We re-estimate the model excluding environmental variables to emphasize the significance of environmental variables. Appendix 2 shows the estimation results. Seven estimates are insignificant, and the log-likelihood function value equals 3016.81, much lower than 18,296.3 in Table 4, as expected. The estimates of proportional parameters α_1 and α_2 equal 0.3861 and 0.00041, respectively, but only α_1 is significant. Compared with Table 4, the estimates of these two fractional parameters are much higher, indicating that the models omitting environmental variables tend to overestimate proportional parameters.

5.2. Efficiency scores

The coefficient estimates in the previous subsection are used to calculate the average efficiency score and measures of scale and scope economies. Table 5 presents these figures according to the scale of the sample firms, production stages, and whether including environmental variables. The environmental variables affect those measures, especially for economies of scale and scope measures. As for the efficiency scores, the effects are relatively mild.

For the first sector, the average technical efficiency scores of large and small firms are 0.71 and 0.75, respectively, meaning that the intermediate output (number of cases) can be increased by 29% and 25% to achieve their production frontiers, given their current input mixes. For the second sector, the average cost efficiency scores of the two groups of firms are equal to 0.48 and 0.63, respectively. Their total cost can be cut by 52% and 37%, respectively, to attain their cost frontiers, given their existing output levels.

We discover that the key source of inefficiency stems from the second sector of the firms, which is also the primary production sector that consumes most of the sample firms' resources. We recommend that the sample firms promote their cost efficiency in the second stage by saving input usage and improve managerial abilities in the first stage by undertaking more cases.

Next, we test whether the average efficiencies of both types of firms in the two production stages differ significantly. The null hypotheses that their production and cost efficiencies in both stages are indifferent are decisively rejected. One can conclude that small firms are more efficient than large ones in both stages.[17] Therefore, the data reject H1 and accept H2. Smaller audit firms outperform larger ones.

Copley and Doucet (1993) and Numan and Willekens (2012) claimed that audit market competition influences audit firms' behaviors and operation strategies. Greenwood et al. (2005) found that market competition prompts audit firms' efficiency. Starting in 1988, the government of Taiwan significantly increased the issuance of accountant licenses, making many new accountants enter the market. This condition alters the market structure, raises the degree of market competition, and squeezes the profits of the member firms.

Table 5. Average measures of scale and scope economies.

	Large firms		Small firms	
	With environments	Without environments	With environments	Without environments
TE in the first stage	0.7076	0.7776	0.7542	0.6290
	(0.1717)	(0.0713)	(0.0120)	(0.1286)
CE in the second stage	0.4769	0.4532	0.6334	0.6838
	(0.1507)	(0.0738)	(0.1620)	(0.0977)
RTS1	1.5389	0.8015	1.0707	0.2822
	(0.2507)	(0.2781)	(0.1292)	(0.1434)
RTS2	0.8835	0.6368	0.5178	0.3354
	(0.1735)	(0.1427)	(0.1436)	(0.1215)
SC	78.0906	436.8302	196.6276	496.0289
	(40.3657)	(158.6826)	(83.5369)	(168.8502)
Sample size	310		3,100	

	All firms	
	With environments	Without environments
TE in the first stage	0.7500	0.6425
	(0.0546)	(0.1316)
CE in the second stage	0.6192	0.6629
	(0.1671)	(0.1165)
RTS1	1.1132	0.3294
	(0.1974)	(0.2191)
RTS2	0.5510	0.3628
	(0.1804)	(0.1509)
SC	185.85146	290.39723
	(87.4820)	(95.5846)
Sample size	3,410	

Note: Numbers in the parentheses are standard deviations.

It is usually true that competition enhances the operating efficiency of organizations due to the pressures of economic losses and survival. Small firms own meager resources relative to large firms, and thus, they are less likely to be viable if inefficiency endures for some time. These firms must strive to promote their managerial abilities by expanding the output – the number of cases – in the first production stage and make efforts to save production costs by diminishing the input used in the second production stage. Small firms must adapt to the fierce competition in the audit market and respond to the changing market conditions quickly. Our empirical results seem to be congruent with the real audit market in Taiwan.

Taking partial derivatives of Equation 19 and Equation 20 with respective to the environmental variables, we can quantify the marginal effects of those variables on efficiency scores. Table 6 shows a one-percentage increase in the environmental variables on efficiency scores. The signs are consistent with their coefficient estimates, i.e. a positive coefficient estimate corresponds to a negative marginal effect, and vice versa. Among them, the variable of female exerts the largest negative effect on technical efficiency in the first stage, followed by the variable of Share. The variable of cpi_rate has the largest adverse impact on cost efficiency, while the variable of HHI has the largest favorable impact on cost efficiency.

5.3. Scale and scope economies

Using Equation 14 and Equation 15, we compute the scale economies measures for the two sectors, RTS1 and RTS2, respectively. The bottom half of Table 5 reveals that the average

Table 6. Average marginal effects.

First stage: Production frontier			Second stage: Cost frontier		
Environmental variable	Average marginal effect	Standard error	Environmental variable	Average marginal effect	Standard error
female	−0.3379	0.0670	Fe_worker	0.0018	0.0013
MA	−0.1217	0.0241	PhD	0.0028	0.0021
Share	−0.2937	0.0583	HHI	0.0886	0.0660
growth	−0.0521	0.0103	cpi_rate	−0.1747	0.1302
			divy	0.0096	0.0071

RTS1 and RTS2 (standard deviations) are, respectively, 1.1132 (0.1974) and 0.5510 (0.1804),[18] indicating that both sectors are adopting increasing returns to scale technology. The sample firms can reduce their long-run average costs by expanding the operation scale in both sectors. Mergers and acquisitions may be a feasible and faster way to realize such benefits. Please see Francis (1984), Francis and Stokes (1986), Banker, Chang, and Kao (2002), Banker, Chang, and Natarajan (2005), and Fung, Gul, and Krishnan (2012).

If the environmental variables are precluded from the regression model, then the average RTS1 and RTS2 are 0.33 and 0.36 (shown in the bottom half of Table 5), respectively. Now, the mean RTS1 is less than unity, implying that the operation scale of the sample firms in the first stage is too large and should be reduced to lower their long-run average costs. The omission of the environmental variables tends to impact the scale economies measure of the first stage substantially, while the impact on that of the second stage is relatively small. One can draw similar conclusions for RTS1 and RTS2 for the groups of small and large firms, according to the upper half of Table 5.

The scope economies of the second sector are calculated using Equation 16. Regardless of whether environmental variables are considered, the average SC values of large and small firms are all positive but substantially different. Economies of scope prevail in Taiwan's audit industry, recommending that the sample firms diversify their products. The resulting advantages of resource sharing allow the firms to save input usage and cut costs. The models without including environmental variables are inclined to overestimate the SC values. These outcomes tend to favor the inclusion of environmental variables to explain the inefficiencies in the two production stages.

We partition the entire sample into two groups, i.e. low and high diversification firms, where the former has the product diversification index, divy, below the average value of 0.52, and the latter more than 0.52. Recall that this index is constructed based on audit fees (y_1) and non-audit fees (y_2). The test statistic of H3A is insignificant, meaning that the degree of product diversification has no significant effects on the technical efficiency of the first sector. This conclusion is acceptable because y_1 and y_2 are the final outputs of the second stage, which are not directly related to the production efficiency of the first stage. Conversely, the test statistics of H3B reach the 10% significance level, verifying that product diversification helps promote the sample firms' cost efficiency in the second production stage.[19]

It is insightful to conduct the same tests under the condition of eliminating environmental variables from the regression models. Both test statistics of H3A and H3B are negative and significant at the 1% level, indicating that highly diversified firms in both stages are less technically efficient, supporting product specialization instead. This outcome lacks economic or auditing justifications and is inconsistent with intuition.

In summary, models that ignore environmental variables are apt to distort estimates of efficiencies and measures of scale and scope economies and mislead the potential benefits of product diversification. Therefore, regression models should contain environmental variables to account for efficiencies, which is particularly useful in regulatory analysis to measure the effects of, e.g. mergers and acquisitions.

We finally estimate the traditional one-stage cost frontier with the same environmental variables as in Equation 20. We do not show the parameter estimates to save space and use them to calculate the average value of CE. The average cost efficiency scores of large and small firms are equal to 0.8139 and 0.8159, respectively. The difference between them is insignificant, which is inconsistent with the two-stage model, where small firms outperform large ones in both stages. The results highlight the usefulness of decomposing the production process into different stages.

6. Conclusions

This study applies a network stochastic frontier model to illustrate the production process for Taiwan's audit firms, which consists of two production sectors and is closer to the firms' actual production manner. The first sector is assumed to employ a fraction of partners and managers and a fraction of fixed assets to undertake cases, defined as the sector's single and intermediate output. We use a production frontier to depict the production technology. The second sector is presumed to employ the remaining proportions of partners and managers, as well as fixed assets and the number of cases and staff to fulfill the entrusted cases by clients to earn approval and non-audit fees. Here, we specify a cost frontier to signify the production technology.

Since the error terms of both sectors are composed errors with skewed normal distributions, we select the copula methods to derive their joint PDFs and the corresponding log-likelihood function. To estimate the fractional parameters, we must construct an economic model that contains five equations and propose an estimation procedure. We suggest estimating the simultaneous estimations by ML to obtain all parameters of interest. The estimation results can be used further to perform Hausman's test. Notably, the simultaneous equations model is quite complicated, making the convergence of the log-likelihood function hardly achievable and time-consuming.

The estimates of the two proportional parameters are equal to 0.0795 and 0.1360E-05, respectively, and both are significantly different from zero. These estimates are reasonable and consistent with the reality. However, if the inefficiency terms fail to link with environmental variables, those two estimates are much higher, and only the first proportional parameter is significantly different from zero.

The parameter estimates are next used to calculate the efficiency scores for both sectors. The second sector is less technically efficient than the first sector. Thus, the sample firms should improve their cost efficiency and raise their managerial abilities. Small firms outperform large firms in both sectors, which reject H1 and accept H2.

We characterize the two production processes as increasing returns to scale, arising from an expansion of scale permitting a greater division of labor and specialization of function. The long-run average costs decrease with the growing production scale as a result. The presence of scope economies in the second sector advocates that product diversification helps reduce the production costs of the sample firms. Moreover, based on the product diversification index divy, we reject H3A and accept H3B. The degree of product diversification positively correlated with the cost efficiency of the second sector. Finally, models ignoring environmental variables are likely to yield misleading results.

Since our simultaneous equations model is already complex, we do not consider the issue of endogenous inputs of the production frontier. The problem is worth investigating in future research. Readers are advised to refer to Amsler, Prokhorov, and Schmidt (2016, 2017). With the advent of artificial intelligence and the big data era, new technological tools will replace data analysis of audit firms' work procedures and working papers. Whether this evolution affects the production efficiency and scale and scope economies of the sample firms are worth further study in the future.

Notes

1. Partners and managers are responsible for controlling audit quality of client engagements and the audit opinions on financial statements. The final audit opinion on the financial statements is approved, signed, and sealed by the partner, while the manager provides audit recommendations for the partner's reference.
2. A team leader directs staff to conduct the actual fieldwork of the assigned audit cases, review audit working papers, and perform initial quality control procedures.
3. Similar problems occur in the banking industry. For details, please see, e.g. Holod and Lewis (2011) and T.-H. Huang, Lin, and Chen (2017).
4. This scalar measure deals directly with the input excesses and the output shortfalls of the DMU concerned.
5. Battese, Rao, and O'Donnell (2004) construct the metafrontier production function based on mathematical programming techniques, which excludes random disturbance term. Conversely, C. J. Huang, Huang, and Liu (2014) developed the stochastic metafrontier model, including random terms.
6. According to the microeconomic theory, one can prove that both production and cost functions represent the same production technology of a firm by duality.
7. Lai and Huang (2013) advise using the Hausman (1987) test by comparing the difference in the estimators of the separate and joint estimation of Equation 8. If the copula density is mis-specified, then the two sets of estimators will have different probability bounds. The test statistic can be shown to have a chi-squared distribution with K degrees of freedom, i.e.
$$H = \left(\hat{\theta}_S - \hat{\theta}_J\right)' \left[\widehat{Cov}\left(\hat{\theta}_S\right) - \widehat{Cov}\left(\hat{\theta}_J\right)\right]^{-1} \left(\hat{\theta}_S - \hat{\theta}_J\right) \sim \chi^2_K,$$
where subscripts S and J denote separate and joint estimation, $\hat{\theta}$ is the estimated vector of parameters, and K is the number of common parameters in the separate and joint estimation. The rejection of the H statistic supports the use of the Gaussian copula, and hence of Equation 8, to implement the empirical study.
8. The PDF of ω_i can be demonstrated to be $f(\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_{\omega_i}} e^{\frac{-1}{2}\left(\frac{\omega_i}{\sigma_{\omega_i}}\right)^2} / \left[1 - \Phi\left(-\frac{\gamma_i' q_i}{\sigma_{\omega_i}}\right)\right]$. Please note that Equation 9 deviates from Amsler, Prokhorov, and Schmidt (2017); thus, the PDFs of ε_1 and ε_2, as shown in Equation 12, differ from theirs.
9. The simulated maximum likelihood method suggested by Greene (2003) and Amsler, Prokhorov, and Schmidt (2014) is another way to solve the integration problem, which was later used by Amsler, Prokhorov, and Schmidt (2016, 2017).
10. Since all variables of the translog cost function must be positive, we have to substitute the values of zeros for Y_1 and Y_2 in the first two terms of the numerator in Equation 16 by 10% of the smallest sample values of these two variables. Please refer to Mester (1987, 1993), T.-H. Huang (1998), and T.-H. Huang, Chang, and Chiu (2009).

11. One cannot apply the two-stage model to examine the performance of one-person audit firms, which is a restriction of the model. In other words, researchers can only apply our model to investigate the performance of multi-person audit firms.
12. Since the TEJ database gives the same sample audit firm a different firm number each year, it restricts researchers from constructing panel data. We, therefore, choose not to include the variable of time trend in the regression equations as the regressors.
13. According to the quiet life hypothesis, firms in highly concentrated markets often lack competitive pressure due to their high market share, resulting in low production efficiency. See, for example, Berger and Hannan (1998) and Koetter, Kolari, and Spierdijk (2012).
14. There are four share equations, but we can only include three of them to avoid the singularity of the error covariance matrix.
15. Although the estimate of α_2 is quite small, it appears reasonable since undertaking cases tends to be labor-intensive, instead of capital-intensive.
16. Note that all ten dependence parameters are significantly estimated at the 1% level, indicating that the interactions among the five error terms, $\varepsilon_1, \ldots, \varepsilon_5$, should be considered in the econometric model. The most relevant one is the dependence parameter between ε_1 and ε_2, Ω_{12}. However, it is difficult for us to identify whether the dependence can be ascribable to v_1 and v_2, u_1 and u_2, v_1 and u_2, u_1 and v_2, or all the above four cases.
17. To test the null hypotheses that both forms of firms have equal mean production (cost) efficiency in the two production stages, we obtain the t-test statistics equaling 4.78 and 6.89, respectively, under the assumption that the variances of both types of firms are equal. Both test statistics are significantly different from zero at the 1% level. As a sensitivity analysis, we re-classify the sample firms into small and large firms using 20 employees as the dividing line. The results are intact, i.e. small firms outperform large ones in the two production stages.
18. These two average values are all positive, confirming that the inputs' marginal products and outputs' marginal costs are positive, indirectly verifying that the regression coefficient estimates satisfy the requirements of the microeconomic theory.
19. Based on Table 4, the coefficient estimate of divy in the second stage is negative and significant at the 1% level, which supports that firms with a higher degree of product diversification in the second production stage have higher cost efficiency.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Yi-Chun Lin http://orcid.org/0000-0002-3711-3477
Kuo-Jui Huang http://orcid.org/0000-0002-8342-3329

References

Aigner, D., C. A. K. Lovell, and P. Schmidt. 1977. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6 (1): 21–37. https://doi.org/10.1016/0304-4076(77)90052-5.

Amsler, C., A. Prokhorov, and P. Schmidt. 2014. "Using Copulas to Model Time Dependence in Stochastic Frontier Models." *Econometric Reviews* 33 (5–6): 497–522. https://doi.org/10.1080/07474938.2013.825126.

Amsler, C., A. Prokhorov, and P. Schmidt. 2016. "Endogeneity in Stochastic Frontier Models." *Journal of Econometrics* 190 (2): 280–288. https://doi.org/10.1016/j.jeconom.2015.06.013.

Amsler, C., A. Prokhorov, and P. Schmidt. 2017. "Endogenous Environmental Variables in Stochastic Frontier Models." *Journal of Econometrics* 199 (2): 131–140. https://doi.org/10.1016/j.jeconom.2017.05.005.

Atkinson, S. E., and C. Cornwell. 1993. "Measuring Technical Efficiency with Panel Data: A Dual Approach." *Journal of Econometrics* 59 (3): 257–261. https://doi.org/10.1016/0304-4076(93)90025-Z.

Atkinson, S. E., and C. Cornwell. 1994. "Parametric Estimation of Technical and Allocative Inefficiency with Panel Data." *International Economic Review* 35 (1): 231–243. https://doi.org/10.2307/2527099.

Ball, R. 2009. "Market and Political/Regulatory Perspectives on the Recent Accounting Scandals." *Journal of Accounting Research* 47 (2): 277–323. https://doi.org/10.1111/j.1475-679X.2009.00325.x.

Banker, R. D., H. Chang, and Y.-C. Kao. 2002. "Impact of Information Technology on Public Audit Firm Productivity." *Journal of Information Systems* 16 (2): 209–222. https://doi.org/10.2308/jis.2002.16.2.209.

Banker, R. D., H. Chang, and R. Natarajan. 2005. "Productivity Change, Technical Progress, and Relative Efficiency Change in the Public Accounting Industry." *Management Science* 51 (2): 291–304. https://doi.org/10.1287/mnsc.1040.0324.

Banker, R. D., H. Chang, and R. Natarajan. 2007. "Estimating DEA Technical and Allocative Inefficiency Using Aggregate Cost or Revenue Data." *Journal of Productivity Analysis* 27 (2): 115–121. https://doi.org/10.1007/s11123-006-0027-1.

Banker, R. D., A. Charnes, and W. W. Cooper. 1984. "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science* 30 (9): 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078.

Barton, J. 2005. "Who Cares About Auditor Reputation?" *Contemporary Accounting Research* 22 (3): 549–586. https://doi.org/10.1506/C27U-23K8-E1VL-20R0.

Battese, G. E., D. S. P. Rao, and C. J. O'Donnell. 2004. "A Metafrontier Production Function for Estimation of Technical Efficiencies and Technology Gaps for Firms Operating Under Different Technologies." *Journal of Productivity Analysis* 21 (1): 91–103. https://doi.org/10.1023/B:PROD.0000012454.06094.29.

Berardino, J. F. 2000. "Session Five, Hearing Testimony: Auditor Independence." In *United States Securities & Exchange Commission Hearing on Auditor Independence*. Washington, D.C.: Securities and Exchange Commission. https://www.sec.gov/files/rules/extra/audmin.htm.

Berger, A. N., and T. H. Hannan. 1998. "The Efficiency Cost of Market Power in the Banking Industry: A Test of the "Quiet Life" and Related Hypotheses." *Review of Economics and Statistics* 80 (3): 454–465. https://doi.org/10.1162/003465398557555.

Black, B. S. 2000. "The Legal and Institutional Preconditions for Strong Securities Markets." *UCLA Law Review* 48:781–855. https://doi.org/10.2139/ssrn.182169.

Chang, B.-G., T.-H. Huang, and C.-Y. Kuo. 2015. "A Comparison of the Technical Efficiency of Audit Firms Among the US, China, and Taiwan Under the Framework of a Stochastic Metafrontier Production Function." *Journal of Productivity Analysis* 44 (3): 337–349. https://doi.org/10.1007/s11123-014-0397-8.

Chang, H., J. Chen, R.-R. Duh, and S.-H. Li. 2011. "Productivity Growth in the Public Accounting Industry: The Roles of Information Technology and Human Capital." *Auditing: A Journal of Practice & Theory* 30 (1): 21–48. https://doi.org/10.2308/aud.2011.30.1.21.

Chang, H., H. L. Choy, W. W. Cooper, B. R. Parker, and T. W. Ruefli. 2009. "Measuring Productivity Growth, Technical Progress, and Efficiency Changes of CPA Firms Prior to and Following the Sarbanes–Oxley Act." *Socio-Economic Planning Sciences* 43 (4): 221–228. https://doi.org/10.1016/j.seps.2008.11.002.

Chang, H., H. L. Choy, W. W. Cooper, and T. W. Ruefli. 2009. "Using Malmquist Indexes to Measure Changes in the Productivity and Efficiency of US Audit Firms Before and After the sarbanes–Oxley Act." *Omega* 37 (5): 951–960. https://doi.org/10.1016/j.omega.2008.08.004.

Chang, H.-H., H.-L. Choy, W. W. Cooper, and M.-H. Lin. 2015. "Service Mix and Productivity: Evidence from Accounting Firms." *NTU Management Review* 25 (3): 197–226. https://doi.org/10.6226/NTUMR.2015.AUG.R12032.

Charnes, A., W. W. Cooper, and E. Rhodes. 1978. "Measuring the Efficiency of Decision-Making Units." *European Journal of Operational Research* 2 (6): 429–444. https://doi.org/10.1016/0377-2217(78)90138-8.

Copley, P. A., and M. S. Doucet. 1993. "The Impact of Competition on the Quality of Governmental Audits." *Auditing: A Journal of Practice & Theory* 12 (1): 88–98.

Dopuch, N., M. Gupta, D. A. Simunic, and M. T. Stein. 2003. "Production Efficiency and the Pricing of Audit Services." *Contemporary Accounting Research* 20 (1): 47–77. https://doi.org/10.1506/6UDH-HM5M-3W63-PKJP.

Duncan, R. B. 1972. "Characteristics of Organizational Environments and Perceived Environmental Uncertainty." *Administrative Science Quarterly* 17 (3): 313–327. https://doi.org/10.2307/2392145.

Färe, R., and S. Grosskopf. 2000. "Network DEA." *Socio-Economic Planning Sciences* 34 (1): 35–49. https://doi.org/10.1016/S0038-0121(99)00012-9.

Francis, J. R. 1984. "The Effect of Audit Firm Size on Audit Prices: A Study of the Australian Market." *Journal of Accounting and Economics* 6 (2): 133–151. https://doi.org/10.1016/0165-4101(84)90010-7.

Francis, J. R., and D. J. Stokes. 1986. "Audit Prices, Product Differentiation, and Scale Economies: Further Evidence from the Australian Market." *Journal of Accounting Research* 24 (2): 383–393. https://doi.org/10.2307/2491141.

Francis, J. R., and E. R. Wilson. 1988. "Auditor Changes: A Joint Test of Theories Relating to Agency Costs and Auditor Differentiation." *The Accounting Review* 63 (4): 663–682.

Frendy,and D. Hu. 2014. "Japanese Stock Market Reaction to Announcements of News Affecting Auditors' Reputation: The Case of the Olympus Fraud." *Journal of Contemporary Accounting & Economics* 10 (3): 206–224. https://doi.org/10.1016/j.jcae.2014.08.004.

Fukuyama, H., and W. L. Weber. 2015. "Measuring Japanese Bank Performance: A Dynamic Network DEA Approach." *Journal of Productivity Analysis* 44 (3): 249–264. https://doi.org/10.1007/s11123-014-0403-1.

Fung, S. Y. K., F. A. Gul, and J. Krishnan. 2012. "City-Level Auditor Industry Specialization, Economies of Scale, and Audit Pricing." *The Accounting Review* 87 (4): 1281–1307. https://doi.org/10.2308/accr-10275.

Gerloff, E. A. 1985. *Organizational Theory and Design: A Strategic Approach for Management*. New York: McGraw-Hill.

Greene, W. H. 2003. "Simulated Likelihood Estimation of the Normal-Gamma Stochastic Frontier Function." *Journal of Productivity Analysis* 19 (2): 179–190. https://doi.org/10.1023/A:1022853416499.

Greenwood, R., S. X. Li, R. Prakash, and D. L. Deephouse. 2005. "Reputation, Diversification, and Organizational Explanations of Performance in Professional Service Firms." *Organization Science* 16 (6): 661–673. https://doi.org/10.1287/orsc.1050.0159.

Grosskopf, S., K. Hayes, L. L. Taylor, and W. Weber. 2015. "Centralized or Decentralized Control of School Resources? A Network Model." *Journal of Productivity Analysis* 43 (2): 139–150. https://doi.org/10.1007/s11123-013-0379-2.

Hausman, J. A. 1987. "Specification Tests in Econometrics." *Econometrica* 46 (6): 1251–1271. https://doi.org/10.2307/1913827.

Holod, D., and H. F. Lewis. 2011. "Resolving the Deposit Dilemma: A New DEA Bank Efficiency Model." *Journal of Banking & Finance* 35 (11): 2801–2810. https://doi.org/10.1016/j.jbankfin.2011.03.007.

Huang, C. J., T.-H. Huang, and N.-H. Liu. 2014. "A New Approach to Estimating the Metafrontier Production Function Based on a Stochastic Frontier Framework." *Journal of Productivity Analysis* 42 (3): 241–254. https://doi.org/10.1007/s11123-014-0402-2.

Huang, T.-H. 1998. "Estimating Scale and Scope Economics in Taiwan's Banking Sector Using a Stochastic Frontier Cost Function." *Taiwan Economic Review* 26 (2): 209–241.

Huang, T.-H., B.-G. Chang, and Y.-F. Chiu. 2009. "An Examination of Banks' Efficiency for Six Asian Countries Using the Metafrontier Cost Function." *Academia Economic Papers* 37 (1): 61–100. https://doi.org/10.29628/AEP.200903.0003.
Huang, T.-H., B.-G. Chang, and C.-Y. Kuo. 2019. "Comparing the Metafrontier Malmquist Productivity Changes of Public Audit Firms Across Countries." *Asia-Pacific Journal of Accounting & Economics* 26 (5): 589–608. https://doi.org/10.1080/16081625.2017.1354712.
Huang, T.-H., K.-C. Chen, and C.-I. Lin. 2018. "An Extension from Network DEA to Copula-Based Network SFA: Evidence from the U.S. Commercial Banks in 2009." *The Quarterly Review of Economics and Finance* 67:51–62. https://doi.org/10.1016/j.qref.2017.04.007.
Huang, T.-H., C.-I. Lin, and K.-C. Chen. 2017. "Evaluating Efficiencies of Chinese Commercial Banks in the Context of Stochastic Multistage Technologies." *Pacific-Basin Finance Journal* 41:93–110. https://doi.org/10.1016/j.pacfin.2016.12.008.
Huang, T.-H., C.-I. Lin, and R.-C. Wu. 2019. "Assessing the Marketing and Investment Efficiency of Taiwan's Life Insurance Firms Under Network Structures." *The Quarterly Review of Economics and Finance* 71:132–147. https://doi.org/10.1016/j.qref.2018.07.002.
Juo, J.-C., T.-T. Fu, M.-M. Yu, and Y.-H. Lin. 2016. "Non-Radial Profit Performance: An Application to Taiwanese Banks." *Omega* 65:111–121. https://doi.org/10.1016/j.omega.2016.01.003.
Kao, C. 2009. "Efficiency Decomposition in Network Data Envelopment Analysis: A Relational Model." *European Journal of Operational Research* 192 (3): 949–962. https://doi.org/10.1016/j.ejor.2007.10.008.
Kao, C. 2014. "Efficiency Decomposition in Network Data Envelopment Analysis with Slacks-Based Measures." *Omega* 45:1–6. https://doi.org/10.1016/j.omega.2013.12.002.
Kao, C., and S.-N. Hwang. 2010. "Efficiency Measurement for Network Systems: IT Impact on Firm Performance." *Decision Support Systems* 48 (3): 437–446. https://doi.org/10.1016/j.dss.2009.06.002.
Knechel, W. R., P. Rouse, and C. Schelleman. 2009. "A Modified Audit Production Framework: Evaluating the Relative Efficiency of Audit Engagements." *The Accounting Review* 84 (5): 1607–1638. https://doi.org/10.2308/accr.2009.84.5.1607.
Koetter, M., J. W. Kolari, and L. Spierdijk. 2012. "Enjoying the Quiet Life Under Deregulation? Evidence from Adjusted Lerner Indices for U.S. Banks." *Review of Economics and Statistics* 94 (2): 462–480. https://doi.org/10.1162/REST_a_00155.
Kutlu, L. 2010. "Battese-Coelli Estimator with Endogenous Regressors." *Economics Letters* 109 (2): 79–81. https://doi.org/10.1016/j.econlet.2010.08.008.
Laeven, L., and R. Levine. 2007. "Is There a Diversification Discount in Financial Conglomerates?" *Journal of Financial Economics* 85 (2): 331–367. https://doi.org/10.1016/j.jfineco.2005.06.001.
Lai, H.-P., and C. J. Huang. 2013. "Maximum Likelihood Estimation of Seemingly Unrelated Stochastic Frontier Regressions." *Journal of Productivity Analysis* 40 (1): 1–14. https://doi.org/10.1007/s11123-012-0289-8.
Lewis, H. F., S. Mallikarjun, and T. R. Sexton. 2013. "Unoriented Two-Stage DEA: The Case of the Oscillating Intermediate Products." *European Journal of Operational Research* 229 (2): 529–539. https://doi.org/10.1016/j.ejor.2013.02.058.
Lewis, H. F., and T. R. Sexton. 2004. "Network DEA: Efficiency Analysis of Organizations with Complex Internal Structure." *Computers & Operations Research* 31 (9): 1365–1410. https://doi.org/10.1016/S0305-0548(03)00095-9.
Lin, Y.-C., and T.-H. Huang. 2023. "Studying Economic Efficiency of Taiwan's Accounting Firms Under the Stochastic Meta-Frontier Framework with Endogenous Inputs." *Taiwan Economic Forecast and Policy* 53 (2): 105–155. https://www.airitilibrary.com/Article/Detail/17298849-N202304010008-00003.
Meeusen, W., and J. Van Den Broeck. 1977. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review* 18 (2): 435–444. https://doi.org/10.2307/2525757.

Melancon, B. C. 1998. “The Changing Strategy for the Profession, the CPA and the AICPA: What This Means for the Education Community.” *Accounting Horizons* 12 (4): 397–406.
Mester, L. J. 1987. “A Multiproduct Cost Study of Savings and Loans.” *The Journal of Finance* 42 (2): 423–445. https://doi.org/10.1111/j.1540-6261.1987.tb02575.x.
Mester, L. J. 1993. “Efficiency in the Savings and Loan Industry.” *Journal of Banking & Finance* 17 (2–3): 267–286. https://doi.org/10.1016/0378-4266(93)90032-9.
Morris, T., and L. Empson. 1998. “Organisation and Expertise: An Exploration of Knowledge Bases and the Management of Accounting and Consulting Firms.” *Accounting, Organizations and Society* 23 (5–6): 609–624. https://doi.org/10.1016/S0361-3682(98)00032-4.
Nayyar, P. R. 1990. “Information Asymmetries: A Source of Competitive Advantage for Diversified Service Firms.” *Strategic Management Journal* 11 (7): 513–519. https://doi.org/10.1002/smj.4250110703.
Numan, W., and M. Willekens. 2012. “An Empirical Test of Spatial Competition in the Audit Market.” *Journal of Accounting and Economics* 53 (1–2): 450–465. https://doi.org/10.1016/j.jacceco.2011.10.002.
O’Donnell, C. J., and T. J. Coelli. 2005. “A Bayesian Approach to Imposing Curvature on Distance Functions.” *Journal of Econometrics* 126 (2): 493–523. https://doi.org/10.1016/j.jeconom.2004.05.011.
Prieto, A. M., and J. L. Zofio. 2007. “Network DEA Efficiency in Input–Output Models: With an Application to OECD Countries.” *European Journal of Operational Research* 178 (1): 292–304. https://doi.org/10.1016/j.ejor.2006.01.015.
Scott, M. C. 1998. *The Intellect Industry: Profiting and Learning from Professional Services Firms*. New York: Wiley.
Simar, L., and P. W. Wilson. 2007. “Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes.” *Journal of Econometrics* 136 (1): 31–64. https://doi.org/10.1016/j.jeconom.2005.07.009.
Sklar, M. 1959. “Fonctions de Répartition À N Dimensions Et Leurs Marges.” *Publications de l’Institut Statistique de l’Université de Paris* 8 (3): 229–231. https://hal.science/hal-04094463v1 .
Starbuck, W. H. 1992. “Learning by Knowledge-Intensive Firms.” *Journal of Management Studies* 29 (6): 713–740. https://doi.org/10.1111/j.1467-6486.1992.tb00686.x.
Tone, K. 2001. “A Slacks-Based Measure of Efficiency in Data Envelopment Analysis.” *European Journal of Operational Research* 130 (3): 498–509. https://doi.org/10.1016/S0377-2217(99)00407-5.
Tone, K. 2002. “A Slacks-Based Measure of Super-Efficiency in Data Envelopment Analysis.” *European Journal of Operational Research* 143 (1): 32–41. https://doi.org/10.1016/S0377-2217(01)00324-1.
Tone, K., and M. Tsutsui. 2010. “Dynamic DEA: A Slacks-Based Measure Approach.” *Omega* 38 (3–4): 145–156. https://doi.org/10.1016/j.omega.2009.07.003.
Tone, K., and M. Tsutsui. 2014. “Dynamic DEA with Network Structure: A Slacks-Based Measure Approach.” *Omega* 42 (1): 124–131. https://doi.org/10.1016/j.omega.2013.04.002.
Tran, K. C., and E. G. Tsionas. 2013. “GMM Estimation of the Stochastic Frontier Model with Endogenous Regressors.” *Economics Letters* 118 (1): 233–236. https://doi.org/10.1016/j.econlet.2012.10.028.
Tran, K. C., and E. G. Tsionas. 2015. “Endogeneity in Stochastic Frontier Models: Copula Approach Without External Instruments.” *Economics Letters* 133:85–88. https://doi.org/10.1016/j.econlet.2015.05.026.
Trebilcock, M. J., and L. Csorgo. 2001. “Multi-Disciplinary Professional Practices: A Consumer Welfare Perspective.” *The Dalhousie Law Journal* 24 (2): 1–36.
Tsay, W.-J., C. J. Huang, T.-T. Fu, and I.-L. Ho. 2013. “A Simple Closed-Form Approximation for the Cumulative Distribution Function of the Composite Error of Stochastic Frontier Models.” *Journal of Productivity Analysis* 39 (3): 259–269. https://doi.org/10.1007/s11123-012-0283-1.

Wang, K., W. Huang, J. Wu, and Y.-N. Liu. 2014. “Efficiency Measures of the Chinese Commercial Banking System Using an Additive Two-Stage DEA.” *Omega* 44:5–20. https://doi.org/10.1016/j.omega.2013.09.005.

Watts, R. L., and J. L. Zimmerman. 1983. “Agency Problems, Auditing, and the Theory of the Firm: Some Evidence.” *The Journal of Law and Economics* 26 (3): 613–633. https://doi.org/10.1086/467051.

Yang, C., and H.-M. Liu. 2012. “Managerial Efficiency in Taiwan Bank Branches: A Network DEA.” *Economic Modelling* 29 (2): 450–461. https://doi.org/10.1016/j.econmod.2011.12.004.

Zha, Y., N. Liang, M. Wu, and Y. Bian. 2016. “Efficiency Evaluation of Banks in China: A Dynamic Two-Stage Slacks-Based Measure Approach.” *Omega* 60:60–72. https://doi.org/10.1016/j.omega.2014.12.008.

Appendices

Appendix 1

Approximate integrations of $I^{app}(Q_1)$ and $I^{app}(Q_2)$

$$
\begin{aligned}
I^{app}(Q_1) = {} & \frac{1}{bk_1\sqrt{2}}\left\{1 + erf\left[\frac{(a+bQ_1)k_1}{2}\frac{1+sign(\theta_1)}{2} + \frac{k_2}{\sqrt{2}}\right]\right\} + \\
& \frac{1}{4b}\sqrt{\frac{2}{k_1^2 - 2c_2}}\exp\left[\frac{(k_1k_2 - c_1)^2}{2(k_1^2 - 2c_2)} - \frac{k_2^2}{2}\right] \times \\
& \left\{1 - erf\left[\frac{k_1k_2 - c_1}{\sqrt{2(k_1^2 - 2c_2)}} + \frac{(a+bQ_1)\sqrt{k_1^2 - 2c_2}}{2}\frac{1+sign(\theta_1)}{2}\right]\right\} - \\
& \frac{1}{4b}\sqrt{\frac{2}{k_1^2 - 2c_2}}\exp\left[\frac{(k_1k_2 + c_1)^2}{2(k_1^2 - 2c_2)} - \frac{k_2^2}{2}\right] \times \\
& \left\{erf\left[\frac{k_1k_2 + c_1}{\sqrt{2(k_1^2 - 2c_2)}}\right] - erf\left[\frac{k_1k_2 + c_1}{\sqrt{2(k_1^2 - 2c_2)}} + \frac{(a+bQ_1)\sqrt{k_1^2 - 2c_2}}{2}\right]\right\}\frac{1-sign(\theta_1)}{2},
\end{aligned}
\tag{A1}
$$

where $a = \frac{\gamma_1' q_1 \sigma_1}{\sigma_{v_1}\sigma_{\omega_1}}$, $b = \frac{\lambda}{\sigma_1} > 0$, $\lambda = \frac{\sigma_{\omega_1}}{\sigma_{v_1}} > 0$, $c = \frac{1}{\sigma_1} > 0$, $k_1 = \frac{\sqrt{2}c}{b} > 0$, $k_2 = \frac{-ac}{b}$, $\theta_1 = \frac{a+bQ_1}{\sqrt{2}}$, and $sign(\theta_1) = -1, 0, 1$, if $\theta_2 <, =, > 0$. Notation $erf(\cdot)$ is called the error function, which is defined as follows:

$$
\begin{aligned}
erf(z) &= \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt = 2\int_0^{\sqrt{2}z}\phi(t)dt = 2\Phi\left(\sqrt{2}z\right) - 1 \\
&\approx 1 - \exp\left(c_1 z + c_2 z^2\right) = g(z),
\end{aligned}
\tag{A2}
$$

where $z \geq 0$, and $c_1 = -1.09500814703333$ and $c_2 = -0.75651138383854$ are the two parameters obtained by the simulation approach of Tsay et al. (2013), which lets $g(z)$ be very close to $erf(z)$.

Appendix 2

Coefficient estimates without environmental variables.

First stage: Production frontier			Second stage: Cost frontier		
Variable	Coefficient estimate	Standard error	Variable	Coefficient estimate	Standard error
α_1	.3861***	.0306	Intercept	11.5079***	.1155
α_2	.4107E-03	.6148E-02	$\ln(p_2/p_1)$	.5638***	.0135
Intercept	5.1074***	.6482	$\ln(p_3/p_1)$	.4311***	.0144
$\ln x_1$	.1047	.1213	$\ln(p_4/p_1)$	−.1064***	.0172
$\ln x_2$	.0418	.1684	$0.5[\ln(p_2/p_1)]^2$	.0804***	.1164E-02
$0.5(\ln x_1)^2$	.2348***	.0134	$0.5[\ln(p_3/p_1)]^2$	.0831***	.1803E-02
$0.5(\ln x_2)^2$	−.0109*	.6468E-02	$0.5[\ln(p_4/p_1)]^2$	.1118***	.1747E-02
$\ln x_1 \ln x_2$	.7721E-02	.7609E-02	$\ln(p_2/p_1)\ln(p_3/p_1)$	−.0283***	.9415E-03
			$\ln(p_2/p_1)\ln(p_4/p_1)$	−.0466***	.1244E-02
Ω_{12}	.8519***	.5647E-02	$\ln(p_3/p_1)\ln(p_4/p_1)$	−.0356***	.1488E-02
Ω_{13}	−.0423*	.0228	$\ln y_1$	−.9135***	.0183
Ω_{14}	.4095***	.0170	$\ln y_2$	−.3438***	.0154
Ω_{23}	.0756***	.0204	$0.5(\ln y_1)^2$	.0953***	.1357E-02
Ω_{24}	.1074***	.0188	$0.5(\ln y_2)^2$	.0510***	.1839E-02
Ω_{34}	−.1364***	.0225	$\ln y_1 \ln y_2$	−.0190***	.1236E-02
Ω_{15}	−.2371***	.0166	$\ln(p_2/p_1)\ln y_1$	−.4145E-02***	.1077E-02
Ω_{25}	−.2265E-02	.0186	$\ln(p_2/p_1)\ln y_2$	−.4575E-02***	.8963E-03
Ω_{35}	−.0883***	.0231	$\ln(p_3/p_1)\ln y_1$	.8016E-03	.9045E-03
Ω_{45}	−.6807***	.8269E-02	$\ln(p_3/p_1)\ln y_2$	.3519E-02***	.7789E-03
σ_3	.0785***	.1194E-02	$\ln(p_4/p_1)\ln y_1$	.9551E-02***	.1025E-02
σ_4	.1002***	.1163E-02	$\ln(p_4/p_1)\ln y_2$	−.1227E-02	.8366E-03
σ_5	.0982***	.1142E-02	σ_2	.6157***	.0132
σ_1	.99998***	.0242	λ_2	1.1742***	.0634
λ_1	1.0841***	.0623			
Log-likelihood	3,016.81				
Sample size	3,410				

Note: 1. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.
2. σ_3, σ_4, and σ_5 are estimated standard deviations of the error terms in three share equations.
3. Ω_{ij}, $i, j = 1, \ldots, 5$, $i < j$, is the estimated off-diagonal element of the correlation coefficient matrix of Ω in Equation 8.