# Measuring Technical, Allocative inefficiency, and Cost Inefficiency by Applying Duality Theory

Tung Liu*

6/16/2020

* Professor of Economics, Ball State University, Indiana 47306, U.S. email: tliu@bsu.edu.

**Measuring Technical, Allocative inefficiency, and Cost Inefficiency by Applying Duality Theory**

**Abstract**

Total cost inefficiency comprises technical inefficiency and allocative inefficiency. Differentiating between the two components of total cost inefficiency has been challenging both theoretically and empirically. Current literature has solved this problem for specific cases, but not in general. Using the duality of production and cost functions, this paper proposes a new approach to solve this problem for general cases. Firstly, we define two types of allocative inefficiency and their input cost share inequalities. Secondly, a three-stage procedure is developed to estimate total cost inefficiency and its two components. Finally, an empirical presentation based on the existing data supplements the theory.

JEL Subject Code: D21, D22, D24

Keywords: Cost inefficiency; Technical inefficiency; Allocative inefficiency; Duality theory

**Measuring Technical, Allocative Inefficiency, and Cost Inefficiency by Applying Duality Theory**

I. Introduction

Although classical economic theory assumes that the production possibility frontier incorporates all output and input combinations, technical inefficiency can occur when the observed combinations are not situated on the frontier. When data of output and inputs are available, the methods which have been used to estimate technical inefficiency in the production function include the stochastic frontier models (Aigner *et al*., 1977; Battese and Coelli, 1988 and 1992; Greene, 2005; Kumbhakar and Lovell, 2000) among others.

Technical inefficiency is said to occur when the observed output is less than the optimal output for a given level of inputs, or when firms use more inputs than they should for a given level of output. In either case, technical inefficiency leads to cost inefficiency, which is the difference between the observed total cost and the minimized cost derived from the cost function. Furthermore, even if there is no technical inefficiency, cost inefficiency can still occur if the cost minimization condition does not hold. This type of cost inefficiency is called cost allocative inefficiency (CAI). Hence, total cost inefficiency comprises both technical inefficiency and CAI.

Stochastic frontier models have been applied to estimate the empirical aspects of cost inefficiency. One major problem, however, is that this modelling approach cannot differentiate between the two components of cost inefficiency. This empirical difficulty is known as the “Greene problem” (Greene, 1980; Bauer, 1990a) and several solutions have been proposed. Kopp and Diewert (1982) applied a duality theory to the cost function to estimate these two components. Another popular solution, which is also based on duality theory, is to include the first-order condition of the cost minimization problem in estimating the production function or the cost function (Schmidt and Lovell, 1979; Kumbhakar, 1997; Kumbhakar and Lovell, 2000; Kumbhakar and Wang, 2006b). Although this issue has been solved for some specific production and cost functions, such as a Cobb-Douglas production function (Schmidt and Lovell, 1979) or a translog cost function (Kumbhakar, 1997), a general solution has not been provided.

This paper proposes a general solution in measuring total cost inefficiency and its two components. In Section II, we briefly review inefficiency and duality theory. For the review of duality theory, we emphasize that the optimal solution for the cost minimization problem can be studied from either the production function side or the cost function side. In Section III, by using the duality relationship between production and cost functions in the cost minimization problem, we define two types of CAI assuming there is no technical inefficiency. The first type of CAI is caused by the difference between the observed input prices and input shadow prices derived from the production function. The second type of CAI is caused by the difference between the observed inputs and optimal inputs derived from the cost function. We derive the input cost share inequalities for each type of CAI and two measures of CAI. In Section IV, we extend the analysis of the two types of CAI to the case when there is technical inefficiency and derive two forms of cost inefficiency decomposition. In Section V, we explain the appropriate use of input cost share equations in estimations. In Section VI, we construct a three-stage procedure to estimate the two components of cost inefficiency. We also show an empirical application by using the data from electric power companies used in Nerlove (1963) and Christensen and Greene (1976). Finally, we conclude our findings in the last section.

## II. Inefficiency, Cost Minimization, and Duality

Duality theory is frequently applied to measure the two components of cost inefficiency (Kopp and Diewert, 1982; Schmidt and Lovell, 1979; Kumbhakar, 1997; Kumbhakar and Wang, 2006a, 2006b). This section reviews the basic concepts of technical and cost inefficiency and the relationship between inefficiency and duality theory. Consider the production with a single-product output and $k$ inputs. Let $y = f(x)$ be the production function, where $y$ is the output and $x$ is a $k \times 1$ vector with the quantities of $k$ inputs. Technical inefficiency occurs when the observed $(y, x)$ are not on the production possibility frontier. A production function with technical inefficiency for firm $i$ can be specified as

$$y_i = f(x_i)e^{-u_{pi}}, \qquad (1)$$

where $y_i$ is the output for firm $i$; $x_i = (x_{i1}, x_{i2}, \cdots, x_{ik})$ is an input vector with $k$ inputs and the element $x_{ij}$ is the quantity of input $j$ for firm $i$; $e^{-u_{pi}}$ is the technical inefficiency term. When $u_{pi} > 0$, the technical inefficiency (TE) for firm $i$ is

$$TE_i = \frac{y_i}{f(x_i)} = e^{-u_{pi}} < 1. \quad (2)$$

The stochastic frontier model to measure this technical inefficiency can be shown as (Aigner *et al.*, 1977; Battese and Coelli, 1988, 1992; Greene, 2005):

$$\ln y_i = \ln f(x_i; \beta) + v_{pi} - u_{pi}, \quad (3)$$

where $\beta$ is a vector of parameters, $v_{ci}$ is a two-sided random error term for noise, and $u_{ci}$ is a one-side random error with a positive mean to measure technical inefficiency. With the data values of $(y_i, x_i)$, $i = 1, 2, \cdots n$ from $n$ firms, the parameter vector $\beta$ can be estimated by the maximum likelihood estimators (MLE).

The analysis of technical inefficiency in Equations (1) – (3) does not involve input prices. When input prices are given, cost inefficiency can be analyzed. Denote the input prices as $w$, a $k \times 1$ vector of $k$ inputs. When input prices $w$ and $y = f(x)$ are given, the cost function $C(y, w)$ is derived from the following cost minimization problem:

$$C(y, w) = \min_{x} w^T x \text{ subject to } y = f(x). \quad (4)$$

Cost inefficiency for firm $i$ occurs when the firm's actual total cost is greater than the firm's minimized cost. Input prices for firm $i$ are denoted as $w_i = (w_{i1}, w_{i2}, \cdots, w_{ik})$, where $w_{ij}$ is the price of input $j$ for firm $i$. For firm $i$, the observed total cost is $TC(x_i, w_i) = \sum_j w_{ij} x_{ij}$ and the minimized cost is $C(y_i, w_i)$. The cost inefficiency for firm $i$ is defined as (Farrell, 1957):

$$CE_i = \ln \frac{TC(x_i, w_i)}{C(y_i, w_i)} = \ln TC(x_i, w_i) - \ln C(y_i, w_i) \geq 0. \quad (5)$$

Cost inefficiency occurs when $TC(x_i, w_i) > C(y_i, w_i)$. With the data of the total cost, output quantity, and the input prices for firm $i = 1,2, \cdots, n$, a stochastic cost frontier model for estimating cost inefficiency can be shown as (Schmidt and Lovell, 1979; Kumbhakar, 1997; Kumbhakar and Lovell, 2000):

$$\ln TC(x_i, w_i) = \ln C(y_i, w_i; \gamma) + v_{ci} + u_{ci}, \tag{6}$$

where $\gamma$ is a parameter vector, $v_{ci}$ is a two-sided random error term for noise, and $u_{ci}$ is a one-side random error with a positive mean for the unobserved cost inefficiency.

One major issue when using Equation (6) to estimate cost inefficiency is that the inefficiency term $u_{ci}$ includes two types of inefficiency: technical inefficiency and CAI. Technical inefficiency occurs when the optimal output is greater than the actual output ($f(x_i) > y_i$) with given inputs $x_i$ (output-oriented inefficiency) or when the firm uses more inputs to produce the same output $y_i$ (input-oriented inefficiency). In either case, it leads to cost inefficiency, such that $TC(x_i, w_i) > C(y_i, w_i)$. Furthermore, even if there is no technical inefficiency, cost inefficiency can still occur due to CAI, which occurs when the first-order condition for the cost minimization problem of Equation (4) does not hold.

Before exploring CAI, we first examine the first-order condition. We simplify the notations by omitting the firm indicator *i* from the subscript of $x_{ij}$ and $w_{ij}$, and denote $x_j$ and $w_j$ as the quantity and the price of input *j*, respectively. The Lagrangian form for Equation (4) is

$$L(x, \lambda) = \sum_j w_j x_j + \lambda(y - f(x)), \tag{7}$$

where $\lambda$ is the Lagrange multiplier. The first-order condition is

$$w_j = \lambda \frac{\partial f}{\partial x_j} = \lambda \eta_j \frac{y}{x_j} \tag{8}$$

where $\eta_j = \frac{\partial f}{\partial x_j}\frac{x_j}{y}$ is the output elasticity of input $j$. Choosing the first input price as the numeraire, the relative input price of input $j$ is

$$\frac{w_j}{w_1} = \frac{\eta_j}{\eta_1}\frac{x_1}{x_j}. \tag{9}$$

Denote the optimal solution of the minimized total cost for Equation (4) as $C(y, w)$. The optimal quantity or the input demand function for input $j$ is uniquely determined by Shephard's lemma:

$$x_j^* = \frac{\partial C(y,w)}{\partial w_j}. \tag{10}$$

The optimal cost share of input $j$ from the cost function is

$$s_j^C = \frac{\partial \ln C(y,w)}{\partial \ln w_j} = \frac{\partial C(y,w)}{\partial w_j}\frac{w_j}{C(y,w)} = \frac{w_j x_j^*}{\sum_l w_l x_l^*}. \tag{11}$$

We can derive the economies of scale in the optimization process from Equation (8). Multiplying both sides of Equation (8) by $x_j^*$ and summing all input $j$, gives

$$TC = \lambda\,\eta\,y, \tag{12}$$

where $\eta = \sum_j \eta_j$ is economies of scale or returns to scale. The optimal solution for the Lagrangian multiplier in Equation (7) is $\lambda = \frac{\partial TC}{\partial y}$, which is the marginal cost. Then

$$\eta = \lambda^{-1} TC/y = \left(\frac{\partial TC}{\partial y}\cdot\frac{y}{TC}\right)^{-1} = \theta^{-1}, \tag{13}$$

where $\theta = \frac{\partial TC}{\partial y}\cdot\frac{y}{TC}$ is the elasticity of total cost with respect to output (Hanoch, 1975; Liu and Li, 2008). $\theta^{-1}$ is the economies of scale defined from the cost function (Christensen and Greene, 1976). Since $x_j^*$, $s_j^c$, and $\theta$ are all defined by the given $(y, w)$ and $C(y, w)$, denote these as $x^*(y, w)$, $s_j^c(y, w)$, and $\theta(y, w)$, respectively; each of $x_j^*$, $s_j^c$, and $\theta$ is a function of $(y, w)$.

Equations (10) – (13) are derived from the first-order condition based on the cost function. An alternative way to examine the first-order condition is from the production function side. Consider the first-order condition in Equations (8) and (9) from the production function and define input shadow prices as

$$w_j^* = \lambda \frac{\partial f}{\partial x_j} = \lambda \eta_j \frac{y}{x_j}. \tag{14}$$

$$\frac{w_j^*}{w_1^*} = \frac{\eta_j}{\eta_1} \frac{x_1}{x_j}. \tag{15}$$

Without loss of generality, denote $w^*$ as a vector of relative input shadow prices, such that $w_1^* = 1$, $w_2^* = w_2^*/w_1^*$, $\cdots$, and $w_k^* = w_k^*/w_1^*$ . Multiplying both sides of Equation (14) by $x_j$ and summing all input $j$, it gives the total cost based on input shadow prices $TC(x, w^*) = \sum_j w_j^* x_j$ and $TC(x, w^*) = \lambda \eta y$. Multiplying both sides of Equation (14) by $x_j$ to get $w_j^* x_j = \lambda \eta_j y$ and dividing this equation by $TC = \lambda \eta y$, it gives the cost share of input *j* based on the production function.

$$s_j^p = \frac{w_j^* x_j}{TC} = \frac{\lambda \eta_j y}{\lambda \eta y} = \frac{\eta_j}{\eta}. \tag{16}$$

This cost share equation does not require input price information. Since $w_j^*$, $s_j^p$, and $\eta$ are all defined by the given $(y, x)$ and $y = f(x)$, denote these as $w_j^*(y, x)$, $s_j^p(y, x)$, and $\eta(y, x)$, respectively. Comparing two optimal cost share equations, the cost share equation defined by the cost function (Equation (11)) is more widely used, whereas the one defined by the production function (Equation (16)) is relatively unknown with only a few applications (Kim, 1992; Li and Liu, 2011), although Kumbhakar and Lovell (2000, p284) noted the ratio of $\eta_j$ to $\eta$ as the input's normalized output elasticity.

Duality theory establishes the relationship between production and cost functions. In this theory, the cost function is derived from the cost minimization problem with a given production function, and it contains all the information necessary to reconstruct the production function (McFadden, 1978). Therefore, the input shadow prices $w^*$ from the production function should equal to the given input prices $w$, and optimal inputs $x^*$ from the cost function should equal to the given inputs $x$ when the first-order condition holds. This also implies that

$$s_j^c(y,w) = s_j^p(y,x) \text{ for all } j, \tag{17}$$

which shows that the cost share of input $j$ from the cost function is the same as its cost share from the production function. The relationships that $w = w^*$, $x = x^*$, $s_j^c(y,w) = s_j^p(y,x)$ and $\eta(y,x) = \theta(y,w)^{-1}$ are true only if the first-order condition holds. When the first-order condition defined from the cost function and the production function does not hold, we define two types of CAI in the next section.

## III. Cost Allocative Inefficiency without Technical Inefficiency

Our discussion of CAI in this section assumes there is no technical inefficiency. CAI occurs when the first-order condition does not hold. Based on the duality of production and cost functions, the first-order condition can be examined either from the production function or from the cost function. From the production function, the first-order condition gives input shadow prices; from the cost function, it gives optimal inputs. Using input shadow prices and optimal inputs, two types of CAI are defined.

Assume both production function $y = f(x)$ and cost function $C(y,w)$ are given. Denote the observed output, and input quantities and prices as $y^0$, $x^0$, and $w^0$, respectively, forming the triplet $(y^0, x^0, w^0)$. The first-order condition is related to all three variables in this triplet. CAI occurs when there is a mismatch of these three variables. Consider three pairs of two variables from this triplet: $(x^0, w^0)$, $(y^0, x^0)$, and $(y^0, w^0)$. The first pair $(x^0, w^0)$ gives the observed total cost $TC(x^0, w^0) = \sum_j w_j^0 x_j^0$, which is a numerical sum without any economic functions involved. The second pair $(y^0, x^0)$

involves the production function $y = f(x)$. Given $y^0 = f(x^0)$, the input shadow prices $w^*(y^0, x^0)$ are derived from Equation (14). These three variables, $y^0, x^0$, and $w^*$, form the production triplet $(y^0, x^0, w^*)^p$. The first type of CAI occurs when the observed input prices differ from input shadow prices ($w^0 \neq w^*$). This gives a mismatch in triplet form: $(y^0, x^0, w^0) \neq (y^0, x^0, w^*)^p$.

The third pair $(y^0, w^0)$ involves the cost function $C(y, w)$. Given $C(y^0, w^0)$, optimal inputs $x^*(y^0, w^0)$ are derived from Equation (10). These three variables, $y^0, x^*$, and $w^0$, form the cost triplet $(y^0, x^*, w^0)^c$. The second type of CAI occurs when the observed inputs differ from optimal inputs ($x^0 \neq x^*$). This gives another mismatch in triplet form: $(y^0, x^0, w^0) \neq (y^0, x^*, w^0)^c$. With CAI, we conclude that $(y^0, x^0, w^0)$ is not an efficient triplet, whereas both production and cost triplets are efficient triplets, and the three variables in each efficient triplet meet the first-order condition.

Current studies on CAI mainly discuss the first type of CAI, where $w^0 \neq w^*$. The second type of CAI is mostly ignored. The primary reason for this omission is that $w^0 \neq w^*$ implies $x^0 \neq x^*$ based on Equations (9) and (10). Thus, it appears to be redundant to consider the second type of CAI. However, in the following, we show that the two types of CAI have different implications upon two important elements in the CAI analysis.

The first important element in the CAI analysis is to incorporate the optimal input cost share equations from the cost function (Equation (11)) into a system of equations with the cost function and input cost share equations. This system of equations approach is commonly used for the models without CAI (Christensen and Greene, 1976) and with CAI (Kumbhakar, 1997). We examine input cost share equations for each type of CAI. For the first type of CAI, the observed relative price of input $j$ is different from its input shadow price, $\frac{w_j^0}{w_1^0} \neq \frac{w_j^*}{w_1^*}$. Using Equations (15) and (16), it gives the following cost share inequality.

$$\frac{w_j^0}{w_1^0} \neq \frac{\eta_j^0 x_1^0}{\eta_1^0 x_j^0} \Rightarrow \frac{w_j^0 x_j^0}{w_1^0 x_1^0} \neq \frac{\eta_j^0}{\eta_1^0} \Rightarrow \frac{w_j^0 x_j^0 \big/ \sum_l w_l^0 x_l^0}{w_1^0 x_1^0 \big/ \sum_l w_l^0 x_l^0} \neq \frac{\eta_j^0 \big/ \eta}{\eta_1^0 \big/ \eta} \Rightarrow \frac{s_j^0}{s_1^0} \neq \frac{s_j^p}{s_1^p}, \tag{18}$$

where $s_j^0(x^0, w^0) = \frac{w_j^0 x_j^0}{\sum_l w_l^0 x_l^0}$ is the observed cost share of input $j$ and $s_j^p(y^0, x^0) \; = \frac{w_j^* x_j^0}{\sum_l w_l^* x_l^0}$ is its optimal cost share from the production function (Equation (16)). If $s_1^0 = s_1^p$, then $s_j^0 \neq s_j^p$ for some $j$. If $s_1^0 \neq s_1^p$, then $s_j^0 \neq s_j^p$ for some $j$ since $\sum_l s_l^0 = \sum_l s_l^p = 1$. And

$$s_j^0(x^0, w^0) \neq s_j^p(y^0, x^0)\text{, for some } j \tag{19}$$

$$s_j^0(x^0, w^0) \neq s_j^c(y^0, w^*)\text{, for some } j, \tag{20}$$

where $s_j^c(y^0, w^*) = \frac{w_j^* x_j^0}{\sum_l w_l^* x_l^0}$ is the optimal cost share of input $j$ from the cost function. Equation (19) is derived from Equation (18). The first type of CAI is associated with the production triplet $(y^0, x^0, w^*)^p$, which is an efficient triplet. Based on the duality of the production function $y^0 = f(x^0)$ and the cost function $C(y^0, w^*)$ for this triplet, Equation (17) gives $s_j^p(y^0, x^0) = s_j^c(y^0, w^*) = \frac{w_j^* x_j^0}{\sum_l w_l^* x_l^0}$, and thereby Equation (20).

Note that the observed input prices $w^0$ are not involved in the optimal cost share equation $s_j^c$ in Equation (20). This implies $s_j^0 \neq s_j^c(y^0, w^0)$ is irrelevant to the first type of CAI and we cannot conclude $s_j^0 \neq s_j^c(y^0, w^0)$. This result is reasonable since $w^0 \neq w^*$ due to the first type of CAI and the optimal cost share of input $j$ should be a function of $w^*$ instead of $w^0$. Although $s_j^c(y^0, w^*)$ has been commonly used in current CAI studies, the inequality in Equation (20) has not been recognized in general cases since its derivation requires two additional steps. The first additional step is to derive Equation (19). The derivation of this inequality requires the use of the input cost share equation from the production function, $s_j^p(y^0, x^0)$. However, this measure of the cost share $s_j^p(y^0, x^0)$ is rarely known, as stated previously. The second additional step is from Equation (19) to Equation (20), which requires the application of duality theory.

For the second type of CAI, where $x^0 \neq x^*$, there is a different set of input cost share inequality. The second type of CAI is associated with the first-order condition from the cost function; we consider the optimal inputs $x^*$, the cost triplet $(y^0, x^*, w^0)^c$, and the optimal cost share of input $j$ from the cost function, $s_j^c(y^0, w^0) = \frac{w_j^0 x_j^*}{\sum_l w_l^0 x_l^*}$. Then the cost share inequalities are:

$$s_j^0(x^0, w^0) \neq s_j^c(y^0, w^0), \text{ for some } j. \tag{21}$$

$$s_j^0(x^0, w^0) \neq s_j^p(y^0, x^*), \text{ for some } j. \tag{22}$$

To prove Equation (21), we prove $\frac{w_j^0 x_j^0}{\sum_l w_l^0 x_l^0} \neq \frac{w_j^0 x_j^*}{\sum_l w_l^0 x_l^*}$. For the denominators in this inequality, we have $\sum_l w_l^0 x_l^0 > \sum_l w_l^0 x_l^*$ based on the second type of CAI. $\frac{w_j^0 x_j^0}{\sum_l w_l^0 x_l^0} = \frac{w_j^0 x_j^*}{\sum_l w_l^0 x_l^*}$ for all $j$ holds only if $x_j^0 > x_j^*$ for all $j$. Since there is no technical inefficiency, both $x^0$ and $x^*$ should be on the same production possibility frontier, i.e. $y^0 = f(x^0) = f(x^*)$. If $x_j^0 > x_j^*$ for all $j$, $x^0$ and $x^*$ cannot give the same output $y^0$ if $\frac{\partial f(x)}{\partial x_j} > 0$. Therefore, $\frac{w_j^0 x_j^0}{\sum_l w_l^0 x_l^0} \neq \frac{w_j^0 x_j^*}{\sum_l w_l^0 x_l^*}$ for some $j$, thereby proving Equation (21). The optimal cost share $s_j^c$ in Equation (21) is derived from the cost function with the cost triplet $(y^0, x^*, w^0)^c$. Using duality theory, Equation (17) gives $s_j^p(y^0, x^*) = s_j^c(y^0, w^0)$, and thereby Equation (22). The inequality in Equation (21) seems a natural result from CAI. Bauer (1990b, p.292) noted that $s^0 = s^c(y^0, w^0)$ if the firm is allocatively efficient, which is an indirect reference to Equation (21). Our proof shows that this inequality holds only for the second type of CAI.

We conclude that the two types of CAI give two different sets of cost share inequalities, i.e. Equations (19) and (20) as compared to Equations (21) and (22). This conclusion provides an insight to the first element in the CAI analysis. In the current literature, where only the first type of CAI is considered, the use of input cost share inequalities requires the information of input shadow prices $w^*$. An advantage of using the second type of CAI is that input cost share inequalities do not require the information of $w^*$; only the information of the observed input prices $w^0$ is enough in estimations. In one special case, this advantage or difference disappears. For a production function with a constant output elasticity for all values of $x_j$ and for all $j$ as in a Cobb-Douglas production function, $\eta_j$ is a constant for all $j$ and $\eta$ is a constant, which imply $s_j^p(y^0, x^0) = s_j^p(y^0, x^*)$. The two types of CAI share the same set of cost share inequalities.

The second important element and the key issue in the CAI analysis is to measure the size of inefficiency. For the second type of CAI, this measure can be easily defined. Because $x^*$ is the optimal solution from the cost minimization problem with the given $(y^0, w^0)$ and $y^0 = f(x^0) = f(x^*)$, the total cost $TC(x^0, w^0) = \sum_j w_j^0 x_j^0$ is greater than the minimized cost $C(y^0, w^0) = \sum_l w_l^0 x_l^*$. The measure of this type of CAI is

$$CAI = \ln \frac{TC(x^0, w^0)}{C(y^0, w^0)} = \ln \frac{\sum_j w_j^0 x_j^0}{\sum_j w_j^0 x_j^*} = \ln \sum_j w_j^0 x_j^0 - \ln \sum_j w_j^0 x_j^* > 0. \qquad (23)$$

A standard stochastic cost frontier model can be applied to the measure of this type of CAI.

However, it is worth noting that there is no similar measure of inefficiency for the first type of CAI. For the first type of CAI and its production triplet $(y^0, x^0, w^*)^p$, the minimized cost is $C(y^0, w^*)$. Suppose CAI is measured by the ratio of the observed total cost to this minimized cost, i.e. $CAI = \ln \frac{TC(x^0, w^0)}{C(y^0, w^*)} = \ln \frac{\sum_j w_j^0 x_j^0}{\sum_j w_j^* x_j^0}$. Because $(y^0, x^0, w^*)^p$ is an efficient triplet, the total cost is minimized at $x^0$ with respect to any other inputs $x' \neq x^0$ with the given $(y^0, w^*)$; i.e., $C(y^0, w^*) < TC(x', w^*)$ and $y^0 = f(x')$. The observed total cost $TC(x^0, w^0)$ is not in the feasible solutions of this cost minimization problem. Therefore, there is no conclusion about the ratio of $TC(x^0, w^0)$ to $C(y^0, w^*)$, and a stochastic cost frontier model with the data of $TC(x^0, w^0)$ and $(y^0, w^*)$ cannot be used to derive a CAI measure. Researchers cannot solve the "Greene Problem" in general cases if the studies focus only on the first type of CAI.

Instead of using $\ln \frac{TC(x^0, w^0)}{C(y^0, w^*)}$, an alternative measure of CAI can be developed for the given shadow prices $w^*$ and the minimized cost $C(y^0, w^*)$. The original triplet for the cost inefficiency analysis is $(y^0, x^0, w^0)$, where all variables are observed. Replacing both observed input quantities and prices with optimal inputs $x^*$ and input shadow prices $w^*$, a new triplet $(y^0, x^*, w^*)$ is formed. Since this triplet is not an efficient triplet, an alternative analysis of CAI can be derived based on $(y^0, x^*, w^*)$.

We apply the procedure of analyzing CAI for $(y^0, x^0, w^0)$ to the analysis of CAI for $(y^0, x^*, w^*)$. Consider three pairs of two variables from the triplet $(y^0, x^*, w^*)$:

$(x^*, w^*)$, $(y^0, x^*)$, and $(y^0, w^*)$. The first pair $(x^*, w^*)$ gives the "observed" total cost $TC(x^*, w^*) = \sum_l w_l^* x_l^*$. For the second pair $(y^0, x^*)$, using the production function $y^0 = f(x^*)$, it gives the input "shadow" prices $w^0$ and a new production triplet $(y^0, x^*, w^0)^p$. The results of $w^0 \neq w^*$ and $(y^0, x^*, w^*) \neq (y^0, x^*, w^0)^p$ are defined as the first type of CAI. For the third pair $(y^0, w^*)$, using the cost function $C(y^0, w^*)$, it gives the "optimal" inputs $x^0(y^0, w^*)$ and a new cost triplet $(y^0, x^0, w^*)^c$. The results of $x^0 \neq x^*$ and $(y^0, x^*, w^*) \neq (y^0, x^0, w^*)^c$ are defined as the second type of CAI.

Similar to Equations (19) – (22), comparing the "observed" cost share of input $j$, $s_j^0(x^*, w^*) = \frac{w_j^* x_j^*}{\sum_l w_l^* x_l^*}$, with its optimal cost shares for $(y^0, x^*, w^*)$, the associated cost share inequalities for the two types of CAI based on $(y^0, x^*, w^0)^p$ and $(y^0, x^0, w^*)^c$ are:

$$s_j^0(x^*, w^*) \neq s_j^p(y^0, x^*) \text{ , or } s_j^0(x^*, w^*) \neq s_j^c(y^0, w^0) \text{ for some } j \tag{24}$$

$$s_j^0(x^*, w^*) \neq s_j^c(y^0, w^*), \text{ or } s_j^0(x^*, w^*) \neq s_j^p(y^0, x^0) \text{ for some } j, \tag{25}$$

where $s_j^p(y^0, x^*) = s_j^c(y^0, w^0) = \frac{w_j^0 x_j^*}{\sum_l w_l^0 x_l^*}$ and $s_j^c(y^0, w^*) = s_j^p(y^0, x^0) = \frac{w_j^* x_j^0}{\sum_l w_l^* x_l^0}$. Again, there is no measure of the first type of CAI based on $w^0 \neq w^*$ and $(y^0, x^*, w^0)^p$. But a measure of the second type of CAI based on $(y^0, x^0, w^*)^c$ can be defined as

$$CAI^{w^*} = \ln \frac{TC(x^*, w^*)}{C(y^0, w^*)} = \ln \frac{\sum_j w_j^* x_j^*}{\sum_j w_j^* x_j^0} = \ln \sum_j w_j^* x_j^* - \ln \sum_j w_j^* x_j^0 \geq 0. \tag{26}$$

It is impractical to use this measure of CAI since the observed total cost $TC(x^0, w^0)$ is not involved in the analysis.

Graphical representation of CAI with two inputs is useful to understand the nature and properties of CAI (Kopp and Diewert, 1982; Kumbhakar and Lovell, 2000, p. 160 and 221; Kumbhakar and Wang, 2006b). We use Figure 1 to further clarify the above discussion of the two types of CAI and the two important implications. Given two inputs $x_1$ and $x_2$, denote the observed input quantities and prices as $x^0 = (x_1^0, x_2^0)$ and $w^0 = (w_1^0, w_2^0)$, respectively. Without technical inefficiency, $x^0$ is a point on the isoquant $L(y^0)$.

The isocost line $AA$' with a total cost of $\sum_j w_j^0 x_j^0$ and a slope of $-\frac{w_1^0}{w_2^0}$ passes through $L(y^0)$ at $x^0$. Because the isocost line $AA'$ crosses the isoquant $L(y^0)$ at $x^0$, we can find another isocost line which is tangent to $L(y^0)$ at $x^0$. This isocost line is $BB'$ with a slope of $-\frac{w_1^*}{w_2^*}$, where $(w_1^*, w_2^*)$ are input shadow prices. The cross of *AA*' and *BB*' at $x^0$ implies $\frac{w_1^0}{w_2^0} \neq \frac{w_1^*}{w_2^*}$, which is the first type of CAI.

The second type of CAI is related to points $x^0$ and $x^*$ resulting from the given $w^0$. Given $w^0$, the isocost line *AA*' passes through $L(y^0)$ at $x^0$. Hence, $x^0$ is not the optimal inputs point given $w^0$. The optimal input point given $w^0$ is at $x^*$, where the isocost line *DD*' with the slope of $-\frac{w_1^0}{w_2^0}$ is tangent to $L(y^0)$ at $x^*$. Since *AA*' and *DD*' are parallel, the second type of CAI can be measured by the distance between these two lines. Figure 1 shows there is no measure of the size of the first type of CAI. The two total costs associated with the first type of CAI are: the observed total cost in *AA*' and the optimal total cost given $w^*$ in $BB$'. The difference between these two total costs cannot be used to measure the size of the first type of CAI because the difference can be either positive, negative, or zero; it is positive with OA > OB in terms of $x_2$ and it is negative with OB' > OA' in terms of $x_1$.

With $w^0 \neq w^*$, we demonstrate an alternative measure of CAI involving $w^*$ and $(y^0, x^*, w^*)$ (Equation (26)) in Figure 1. Define another "observed" total cost as $TC(x^*, w^*) = \sum_j w_j^* x_j^*$. The isocost line *EE*' with this total cost and the slope of $-\frac{w_1^*}{w_2^*}$ passes through $L(y^0)$ at $x^*$. For the triplet $(y^0, x^*, w^*)$, the associated first type of CAI is the cross of two isocost lines *DD*' and *EE*' at $x^*$ with two different slopes; the associated second type of CAI is to compare $x^0$ and $x^*$ on $L(y^0)$ with the given $w^*$. *BB*' and *EE*' are parallel and the distance between these two lines is an alternative measure of CAI.

In addition to the measures of CAI, Figure 1 also shows two different sets of input cost share inequalities corresponding with the two types of CAI. For the first type of CAI, we consider the cost share of $x_1$ and compare the observed cost share of $x_1$ with its optimal cost share given $w^*$. At $x^0$, the observed cost share of input $x_1$ is $s_1^0(x^0, w^0) = \frac{w_1^0 x_1}{TC(x^0, w^0)} = \frac{x_1^0}{OA'}$, and its optimal cost share given $w^*$ is calculated according to *BB*' and it is

$s_1^c(y^0, w^*) = \frac{w_1^* x_1^0}{C(y^0, w^*)} = \frac{x_1^0}{OB'}$. Since $OA' < OB'$, it gives $\frac{x_1^0}{OA'} > \frac{x_1^0}{OB'}$. Hence, $s_1^0 \neq s_1^c$ and thereby $s_2^0 \neq s_2^c$. This shows the cost share inequalities for the first type of CAI. For the second type of CAI, we consider the cost share of $x_2$ and compare the observed cost share of $x_2$ with its optimal cost share given $w^0$. The observed cost share of input $x_2$ at $x^0$ is $s_2^0(x^0, w^0) = \frac{w_2^0 x_2^0}{TC(x^0, w^0)} = \frac{x_2^0}{OA}$. Given $w^0$, the optimal input point is at $x^*$, and the optimal cost share of input $x_2$ is $s_2^c(y^0, w^0) = \frac{w_2^0 x_2^0}{C(y^0, w^0)} = \frac{x_2^*}{OD}$. Since $OD < OA$ and $x_2^* > x_2^0$, it gives $\frac{x_2^*}{OD} > \frac{x_2^*}{OA} > \frac{x_2^0}{OA}$. Hence, $s_2^0 \neq s_2^c$ and thereby $s_1^0 \neq s_1^c$. This shows the cost share inequalities for the second type of CAI.

Based on the first-order condition from the production function and the cost function, this section provided both the derivation and the graphical demonstration of the two types of CAI, their associated cost share inequalities, and two measures of CAI when there is no technical inefficiency. Although the first type of CAI is commonly used in current literature, we conclude that the second type of CAI has two advantages over the first type of CAI. Firstly, input cost share inequalities for the second type of CAI do not require the knowledge of input shadow prices. Secondly, and more importantly, the second type of CAI can be estimated by the standard stochastic cost frontier model.

From duality theory, one type of CAI implies the other type of CAI since $w^0 \neq w^*$ implies $x^0 \neq x^0$, and vice versa. In addition to the modelling differences and necessaries as discussed above, there are economic reasons to consider two types of CAI. For the first type of CAI, the firms determine the level of inputs based on $w^*$ instead of $w^0$. The reasons for such a decision can include imperfect input price information and price regulations. When the input price information is imperfect, the firms may perceive market input price as $w^*$ instead of $w^0$. Therefore, $x^0$ is the "optimal" solution to the firms since they only see $w^*$. Comparing the two basic inputs, labor and capital, labor wages are usually known to the firms as the labor market are mostly perfectly competitive, whereas implicit rental price for capital is more complex due to the involvement of risks. With uncertain outcomes in investing physical capitals, the firms' investment decisions are based on the self-projected $w^*$. In the case of price regulations, the regulated input prices $w^*$ can be different from the market input prices $w^0$. Because of price regulations, the firms produce outputs with the input level at $x^0$, which is the inefficient use of inputs given $w^0$. For the second type of CAI, the input prices are given as $w^0$ and there is no misunderstanding of the input price information by the firms. The inefficiency is purely reflected by the misuse of inputs; the misuse may be associated with the constraints in using input resources, such as quantity regulations, and poor managements.

The above analysis on the two types of CAI assumes that there is no technical inefficiency. In the next section, we continue the CAI analysis, but this time, with technical inefficiency. The needs on the use of the two types of CAI and the advantages of using the second type of CAI are even stronger in this case.

## IV. Cost Allocative Inefficiency with Technical Inefficiency

In this section, we extend the previous CAI analysis to the case when technical inefficiency occurs. Assume both production and cost functions are given. We denote the observed output, and input quantities and prices as $(y^0, x^0, w^0)$. When technical inefficiency occurs, $(y^0, x^0)$ will no longer be on the production possibility frontier. For the output-oriented technical inefficiency, $y^* > y^0$, where $y^* = f(x^0)$ is the optimal output. We begin the CAI analysis by isolating technical inefficiency from total cost inefficiency.

We can isolate technical inefficiency by replacing the "output-oriented" inefficient output $y^0$ in the original triplet $(y^0, x^0, w^0)$ with the optimal output $y^*$ for the CAI analysis. Since $(y^*, x^0, w^0)$ does not have technical inefficiency, we follow the same procedure in the previous section to analyze CAI. For the triplet $(y^*, x^0, w^0)$, the two types of CAI are: $w^0 \neq w^*$ and $x^0 \neq x^*$. Here, $w^*$ is a vector of input shadow prices derived from $y^* = f(x^0)$ and $x^*$ is a vector of the optimal inputs derived from $C(y^*, w^0)$, where $\frac{w_j^*}{w_1^*} = \frac{\eta_j^0}{\eta_1^0}\frac{x_1^0}{x_j^0}$, $\eta_j^0 = \frac{\partial f(x^0)}{\partial x_j}\frac{x_j^0}{y^*}$, and $x^* = \frac{\partial C(y^*,w^0)}{\partial w_j}$. The efficient triplets corresponding to the two types of CAI are: the production triplet $(y^*, x^0, w^*)^p$ and the cost triplet $(y^*, x^*, w^0)^c$. Replacing $y^0$ in Equations (19) – (22) with $y^*$, the cost share inequalities are

$$s_j^0(x^0, w^0) \neq s_j^p(y^*, x^0) \text{ or } s_j^0(x^0, w^0) \neq s_j^c(y^*, w^*) \text{ for some } j. \quad (27)$$

$$s_j^0(x^0, w^0) \neq s_j^c(y^*, w^0) \text{ or } s_j^0(x^0, w^0) \neq s_j^p(y^*, x^*), \text{ for some } j. \quad (28)$$

The inequalities in Equation (27) are derived from $w^0 \neq w^*$ and $(y^*, x^0, w^*)^p$; the inequalities in Equation (28) are derived from $x^0 \neq x^*$ and $(y^*, x^*, w^0)^c$. Like Equation (23), the measure of the second type of CAI is

$$CAI = \ln TC(x^0, w^0) - \ln C(y^*, w^0) = \ln \sum_j w_j^0 x_j^0 - \ln \sum_j w_j^0 x_j^*, \tag{29}$$

where both $x^0$ and $x^*$ have the same output level $y^*$, but $x^*$ are the optimal inputs associated with the minimized cost $C(y^*, w^0)$. $CAI > 0$ when the observed inputs $x^0$ differ from the optimal inputs $x^*$ with given $w^0$.

After the derivation of the measure of CAI in Equation (29), we examine the technical inefficiency component (CTE) in total cost inefficiency. Note that the analysis of the CAI is based on $(y^*, x^0, w^0)$. Comparing this triplet with the observed triplet $(y^0, x^0, w^0)$, we derive a measure of CTE. Given $w^0$ and the cost function $C(y, w)$, a measure of CTE can be shown as

$$CTE = \ln C(y^*, w^0) - \ln C(y^0, w^0) = \ln \sum_j w_j^0 x_j^* - \ln \sum_j w_j^0 x_j^{**}, \tag{30}$$

where $x_j^{**} = \frac{\partial C(y^0, w^0)}{\partial w_j}$ is the optimal input derived from $C(y^0, w^0)$. Since the cost function is increasing in output, $CTE > 0$ when $y^* > y^0$.

The measure of CTE in Equation (30) depends on the level of input prices. Take the partial derivative of CTE with respective to $\ln w_j^0$ in Equation (30). It gives

$$\frac{\partial CTE}{\partial \ln w_j^0} = s_j^c(y^*, w^0) - s_j^c(y^0, w^0) = s_j^p(y^*, x^*) - s_j^p(y^0, x^{**}). \tag{31}$$

The second equality is derived from $s_j^c(y^*, w^0) = s_j^p(y^*, x^*)$ and $s_j^c(y^0, w^0) = s_j^p(y^0, x^{**})$, which are the application of the duality for $(y^*, x^*, w^0)^c$ and $(y^0, x^{**}, w^0)^c$, respectively. In a special case where the production function has a fixed output elasticity for all inputs, i.e. $\eta_j(y, x) = \eta_j(y', x')$, for all $j$, the optimal cost share of input $j$ from the production function is a constant and $\frac{\partial CTE}{\partial \ln w_j^0} = 0$. The change in the input price has no impact on the size of the cost inefficiency due to technical inefficiency.

Using CAI from Equation (29) and CTE from Equation (30), the total cost inefficiency defined in Equation (5) has the following decomposition.

$$\begin{aligned} CE &= \ln TC(x^0, w^0) - \ln C(y^0, w^0) \\ &= \ln TC(x^0, w^0) - \ln C(y^*, w^0) + \ln C(y^*, w^0) - \ln C(y^0, w^0) \\ &= \ln \textstyle\sum_j w_j^0 x_j^0 - \ln \sum_j w_j^0 x_j^* + \ln \sum_j w_j^0 x_j^* - \ln \sum_j w_j^0 x_j^{**} \\ &= CAI + CTE. \end{aligned} \tag{32}$$

This shows that total cost inefficiency includes CAI and technical inefficiency. This decomposition is only applicable to the second type of CAI.

We now derive an alternative decomposition of total cost inefficiency with respect to input shadow prices $w^*$. Replacing $x^0$ and $w^0$ in the original triplet $(y^0, x^0, w^0)$ with $x^*$ and $w^*$, the alternative decomposition is based on $(y^0, x^*, w^*)$. Define the alternative total cost inefficiency from this triplet as $CE^{w^*} = \ln \frac{TC(x^*, w^*)}{C(y^0, w^*)}$. With $y^* = f(x^0) = f(x^*) > y^0$, the CAI analysis of this new triplet is based on $(y^*, x^*, w^*)$. Similar to Equations (26) and (30), the alternative measures of CAI and CTE are $CAI^{w^*} = \ln \frac{TC(x^*, w^*)}{C(y^*, w^*)}$ and $CTE^{w^*} = \ln \frac{C(y^*, w^*)}{C(y^0, w^*)}$, respectively; similar to Equation (32), the decomposition of the alternative total cost inefficiency is

$$\begin{aligned} CE^{w^*} &= \ln TC(x^*, w^*) - \ln C(y^0, w^*) \\ &= \ln TC(x^*, w^*) - \ln C(y^*, w^*) + \ln C(y^*, w^*) - \ln C(y^0, w^*) \\ &= \ln \textstyle\sum_j w_j^* x_j^* - \ln \sum_j w_j^* x_j^0 + \ln \sum_j w_j^* x_j^0 - \ln \sum_j w_j^* x_j^{***} \\ &= CAI^{w^*} + CTE^{w^*}, \end{aligned} \tag{33}$$

where $x_j^{***} = \frac{\partial C(y^0, w^*)}{\partial w_j}$.

When technical inefficiency occurs, we derive two forms of cost inefficiency decomposition (Equations (32) and (33)). The decomposition based on the second type of CAI (Equation (32)) has not been examined before in the literature. While the decomposition with input shadow prices $w^*$ has been studied with specific production and cost functions (Schmidt and Lovell, 1979; Kumbhakar, 1997), no general solution has been found. A major difficulty in deriving a decomposition based on input shadow prices is that there is no general measure of the first type of CAI. Although Equation (33) may provide a solution to the decomposition with input shadow prices, the observed total cost is irrelevant in this design. Hence, the use of Equation (33) is only of theoretical interest.

Figure 2 demonstrates the two types of CAI with two inputs $x_1$ and $x_2$ when there is technical inefficiency. In this figure, the observed $x^0$ lies above the isoquant $L(y^0)$, which indicates technical inefficiency. The isoquant containing $x^0$ is $L(y^*)$, $y^* = f(x^0) > y^0$. The analysis of CAI is based on $L(y^*)$ and the triplet $(y^*, x^0, w^0)$. As in Figure 1, $L(y^*)$ and the isocost lines *AA*', *BB*', and *DD*' are used to show the two types of CAI. The cross of two isocost lines *AA*' and *BB*' at point $x^0$ indicates the first type of CAI. *AA*' and *DD*' are parallel and the distance between these two lines measures the second type of CAI. To demonstrate CTE, we compare two isoquant curves: $L(y^0)$ and $L(y^*)$. Based on $w^0$ and $w^*$, two isocost lines are tangent to $L(y^0)$. Given $w^0$, the isocost line *FF*' is tangent to $L(y^0)$ at $x^{**}$ with the slope of $-\frac{w_1^0}{w_2^0}$. Because *AA*', *DD*', and *FF*' are parallel, the measure of CTE is the distance between *DD*' and *FF*', and the decomposition of total cost inefficiency is represented by the distances derived from these three lines. Given $w^*$, the isocost line *GG*' is tangent to $L(y^0)$ at $x^{***}$ with the slope of $-\frac{w_1^*}{w_2^*}$. The isocost line *EE*' with the slope of $-\frac{w_1^*}{w_2^*}$ passes through $L(y^*)$ at $x^*$. *EE*', *BB*', and *GG*' are parallel; the distances from these three lines show the alternative decomposition of total cost inefficiency.

The two types of CAI have significant implications on two elements in analyzing cost inefficiency: input cost share equations and the measure of cost inefficiency. These implications are applied to the cases with and without technical inefficiency. In the following two sections, we examine their practical uses. First, we explain how to use input cost share equations appropriately in empirical estimations when CAI occurs. Second, we derive a three-stage procedure to estimate the two components of total cost inefficiency. An empirical application of this three-stage procedure is also demonstrated.

## V. Use of Cost Share Equations in Estimations

The system of cost and input cost share equations is commonly used for cost analysis both with and without CAI (Christensen and Greene, 1976; Schmidt and Lovell, 1979; Greene, 1980; Bauer, 1990a; Kumbhakar, 1997; Kumbhakar and Lovell, 2000, p. 146; and among others). The reasons to include cost share equations in the system of equations can be summarized as: 1) it provides efficient estimates (Christensen and Greene, 1976), 2) it provides additional constraints to the parameters of the production or cost function under investigation (Berndt and Christensen, 1973), and 3) it provides a method to model CAI (Schmidt and Lovell, 1979). For the system of equations approach, the use of cost share equations is based on $s_j^0 = s_j^c$, where $s_j^0$ is the observed cost share and $s_j^c$ is the optimal cost share derived from the cost function. The main issue in using $s_j^0 = s_j^c$ is that this equality may and may not be true if there is allocative inefficiency (Schmidt and Lovell, 1979; Kumbhakar, 1997; Kumbhakar and Wang, 2006a). We examine this issue using the input cost share inequalities derived from the two types of CAI.

For a given cost function, the optimal cost share of input $j$ is a function of output and input prices. The above cost inefficiency analysis involves two output levels (the observed output $y^0$ and the optimal output $y^*$) and two input prices (the observed input prices $w^0$ and input shadow prices $w^*$). This gives four combinations of outputs and input prices: $(y^*, w^*)$, $(y^*, w^0)$, $(y^0, w^*)$, and $(y^0, w^0)$. The cost share equations from the cost function corresponding to these four combinations are: $s_j^c(y^*, w^*)$, $s_j^c(y^*, w^0)$, $s_j^c(y^0, w^*)$, and $s_j^c(y^0, w^0)$, respectively. Based on duality theory, each cost share equation from the cost function is equal to the cost share equation from the production function (Equation (17)) for an efficient triplet. Consider four efficient triplets: $(y^*, x^0, w^*)^p$, $(y^*, x^*, w^0)^c$, $(y^0, x^{**}, w^*)^c$, and $(y^0, x^{***}, w^0)^c$. It gives four cost share equations from the cost function and four equations from the production function. The observed cost share equation and these eight cost share equations are summarized in Table 1. Taking cost inefficiency into the consideration, the comparison of the observed cost share with each of these four cost share equations gives four possible cost share inequalities: $s_j^0 \neq s_j^c(y^*, w^*)$ in Equation (27), $s_j^0 \neq s_j^c(y^*, w^0)$ in Equation (28), and the following two:

$$s_j^0(x^0, w^0) \neq s_j^c(y^0, w^0) \tag{34}$$

$$s_j^0(x^0, w^0) \neq s_j^c(y^0, w^*). \tag{35}$$

We have proved that the inequalities in Equations (27) and (28) are true when two types of CAI occur; these two inequalities are related only to $y^*$, not $y^0$. To check if the cost share inequality of Equation (34) is true, consider the following inequality derived from cost inefficiency decomposition in Equation (32):

$$TC(x^0, w^0) > C(y^*, w^0) > C(y^0, w^0), \tag{36}$$

where the first inequality is due to the second type of CAI and the second inequality is due to technical inefficiency. An intuitive inference from Equation (36) suggests that total cost inequalities implies cost share inequalities $s_j^0(x^0, w^0) \neq s_j^c(y^*, w^0) \neq s_j^c(y^0, w^0)$. We show this intuitive inference may not be true. The first cost share inequality $s_j^0(x^0, w^0) \neq s_j^c(y^*, w^0)$ for some $j$ is true due to the second type of CAI (Equation (28)). When both CTE and CAI occur, it is possible that

$$s_j^0(x^0, w^0) = s_j^c(y^0, w^0) \text{ for all } j, \text{ or } \frac{w_j^0 x_j^0}{\sum_l w_l^0 x_l^0} = \frac{w_j^0 x_j^{**}}{\sum_l w_l^0 x_l^{**}} \text{ for all } j, \tag{37}$$

where $s_j^c(y^0, w^0) = \frac{w_j^0 x_j^{**}}{\sum_l w_l^0 x_l^{**}}$. For the denominators in Equation (37), Equation (36) implies that $\sum_l w_l^0 x_l^0 > \sum_l w_l^0 x_l^{**}$. The equality in Equation (37) can be true if $x_j^0 \geq x_j^{**}$ for all $j$, which may occur when there is technical inefficiency such that $y^* = f(x^0) > y^0 = f(x^{**})$. Therefore, the inequality in Equation (34) may not hold even if CAI occurs; the addition of technical inefficiency contributes to the equality. The alternative decomposition of cost inefficiency in Equation (33) gives

$$TC(x^*, w^*) > C(y^*, w^*) > C(y^0, w^*). \tag{38}$$

Similar to the proof of Equation (37), $s_j^0(x^*, w^*) = s_j^c(y^0, w^*)$ can be true. If $s_j^0(x^*, w^*) = s_j^0(x^0, w^0)$, then $s_j^0(x^0, w^0) = s_j^c(y^0, w^*)$ can be true. The inequality in Equation (35) may not hold as well even if CAI occurs.

Because Equations (34) and (35) may or may not be true, there is no definite conclusion on how to use cost share equations correctly when technical inefficiency is added to total cost inefficiency and the observed output $y^0$ is used in cost share equations. However, Equations (27) and (28) show that there are two sets of cost share inequalities when CAI occurs. The misallocation of inputs can be identified with these two inequalities.

Kumbhakar and Wang (2006a) used simulations to demonstrate that traditional estimations of the stochastic cost frontier model are biased if CAI is ignored explicitly in specifying the cost function. In the case without technical inefficiency, we use Equations (20) and (21) to show how the bias issue occurs if the cost share equations from the first-order condition are incorporated in estimating the cost function. Consider a translog cost function, shown as follows (Christensen and Greene, 1976),

$$\begin{aligned}\ln C = \delta_0 + \gamma_y \ln y + \tfrac{1}{2}\gamma_{yy}(\ln y)^2 + \textstyle\sum_j \alpha_j \ln w_j + \tfrac{1}{2}\sum_j \sum_l \gamma_{jl} \ln w_j \ln w_l \\ + \textstyle\sum_j \gamma_{yj} \ln w_j \ln y.\end{aligned} \tag{39}$$

From $\frac{\partial \ln C}{\partial \ln w_j \partial \ln w_l} = \frac{\partial \ln C}{\partial \ln w_l \partial \ln w_j}$, $\gamma_{jl} = \gamma_{lj}$. Because the cost function is homogenous of degree one in input prices,

$$\textstyle\sum_j \alpha_j = 1, \sum_j \gamma_{jl} = \sum_l \gamma_{lj} = \sum_j \sum_l \gamma_{jl} = 0, \text{ and } \sum_j \gamma_{yj} = 0. \tag{40}$$

The optimal cost share of input $j$ is

$$s_j^c = \frac{\partial \ln C}{\partial \ln w_j} = \alpha_j + \sum_l \gamma_{jl} \ln w_l + \gamma_{yj} \ln y. \tag{41}$$

Take the sum for all $j$ to get total cost share of one, and

$$\sum_j \alpha_j + \sum_l \sum_j \gamma_{jl} \ln w_l + \sum_j \gamma_{yj} \ln y = 1. \tag{42}$$

Based on the equality of the coefficients on both sides of the equation, $\sum_j \alpha_j = 1$, $\sum_j \gamma_{jl} = 0$ for all $l$, and $\sum_j \gamma_{yj} = 0$. These constraints are the same as the constraints for the homogenous of degree one in input prices (Equation (40)).

Denote the observed data $(y^0, x^0, w^0)$ for firm $i$ as $(y_i, x_i, w_i)$, $i = 1,2, \cdots n$. Each $x_i$ and $w_i$ is a $k \times 1$ vector with the components of $x_{ij}$ and $w_{ij}$ as the observed input quantity and price for input $j$ and firm $i$, $j = 1,2, \cdots k$, respectively. The system of equations for estimation can be represented as

$$\ln TC_i = \delta_0 + \gamma_y \ln y_i + \frac{1}{2}\gamma_{yy} \ln y_i^2 + \sum_j \alpha_j \ln w_{ij} + \frac{1}{2}\sum_j \sum_l \gamma_{jl} \ln w_{ij} \ln w_{il} + \sum_j \gamma_{yj} \ln w_{ij} \ln y_i + v_{ci} + u_{ci}. \tag{43}$$

$$s_{ij}^0 = \alpha_j^c + \sum_l \gamma_{jl}^c \ln w_{il} + \gamma_{yj}^c \ln y_i + \varepsilon_{ij}^c, \tag{44}$$

where $v_{ci}$ and $\varepsilon_{ij}^c$ are the random error terms for noise, and $u_{ci}$ is a random error term with a positive mean for inefficiency. $TC_i = \sum_l w_{il} x_{il}$ is the observed total cost for firm $i$ and $s_{ij}^0$ is the observed cost share of input $j$ for firm $i$. The coefficients $\alpha_j^c$, $\gamma_{jl}^c$, and $\gamma_{yj}^c$ in Equation (44) are different from those $\alpha$s and $\gamma$s in Equation (43). The sum of all input cost shares for each firm in Equation (44) is one. Similar to Equation (42), it implies $\sum_j \alpha_j^c = 1$, $\sum_j \gamma_{jl}^c = 0$, $\sum_j \gamma_{yj}^c = 0$, and $\sum_j \varepsilon_{ij}^c = 0$. Because $\sum_j \varepsilon_{ij}^c = 0$, one of the cost share equations in Equation (44) is dropped from the system of equations. For the traditional estimation of the system of equations (Equations (43) and (44)), the constraints $\alpha_j^c = \alpha_j$, $\gamma_{jl}^c = \gamma_{jl}$, and $\gamma_{yj}^c = \gamma_{yj}$ are imposed based on the optimal cost share equations (Equation (41)). If the second type of CAI occurs, then $s_j^0(x^0, w^0) \neq s_j^c(y^0, w^0)$ from Equation (21), and $\alpha_j^c \neq \alpha_j$, $\gamma_{jl}^c \neq \gamma_{jl}$, or $\gamma_{yj}^c \neq \gamma_{yj}$ for some $j$. The constraints used in the system of equations cause a bias issue, which exists either with or without the stochastic inefficient term $u_{ci}$.

This bias issue also exists in the estimation when the first type of CAI, where $w^0 \neq w^*$, is considered. For the first type of CAI, the observed input prices $w$s in Equations (43) and (44) are replaced by input shadow prices $w^*$ (Schmidt and Lovell, 1979; Kumbhakar, 1997; Kumbhakar and Tsionas, 2005a, 2005b). From Equation (20), $s_j^0(x^0, w^0) \neq s_j^c(y^0, w^*)$; the constraints of $\alpha_j^c = \alpha_j$, $\gamma_{jl}^c = \gamma_{jl}$, and $\gamma_{yj}^c = \gamma_{yj}$ imposed on Equations (43) and (44) give biased estimates as well. The current literature on CAI solves this bias issue by including CAI in $\varepsilon_{ij}^c$ and $u_{ci}$, such that $\varepsilon_{ij}^c$ and $u_{ci}$ are correlated. However, this solution creates a fundamental issue where CAI cannot be differentiated from CTE in $u_{ci}$. Another and more serious problem with the use of input shadow prices in the stochastic cost frontier model is that the mean of $u_{ci}$ may not be non-negative since the observed total cost may not be greater than the optimal cost $C(y^0, w^*)$ for the first type of CAI.

We conclude that the approach to include input cost share equations in the system of equations has some estimation issues. When CAI occurs without technical inefficiency, cost share inequalities cause biased estimation of the system of equations unless CAI is explicitly specified in the system. When technical inefficiency is added to total cost inefficiency, there is no conclusion on how to use cost share equations correctly. In either case with or without technical inefficiency, there is no general formula to link cost share equations with the cost function in estimation. It is therefore more appropriate to abandon the system of equations and remove cost share equations from the system in estimating cost inefficiency. In the next section, we show how to estimate technical and allocative inefficiency with a three-stage procedure and how to use cost share inequalities in this setting.

## VI. Three-Stage Procedure and Empirical Examples

Based on the analysis of the two types of CAI and the total cost inefficiency decomposition, a procedure to estimate technical, allocative inefficiency, and total cost inefficiency can be derived. Since the estimation of the production function only involves technical inefficiency, the production function can be estimated using traditional stochastic frontier methods (Equation (3)) and the data $(y_i, x_i)$ for firm $i$, $i = 1, \cdots n$. The estimation results give the estimated parameters $\hat{\beta}$, the estimated optimal output $\hat{y}_i^* = f(x_i; \hat{\beta})$, and the estimated technical inefficiency based on $\hat{u}_{pi}$.

With the estimated production function, the first type of CAI can be checked if the data of either input prices or input cost shares are available. When input prices $w^0$ are given, a simple check of the first type of CAI is testing $w^0 = w^*$, where $w^*$ is input shadow prices derived from $\hat{y}_i^* = f(x_i; \hat{\beta})$ and Equation (15). When the data of input cost shares are given, an alternative check is to test the following hypothesis based on Equation (27).

$$E\left(s_j^0\right) = E\left(s_j^p\right), \tag{45}$$

where $E(s_j^0)$ is the mean observed cost share of input $j$ and $E\left(s_j^p\right)$ is the mean optimal cost share of input $j$ from the production function. $E(s_j^0)$ is estimated by $(\sum_i s_{ij}^0)/n$ and $E\left(s_j^p\right)$ is estimated by $\left(\sum_i \hat{s}_{ij}^p\right)/n$, where $s_{ij}^0$ is the observed cost share of input $j$ for firm $i$, $\hat{s}_{ij}^p = \frac{\hat{\eta}_{ij}}{\hat{\eta}_i}$, $\hat{\eta}_{ij} = \frac{\partial f(x_i;\hat{\beta})}{\partial x_{ij}} \frac{x_{ij}}{\hat{y}_i^*}$, and $\hat{\eta}_i = \sum_j \hat{\eta}_{ij}$. When the hypothesis in Equation (45) is rejected for some inputs, there is an evidence of the first type of CAI. However, there is no measure of the first type of CAI.

To measure and check the second type of CAI, the data of input prices are required to estimate the cost function and CAI. With the data $(\hat{y}_i^*, x_i, w_i)$ for firm $i$, $i = 1, \cdots n$, the following stochastic cost frontier model is used.

$$\ln TC(x_i, w_i) = \ln C(\hat{y}_i^*, w_i; \gamma) + v_{ci} + u_{ci}. \tag{46}$$

This model is the same as Equation (6), except that the data $(\hat{y}_i^*, x_i, w_i)$ instead of $(y_i, x_i, w_i)$ are used in this estimation, where $\hat{y}_i^*$ is the estimated optimal output from the first stage. The estimation gives the estimated cost function $C(y, w; \hat{\gamma})$ and $\hat{u}_{ci}$, which is the estimate of the second type of CAI according to Equation (29). Using Equation (28), we check the second type of CAI by testing the following hypothesis.

$$E\left(s_j^0\right) = E\left(s_j^c\right), \tag{47}$$

where $E\left(s_j^c\right)$ is the mean optimal cost share of input $j$. $E\left(s_j^c\right)$ is estimated by $\left(\sum_i \hat{s}_{ij}^c\right)/n$, where $\hat{s}_{ij}^c = \frac{w_{ij}^0 \hat{x}_{ij}^*}{\sum_l w_{il}^0 \hat{x}_{il}^*}$ and $\hat{x}_{ij}^* = \frac{\partial C(\hat{y}_i^*, w_i; \hat{\gamma})}{\partial w_{ij}}$. Because the observed cost share is not equal to the optimal cost share for some inputs when CAI occurs, the system of equations of cost share equations and the stochastic production or cost frontier model are not used. Finally, using the estimated cost function $C(y, w; \hat{\gamma})$, the total cost inefficiency for firm $i$ is estimated by $CE_i = TC(y_i, w_i) - C(y_i, w_i; \hat{\gamma})$ (Equation (5)); the cost inefficiency due to technical and allocative inefficiency are estimated by $CTE_i = TC(\hat{y}_i^*, w_i) - C(y_i, w_i; \hat{\gamma})$ (Equation (30)) and $CAI_i = TC(y_i, w_i) - C(\hat{y}_i^*, w_i; \hat{\gamma})$ (Equation (29)), respectively. These three terms give the cost inefficiency decomposition for firm $i$, $CE_i = CAI_i + CTE_i$.

*Three-Stage Procedure*

The above procedure for estimating the two components of total cost inefficiency and the testing for the two types of CAI can be summarized in a three-stage procedure.

Stage 1: Estimate the production function and test the first type of CAI. Use the stochastic production frontier function (Equation (3)) with the data values of $(y_i, x_i)$ to estimate the production function $y = f(x; \beta)$ and derive the estimated optimal output $\hat{y}_i^* = f(x_i; \hat{\beta})$. Test the first type of CAI with Equation (45).

Stage 2: Estimate the cost function and the second type of CAI. Use the stochastic cost frontier function (Equation (46)) with the data values of the total observed cost $TC(x_i, w_i)$ and $(\hat{y}_i^*, w_i)$ to estimate the cost function $C(y, w; \gamma)$ and the second type of CAI. Test the second type of CAI with Equation (47).

Stage 3: Derive the cost inefficiency decomposition. Use the estimated cost function from Stage 2 and the data values of $TC(x_i, w_i)$, $(\hat{y}_i^*, w_i)$ and $(y_i, w_i)$ to compute total cost inefficiency, and the cost inefficiency due to technical and allocative inefficiency using Equations (5), (30), and (29), respectively.

The first two stages in the above is the two-stage least squares estimation (TSLS) in the simultaneous equations model. Consider the simultaneous equations with production and cost functions and their three variables $(y, x, w)$. Suppose both $y$ and total cost are the endogenous variables, and $x$ and $w$ are the exogenous variables. The first stage estimation of the production function involves $y$ as the endogenous variable and $x$ as the exogenous variable. The fitted value of $y$ from the first stage estimation is used as one of the explanatory variables for the second stage estimation of the cost function. This procedure is the TSLS estimation. Based on duality theory, there is no need for a system of equations with production and cost functions since the information from one function gives the exact same information from the other function. In the case of CAI, it is necessary to consider two separate regressions for two functions and for each regression to include its inefficiency term. The first two stages combine production and cost functions into a system of simultaneous equations. This is more general than treating $y$ as an exogenous variable in the traditional cost function estimation (Christensen and Greene, 1976).

This three-stage procedure is different from the current CAI analysis where the system of the cost function and cost share equations are estimated (Kumbhakar, 1997; Kumbhakar and Tsionas, 2005b). A cost system model specifies CAI in the cost share equations and combines both technical and allocative inefficiency in the inefficiency term of the stochastic cost frontier model. This approach faces the "Greene problem." However, for the second type of CAI, CAI is the only inefficiency term in the stochastic cost frontier model in the three-stage procedure. With the separation of technical inefficiency and CAI in two separate stochastic production and cost frontier models, the three-stage procedure provides a solution to identify technical and allocative inefficiency of cost inefficiency.

*Empirical Examples*

We use the translog production and cost functions to demonstrate the empirics of the three-stage procedure, where the two components of total cost inefficiency are estimated, and the two types of CAI are tested. In the first stage, the following translog production function is considered (Berndt and Christensen, 1973).

$$\ln y = c_0 + \sum_j a_j \ln x_j + \frac{1}{2}\sum_j \sum_l b_{jl} \ln x_j \ln x_l. \tag{48}$$

Derive $\eta_j = \frac{\partial \ln y}{\partial \ln x_j} = a_j + \sum_l b_{jl} \ln x_l$ and $\eta = \sum_j \eta_j$. From Equation (16), $s_j^p = \frac{\eta_j}{\eta}$, the optimal cost share equation for input $j$ is

$$s_j^p = \frac{a_j}{\eta} + \sum_l \frac{b_{jl}}{\eta} \ln x_l. \tag{49}$$

Using $\sum_j s_j^p = 1$ and the symmetry of cross partial derivatives in Equation (48), the constraints for the production function are:

$$\sum_j a_j = \eta, \sum_j b_{jl} = 0, \text{ and } b_{jl} = b_{lj}. \tag{50}$$

With the data $(y_i, x_i)$, the stochastic production frontier model based on Equation (48) is

$$\ln y_i = c_0 + \sum_j a_j \ln x_{ij} + \frac{1}{2}\sum_j \sum_l b_{jl} \ln x_{ij} \ln x_{il} + v_{pi} - u_{pi}. \tag{51}$$

When there is no technical inefficiency and no cost inefficiency, a system of equations with Equation (51) without $u_{pi}$ and the following cost share equation can be considered.

$$s_{ij}^0 = a_j^p + \sum_l b_{jl}^p \ln x_{il} + \varepsilon_{ij}^p. \tag{52}$$

where the coefficients $a_j^p$ and $b_{jl}^p$ are different from those $a$s and $b$s in Equation (51) and $\varepsilon_{ij}^p$ is the random error term. Comparing Equations (49) and (52), $a_j^p = \frac{a_j}{\eta}$, and $b_{jl}^p = \frac{b_{jl}}{\eta}$ if $s_j^0 = s_j^p$. Hence, the constraints in estimating the system of Equations (51) and (52) include Equation (50), $a_j^p = \frac{a_j}{\eta}$, and $b_{jl}^p = \frac{b_{jl}}{\eta}$. These constraints are difficult to implement in estimation since $\eta$ is a function of $y$ and $x$ and the nonlinear constraints $a_j^p = \frac{a_j}{\eta}$ and $b_{jl}^p =$

$\frac{b_{jl}}{\eta}$ are involved. Assume that $\eta$ is a constant parameter. The estimation with this constant parameter and the nonlinear constraints can be carried out using an iterative procedure. The initial estimates of $\hat{a}_j$ and $\hat{\eta} = \sum_j \hat{a}_j$ is derived from an initial estimation of Equation (51) without cost share equations. Then impose all constraints and with $\hat{\eta}$ as a constant in the first iteration of the system of equations. We then use the estimated coefficients $\hat{a}_j$ from this iteration to derive $\hat{\eta} = \sum_j \hat{a}_j$ for the constraints in the next iteration of the system of equations, and continue the iterations until $\hat{\eta}$ converges. Kim (1992) shows an alternative procedure to estimate variable returns to scale $\eta(y, x)$.

When technical and allocative inefficiency occur, the first stage procedure begins with the stochastic production frontier model of Equation (51) without Equation (52) since the estimation of a system of equations with cost share equations is biased when CAI occurs. This first stage estimation gives the estimated optimal output $\hat{y}_i^*$ and the estimated technical inefficiency $\hat{u}_{pi}$. With the estimated production function, the first type of CAI can be checked using Equation (45), where the mean optimal cost share of input $j$ is estimated by $(\sum_i \hat{s}_{ij}^p)/n$, and $\hat{s}_{ij}^p = \frac{\hat{a}_j}{\hat{\eta}} + \sum_l \frac{\hat{b}_{jl}}{\hat{\eta}} \ln x_{il}$.

In Section V, we use a translog cost function to demonstrate the estimation of the stochastic cost frontier model without technical inefficiency. The same procedure is used here for the second stage of the empirical estimation. The second stage estimation begins with the translog stochastic cost frontier model of Equation (43) by replacing $y^0$ in this equation with $\hat{y}^*$ from the first stage estimation. Because the estimation of a system of Equations (43) and (44) is biased when CAI occurs, the cost share equations is not included in the estimation. The constraints in Equation (40) are imposed in the estimation of the stochastic cost frontier equation. The second type of CAI is estimated by $\hat{u}_{ct}$ from the stochastic cost frontier model. Using the estimated cost function, we derive the estimated optimal input cost shares with Equations (41) and conduct the test of the second type of CAI with Equation (47). The economies of scale is estimated with

$$\hat{\eta} = \hat{\theta}^{-1}, \hat{\theta} = \frac{\partial \ln C}{\partial \ln y} = \hat{\gamma}_y + \hat{\gamma}_{yy} \ln \hat{y}^*. \tag{53}$$

In the third stage, the measures of total cost inefficiency, technical and allocative inefficiency are computed using the estimated cost function and Equations (5), (30), and (29), respectively.

Two datasets from Greene (2018) are used in an empirical application of the three-stage procedure with translog production and cost functions. The first dataset is Greene's (2018) Table F6.2, which includes the data of 159 U.S. firms producing electric power in 1955. A subset of the dataset was originally used by Nerlove (1963). The second dataset is Greene's (2018) Table F4.4, which includes the data of 158 firms in 1970. A subset of the dataset was originally used by Christensen and Greene (1976). The datasets include total output, total cost, and input quantities, prices, cost shares of labor, fuel, and capital for each firm.

Table 2 shows the estimated optimal input cost shares of labor, fuel, and capital ($\hat{s}_{labor}$, $\hat{s}_{fuel}$, and $\hat{s}_{capital}$, respectively), the inefficiency estimates, and the economies of scale $\hat{\eta}$ for the standard and the three-stage procedures. For the standard procedure, column (2) shows the results from the seemly unrelated regression estimates (SURE) for the cost function, where the system of Equations (43) and (44) is estimated without the inefficiency term $u_{ci}$; column (3) shows the results from the standard stochastic cost frontier model (Equation (43)). The estimated optimal cost shares of labor, fuel, and capital from the SURE in column (2) should be very close to the observed cost shares in column (1) due to the constrains in the system of equations. The estimated optimal cost shares in column (3) may be biased due to the omission of CAI in specifying the cost function.

For the three-stage procedure in columns (4) – (7), the estimates from the stochastic production frontier model in column (5) are compared with the SURE from the system of equations for production function in column (4), where the SURE is based on Equations (51) and (52) without the inefficiency term $u_{pi}$. The estimations for the SURE for the cost function in column (6) and stochastic cost frontier model in column (7) correspond to those estimations in the standard procedure in columns (2) and (3); the difference is that the standard procedure uses the observed data $(y_i, x_i, w_i)$ and the three-stage procedure uses the data $(\hat{y}_i^*, x_i, w_i)$, where $\hat{y}_i^*$ is derived from the stochastic production frontier model in the first stage. The estimations of the stochastic frontier models in columns (3), (5), and (7) do not include cost share equations to avoid bias. The estimated optimal cost shares from the SURE in columns (4) and (6) should be close to the observed cost shares as in column (2). When CAI occurs, the optimal cost shares from the production function in column (5) and the optimal cost shares from the cost function in column (7) should be different from the observed cost shares in column (1) and they are different from each other.

For the 1955 data, the observed cost shares of labor, fuel, and capital are 10.57%, 46.78%, and 42.68%, respectively. The estimated cost shares for three SURE estimations in columns (2), (4), and (6) are very close to the observed cost shares. The estimated optimal cost shares of labor, fuel, and capital from the production function in column (5) are -20.56%, 75.42%, and 45.14%, respectively; these shares from the cost function in column (7) are 12.57%, 62.77%, and 24.66%, respectively. The negative optimal labor share from the production function is a concern. This implies that the first type of CAI with the use of input shadow prices may give unrealistic optimal cost shares. Hence, our analysis focuses on the optimal cost shares from the cost function. Applying Equation (47), the $t$-statistics in comparing the differences between the three observed input cost shares in column (1) and their corresponding optimal cost shares from the cost function in column (7) are 0.92, -8.80, and 5.68, respectively. The use of labor is close to the optimal level. The fuel share should be higher, and the capital share should be lower to be optimal.

As for the 1970 data, the observed cost shares of labor, fuel, and capital are 13.9%, 63.24%, and 22.64%, respectively, whereas column (7) shows their estimated optimal cost shares from the cost function are 13.98%, 76.07%, and 9.95%, respectively. The corresponding $t$-statistics used to compare the differences in the observed cost shares and their optimal cost shares are -0.135, -23.88, and 16.39, respectively. The results show the observed labor share is close to the optimal level. The fuel share should be higher, and the capital share should be lower to be optimal.

There are two additional observations from these estimates of optimal cost shares. The optimal cost shares from the cost function in column (7) for the 1955 data are close to the observed cost shares in column (1) for the 1970 data. This implies that the firms adjusted their optimal use of inputs over the years. The estimated cost shares from the standard stochastic cost frontier model in column (3) for the 1955 data are 38.89%, 52.13%, and 8.98%. They are very different from their three-stage estimates in column (7). For the 1970 data, the estimated cost shares from the standard stochastic cost frontier model in column (3) are 15.7%, 72.48%, and 11.82%, which are close to their corresponding three-stage estimates in column (7). The large differences of these estimates for the 1955 data and the small differences for the 1970 data may be related to their CAI sizes; a bigger size of CAI may give a larger biased estimation in the standard procedure. The CAI sizes for the two data sets are measured in the following.

The last two rows in Table 2 shows the inefficiency estimates $E(\exp\hat{u}\,|\varepsilon)$ for the stochastic frontier models and the estimates of economies of scale. In 1955, $E(\exp\hat{u}\,|\varepsilon)$ for the standard stochastic cost frontier model in column (3) is 1.446, which includes both technical and cost allocative inefficiency and may be biased. For the three-stage procedure, column (5) shows the estimate of technical inefficiency from the stochastic production frontier model is 0.775, and column (7) shows the estimate of CAI from the cost function is 1.308. In 1970, these three inefficiency estimates in columns (3), (5), and (7) are 1.108, 0.913, and 1.065, respectively. They are close to one whereas those values in 1955 are not. This shows the improvement of both technical and allocative inefficiency.

When CAI occurs, the estimate of economies of scale from the production function should be different from the estimate from the cost function. For the 1955 data, columns (5) and (7) show the estimates from the production and cost functions are 1.092 and 1.122, respectively; they are 1.068 and 1.107 in 1970. The results show increasing return to scale in both years and the estimates from the cost function are higher than those from the production function.

Table 3 shows the measures of cost inefficiency and its two components in two sample periods. We use the estimated cost function from the second stage to estimate the two costs: $C(\hat{y}^*, w^0)$ and $C(y^0, w^0)$, which are the minimized costs based on the estimated optimal output $\hat{y}^*$ and the observed output $y^0$, respectively. For the 1955 data, the mean of the observed total costs of all firms is 14.72, and the means of $C(\hat{y}^*, w^0)$ and $C(y^0, w^0)$ are 13.03 and 10.04, respectively. The mean of total cost inefficiency, $TC - C(y^0, w^0)$, is 4.67; the mean cost inefficiency due to CAI, $TC - C(\hat{y}^*, w^0)$, is 1.69; and the mean cost inefficiency due to technical inefficiency, $C(\hat{y}^*, w^0) - C(y^0, w^0)$, is 2.98. The percentages of total inefficiency, CAI, and technical inefficiency are $4.67/14.72 = 31.73\%$, $1.69/14.72 = 11.48\%$, and $2.98/14.72 = 20.24\%$, respectively. For the 1970 data, the mean of observed total costs is 53.27; the mean minimized costs based on $\hat{y}^*$ and

$y^0$ are 49.9 and 46.55, respectively. The mean of total cost inefficiency is 6.72, with 3.37 and 3.35 due to CAI and technical inefficiency, respectively. The percentages of total inefficiency, CAI, and technical inefficiency are 12.61%, 6.33%, and 6.29%, respectively. The cost inefficiency decreased significantly from 1955 to 1970, and most of the efficiency improvements came from technical efficiency.

## VII. Conclusions

Cost inefficiency is comprised of two sources: technical inefficiency and allocative inefficiency. When a flexible cost functional form is assumed, it is difficult to estimate these two components when the analysis of allocative inefficiency focuses only on input shadow prices. In this paper, a solution of the decomposition of total cost inefficiency for general cases was found.

Based on duality of production and cost functions in the cost minimization problem, the first-order condition can be examined separately from the production function and the cost function sides. From the perspective of the production function, relative input shadow prices and optimal cost shares are derived from the production function with given quantities of output and inputs. From the perspective of the cost function, the optimal inputs and their cost shares are derived from the cost function with given output quantity and input prices. Two types of allocative inefficiency can occur when: 1) input shadow prices from the production function are not the same as the observed input prices; 2) the optimal inputs from the cost function are not the same as the observed inputs. Based on the two types of allocative inefficiency, we derive their corresponding cost share inequalities and two forms of cost inefficiency decomposition. We show that only the second type of allocative inefficiency is measurable, and the decomposition of cost inefficiency with this type of allocative inefficiency can be studied empirically.

This paper proposes a three-stage procedure to estimate the two components of cost inefficiency. In the first stage, the production function and technical inefficiency are estimated using traditional stochastic production frontier model. In the second stage, the cost function and the second type of allocative inefficiency are estimated using the stochastic cost frontier model with the data values of the observed output replaced by the estimated optimal output from the first stage. Once the estimated cost function is derived, the third stage estimates the total cost inefficiency and the cost inefficiencies due to technical and allocative inefficiency. In this three-stage procedure, the stochastic production and cost functions form a system of equations in the first two stages, where technical and allocative inefficiency are specified and estimated separately in two stochastic frontier models. Input cost shares equations are not incorporated into a system of equations with the production or cost function; instead, cost share inequalities are used to determine allocative efficiency of each input.

The mathematic derivations in this paper are not based on specific functional forms of production and cost functions. In this analysis, the production function can differ from a translog function and can have non-constant returns to scale; the cost function can be any nonlinear functions with feasible solutions for the cost minimization problem. Thus, a new methodology to analyze cost inefficiency and measure the two components of total cost inefficiency has been provided.

**References**


Aigner, D., Lovell, C. A. K., & Schmidt, K. P., (1977). Formulation and Estimation of Stochastic Frontier Function Models. *Journal of Econometrics* 6: 21-37.

Battese, G. E. & Coelli, T. J., (1988). Prediction of Firm-Level Technical Efficiencies with a Generalised Frontier Production Function and Panel Data. *Journal of Econometrics* 38: 387-399.

Battese, G. E., & Coelli, T. J., (1992). Frontier Production Functions, Technical Efficiency and Panel Data: with Application to Paddy Farmers in India. *Journal of Productivity Analysis* 3(1): 153-169.

Bauer, P. W. (1990a). Recent Development in the Econometric Estimation of Frontiers. *Journal of Econometrics* 46 (2): 39-56.

Bauer, P. W. (1990b). Decomposing TFP Growth in the Presence of Cost Inefficiency, Nonconstant Returns to Scale, and Technological Progress, Journal of Productivity Analysis 1 (4): 287-299.

Berndt, E. R. & Christensen, L. R. (1973) The Translog Function and the Substitution of Equipment, Structures, and Labor in U.S. Manufacturing 1929 – 68, *Journal of Econometric* 1(1): 81-114.

Christensen, L. R. & Greene, W. H., (1976) Economies of Scale in U.S. Electric Power Generation. *The Journal of Political Economy* 84 (4): 655-676.

Farrell, M. J., (1957). The Measurement of Productivity Efficiency. *Journal of the Royal Statistical Society*, Series A, General 120(3): 253-281.

Greene, W. H., (1980). On the Estimation of a Flexible Frontier Production Model. *Journal of Econometrics* 13 (1): 101-115.

Greene, W. H., (2005). Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model. *Journal of Econometrics* 126: 269-303.

Greene, W. H. (2018). *Econometric Analysis*, 8th Edition, New York, New York: Pearson.

Hanoch, G., (1975). The Elasticity of Scale and the Shape of Average Costs. *The American Economic Review* 65 (3): 492-497.

Kim, H. Y., (1992). The Translog Production Function and Variable Returns to Scale. *The Review of Economics and Statistics* 74 (3): 546-552.

Kopp, R. J., Diewert, W. E., (1982) The Decomposition of Frontier Cost Function Deviations into Measures of Technical and Allocative Efficiency. *Journal of Econometrics* 19 (2): 317-331.

Kumbhakar, S. C., (1997). Modeling Allocative Inefficiency in a Translog Cost Function and Cost Share Equations: An Exact Relationship. *Journal of Econometrics* 76 (1/2): 351-356.

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

Kumbhakar, S. C. & Tsionas, E. G., (2005a). The Joint Measurement of Technical and Allocative Inefficiencies: An Application of Bayesian Inference in Nonlinear Random-Effects. *Journal of the American Statistical Association* 100 (471): 736-747.

Kumbhakar, S. C. & Tsionas, E. G., (2005b). Measuring Technical and Allocative Inefficiency in the Translog Cost System: a Bayesian Approach. *Journal of Econometrics* 126: 355–384.

Kumbhakar, S. C., & Wang, H.-J., (2006a). Pitfalls in the Estimation of a Cost Function that Ignores Allocative Inefficiency: A Monte Carlo analysis. *Journal of Econometrics* 134 (2): 317–340.

Kumbhakar, S. C., & Wang, H.-J., (2006b). Estimation of Technical and Allocative Inefficiency: A Primal System Approach. *Journal of Econometrics* 134 (2): 419–440.

Li, K.-W., & Liu, T., (2011). Economic and Productivity Growth Decomposition: An Application to Post-reform China. *Economic Modelling* 28 (1-2): 366-373.

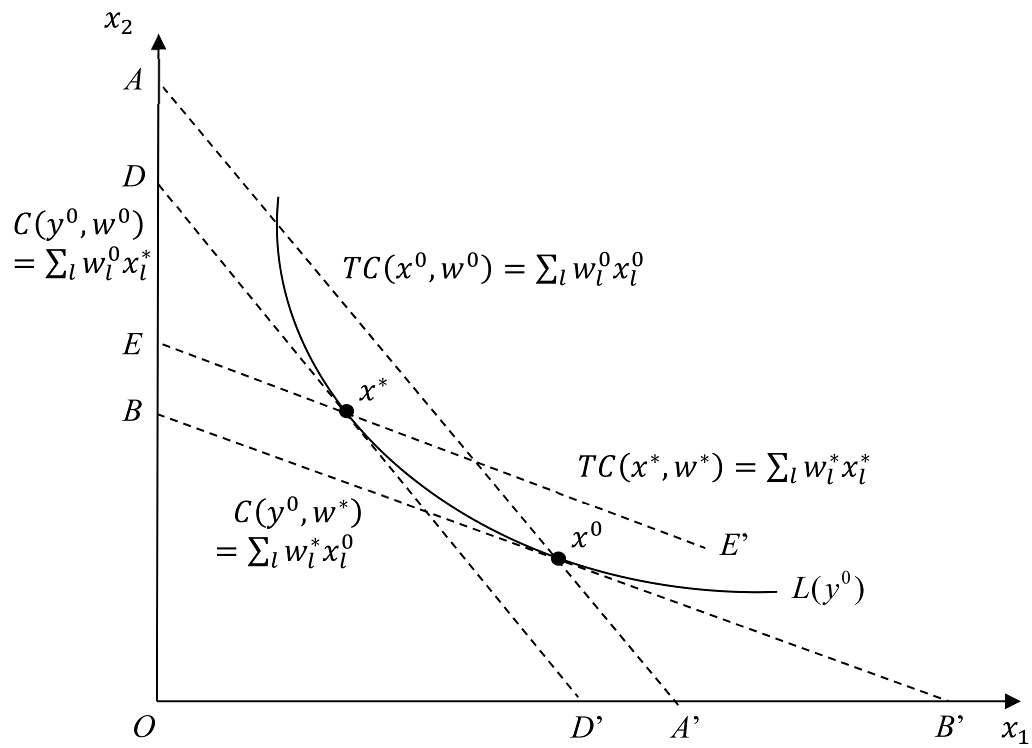
Liu, T., & Li, K.-W., (2008). Revisiting Solow's Decomposition of Economic and Productivity Growth. Ball State University, Working Paper, ECWP200805.

McFadden, D. (1978). "Cost, Revenue and Profit Functions", pp. 3-109 in *Production Economics: A Dual Approach*, Volume 1, M. Fuss and D. McFadden (eds.), Amsterdam: North-Holland.

Nerlove, M. (1963). Returns to Scale in Electricity Supply. *Measurement in Economics-Studies in Mathematical Economics and Econometrics in Memory of Yenhuda Grunfeld*, by Carl F. Christ (ed.). Stanford, California: Stanford University Press.

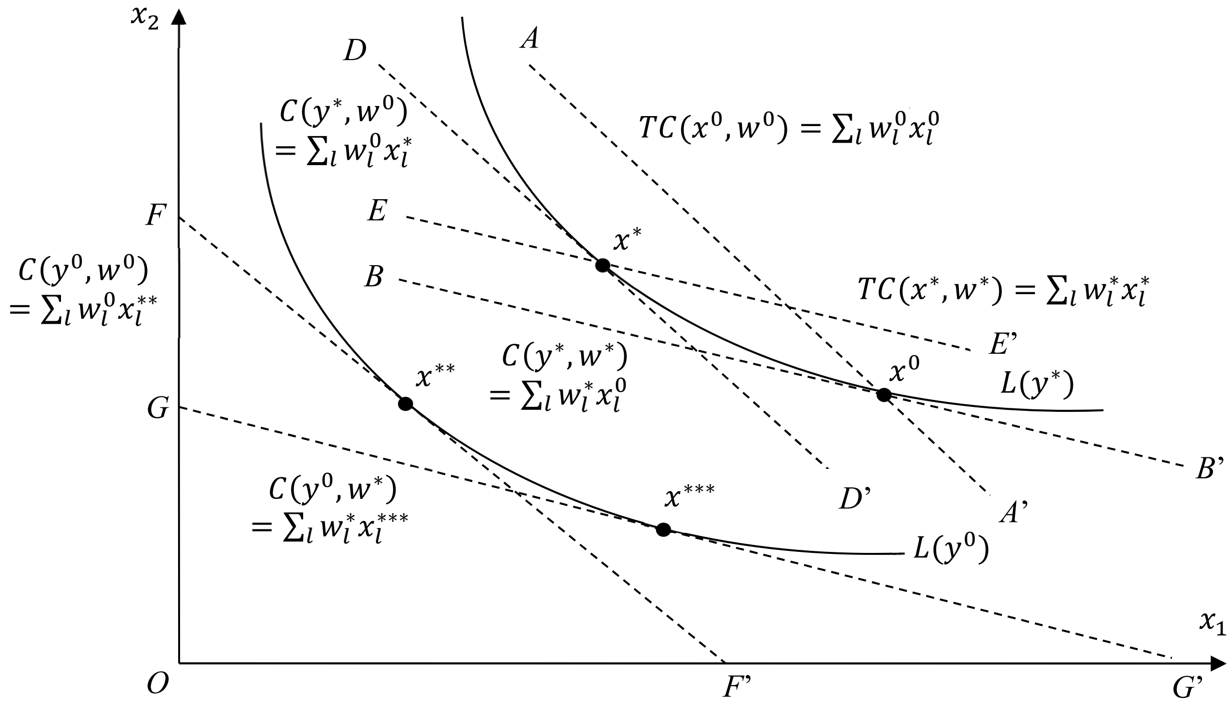
Schmidt, P., & Lovell, C. A. K., (1979). Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers. *Journal of Econometrics* 9: 343-366.

**Figure 1 Cost Allocative Inefficiency without Technical Inefficiency**




*Notes*: $(y^0, x^0, w^0)$ are the observed output, and input quantities and prices. $L(y^0)$ is the isoquant curve with $y^0 = f(x^0) = f(x^*)$. The first type of CAI is related to the cross of two isocost lines at point $x^0$. The isocost line *AA*' passes through $L(y^0)$ at $x^0$ with the slope of $-\frac{w_1^0}{w_2^0}$; the isocost line *BB*' is tangent to $L(y^0)$ at $x^0$ with the slope of $-\frac{w_1^*}{w_2^*}$, where $w^*$ are input shadow prices. The second type of CAI is related to the difference between $x^0$ and $x^*$, where $x^*$ is the optimal inputs point derive from $L(y^0)$ and $w^0$. The isocost line *DD*' is tangent to $L(y^0)$ at $x^*$ with the slope of $-\frac{w_1^0}{w_2^0}$. The distance between *AA*' and *DD*' measures the second type of CAI. An alternative measure of CAI is the distance between *BB*' and *EE*', where the isocost line *EE*' with $TC(x^*, w^*) = \sum_j w_j^* x_j^*$ passes through $L(y^0)$ at $x^*$ with a slope of $-\frac{w_1^*}{w_2^*}$. At $x^0$, the observed cost share of $x_1$ is the ratio of $x_1^0$ to *OA*'. Given $w^*$, the optimal cost share of $x_1$ is the ratio of $x_1^0$ to *OB*'. The difference in the cost shares of $x_1$ is due to the first type of CAI. The observed cost share of $x_2$ is the ratio of $x_2^0$ to *OA*. Given $w^0$, the optimal input is at $x^*$ and the optimal cost share of $x_2$ is the ratio of $x_2^*$ to *OD*. The difference in the cost shares of $x_2$ is due to the second type of CAI.

**Figure 2 Cost Allocative Inefficiency with Technical Inefficiency**




*Notes*: $(y^0, x^0, w^0)$ are the observed output, and input quantities and prices. $L(y^0)$ and $L(y^*)$ are the isoquant curves with $y^* = f(x^0) = f(x^*) > y^0 = f(x^{**}) = f(x^{***})$. For $L(y^*)$, the definitions of four isocost lines *AA*', *BB*', *DD*', and *EE*' and two inputs points $x^0$ and $x^*$ are similar to those defined in Figure 1. $w^*$ are the input shadow prices, which is derived from the slope of *BB*' $\left(-\frac{w_1^*}{w_2^*}\right)$ and *BB*' is tangent to $L(y^*)$ at $x^0$. $x^*$ is the optimal inputs point given $(y^*, w^0)$, where *DD*' is tangent to $L(y^*)$ at $x^*$ with the slope of $-\frac{w_1^0}{w_2^0}$. For $L(y^0)$, $x^{**}$ is the optimal input point given $(y^0, w^0)$, where the iscost line *FF*' is tangent to $L(y^0)$ at $x^{**}$ with the slope of $-\frac{w_1^0}{w_2^0}$. $x^{***}$ is the optimal input point given $(y^0, w^*)$, where the isocost line *GG*' is tangent to $L(y^0)$ at $x^{***}$ with the slope of $-\frac{w_1^*}{w_2^*}$. The different cost shares can be defined as those in Figure 1.

Table 1 Cost Share Equations

| | Cost Shares | Triplets | Figure 2 |
|---|---|---|---|
| (i) | $s_j^0(x^0, w^0) = \frac{w_j^0 x_j^0}{\sum w_l^0 x_l^0}$ | $(y^*, x^0, w^0)$ | *AA'* |
| (ii) | $s_j^p(y^*, x^0) = \frac{\eta_j(y^*,x^0)}{\eta(y^*,x^0)} = \frac{w_j^* x_j^0}{\sum w_l^* x_l^0}$ | $(y^*, x^0, w^*)^p$ | *BB'* |
| (iii) | $s_j^c(y^*, w^*) = \frac{\partial \ln C(y^*,w^*)}{\partial \ln w_j} = \frac{w_j^* x_j^0}{\sum w_l^* x_l^0}$ | $(y^*, x^0, w^*)^p$ | *BB'* |
| (iv) | $s_j^c(y^*, w^0) = \frac{\partial \ln C(y^*,w^0)}{\partial \ln w_j} = \frac{w_j^0 x_j^*}{\sum w_l^0 x_l^*}$ | $(y^*, x^*, w^0)^c$ | *DD'* |
| (v) | $s_j^p(y^*, x^*) = \frac{\eta_j(y^*,x^*)}{\eta(y^*,x^*)} = \frac{w_j^0 x_j^*}{\sum w_l^0 x_l^*}$ | $(y^*, x^*, w^0)^c$ | *DD'* |
| (vi) | $s_j^c(y^0, w^0) = \frac{\partial \ln C(y^0,w^0)}{\partial \ln w_j} = \frac{w_j^0 x_j^{**}}{\sum w_l^0 x_l^{**}}$ | $(y^0, x^{**}, w^0)^c$ | *FF'* |
| (vii) | $s_j^p(y^0, x^{**}) = \frac{\eta_j(y^0,x^{**})}{\eta(y^0,x^{**})} = \frac{w_j^0 x_j^{**}}{\sum w_l^0 x_l^{**}}$ | $(y^0, x^{**}, w^0)^c$ | *FF'* |
| (viii) | $s_j^c(y^0, w^*) = \frac{\partial \ln C(y^0,w^*)}{\partial \ln w_j} = \frac{w_j^* x_j^{***}}{\sum w_l^* x_l^{***}}$ | $(y^0, x^{***}, w^*)^c$ | *GG'* |
| (ix) | $s_j^p(y^0, x^{***}) = \frac{\eta_j(y^0,x^{***})}{\eta(y^0,x^{***})} = \frac{w_j^* x_j^{***}}{\sum w_l^* x_l^{***}}$ | $(y^0, x^{***}, w^*)^c$ | *GG'* |

*Notes*: Assume production function $y = f(x)$ and cost function $C(y, w)$ are given. The observed data are $(y^0, x^0, w^0)$. Let $y^* = f(x^0) > y^0$. The relative shadow price for input $j$ is $\frac{w_j^*}{w_1^*} = \frac{\eta_j(y^*,x^0)}{\eta_1(y^*,x^0)} \frac{x_1^0}{x_j^0}$. $\eta_j(y, x) = \frac{\partial f}{\partial x_j} \frac{x_j}{f}$ is the output elasticity of input $j$. $\eta = \sum \eta_j$ is economies of scale. The optimal inputs derived from four different combinations of outputs and input prices $(y^*, w^*)$, $(y^*, w^0)$, $(y^0, w^0)$, and $(y^0, w^*)$ are: $x_j^0 = \frac{\partial C(y^*,w^*)}{\partial w_j}$, $x_j^* = \frac{\partial C(y^*,w^0)}{\partial w_j}$, $x_j^{**} = \frac{\partial C(y^0,w^0)}{\partial w_j}$, $x_j^{***} = \frac{\partial C(y^0,w^*)}{\partial w_j}$. And $y^* = f(x^*)$ and $y^0 = f(x^{**}) = f(x^{***})$. In the case of two inputs, input cost share inequalities can be demonstrated using the isocost lines in Figure 2. See the footnotes in Figures 1 and 2.

Table 2 The Estimates of Input Cost Shares, Inefficiency, and Economies of Scale

| | | Standard Procedure | | Three-Stage Procedure | | | |
|---|---|---|---|---|---|---|---|
| | Original | Cost Function | | Production Function | | Cost Function | |
| | | SURE | Frontier | SURE | Frontier | SURE | Frontier |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| **1955** | | | | | | | |
| $\hat{s}_{labor}$ | 10.57 | 10.58 | 38.89 | 10.50 | -20.56 | 10.55 | 12.57 |
| | (0.35) | (-0.005) | (-23.69) | (0.454) | (9.566) | (0.110) | (0.920) |
| $\hat{s}_{fuel}$ | 46.78 | 46.90 | 52.13 | 47.21 | 75.42 | 47.11 | 62.77 |
| | (0.99) | (-0.135) | (-3.093) | (-0.684) | (-9.040) | (-0.380) | (-8.797) |
| $\hat{s}_{capital}$ | 42.68 | 42.52 | 8.98 | 42.29 | 45.14 | 42.34 | 24.66 |
| | (0.92) | (0.194) | (14.50) | (0.737) | (-1.902) | (0.410) | (5.681) |
| $E(\exp u_i \mid \varepsilon_i)$ | | | 1.446 | | 0.775 | | 1.308 |
| | | | (0.075) | | (0.008) | | (0.051) |
| $\hat{\eta}$ | | 1.309 | 1.163 | 1.226 | 1.092 | 1.251 | 1.122 |
| | | (0.028) | (0.016) | | (0.007) | (0.014) | (0.010) |
| **1970** | | | | | | | |
| | | Standard Procedure | | Three-Stage Procedure | | | |
| | Original | Cost Function | | Production Function | | Cost Function | |
| | | SURE | Frontier | SURE | Frontier | SURE | Frontier |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| $\hat{s}_{labor}$ | 13.90 | 13.84 | 15.70 | 13.77 | -3.33 | 13.65 | 13.98 |
| | (0.43) | (0.157) | (-2.619) | (0.549) | (32.27) | (0.643) | (-0.135) |
| $\hat{s}_{fuel}$ | 63.24 | 63.35 | 72.48 | 63.55 | 96.41 | 63.54 | 76.07 |
| | (0.66) | (-0.225) | (-10.22) | (-0.623) | (-64.65) | (-0.597) | (-23.88) |
| $\hat{s}_{capital}$ | 22.64 | 22.81 | 11.82 | 22.68 | 6.92 | 22.82 | 9.95 |
| | (0.48) | (-0.401) | (19.93) | (-0.118) | (28.05) | (-0.419) | (16.39) |
| $E(\exp u_i \mid \varepsilon_i)$ | | | 1.108 | | 0.913 | | 1.065 |
| | | | (0.008) | | (0.003) | | (0.005) |
| $\hat{\eta}$ | | 1.132 | 1.110 | 1.160 | 1.068 | 1.117 | 1.107 |
| | | (0.012) | (0.013) | | (0.009) | (0.010) | (0.012) |

*Notes*: $\hat{s}_{labor}$, $\hat{s}_{fuel}$, and $\hat{s}_{capital}$ are the cost shares of labor, fuel, and capital, respectively. Column (1) shows the mean of the original cost shares with the standard errors in the parenthesis, and the rest of the columns show the estimated optimal cost shares from different models. $E(\exp u_i \mid \varepsilon_i)$ is the inefficiency measure from the stochastic frontier model. $\hat{\eta}$ is the estimate of economies of scale. The numbers in the parentheses under the estimates of the three cost shares in columns (2) – (7) are the $t$-statistics comparing the difference between the mean observed cost share and the mean optimal cost share of each input. The numbers in the parentheses under the other estimates are the standard errors for the sample mean.

Table 3 Cost Inefficiency Decomposition Comparisons

| | 1955 | | | | 1970 | | | |
|---|---|---|---|---|---|---|---|---|
| | Total | | CE | Percentage | Total | | CE | Percentage |
| Total cost | 14.72 (1.75) | CE | 4.67 (0.76) | 31.73% | 53.27 (6.93) | CE | 6.72 (1.03) | 12.61% |
| $\hat{C}(\hat{y}^{*}, w^{0})$ | 13.03 (1.57) | CAI | 1.69 (0.23) | 11.48 | 49.9 (6.29) | CAI | 3.37 (0.75) | 6.33 |
| $\hat{C}(y^{0}, w^{0})$ | 10.04 (1.08) | CTE | 2.98 (0.60) | 20.24 | 46.55 (6.12) | CTE | 3.35 (0.59) | 6.29 |

*Notes*: See notes on Table 2. Total cost is the mean of the observed total cost of all firms. $\hat{C}(\hat{y}^{*}, w^{0})$ and $\hat{C}(y^{0}, w^{0})$ are the mean estimates of the optimal costs calculated from the estimated cost function from the three-stage procedure, where $\hat{y}^{*}$ is the estimated optimal output from the production frontier model and $y^{0}$ is the observed output. The numbers in the CE column are the mean estimates of total cost inefficiency, and those due to allocative and technical inefficiency from the three-stage procedure. The percentage is calculated by dividing each type of the cost inefficiency by the observed total cost. The numbers in the parentheses are the standard errors for the sample mean.