RJ10490 (A1111-001) November 1, 2011
Other

IBM Research Report

On a Simple Comparison Procedure in Analyzing Observational Data

Ying Tat Leung, Barbara Jones
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099
USA

On a Simple Comparison Procedure in Analyzing Observational Data

Ying Tat Leung and Barbara Jones
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120
U.S.A.

October 2011

Abstract

In many business and engineering situations, we are interested in the impact of a single factor on the outcome or performance of a system. Ideally, we can conduct a controlled experiment to investigate. However, experimentation is often too expensive or even impossible and we have to resort to an observational approach in which data about the system is collected and subsequently analyzed in the hope of being able to isolate the effect of the factor of interest. We consider a naïve approach in which we perform a simple comparison of the observed system performance at different, known settings of the single factor of interest. Despite the fact that we ignore all the other factors which could be at any number of unknown settings when the observations were taken, this procedure turns out to be more reasonable than it might appear. We analyze this simple comparison procedure in general as well as under the assumption of several different forms of the underlying model of the system performance.

Keywords: performance estimation, single factor impact, Constant Elasticity of Substitution function, randomized experiments

1. INTRODUCTION AND PROBLEM DESCRIPTION

In empirical analyses arising in many different subjects, a common goal is to discern the effect, if any, of a single factor on an end result of interest which we shall call the performance function. In business analysis, a frequently encountered scenario is to investigate the impact of a single factor on business performance. A specific example is that a company investigating the prospect of deploying a customer relationship management (CRM) system would want to examine the difference in selling costs between companies who employ a CRM system and those who do not. In manufacturing engineering, often our goal is to investigate whether a manufacturing process parameter is responsible for a certain aspect of product quality. For example, in plastic parts manufacturing we may be interested in knowing whether the presence of an air draft in the room affects the number of surface defects of a molded plastic part. In engineering design, we might be interested in the effect of a design feature on the performance of the system. For instance, in automobiles we are interested in the effect of an increase in tire size on the stability of a vehicle as measured by the National Highway Transportation Safety Administration's star rating system.

In order to accurately estimate the effect of a single factor, other factors that potentially have an effect on the performance function have to be carefully controlled or accounted for. If it is determined to perform an experiment, the strategy of control can be used. (For factors we cannot control, we can employ the strategies of randomization and blocking. We shall return to randomized experiments in Section 4.) Such is the case for laboratory studies of manufacturing processes or in laboratory tests of automobiles. Here the extensive literature of statistical experimental design can be used to help design an effective set of experiments to simultaneously study the effects of several factors while minimizing the number of experiments needed (see, e.g., Box et al. 2005). In some cases, a controlled experiment is not practical. To illustrate, it will be too expensive to do an experiment on changing the business strategy of a company. In some other cases, to supplement the results of a controlled experiment, the performance of the system of interest as used in real life has to be analyzed. In both types of scenarios, one has to resort to using existing data or to making observations of a working system that will not be intervened. Here we can use the strategy of accounting for the effects of all possible factors, in addition to the one being investigated, so that our conclusion is free of confounding effects. A proper way is to develop a mathematical model of the performance function of all factors involved, estimate the parameters of the model using observed data, and then use the estimated model to conclude whether the effect of a selected factor is significant. For example, in a linear model this can be done by testing whether the model coefficient corresponding to the factor is significantly different from zero. If so, we also know the estimated magnitude of the effect of that factor. In business research, such an approach is commonly used in organizational performance or organizational behavior studies where we are interested in the effect of one or few factors (such as the presence of a CRM system) on the performance of a company. Indeed, one may argue that we have no other choice there, since only results from a real life setting are relevant in the world of business.

This approach of utilizing a model, however, is technically non-trivial. First, the list of factors that have an impact on performance is generally difficult to obtain precisely. In business research in particular, often there are so many factors that affect the performance of a business, some of which may be difficult to find a simple measure to characterize. Missing a factor could completely change the conclusion. Even if all the factors and their measurements have been identified, collecting sufficient data for each of the measurements may not be easy in practice. Further, when all data have been collected, developing an adequate model that fits the data reasonably well is a task that requires specialized expertise. Each class of models has its own set of technical assumptions and methodologies for fitting; typically an entire book of significant technical complexity is dedicated to the basics of one class of models. For instance, for linear models we have, among many others, Draper and Smith (1998), generalized linear models McCullagh and Nelder (1989), structural equation models Bollen (1989).

At the other extreme, a naïve approach is to use a straightforward comparison procedure in which data on business performance are directly compared between cases of different levels of the factor of interest. For example, we may have a database of historical data on a measure of business performance together with the corresponding data on a set of factors potentially affecting the business performance. We are interested in the impact of one of the factors. For simplicity of argument we assume that this factor of interest has two levels, high and low. The database will have two subsets of records, depending on whether the factor of interest is at high or low. We take the average of the performance in each subset, calculate the difference of the two averages, and call this the effect of the factor of interest. That is, we assume that any difference in business performance observed is attributed entirely to the effect of this factor.

The latter, straightforward approach is much more attractive in terms of ease of execution, in view of the complexity of the proper modeling approach discussed earlier. The problem is that the results will seem to be obviously in doubt, since we are ignoring the effect of many other factors and are effectively assuming that they remain constant while only our factor of interest changes. Even in a controlled experiment, the levels of the controlled factors may in fact vary a little due to variations in the physical environment, imperfect instruments, limited precision of human actions, etc. But how bad will the result be? Will the results be somewhat salvageable so that it can at least be used as a first approximation? Will we be so lucky to encounter some conditions under which the effects of other factors will "average out" to be a constant quantity so that the differencing operation will eliminate them? We attempt to provide some insights in this paper.

In Section 2, we develop a general model to precisely describe the problem mentioned. Without specific assumptions, it seems that only limited results can be obtained from the general model. So we resort to analyzing some specific model forms for the performance function in Section 3. Section 4 reviews a thread of research in randomized experiments that is related to our study. Section 5 discusses practical implications and provides some concluding remarks.

2. ANALYSIS OF A SIMPLE COMPARISON PROCEDURE

We are interested in estimating the effect of a selected vector of factors, denoted by $\vec{x} = (x_1,...,x_p)$, on a real-valued, performance function $f(\vec{x}, \vec{y})$, where $\vec{y} = (y_1,..., y_q)$ is the vector of factors other than $\vec{x}$ that impact *f*. Let $\vec{x}_i$ be the value of $\vec{x}$ at level *i*. A routine approach in estimating the effect of $\vec{x}$ on *f* is by evaluating *f* at several discrete settings of $\vec{x}$. In particular, in many applications (e.g., when $\vec{x} = x$, a single factor) it is common to use two settings of $\vec{x}$, keeping $\vec{y} = \vec{y}_0$ constant (or assuming $\vec{y}$ to be constant) and compare their corresponding values of *f*. That is, we want to estimate

$$g(\vec{x}_1, \vec{x}_2, \vec{y}_0) = f(\vec{x}_2, \vec{y}_0) - f(\vec{x}_1, \vec{y}_0) . \tag{2.1}$$

In a real-life application, *f* is typically not known exactly and we can only observe some noisy version of it, denoted by $\hat{f}$. A common assumption is that the *n*-th observation of $\hat{f}$, denoted by $\hat{f}_n$, satisfies the following set of conditions:

$$\begin{aligned} &\mathrm{E}\hat{f}_n(\vec{x}, \vec{y}) = f(\vec{x}, \vec{y}), \\ &\mathrm{var}[\hat{f}_n(\vec{x}, \vec{y})] < \infty, \\ &\hat{f}_n(\vec{x}, \vec{y}) \text{ is i.i.d. given } (\vec{x}, \vec{y}), n = 1,2,3,..., \end{aligned} \tag{2.2}$$

where $\mathrm{var}[\hat{f}_n(\vec{x}, \vec{y})] = E[f_n^2(\vec{x}, \vec{y})] - [\mathrm{E}f_n(\vec{x}, \vec{y})]^2$.

This covers a frequent underlying assumption of additive noise, i.e., $\hat{f}_n(\vec{x}, \vec{y}) = f(\vec{x}, \vec{y}) + \varepsilon_n$, where ε_n is an i.i.d. random variable with $\mathrm{E}(\varepsilon_n) = 0$, and $\mathrm{var}[\varepsilon_n] < \infty$, $n = 1, 2, 3, \ldots$.

To estimate *g* as defined in (2.1), we observe multiple samples of *f* at the two settings of $\vec{x}_1$ and $\vec{x}_2$ but at a fixed $\vec{y} = \vec{y}_0$, to obtain

$$\hat{g}(\vec{x}_1, \vec{x}_2, \vec{y}_0) = \frac{1}{N_2}\sum_{n=1}^{N_2} \hat{f}_n(\vec{x}_2, \vec{y}_0) - \frac{1}{N_1}\sum_{n=1}^{N_1} \hat{f}_n(\vec{x}_1, \vec{y}_0) , \tag{2.3}$$

where N_i is the number of observations of $\hat{f}$ at $(\vec{x}_i, \vec{y}_0)$. This is the ideal case where we can, for example, choose two subsets of records in a database corresponding to $(\vec{x}_1, \vec{y}_0)$ or $(\vec{x}_2, \vec{y}_0)$, then for each set take the average and then the difference between the two averages.

We assume that

$$N_i > 0 \text{ is either fixed or an integer-valued random variable that is independent of } \hat{f}_n \text{ for } n = 1, 2, 3, \ldots \tag{2.4}$$

Under assumptions (2.2) and (2.4), $\hat{g}$ is unbiased, i.e.,

$$\mathrm{E}(\hat{g}(\vec{x}_1, \vec{x}_2, \vec{y}_0)) = f(\vec{x}_2, \vec{y}_0) - f(\vec{x}_1, \vec{y}_0) = g(\vec{x}_1, \vec{x}_2, \vec{y}_0), \tag{2.5}$$

where the expectation is taken w.r.t. $\hat{f}_n$ and N_i, if applicable.

Due to the circumstances discussed above, the settings of $\vec{y}$ in the two scenarios of $\vec{x}_1$ and $\vec{x}_2$ may possibly be different from each other and as well different from $\vec{y}_0$. We model this situation with $\vec{y}$ being a random variable and let $\vec{y}_{in}$ be the (random) level of $\vec{y}$ in the *n*-th observation of $\hat{f}$ when $\vec{x} = \vec{x}_i$. (For this reason we refer to $\vec{y}$ as the unmeasured or unknown variable and $\vec{x}$ the measured or known variable.) Similar to (2.3), we obtain an estimate for g as

$$\tilde{g}(\vec{x}_1, \vec{x}_2, \vec{y}_0) = \frac{1}{N_2}\sum_{n=1}^{N_2} \hat{f}_n(\vec{x}_2, \vec{y}_{2n}) - \frac{1}{N_1}\sum_{n=1}^{N_1} \hat{f}_n(\vec{x}_1, \vec{y}_{1n}). \tag{2.6}$$

We assume that

$$N_i \text{ is either fixed or a random variable that is independent of } \hat{f}_n \text{ and } \vec{y}_{in} \text{ for } i = 1, 2 \text{ and } n = 1, 2, 3, \ldots \tag{2.7}$$

$$\text{For } i = 1 \text{ or } 2, \text{ given } \vec{x}_i, \ \vec{y}_{in} \text{ are i.i.d. for } n = 1, 2, 3, \ldots, \text{ and} \tag{2.8}$$

$$\mathrm{E}(\vec{y}_{in}) = \vec{y}_0; \ \mathrm{E}(\vec{y}_{in}^{\,2}) < \infty \tag{2.9}$$

Effectively we assume that $\vec{y}$ follow a single, fixed distribution, the mean of which is the correct value $\vec{y}_0$, that is independent of the factor of interest $\vec{x}$. If we have absolutely no knowledge of the missing values of $\vec{y}$ or even what factor is missing, this does not seem unreasonable. This is also the most basic scenario to analyze.

2.1. Bias of the Estimator

The estimator $\tilde{g}$ as defined in (2.6) is representative of the output of a naïve comparison procedure introduced in Section 1. We now study this estimator more closely.

Under assumptions (2.2), (2.7), and (2.8), taking advantage of the fact that each of N_1 and N_2 is independent of everything else, we have, for i=1 or 2,

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{N_i}\sum_{n=1}^{N_i}\hat{f}_n(\vec{x}_i,\vec{y}_{in})\right] \\
&= \mathrm{E}\left[\mathrm{E}\left(\frac{1}{N_i}\sum_{n=1}^{N_i}\hat{f}_n(\vec{x}_i,\vec{y}_{in}) \mid N_i\right)\right] \\
&= \mathrm{E}\left[\frac{1}{N_i}N_i\mathrm{E}\hat{f}_1(\vec{x}_i,\vec{y}_{i1})\right] \\
&= \mathrm{E}\hat{f}_1(\vec{x}_i,\vec{y}_{i1}) \\
&= \mathrm{E}\left[\mathrm{E}\left(\hat{f}_1(\vec{x}_i,\vec{y}_{i1}) \mid \vec{y}_{i1}\right)\right] \\
&= \mathrm{E}f(\vec{x}_i,\vec{y}_{i1})
\end{aligned}$$

Therefore, we obtain

$$\mathrm{E}\tilde{g}(\vec{x}_1,\vec{x}_2,\vec{y}_0) = \mathrm{E}f(\vec{x}_2,\vec{y}_{21}) - \mathrm{E}f(\vec{x}_1,\vec{y}_{11})\,. \tag{2.10}$$

From (2.1) and (2.10), we have

$$\begin{aligned}
&\mathrm{E}(\tilde{g}(\vec{x}_1,\vec{x}_2,\vec{y}_0) - g(\vec{x}_1,\vec{x}_2,\vec{y}_0)) = \\
&[\mathrm{E}f(\vec{x}_2,\vec{y}_{21}) - f(\vec{x}_2,\vec{y}_0)] - [\mathrm{E}f(\vec{x}_1,\vec{y}_{11}) - f(\vec{x}_1,\vec{y}_0)]\,.
\end{aligned} \tag{2.11}$$

Hence, a necessary and sufficient condition for $\tilde{g}$ to be unbiased is that

$$[\mathrm{E}f(\vec{x}_2,\vec{y}_{21}) - f(\vec{x}_2,\vec{y}_0)] - [\mathrm{E}f(\vec{x}_1,\vec{y}_{11}) - f(\vec{x}_1,\vec{y}_0)] = 0 \tag{2.12}$$

Under assumption (2.9), a sufficient condition for $\tilde{g}$ to be unbiased, i.e., (2.12) to be true, is that,

$$\mathrm{E}f(\vec{x}_i, \vec{y}_{i1}) = f(\vec{x}_i, \mathrm{E}(\vec{y}_{i1})) \text{ for } i = 1, 2. \tag{2.13}$$

It is well known that (2.13) does not hold in general (e.g., Young 2010). However, (2.13) is still true if

$$\mathrm{E}f(\vec{x}_2, \vec{y}_{21}) - f(\vec{x}_2, \vec{y}_0) = \mathrm{E}f(\vec{x}_1, \vec{y}_{11}) - f(\vec{x}_1, \vec{y}_0). \tag{2.14}$$

Hence, an alternative sufficient condition for $\tilde{g}$ to be unbiased is that

$$[\mathrm{E}f(\vec{x}, \vec{y}_l) - f(\vec{x}, \mathrm{E}\vec{y}_l)] \text{ is independent of } \vec{x}. \tag{2.15}$$

We use these conditions to investigate some specific forms of f in Section 3.

From the RHS of (2.11), we know that the relationship between $\mathrm{E}f(\vec{x}_i, \vec{y}_{i1})$ and $f(\vec{x}_i, \mathrm{E}(\vec{y}_{i1}))$ is critical to the unbiasedness of $\tilde{g}$. When f is convex and the random variables $\vec{y}_i$ are assumed to be bounded by constants, each of the two summands in the RHS of (2.11), i.e., $\mathrm{E}f(\vec{x}_i, \vec{y}_{i1}) - f(\vec{x}_i, \mathrm{E}(\vec{y}_{i1}))$, i = 1, 2, can be bounded from the above by the Edmundson-Madansky upper bound and from below using Jensen's inequality. In fact, tighter bounds than the abovementioned have been developed (Dokov and Morton 2002, 2005). Using these bounds we can derive a bound for (2.11) in a straightforward way. For example, in the simplest case when $\vec{x}_i = x_i$ and $\vec{y}_{in} = y_{in}$ are scalar, using the classic Edmundson-Madansky and Jensen's bounds, we obtain

$$\begin{aligned}&0 \le \mathrm{E}(\tilde{g}(x_1, x_2, y_0) - g(x_1, x_2, y_0)) \\ &\le \frac{b - y_0}{b - a} f(x_2, a) + \frac{y_0 - a}{b - a} f(x_2, b) - f(x_2, y_0)\end{aligned}, \tag{2.16}$$

where a and b are respectively the lower and upper bounds of y_{in}. This yields an interesting insight that $\tilde{g}$ always overestimates g (in expectation) when f is convex.

2.2. Variance of the Estimator

Now we calculate the variance of $\tilde{g}(\vec{x}_1, \vec{x}_2, \vec{y}_0)$. From (2.6), we have

$$
\begin{aligned}
&\tilde{g}^2(\vec{x}_1,\vec{x}_2,\vec{y}_0) = \\
&\frac{1}{N_2{}^2}\left[\sum_{n=1}^{N_2}\hat{f}_n{}^2(\vec{x}_2,\vec{y}_{2n}) + \sum_{n=1}^{N_2}\sum_{m=1,m\neq n}^{N_2}\hat{f}_n(\vec{x}_2,\vec{y}_{2n})\hat{f}_m(\vec{x}_2,\vec{y}_{2m})\right] \\
&+\frac{1}{N_1{}^2}\left[\sum_{n=1}^{N_1}\hat{f}_n{}^2(\vec{x}_1,\vec{y}_{1n}) + \sum_{n=1}^{N_1}\sum_{m=1,m\neq n}^{N_1}\hat{f}_n(\vec{x}_1,\vec{y}_{1n})\hat{f}_m(\vec{x}_1,\vec{y}_{1m})\right] \\
&-\frac{2}{N_1N_2}\left[\sum_{n=1}^{N_2}\sum_{m=1}^{N_1}\hat{f}_n(\vec{x}_2,\vec{y}_{2n})\hat{f}_m(\vec{x}_1,\vec{y}_{1m})\right]
\end{aligned} \tag{2.17}
$$

Taking expectation on both sides and again using a conditional expectation on N_1 and N_2 and their independence of everything else, we have, after some algebra,

$$
\begin{aligned}
&\mathrm{E}\tilde{g}^2(\vec{x}_1,\vec{x}_2,\vec{y}_0) = \mathrm{E}(1/N_2)\operatorname{var}\left(\hat{f}_1(\vec{x}_2,\vec{y}_{21})\right) \\
&+\mathrm{E}(1/N_1)\operatorname{var}\left(\hat{f}_1(\vec{x}_1,\vec{y}_{11})\right) + \left[\mathrm{E}f(\vec{x}_2,\vec{y}_{21}) - \mathrm{E}f(\vec{x}_1,\vec{y}_{11})\right]^2
\end{aligned} \tag{2.18}
$$

Noting (2.10), we have

$$
\begin{aligned}
&\operatorname{var}\left(\tilde{g}(\vec{x}_1,\vec{x}_2,\vec{y}_0)\right) = \\
&\mathrm{E}(1/N_2)\operatorname{var}\left(\hat{f}_1(\vec{x}_2,\vec{y}_{21})\right) + \mathrm{E}(1/N_1)\operatorname{var}\left(\hat{f}_1(\vec{x}_1,\vec{y}_{11})\right)
\end{aligned} \tag{2.19}
$$

The above shows that the estimator $\tilde{g}$ behaves similarly to the simple average sample mean estimator; for example, if $N_1 = N_2 = N$ which is fixed, then the variance of the estimator decreases at the rate of $1/N$. This is as expected from the fact that the estimator is of the form of the difference of two sample averages.

3. BIAS PROPERTY FOR SOME COMMON MODEL FORMS

In this section we examine several model forms of the performance function f. These model forms are commonly encountered in different technical areas.

3.1. Separable Models

We consider the case when f is separable in $\vec{x}$ and $\vec{y}$, in the form of

$$
f(\vec{x},\vec{y}) = Af_1(\vec{x}) + Bf_2(\vec{y}) + Cf_3(\vec{x})f_4(\vec{y}), \tag{3.1}
$$

where A, B, C are non-zero constants. Then

$$\begin{aligned} \mathrm{E}f(\vec{x},\vec{y}) - f(\vec{x},\mathrm{E}\vec{y}) &= B[\mathrm{E}f_2(\vec{y}) - f_2(\mathrm{E}\vec{y})] \\ &+ Cf_3(\vec{x})[\mathrm{E}f_4(\vec{y}) - f_4(\mathrm{E}\vec{y})] \end{aligned} \tag{3.2}$$

Under our assumptions on $\vec{y}$, for (2.14) to be true, we need

$$\begin{aligned} &B[\mathrm{E}f_2(\vec{y}) - f_2(\mathrm{E}\vec{y})] + Cf_3(\vec{x}_2)[\mathrm{E}f_4(\vec{y}) - f_4(\mathrm{E}\vec{y})] = \\ &B[\mathrm{E}f_2(\vec{y}) - f_2(\mathrm{E}\vec{y})] + Cf_3(\vec{x}_1)[\mathrm{E}f_4(\vec{y}) - f_4(\mathrm{E}\vec{y})] \\ &\Leftrightarrow f_3(\vec{x}_2)[\mathrm{E}f_4(\vec{y}) - f_4(\mathrm{E}\vec{y})] = f_3(\vec{x}_1)[\mathrm{E}f_4(\vec{y}) - f_4(\mathrm{E}\vec{y})] \end{aligned}$$

A sufficient condition for the above to be true is that

$$\mathrm{E}f_4(\vec{y}) = f_4(\mathrm{E}(\vec{y})), \tag{3.3}$$

while a necessary and sufficient condition is that

$$f_3(\vec{x}_1) = f_3(\vec{x}_2). \tag{3.4}$$

Condition (3.4) represents the case that $\vec{x}$ has no impact on the cross *x-y* term.

Clearly, (3.3) is satisfied when $f_4(\vec{y})$ is linear w.r.t. $y_1, \ldots, y_n$. In other words, for $\hat{g}$ to be an unbiased estimator for *g*, we require that the cross *x-y* terms of the performance function be linear with respect to factors that are not under investigation. The separable part of the cross *x-y* terms that involves $\vec{x}$, the factors of interest, the terms that involve only $\vec{x}$ or $\vec{y}$ alone, can still be general (e.g., nonlinear).

3.2. Polynomial Models

Polynomial models need no introduction; suffice it to say they are among the most popular classes of empirical models used to fit data. A polynomial model can be viewed as a special class of separable models. Because of its specific structure we are able to draw a stronger conclusion about them. A polynomial in $\vec{z}$ of degree *d* can be written as

$$f(\vec{z}) = \sum_{\gamma \in I} a_\gamma \vec{z}^{\gamma}, \tag{3.5}$$

where

$$
\begin{aligned}
&\vec{z} = (z_1, \cdots, z_T), \gamma = (\gamma_1, \cdots, \gamma_T), I = \{\gamma : \gamma_1 + \cdots + \gamma_T \le d\}, \\
&d = \max_{\gamma \in I : a_\gamma \neq 0} \|\gamma\|, \\
&\|\gamma\| = \gamma_1 + \cdots + \gamma_T, \\
&\vec{z}^{\gamma} = \prod_{i=1}^{T} z_i^{\gamma_i}.
\end{aligned}
$$

Using our notation of $\vec{x}$ being the variables of interest and $\vec{y}$ the other factors that influence performance, we split $\vec{z}$ in the above accordingly so that our polynomial model is as follows.

$$f(\vec{x}, \vec{y}) = \sum_{\gamma \in I} a_\gamma \vec{x}^{\alpha} \vec{y}^{\beta}, \tag{3.6}$$

where

$$
\begin{aligned}
&\vec{x} = (x_1, \cdots, x_p), \vec{y} = (y_1, \cdots, y_q), \alpha = (\alpha_1, \cdots, \alpha_p), \beta = (\beta_1, \cdots, \beta_q), \\
&\gamma = (\alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q), I = \{\gamma : \alpha_1 + \cdots + \alpha_p + \beta_1 + \cdots + \beta_q \le d\}, \\
&d = \max_{\gamma \in I : a_\gamma \neq 0} \|\gamma\|, \\
&\|\gamma\| = \alpha_1 + \cdots + \alpha_p + \beta_1 + \cdots + \beta_q, \\
&\vec{x}^{\alpha} = \prod_{i=1}^{p} x_i^{\alpha_i}, \\
&\vec{y}^{\beta} = \prod_{j=1}^{q} y_j^{\beta_j}.
\end{aligned}
$$

When $d = 1$, the model is a simple linear model in $\vec{x}$ and $\vec{y}$ (with no cross terms), so that condition (2.13) applies and $\tilde{g}$ is unbiased.

When $d = 2$, to calculate the LHS of (2.14), we have

$$\mathrm{E}f(\vec{x}, \vec{y}) - f(\vec{x}, \mathrm{E}\vec{y}) = \sum_{\gamma \in I} a_\gamma \vec{x}^{\alpha} \left(\mathrm{E} \prod_{j=1}^{q} y_j^{\beta_j} - \prod_{j=1}^{q} (\mathrm{E} y_j)^{\beta_j} \right). \tag{3.7}$$

We proceed by calculating the summand for each γ based on the values of the exponents α and β, as shown in Table 3.1. Combining the results in Table 3.1, we see that the RHS of (3.7) is a constant. Therefore condition (2.14) applies and $\tilde{g}$ is unbiased.

It turns out that $d = 2$ is the highest degree for which $\tilde{g}$ is unbiased.

Table 3.1. Calculation of RHS of Equation (3.7)

$\lVert\alpha\rVert$	$\lVert\beta\rVert$	Summand in equation (3.6)
2	0	$a_\gamma \vec{x}^\alpha (1-1) = 0$
1	1	$a_\gamma \vec{x}(\mathrm{E}y_k - \mathrm{E}y_k) = 0$ (for some k, $1 \le k \le q$)
0	2	$a_\gamma \left(\mathrm{E}\prod_{j=1}^{q} y_j^{\beta_j} - \prod_{j=1}^{q} (\mathrm{E}y_j)^{\beta_j} \right) = \text{constant}$
1	0	$a_\gamma \vec{x}^\alpha (1-1) = 0$
0	1	$a_\gamma (\mathrm{E}y_k - \mathrm{E}y_k) = 0$ (for some k, $1 \le k \le q$)
0	0	0

This is not a bad result since linear and quadratic models are among the most commonly used to fit observed data and the fitted model is then used to estimate the effect of a particular factor, or to judge whether the impact of a particular factor is significant (i.e., different from zero).

3.3. The Cobb-Douglas Production Function

A production function is a mathematical function used to represent the output of an economic system as a function of its inputs. Production functions are used frequently in the subject of economics, at both the microeconomic (i.e., individual firm) and macroeconomic (i.e., an economic sector or an entire economy) levels. Production functions are often closed form mathematical expressions discovered over time and shown to adequately represent the actual behavior of some economic systems.

One of the oldest production functions which is still in use today is the Cobb-Douglas production function (Cobb and Douglas 1928). It models production output f (the performance function in our terminology) as a simple function of capital x_1 and labor x_2 as follows.

$$f = \alpha_0 x_1^{\alpha_1} x_2^{\alpha_2},$$

where α_i, $i = 0, 1, 2$, are constants (the model parameters). In its more general form of more than two input factors, we have

$$f = \alpha_0 \prod_{i=1}^{p} x_i^{\alpha_i} \ , \tag{3.8}$$

where f is the production output, x_i are the input factors, including labor or capital, and α_i, $i = 1, \ldots, p$, and $p > 0$ are constants.

We may be interested in the effect of changing say $\vec{x} = (x_1,\ldots,x_{k-1})$ on f at a given value of $\vec{y} = (x_{k+1}, \ldots, x_p)$, $1 < k < p$. Similar to Zellner et al. (1966) and Goldberger (1968), assume that the n-th estimate of f takes on the following form:

$$\hat{f}_n = e^{\varepsilon_n} \alpha_0 \prod_{i=1}^{p} x_i^{\alpha_i} \ , \tag{3.9}$$

where ε_n is an i.i.d. random variable with $\mathrm{E}(\varepsilon_n) = 0, n = 1,2,3\ldots$, and a finite variance. Then the logarithmic form of (3.9) satisfies our previous assumption (2.2) and is linear in $\ln(x_i)$, $i = 1, \ldots, p$. Condition (2.13) is true in the log transformed variables. (One has to be careful in estimation using the log transform as it will shift the mean; see Goldberger 1968.)

If we assume that the n-th estimate of f takes on an alternative form with additive noise:

$$\hat{f}_n = \alpha_0 \prod_{i=1}^{p} x_i^{\alpha_i} + \varepsilon_n , \tag{3.10}$$

f does not satisfy any of the conditions discussed. We shall examine this case as the limiting case of the Constant Elasticity of Substitution function in the next section.

3.4. The Constant Elasticity of Substitution Production Function

Another well-known production function in econometric studies is the Constant Elasticity of Substitution (CES) production function. The first version, which models production output as a function of capital and labor, was introduced by Arrow et al. (1961). It was subsequently generalized by many authors. For example, Uzawa (1962) and McFadden (1963) extended it to include more than two factors under different definitions of elasticity of substitution between factors; Sato (1967) introduced a two-level version, in which an input variable in a CES function is in turn modeled by another CES function. The Cobb-Douglas function is a limiting case of CES and so sometimes the CES function is seen as a theoretical evolution of the Cobb-Douglas function. In practical use, however, both have found their own place in empirical models in economics. Miller (2008) presents a comparison of these two models in a macroeconomic application. The CES function has been shown in Christensen et al. (1973) to be part of a larger class of production functions.

Many variations of the basic form of the CES function (of the original two-factor and multi-factor variety) have been developed; we chose to use a fairly compact but general form as in Dhrymes (1967):

$$Q = \left[\sum_{i=1}^{p} \alpha_i x_i^{\beta} \right]^{1/\beta} \tag{3.11}$$

where $\alpha_i > 0$, $i = 1, 2, \ldots, p$, $\beta \in (-\infty, 1)$, Q is the production output, and x_i are the input factors, such as labor or capital.

The CES model, when used as our performance function f, does not satisfy any of the above conditions for unbiasedness. In this section we study this model to see how biased the estimator $\tilde{g}$ is, for a range of realistic β, and with different distribution functions for the unknown or unmeasured variable.

3.4.1. Fundamentals

The production function can be re-expressed in a way which makes the difference between random and measured variables clearer. As in the previous parts of this paper, we call the measured variables $\vec{x}$, and the random, unmeasured variable $\vec{y}$. For simplicity we assume that $\vec{x} = (x_1, \ldots, x_{p-1})$ and $\vec{y} = y = x_p$ in the CES model (3.11). Then for the case of β less than zero, the production function can be written as:

$$Q = 1 / \left[\sum_{i=1}^{p-1} \alpha_i / x_i^{|\beta|} + \alpha_y / y^{|\beta|} \right]^{1/|\beta|} . \tag{3.12}$$

We multiply numerator and denominator by y to get:

$$Q = y / \left[\alpha_y + y^{|\beta|} \sum_{i=1}^{p-1} \alpha_i / x_i^{|\beta|} \right]^{1/|\beta|} \tag{3.13}$$

Now we define a variable, Γ, which is the ratio of the contribution of the unmeasured variable to the measured ones, that is, the ratio of the average of the y term to the sum of the x terms. Writing the average of y as $\overline{y}$, we get for Γ:

$$\Gamma = \left(\alpha_y / \overline{y}^{|\beta|} \right) / \left(\sum_{i=1}^{p-1} \alpha_i / x_i^{|\beta|} \right) \tag{3.14}$$

Expressed in terms of Γ, the equation for Q can be written:

$$\frac{Q}{Q(\alpha_y = 0)} = \frac{y}{\left(\Gamma \overline{y}^{|\beta|} + y^{|\beta|} \right)^{1/|\beta|}} \tag{3.15}$$

Where $Q(\alpha_y = 0)$ is the value of the CES function evaluated as if the unknown variable y does not exist., or does not contribute to the production function, i.e.,

$$Q(\alpha_y = 0) \equiv 1 / \left(\sum_{i=1}^{p-1} \alpha_i / x_i^{|\beta|} \right)^{1/|\beta|} = \left[\sum_{i=1}^{p-1} \alpha_i x_i^{\beta} \right]^{1/\beta} \tag{3.16}$$

We now assume the variable y to be positive definite and to have a fixed distribution. We call this probability density function $\rho(y)$, i.e.,

$$\begin{gathered} \int_0^\infty dy \rho(y) = 1 \\ \overline{y} = \int_0^\infty dy \rho(y) y \end{gathered} \tag{3.17}$$

Now, taking the expectation of both sides of equation (3.15), we obtain:

$$\frac{EQ}{Q(\alpha_y = 0)} = \int_0^\infty dy \rho(y) \frac{y}{\left(\Gamma \overline{y}^{-|\beta|} + y^{|\beta|}\right)^{1/|\beta|}} \tag{3.18}$$

For the given class of CES function defined by β, the quantity on the right hand side of (3.18) is a function of only a single parameter Γ, for a fixed distribution ρ.

Since for all but a few values of β the integral on the right hand side of (3.18) cannot be evaluated in analytic form, we calculate the integral numerically. For the range of β and distribution functions we chose, the integral ceases to have signification contribution for y larger than 100, and so the upper limit itself was not a problem. The challenge in evaluating integral (3.18) comes in careful multiple-precision evaluation of the integrand, especially at moderate values of the argument. There, the nonlinear nature of the quantity in the denominator, combined with rapidly falling values of the distribution function, mean that a good integration routine must be used. If these precautions are taken, the quantity in (3.18) can be evaluated as a function of Γ.

Notice from the form of Q in (3.13) that for $\beta < 0$, including additional terms to the production function, for fixed values of the previous terms, results in a smaller value of the production function. Therefore Q is less than $Q(y = \infty)$, and their ratio is less than one. This range of zero to one is also consistent with the form of the integrand in (3.18). For all values of β and all distributions of y, the quantity in (3.18) is a monotonically decreasing function of Γ.

For $\beta > 0$, the equations are expressed differently. Keeping the same definition of Γ as the ratio of the average contribution of the unmeasured variable to the measured ones, we get

$$\Gamma = \left(\alpha_y \overline{y}^{-\beta}\right) / \left(\sum_{i=1}^{p-1} \alpha_i x_i^\beta\right) \tag{3.19}$$

and

$$\frac{Q}{Q(\alpha_y = 0)} = \left[1 + \frac{\Gamma}{\bar{y}^{\beta}} y^{\beta}\right]^{1/\beta} \tag{3.20}$$

where as before $Q(\alpha_y = 0)$ is the value of the production function with the *y*-term removed.

This gives the expectation of *Q* as

$$\frac{EQ}{Q(\alpha_y = 0)} = \int_0^{\infty} dy\, \rho(y) \left[1 + \frac{\Gamma}{\bar{y}^{\beta}} y^{\beta}\right]^{1/\beta} \tag{3.21}$$

To understand the effects of not precisely knowing or measuring *y*, from (2.13) we are interested in the difference

$$\mathrm{E}Q(\vec{x}_i, \vec{y}_{i1}) - Q(\vec{x}_i, \mathrm{E}(\vec{y}_{i1}))\,. \tag{3.22}$$

We have the expectation of *Q* with the results of evaluating (3.18) for $\beta < 0$ and (3.21) for $\beta > 0$. To determine the second term in (3.22), we go back to (3.15) and (3.20) and substitute $\mathrm{E}(\vec{y}_{i1}) = \bar{y}$ for *y*. As can be easily shown, this gives the same form for both $\beta < 0$ and $\beta > 0$:

$$\frac{Q(\bar{y})}{Q(\alpha_y = 0)} = (1+\Gamma)^{1/\beta}\ . \tag{3.23}$$

We think it is more useful to express the difference (3.22) in terms of the quantities E*Q* and *Q*(E(*y*)), rather than keeping the quantity $Q(\alpha_y = 0)$ in the equations, which assumes it is possible to evaluate the production function with the quantity *y* entirely absent. Accordingly we calculate the quantity

$$\Delta Q = [\mathrm{E}Q(\vec{x}_i, \vec{y}_{i1}) - Q(\vec{x}_i, \mathrm{E}(\vec{y}_{i1}))] / \mathrm{E}Q(\vec{x}_i, \vec{y}_{i1}) \tag{3.24}$$

which is related to equation (2.13), and which would be zero if equation (2.11) would hold. Because like (3.15) and (3.20) it is dimensionless, it can be expressed as a calculated percentage. To the degree that this quantity deviates from zero, it gives an indication of how poor the approximation of using the estimator $\tilde{g}$ is.

3.4.2. Numerical Results

To gain some insights on the extent of the deviation of ΔQ from zero in a practical situation, we calculate ΔQ numerically for some selected parameter values. Since β is a key parameter that impacts ΔQ, we chose its values carefully so that they are representative of estimates that would be obtained in a typical application of the CES function. To this end, we surveyed a number of studies fitting a CES function to real data. Not all applications are directly usable due to the usage of different forms of the CES function; nonetheless we were able to obtain a practical range of β values from a reasonable number of fitted CES functions (Berndt 1976, Chirinko 2008, Salem 2004, Soda and Vichi 1976).

We evaluated difference quantity (3.24) for 5 values of β (β= 1/2, -1/2, -1, -2, and -5), and 5 types of distribution functions for y: Gamma distribution with three parameter values of A (upper case alpha; A= 2, 1, and ½); and normal distribution peaked at zero, and peaked at a finite value of y. For each of these, quantity (3.24) was calculated at a full range of Γ (defined in (3.19)). Details of these results follow.

The Gamma distribution function is defined for $y > 0$ as:

$$\rho(y) = \frac{B^{-A} y^{A-1} e^{-y/B}}{\Gamma(A)} \tag{3.25}$$

where $\Gamma(A)$ is the gamma function, required for normalization. We kept the Gamma distribution parameter B = 1 (upper case beta). For our three choices of A, the distribution functions have the specific forms:

$$\begin{aligned} \rho_{A=2}(y) &= ye^{-y} \\ \rho_{A=1}(y) &= e^{-y} \\ \rho_{A=1/2}(y) &= \frac{1}{\sqrt{\pi}} y^{-1/2} e^{-y} \end{aligned} \tag{3.26}$$

The mean of each function, $\bar{y}$, is just equal to A. The distribution for $A = 2$ is a gently peaked function with a maximum at $y = 1$. For $A = 1$, the familiar exponential has a maximum at 0 and falls monotonically; and for $A = 1/2$, the distribution diverges at the origin before falling more sharply than the other two. For these distributions, the quantity in (3.24) can be evaluated in closed form for the case of $\beta = 1/2$. For the other values of β, E(Q) was calculated numerically. The results of evaluating (3.24) are shown in figures 1, 2, and 3, for the three Gamma distributions, and in figure 4 there is a comparison of the three distributions on one plot, for the case of $\beta = -2$.

The first thing to notice is that in no case does the deviation of (3.22) from zero, as a ratio with the value of EQ, even reach unity. In many cases, the value is quite small indeed. Some trends can be seen. For the fixed distribution of y, equation (3.24) is nearest to zero, for all values of Γ, the closer that β is to its upper range value of 1. (At the point at which $\beta = 1$, the model becomes linear, and reduces to case 3.1.) It becomes increasingly worse as β becomes more negative. The value of the expression (3.24) at $\beta = -\infty$ can be derived analytically for each distribution. In this limit, there is no dependence on Γ; (3.24) becomes a step function, zero at $\Gamma = 0$ and a finite constant everywhere else. In this limiting case the CES function becomes the Leontief function. We show this maximum limit as a dashed line on the plots. The $\beta = -\infty$ values are: for Gamma distribution $A = 2$, -37%; for $A = 1$, -58%; and for $A = 0.5$, -94%.

Secondly, as the distribution of y becomes sharper (going from Gamma function with $A = 2$ to 1 and then 1/2), for any β, the percentage deviation from zero increased, reaching nearly 80% with $A = 1/2$ for $\beta = -5$ at its maximum.

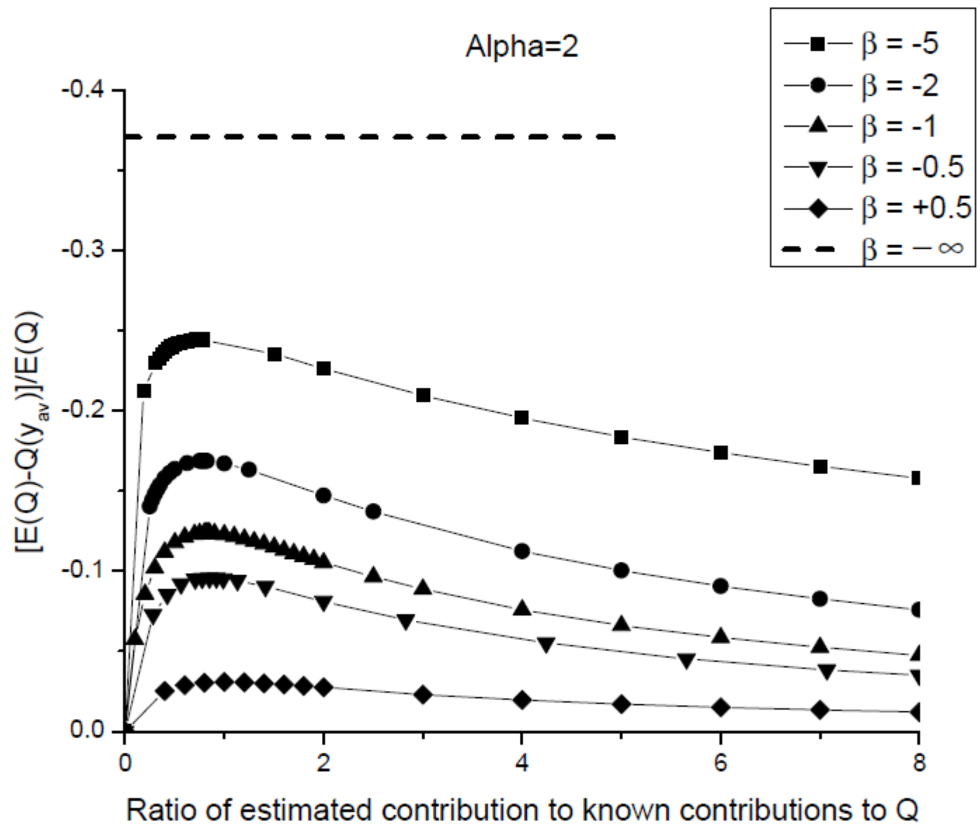

Figure 1: Calculation of ΔQ, Eq (3.24), as a function of the scaling variable Γ (Eq. 3.14), for a range of values of the CES exponent β. A Gamma distribution with $A = 2$ was used to model the variable y. The limit of negative-infinite β is marked as a dashed line. The closer ΔQ is to zero, the better the approximation of the estimator $\tilde{g}$. Large negative β give more inaccurate results with this approximation than small β do. Even $\beta = -5$ is still less than 25%.

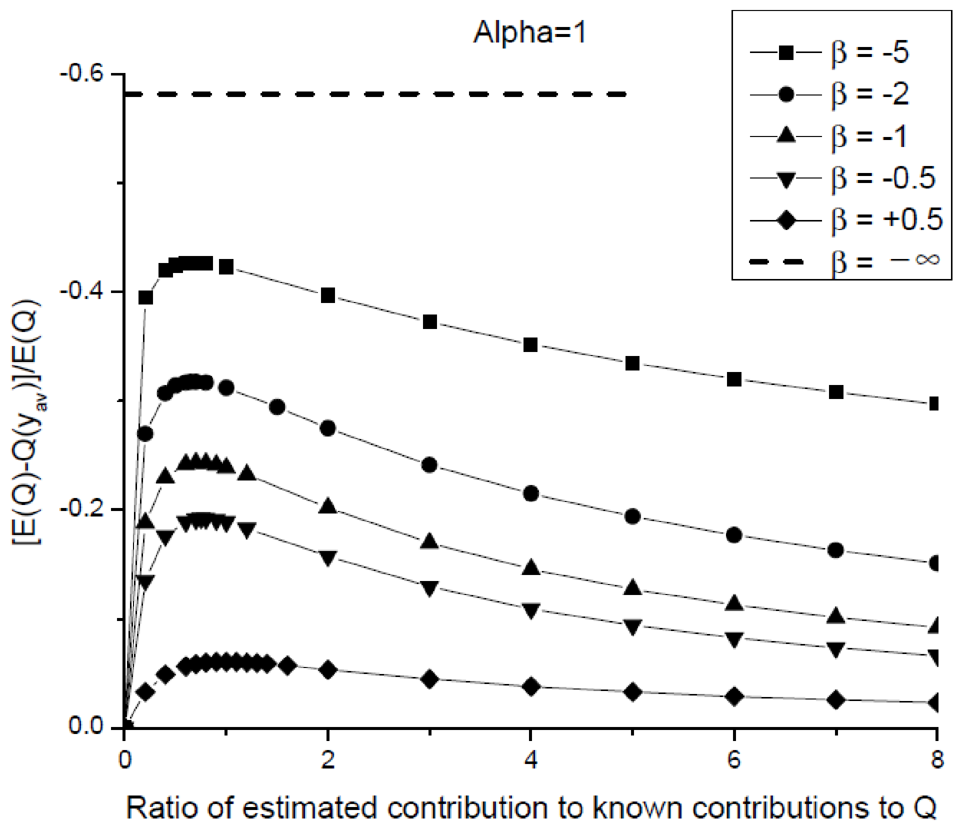

Figure 2: Calculation of ΔQ, Eq (3.24), as a function of the scaling variable Γ (Eq. 3.14), for a range of values of the CES exponent β. The limit of negative-infinite β is marked as a dashed line. A Gamma distribution with $A = 1$ was used to model the variable y. The closer ΔQ is to zero, t the better the approximation of the estimator $\tilde{g}$. Larger negative β give more inaccurate results than small β.

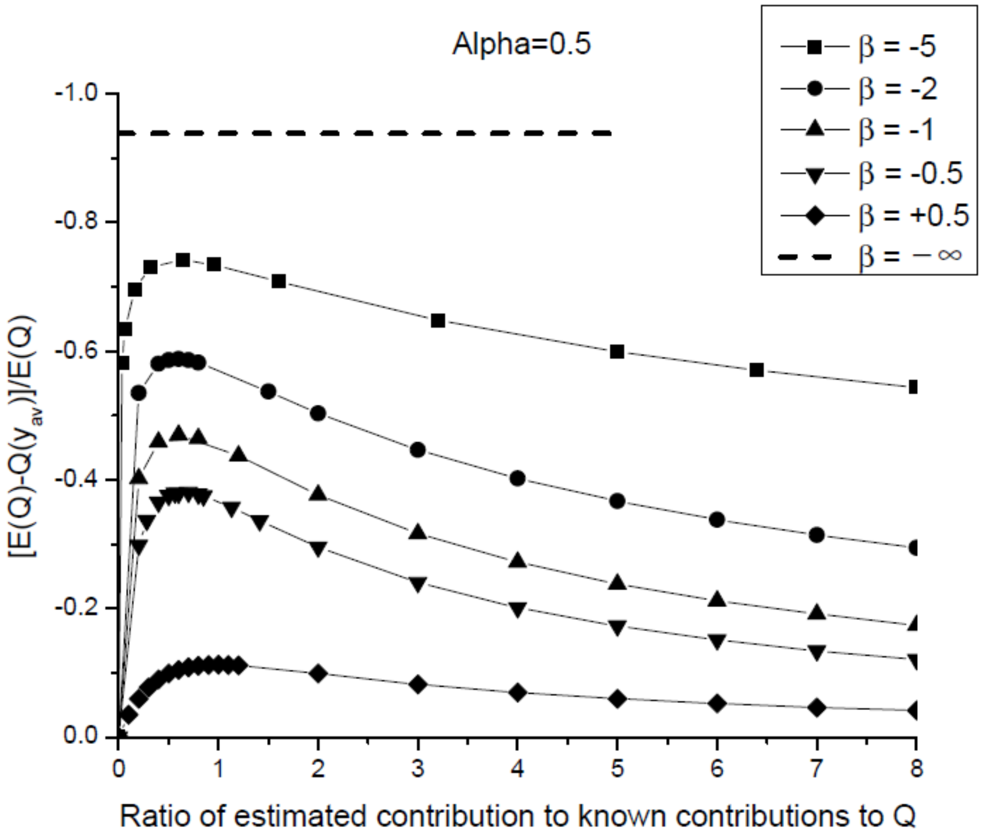

Figure 3: Calculation of ΔQ , Eq (3.24), as a function of the scaling variable Γ (Eq. 3.l4), for a range of values of the CES exponent β. The limit of negative-infinite β is marked as a dashed line. A Gamma distribution with $A = 0.5$ (which diverges at $y = 0$) was used to approximate the variable y. The smaller ΔQ , the better the approximation of the estimator $\tilde{g}$. Larger negative β are seen to give rising inaccurate results for this sharp distribution, with nearly 100% variance at β = -∞.

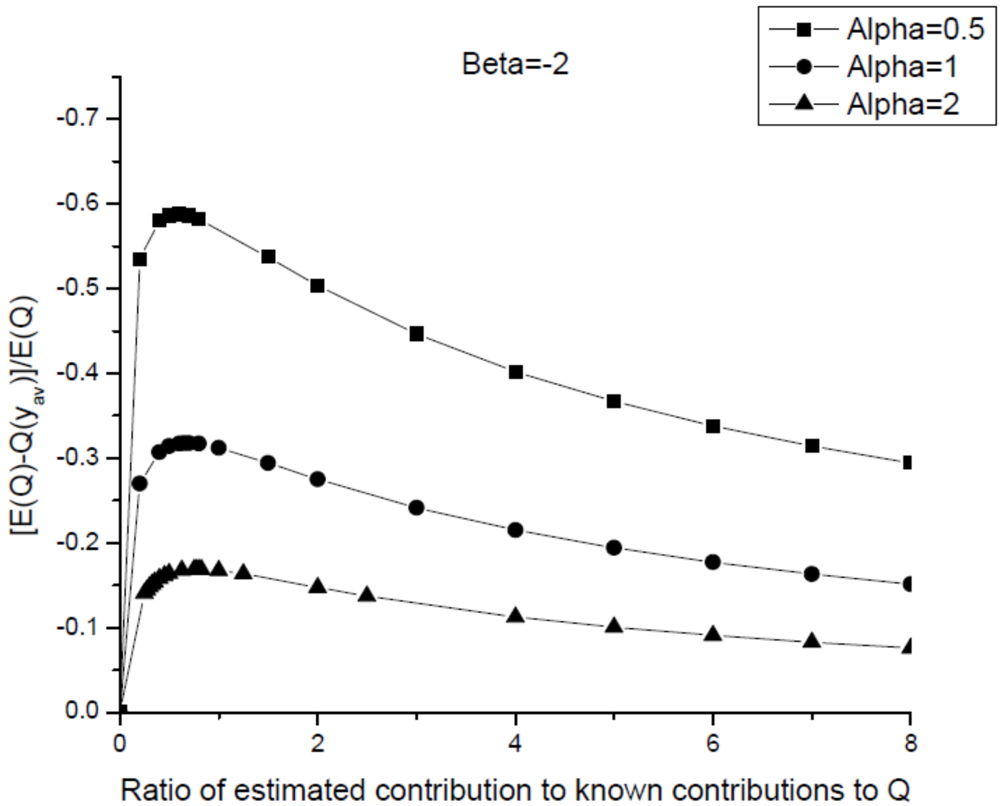

Figure 4: Calculation of ΔQ, Eq (3.24), as a function of the scaling variable Γ (Eq. 3.14), for a fixed value of the CES exponent $\beta = 2$, for a range of sharpness of the distribution of y. A Gamma distribution was used (Eq. 3.26), with values of parameter $A = 2$, 1, and 0.5. As A decreases, the distribution sharpens, with $A = 0.5$ divergent at $y = 0$. The closer ΔQ is to zero, the better the approximation of the estimator $\tilde{g}$. Smoother distributions clearly give more accurate results than sharper ones.

Finally, keeping the distribution and β fixed, and studying the deviation as a function of the ratio Γ, we found unexpected scaling behavior. With this choice of the independent variable, all the calculations of ΔQ, for all values of β and distributions of y considered, have a maximum at the nearly the same spot. This location is the value $\Gamma = 1$, at which point the sum of the x terms is equal to the average value of the y.

We therefore find that the approximation of neglecting the variable(s) y is worst when the measured and unmeasured variables have equal weight in the CES function. For cases when the average of the unmeasured quantities is smaller than the measured, the approximation is a very good one, and (3.24) rapidly goes to zero as the average of y decreases, as expected. However, (3.24) also decreases for cases in which the average of y is much larger than the contribution of the x. This result can be explained as the limiting case in which the contributing variables are essentially random, with a known distribution and average, and the measured quantities have very little effect.

To see whether these results are specific to the Gamma distribution, we also tried another distribution, a Normal distribution defined for only positive values, and hence peaked at the origin, but with a zero slope, and then a rapid fall-off. Hence in shape, the distribution falls between $A = 1$ and $A = 2$ of the Gamma distribution. Because of the dimensionless value of the integrand, one can show that width of the distribution cancels out, and that the results are independent of the standard deviation of the distribution. The results for the normal distribution are shown in figure 5, including a dashed line for $\beta = -\infty$.

One can see a considerable similarity with the $A = 1$ and $A = 2$ Gamma distributions, with the maximum for the normal distribution reaching a value about midway between the two, and even the falloff for large Γ at the same rate. The value at $\beta = -\infty$ reaches -43%, again midway between the limiting values for $A = 1$ and $A = 2$. This gives some confidence to our conclusion that the smoother the distribution of y, the more accurate it is to use the estimator $\tilde{g}$. Moreover, similar distributions can give similar results, and so outside cases with distributions similar to ours may be able to fairly accurately use our figures to extract values for (3.24).

To give some numerical values, for a normal distribution of y, (3.24) has a peak value for $\beta = +0.5$ of only 5%; for $\beta = -0.5$ it is 13%, increasing with decreasing β until it reaches a value of 30% for $\beta = -5$. Similar results are obtained for a Gamma function distribution with parameter $A = 2$, a similarly smooth function as a normal distribution. For a more sharply peaked distribution, such as a Gamma function with parameter $A = 1/2$, (3.22) has a maximum value for $\beta = +0.5$ of 10%, increasing with decreasing β until it reaches a value of 75% for $\beta = -5$.

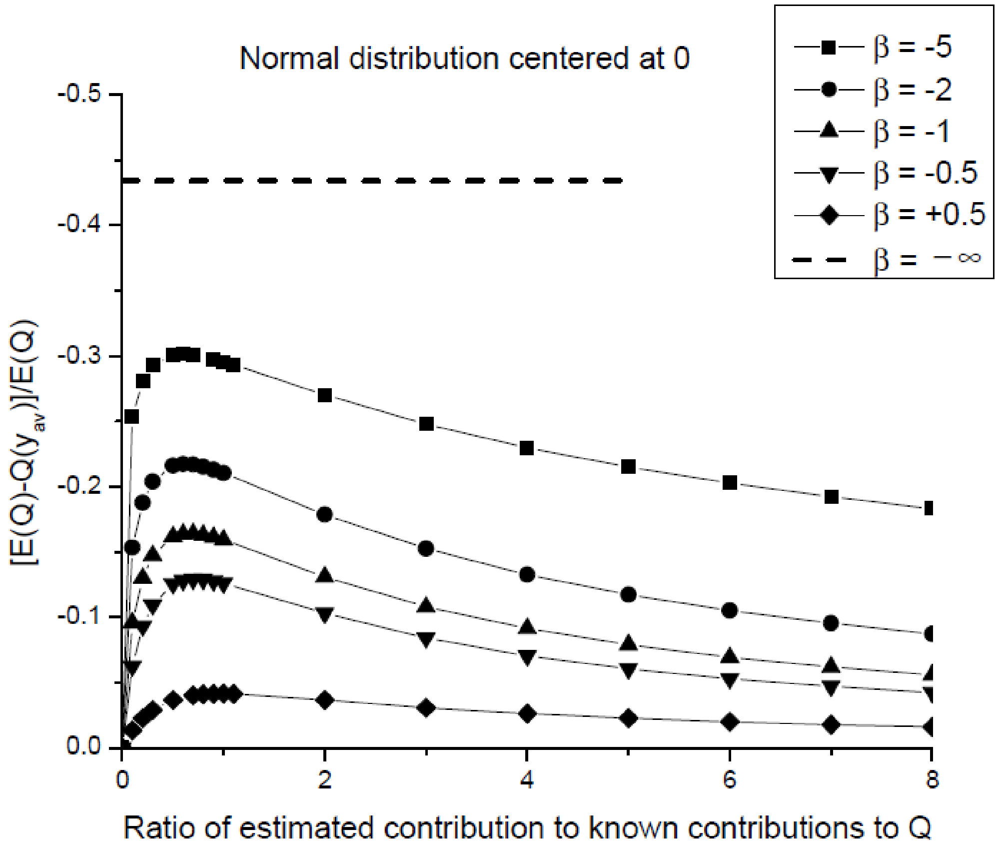

Figure 5: Calculation of ΔQ, Eq (3.24), as a function of the scaling variable Γ (Eq. 3.14), for a range of values of the CES exponent β. A Normal distribution centered at the origin (with $y \geq 0$) was used to model the variable *y*. The limit of negative-infinite β is marked as a dashed line. The smaller ΔQ, the better the approximation of the estimator $\tilde{g}$. Larger negative β give more inaccurate results with this approximation than small β do.

Our final set of distributions has a normal distribution with a nonzero offset from the origin. This is the more general form of the normal distribution, and has the form:

$$\rho_{\text{normal}}(y) = Ke^{-(y-\mu)^2/\sigma^2}, \text{ y>0} \tag{3.27}$$

Here *K* is the normalization constant, which has to be calculated numerically using:

$$K\int_0^\infty dy\, e^{-(y-\mu)^2/\sigma^2} = 1 \tag{3.28}$$

Making the change of variables $z = y/\sigma$, we can express K as:

$$K = \frac{1}{\sigma \int_0^\infty dz\, e^{-(z-\frac{\mu}{\sigma})^2}} \equiv \frac{1}{\sigma I(\mu/\sigma)} \tag{3.29}$$

We also need the average of y, $\overline{y}$:

$$\overline{y} = \frac{1}{\sigma I(\mu/\sigma)} \int_0^\infty dy\, y e^{-(y-\mu)^2/\sigma^2} \tag{3.30}$$

Again making the change of variables $z = y/\sigma$ in the upper integral, $\overline{y}$ can be expressed:

$$\overline{y} = \frac{\sigma}{I(\mu/\sigma)} \int_0^\infty dz\, z e^{-(z-\frac{\mu}{\sigma})^2} \equiv \sigma r(\mu/\sigma) \tag{3.31}$$

We then obtain for the expectation of Q from (3.18) for $\beta < 0$,

$$\begin{aligned} &\frac{EQ}{Q(\alpha_y = 0)} = \\ &\frac{1}{\sigma I(\mu/\sigma)} \int_0^\infty dy\, e^{-(y-\mu)^2/\sigma^2} \frac{y}{\left(\Gamma \sigma^{|\beta|} r^{|\beta|}(\mu/\sigma) + y^{|\beta|}\right)^{1/|\beta|}} \end{aligned} \tag{3.32}$$

Making the change of variables $z = y/\sigma$ finally gives for EQ

$$\frac{EQ}{Q(\alpha_y = 0)} = \frac{1}{I(\mu/\sigma)} \int_0^\infty dz\, e^{-(z-\mu/\sigma)^2} \frac{z}{\left(\Gamma r^{|\beta|}(\mu/\sigma) + z^{|\beta|}\right)^{1/|\beta|}} \tag{3.33}$$

So that EQ is seen to only depend on the ratio μ/σ, not on either separately. In similar manner from (3.21) the same conclusion can be reached for $\beta > 0$. From the equation (3.23), it is clear that $Q(\overline{y})$ is independent of either μ or σ. Hence the quantity of interest, (3.24), depends only on the ratio μ/σ.

The case of $\mu/\sigma = 0$ was covered above for the (half) Normal distribution centered at zero. Here we choose $\mu/\sigma = 1$ and 5 to give examples of distributions which are peaked just one sigma away from the origin, versus ones which have essentially the entire normal distribution contained on the positive side. To calculate (3.24), for each ratio of μ/σ we first numerically calculate $I(\mu/\sigma)$ and $h(\mu/\sigma)$ from equations (3.29) and (3.31), respectively. For interest, $\bar{y} = 1.11\sigma$ for $\mu/\sigma = 1$, and $\bar{y} = 5.0\sigma$ for $\mu/\sigma = 5$, showing how nearly the latter case is to a complete Normal distribution. For the zero-offset Normal distribution above, $\bar{y} = \sigma/\sqrt{\pi} = 0.56\sigma$.

The results of calculating quantity (3.24) for the two cases of μ/σ are shown in figures 6 and 7. We only show a few representative values of β, since by now the trends are clear.

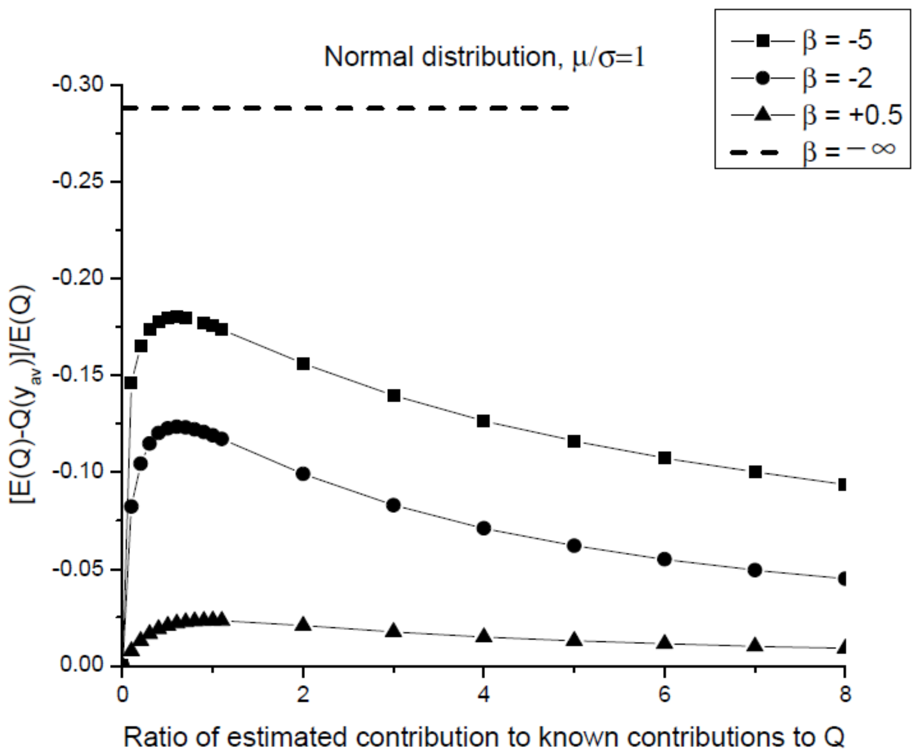

Figure 6: Calculation of ΔQ, Eq (3.24), as a function of the scaling variable Γ (Eq. 3.14), for a range of values of the CES exponent β. The limit of negative-infinite β is marked as a dashed line. A Normal distribution (with $y \geq 0$) with a center-to-width ratio of 1 was used to model the variable y. The smaller ΔQ, the better the approximation of the estimator $\tilde{g}$. Larger negative β give more inaccurate results than small β.

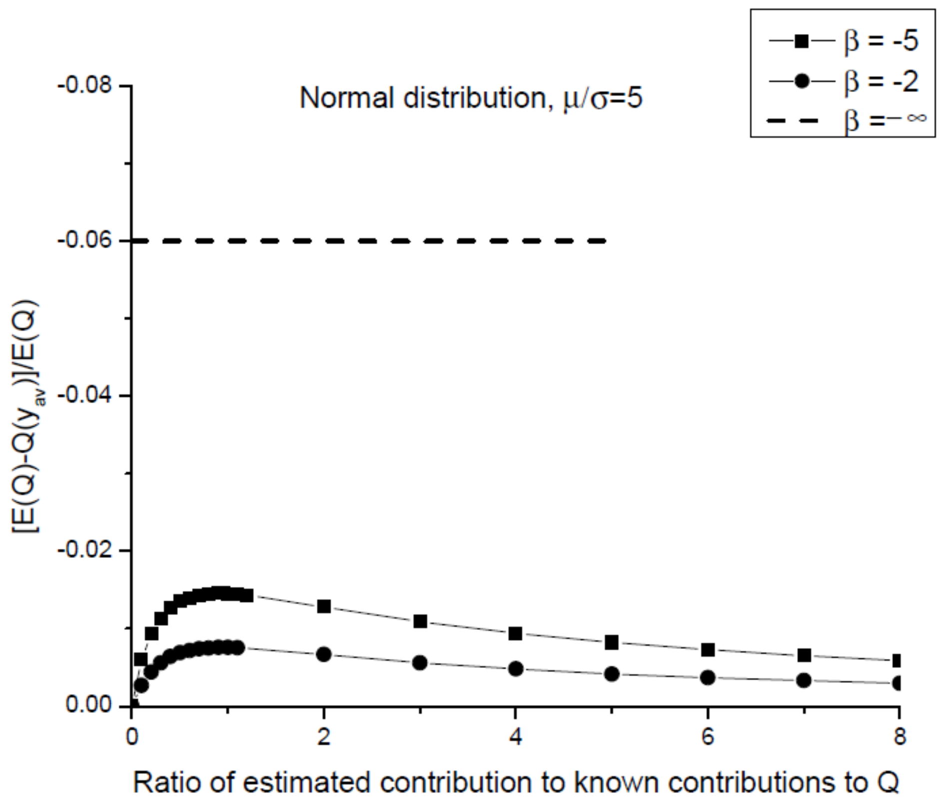

Figure 7: Calculation of ΔQ, Eq (3.24), as a function of the scaling variable Γ (Eq. 3.14), for a few representative values of the CES exponent β. The limit of negative-infinite β is marked as a dashed line. A Normal distribution (with $y \geq 0$) with a center-to-width ratio of 5 was used to model the variable *y*. The very small values of ΔQ, even at negative infinite β, show this distribution to give very accurate estimator $\tilde{g}$.

What is particularly surprising is the rate at which our quantity of merit (3.24) improves as the peak of the Normal distribution shifts away from the origin. At only $\mu/\sigma = 1$, already the percent deviation has fallen from a maximum of roughly 30% for $\beta = -5$ at $\mu/\sigma = 0$, to a maximum of 17.5% for $\mu/\sigma = 1$, and then down to a negligible 1.5% at $\mu/\sigma = 5$. For less negative β, as previous trends showed, the deviation is even smaller. We do not even show the $\beta = +0.5$ case for $\mu/\sigma = 5$, because it is so small. The limiting case of $\beta = -\infty$, again shown as a dashed line on each plot, is at 29% for $\mu/\sigma = 1$, and an amazingly small 6% for $\mu/\sigma = 5$.

These results imply that if the distribution of the unknown variable can be found to be Normal, the larger the ratio of center to width in the distribution, the better will be the approximation of the estimator $\tilde{g}$. In the limit that the distribution is centered sufficiently far from the origin that it is essentially completely contained on the positive side of the axis, then the deviations from the exact answer approach zero.

One further point to note is that in every case, the deviation from zero of ΔQ is negative. This is true for all distributions and values of the exponent β considered. Referring to our earlier equation 2.16, we see this is consistent with the fact that the CES function is concave for all β in the allowed range of < 1.

3.4.3. The Limit of β at Zero

There is an important issue which appears to be rarely addressed when discussing the CES function, especially general forms such as (3.11) we are using. This issue is whether this function is continuous and well-defined at all values of β in its range $\beta \leq 1$, and in particular at $\beta = 0$.

Examination of equation (3.11) defining the CES function will readily reveal that unless $\sum_{i=1}^{p} \alpha_i = 1$ the general CES function will be divergent and discontinuous as a function of the exponent β as it passes through zero. If the sum is less than one, the CES function will diverge as β approaches zero from below, and will approach zero as β approaches zero from above. The reverse is true when the sum greater than one. Only if the sum is unity will the CES function be defined at $\beta = 0$, and in that case it can be mapped to the Cobb-Douglass function. Therefore in many applications of the CES function this is indeed an assumption.

Setting the sum of α's to be one, even after the fact as we have done here, does not affect the results we have presented so far. It could be thought that such a restriction would eliminate variables, and cause simplifications. However, this is not the case. The CES function is still a function of all the other $(p - 1)$ α's, plus β. The variable Γ (3.14, 3.19) which works so well as a scaling variable for this problem is seen from its definition to be a complex function of all the α's and x's, plus the average of y. In general, Γ is a function of $(2p - 1)$ variables (all $(p - 1)$ x's and their α's, plus α_y). If we eliminate one variable due to a sum rule on the α's, then Γ is still a function of $(2p - 2)$ variables; it doesn't help much. Even if we go down to just two variables x and y, with a single α between them, Γ is still a nontrivial function: $\Gamma = \alpha \bar{y}^{\beta} / (1-\alpha) x^{\beta}$. All the plots in section 3.4.2 still hold; pick α and pick x, and you can get Γ and, finding that value on the x-axis, pick the value of ΔQ off the plot. All expressions for ΔQ continue to be very general, depending on the sharpness of the distribution of y.

3.4.4. Other Issues of Interest

In addition to the criterion for being able to use the estimator $\tilde{g}$, Eq(3.24), we also examined the behavior of the CES function with regard to constancy of x, an alternative test (Eq. 2.16). We did not find any regions in which the CES function met this criterion. As a function of x, ΔQ is zero at $x = 0$, and then rises with increasing x. For β greater than zero, it continues to rise with x, and for β less than zero, it reaches a maximum at an x around unity, and then decreases toward zero for very large x. We thus note that (2.14) is a more stringent criterion than (2.11).

A second question of note is the use of relative error in looking at ΔQ, (3.24), as we have been doing in the previous sections of numerical results. Relative error can become less useful in the range where the size of the data itself is small. Accordingly, we have directly calculated the value of EQ for the case of the smallest relative ΔQ, namely $\beta = +0.5$. The main question we want to answer is, is the value of (3.24) small for $\beta = +0.5$ because it really does have small deviations, or because we are inadvertently dividing by a large EQ?

To address this question and calculate the value of EQ, we direct our attention to equations (3.18) and (3.21), which are expressions for EQ for β less than zero and greater than zero, respectively, which can be calculated numerically as a function of the single parameter Γ. There however is the factor $Q(\alpha_y = 0)$ in the denominator, which prevents this from being a completely separate calculation of EQ. This factor is actually discontinuous as β approaches zero from above and below, depending on whether the sum of α_x is less than or greater than one. This is of course related to the point made in the previous section 3.4.3. If the sum of all the α's is taken to be one, then removing α_y will necessarily make the remaining sum less than one for positive α_y, meaning that $Q(\alpha_y = 0)$ will go to infinity as β approaches zero from below, and become zero as β approaches zero from above. Of course, $Q(\alpha_y = 0)$ appears in all three terms of (3.24), and so cancels out, which is why our results for ΔQ are smooth and finite. However, to calculate EQ directly, we must leave off $Q(\alpha_y = 0)$.

Without the factor of $Q(\alpha_y = 0)$ in the denominator, EQ becomes a function of many variables. To simplify matters and get a numerical answer, we focused on the case of only two variables, y and x, with coefficients α and $(1-\alpha)$. Then EQ is just a function of x and α, and we performed a matrix of calculations of EQ, spread over reasonable values of x and α, with α ranging from zero to 1 and x from zero to 10. (We used the Gamma distribution with A equal to two (3.26) for a well-behaved distribution of y's.)

We found that EQ takes a maximum value equal to x, in the limiting case that α tends to zero (as one would expect, since the CES function reduces to x if all the other coefficients are zero). However, in this limit the numerator of (3.24) is identically zero – there is no variance in the value of Q if all variables are known. For all other values of α and x, EQ is always a moderate value in the neighborhood of 0.5 to 3.0. We find ΔQ calculated in the same way thus retains its very small values, with or without being divided by EQ.

In general for all values of β, not just for β equals 0.5, in the simple case of just two variables, we find that EQ scales with x for small α, the limit when the numerator is always small. Also EQ becomes uniformly the average of y for α near one. For the more realistic case of a multiple of x's, finding all coefficients $\alpha_{x,i}$ at one or the other limit becomes increasingly more unlikely, and EQ is even more limited to range of values of order unity. We conclude that dividing by EQ is in no way a misleading way to normalize our quantities.

4. RELATED RESULTS IN RANDOMIZED EXPERIMENTS

4.1. Relationship to an Estimation Problem

To the best of our knowledge, we are not aware of any other study investigating the specific issue at hand. However, there is a related thread of work in statistical analysis of randomized experiments. A prominent application of randomized experiments is clinical trials in health care. In clinical trials for a treatment, a standard procedure is to select a sample of patients, assign a patient the treatment A or B (e.g., B can be a control, i.e., no treatment A) using some randomization method, then analyze what effect the treatment has on the sampled patients. There are a number of ways to analyze the treatment effect; see Lachin (1988) for an overview. One approach of our interest is the utilization of a regression model. Following the set-up of Gail et al.

(1984) but using our notation used throughout the present paper, let $\hat{f}_n\ (x, y)$ be the response of a patient, x = +1 or -1 be treatment A or B, y be a factor (usually called covariates in the clinical trial literature) that impacts the response, n = 1, …, N be the index of the patient in the sample. Gail et al. (1984) assumes that the response of the patient follows the model:

$$\mathrm{E}\hat{f}_n\ (x, y) = f(x, y) = h(\mu + \alpha x + \beta y)\,, \tag{4.1}$$

where h is the link function in generalized linear models. Our interest is to estimate the parameters μ, α, and β, and in particular α which represents the effect of the treatments. Denote the estimator of α by $\hat{\alpha}$.

Now assume that the covariate was not measured or ignored in the model. Then the analysis is based on the wrong model

$$\mathrm{E}\hat{f}_n\ (x, y) = h(\mu^* + \alpha^* x)\,, \tag{4.2}$$

Let $\hat{\mu}^*$ and $\hat{\alpha}^*$ denote estimators of μ^* and α^* respectively. The important issue then is under what conditions will $\hat{\mu}^*$ and $\hat{\alpha}^*$ be consistent with or unbiased with respect to the correct values of μ and α? Using the method of moments for estimation and due to the randomized assignment of treatments to patients in clinical trials, the sample average responses of the two groups of patients with treatments A or B converge, respectively, to

$$\mathrm{E}\left(\hat{f}_n\ (x, y) \mid x = 1\right) = \mathrm{E}_y h(\mu + \alpha + \beta y) \tag{4.3}$$

$$\mathrm{E}\left(\hat{f}_n\ (x, y) \mid x = -1\right) = \mathrm{E}_y h(\mu - \alpha + \beta y)\,, \tag{4.4}$$

where y is assumed to be i.i.d. with E(y) = 0 and E(y^2) < ∝, and E_y denotes expectation with respect to y, even though we think mistakenly that the sample average response is converging to (4.2). We can see that this problem is equivalent to the problem studied in the present paper, with y_0 = Ey = 0, x_1 = -1, x_2 = 1, except that the present paper is interested in the difference in the responses:

$$\begin{aligned}&g(x_1=-1,x_2=1,y_0=0)=\\&f(x_2=1,y_0=0)-f(x_1=-1,y_0=0)\end{aligned}, \tag{4.5}$$

corresponding to (2.1) above, while Gail et al. (1984) is interested in the parameter values of the model *h*.

In the special case of *h* being linear, i.e.,

$$h(\eta)=a\eta+b\,, \tag{4.6}$$

$\hat{\alpha}^*$, using the method of moments say, converges to the difference in the responses g(-1,1,0) as defined in (4.5). Our results for linear models (under separable functions above) are consistent with Theorem 3 in Gail et al. (1984) which states that $\hat{\mu}^*$ and $\hat{\alpha}^*$ converge to μ and α correctly in this case.

In the case of h being exponential, i.e.,

$$h(\eta)=c\exp(a\eta)+b\,, \tag{4.7}$$

Theorem 5 in Gail et al. (1984) states that only $\hat{\alpha}^*$ but not $\hat{\mu}^*$ converges correctly. This means that the quantity of interest in the present paper,

$$g(-1,1,0)=c\left[\exp(a(\mu+\alpha))-\exp(a(\mu-\alpha))\right], \tag{4.8}$$

will not be estimated correctly. To verify this using our results, we check that condition (2.12) is not true.

Noting Ey = 0, we have

$$\begin{aligned}&\mathrm{E}f(x,y)-f(x,\mathrm{E}y)\\&=c\left[\mathrm{E}\exp(a(\mu+\alpha x+\beta y))-\exp(a(\mu+\alpha x))\right]\end{aligned}. \tag{4.9}$$

For condition (2.12) to apply at x_1 = -1, x_2 = 1, and y_0 = 0, we need

$$\mathrm{E}\exp(a(\mu+\alpha+\beta y))-\exp(a(\mu+\alpha))=\mathrm{E}\exp(a(\mu-\alpha+\beta y))-\exp(a(\mu-\alpha))$$

$$\Leftrightarrow \mathrm{E}\left[\exp(a(\mu+\alpha+\beta y))-\exp(a(\mu-\alpha+\beta y))\right]=\exp(a(\mu+\alpha))-\exp(a(\mu-\alpha))$$

$$\begin{aligned}&\Leftrightarrow \mathrm{E}\left[\exp(a\beta y)\right]\left[\exp(a(\mu+\alpha))-\exp(a(\mu-\alpha))\right]=\\&\exp(a(\mu+\alpha))-\exp(a(\mu-\alpha))\end{aligned} \tag{4.10}$$

This is not true unless we have the trivial cases of $\beta = 0$ or $\alpha = 0$ or $y \equiv y_0 = 0$, which are precisely the conditions of Corollary 1 in Gail et al. (1984). In non-trivial cases (4.10) is not true and condition (2.12) is not satisfied.

Subsequent to Gail et al. (1984), other forms of *f* were studied using a similar set-up, for example, by Begg and Lagakos (1990) when *f* is a logistic regression model, and by Hauck et al. (1991) for estimating odds ratios. A recent review is in Fergusson et al. (2009).

4.2. Discussion

Comparing our study to the results in this thread of work in randomized experiments, we make the following observations.

1. Although the motivation of our work is quite different from those in randomized experiments, the spirit of both is the same and the conclusions are similarly positive.
2. The results of the present paper and the papers in randomized experiments are of course subject to the usual assumption of independent random samples. One can argue that in randomized experiments such as clinical trials the i.i.d. sample assumption is inherently easier to be satisfied, due to our own action of randomization and the fact that human subjects can relatively easily be chosen so that they are independent. In business analysis, independence is less straightforward but is still possible with careful selection of observation targets. In addition, if even such a basic scenario were not favorable, chances for more complicated situations would not have been good.
3. The results of the present paper overlap with the results in Gail et al. (1984) when *f* is linear.
4. We are able to derive a set of more general sufficient conditions (2.13-2.14), intuitively because we limit ourselves to the difference in *f* at two values of *x*, rather than estimating the model parameters.

5. For the same reason, we are able to obtain positive results for a wide class of functions (separable functions and polynomials of second degree), significantly extending the result for linear, separable functions.
6. We make no assumptions about the values of the measured variable or the distribution of the unmeasured variable, while Gail et al use a measured variable with values of +/- 1 (due to their focus on treatment analysis) and an average of zero for the unmeasured variable. Given their generalized linear model, all their results would still be valid with a finite value of the average, since this could be folded into the constant term. Their method is limited, however, to a limited number of values of the measured variable, in order that a solution for the coefficients can be found.
7. Similar to the literature in randomized experiments, we analyze some specific forms of *f*, but focus on functions that are commonly seen in business or economics (Cobb-Douglas and CES functions).
8. Gail et al, using their measure, find the biases are small for generalized linear models if:

   a) the coefficient of the unmeasured variable is small, or

   b) if the unmeasured variable does not vary much.

   We address (a) in much detail in the analysis of the CES model. If the coefficient of the unmeasured variable is small, this means our Γ (defined in 3.19) is small. From our plots we can see that of course at Γ equals zero, the bias is zero, but as that coefficient increases, the bias can increase rapidly, and in fact we measure the effect of the size of this coefficient over the entire range, from small to large. We show it is a function of the smoothness of the potential whether the bias stays small over the whole range, or whether it increases sharply for a sharp distribution.

   Similarly, we address (b) in more detail. If the unmeasured variable does not vary much, then it can be represented by its average plus small deviations – in other words a very smooth distribution of y. For the CES model, we show specifically how the bias can vary with the amount that the unmeasured variable varies. We find that even if the unmeasured variable varies, then there are ways in which the bias can still be small.

5. CONCLUDING REMARKS

We analyze the situation where we are interested in the difference in a performance function f at two different values of a variable of interest, x, while keeping other variables, y, constant, but in fact the y are randomly varying in our observations of the performance function. To compensate for the latter, we know that one way is to build a model that takes into account all the varying factors. We ask the question, if we were to assume that the y had remained constant (effectively ignoring them) and calculate the simple average difference of f at different values of x, would this be of any value? We find that the answer to this is yes, in a surprisingly broad range of circumstances.

Specifically, we show, for a general performance function, conditions under which the estimated difference is unbiased with respect to the correct difference at the average of y. Using these results we derive the particular conditions under which the above is true for certain forms of the performance function commonly seen in business and economic analysis, including separable functions, polynomial functions, Cobb-Douglas functions, and Constant Elasticity of Substitution functions. For the first three families of functions, the conditions do not seem very restrictive in practice:

1. For separable functions we require that either x has no impact on the cross x-y term, or that the y-part in the cross x-y term is linear.
2. For polynomial functions we require that the polynomial is second order or below.
3. For Cobb-Douglas functions we require the assumption of exponential noise.

For CES functions, the estimated difference is always biased and we study the relative magnitude of this bias numerically. Using a gamma distribution for y, we find that, under a wide range of parameter values for the y distribution and the CES function, the relative magnitude of the bias is reasonably small, so that the simple difference can be used as an approximation to the correct value. In particular, we find that the approximation is likely to be worse the more sharply peaked or divergent the distribution for y is, and the more nonlinear the model (large β for the CES case). Moreover, the approximation is worse when the contributions to the quantity of interest from both x and y (on average) are roughly the same. For most reasonably smooth distributions of y, the relative error do not much go beyond 40%, even for relatively large values of β. For data which are better fit by β in the range from -1 to 1, the deviations never get really large, and the assumption of ignoring the fact that y was not constant is a surprisingly good one.

Calculation of the deviation of the model, in the way in which we provide, can be used not only to provide a qualitative measure of the accuracy of the approximation, but to give even more quantitative valuation of the expectation of the quantity of interest, if the specific distribution of the unmeasured variable y is known.

In practice, our situation is not all that uncommon. For example, it happens when variables are not completely under our control, or some variables are very difficult or expensive to measure and so we do not have data on them, or we are using a set of historical data collected previously for some other purposes, and data on some variables that could impact our performance function were simply not collected, or we simply do not have the time or technical resources to develop a proper model taking into account all factors, but rather use the simplest approach of before-and-after differencing to do our analysis, assuming all other variables remain constant.

Our results are also useful in the following situation. Say we want to make use of an existing set of historical data to do an analysis that comes up after the data were collected. Our new analysis requires a variable that the existing data set does not provide and it may be difficult to go back to collect the detailed data of the variable. However, an average of the missing variable can be obtained or estimated without much effort. The results in this paper indicate that the average difference in the performance function at two levels of a variable measured is not a bad estimate for the correct difference at the average value of the variable unmeasured. While it is impossible to obtain the difference as a function of the unmeasured variable from such a data set, we can at least get an estimate at a single point of the unmeasured variable.

APPENDIX A. POLYNOMIAL MODELS OF HIGHER DEGREES

In this section we calculate the LHS of equation (4.6) when a polynomial model is of degree 3 or higher. The main result is stated as follows.

Proposition A.1. Let $h_{\vec{x}}(d)$ denote a polynomial in $\vec{x}$ of degree *d*. Using the same notations as in equation (4.5) and assuming that $\vec{y}$ is a random vector with finite moment of all orders ≤ *d*,

$\mathrm{E}f(\vec{x},\vec{y}) - f(\vec{x},\mathrm{E}\vec{y}) = h_{\vec{x}}(d-2)$, for $d \geq 3$.

Proof.

We prove by induction on *d*.

Case of *d* = 3:

From the definition of *f* in (4.5), we have

$$\mathrm{E}f(\vec{x},\vec{y}) - f(\vec{x},\mathrm{E}\vec{y}) = \sum_{\gamma\in I} a_\gamma \vec{x}^{\alpha}\left(\mathrm{E}\prod_{j=1}^{q} y_j^{\beta_j} - \prod_{j=1}^{q}(\mathrm{E}y_j)^{\beta_j}\right) \tag{A.1}$$

We calculate the summand for each γ based on the values of the exponents α and β, as shown in Table A.1. Combining the results in Table A.1, we see that the RHS of (A.1) is a polynomial in $\vec{x}$ of degree 1.

Case of *d* = *D*:

We now assume that the result is true for *d* = *D* and show that the result is true for *d* = *D* + 1. Let

$$I_0 = \{\gamma : \alpha_1 + \cdots + \alpha_p + \beta_1 + \cdots + \beta_q \le D + 1\},$$
$$I_1 = \{\gamma : \alpha_1 + \cdots + \alpha_p + \beta_1 + \cdots + \beta_q \le D\},$$
$$I_2 = \{\gamma : \alpha_1 + \cdots + \alpha_p + \beta_1 + \cdots + \beta_q = D + 1\}$$

From the definition of f, when $d = D + 1$,

$$f(\vec{x}, \vec{y}) = \sum_{\gamma \in I_0} a_\gamma \vec{x}^\alpha \vec{y}^\beta = \sum_{\gamma \in I_1} a_\gamma \vec{x}^\alpha \vec{y}^\beta + \sum_{\gamma \in I_2} a_\gamma \vec{x}^\alpha \vec{y}^\beta \tag{A.2}$$

Then,

$$\begin{aligned} \mathrm{E}f(\vec{x}, \vec{y}) - f(\vec{x}, \mathrm{E}\vec{y}) &= \sum_{\gamma \in I_1} a_\gamma \vec{x}^\alpha \left(\mathrm{E}\prod_{j=1}^{q} y_j^{\beta_j} - \prod_{j=1}^{q} (\mathrm{E}y_j)^{\beta_j} \right) \\ &+ \sum_{\gamma \in I_2} a_\gamma \vec{x}^\alpha \left(\mathrm{E}\prod_{j=1}^{q} y_j^{\beta_j} - \prod_{j=1}^{q} (\mathrm{E}y_j)^{\beta_j} \right) \end{aligned} \tag{A.3}$$

The first sum on the RHS of (A.3) is the case of $d = D$ and so is a polynomial of degree $(D – 2)$ by our induction assumption. We calculate the summand in the second sum of (A.3) in Table A.2. Combining the results in Table A.2, we see that the second sum in (A.3) is a polynomial of degree $(D – 1)$. Therefore (A.3) is a polynomial of degree $(D – 1)$.

<u>Table A.1</u>. Calculation of RHS of Equation (A.1) for $d = 3$

$\|\alpha\|$	$\|\beta\|$	Summand in equation (A.1)
3	0	$a_\gamma \vec{x}^\alpha (1-1) = 0$
2	1	$a_\gamma \vec{x}^\alpha (\mathrm{E}y_k - \mathrm{E}y_k) = 0$ (for some k, $1 \le k \le q$)
1	2	$a_\gamma x_l \left(\mathrm{E}\prod_{j=1}^{q} y_j^{\beta_j} - \prod_{j=1}^{q} (\mathrm{E}y_j)^{\beta_j} \right) = \text{constant} \times x_l$, for some l.
0	3	$a_\gamma \vec{x}^\alpha (1-1) = 0$
$\|\alpha\| + \|\beta\| \le 2$		Constant (see Table 4.1)

Table A.2. Calculation of Second Sum in RHS of Equation (A.3)

$\|\alpha\|$	$\|\beta\|$	Summand in second sum in equation (A.3)
$D+1$	0	0
D	1	0
$D-1$	2	Polynomial in $\vec{x}$ of degree ($D-1$)
$D-2$	3	Polynomial in $\vec{x}$ of degree ($D-2$)
…	…	…
1	D	Polynomial in $\vec{x}$ of degree 1
0	$D+1$	Constant

REFERENCES

Arrow, K.J., Chenery, H.B., Minhas, B.S., and Solow, R.M. (1961), "Capital-labor substitution and economic efficiency," *The Review of Economics and Statistics* 43, 3, 225-250.

Begg, M.D., and Lagakos, S. (1990), "On the consequences of model misspecification in logistic regression," *Environmental Health Perspectives* 87, 69-75.

Berndt, E.R. (1976), "Reconciling alternative estimates of the elasticity of substitution," *The Review of Economics and Statistics* 58, 1, 59-68.

Bollen, K.A. (1989), *Structural Equations with Latent Variables*, New York, NY: John Wiley.

Box, G.E.P., Hunter, W.G., and Hunter, J.S. (2005), *Statistics for Experimenter: Design, Innovation, and Discoverys*, 2nd edition, New York, NY: John Wiley.

Chirinko, R.S. (2008), "σ: The long and short of it," *Journal of Macroeconomics* 30, 2, 671-686.

Christensen, L.R., Jorgenson, D.W., and Lau, L.J. (1973), "Transcendental logarithmic production frontiers," *The Review of Economics and Statistics* 55, 1, 28-45.

Cobb, C., and Douglas, P. (1928), "A theory of production," *American Economic Review* 18, 1, 139-165.

Draper, N., and Smith, H. (1998), *Applied Regression Analysis*, 3rd Edition, New York, NY: John Wiley.

Dokov, S.P., and Morton, D.P. (2002), "Higher-order upper bounds on the expectation of a convex function," *Stochastic Programming* preprint, downloaded at http://edoc.hu-berlin.de/docviews/abstract.php?id=26677.

Dokov, S.P., and Morton, D.P. (2005), "Second-order lower bounds on the expectation of a convex function," *Mathematics of Operations Research* 30, 3, 662-677.

Dhrymes, P.J. (1967), "Adjustment dynamics and the estimation of the CES class of production functions," *International Economic Review* 8, 2, 209-217.

Fergusson, D., Ramsay, T., and Whitmore, G.A. (2009), "Bias in logistic regression due to omitted covariates," presented at the CANNeCTIN Biostatistics and Methodological Innovation Working Group Seminar Series on Advanced Issues in Clinical Trials Methodology, November 2009, downloaded at http://www.cannectin.ca/default.cfm?id=25.

Gail, M.H., Wieand, S., and Piantadosi, S. (1984), "Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates," *Biometrika* 71, 3, 431-444.

Goldberger, A.S. (1968), "The interpretation and estimation of Cobb-Douglas functions," *Econometrica* 35, 3-4, 464-472.

Hauck, W.W., Neuhaus, J.M., Kalbfleisch, J.D., and Anderson, S. (1991), "A consequence of omitted covariates when estimating odds ratios," *Journal of Clinical Epidemiology* 44, 1, 77-81.

Lachin, J.M. (1988), "Statistical properties of randomization in clinical trials," *Controlled Clinical Trials* 9, 289-311.

McCullagh, P., and Nelder, J.A. (1989), *Generalized Linear Models*, 2nd Edition, New York, NY: Chapman & Hall/CRC.

McFadden, D. (1963), "Constant elasticity of substitution production functions," *The Review of Economic Studies* 30, 2, 73-83.

Miller, E. (2008), "An assessment of CES and Cobb-Douglas production function," working paper 2008-05, Congressional Budget Office, United States of America.

Salem, H.H. (2004), "The estimation of the elasticity of substitution of a CES production function: Case of Tunisia," *Economics Bulletin* 28, 7, A1.

Sato, K. (1967), "A two-level constant-elasticity-of-substitution production function," *The Review of Economic Studies* 34, 2, 201-218.

Soda, G., and Vichi, E.G. (1976), "Least squares estimation of C.E.S. production function's non-linear parameters," *Proceedings of APL76 8th International Conference on APL*, New York, NY.

Uzawa, H. (1962), "Production functions with constant elasticities of substitution," *The Review of Economic Studies* 29, 4, 291-299.

Young, P. (2010). "Jackknife and Bootstrap resampling methods in statistical analysis to correct for bias," lecture notes, Physics Department, University of California – Santa Cruz, Santa Cruz, CA.

Zellner, A., Kmenta, J., and Dreze, J. (1966), "Specification and estimation of Cobb-Douglas production function models," *Econometrica* 34, 4, 784-795.