IDENTIFYING OUTLIERS IN MULTI-OUTPUT MODELS

KEVIN J. FOX

School of Economics, University of New South Wales, Sydney, NSW 2052, Australia. E-Mail: K.Fox@unsw.edu.au

ROBERT J. HILL

School of Economics, University of New South Wales, Sydney, NSW 2052, Australia. E-Mail: R.Hill@unsw.edu.au

W. ERWIN DIEWERT

Department of Economics, University of British Columbia, Vancouver, BC V6T 1Z1, Canada. E-Mail: diewert@econ.ubc.ca

August 2003

Abstract

A firm may be different from other firms either in terms of the mix or scale of its input-output vector. This paper develops separate mix and scale measures of dissimilarity, and shows that these can be additively aggregated into a measure of absolute dissimilarity. The mix measure is particularly relevant in the context of frontier analysis since it will identify the firms that exert the most influence on the resulting efficiency scores.

J.E.L. Classification Numbers: C44, C52

Keywords: Outliers, Dissimilarity Indexes, Extreme Observations, Frontier analysis, Efficiency

---

This paper has benefited from comments from two anonymous referees, Bert Balk, Quentin Grafton, Paul Wilson, and participants at the Georgia Productivity Workshop. We gratefully acknowledge the programming assistance of Richard Gerlach, and the computing assistance of Patrick Young and John Piggott.

1 Introduction

There are many practical applications which require economists to compare two observations. If each of these observations is a vector, then the problem of comparison is more complex than otherwise. Allen and Diewert (1981) introduced a criterion to measure the dissimilarity between vectors, which we develop to identify firms which are "outliers" in the context of efficiency analysis. Specifically, our method identifies firms which are extreme observations in at least one of two well-defined senses. While the method is applicable to the general problem of outlier detection, the application of the method is particularly attractive in contexts such as efficiency analysis.

Nonparametric efficiency analysis methods, based on constructing a best-practice frontier using linear programming techniques, yield neither the OLS residuals nor parameters which are typically used by outlier diagnostics; see e.g., Färe *et al.* (1994), Coelli *et al.* (1998), Hjalmarsson *et al.* (1996), Färe *et al.* (1992). Moreover, the detection of outliers can be complicated by the existence of multiple outputs. The results generated by frontier-based efficiency models are particularly sensitive to outliers, since frequently it is the outliers that define the frontier. Hence it is perhaps surprising that the detection of outliers has not received more attention in the efficiency measurement literature. One notable exception is a paper by Wilson (1993), which generalizes the outlier measure proposed by Andrews and Pregibon (1978) to the case of multiple outputs. Wilson also provides a useful survey of the outlier literature, while a detailed review of the theory of multi-output production models can be found in Färe and Primont (1995).

"Outliers" are observations which are different, in some sense, from the other observations in the sample. For example, Davies and Gather (1993) defined outliers as those observations which have a different distribution from some assumed distribution for the "non-outliers." However, there is no standard definition of what constitutes an outlier. Fieller (1993) notes that there are two common themes to proposed definitions: (a) outliers are extreme observations in the sample; and (b) they are observations that are sufficiently extreme as to have an apparently low probability of occurrence or are surprising in some other way, even when adjudged as the extremes of the sample.

Observations can be defined in multiple dimensions, so they may be different from one another due to e.g., measurement error in some subset or all dimensions, a difference in scale in all dimensions, or a difference in scale in only some dimensions. While it is not (usually) possible to determine whether or not measurement error is the reason for data anomalies from only observing the data, for conducting meaningful analysis it is important to be able to at least identify the observations which are different in some well-defined sense from the other observations.

The distinction between observations in terms of a (multidimensional) measure of their average relative scale and a measure of dissimilarity in a subset of dimensions appears to be an important one. The first concept can be described simply as one of scale. An observation may be described as a "scale outlier" if it is relatively larger (or smaller) in all, or many, dimensions than other observations. The second concept can be described as one of mix. An observation may be described as a "mix outlier" if it has an unusual combination in terms of the size of vector elements relative to other firms. Thus its scale across all dimensions may not be at all dissimilar from the average, yet across individual dimensions it may have a high variance in its scale relative to the average. Of course, an observation could be both a scale and a mix outlier. The mix measure is particularly relevant in the context of frontier analysis, as it may identify the firms that exert the most influence on the efficiency scores.

Our scale and mix measures of dissimilarity sum to a measure of absolute dissimilarity (Diewert, 2002). Hence, another way of thinking about our method is that we decompose an index of absolute dissimilarity into mix (or "relative dissimilarity") and scale components.

Our method does not draw a distinction between observations on the best-practice frontier and beneath the frontier, allowing measurement and other errors in all observations to become apparent. Hence, as in the approach of Wilson (1993), we treat observations symmetrically. This is not possible if a sensitivity-analysis approach to outlier detection is taken, e.g. if efficient observations are deleted and the effect of their absence on average efficiency scores is used as a criterion. While an observation on the interior of a frontier will not affect the efficiency scores of other observations, it will affect average efficiency scores. In addition, if the observation is an outlier in some sense, then its own efficiency evaluation may be misleading.

We emphasise that we do not propose a statistical test, or equivalent algorithm, for the acceptance or rejection of an observation as an outlier. This would be equivalent to having a test which could determine if e.g., an observation was measured with error. This is clearly impossible. The mechanical application of such a test could quite possibly lead to the deletion of legitimate observations which just happen to be "extreme" relative to other observations. The purpose of identifying extreme observations is so that they can then be checked for accuracy, or the results of some empirical analysis can be qualified by noting the existence of disparate observations. That is, the identification of outliers and deciding what to do with them are two separate tasks. The latter requires the researcher to take account of prior knowledge and the empirical context.

This paper is organised as follows. Section 2 introduces the concepts of mix, scale and absolute dissimilarity, and derives our measures of mix and scale dissimilarity as the only measures that satisfy reasonable sets of axioms for such measures. That is, they are uniquely defined relative to these axioms. The resulting measures are "economics" oriented, in the sense that the distances between two vectors must be invariant to changes in the units of measurement. We show that these measures can be added to yield a measure of absolute dissimilarity. Section 3 generalizes our method to the case when there are more than two vectors being compared. In section 4 we compare our method with that of Wilson (1993) using both mock data sets and the Charnes *et al.* (1981) data set. Section 5 concludes.

2 Mix, Scale and Absolute Dissimilarity

Suppose K firms use the same M inputs to produce N outputs. Let x_{ki} denote the input or output i used by firm k. If $i = 1, \ldots, M$, then x_{ki} is an input, while if $i = M+1, \ldots, M+N$ then x_{ki} is an output, and $x_{ki} > 0$ for $i = 1, \ldots, M+N$.

Consider the ordinary least squares (OLS) residual sum of squares obtained by running the following regression:

$$\ln\left(\frac{x_{ki}}{x_{ji}}\right) = \alpha + \varepsilon_i, \quad x_{ji}, x_{ki} > 0, i = 1, \ldots, M+N. \tag{1}$$

The regression in (1) can be visualised in a two-dimensional plot, with $\ln(x_{ki}/x_{ji})$ on the y-axis and $i = 1, \ldots, M+N$ on the x-axis. The coefficient α represents the mean of the dependent variable, and the regression equation consists of a horizontal line through the data at this point. Hence, α is a summary measure of the relative scale of the vectors being compared. The scale dissimilarity measure proposed in this paper is a function of $\hat{\alpha}$, the OLS estimate of α. Deviations from the horizontal line $\ln(x_{ki}/x_{ji}) = \hat{\alpha}$ are measured by the regression residuals, denoted by $\hat{\varepsilon}_i$, for $i = 1, \ldots, M+N$. Squaring and summing these residuals results in a measure of the variance of the differences in logs between corresponding elements of the two vectors. This is intuitively what we mean by a mix dissimilarity measure.

If price data also existed in such a production context, the vectors of prices could be added as additional (positive) "input" or "output" vectors, or alternatively, separate dissimilarity measures could be computed for the price indexes. The latter approach is probably preferable since typically the scale of price differences will be very different from the scale of quantity differences.

In what follows, these measures of dissimilarity are derived from axioms under which they are uniquely defined. Although our mix dissimilarity measure is likely to be more useful in many applications (including efficiency analysis), it is convenient to focus first on our scale dissimilarity measure.

2.1 Scale Dissimilarity

Six desirable axioms for a scale measure of dissimilarity, $s(x_j, x_k)$, between the input-output vectors of firms j and k, x_j and x_k, are listed below.

(s1) $s(x_k, x_k) = 0$.

(s2) $s(x_j, x_k) \geq 0$.

(s3) $s(x_j, x_k) = s(x_k, x_j)$.

(s4) $s(x_k, \delta x_k) = s(x_k, (1/\delta)x_k) = f[\max(\delta, 1/\delta)]$, where f is a strictly increasing function and δ is a positive scalar.[1]

(s5) $s(P[x_j, x_k]) = s(x_j, x_k)$, where $P[x_j, x_k]$ is a common permutation of the input-output vectors x_j, x_k, i.e., the vectors are rearranged in the same way.

(s6) $s(x_j, x_k) = s(\tilde{x}_j, \tilde{x}_k)$, where $\tilde{x}_{ji} = \lambda_i x_{ji}$, $\tilde{x}_{ki} = \lambda_i x_{ki}$ and λ_i is a positive scalar, for all i (invariance to changes in the units of measurement).

Axiom (s1) is a very basic requirement that the scale dissimilarity measure should equal zero if the two vectors being compared are identical in all elements. (s2) says that a scale dissimilarity measure should always be greater than or equal to zero. (s3) says that the scale score should not be dependent on the order in which the vectors are considered in calculating the scale dissimilarity. (s4) requires that small and large scale outliers are treated symmetrically. (s5) says that the scale dissimilarity measure should not depend on the ordering of the data. (s6) is required so that the results are independent of the units of measurement.

Consider the following definition of scale dissimilarity:

$$\text{Definition}: \quad s_{\mathcal{M}}(x_j, x_k) \equiv \left\{ \ln \left[\frac{\mathcal{M}(x_k)}{\mathcal{M}(x_j)} \right] \right\}^2, \tag{2}$$

[1] It follows from (s1) that $f(1) = 0$.

where $\mathcal{M}(x)$ denotes a symmetric mean of the components of x. It is assumed that a symmetric mean satisfies at least the following axioms:[2]

(M1) $\mathcal{M}(\lambda, \lambda, \ldots, \lambda) = \lambda$, where $\lambda > 0$.

(M2) $\mathcal{M}(\lambda x) = \lambda \mathcal{M}(x)$, where λ denotes a scalar.

(M3) $\mathcal{M}(Px) = \mathcal{M}(x)$ where Px is a permutation of the components of the input-output vector x.

(M4) $\mathcal{M}(x_j) < \mathcal{M}(x_k)$ if $0_{N+M} \ll x_j < x_k$.[3]

(M5) $\mathcal{M}(x)$ is continuous over $0_{N+M} \ll x$.

The rationale for the scale dissimilarity measure in (2) is that $\mathcal{M}(x_j)$ and $\mathcal{M}(x_k)$ are the average scale of firms j and k. Scale dissimilarity, therefore, is an increasing function of the ratio of their average scales. As long as $\mathcal{M}(x)$ satisfies (M1)-(M5), our scale dissimilarity measure in (2) will satisfy (s1)-(s5). Further restrictions, however, must be placed on $\mathcal{M}(x)$ if (s6) is to be satisfied. We return to this point later.

2.2 Mix Dissimilarity

Seven desirable axioms for a mix measure of dissimilarity, $m(x_j, x_k)$, between the input-output vectors of firms j and k are listed below.

(m1) $m(x_k, x_k) = 0$.

(m2) $m(x_j, x_k) \geq 0$.

(m3) $m(x_j, x_k) = m(x_k, x_j)$.

(m4) $m(\delta x_k, x_k) = m(x_k, \delta x_k) = 0$ where δ is a positive scalar.

(m5) $m(\lambda x_j, \mu x_k) = m(x_j, x_k)$ where λ and μ are positive scalars.

[2] A detailed discussion of the properties of symmetric means can be found in Diewert (1993).

[3] The notation 0_{N+M} denotes a column vector of zeros, with $N + M$ elements.

(m6) $m(P[x_j, x_k]) = m(x_j, x_k)$, where $P[x_j, x_k]$ is a common permutation of the input-output vectors x_j, x_k, i.e., the vectors are rearranged in the same way.

(m7) $m(x_j, x_k) = m(\tilde{x}_j, \tilde{x}_k)$, where $\tilde{x}_{ji} = \lambda_i x_{ji}$, $\tilde{x}_{ki} = \lambda_i x_{ki}$ and λ_i is a positive scalar, for all i (invariance to changes in the units of measurement).

Axiom (m1) is a very basic requirement that the mix dissimilarity measure should be equivalent to zero if the two vectors being compared are identical in all elements. (m2) says that a mix dissimilarity measure should always be non-negative, which avoids the problem of having to give an interpretation to a negative mix score. (m3) says that the mix score should be independent of the order in which the vectors are considered. (m4) is a generalisation of (m1). It says that the mix dissimilarity measure must have zero value if the two vectors are proportional to each other. That is, if each element of one vector varies by the same proportion relative to the corresponding element in the other vector, then the vectors have the same "mix" and so should have a score of zero. (m5) states that the mix score must be unaffected by rescaling of either of the vectors. This is required since the "mix" of a vector should not depend on its scale. (m6) says that a reorganisation of the elements of both vectors in the same fashion should not affect the mix dissimilarity. That is, it should not matter in which order the data are organised in the vectors. (m7) is required so that the results are independent of the units of measurement of the data.

Consider the following definition of mix dissimilarity:

$$\text{Definition :} \quad m_{\mathcal{M}}(x_j, x_k) \equiv \frac{1}{M+N} \sum_{i=1}^{M+N} \left[\ln\left(\frac{x_{ki}}{x_{ji}}\right) - \ln\left(\frac{\mathcal{M}(x_k)}{\mathcal{M}(x_j)}\right) \right]^2, \tag{3}$$

where $\mathcal{M}(x)$ again denotes a symmetric mean.

The rationale for this mix dissimilarity measure is that it measures the extent to which the ratios of individual inputs or outputs used by the two firms differ from the average ratio. As long as $\mathcal{M}(x)$ satisfies (M1)-(M5), our mix dissimilarity measure in (3) will satisfy axioms (m1)-(m6). Further restrictions must be placed on $\mathcal{M}(x)$ if (m7) is to be satisfied.

2.3 Ideal Specifications of Mix and Scale Dissimilarity

We now demonstrate that there exist unique expressions for the theoretical scale and mix dissimilarity measures introduced in the previous sections. These expressions are "ideal" in the sense that they are the only ones for which the mix and scale measures satisfy all their respective axioms. This is shown through the following two theorems.

Theorem 1: Suppose the symmetric mean function $\mathcal{M}(x)$ satisfies axioms (M1) to (M5) and the scale dissimilarity measure $s_{\mathcal{M}}(x_j, x_k)$ is defined by (2). Suppose further that $s_{\mathcal{M}}(x_j, x_k)$ satisfies the invariance to changes in the units of measurement axiom (s6). Then the mean function $\mathcal{M}(x)$ must be the geometric mean function, $\mathcal{M}_G(x)$ defined as follows:

$$\mathcal{M}_G(x) \equiv \prod_{i=1}^{M+N} x_i^{1/(M+N)}. \tag{4}$$

Proof: Since $s_{\mathcal{M}}(x_j, x_k)$ defined by (2) satisfies the axiom (s6), then by setting $\lambda_i = 1/x_{ji}$ for $i = 1, \ldots, M+N$, we obtain the following equation for all $x_j \gg 0_{M+N}$ and $x_k \gg 0_{M+N}$:

$$\left\{\ln\left[\frac{\mathcal{M}(x_k)}{\mathcal{M}(x_j)}\right]\right\}^2 = \left\{\ln\left[\frac{\mathcal{M}(x_{j1}^{-1}x_{k1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N})}{\mathcal{M}(1, \ldots, 1)}\right]\right\}^2$$

$$= \{\ln[\mathcal{M}(x_{j1}^{-1}x_{k1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N})]\}^2, \tag{5}$$

using (M1).

Case 1: $\mathcal{M}(x_j) = \mathcal{M}(x_k)$.

In this case, the left-hand side of (5) is 0 and hence the right-hand side must also be 0 and so we must have $\ln[\mathcal{M}(x_{j1}^{-1}x_{k1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N})$ equal to 0. Thus in this case, we have

$$\frac{\mathcal{M}(x_k)}{\mathcal{M}(x_j)} = \mathcal{M}(x_{j1}^{-1}x_{k1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N}). \tag{6}$$

Case 2: $\mathcal{M}(x_k) = u_k > u_j = \mathcal{M}(x_j) > 0$. In this case, we have $\mathcal{M}(x_k)/\mathcal{M}(x_j) > 1$, or $\ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)] > 0$. Taking positive square roots on both sides of (5) leads to the following equation:

$$\ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)] = [\{\ln \mathcal{M}(x_{j1}^{-1}x_{k1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N})\}^2]^{1/2}. \tag{7}$$

Now let x_k^* move in a continuous path along the level surface with height u_k from the point x_k to the point $u_k 1_{M+N}$. Let x_j^* move in a continuous path along the level surface with height u_j from the point x_j to the point $u_j 1_{M+N}$. Using axioms (M1), (M4) and (M5) on $\mathcal{M}(x)$, it can be seen that the left-hand side of (7) will remain constant at a value equal to $\ln[u_k/u_j] > 0$. Hence the right-hand side of (7) must also remain constant and be equal to the same positive value, $\ln[u_k/u_j]$, which is also equal to $[\{\ln \mathcal{M}(u_k/u_j, \ldots, u_k/u_j)\}^2]^{1/2} = [\{\ln[u_k/u_j]\}^2]^{1/2}$ using axiom (M1). Thus

$$|\ln \mathcal{M}(x_{j1}^{-1}x_{k1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N})| = |\ln \mathcal{M}(u_k/u_j, \ldots, u_k/u_j)|$$

$$= |\ln[u_k/u_j]| = \ln[u_k/u_j] > 0. \tag{8}$$

Equation (8) and the continuity of $\mathcal{M}(x)$ show that $\ln \mathcal{M}(x_{j1}^{-1}x_{j1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N})$ cannot be negative and hence the right hand side of (7) must be equal to the positive number $\ln \mathcal{M}(x_{j1}^{-1}x_{k1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N})$. Exponentiating both sides of the resulting equation shows that (6) must hold in this case.

Case 3: $\mathcal{M}(x_j) = u_j > u_k = \mathcal{M}(x_k) > 0$. Case 3 is analogous to case 2. Again we find that (6) must hold for this case.

Equation (6) is equivalent to the following functional equation: for all x_j and $x_k \gg$

0_{M+N}:

$$\mathcal{M}(x_k) = \mathcal{M}(x_j)\mathcal{M}(x_{j1}^{-1}x_{k1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N}). \tag{9}$$

Now define the positive variables $z_i \equiv x_{ji}^{-1}x_{ki}$ for $i = 1, \ldots, M+N$ and note that $x_{ki} = x_{ji}z_i$ for $i = 1, \ldots, M+N$. Substituting these two sets of equations into (9) leads to the following functional equation:

$$\mathcal{M}(x_{j1}z_1, \ldots, x_{j,M+N}z_{M+N}) = \mathcal{M}(x_j)\mathcal{M}(z) \text{ for all } x_j \gg 0_{M+N} \text{ and } z \gg 0_{M+N}. \tag{10}$$

Using (10), axiom (M1) on $\mathcal{M}(x)$ and Theorem 3.6.7 in Eichhorn (1978; 67), we can deduce that $\mathcal{M}(x)$ must have the following Cobb-Douglas functional form:

$$\mathcal{M}(x_1, \ldots, x_{M+N}) = \prod_{i=1}^{M+N} x_i^{\alpha_i}, \tag{11}$$

where the α_i are positive constants. However, axioms (M2) and (M3) on $\mathcal{M}(x)$ imply that these positive weights α_i are all equal to $1/(M+N)$. Thus (11) becomes (4), the equally weighted geometric mean function. Q.E.D.

Theorem 2: Suppose the symmetric mean function $\mathcal{M}(x)$ satisfies the axioms (M1)-(M5). Let the mix dissimilarity function, $m_{\mathcal{M}}(x_j, x_k)$, which depends on $\mathcal{M}(x)$, be defined by (3) above. If $m_{\mathcal{M}}(x_j, x_k)$ satisfies the invariance to changes in the units of measurement axiom (m7), then $\mathcal{M}(x)$ must be the unweighted geometric mean function $\mathcal{M}_G(x)$ defined earlier by (4).

Proof: Using definition (3) and the axiom (m7) with $\lambda_i = 1/x_{ji}$ for $i = 1, \ldots, M+N$, we find that $\mathcal{M}(x)$ must satisfy the following functional equation for all $x_j \gg 0_{M+N}$ and $x_k \gg 0_{M+N}$:

$$\frac{1}{M+N}\sum_{i=1}^{M+N}\left\{\ln\left[\frac{x_{ki}}{x_{ji}}\right] - \ln\left[\frac{\mathcal{M}(x_k)}{\mathcal{M}(x_j)}\right]\right\}^2$$

$$= \frac{1}{M+N} \sum_{i=1}^{M+N} \left\{ \ln \left[\frac{x_{ki}}{x_{ji}} \right] - \ln \left[\frac{\mathcal{M}(x_{j1}^{-1} x_{k1}, \ldots, x_{j,M+N}^{-1} x_{k,M+N})}{\mathcal{M}(1, \ldots, 1)} \right] \right\}^2 . \tag{12}$$

Using (M1) which implies $\mathcal{M}(1, \ldots, 1) = 1$, equation (12) simplifies to:

$$-2 \sum_{i=1}^{M+N} \ln(x_{ki}/x_{ji}) \{ \ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)] - \ln \mathcal{M}(x_{j1}^{-1} x_{k1}, \ldots, x_{j,M+N}^{-1} x_{k,M+N}) \}$$

$$+(M+N)\{ (\ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)])^2 - [\ln \mathcal{M}(x_{j1}^{-1} x_{k1}, \ldots, x_{j,M+N}^{-1} x_{k,M+N})]^2 \} = 0,$$

or

$$2[\ln \mathcal{M}_G(x_k) - \ln \mathcal{M}_G(x_j)] \{ \ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)] - \ln \mathcal{M}(x_{j1}^{-1} x_{k1}, \ldots, x_{j,M+N}^{-1} x_{k,M+N}) \}$$

$$= \{ (\ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)])^2 - [\ln \mathcal{M}(x_{j1}^{-1} x_{k1}, \ldots, x_{j,M+N}^{-1} x_{k,M+N})]^2 \}$$

$$= \{ \ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)] - \ln \mathcal{M}(x_{j1}^{-1} x_{k1}, \ldots, x_{j,M+N}^{-1} x_{k,M+N}) \} \{ \ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)]$$

$$+ \ln \mathcal{M}(x_{j1}^{-1} x_{k1}, \ldots, x_{j,M+N}^{-1} x_{k,M+N}) \}. \tag{13}$$

There are two cases to consider where equation (13) can be satisfied.

Case 1: Obviously, (13) will be satisfied if $\mathcal{M}(x)$ satisfies the following functional equation:

$$\ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)] - \ln \mathcal{M}(x_{j1}^{-1} x_{k1}, \ldots, x_{j,M+N}^{-1} x_{k,M+N}) = 0 \text{ for all } x_j, x_k \gg 0_{M+N}. \tag{14}$$

But (14) is equivalent to (6), which was considered in Theorem 1, and we thus have the result that $\mathcal{M}(x)$ must be the geometric mean function in (4).

Case 2: For some $x_j^* \gg 0_{M+N}$ and $x_k^* \gg 0_{M+N}$, we have

$$\ln[\mathcal{M}(x_k^*)/\mathcal{M}(x_j^*)] - \ln \mathcal{M}(x_{j1}^{*-1} x_{k1}^*, \ldots, x_{j,M+N}^{*-1} x_{k,M+N}^*) \neq 0. \tag{15}$$

Using the continuity axiom (M5), there will exist a neighbourhood around x_j^* and x_k^*

where the inequality (15) will continue to hold. Hence for x_j and x_k in this neighbourhood, we can divide both sides of (13) by the nonzero common factor $\ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)] - \ln \mathcal{M}(x_{j1}^{-1}x_{k1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N})$ and we find that $\mathcal{M}$(x) must satisfy the following functional equation in this neighbourhood:

$$2[\ln \mathcal{M}_G(x_k) - \ln \mathcal{M}_G(x_j)] = \ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)] + \ln \mathcal{M}(x_{j1}^{-1}x_{k1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N}),$$

or

$$\begin{aligned}\ln \mathcal{M}_G(x_k) &= \ln \mathcal{M}_G(x_j) + (1/2)\ln[\mathcal{M}(x_k)/\mathcal{M}(x_j)] \\ &+ (1/2)\ln \mathcal{M}(x_{j1}^{-1}x_{k1}, \ldots, x_{j,M+N}^{-1}x_{k,M+N}). \qquad (16)\end{aligned}$$

However, equation (16) can hold as an identity in the neighbourhood of x_j^* and x_k^* only if $\mathcal{M}(x) = \mathcal{M}_G(x)$.

Thus in both cases we find that $\mathcal{M}(x)$ must equal the equally weighted geometric mean $\mathcal{M}_G(x)$ defined by (4). Q.E.D.

Now define the scale measure $s^*(x_j, x_k)$ as $s_{\mathcal{M}}(x_j, x_k)$ defined in (2) with $\mathcal{M}(x)$ equal to the geometric mean $\mathcal{M}_G(x)$ defined in (4).

$$s^*(x_j, x_k) \equiv \left[\frac{1}{M+N}\sum_{i=1}^{M+N} \ln\left(\frac{x_{ki}}{x_{ji}}\right)\right]^2. \qquad (17)$$

From Theorem 1, it has been shown that the scale measure (17) satisfies axioms (s1)-(s6). In addition, a measure of scale dissimilarity of the form (2) satisfies these axioms if only if it is equal to $s^*(x_j, x_k)$ in (17).

Similarly, define the mix measure $m^*(x_j, x_k)$ as $m_{\mathcal{M}}(x_j, x_k)$ defined in (3) with $\mathcal{M}(x)$ equal to the geometric mean $\mathcal{M}_G(x)$ defined in (4).

$$m^*(x_j, x_k) \equiv \frac{1}{M+N}\sum_{l=1}^{M+N}\left[\ln\left(\frac{x_{kl}}{x_{jl}}\right) - \frac{1}{M+N}\sum_{i=1}^{M+N}\ln\left(\frac{x_{ki}}{x_{ji}}\right)\right]^2 \qquad (18)$$

From Theorem 2, it has been shown that the mix measure (18) satisfies axioms (m1)-(m7). In addition, a mix measure of dissimilarity of the form (3) satisfies these axioms if and only if it is equal to $m^*(x_j, x_k)$ in (18).

It also can be shown that $s^*(x_j, x_k) = \hat{\alpha}^2$, where $\hat{\alpha}$ is the OLS estimate of α in (1),[4] and that

$$m^*(x_j, x_k) = \frac{1}{M+N} \sum_{i=1}^{M+N} \hat{\varepsilon}_i^2,$$

where $\hat{\varepsilon}_i$ denotes the OLS residual on element i in equation (1). This metric was first proposed by Allen and Diewert (1981) to measure the dissimilarity between price vectors in an index number context. Alternatively, it can also be defined as follows:

$$m^*(x_j, x_k) = \frac{1}{M+N} |\hat{x}_k - \hat{x}_j|^2,$$

where $\hat{x}_{li} = \ln x_{li} - (\sum_{i=1}^{M+N} \ln x_{li})/(M+N)$, for $l = j, k$. In other words, $m^*(x_j, x_k)$ is proportional to the square of the Euclidean distance between the points $(\hat{x}_{j1}, \ldots, \hat{x}_{j,M+N})$ and $(\hat{x}_{k1}, \ldots, \hat{x}_{k,M+N})$, both of which lie on the hyperplane $\sum_{i=1}^{M+N} \hat{x}_{li} = 0$.

2.4 Absolute Dissimilarity

Using the scale measure from equation (17) and the mix measure from equation (18), we can decompose an absolute measure of dissimilarity, $AD(x_j, x_k)$, as follows:

$$\begin{aligned} AD(x_j, x_k) &\equiv \frac{1}{M+N} \sum_{i=1}^{M+N} \left[\ln\left(\frac{x_{ki}}{x_{ji}}\right)\right]^2 \\ &= \frac{1}{M+N} \sum_{l=1}^{M+N} \left[\ln\left(\frac{x_{kl}}{x_{jl}}\right) - \frac{1}{M+N} \sum_{i=1}^{M+N} \ln\left(\frac{x_{ki}}{x_{ji}}\right)\right]^2 \\ &\quad + \left[\frac{1}{M+N} \sum_{i=1}^{M+N} \ln\left(\frac{x_{ki}}{x_{ji}}\right)\right]^2 \\ &= m^*(x_j, x_k) + s^*(x_j, x_k). \end{aligned} \tag{19}$$

[4] Note that if $x_k = \lambda x_j$ (where λ is a scalar) in the regression equation (1), then $s^*(x_j, x_k) = \{\ln[\max(\lambda, \lambda^{-1})]\}^2$. Thus the transformation of s^* that would recover $\max(\lambda, \lambda^{-1})$ is $\exp[s^*(x_j, x_k)^{1/2}]$.

This is what Diewert (2002) calls the "log squared index of absolute dissimilarity." As with our mix and scale measures of dissimilarity, $AD(x_j, x_k)$ in (19) can be shown to satisfy a list of reasonable axioms for an aggregate measure of dissimilarity between two vectors (Diewert, 2002).[5]

It is interesting to note that our absolute, mix and scale dissimilarity measures could have been derived starting from a Jevons (1865, 1884) index framework. The Jevons quantity index, $Q^J(x_j, x_k)$ can be written as follows:

$$\ln Q^J(x_j, x_k) = \frac{1}{M+N} \sum_{i=1}^{M+N} \ln \left(\frac{x_{ki}}{x_{ji}} \right). \tag{20}$$

As a quantity index, it measures the scale of the vector x_k relative to x_j. Using (17) and (20), the Jevons quantity index can be related to our measure of scale dissimilarity:

$$[\ln Q^J(x_j, x_k)]^2 = s^*(x_j, x_k). \tag{21}$$

Similarly, using (18) and (20), the Jevon's quantity index can also be related to our measure of mix dissimilarity.

To summarise, we have specified measures for determining the mix and scale dissimilarity of vectors. These are *uniquely* determined relative to axioms which characterise desirable properties for such measures. An alternative derivation of these measures from index-number theory has been noted. Further, these measures which reflect different characteristics of dissimilarity can be additively aggregated into a known index of absolute dissimilarity.

[5]Note that (19) can be reorganised to yield the familiar variance decomposition, which is that variance equals the expected value of the square minus the square of the expected value.

3 Ranking Outliers

The mix and scale dissimilarity measures discussed thus far are bilateral in nature. This section shows how they can be developed to provide multilateral comparisons of input-output vectors for $K > 2$ firms. An alternative outlier detection method, due to Wilson (1993), is also discussed in this section. In the next section, these outlier detection methods are then compared.

3.1 Construction of Mix and Scale Rankings

Let $\mathcal{K}$ denote the full set of firms, $\mathcal{K} = \{1, \ldots, K\}$. Also let $\mathcal{K}_n^c$ denote a subset of n firms. The possible combinations of n firms are indexed by $c = 1, \ldots, C$ where $C = \begin{pmatrix} K \\ n \end{pmatrix}$. To construct a ranking of outliers based on the mix and scale dissimilarity measures, it is necessary to compare the input-output vectors of all firms with the same reference input-output vector. The ith element of the reference input-output vector denoted by x_{Ri} is obtained by taking the geometric mean of the ith element of the input-output vectors of the $K - n$ firms not in the subset $\mathcal{K}_n^c$, i.e.:

$$x_{Ri} = \left(\prod_{k \notin \mathcal{K}_n^c} x_{ki} \right)^{1/(K-n)} . \tag{22}$$

The mix and scale measures from (18) and (17) are then measured relative to the reference input-output vector, as follows:

$$m^*(x_R, x_k) = \frac{1}{M+N} \sum_{l=1}^{M+N} \left\{ \ln\left(\frac{x_{kl}}{x_{Rl}}\right) - \frac{1}{M+N} \sum_{i=1}^{M+N} \ln\left(\frac{x_{ki}}{x_{Ri}}\right) \right\}^2 , \tag{23}$$

$$s^*(x_R, x_k) = \left[\frac{1}{M+N} \sum_{i=1}^{M+N} \ln\left(\frac{x_{ki}}{x_{Ri}}\right) \right]^2 . \tag{24}$$

To detect mix outliers, proceed as follows.

1. Choose the value of n.

2. For all possible subsets $\mathcal{K}_n^c$, calculate $m^*(x_R, x_k)$ in (23) for all firms.

3. Calculate $K_n^c = (1/n)\sum_{k\in\mathcal{K}_n^c} m^*(x_R, x_k)$ for each subset $\mathcal{K}_n^c$. This yields $C = \begin{pmatrix} K \\ n \end{pmatrix}$ average mix scores, one for each combination of n observations.

4. Define the set of n outliers as the elements of $\mathcal{K}_n^c$ that solve the following optimization problem: $max_{c=1,\ldots,C}\{K_n^c\}$.

Using this algorithm we can determine the mix outlier if $n = 1$, the two mix outliers if $n = 2$, and so forth up to the (arbitrarily determined) maximum of n_{max} possible outliers. Similarly, scale outliers can be identified by replacing $m^*(x_R, x_k)$ with $s^*(x_R, x_k)$ in the above algorithm, and using (24) instead of (23). It is useful to consider groups of outliers, i.e. $n > 1$, due to the possibility of masking effects. Masking arises if there are groups of outliers. Hence, groups of outliers are hard to detect unless n is allowed to vary. However, as noted by Fieller (1993), "[o]ne of the fundamental difficulties in handling multiple outliers is that conventional definitions all essentially require prespecification of the number of outliers—extremeness of [n] outliers can only be assessed if [n] is known." As information on the exact number of outliers typically does not exist, it is therefore useful to consider different values of n.[6]

Rousseeuw and van Zomeren (1990) argued for adjusting the Mahalanobis distance to allow for the possibility of masking when trying to rank single observations as outliers. Our approach more directly addresses the masking problem by considering all possible combinations of groups of observations as potential outlier groups.[7]

[6]This strategy also takes into account the possibility of "swamping" (Fieller, 1976). "Swamping is said to occur when the nonregular observations cause nonoutliers to be identified as outliers" (Davies and Gather, 1993; 783).

[7]Software for calculating the mix, scale and absolute measures of dissimilarity (for different n) is available from the authors or the web site for the Economic Measurement Group at the University of New South Wales: www.economics.unsw.edu.au/research/EMG/EMGindex.htm.

3.2 The Wilson-Andrews-Pregibon (WAP) Measure

The measure of Andrews and Pregibon (1978) identifies as outliers those observations which contribute the largest proportion of the volume of the full data set. The Andrews-Pregibon (AP) measure is generalized to the case of multiple outputs by Wilson (1993). Let X denote the input matrix, with the first column being a column of ones, and Y the output matrix. The AP measure can be written as follows in the single output context:

$$\begin{aligned} R_n^c(X_*) &= [D_n^c|X_*'X_*|]|X_*'X_*|^{-1}, \\ &= R_n^c(X)[D_n^c(e'e)](e'e)^{-1}, \end{aligned} \tag{25}$$

where the subscript n denotes the number of observations (rows) deleted from $X_* = [XY]$, by the operator D_n^c, while the superscript c indexes a particular combination of n observations deleted from the set. Y is $K \times 1$, and e is the vector of OLS residuals of a regression of Y on X. The AP measure selects as outliers the n elements that solve the following optimization problem: $\min_{c=1,\ldots,C}\{R_n^c(X_*)\}$.

Wilson's generalization of this measure to the multi-output context ($N > 1$, $M > 1$) gives what we call the Wilson-Andrews-Pregibon (WAP) measure:

$$R_n^c(X_*) = R_n^c(X)[D_n^c|\Omega|]|\Omega|^{-1}, \tag{26}$$

where $\Omega = [e_p'e_q]$ for $p, q = 1, \ldots, K$, with e_p and e_q the OLS residuals from regressing Y_p and Y_q on X, respectively.

In the next section we compare the results from using the WAP outlier detection method with our method in order to highlight the extra information that our method can provide.

4 Dissimilarity Measures and DEA Models

This section compares the above outlier detection methods using two mock data sets, and an actual data set. It is demonstrated how the methods can be used to help identify errors in the data, such as data-entry error.

4.1 Illustrative Examples

We begin by comparing the mix, scale, absolute dissimilarity (AD) and Wilson-Andrews-Pregibon (WAP) outlier measures using two mock data sets. Each data set has only a single input and output. The first data set is depicted in Figure 1. The constant and nonincreasing returns to scale best-practice frontiers as determined by Data Envelopment Analysis (DEA), a linear-programming technique, are also depicted in Figure 1. Efficiency is measured either in the input direction (the distance an input vector has to be contracted until the frontier is reached), or the output direction (the distance an output vector has to be expanded until the frontier is reached).

The mix, scale and WAP outlier scores which correspond to the data of Figure 1, for the case where only a single observation is deleted, are given in Table 1. All three methods are able to identify observations as outliers even if they are beneath the best-practice frontier. While these observations do not affect the efficiency scores of other observations, their detection can alert researchers to their dissimilarity with other observations. If this dissimilarity is due to measurement or other errors, then it is possible to avoid incorrect conclusions regarding their efficiency.

According to both the scale and WAP measures, the biggest outlier in Figure 1 is observation 3. In contrast, the biggest mix outlier is observation 1. In the context of DEA, the most influential observation is 1, in the sense that removing this observation from the set will have the biggest impact on the relative efficiency scores (measured in the input direction) of the other observations in the sample. This is an important advantage of the mix measure since influential observations (in terms of forming part of the efficient frontier) in DEA will typically be mix outliers, although not necessarily scale and WAP outliers.

The rank correlation coefficients between the mix, scale and WAP measures are given in Table 2. The correlation coefficients range between 0.119 and 0.189. Hence at least for the data in Figure 1, there is little correlation between the mix, scale and WAP outlier rankings.

In Figure 2 we consider a special case where all the firms are located on the same ray. The results in Table 3 provide interesting insights into each of the measures. In this case the WAP measure is not defined because the $X_*'X_*$ matrix in (25) is singular. The mix scores are all zero, since the firms all lie on the same ray. Hence, absolute dissimilarity between the firms is given solely by the scale measure. Finally, the scale scores increase symmetrically as one moves away from the firm in the middle.

Returning to Figure 1, it is also interesting to compare the rankings obtained when $n = 1$ with the groupings of outliers obtained when n is allowed to vary. The results obtained when n is varied between 1 and 6 are shown in tables 4 and 5. For all three measures, the rankings in Table 1 differ quite substantially from the groupings in tables 4 and 5. This is because of masking. For example, according to the mix measure, observations 1, 2, 4 and 5 are a potential group of outliers. Such groupings of outliers can only be discerned by deleting multiple observations. However, there is a potential weakness that arises as a result of varying n. Intuitively, one would expect the scale measure to select both the largest and smallest scale observations as outliers in Figure 1. This is exactly what the scale measure does when $n = 1$. Unfortunately, when n is varied to control for masking, the scale measure starts selecting observations only at one or other extreme. Hence it is not clear that it is always desirable to control for masking. This point is equally applicable to the mix, AD and WAP measures. Of course, allowing n to vary has the additional disadvantage of dramatically increasing the number of computations.

4.2 Application to a Multi-Output Data Set

The outlier measures developed in this paper are illustrated here using the Charnes *et al.* (1981) data set, which consists of 5 inputs and 3 outputs for 70 public schools in the United States.[8]

The five inputs are:

(i) Education level of mother as measured in terms of percentage of high school graduates among female parents ($x1$).

(ii) Highest occupation of a family member according to a pre-arranged rating scale ($x2$).

(iii) Parental visit index representing the number of visits to the school site ($x3$).

(iv) Parent counselling index calculated from the data on time spent with child on school-related topics such as reading together, etc. ($x4$).

(v) Number of teachers at a given site ($x5$).

The three outputs are:

(i) Total reading score as measured by the Metropolitan Achievement Test ($y1$).

(ii) Total Mathematics Score as measured by the Metropolitan Achievement Test ($y2$).

(iii) Coopersmith Self-Esteem Inventory, intended as a measure of self-esteem ($y3$).

Pairwise plots of these data are given in Figure 3. From this figure we can see that there does not appear to be much of a masking problem, i.e., there does not appear to be a group of outliers of size $n < K/2$ which share similar attributes with each other, but not with the rest of the sample.

An inspection of Figure 3 also reveals the difficulties involved with attempting a visual identification of outliers when there are many variables (dimensions). In fact, there is a deliberate data-entry error in the figure. We challenge the reader to find this error using visual inspection of these plots. The answer will be given later in this section.

Table 6 reports the mix, scale, AD and WAP outlier rankings for the Charnes *et al.* data set (without the data entry error), obtained by deleting only a single observation. We see that scale dominates the AD ranking and that WAP gives quite a different

[8]For more details on the data set and its construction, see Charnes *et al.* (1981).

ranking from the other measures. Some rank correlation coefficients are given in Table 7. In absolute value, the correlation coefficients are all less than 0.15. In particular, the correlation between the WAP ranking and both the mix and scale ranking is very close to zero. These results confirm the previous results obtained using the mock data in Figure 1. Clearly, therefore, the different outlier measures are each capturing different aspects of the outlier problem.

The WAP ranking selects observations on the basis of their contribution to the volume of the full data set. Hence, like the AD measure, it does not provide much information regarding exactly how a particular observation differs from the others in the sample. In contrast, the mix and scale measures are very specific as to the reason why the observations are being selected as potential outliers.

The groupings of outliers obtained by varying n from 1 to 4 are shown in Table 8. The AD ranking is the same as the scale ranking. A comparison between tables 6 and 8 reveals that the results obtained from the Charnes *et al.* data set are not affected much by masking. In particular, the top four outliers in Table 6 for both the mix and scale outlier measures correspond exactly with the group of outliers in Table 8. This result is in stark contrast to the results obtained for the mock data in Figure 1, where the selected outliers were highly sensitive to the choice of n.

Outliers reported in Table 8 which are not on the frontier, using a standard non-increasing returns to scale DEA frontier (see e.g., Coelli et al., 1998), are observations 32, 33, 66 and 67. These observations would not have been identified as potential outliers if a sensitivity analysis approach was employed, as this approach is restricted to considering only firms on the frontier as potential outliers.

As noted in the introduction, deciding what to do with outliers depends on the underlying cause of their "extremeness." This cannot be determined by merely observing the data. If, for example, the cause is found to be data-entry error, then either the error should be corrected or the firm in question removed from the data set. However, the first step is to identify the outliers. Our mix and scale measures provide two new ways of doing this in multiple dimensions, and yield information about the way in which the observations are extreme.

As an illustration of this, we return to the deliberate data-entry error in the data plotted in Figure 3. The error is in $x1$ for observation 5. The actual value is 11.62, but it was entered as 21.62 — variable $x1$ for observation 4 is the similar number 24.96, so this type of error is not entirely unlikely. As will be obvious to the reader, this type of error is very hard to identify visually. Also, a statistical test or algorithm that results in the mechanical exclusion of observation 5 is obviously undesirable. Applying our dissimilarity measures to the data set with the error, we now get observation 5 as the highest ranked mix outlier (rather than observation 66 as in Table 6), with observation 59 still the most highly ranked outlier in terms of scale and absolute dissimilarity. A subsequent check of the data in this case should lead to the detection of the entry error.[9]

5 Conclusion

We have introduced two new and complementary methods for detecting extreme observations. Our methods provide a natural interpretation of the division of such outliers into mix and scale components. Other methods typically do not draw this distinction. The introduced methods are applicable to any outlier-detection context. An obvious application is to the detection of tax fraud.[10] However, they are particularly useful for detecting outliers in multi-output data sets. Also, in the context of frontier analysis, our mix and scale measures have the advantage of being able to detect outliers beneath the frontier.

[9]This example is not as contrived as it may seem. It was inspired by the actual experience of helping another researcher with the preliminary analysis of a large data set. Several data-entry errors were detected by both the mix and scale measures.

[10]In this context, outliers are defined over the space of tax characteristics, such as income, deductions, and exemptions. A useful reference on tax fraud is Andreoni *et al.* (1998).

REFERENCES

Allen, R.C. and W.E. Diewert (1981), "Direct Versus Implicit Superlative Index Number Formulae," *Review of Economics and Statistics* 63, 430-435.

Andrews, D.F. and D. Pregibon (1978), "Finding the Outliers that Matter," *Journal of the Royal Statistical Society, Series B* 40, 85–93.

Andreoni, J., B. Erard and J. Feinstein (1998), "Tax Compliance," *Journal of Economic Literature* Vol. XXXVI, No. 2, 818–860.

Charnes, A., W.W. Cooper and E. Rhodes (1981), "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through," *Management Science* 27(6), 668-697.

Coelli, T., D.S. Prasada Rao and G.E. Battese (1998), *An Introduction to Efficiency and Productivity Analysis*, Boston/London/Dordrecht: Kluwer Academic Publishers.

Davies, L. and U. Gather (1993), "The Identification of Multiple Outliers," *Journal of the American Statistical Association* 88, 782-792.

Diewert, W. E. (1993), "Symmetric Means and Choice Under Uncertainty," in W. E. Diewert and A. O. Nakamura, (ed.), Essays in Index Number Theory - Volume 1, Amsterdam, North Holland, 355-433.

Diewert, W.E. (2002), "Similarity and Dissimilarity Indexes: An Axiomatic Approach," Discussion Paper No. 02-10, Department of Economics, University of British Columbia.

Eichhorn, W. (1978), *Functional Equations in Economics*, London, Addison-Wesley.

Fieller, N.R.J. (1976), *Some Problems Related to the Rejection of Outlying Observations*, unpublished PhD. thesis, University of Hull, Department of Statistics.

Fieller, N.R.J. (1993), Discussion of Davies and Gather (1993), *Journal of the American Statistical Association* 88, 794-795.

Färe, R., S. Grosskopf and S-K. Li (1992), "Linear Programming Models for Firm and Industry Performance," *Scandinavian Journal of Economics* 94(4), 599–608.

Färe, R., S. Grosskopf and C.A.K. Lovell (1994), *Production Frontiers*, Cambridge: Cambridge University Press.

Färe, R., and D. Primont (1995), *Multi-Output Production and Duality: Theory and Applications*, Boston/London/Dordrecht: Kluwer Academic Publishers.

Hjalmarsson, L., S.C. Kumbhakar and A. Heshmati (1996), "DEA, DFA and SFA: A Comparison," *Journal of Productivity Analysis*, 7(2-3), 303-327.

Jevons, W.S. (1865), "Variations of Prices and the Value of Currency Since 1762," *Journal of the Royal Statistical Society* 28 (June), 294–325.

Jevons, W.S. (1884), *Investigations in Currency and Finance*, London: Macmillan. (Originally published in 1865).

Rousseeuw, P.J. and B.C. van Zomeren (1990), "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association* 85(411), 633-639.

Wilson, P.W. (1993), "Detecting Outliers in Deterministic Nonparametric Frontier Models with Multiple Outputs," *Journal of Business and Economic Statistics* 11(3), 319-323.

Table 1: Ranking of Outliers for $n = 1$ in the Mock Data of Figure 1

Mix	Mix Score	Scale	Scale Score	AD	AD Score	WAP	WAP Score
1	0.181	3	0.681	3	0.721	3	0.296
11	0.060	4	0.360	1	0.377	1	0.669
3	0.039	1	0.197	4	0.370	2	0.689
5	0.034	12	0.144	12	0.148	6	0.718
2	0.034	6	0.125	6	0.147	11	0.722
9	0.025	10	0.068	11	0.081	4	0.770
6	0.021	7	0.022	10	0.069	5	0.816
4	0.010	11	0.021	5	0.050	9	0.821
12	0.004	5	0.016	2	0.039	12	0.844
7	0.001	2	0.005	9	0.026	10	0.863
8	0.001	9	0.001	7	0.024	7	0.879
10	0.000	8	0.000	8	0.002	8	0.913

Note: n is the size of the group of outliers, and observations have been ordered according to their mix, scale and absolute dissimilarity (AD) scores respectively. WAP stands for the measure of Andrews and Pregibon (1978), generalized by Wilson (1993), which is a negative objective function.

Table 2: Rank Correlation Coefficients for Mock Data in Figure 1

	Mix,Scale	Mix,WAP	Scale,WAP
correlation	0.119	0.119	0.189

Table 3: Scores for $n = 1$ in the Mock Data of Figure 2

	Mix Score	Scale Score	WAP Score
1	0	0.259	—
2	0	0.193	—
3	0	0.133	—
4	0	0.081	—
5	0	0.039	—
6	0	0.017	—
7	0	0.000	—
8	0	0.017	—
9	0	0.039	—
10	0	0.081	—
11	0	0.133	—
12	0	0.193	—
13	0	0.259	—

Note: See the note to Table 1.

Table 4: Groups of Outliers in the Mock Data of Figure 1: Mix and Scale

n	Mix	Mix Score	Scale	Scale Score
1	1	0.181	3	0.681
2	1,5	0.124	3,12	0.481
3	1,5,2	0.115	3,12,10	0.401
4	1,5,2,4	0.107	3,12,10,11	0.363
5	1,5,2,4,10	0.095	3,12,10,11,2	0.339
6	1,5,2,4,10,7	0.087	1,4,5,6,7,9	0.332

Note: n is the size of the group of outliers. The mix and scale scores are the average for the n observations, and the objective is to find the largest values of the scores for each n.

Table 5: Groups of Outliers in the Mock Data of Figure 1: AD and WAP

n	AD	AD Score	WAP	WAP Score
1	3	0.721	3	0.296
2	3,12	0.506	3,1	0.174
3	3,10,12	0.424	3,1,11	0.096
4	3,10,11,12	0.402	3,1,11,12	0.048
5	2,3,10,11,12	0.369	3,1,11,2,5	0.017
6	3,8,9,10,11,12	0.362	3,2,4,6,10,12	0.005

Note: n is the size of the group of outliers. The absolute dissimilarity (AD) scores are the sum of the mix and scale scores, and the objective is to find the largest values of the scores for each n. The WAP scores are minimum values.

Table 6: Ranking of Outliers for $n = 1$ in Charnes et al. (1981) Data

Rank	Mix	Scale	AD	WAP
1	66	59	59	59
2	48	32	32	44
3	15	69	69	33
4	56	5	5	66
5	69	62	62	35
6	49	44	44	54
7	68	29	29	68
8	5	61	61	67
9	61	38	48	8
10	67	48	38	50
11	51	45	54	1
12	32	54	45	52

Note: n is the size of the group of outliers.

Table 7: Rank Correlation Coefficients for Charnes et al. (1981) Data

	Mix,Scale	Mix,WAP	Scale,WAP
correlation	0.149	-0.002	0.014

Table 8: Groups of Outliers in Charnes et al. (1981) Data

n	Mix	Mix Score	Scale	Scale Score	WAP
1	66	0.218	59	2.64	59
2	15,48	0.221	59,32	2.04	59,44
3	15,48,56	0.224	59,32,69	1.84	59,44,33
4	15,48,56,66	0.225	59,32,69,5	1.69	59,44,66,67

Note: n is the size of the group of outliers, and the mix and scale scores are the average for the n observations. The results for the WAP measure are taken from Wilson (1993).

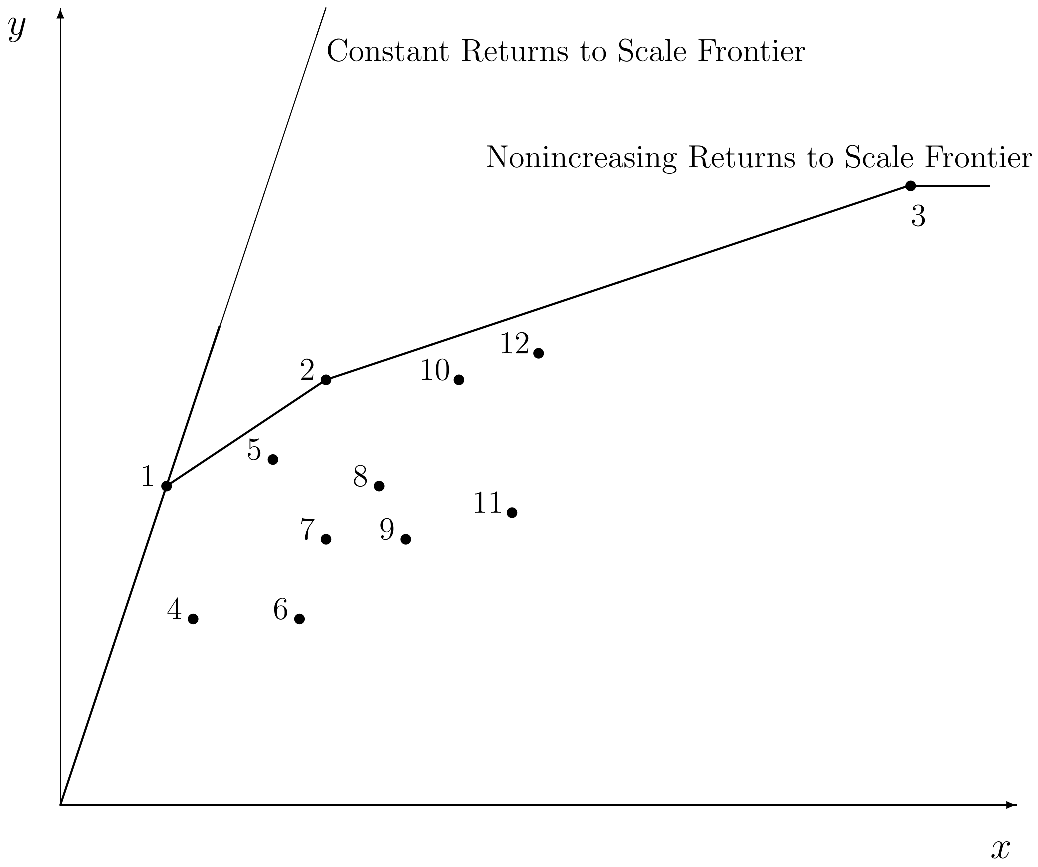

Figure 1: Mock Data I

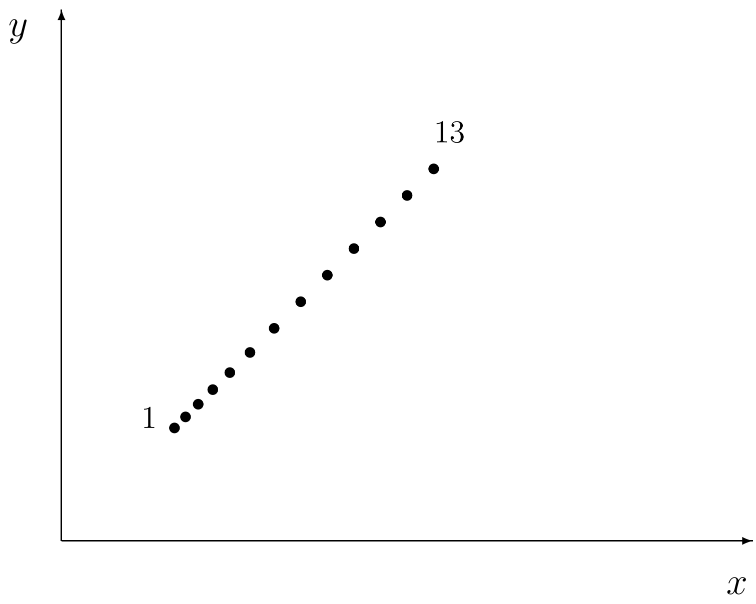

Figure 2: Mock Data II

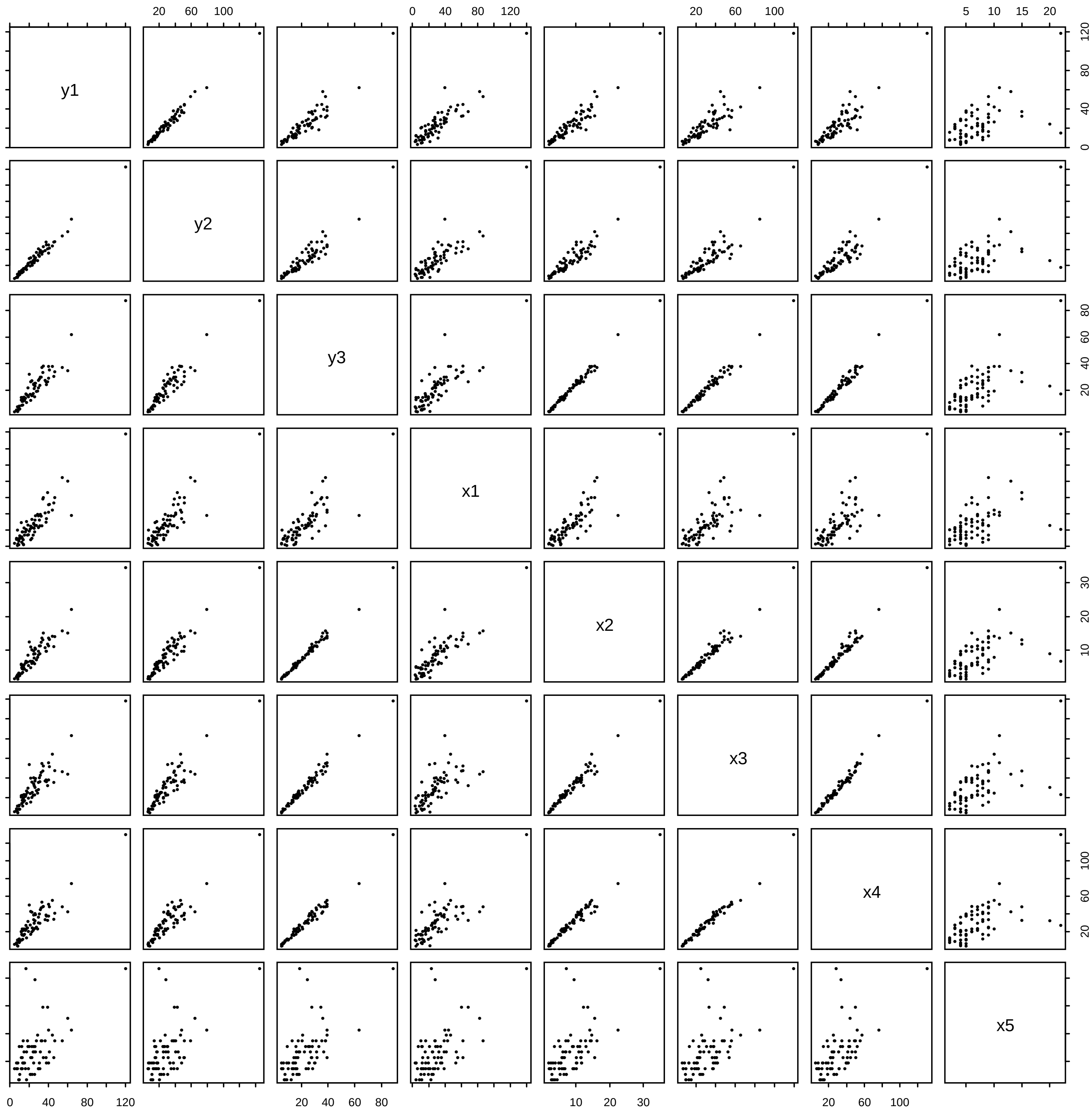

Figure 3: Pairwise Plots of the Charnes et al. (1981) Data