CHAPTER 4

EFFICIENCY AND DATA ENVELOPMENT ANALYSIS

This chapter has four objectives:

1. To justify – Section 1 – the use, in the analysis, of only one DEA model in its plainest and most simple version, when there are a very large number of different Data Envelopment Analysis models and model variants to choose from – refer to Appendix A,

2. To describe how efficiency is measured by the model chosen – the CCR model – and to describe the model itself – Sections 2, 3 and 4,

3. Through a series of simple examples to illustrate how structural properties in the data impact on the measurement of efficiency – Sections 5, 6 and 7, and

4. To identify how measurement of efficiency can be used to identify structural properties in the data such as correlation and dispersion – Section 8.

4.1 The Choice of Model

The CCR model was the first DEA model – Charnes et al. (1978). It was selected for use in this dissertation because the results from the model were not qualitatively distinct from the results obtained using three other DEA models. Price ratio constraints – see Section 3 of Appendix A – were tested and found to have no impact on the analysis. Models and model variants that allow for measurement of returns to scale were not used because the data – see Chapter 3 – are all scaled either to quantities per person or quantities per town.

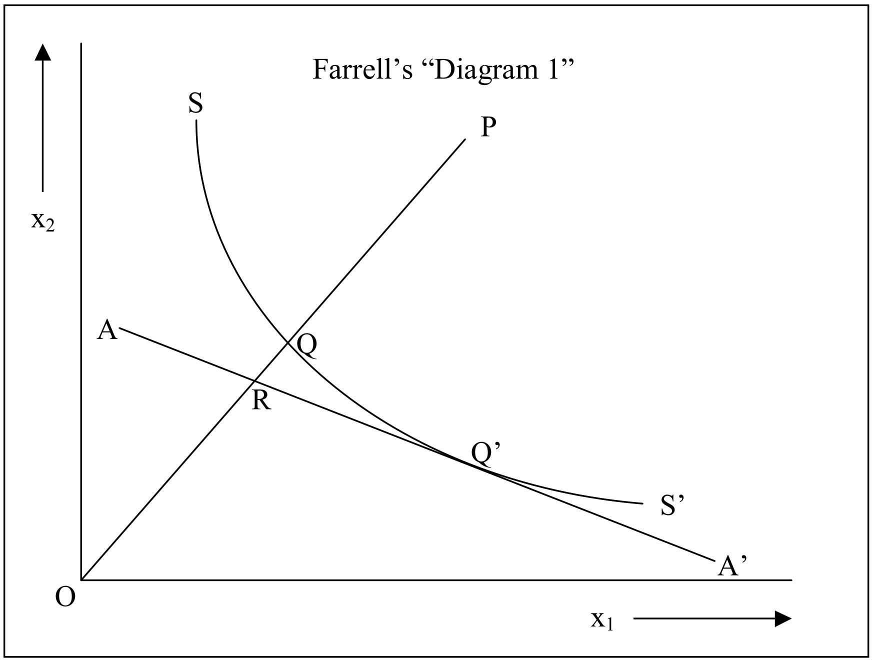
Since CCR model measurements of efficiency are the core of the analysis, an understanding of the nature of these measurements is key to an understanding of the analysis.

4.2 Farrell's Radial Measurement of Efficiency

Koopmans (1951) modeled an entire economy and provided a definition of a technically efficient production frontier. Technical efficiency is achieved when a producer is able to select the prices of inputs and outputs in such a way as to make a zero profit as all other producers make a zero profit or a loss. Taking the frontier and all points behind the frontier gives a production possibilities set – the union of the technically efficient DMUs and the technically inefficient DMUs.

Farrell (1957) was concerned with measurement of the difference between efficient and inefficient points. Farrell's "Diagram 1" is reproduced as Figure 4.01, with the modification that the axes representing the quantities of the primary factors have been labeled as χ_1 and χ_2 to conform with the notation used in this dissertation and set out in Section 3 of this Chapter. Primary factors are the raw materials: inputs to the production process.

Figure 4.01 – Farrell's "Diagram 1".



SS' is an isoquant representing combinations of the primary factors x_1 and x_2 required to produce a single output y_1 at a given level, q, with 100 percent technical efficiency.

AA' is a market price line whose slope is the ratio of the market prices of the primary factors. Production of quantity, q, of the output at Q' (where AA' is tangent to SS') is both technically efficient and price efficient. Q' represents the combination of the primary factors which results in the production of quantity, q, at the lowest cost.

A producer operating at P is both technically inefficient and price inefficient at the market prices implicit in AA'. P can improve its technical efficiency by reducing its usage of either or both of the primary factors. Assuming proportionate reductions in each primary factor, P would operate technically efficiently at Q. However to be as cost efficient as Q' at the market prices implied by AA' would require P to be operating at R if it were possible for P to further proportionately reduce its use of the primary factors.

Alternatively, P, operating at Q would be price efficient if AA' were tangent to SS' at Q. If a producer on SS' is allowed to select its own "market prices" such that AA' is tangent to SS' where the producer sits on SS' then the producer will be both 100 percent Technically and 100 percent Price Efficient. Put in another way, if a producer is allowed to choose its own market prices and it can achieve 100 percent Price Efficiency at those prices, then it is 100 percent Technically Efficient.

Farrell takes the ratio OR/OQ as the measure of Price Efficiency of P and OQ/OP as the measure of Technical Efficiency of P. The Overall Efficiency of P is the Price Efficiency * Technical Efficiency = OR/OQ * OQ/OP = OR/OP. Both measures are known as "radial" measures since they are derived from the ratio of the lengths of radii – lines from the geometric origin.

Reduction from P's usage to Q's usage involves reductions in x_2 and x_1 proportional to their actual usage, since PQO is a straight line. This quality is known as "equi-proportional reduction".

Actual examples of the production of output y_1 from primary factors x_1 and x_2 would be plotted as a series of points, rather than a continuous curve. Drawing a set of lines between a subset of the points, such that no point was closer to the origin than the lines, would give an estimation of SS'. Inefficient producers can then derive an estimated technical efficiency score by reference to the estimate of SS'.

Farrell addressed a single-output / dual-input case. As we shall see, the CCR model generalized this concept to multiple inputs and multiple outputs, but first the notation will be formalized in the context of Farrell and generalized.

4.3 Notation

A producer of outputs from primary factors will be known as a Decision Making Unit ("DMU").

The number of DMUs will be represented by n.

The subscript j will be used for DMUs numbered from 1 to n.

Output will be generalized and more than one output will be considered. The number of outputs will be represented by s.

The subscript r will be used for outputs numbered from 1 to s.

The quantities of each output used will be represented by the parameter y, so y_{rj} will represent the quantity of the rth output produced by the jth DMU.

Each DMU is allowed to choose the prices that result in its Price Efficiency being equal to its Technical Efficiency.

The output prices for a DMU are represented by u, which like the outputs themselves are indexed by r and by j, thus: μ_{rj}

The sum of the priced outputs will therefore be: $\sum_{r=1}^{s} \mu_{rj} y_{rj}$ for any given DMU, j. In other contexts this might be thought of as a weighted average of the outputs.

The number of inputs will be represented by m. The subscript i will be used for inputs numbered 1 to m. Inputs will be represented by the parameter x, so x_{ij} will represent the quantity of the ith input used by the jth DMU.

The input prices for a DMU are represented by v, which, like the inputs themselves, are indexed by i and by j, thus: v_{ij}

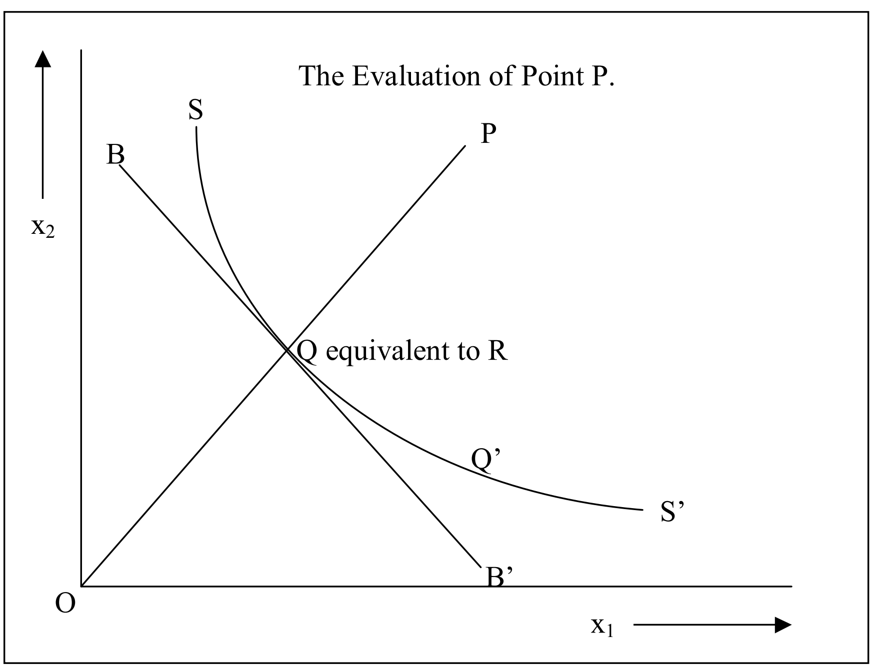
The sum of the priced inputs will therefore be: $\sum_{i=1}^{m} v_{ij} x_{ij}$ for any given DMU, j.

Each DMU is evaluated separately, so when the parameter or price is specifically the parameter or price for the DMU under consideration, the subscript "0" replaces "j".

4.4 The CCR Model

In Section 2, point P was evaluated at prices (AA') that allowed DMU Q' on SS' to be both Technically and Price Efficient. Prices that would allow P to operate as efficiently as Q – if P managed to reduce input usage of x_1 and x_2 proportionately to be operating at the same point as Q – are given by BB' the line tangent to SS' at Q – See Figure 4.02. Q, the Technically Efficient point is also R, from Figure 4.01, the Price Efficient point at which P would operate: if it could.

Figure 4.02 – The Evaluation of Point P.



Koopmans (1951) defined efficiency as arising when the sum of the Priced Outputs equals the sum of the Priced Inputs. The ratio is one for an efficient DMU. Using the notation described in Section 2 and substituting Q for the subscript j we can express the efficiency of Q, an efficient DMU, as: $h_Q = \frac{\sum_{i=1}^{m} v_{iQ} x_{iQ}}{\sum_{r=1}^{s} u_{rQ} y_{rQ}} = 1$. This expression of efficiency allows for up to s outputs: only one output is assumed in Farrell's model.

Evaluation of P at the market prices for Q gives P an efficiency[23] of:

$$h_P = \frac{\sum_{i=1}^{m} v_{iQ} x_{iP}}{\sum_{r=1}^{s} u_{rQ} y_{rP}} \geq 1$$

The Technical Efficiency and Price Efficiency of P, OP/OQ, is equivalent to h_P/h_Q. So, if both P and Q are evaluated at the "best" prices for them: P will be judged inefficient in relation to Q to the extent OP/OQ.

In order to solve for the efficiency of all of the DMUs requires a solution for *each* DMU of two problems:

---

[23] or we can take the reciprocal $\frac{1}{h_P} = \frac{\sum_{r=1}^{s} u_{rQ} y_{rP}}{\sum_{i=1}^{m} v_{iQ} x_{iP}} \leq 1$ which has the intuitively pleasing property of making 1 equivalent to full efficiency and numbers less than 1 and greater than zero represent less than full efficiency. The average of a set of reciprocals is equivalent to the reciprocal of the average of a set of values, so representing efficiency and average efficiencies from the CCR model on a scale from zero to one is simply a matter of calculating an average and taking the reciprocal.

1. What prices allow the DMU being evaluated to achieve the highest level of efficiency relative to all the other DMUs?

2. What is an Isoquant when there is more than one output?

Each DMU, in turn, is evaluated against its own Isoquant, which is to say at prices that allow the sum of the priced inputs to be equal to 1. $\sum_{r=1}^{s} u_{r0} y_{r0} = 1$. The DMU being evaluated then compares itself to the other DMUs at this Isoquant.

If we let the sum of the Priced Outputs for any given DMU, 0, be equal to 1:

$\sum_{r=1}^{s} u_{r0} y_{r0} = 1$, then DMU 0 is efficient if: $h_0 = \dfrac{\sum_{i=1}^{m} v_{i0} x_{i0}}{\sum_{r=1}^{s} u_{r0} y_{r0}} = \dfrac{\sum_{i=1}^{m} v_{i0} x_{i0}}{1} = \sum_{i=1}^{m} v_{i0} x_{i0} = 1$

but DMU, 0, is inefficient if : $\sum_{i=1}^{m} v_{i0} x_{i0} = h_0 > 1$.

For any given DMU, 0, minimizing $\sum_{i=1}^{m} v_{i0} x_{i0} = h_0$, while at the same time ensuring that:

1. no other DMU can do better than DMU, 0, at the prices for DMU, 0, and

2. $\sum_{r=1}^{s} u_{r0} y_{r0} = 1$

will yield the prices for DMU, 0, that allow it to achieve its highest level of efficiency relative to the other DMUS and will allow calculation of an efficiency score. This gives an algorithm that can be applied to the n DMUs as follows:

For 0 from 1 to n, solve:

$$\min h_0 = \sum_{i=1}^{m} v_{i0} x_{i0}$$

$s.t.$

(1) $-\sum_{r=1}^{s} \mu_{r0} y_{rj} + \sum_{i=1}^{m} v_{i0} x_{ij} \geq 0 \quad \forall j$

(2) $\sum_{r=1}^{s} \mu_{r0} y_{r0} = 1$

(3) $\mu_{r0}, v_{i0} \geq 0$

(4) $y_{rj}, x_{ij} \geq 0 \quad \forall i, r$ and j

The result is an efficiency score h_0 for each DMU and a set of input prices and a set of output prices for each DMU.

Consider P and Q in this framework. There are two DMUs: so n = 2. For Q one first evaluates Q against P at Q's prices:

$$\min h_Q = \sum_{i=1}^{m} v_{iQ} x_{iQ}$$

$s.t.$

(1Q) $-\sum_{r=1}^{s} \mu_{rQ} y_{rQ} + \sum_{i=1}^{m} v_{iQ} x_{iQ} \geq 0$

(1P) $-\sum_{r=1}^{s} \mu_{rQ} y_{rP} + \sum_{i=1}^{m} v_{iQ} x_{iP} \geq 0$

(2) $\sum_{r=1}^{s} \mu_{rQ} y_{rQ} = 1$

(3) $\mu_{rQ}, v_{iQ} \geq 0$

Expanding the constraints as (1Q) and (1P) make it explicit that Q is being evaluated against P at Q's prices.

Then one evaluates P against Q at P's prices:

$$\min h_P = \sum_{i=1}^{m} v_{iP} x_{iP}$$

$s.t.$

$$(1Q) \qquad -\sum_{r=1}^{s} \mu_{rP} y_{rQ} + \sum_{i=1}^{m} v_{iP} x_{iQ} \geq 0$$

$$(1P) \qquad -\sum_{r=1}^{s} \mu_{rP} y_{rP} + \sum_{i=1}^{m} v_{iP} x_{iP} \geq 0$$

$$(2) \qquad \sum_{r=1}^{s} \mu_{rP} y_{rP} = 1$$

$$(3) \qquad \mu_{rP}, v_{iP} \geq 0$$

The model, in effect, chooses the prices at which a DMU does the best it can: relative to the other DMUs at those prices.

4.5 The Radial Measurement of Efficiency

Example One will show how the CCR model adjusts for pure scale differences in the inputs and outputs of two DMUs whose inputs and outputs are positively correlated. Example Two will show how the CCR model adjusts for pure scale differences where inputs and outputs are negatively correlated. Average efficiency in Example One is high and average efficiency in Example Two is low. In other words high efficiency can be a measure of positive correlation and low efficiency can be a measure of negative correlation between sets of parameters: inputs and outputs.

These two examples also demonstrate the impact of a total lack of statistical independence between the values of the inputs for any given DMU and also between the values of the outputs for any given DMU.

4.5.1. Example One

Take two DMUs, A and B, with inputs highly positively correlated to outputs – Example One:

- Assume A has inputs of 1 and 1 and outputs of 1 and 1.

  $\left(x_{1A}=1,\ x_{2A}=1,\ y_{1A}=1,\ y_{2A}=1\right)$

- Assume B has inputs of 10 and 10 and outputs of 10 and 10.

  $\left(x_{1B}=10, x_{2B}=10, y_{1B}=10, y_{2B}=10\right)$

If A adopts prices of ½, $\left(u_{1A}=\frac{1}{2}, u_{2A}=\frac{1}{2}, v_{1A}=\frac{1}{2}, v_{2A}=\frac{1}{2}\right)$ then the sum of the priced inputs is $\sum_{i=1}^{m} v_{iA}x_{iA}=\frac{1}{2}1+\frac{1}{2}1=1$, and the sum of the priced outputs is $\sum_{r=1}^{s} u_{rA}y_{rA}=\frac{1}{2}1+\frac{1}{2}1=1$.

If B uses the same prices, then the sum of the priced inputs is $\sum_{i=1}^{m} v_{iA}x_{iB}=\frac{1}{2}10+\frac{1}{2}10=10$, and the sum of the priced outputs is $\sum_{r=1}^{s} u_{rA}y_{rB}=\frac{1}{2}10+\frac{1}{2}10=10$.

In each case, A and B, the ratio of priced inputs to priced outputs will be 1 – the two DMUs are equally efficient. When evaluating DMU A, constraint (2) will apply to DMU A, which could choose prices of ½. When evaluating DMU B, constraint (2) will now apply to DMU B, which will choose prices of 1/20. Note that the ratio of the prices is the same in either case: both A and B are evaluated against a price line with a slope of minus one.

B's greater scale has been incorporated into the prices it selects for itself in the CCR model. High positive correlation is evidenced in the high efficiency scores because the CCR model adjusts for pure scale differences between the DMUs.

Example Two illustrates the way in which the CCR model adjusts when the scale differences imply inefficiency or negative correlation between the inputs and the outputs.

4.5.2 Example Two

Take two DMUs, C and D, with inputs highly negatively correlated to outputs:

- Assume C has inputs of 1 and 1 and outputs of 10 and 10.
- Assume D has inputs of 10 and 10 and outputs of 1 and 1.

In this case the prices make no difference to the outcome. If C adopts prices of ½ for inputs and 1/20 for outputs, then for C, the efficiency score is given by:

$$h_C = \frac{\sum_{i=1}^{m} v_{iC} x_{iC}}{\sum_{r=1}^{s} u_{rC} y_{rC}} = \frac{\frac{1}{2}*1 + \frac{1}{2}*1}{\frac{1}{20}*10 + \frac{1}{20}*10} = 1$$

Scaling D to C's level of output assumes multiplication throughout by 10, so, for D the efficiency score, at C's prices is given by:

$$h_D = \frac{\sum_{i=1}^{m} v_{iD} 10 x_{iD}}{\sum_{r=1}^{s} u_{rD} 10 y_{rD}} = \frac{\frac{1}{2}*100 + \frac{1}{2}*100}{\frac{1}{20}*10 + \frac{1}{20}*10} = \frac{100}{1} = 100$$

The effect of this scaling difference is reflected in the efficiency scores.

4.5.3 Example Three

For Example Three – consider two DMUs E and F with inputs highly negatively correlated to outputs but more variation in the quantities. The input parameters and the output parameters for each DMU continue to be statistically dependent on each other, but not as absolutely as in Examples One and Two. This allows the two DMUs some freedom to choose two distinct sets of prices and reduces the measured inefficiency of the second DMU.

- Assume E has inputs of 2 and 1 and outputs of 10 and 8.
- Assume F has inputs of 8 and 10 and outputs of 1 and 2.
- Assume E and F are the only DMUs

If both E and F adopt prices of 0, ½, 0.1 and 0 respectively then for E and F the efficiency scores are given by:

$$h_E = \frac{\sum_{i=1}^{m} v_{iE} x_{iE}}{\sum_{r=1}^{s} u_{rE} y_{rE}} = \frac{0*2 + \frac{1}{2}*1}{\frac{1}{10}*10 + 0*8} = 1$$, and

$$h_F = \frac{\sum_{i=1}^{m} v_{iE} x_{iF}}{\sum_{r=1}^{s} u_{rE} y_{rF}} = \frac{0*8 + \frac{1}{2}*10}{\frac{1}{10}*1 + 0*2} = \frac{5}{0.1} = 50$$

Alternatively, if both E and F adopt prices of 2, 0, 0, and 1/2 respectively then for F and E the efficiency scores are given by:

$$h_F = \frac{\sum_{i=1}^{m} v_{iF} x_{iF}}{\sum_{r=1}^{s} u_{rF} y_{rF}} = \frac{2*8 + 0*10}{0*1 + \frac{1}{2}*2} = \frac{16}{1} = 16$$, and

$$h_E = \frac{\sum_{i=1}^{m} v_{iF} x_{iE}}{\sum_{r=1}^{s} u_{rF} y_{rE}} = \frac{2*2 + 0*1}{0*10 + \frac{1}{2}*8} = \frac{4}{4} = 1$$

16 is closer to 1 than is 50: so F will "choose" evaluation at prices 2, 0, 0, and 1/2 respectively. E is indifferent between the two sets of prices.

Comparison of the parameter values in Example Three with those in Example Two and then of the two sets of efficiency scores shows the extent to which the small variation between the parameters for each DMU impacts on efficiency scores.

C had inputs of 1 and 1 and outputs of 10 and 10 and had efficiency of 1. D had inputs of 10 and 10 and outputs of 1 and 1 and had efficiency of 100.

E had inputs of 2 and 1 and outputs of 10 and 8 and efficiency of 1. F had inputs of 8 and 10 and outputs of 1 and 2 and had efficiency of 16.

4.5.4 Example Four

When the standard deviations for each input parameter are relatively lower than in the previous examples, the DMUs may be able to choose three sets of prices at which they are both efficient. The more sets of prices they can choose, the more unlikely it is that the addition of a third DMU, of similar scale, will cause either one of them to be deemed inefficient. This makes it desirable to have a large number of DMUs in a model.

Example Four considers two DMUs G and H. Both are evaluated as 100 percent efficient at three different sets of prices. Applying these prices in turn allows G to be considered first to be twice the size of H; then the same size as H, and finally, half the size of H.

- Assume G has inputs of 2 and 1 and outputs of 1 and 2.
- Assume H has inputs of 1 and 2 and outputs of 2 and 1.
- Assume G and H are the only DMUs

If both G and H adopt prices of 1/3, 1/3, 1/3 and 1/3 respectively then G and H have efficiency scores given by:

$$h_G = \frac{\sum_{i=1}^{m} v_i x_{iG}}{\sum_{r=1}^{s} u_r y_{rG}} = \frac{\frac{1}{3}*2+\frac{1}{3}*1}{\frac{1}{3}*1+\frac{1}{3}*2} = 1 \text{ , and } h_H = \frac{\sum_{i=1}^{m} v_i x_{iH}}{\sum_{r=1}^{s} u_r y_{rH}} = \frac{\frac{1}{3}*1+\frac{1}{3}*2}{\frac{1}{3}*2+\frac{1}{3}*1} = 1$$

If both G and H adopt prices of 0, 1, 1 and 0 respectively then G and H have efficiency scores given by:

$$h_G = \frac{\sum_{i=1}^{m} v_i x_{iG}}{\sum_{r=1}^{s} u_r y_{rG}} = \frac{0*2+1*1}{1*1+0*2} = 1 \text{, and } h_H = \frac{\sum_{i=1}^{m} v_i x_{iH}}{\sum_{r=1}^{s} u_r y_{rH}} = \frac{0*1+1*2}{1*2+0*1} = 1$$

If both G and H adopt prices of 1, 0, 0 and 1 respectively then G and H have efficiency scores given by:

$$h_G = \frac{\sum_{i=1}^{m} v_i x_{iG}}{\sum_{r=1}^{s} u_r y_{rG}} = \frac{1*2+0*1}{0*1+1*2} = 1 \text{ , and } h_H = \frac{\sum_{i=1}^{m} v_i x_{iH}}{\sum_{r=1}^{s} u_r y_{rH}} = \frac{1*1+0*2}{0*2+1*1} = 1$$

G and H can both be efficient at three different sets of prices. With the first set of prices, they are both assumed to be operating at the same scale. At the second set of prices G produces one unit of output from one unit of input and H produces two units of output from one unit of input. At the third set of prices the G produces twice as much as H.

If G and H were being evaluated on three inputs and three outputs the number of sets of prices at which they both would be evaluated as efficient would increase. Consider H with inputs (1, 2, and 1) and outputs of (2, 1 and 2), the inputs pricing choices are: (1, 0, 0) or (0, ½, 0) or (0, 0, 1) or (¼,¼,¼) or (½, 0, ½) or (0, ⅓, ⅓) or… with a similar set of pricing choices for the outputs.

The greater the number of inputs parameters and the greater the number of outputs parameters the greater the probability that a DMU will be evaluated as fully efficient. The addition of more DMUs may limit the individual DMU's ability to use combinations of prices to "achieve" full efficiency.

So, the average efficiency scores depend on the distributions of the parameters, the number of the parameters and the number of DMUs evaluated.

4.6 Units of Measurement and Translation of Parameters

The CCR model is *not* invariant to a translation of the values of the parameters either in terms of the scale of the efficiency scores or in the rank order of efficiency of the DMUs. The CCR model is not "Translation Invariant". In other words, the distribution of the parameters, their means and standard deviations, impacts on the efficiency scores. Translation of the data changes the means, without changing the standard deviations.

In order to assess the impact on efficiency scores of differential scaling of means and standard deviations, the parameters are split into their means and standard deviations in the following analysis.

Recall that the CCR efficiency score h_o for any given DMU, 0, is given by the ratio of priced inputs to priced outputs:

$$h_o = \frac{\sum_{i=1}^{m} v_{io} x_{io}}{\sum_{r=1}^{s} u_{ro} y_{ro}} .$$

Multiplying through by the denominator on the right hand side gives:

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} v_{io} x_{io}$$

The values of the inputs and outputs can be decomposed into their means, standard deviations and z-scores, for the values from all of the DMUs from 1 to n, such that for the “0”[th] DMU: $x_{io} = \mu_i + (z_{io}\sigma_i)$ and $y_{ro} = \mu_r + (z_{ro}\sigma_r)$

If all the inputs are multiplied by n, this is the same as multiplying the means and the standard deviations of the inputs by n, the prices of the inputs adjust by dividing by 1/n and the efficiency score is unaffected, thus:

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n} n x_{io} = \frac{1}{n} \sum_{i=1}^{m} v_{io} n x_{io} = \frac{n}{n} \sum_{i=1}^{m} v_{io} x_{io} = \sum_{i=1}^{m} v_{io} x_{io}$$

If the standard deviations of the inputs are multiplied by 13.33 and the means of the inputs are multiplied by 5.65, then the prices adjust, but they do so by finding new relationships between the prices, rather than adopting a new scale, thus:

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} \left(5.65\mu_i + 13.33 * z_{io}\sigma_i\right) \neq \sum_{i=1}^{m} v_{io} x_{io}$$, where n_{io}

is the adjustment to the price for input i for unit o.

Even if one knew the distribution of the z-scores it would still be difficult to know what the impact on the prices and the efficiency scores of all the DMUs would be.

The “units of measurement effect” can be separated from the “translation effect”. If the multiple of the standard deviation is split into two components, 5.65 and 7.68: the following results:

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} \left(5.65 * \mu_i + (5.65 + 7.68) * (z_{i0} * \sigma_i)\right)$$

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} \left(5.65 * \mu_i + 5.65 * z_{i0} * \sigma_i + 7.68 * z_{i0} * \sigma_i\right)$$

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} \left(5.65 * (\mu_i + z_{i0} * \sigma_i) + 7.68 * z_{i0} * \sigma_i\right)$$

From this it is clear that if the σ_i are small relative to the μ_i then the impact of the translation component of the scaling $(7.68 * z_{i0} * \sigma_i)$ should be small relative to the units component of the scaling $\left(5.65 * \left(\mu_i + z_{i0} * \sigma_i\right)\right) = \left(5.65 * x_{i0}\right)$.

The key components are the z-scores for each DMU. The relationships between the inputs z-scores and between the outputs z-scores can have a large impact. Consider DMU 0, with 4 inputs with z-scores of –2, -1, 1, and 2 and plug these values into the calculation of efficiency. Assume that the standard deviations are the same for each input parameter.

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} \left(5.65 * \left(\mu_i + z_{i0} * \sigma_i\right) + 7.68 * z_{i0} * \sigma_i\right)$$

and since $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$ is assumed σ replaces σ_i

and $\left(\mu_i + z_{i0} * \sigma_i\right) = x_{i0}$

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = 5.65 \sum_{i=1}^{m} \frac{v_{io} x_{i0}}{n_{io}} + 7.68 * \sigma * \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} (* z_{i0})$$

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = 5.65 \sum_{i=1}^{m} \frac{v_{io} x_{i0}}{n_{io}} + 7.68 * \sigma * \left(\frac{v_{1o}}{n_{1o}} * (-2) + \frac{v_{2o}}{n_{2o}} * (-1) + \frac{v_{3o}}{n_{3o}} * 1 + \frac{v_{4o}}{n_{4o}} * 2 \right)$$

If the prices $\frac{v_{1o}}{n_{1o}} = \frac{v_{2o}}{n_{2o}} = \frac{v_{3o}}{n_{3o}} = \frac{v_{4o}}{n_{4o}}$ then the impact of the distribution of the z-scores will be to nullify the impact of the translation component of scaling

since: $7.68 * \sigma_i * \left(\frac{v_{1o}}{n_{1o}} * (-2) + \frac{v_{2o}}{n_{2o}} * (-1) + \frac{v_{3o}}{n_{3o}} * 1 + \frac{v_{4o}}{n_{4o}} * 2 \right) = 0$

But, the prices are unlikely to equal each other. With respect to test scores a better assumption might be that they are not statistically independent and therefore; the test score z-scores for a DMU are likely to be of a similar scale and so the translation component is likely to have some effect.

This analysis decomposed the inputs side of the equation $h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} v_{io} x_{io}$.

The same analysis can be applied to the output side of the equation.

4.7 Panel Data

This dissertation looks at trends in education over time. In other words it deals with "Panel Data". Assuming that, over time, there are no scale changes in the data being analyzed; then there are two main approaches to analysis of panel data with the CCR model.

In the first approach, each period is analyzed in a separate model and the results are compared across time. In this approach each DMU is compared with every other DMU from its own time period.

In the second approach, all of the periods are combined in a single model. Every DMU in every time period is compared with every DMU in all time periods.

A simple three-year and two-DMU example will illustrate the issues inherent in the two approaches.

Taking the first approach, first, and looking at the efficiency scores from three individual analyses, one for each year for two DMUs: A and B. Assume that: in year 1, A produced 4 from 3, and B produced 4 from 4. (What the input is and what the output is: is not important). A did better than B since 3/4 < 4/4. The model prices A such that A's priced ratio = 1 to conform to Koopmans definition of efficiency. So A's efficiency score is 1 and B's efficiency score is 1.333. Average efficiency is 1.166.

Assume B improves in year 2, such that it also produces 4 from 3: both A and B have efficiency scores of 1. Average efficiency improved (8 comes from 6) to 1.

Assume, that in year 3, A had become less efficient producing 3 from 3 but B reverts to producing 3 from 4. A is now as inefficient as B, but the model shows them to have efficiency scores of 1. Average efficiency in the third year is 1, an apparent improvement over the first year, although A and B are now producing 7 from 7 as opposed to 8 from 6 in the second year or 8 from 7 in the first year.

The average efficiency score is revealing something about the change in relative efficiency between DMU A and DMU B. A is more efficient than B in year 1, so the average efficiency score is higher (reflecting a lower level of efficiency) than it is for the two years in which A and B are equally efficient.

Now, taking the second approach and looking at the efficiency scores from a single analysis of both DMUs and three years. There are six "DMUs". A1, B1, A2, B2, A3 and B3. A1, A2, B2 now define full efficiency: each getting 4 from 3. B1 scores 1.333 as before. A3 and B3 now score 1.333 (relative to A1, A2 and B2). Average efficiency in year one is 1.1666. Average efficiency in year two is 1. Average efficiency in year three is 1.333.

From this simple example one can conclude that if the issue at stake is absolute change in efficiency over time, then the evaluation against the six "DMUs" provides good measurements. If, on the other hand, the issue is relative efficiency in each year and how this changes, then evaluating each year separately provides good measurements.

It is also instructive to note that average efficiency can increase in the case where some efficient DMUs loose efficiency or in the case where some inefficient DMUs become more efficient.

4.8 Conclusions

The eight properties or characteristics in the data that impact on the measurement of efficiency using the CCR model are:

1. The degree and direction of the correlation between the inputs (as priced) and the outputs (as priced),
2. The distributions of the data – the z-scores or the degree of statistical independence between the inputs parameters, and the statistical independence between the outputs parameters,
3. The number of DMUs evaluated,
4. The number of input parameters and the number of output parameters,
5. The freedom that some DMUs have to choose prices that are zero for some of the inputs and some of the outputs,
6. The units or absolute scale of the data – the means,
7. The standard deviations of the data, and
8. The treatment of Panel Data.

These factors can be used good effect by controlling some and thus measuring others.