CHAPTER 4

THE DEVELOPMENT OF THE DSDF APPROACH AND DATA COLLECTION

4.1 Introduction

As has been discussed in the previous chapter, Directional Distance Function (DDF) is a recognized technique for measuring efficiency while incorporating undesirable output. This approach allows for desirable output to be expanded while undesirable output is contracted simultaneously. Despite gaining popularity because of the incorporation of undesirable output, this approach also has drawbacks. The drawbacks of the DDF approach are that the direction vector to the production boundary is fixed arbitrarily and this model does not take into account non-zero slacks in the efficiency measurement. Therefore, the major section in this chapter is about the extension of the previous framework of the DDF technique to introduce a new slacks-based measure of efficiency called the Directional Slack-based Distance Function (DSDF) model. This new approach may determine the optimal direction to the frontier for each unit of analysis and provides dissimilar expansion and contraction factors to achieve a more reasonable efficiency score. In addition, the use of the new approach may also establish target values for the reduction/expansion of output in order for the inefficient DMUs to achieve full eco-efficiency.

In efficiency measurement, the ability to distinguish the top performance is important in order to understand the quality of their performance. The useful application of a super-efficiency model was implemented due to the failure of standard DSDF model to rank the efficient set of the DMUs attaining an efficiency score of unity in this study. This super-efficiency score can distinguish between efficient observations.

To complete the analysis, it would be an advantage to extend the understanding of the productivity change over the years through the Malmquist Luenberger Productivity Index (MLPI). The computed index, which quantifies the productivity change can be decomposed into the measurement of eco-efficiency change and technological change between a fixed based year (*t*) and a target year (*t*+1). Efficiency and productivity measurement are widely used and can be put to work together, as to complement each other. In this study, the DEA, DDF and DSDF models may present the results of the efficiency measurement in a particular year, in other words, static performance while the Malmquist Lunberger index measures the performance over time.

The remainder of this chapter is organized in the following manner. This chapter will start with the extension from the previous framework of efficiency analysis to introduce a new slacks-based measure of efficiency called the Directional Slack-based Distance Function (DSDF) approach in Section 4.2. To overcome the problem with fully efficient using the DSDF approach, the super-efficiency model is suggested and investigated in Section 4.3. Further, this chapter also discusses the Malmquist Luenberger productivity index (MLPI) in Section 4.4 in order to study the productivity change for the study period of 2001 to 2010. Section 4.5 verifies the variable selections followed by data source. Section 4.6 summarizes the chapter.

4.2 Directional Slack-based Distance Function (DSDF)

Based on the original Slack-Based Measure (SBM) model proposed by Tone (2001), Färe and Grosskopf (2010a; 2010b) develop the efficiency measurement with an additive structure of input and output slacks through addition and subtraction from their respective inequality as follows:

$$
\begin{aligned}
&\text{Max } \alpha_m = \gamma_{x1} + \ldots + \gamma_{xI} + \cdots + \gamma_{y1} + \cdots + \gamma_{yJ} \\
&\text{Subject to} \\
&\qquad \sum_{n=1}^{N} z_n x_{in} \leq x_{im} - \gamma_{xi} \,.\, 1; \;\; i = 1,2,\ldots,I \\
&\qquad \sum_{n=1}^{N} z_n y_{jn} \geq y_{jm} + \gamma_{yj} \,.\, 1 \;; \;\; j = 1,2,\ldots,J \\
&\qquad z_n, \gamma_{xi}, \gamma_{yj} \;\geq 0 \;; \;\; n = 1,2,\ldots,N
\end{aligned} \tag{4.1}
$$

Where

z_n = intensity variables

x_{in} = i^{th} input of the n^{th} DMU

x_{im} = i^{th} input of the m^{th} DMU

y_{jn} = j^{th} desirable output of the n^{th} DMU

y_{jm} = j^{th} desirable output of the m^{th} DMU

$\gamma_{xi}\,.\,1$ = number of units of each type of input that can be decreased for m^{th} DMU

$\gamma_{yj}\,.\,1$ = number of units of each type of output that can be increased for m^{th} DMU

The vectors γ_{xi} and γ_{yj} indicate that the input and output can be decreased and increased, respectively, and are called slacks. The results of γ_{xi} and γ_{yj} are independent of the unit of measurement, and therefore, they may be summed in objective function. In this development, Färe and Grosskopf demonstrate a Slack Based Measure (SBM) of efficiency based on the Directional Distance Function (DDF) model incorporating input and desirable output variables.

Based on the works of Färe and Grosskopf (2010a; 2010b), a new slack-based measure of efficiency called the Directional Slack-based Distance Function (DSDF) approach is developed. This new development incorporates the undesirable outputs in order to measure an appropriate direction for each inefficient DMU to attain full efficiency.

There are three aspects that are worth emphasizing in this new development. First, the original DDF model (formulation (3.20) in previous chapter) is modified so that each output bundle can have a different scale direction to the production boundary. The objective function of the DDF model (3.20), which is the single contraction/expansion factor, β_m has been replaced with the summation of γ_{yj}, the slack for desirable output, i.e. the expansion factor for desirable output, and γ_{uk}, the slack for undesirable output, i.e. the contraction factor for undesirable output in the DSDF approach, in formulation (4.2) below. This linear program is based on the slacks-based measure of efficiency. DMU *m* is efficient if and only if the optimal objective for model (4.2) is zero. Note that model (4.2) is unit invariant, which means that its optimal value does not depend on the units of measurement in desirable and undesirable output variables. Model (4.2) computes the efficiency score based on the desirable and undesirable output slacks. With these slack results, directions for improvement are easily obtained for each desirable and undesirable outputs measure. The DSDF model formulation for DMU *m*, which has been adopted from Färe and Grosskopf (2010a; 2010b) is as follows:

$$
\begin{aligned}
&\text{Max } \sigma_m = \sum_{j=1}^{J} \gamma_{yj} + \sum_{k=1}^{K} \gamma_{uk} \\
&\text{Subject to} \\
&\qquad \sum_{n=1}^{N} z_n x_{in} \leq x_{im}\,; \;\; i = 1,2,\dots,I \\
&\qquad \sum_{n=1}^{N} z_n y_{jn} \geq y_{jm} + \gamma_{yj}\,.\,1\,; \;\; j = 1,2,\dots,J \\
&\qquad \sum_{n=1}^{N} z_n u_{kn} = u_{km} - \gamma_{uk}\,.\,1\,; \;\; k = 1,2,\dots,K \\
&\qquad z_n, \gamma_{yj}, \gamma_{uk} \geq 0\,; \;\; n = 1,2,\dots,N
\end{aligned}
\tag{4.2}
$$

Note that in equation (4.1), the slack variables only involve input and desirable output. As for equation (4.2), the undesirable output is incorporated. For this equation, the slack for input is removed implying that input slack has not been computed. The slacks are only computed for desirable and undesirable outputs where the slack for the desirable output is allowed to be expanded and the undesirable output contracted.

It can be verified that $0 < \sigma_m \leq 1$ and also satisfies the properties of units invariance and monotone as has been validated by Tone (2001) in the original slack based measure model for efficiency measurement. Färe and Grosskopf (2010a) also confirmed these two properties in their model on slack based measure with directional distance function approach. The two properties are as follows:

(P1) Units invariant: the measure should be invariant with respect to the units of data.

(P2) Monotone: the measure should be monotone decreasing in each slack in desirable and undesirable outputs.

The slack variables γ_{yj} and γ_{uk} are used to identify and estimate the causes of inefficiency. Since σ_m is the inefficiency score, to obtain the eco-efficiency score from model (4.2), it can be calculated as follows:

$$\varphi_m = 1 - \sigma_m \tag{4.3}$$

Note that σ_m is between 0 and 1, thus, the eco-efficiency score with DSDF approach (φ_m) will also fall into the 0 and 1 closed interval.

In order to make the resulting model unit-invariant, a possible alternative that has been used in this study is normalizing the data (See for example Xu et al., 2012 who used the data normalization method to get a unit invariant result in their study). The difficulty of very large scale variables can occur for all mathematical models, in which many studies leverage different variables to the same level, or normalize them. In this study, the data set has been normalized by dividing them by the maximum value of each data set. This normalization procedure is applied so that a meaningful efficiency or inefficiency measure could be constructed.

Second, the optimal solution to equation (4.2) is used to derive the direction vector to the production boundary. When $\sum_{j=1}^{J} \gamma_{yj} + \sum_{k=1}^{K} \gamma_{uk} > 0$, the scale direction for the desirable output *j* and undesirable output *k* for the DMU assessed can be obtained by the following equation:

$$SD_j = \frac{\gamma_{yj}^*}{\sum_{j=1}^{J} \gamma_{yj} + \sum_{k=1}^{K} \gamma_{uk}}$$

and

$$SD_k = \frac{\gamma_{uk}^*}{\sum_{j=1}^{J} \gamma_{yj} + \sum_{k=1}^{K} \gamma_{uk}} \tag{4.4}$$

Where SD_j = scale direction for desirable output *j* and SD_k = scale direction for undesirable output *k*.

If γ_{yj}^* and γ_{uk}^* are equal to 0, it denotes that the particular DMU is located on the efficient frontier, then the direction vectors SD_j and SD_k can be chosen arbitrarily.

Equation (4.4) explains that each slack of the desirable and undesirable outputs from model (4.2) is divided by the additive structure of the desirable and undesirable output slacks on the denominator which can provide dissimilar direction for each desirable and undesirable output. This dissimilar direction may overcome the drawback of DDF model where the direction is fixed arbitrarily. The total of scale direction SD_j and SD_k must also be equal to 1 ($SD_j + SD_k = 1$) which ensures compactness so that an appropriate scale direction for each desirable and undesirable output variables can be obtained. The minimum and maximum direction for SD_j and SD_k is between 0 and 1 ($0 \leq SD_j$, $SD_k \leq 1$).

Third, from the scale directions obtained, the target value for each DMU can be measured. The DSDF approach can also be utilized for target setting to determine the target value for inefficient DMUs in order to obtain full eco-efficiency. The target value is measured by the summation of multiplication of the intensity variable (z_n) from formulation (4.2) with the actual value of desirable (y_{jn}) and undesirable (u_{kn}) outputs for each DMU, as below:

$$\sum_{n=1}^{N} z_n y_{jn}$$

and

$$\sum_{n=1}^{N} z_n u_{kn} \tag{4.5}$$

The target value will be similar to the actual value if the DMU m obtains a zero value for the objective function in model (4.2). In other words, the DMU m is 100 percent fully efficient.

To demonstrate the DSDF model, a numerical example has been used by using single desirable and undesirable outputs while consuming the same set of inputs. Table 4.1 presents the numerical example for five DMUs with single desirable (y) and undesirable (u) output. For this example, the VRS model is used as the convexity condition under VRS model may illustrate a clearer picture of DSDF model. Under the

VRS model, it is clear that DMUs A, B and C are efficient while the other two DMUs (D and E) are clearly inefficient (see Figure 4.1).

Employing the DDF model (equation (3.20)) with the direction vector (g_y, $-g_u$) = (y,$-u$), DMU D is projected onto the efficient frontier at D' = (2, 1) and DMU E is projected onto the efficient frontier at E' = (3.3, 3.1). For DMUs D and E, the efficiency score associated with the direction vector of (y,$-u$) are 0.67 percent and 0.64 percent, respectively.

Employing the DSDF model (equation (4.2)), DMU D and E are projected onto D" = (2.5, 1.5) and E" = (3, 2), respectively. An appropriate scale direction for DMU D and E computed from equation (4.4) is (1, 0) and (0.19, -0.81), respectively. The efficiency scores with the respective direction vector for D and E are 0.63 percent and 0.36 percent, respectively. Table 4.1 presents information on the numerical example using DDF and DSDF approaches for single desirable and undesirable outputs.

Table 4.1: Numerical example of DDF and DSDF

DMU	y	u	DDF			DSDF		
			$1-\beta_m$	y	u	$1-\sigma_m$	y	u
A	2	1	1	-	-	1	-	-
B	3	2	1	-	-	1	-	-
C	4	5	1	-	-	1	-	-
D	1	1.5	0.67	2	1	0.63	2.5	1.5
E	2.5	4.6	0.64	3.3	3.1	0.36	3	2

* Note that (y,u) for DDF column is projection onto D' and E' while DSDF column is projection onto D" and E".

Figure 4.1 demonstrates how the DSDF approach measures the direction for inefficient DMUs to achieve the efficiency frontier. The DSDF model expands and contracts the desirable and undesirable outputs by a different proportion and this model also determines the optimal direction to the frontier for each of the inefficient DMUs. The direction in this approach is determined by the additive slack of the desirable and undesirable output.

It can be seen that by using the DSDF model, the scale direction for DMU D is (1,0) implying that DMU D needs to increase the desirable (*y*) output but not to decrease the undesirable (*u*) output to achieve a frontier at D”. While the scale direction for DMU E is (0.19, -0.81) implying that DMU E needs to increase desirable (*y*) output by the scale of 0.19 and decrease the undesirable (*u*) output by the scale of 0.81 simultaneously in order to achieve a frontier at E”. These scales are given by the assumption from equation (4.4). Next, equation (4.5) is used to get the projection of DMU D” at (2.5, 1.5) and E” at (3, 2) in order to obtain full eco-efficiency of these two DMUs. For instance, DMU D needs to increase desirable (*y*) output from 1 to 2.5 while undesirable (*u*) output remains the same at 1.5. On the other hand, DMU E needs to increase desirable (*y*) output from 2.5 to 3 and decrease undesirable (*u*) output from 4.6 to 2.

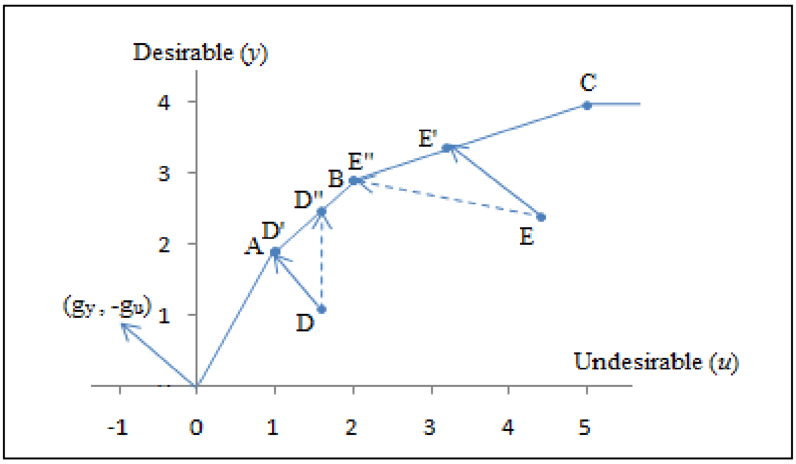


Figure 4.1: DDF and DSDF direction vector

To compare between the DDF and DSDF approaches, some methodological reasons can be taken into consideration for the differences between these two approaches. The original concept of the efficiency score in the DDF approach is determined by the method of ratio. The ratio of EF/EG can be found in Figure 3.6. Giving the expansion of desirable output and reduction of undesirable output simultaneously with an arbitrary direction (g = (*y*, -*u*)) may provide an inappropriate direction for each output variable. This is the drawback of using the DDF approach as there are no standard techniques concerning how to determine the direction vector. This is because a different direction vector may provide a different efficiency score (Bian, 2008).

The new model with a Directional Slack-based Distance Function (DSDF) can determine an appropriate direction while obtaining a more reasonable eco-efficiency score employing the slacks-based measure. The efficiency score in this model is different from the original concept of the DDF model whereby it is determined by the additive slack of the desirable and undesirable outputs. The additive slack of the direction that is under non-radial measure is more appropriate because the DMUs can expand and contract the desirable and undesirable outputs by the different proportions given by the assumption. The proposed method will be particularly useful when the DMU want to identify the amount of undesirable output that needs to be reduced to attain full efficiency and provides a reasonable direction for the decision makers to achieve a higher target in their productivity.

4.3 Super DSDF Eco-efficiency (SDEE)

Referring back to eco-efficiency measurement with the DSDF approach, which was proposed in the previous section, a problem may occur when most of the DMUs are fully efficient or achieve a score of 1. The discrimination power among DMUs becomes problematic when the analysis has a small sample size (Chen et al., 2012). In this section, the DSDF approach with a super-efficiency model to rank the extreme DSDF score of 1 is extended. Unlike the conventional measures of super-efficiency, which is only applicable for standard input and desirable output factors using the DEA model, the super-efficiency model in this study will deal with both desirable and undesirable output factors directly as well as input factor. As far as the author is concerned, Chen et al. (2012) was the first to make an attempt to introduce the super-efficiency model with the incorporation of undesirable output directly. Chen et al. handled the situation when input and output generate both the desirable and undesirable factors.

The development of Super DSDF Eco-efficiency (SDEE) model in this section will follow the super-efficiency measurement proposed by Du et al. (2010), since Du et al. applied the technique of slack-based measure in their super-efficiency model. The model proposed by them is as follows:

$$\text{Min} \sum_{i=1}^{I} t_{im}^{-} + \sum_{j=1}^{J} t_{jm}^{+}$$

Subject to

$$\sum_{n=1, n\neq m}^{N} z_n x_{in} \leq x_{im} + t_{im}^{-}\ ;\ \ i = 1,2,\ldots,I$$

$$\sum_{n=1, n\neq m}^{N} z_n y_{jn} \geq y_{jm} - t_{jm}^{+}\ ;\ \ j = 1,2,\ldots,J$$

$$z_n, t_{im}^{-}, t_{jm}^{+} \geq 0\ ;\ \ n = 1,2,\ldots,N \tag{4.6}$$

Where t_{im}^{-} = input slack needs to be increased and t_{jm}^{+} = output slack needs to be decreased. The '–' has been assigned to the input slack vector and '+' has been assigned to the output slack vector.

Figure 4.2 below illustrates the 2-dimensional super-efficient frontier concerning desirable and undesirable outputs. For this illustration, the VRS model is employed as the convexity condition under the VRS model may illustrate a clearer picture of super-efficiency. Suppose that three efficient DMUs, namely A, B and C are composed of the original efficient frontier under the VRS model. Now, the DMU B is under evaluation and thus excluded from the efficient frontier. Then, the resulting efficient frontier AC is defined as the super-efficient frontier for DMU B. Using the super-efficiency technique, DMU B will increase or/and decrease undesirable output (*y*) and desirable output (*u*), respectively, to reach the frontier AC.

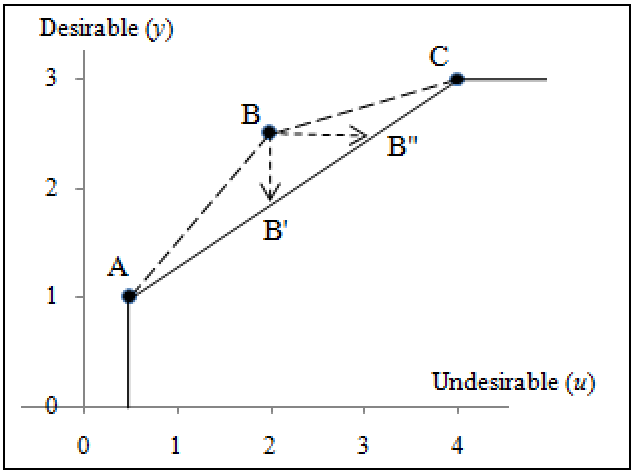


Figure 4.2: Super-efficiency frontier

Source: Johnson and McGinnis (2009)

Recall the DSDF model (4.2) proposed in previous section to compute eco-efficiency. Suppose DMU m is efficient. To obtain the super-efficiency (SE) of DMU m, the transformation to Super DSDF Eco-efficiency (SDEE) model for DMU m is as follows:

$$\text{Min } m_{SE} = \sum_{j=1}^{J} \delta_{yj} + \sum_{k=1}^{K} \delta_{uk}$$

Subject to

$$\sum_{n=1,n\neq m}^{N} z_n x_{in} \leq x_{im} \ ; \ i = 1,2,\dots,I$$

$$\sum_{n=1,n\neq m}^{N} z_n y_{jn} \geq y_{jm} - \delta_{yj} \,.\, 1 \ ; \ j = 1,2,\dots,J$$

$$\sum_{n=1,n\neq m}^{N} z_n u_{kn} \leq u_{km} + \delta_{uk} \,.\, 1 \ ; \ k = 1,2,\dots,K$$

$$z_n, \delta_{yj}, \delta_{uk} \geq 0 \ ; \ n = 1,2,\dots,N$$

(4.7)

In model (4.7), δ_{yj} and δ_{uk} are desirable and undesirable output slacks for this minimization objective function, respectively.

In model (4.7), three modifications have been made from the previous model (4.2). First, for each DMU being evaluated, the objective of the above model is to minimize the unit of slack for desirable and undesirable outputs. The objective also needs to be transformed from maximization to minimization so that the resulting model is bounded. Second, the DMU under evaluation (m) needs to be removed from the reference set, as illustrated in Figure 4.2. Third, the desirable output (δ_{yj}) and undesirable output (δ_{uk}) slacks allows the desirable output j of DMU m to decrease by δ_{yj} and allows the undesirable output *k* of DMU *m* to increase by δ_{uk} .

The constraints for input, desirable and undesirable outputs should be modified because the undesirable output need to be increased while the desirable output need to be decreased for DMU m to reach the frontier constructed by the remaining efficient DMUs. Model (4.7) for super-efficiency is only applied to the efficient DMUs so that they can be distinguished among them through the score obtained in order to rank their performance. Then, to obtain the super-efficiency score for DSDF (α_{SE}) is formulated as 1 + m_{SE}. Note that α_{SE} is greater than 1 to exhibit the super-efficiency score for DMU m. As for the inefficient DMUs, the eco-efficiency measurements have been assessed using model (4.3).

To demonstrate the SDEE model (4.7), a numerical example has been used by using single desirable and undesirable outputs while consuming the same set of input. Table 4.2 presents the numerical example for six DMUs with single desirable (y) and undesirable (u) outputs. For this example, the VRS model is employed as the convexity condition under the VRS model may illustrate a clearer picture of super-efficiency. Using model (4.2) with the additional convexity constraint of intensity variable

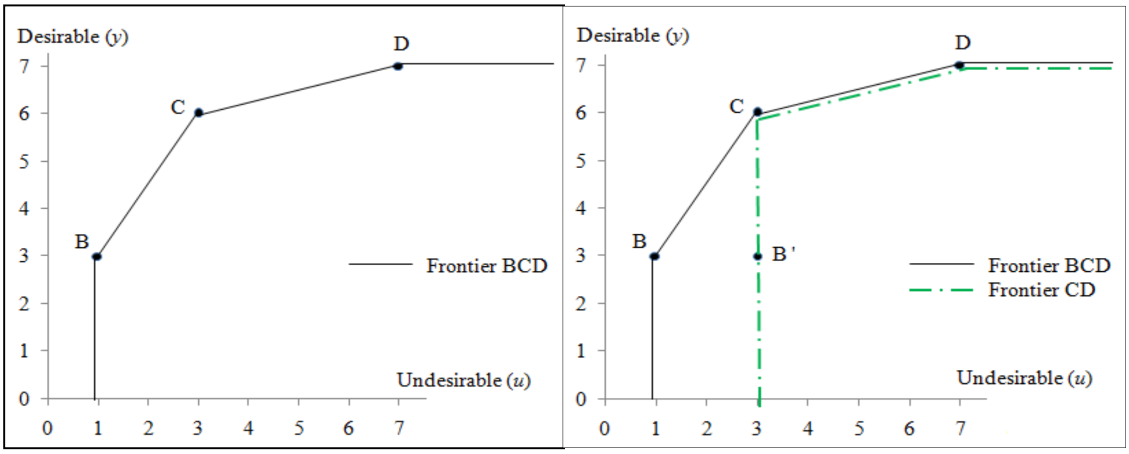
($\sum_{n=1}^{N} z_n = 1$) for the VRS model, DMU A, E and F exhibit eco-inefficiency with reported scores $(1 - \sigma_m)$ in the fourth column while the rest are efficient.

To measure the super-efficiency (α_{SE}) for DMU B, C and D, model (4.7) (m_{SE}) with the additional constraint of the intensity variable ($\sum_{n=1}^{N} z_n = 1$) has been applied. From the results reported in the sixth column, the super-efficiency score (α_{SE}) is obtained by $1 + m_{SE}$. It can be seen in the sixth column that DMU B obtains the highest score with 1.286, followed by DMU C and DMU D with 1.238 and 1.143, respectively. Now, based on these results, their performances can be ranked in the seventh column as first, second and third for DMU B, C and D, respectively. The slack value for desirable (δ_{yj}) and undesirable (δ_{uk}) outputs reported in the eighth and the ninth column can be clearly illustrated in Figure 4.3.

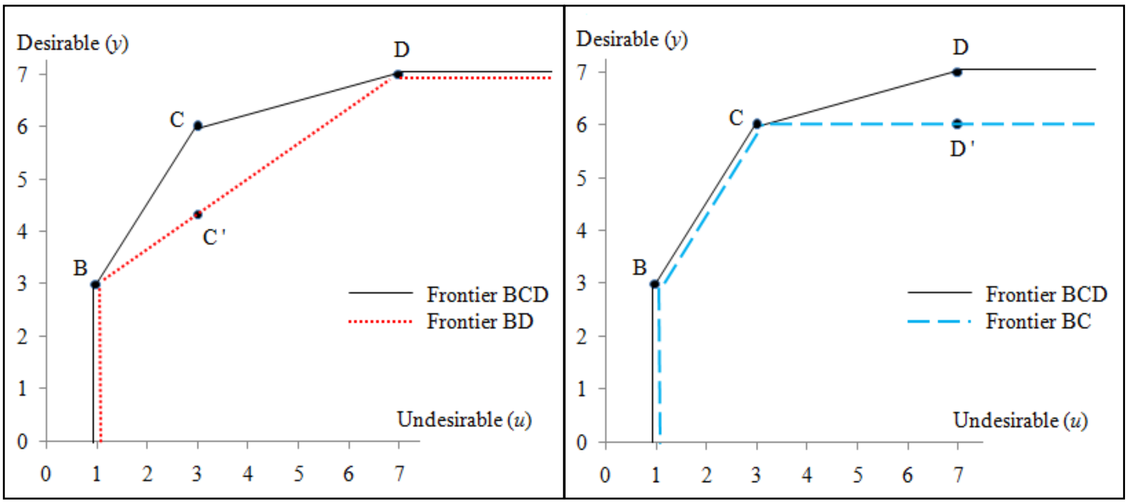
Given that this example consumes the same set of input, the input slack is not computed. From Table 4.2, it can be found that undesirable output slack (δ_{uk}) value for DMU B is 2 implying that DMU B can increase undesirable output (u) value from 1 to 3. While the desirable output slack (δ_{yj}) value for DMU C and D are 1.667 and 1 implying that DMU C and D can decrease the desirable (y) output value from 6 and 7 to 4.33 and 6, respectively. Thus, from the slack value gauged, DMU B, C and D are projected onto B' = (3,3), C' = (4.33, 3) and D' = (6,7), respectively, in Figure 4.3. The original frontier of eco-efficiency as well as the frontier of super-efficiency for each efficient DMUs (B, C and D) are also illustrated in Figure 4.3, i.e. frontier CD, BD and BC are defined for super-efficiency DMU B, C and D, respectively.

Table 4.2: Numerical example of super-efficiency

DMU	y	u	$1-\sigma_m$	m_{SE}	α_{SE}	Rank	δ_{yj}	δ_{uk}
A	2	2	0.64			5		
B	3	1	1.00	0.286	1.286	1	0	2
C	6	3	1.00	0.238	1.238	2	1.667	0
D	7	7	1.00	0.143	1.143	3	1	0
E	4	6	0.29			6		
F	3	2	0.79			4		



a) Eco-efficiency frontier for DMU B,C,D b) Frontier CD for super-efficiency DMU B



c) Frontier BD for super-efficiency DMU C d) Frontier BC for super-efficiency DMU D

Figure 4.3: Eco-efficiency frontier and super-efficiency frontier for DMU B, C and D

Source: Johnson and McGinnis (2009)

To generalize the model, equation (4.7) above can also include undesirable input as follows:

$$\text{Min } m_{SE} = \sum_{f=1}^{F} \delta_{vf} + \sum_{i=1}^{I} \delta_{xi} + \sum_{j=1}^{J} \delta_{yj} + \sum_{k=1}^{K} \delta_{uk}$$

Subject to

$$\sum_{n=1, n\neq m}^{N} z_n v_{fn} \geq v_{fm} - \delta_{vf} . 1; \; f = 1,2, \dots, F$$

$$\sum_{n=1, n\neq m}^{N} z_n x_{in} \leq x_{im} + \delta_{xi} . 1 ; \; i = 1,2, \dots, I$$

$$\sum_{n=1, n\neq m}^{N} z_n y_{jn} \geq y_{jm} - \delta_{yj} . 1 ; \; j = 1,2, \dots, J$$

$$\sum_{n=1, n\neq m}^{N} z_n u_{kn} \leq u_{km} + \delta_{uk} . 1 ; \; k = 1,2, \dots, K$$

$$z_n, \delta_{vf}, \delta_{xi}, \delta_{yj}, \delta_{uk} \geq 0 ; \; n = 1,2, \dots, N \tag{4.8}$$

In model (4.8), one more constraint has been added to undesirable input (*v*). In this model, desirable input (δ_{xi}) and undesirable output (δ_{uk}) slacks allow the desirable input *i* and undesirable output *k* of DMU *m* to increase by δ_{xi} and δ_{uk}, respectively. While undesirable input (δ_{vf}) and desirable output (δ_{yj}) slacks allow the undesirable input *f* and desirable output *j* of DMU *m* to decrease by δ_{vf} and δ_{yj}, respectively. The undesirable input and desirable output are decreased while desirable input and undesirable output are increased so that the DMU being evaluated can be projected optimally close to the frontier constructed by the remaining DMUs. Examples of undesirable input are fines in the case of library systems, time duration to reconnect the electricity supply failure and the amount of waste to be treated in the waste treatment process (Seiford & Zhu, 2002).

It has been noted that under certain conditions the standard super-efficiency model may not be solved and is said to have an infeasible solution especially when the super-efficiency model is under VRS condition (Du et al., 2010; Johnson & McGinnis, 2009; Lee et al., 2011; Lovell & Rouse, 2003; Seiford & Zhu, 1999). Even though this study utilizes the CRS assumption, in which infeasibility does not appear, it would be better to understand the issue of the infeasibility problem in super-efficiency. The infeasibility problem under the VRS condition is illustrated in Figure 4.4 below. The frontier used to measure the super-efficiency of DMU C only uses DMU A and B to construct the frontier. From the illustration, it can be seen that DMU C may appear infeasible solution to reach frontier AB.

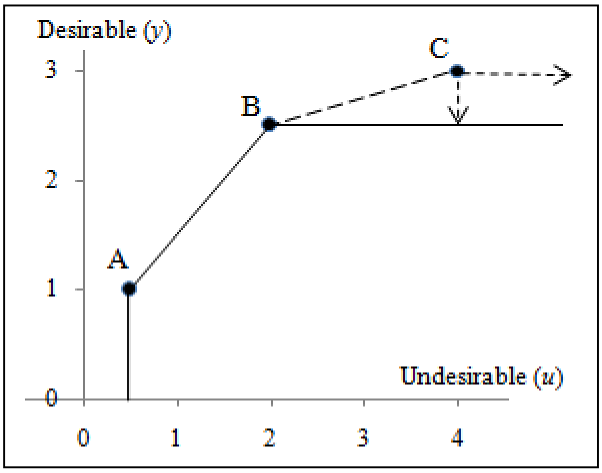


Figure 4.4: Infeasibility problem in super-efficiency model

To better understand the cause of an infeasible result for a DEA model, the linear programming can be examined. Looking at model (4.7), there are three types of constraints those related to the input, to the desirable output and to the undesirable output. By taking undesirable output constraint as an example, the equation is as below:

$$\sum_{n=1, n\neq m}^{N} z_n u_{kn} \leq u_{km} \tag{4.9}$$

From this equation, if $u_{kn} \leq u_{km}$ for all undesirable outputs for a given n then a solution to the super-efficiency model can always be found. But, if u_{km} is less than all u_{kn} values for any of the undesirable outputs in the reference set, the constraint associated with that undesirable output cannot be satisfied and the problem is infeasible.

Nevertheless, using the Super DSDF Eco-efficiency (SDEE) with the modification in model (4.7), the efficiency scores are always satisfiable and thus infeasibility is not possible using both models CRS or VRS. Adopted from Du et al. (2010), the following theorem indicating that the super DSDF eco-efficiency is also feasible.

Theorem 1. Slacks-based super-efficiency models are always feasible under the constant return to scale as well as variable return to scale assumption.

Proof. Du et al. (2010) show that slack-based super-efficiency model is feasible whereby referring to their model (4.6) for any positive set of z_n, n = 1,2,…,N, $n \neq$ m, they define:

$$t_{im}^{-} = \max\{x_{im}, \textstyle\sum_{n=1,n\neq m}^{n} z_n x_{in}\} - x_{im} \geq 0 \text{ for all } i = 1, 2, \ldots, I. \quad (4.10)$$

$$t_{jm}^{+} = y_{jm} - \min\{y_{jm}, \textstyle\sum_{n=1,n\neq m}^{n} z_n y_{jn}\} \geq 0 \text{ for all } j = 1, 2, \ldots, J. \quad (4.11)$$

They then have:

$$x_{im} + t_{im}^{-} = \max\{x_{im}, \textstyle\sum_{n=1,n\neq m}^{n} z_n x_{in}\} \geq \textstyle\sum_{n=1,n\neq m}^{n} z_n x_{in} \quad (4.12)$$

$$y_{jm} - t_{jm}^{+} = \min\{y_{jm}, \textstyle\sum_{n=1,n\neq m}^{n} z_n y_{jn}\} \leq \textstyle\sum_{n=1,n\neq m}^{n} z_n y_{jn} \quad (4.13)$$

In model (4.7), the input (*x*), desirable (*y*) and undesirable (*u*) output constraints are also feasible. For any positive set of z_n, n = 1,2,…,N, $n \neq$ m, it can be defined as follow:

$$\delta_{xi} = \max\{x_{im}, \textstyle\sum_{n=1,n\neq m}^{N} z_n x_{in}\} - x_{im} \geq 0 \text{ for all } i = 1, 2, \ldots, I \quad (4.14)$$

$$\delta_{yj} = y_{jm} - \min\{y_{jm}, \textstyle\sum_{n=1,n\neq m}^{N} z_n y_{jn}\} \geq 0 \text{ for all } j = 1, 2, \ldots, J \quad (4.15)$$

$$\delta_{uk} = \max\{u_{km}, \textstyle\sum_{n=1,n\neq m}^{N} z_n u_{kn}\} - u_{km} \geq 0 \text{ for all } k = 1, 2, \ldots, K \quad (4.16)$$

We then have:

$$x_{im} + \delta_{xi} = \max\{x_{im}, \textstyle\sum_{n=1,n\neq m}^{N} z_n x_{in}\} \geq x_{im}, \textstyle\sum_{n=1,n\neq m}^{N} z_n x_{in} \quad (4.17)$$

$$y_{jm} - \delta_{yj} = \min\{y_{jm}, \textstyle\sum_{n=1,n\neq m}^{N} z_n y_{jn}\} \leq y_{jm}, \textstyle\sum_{n=1,n\neq m}^{N} z_n y_{jn} \quad (4.18)$$

$$u_{km} + \delta_{uk} = \max\{u_{km}, \textstyle\sum_{n=1,n\neq m}^{N} z_n u_{kn}\} \geq u_{km}, \textstyle\sum_{n=1,n\neq m}^{N} z_n u_{kn} \quad (4.19)$$

When applying super-efficiency, the value for efficiency score will be greater than unity. The ranking of DMUs is based on the super-efficiency scores obtained. After computing the super-efficiency, the performance of all extremely efficient DMUs is now able to be distinguished. The highest of super-efficiency will be ranked first and the lowest super-efficiency will be ranked last among efficient DMUs. Apart from the ability to differentiate the performance of efficient DMUs, this approach may also assist the decision maker at the management level to undertake further analysis on resource allocation (Chen et al., 2012).

4.4 Malmquist Luenberger Productivity Index (MLPI)

As noted in the introduction to this chapter, the measures of efficiency of DMU provided in the DEA, DDF and DSDF models only present the efficiency of static performance. However, only concentrating on static efficiency estimates provides an incomplete view of DMUs performance over time. For this reason, the Malmquist Luenberger Index will be utilized to measure the movement of DMUs with regards to technological changes and eco-efficiency changes.

The ML index defined by Chung, et al. (1997) using DSDF model can be formulated as below

$$ML_t^{t+1} =$$

$$\left[\frac{(1+\overrightarrow{DS}_o^{t+1}(x^t,y^t,u^t;y^t,-u^t))}{(1+\overrightarrow{DS}_o^{t+1}(x^{t+1},y^{t+1},u^{t+1};y^{t+1},-u^{t+1}))}\,\frac{(1+\overrightarrow{DS}_o^{t}(x^t,y^t,u^t;y^t,-u^t))}{(1+\overrightarrow{DS}_o^{t}(x^{t+1},y^{t+1},u^{t+1};y^{t+1},-u^{t+1}))}\right]^{\frac{1}{2}} \tag{4.20}$$

Equation (4.20) can be further decomposed into two measured components of productivity change, which are eco-efficiency change (MLEFFC) and technological change (MLTC). MLEFFC represents a movement towards the best practice frontier while MLTC represents a shift in technology between *t* and *t+1*.

$$MLEFFC_t^{t+1} = \left[\frac{(1+\overrightarrow{DS}_O^t(x^t,y^t,u^t;y^t,-u^t))}{(1+\overrightarrow{DS}_O^{t+1}(x^{t+1},y^{t+1},u^{t+1};y^{t+1},-u^{t+1}))}\right] \tag{4.21}$$

$$MLTC_t^{t+1} =$$

$$\left[\frac{(1+\overrightarrow{DS}_O^{t+1}(x^t,y^t,u^t;y^t,-u^t))}{(1+\overrightarrow{DS}_O^t(x^t,y^t,u^t;y^t,-u^t))}\,\frac{(1+\overrightarrow{DS}_O^{t+1}(x^{t+1},y^{t+1},u^{t+1};y^{t+1},-u^{t+1}))}{(1+\overrightarrow{SD}_O^t(x^{t+1},y^{t+1},u^{t+1};y^{t+1},-u^{t+1}))}\right]^{\frac{1}{2}} \tag{4.22}$$

For each observation, four distance functions must be calculated in order to measure the ML productivity index. Two distance functions use observation and technology for time period *t* and *t+1* i.e. $\overrightarrow{DS}_O^t(x^t,y^t,u^t;y^t,-u^t)$ and $\overrightarrow{DS}_O^{t+1}(x^{t+1},y^{t+1},u^{t+1};y^{t+1},-u^{t+1})$, while another two use the mixed period of *t* and *t+1*, i.e. $\overrightarrow{DS}_O^t(x^{t+1},y^{t+1},u^{t+1};y^{t+1},-u^{t+1})$ and $\overrightarrow{DS}_O^{t+1}(x^t,y^t,u^t;y^t,-u^t)$. $\overrightarrow{DS}_O^t(x^{t+1},y^{t+1},u^{t+1};y^{t+1},-u^{t+1})$ compares (y^{t+1},u^{t+1}) with the production frontier at time *t* while $\overrightarrow{DS}_O^{t+1}(x^t,y^t,u^t;y^t,-u^t)$ compares (y^t,u^t) with the production frontier at time *t+1*. Using the DSDF approach in model (4.2), the solution of the four distance functions can be solved as follows:

$$\overrightarrow{DS}_O^t(x^t,y^t,u^t;y^t,-u^t) = \text{Max} \sum_{j=1}^{J}\gamma_{yj}^t + \sum_{k=1}^{K}\gamma_{uk}^t$$

Subject to

$$\begin{aligned}
&\sum_{n=1}^{N} z_n^t x_{in}^t \le x_{im}^t\,; \;\; i = 1,2,\dots,I\\
&\sum_{n=1}^{N} z_n^t y_{jn}^t \ge y_{jm}^t + \gamma_{yj}^t\,.1\,; \;\; j = 1,2,\dots,J\\
&\sum_{n=1}^{N} z_n^t u_{kn}^t = u_{km}^t - \gamma_{uk}^t.1\,; \;\; k = 1,2,\dots,K\\
&z_n^t, \gamma_{yj}^t, \gamma_{uk}^t \ge 0\,; \;\; n = 1,2,\dots,N
\end{aligned} \tag{4.23}$$

$$\overrightarrow{DS}_o^{t+1}(x^{t+1}, y^{t+1}, u^{t+1}; y^{t+1}, -u^{t+1}) = \text{Max} \sum_{j=1}^{J} \gamma_{yj}^{\mathrm{t+1}} + \sum_{k=1}^{K} \gamma_{uk}^{\mathrm{t+1}}$$

Subject to

$$\sum\nolimits_{n=1}^{N} z_n^{\mathrm{t+1}} x_{in}^{\mathrm{t+1}} \leq x_{im}^{\mathrm{t+1}} \; ; \;\; i = 1,2, \dots, I$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t+1}} y_{jn}^{\mathrm{t+1}} \geq y_{jm}^{\mathrm{t+1}} + \gamma_{yj}^{\mathrm{t+1}} \, .1 \;\; ; \;\; j = 1,2, \dots, J$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t+1}} u_{kn}^{\mathrm{t+1}} = u_{km}^{\mathrm{t+1}} - \gamma_{uk}^{\mathrm{t+1}} .1 \; ; \;\; k = 1,2, \dots, K$$

$$z_n^{\mathrm{t+1}}, \gamma_{yj}^{\mathrm{t+1}}, \gamma_{uk}^{\mathrm{t+1}} \geq 0 \; ; \;\; n = 1,2, \dots, N \tag{4.24}$$

$$\overrightarrow{DS}_o^{t}(x^{t+1}, y^{t+1}, u^{t+1}; y^{t+1}, -u^{t+1}) = \text{Max} \sum_{j=1}^{J} \gamma_{yj}^{\mathrm{t+1}} + \sum_{k=1}^{K} \gamma_{uk}^{\mathrm{t+1}}$$

Subject to

$$\sum\nolimits_{n=1}^{N} z_n^{\mathrm{t}} x_{in}^{\mathrm{t}} \leq x_{im}^{\mathrm{t+1}} \; ; \;\; i = 1,2, \dots, I$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t}} y_{jn}^{\mathrm{t}} \geq y_{jm}^{\mathrm{t+1}} + \gamma_{yj}^{\mathrm{t+1}} \, .1 \;\; ; \;\; j = 1,2, \dots, J$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t}} u_{kn}^{\mathrm{t}} = u_{km}^{\mathrm{t+1}} - \gamma_{uk}^{\mathrm{t+1}} .1 \; ; \;\; k = 1,2, \dots, K$$

$$z_n^{\mathrm{t}}, \gamma_{yj}^{\mathrm{t+1}}, \gamma_{uk}^{\mathrm{t+1}} \geq 0 \; ; \;\; n = 1,2, \dots, N \tag{4.25}$$

$$\overrightarrow{DS}_o^{t+1}(x^{t}, y^{t}, u^{t}; y^{t}, -u^{t}) = \text{Max} \sum_{j=1}^{J} \gamma_{yj}^{\mathrm{t}} + \sum_{k=1}^{K} \gamma_{uk}^{\mathrm{t}}$$

Subject to

$$\sum\nolimits_{n=1}^{N} z_n^{\mathrm{t+1}} x_{in}^{\mathrm{t+1}} \leq x_{im}^{\mathrm{t}} \; ; \;\; i = 1,2, \dots, I$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t+1}} y_{jn}^{\mathrm{t+1}} \geq y_{jm}^{\mathrm{t}} + \gamma_{yj}^{\mathrm{t}} \, .1 \;\; ; \;\; j = 1,2, \dots, J$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t+1}} u_{kn}^{\mathrm{t+1}} = u_{km}^{\mathrm{t}} - \gamma_{uk}^{\mathrm{t}} .1 \; ; \;\; k = 1,2, \dots, K$$

$$z_n^{\mathrm{t+1}}, \gamma_{yj}^{\mathrm{t}}, \gamma_{uk}^{\mathrm{t}} \geq 0 \; ; \;\; n = 1,2, \dots, N \tag{4.26}$$

In the Malmquist Lunberger Productivity Index (MLPI), the issue of infeasible solution has been discussed by other researchers (Färe et al., 2001; Jeon & Sickles, 2004; Oh, 2010). (Refer to previous chapter on the discussion of infeasibility problem for mixed period in MLPI). The infeasibility solution may also occur for MLPI when calculated by the DSDF model for two distance functions of mixed period, i.e. *t* and *t*+1.

To overcome the infeasibility problem for a mixed period in the DSDF approach, two stage analyses with multiple year “window” of data, as has been suggested by Färe et al. (2001), is employed to form a frontier of reference technology.

In the first stage, four distance functions are calculated using the new model of DSDF, i.e. equation (4.23), (4.24), (4.25) and (4.26). For mixed period calculation, which is equation (4.25) and (4.26), three-year data are used to construct the reference technology. According to Färe et al. (2001), all of the production frontiers that are calculated are derived using observations from that year and the previous two years. In other words, the reference technology for time period *t* would be constructed from data in *t*, *t* – 1 and *t* – 2 and period *t* + 1 would be constructed from data in *t*, *t* + 1 and *t* – 1. For instance, the reference technology for time period 2003 would be constructed from data between 2001 and 2003 and period 2004 would be constructed from data between 2002 and 2004. Figure 4.5 below illustrates the reference frontier using a three-year window of data.

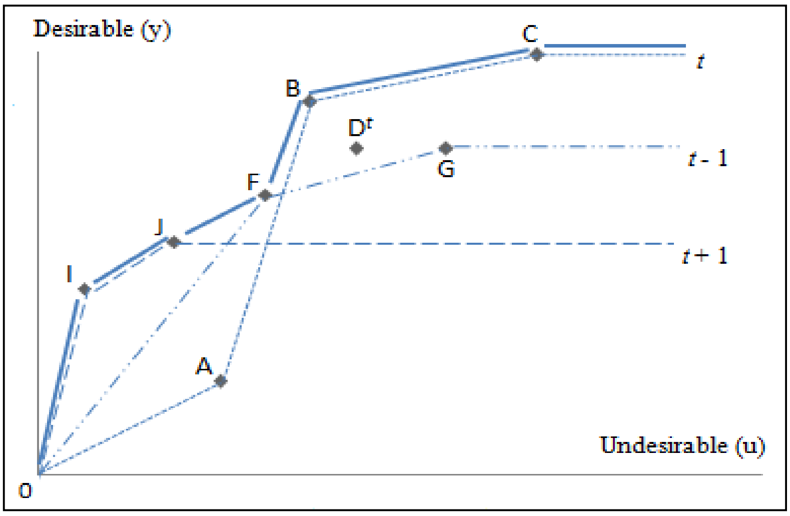


Figure 4.5: Reference frontier using a three-year window of data

As illustrated in Figure 4.5, the frontier of t is bounded by 0ABC, the frontier of t - 1 is bounded by 0FG and the frontier of t + 1 is bounded by 0IJ. To observe DMU D from period t, a three-year window of data is used to form the production frontier, i.e. t, t - 1 and $t+1$. Hence, the frontier for the three-year window of data is bounded by 0IJFBC. Using 0IJFBC frontier, the solution for DMU D for model $\overrightarrow{DS}_o^{t+1}(x^t, y^t, u^t; y^t, -u^t)$ is now feasible and can be measured.

However, the solution using a multiple year "window" of data as the reference technology simply reduces the number of infeasible solutions. There are some circumstances where the infeasible solution still exists, especially when the DMU observed is beyond the reference technology i.e. $\overrightarrow{DS}_o^{t}(x^{t+1}, y^{t+1}, u^{t+1}; y^{t+1}, -u^{t+1})$. To solve the infeasible problem, second stage analysis will be calculated using the concept of super-efficiency measurement (refer back the discussion on super-efficiency in this chapter). Using super-efficiency frontier, the infeasible DMU will increase the undesirable output and decrease the desirable output to reach the production frontier. This second stage analysis is only applied to the infeasible solution that occurs during the first stage analysis. Four distance functions are re-calculated using the Super DSDF Eco-efficiency (SDEE) model. The four distance functions that need to be re-calculated are as follows:

$$\overrightarrow{DS}_o^{t}(x^t, y^t, u^t; y^t, -u^t) = \text{Min} \sum_{j=1}^{J} \gamma_{yj}^t + \sum_{k=1}^{K} \gamma_{uk}^t$$

Subject to

$$\sum_{n=1}^{N} z_n^t x_{in}^t \leq x_{im}^t \; ; \;\; i = 1,2, \dots , I$$

$$\sum_{n=1}^{N} z_n^t y_{jn}^t \geq y_{jm}^t - \gamma_{yj}^t \, .1 \; ; \;\; j = 1,2, \dots , J$$

$$\sum_{n=1}^{N} z_n^t u_{kn}^t \leq u_{km}^t + \gamma_{uk}^t . 1 \; ; \;\; k = 1,2, \dots , K$$

$$z_n^t, \gamma_{yj}^t, \gamma_{uk}^t \geq 0 \; ; \;\; n = 1,2, \dots , N \qquad (4.27)$$

$$\overrightarrow{DS}_o^{t+1}(x^{t+1}, y^{t+1}, u^{t+1}; y^{t+1}, -u^{t+1}) = \text{Min} \sum_{j=1}^{J} \gamma_{yj}^{\mathrm{t+1}} + \sum_{k=1}^{K} \gamma_{uk}^{\mathrm{t+1}}$$

Subject to

$$\sum\nolimits_{n=1}^{N} z_n^{\mathrm{t+1}} x_{in}^{\mathrm{t+1}} \leq x_{im}^{\mathrm{t+1}} \; ; \;\; i = 1,2, \dots, I$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t+1}} y_{jn}^{\mathrm{t+1}} \geq y_{jm}^{\mathrm{t+1}} - \gamma_{yj}^{\mathrm{t+1}} \,.\, 1 \;\; ; \;\; j = 1,2, \dots, J$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t+1}} u_{kn}^{\mathrm{t+1}} \leq u_{km}^{\mathrm{t+1}} + \gamma_{uk}^{\mathrm{t+1}} . 1 \; ; \;\; k = 1,2, \dots, K$$

$$z_n^{\mathrm{t+1}}, \gamma_{yj}^{\mathrm{t+1}}, \gamma_{uk}^{\mathrm{t+1}} \geq 0 \; ; \;\; n = 1,2, \dots, N \qquad (4.28)$$

$$\overrightarrow{DS}_o^{t}(x^{t+1}, y^{t+1}, u^{t+1}; y^{t+1}, -u^{t+1}) = \text{Min} \sum_{j=1}^{J} \gamma_{yj}^{\mathrm{t+1}} + \sum_{k=1}^{K} \gamma_{uk}^{\mathrm{t+1}}$$

Subject to

$$\sum\nolimits_{n=1}^{N} z_n^{\mathrm{t}} x_{in}^{\mathrm{t}} \leq x_{im}^{\mathrm{t+1}} \; ; \;\; i = 1,2, \dots, I$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t}} y_{jn}^{\mathrm{t}} \geq y_{jm}^{\mathrm{t+1}} - \gamma_{yj}^{\mathrm{t+1}} \,.\, 1 \;\; ; \;\; j = 1,2, \dots, J$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t}} u_{kn}^{\mathrm{t}} \leq u_{km}^{\mathrm{t+1}} + \gamma_{uk}^{\mathrm{t+1}} . 1 \; ; \;\; k = 1,2, \dots, K$$

$$z_n^{\mathrm{t}}, \gamma_{yj}^{\mathrm{t+1}}, \gamma_{uk}^{\mathrm{t+1}} \geq 0 \; ; \;\; n = 1,2, \dots, N \qquad (4.29)$$

$$\overrightarrow{DS}_o^{t+1}(x^{t}, y^{t}, u^{t}; y^{t}, -u^{t}) = \text{Min} \sum_{j=1}^{J} \gamma_{yj}^{\mathrm{t}} + \sum_{k=1}^{K} \gamma_{uk}^{\mathrm{t}}$$

Subject to

$$\sum\nolimits_{n=1}^{N} z_n^{\mathrm{t+1}} x_{in}^{\mathrm{t+1}} \leq x_{im}^{\mathrm{t}} \; ; \;\; i = 1,2, \dots, I$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t+1}} y_{jn}^{\mathrm{t+1}} \geq y_{jm}^{\mathrm{t}} - \gamma_{yj}^{\mathrm{t}} \,.\, 1 \;\; ; \;\; j = 1,2, \dots, J$$

$$\sum_{n=1}^{N} z_n^{\mathrm{t+1}} u_{kn}^{\mathrm{t+1}} \leq u_{km}^{\mathrm{t}} + \gamma_{uk}^{\mathrm{t}} . 1 \; ; \;\; k = 1,2, \dots, K$$

$$z_n^{\mathrm{t+1}}, \gamma_{yj}^{\mathrm{t}}, \gamma_{uk}^{\mathrm{t}} \geq 0 \; ; \;\; n = 1,2, \dots, N$$

(4.30)

4.5 Specification on Variables Selection

It is important that an efficiency measurement must be accurate and its measures rely on the accuracy of variables. This section will provide the discussion on the important variables in the analysis activities, which are the unit of assessment, inputs as well as outputs.

4.5.1 Determination of Decision Making Unit (DMU)

The manufacturing industry has been chosen as a context of the study since this sector is the second largest contributor to the Gross Domestic Product (GDP) of Malaysia, and also one of the main contributors to environmental pollution (Department of Statistics Malaysia, 2008). According to the Department of Statistics, Malaysia, the definition of manufacturing follows the "Malaysia Standard Industrial Classification (MSIC) 2000" which can be defined as the physical or chemical transformation of materials or components into new products, whether the work is performed by power-driven machines or by hand, whether it is done in a factory or in the worker's home, and whether the products are sold at wholesale or retail.

The unit of assessment for this study will consider 15 regions throughout Malaysia known as states (including the Federal Territories of Kuala Lumpur and Labuan). The list of DMUs is shown in the table below:

Table 4.3: Lists of DMUs

No	Name of States	No	Name of States
1	Johor	9	Perlis
2	Kedah	10	Selangor
3	Kelantan	11	Terengganu
4	Melaka	12	Sabah
5	Negeri Sembilan	13	Sarawak
6	Pahang	14	Kuala Lumpur
7	Pulau Pinang	15	Labuan
8	Perak		

The 15 DMUs listed in Table 4.3 above have been categorized by industrial grouping of the state as in Table 4.4, i.e. Free Industrial Zone (FIZ) and Non-Free Industrial Zone (N-FIZ). The states that fall under the two categories are provided in Table 4.4 below.

Table 4.4: Categories of FIZ and N-FIZ states

FIZ	N-FIZ
1. Johor	1. Kedah
2. Melaka	2. Kelantan
3. Pulau Pinang	3. Negeri Sembilan
4. Perak	4. Pahang
5. Selangor	5. Perlis
	6. Terengganu
	7. Sabah
	8. Sarawak
	9. Kuala Lumpur
	10. Labuan

Under the provision of Section 3(1) of the Free Zones Act 1990, the Minister of Finance declared a Free Industrial Zone (FIZ) (which replaced the original FTZs (Free Trade Zone)) area, which was mainly designed to promote entrepot trading, and were especially established for manufacturing companies that produce or assemble products that are mainly for export. A Free Industrial Zone comprises a free commercial zone for commercial activities, which include trading (except retail trading), breaking bulk, grading, repacking, relabeling as well as transit for manufacturing activities.

The FIZ are special areas where the normal trade regulations do not apply. In other words, a free zone is referred to as a special area in which foreign or domestic companies may manufacture or assemble goods for export without being subjected to the normal custom duties on imported raw materials or exported products. Furthermore, the FIZ companies are also exempted from the payment of sales tax, excise duty and service tax.

4.5.2 Determination of Input and Output Variables

For this purpose, suggestions made by Golany and Roll (1989) that the ratio should be between two and three times the number of DMUs in relation to the total number of inputs and outputs need to be considered. These suggestions are further strengthened by Dyson et al. (2001) whereby they describe the rule of thumb to achieve a reasonable level of discrimination is that the number of unit DMUs should be at least $2(m \times s)$ where m and s are the product of the number of inputs and the number of outputs.

The selection of the initial list of variables considers a few criteria including taking the advice of the experts, looking at previous use in the literature and the evaluation of data availability. Having made the preliminary considerations, the initial list of input variables are number of establishment, intermediate input, total employment, salaries and wages paid as well as value of assets. However, from the entire list above, only two variables will be taken into account in this study for a parsimonious model.

The inputs are operating expenditure (opex) and capital. Operating expenditure covers all costs involved in the production process of the manufacturing sector including material, salaries and wages as well as electricity. Value of assets has been used as the proxy to the capital. Assets cover all goods, new or used, tangible or intangible, which have normal economic life span of more than one year (e.g. land, building, machinery and equipment, including transport equipment). The value reported is as at the end of the reference year and is according to the books of accounts of the reporting unit. It includes additions during the year and excludes assets disposed off during the year. It is net of depreciation.

All the above items are the main factors in production activities. Based on the previous studies on efficiency and productivity, it is typical to use operating expenditure and capital as inputs (Boyd et al., 2002; Ball et al., 2004; Färe et al., 2006; Managi 2006; Telle & Larsson 2007; Watanabe & Tanaka 2007).

In this study, two outputs were employed, one desirable and the other undesirable. A single desirable output is sales in the manufacturing industry while undesirable output is carbon dioxide (CO_2) emissions. Sales of manufactured products refer to the revenue from sales of products during the reference year irrespective of when the products are produced. Sales have been decided in this study to represent the desirable output even though previous literature mostly uses value added as their desirable output. Sales are more appropriate to be an indicator of the behaviour of real revenue in production (Nagar & Rajan, 2001). As for undesirable output, it has been determined that among the industrial sources of air pollution, CO_2 is the main by-product of industrial activities as the combustion of fossil fuels in the manufacturing process produces CO_2 (Lahiji & Rahim, 2011; Oggioni et al., 2011; Wu et al., 2010). Therefore, CO_2 emission has been included as an undesirable output in this analysis. The list of the input and output variables on the previous eco-efficiency studies are summarized in a table format in the Appendix A.

Even though the manufacturing sector also releases water pollution during the production activities, only the element of air pollution is taken into consideration in this study. The element of air pollution is more preferable than water pollution because of the difficulty in getting the data. Furthermore, fuel combustion in production activities is the largest contributor to air pollutant emissions. In addition, the analysis also needs to consider the limitation in terms of input and output variables so that the efficiency results are presented with a reasonable level of discrimination (Meng et al., 2008).

Studies on the manufacturing sector, especially for efficiency measurement should incorporate the emission of pollutants (undesirable output) in the analysis because not only desired outputs are produced but the factors that can contribute to poor environmental performance are also produced simultaneously. Among the elements that have been frequently used by previous studies as an undesirable output is CO_2 (Aiken & Pasurka, 2003; Arocena & Waddams Price, 2002; Färe et al., 1989; Färe et al., 2007; Korhonen & Luptacik, 2004; Lu & Lo, 2007; Sarkis & Talluri, 2004).

The table below provides a summary of the information on the input and output sets that will be employed in this technical efficiency and eco-efficiency measurement study.

Table 4.5: List of variables

Variables	Name of variables	Unit measurement	Symbol
Input	Operating cost	RM '000	x_1
	Capital	RM '000	x_2
Desirable output	Sales	RM '000	y
Undesirable output	Carbon dioxide (CO_2)	'000 tonne	u

4.5.3 Data Source

Having determined the list of variables, the next step is to collect the relevant data accurately. It is important that the data used in the efficiency measurement must be accurate. Hence, the data must be collected from the right sources and proper data collection procedures must be observed to ensure accuracy.

All the data for this study were obtained from the Department of Statistics, Malaysia. The data for CO_2 emissions are calculated based on fuel combustion. Since manufacturing sector has been chosen as a context of the study, therefore, all the data on fuel combustion in the manufacturing sector in each state has been collected as no state level data are available for the amount of CO_2 released. Fuel combustion, such as diesel oil, petrol (gasoline), fuel oil, and natural gas of the manufacturing sector, by each state, was used to calculate the carbon dioxide (CO_2) emissions. The calculations, guided by the 2006 Intergovernmental Panel on Climate Change (IPCC) guidelines for National Greenhouse Gas Inventories (Eggleston et al. 2006) are based on the total amount of fuels combusted and the averaged carbon content of the fuels. The IPCC methodology breaks the calculations of carbon dioxide emissions from fuel combustion into six steps – estimating apparent fuel consumption in original units, converting to a common energy unit, multiplying by emission factors to compute the carbon content, computing carbon stored, correcting for carbon unoxidised, and converting carbon oxidized to CO_2 emissions. A detail explanation of these six-step calculations is shown in Appendix B (See for example, Kumar Mandal & Madheswaran, 2010; Léonardi & Baumgartner, 2004; Worrell et al., 2001; for estimating the CO_2 emissions based on IPCC guidelines). During the collection and comparison of the data, visits were made to the Department of Statistics, Malaysia (DOSM) to access the data which are regarded as confidential.

Another important aspect to be considered in selecting the DMUs is the availability and accuracy of the data to be used in the evaluation. Once the inputs and outputs have been defined for DEA assessment, data on those inputs and outputs must be obtained for all DMUs. Besides data availability, their accuracy is equally important. If any data are found to be unreliable, an alternative assessment should be carried out. In addition to the availability of the data, it is equally important that these data are taken at a reasonable time period.

Table 4.6: Descriptive statistics of the data set for 15 states from 2001 to 2010

Year	Variables	Operating Expenditure RM '000	Capital RM '000	Sales RM '000	CO_2 '000 tonne
2001	Maximum	98489050	41149188	110187793	2540
	Minimum	637460	583111	737774	47
	Average	23079057	10673011	25584108	659
	Std Deviation	28914976	10716434	32076695	699
2002	Maximum	115541800	44537183	129504347	2357
	Minimum	716557	371312	861500	44
	Average	26551919	11799909	29482419	748
	Std Deviation	32855414	11568096	36426981	700
2003	Maximum	118941263	42631375	132679760	2559
	Minimum	771652	370694	921300	50
	Average	29875819	12239790	33181105	807
	Std Deviation	35394667	11184926	38968967	764
2004	Maximum	138557816	43296887	154010151	2661
	Minimum	797744	451252	948865	62
	Average	34743776	12322515	38520113	905
	Std Deviation	40174995	11567755	43892852	800
2005	Maximum	151595030	47056890	164155467	3183
	Minimum	706187	496388	1013014	78
	Average	38513291	12727631	42410008	931
	Std Deviation	43760150	12161660	46977265	879
2006	Maximum	159953895	48612138	171588785	2611
	Minimum	780788	462608	911110	63
	Average	41588413	12892192	45781038	816
	Std Deviation	46232247	12369785	49106817	713
2007	Maximum	158596681	43609154	171450677	2409
	Minimum	844913	470876	957685	34
	Average	43192341	12175438	47723233	737
	Std Deviation	49690438	11562161	53253651	652
2008	Maximum	184802175	55178878	201259643	2778
	Minimum	1004197	378155	1124429	34
	Average	47367807	13433957	52330804	792
	Std Deviation	52272212	14151672	56266812	720
2009	Maximum	174940811	51688873	187570618	2194
	Minimum	986175	425813	1050261	30
	Average	42755212	13057985	47003084	653
	Std Deviation	48195752	13274629	51469812	617
2010	Maximum	200822159	54891748	216552412	2032
	Minimum	1126528	429870	1142301	32
	Average	48005092	14027064	53333030	656
	Std Deviation	53763752	14161096	57790205	583

For the purpose of this study, the time frame of data has been set for a ten year period from 2001 until 2010. The pattern of observation for all input and output factors are best studied over a longer period because it may give an idea about important changes occurring within them. On the other hand, a short observation period may generate an incomplete picture of the DMU activities. Table 4.6 provides the descriptive statistics summary of the data set employed in this study.

Data for operating expenditure, capital and sales shown in Table 4.6 are presented in the real value expressed in monetary term (that is, in units of Malaysian currency – RM). These values have been adjusted from a nominal value to remove the effects of general price level changes between 2001 and 2010. When considering time series analysis, real values are important to be adjusted for the inflation since the measurement of purchasing power of any price changes over time. According to the Department of Statistics, Malaysia, these real values have been adjusted based on the Producer Price Index (PPI). Index in PPI is measured based on the movement that simplifies it in numerical values. The index is set to 100 and then movements are measured through a base period. For PPI, the base period is set at 2005.

To validate the input and output variables, the Pearson's coefficient of correlation is constructed. Table 4.7 presents the results of the Pearson correlation analysis. The table exhibits the relationship between all possible pairs of variables included in this analysis.

In general, all the correlation coefficients reveal meaningful correlations within pairs of variables. For instance, the correlation between operating expenditure and sales is 0.989 in 2001, implying a high score denoting a strong relationship. Most of the correlation values are high, at more than 0.7. The sales, which are desirable output variables, show relatively high correlation with the input variable of operating expenditure between 2001 and 2007 with more than 0.97. From 2008 up to 2010, the sales show highly correlated with the input variable of capital with a score of 0.968, 0.969 and 0.965, respectively. In all cases, the correlation is significant at the 0.05 level. All the correlation coefficient values are greater than 0.8 except for the capital and carbon dioxide in 2003, which is only 0.787. Therefore, the relation can be interpreted as excellent relationships. The significant correlation between the input and output variables shows that the DEA model developed does capture the important factors that influence efficiency, hence, producing reliable results.

Table 4.7: Results of correlation analysis from 2001 to 2010

Year	Variables	Operating expenditure	Capital	Sales	Carbon Dioxide
2001	Operating expenditure	1	0.930	0.989	0.945
	Capital		1	0.932	0.946
	Sales			1	0.942
	Carbon Dioxide				1
2002	Operating expenditure	1	0.928	0.984	0.942
	Capital		1	0.932	0.870
	Sales			1	0.937
	Carbon Dioxide				1
2003	Operating expenditure	1	0.864	0.984	0.950
	Capital		1	0.877	0.787
	Sales			1	0.948
	Carbon Dioxide				1
2004	Operating expenditure	1	0.875	0.981	0.932
	Capital		1	0.891	0.801
	Sales			1	0.938
	Carbon Dioxide				1
2005	Operating expenditure	1	0.903	0.980	0.892
	Capital		1	0.916	0.926
	Sales			1	0.913
	Carbon Dioxide				1
2006	Operating expenditure	1	0.893	0.990	0.949
	Capital		1	0.909	0.889
	Sales			1	0.954
	Carbon Dioxide				1
2007	Operating expenditure	1	0.892	0.971	0.875
	Capital		1	0.911	0.964
	Sales			1	0.892
	Carbon Dioxide				1

2008	Operating expenditure	1	0.957	0.985	0.879
	Capital		1	0.968	0.963
	Sales			1	0.904
	Carbon Dioxide				1
2009	Operating expenditure	1	0.966	0.987	0.865
	Capital		1	0. 969	0.935
	Sales			1	0.879
	Carbon Dioxide				1
2010	Operating expenditure	1	0.952	0.985	0.819
	Capital		1	0.965	0.912
	Sales			1	0.830
	Carbon Dioxide				1

4.6 Conclusion

Based on the work of the slack-based measure model by Färe and Grosskopf (2010a; 2010b), a DSDF approach was developed in this chapter that incorporates the undesirable output in order to measure an appropriate direction for each inefficient DMU to attain full eco-efficiency. This new approach determines the optimal direction to the frontier for each unit of analysis and provides dissimilar expansion and contraction factors en route to achieving a more accurate efficiency score.

To provide better discrimination between the DMUs, a super-efficiency with DSDF was inverted to rank the extreme DSDF score of one (fully efficient units). Using the super-efficiency approach, the performance of all extreme DMUs can be distinguished. Finally, to complete the analysis, the MLPI was used to measure the productivity performance over the period of time. To overcome the infeasibility problem for a mixed period in DSDF approach, two stage analyses with a multiple year “window” of data is employed to form a frontier of reference technology. With the MLPI, the efficiency for each state can be identified whether improved/deteriorated over time.

The following chapter will outline the employment of these techniques to examine the technical efficiency, eco-efficiency as well as the productivity change of the Malaysian manufacturing sector during the study period. This process involves running the data on the General Algebraic Modeling System (GAMS) software and obtaining the computed indices. The results will be further explored and presented in the next chapter.