Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP01/2010

DPIN Version 1.0: A Program for Decomposing Productivity Index Numbers

C.J. O'Donnell

Date: March 2010

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

DPIN VERSION 1.0: A PROGRAM FOR DECOMPOSING PRODUCTIVITY INDEX NUMBERS

by

C.J. O'Donnell

DPIN Version 1.0 is a fast user-friendly program for computing and **D**ecomposing **P**roductivity **I**ndex **N**umbers. The program uses the conceptual framework developed by O'Donnell (2008) and the data envelopment analysis (DEA) programs developed by O'Donnell (2009). This paper describes the conceptual framework and the main features of the software.

1. TOTAL FACTOR PRODUCTIVITY INDEXES

In the case of a firm[1] that uses a single input to produce a single output, productivity is usually measured as the output-input ratio. In the more general case of a firm that produces several outputs using several inputs, it is common to measure *total factor productivity* (TFP) as the ratio an aggregate output to an aggregate input. Several methods for computing these aggregates are described by O'Donnell (2008). Let Q_t and X_t denote the aggregate output and input of a firm in period t. Then the TFP of the firm in that period is simply

$$TFP_t = \frac{Q_t}{X_t}. \qquad (1)$$

The associated index number that measures the TFP of a firm in period t relative to its TFP in period 0 is:

$$TFP_{0t} = \frac{TFP_t}{TFP_0} = \frac{Q_t / X_t}{Q_0 / X_0} = \frac{Q_{0t}}{X_{0t}} \qquad (2)$$

where $Q_{0t} = Q_t / Q_0$ is an is an output quantity index and $X_{0t} = X_t / X_0$ is an input quantity index. Thus, TFP growth can be viewed as a measure of output growth divided by a measure of input growth.

Any TFP index that can be expressed in terms of aggregate quantities as in equation (2) is said to be *multiplicatively-complete* (O'Donnell, 2008). The class of multiplicatively-complete TFP indexes includes the well-known Paasche, Laspeyres, Fisher, Tornquist and Hicks-Moorsteen TFP indexes. Of these, only the Hicks-Moorsteen index can be computed without price data (all that is required is data on quantities). Another TFP index that can be computed without price data is the Malmquist TFP index of Caves, Christensen and Diewert (1982). However, the Malmquist TFP index is not multiplicatively-complete and cannot, except in restrictive special cases, be regarded as a valid measure of productivity change. **The DPIN software computes the Hicks-Moorsteen TFP index.**

2. THE COMPONENTS OF PRODUCTIVITY CHANGE

O'Donnell (2008) demonstrates that all multiplicatively-complete TFP indexes can be decomposed into a measure of technical change and several measures of efficiency change. Figure 1 illustrates the basic idea in aggregate quantity space. In this figure, the TFP of a firm in period 0 is given by the slope of the ray passing through the origin and point A, while the TFP of the firm in period t is given by the slope of the ray passing the origin and point Z. Let lower-case a and z denote the angles between the horizontal axis and the rays passing through points A and Z. Then the TFP index that measures the change in TFP between the two periods can be compactly written $TFP_{0t} = \tan z / \tan a$. This ability to write a multiplicatively-complete TFP index as the ratio of (tangent) functions of angles in aggregate quantity space is used by O'Donnell (2008) to conceptualise several alternative decompositions of TFP change. For example, let $\tan e$ denote the TFP at *any* non-negative point E. Then it is clear, both mathematically and from Figure 1, that the change in the TFP of the firm between periods 0 and t can be decomposed as $TFP_{0t} = (\tan z / \tan e)(\tan e / \tan a)$.

---

[1] `Firm' is a generic term used to refer to an individual, firm, business, state, country or any other decision-making unit.

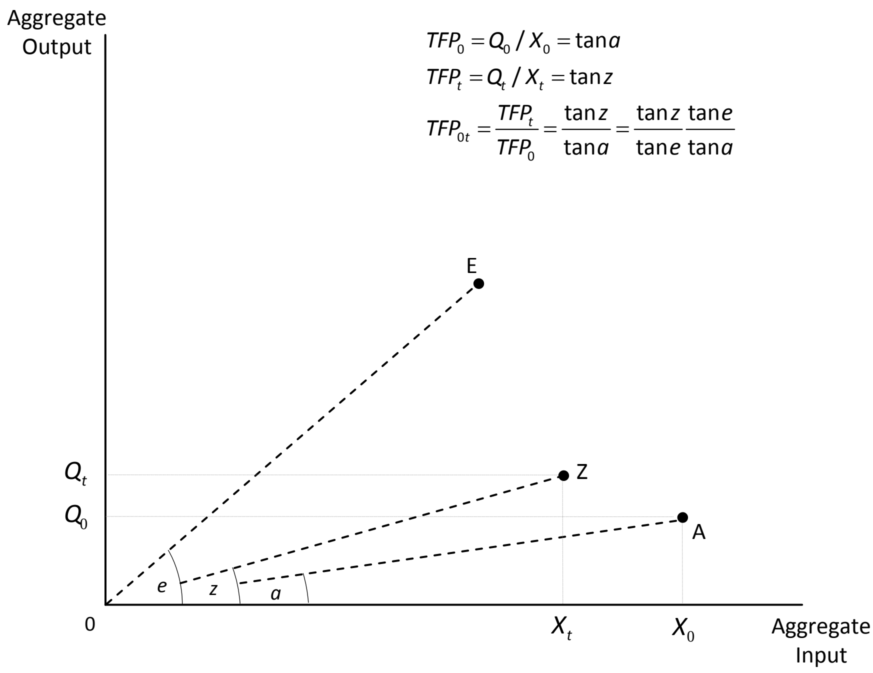

Figure 1. Measuring and Decomposing TFP Change

Within this framework, a potentially infinite number of points E can be used to effect a decomposition of a multiplicatively-complete TFP index. O'Donnell (2008) focuses on those points that feature in measures of efficiency that are common in the economics literature. Expressed in terms of aggregate quantities, the efficiency measures that feature prominently in input-oriented decompositions of TFP change are:

- *Input-oriented Technical Efficiency* (ITE) which measures the difference between observed TFP and the maximum TFP that is is possible while holding the input mix, output mix and output level fixed. This concept is illustrated in Figure 2, where the curve passing through points B and D is the frontier of a "mix-restricted" production possibilities set. The production possibilities set is mix-restricted in the sense that it only contains (aggregates) of input and output vectors that can be written as scalar multiples of the input and output vectors at point A. ITE is a ratio measure of the horizontal distance from point A to point B. Equivalently, it is a measure of the difference in TFP at points A and B: $ITE_0 = \tan a / \tan b$.

- *Input-oriented Scale Efficiency* (ISE) which measures the difference between TFP at a technically-efficient point and the maximum TFP that is possible while holding the input and output mixes fixed (but allowing the levels to vary). This measure of efficiency is represented in Figure 2 as a movement from point B to point D: $ISE_0 = \tan b / \tan d$. O'Donnell (2008) refers to point D as the point of mix-invariant optimal scale (MIOS).

- *Residual Mix Efficiency* (RME) which measures the difference between TFP at a point on a mix-restricted frontier and the maximum TFP possible when input and output mixes (and levels) can vary. This measure of efficiency is represented in Figure 2 as a movement from point D to point E: $RME_0 = \tan d / \tan e$. The curve passing through E is the frontier of an unrestricted production possibilities set (unrestricted in the sense that there are no restrictions on input or output mix). The use of the term "mix" in this context is self-evident – the movement from point D to point E is a movement from an optimal point on a mix-restricted frontier to an optimal point on a mix-unrestricted frontier, so the difference in TFP is essentially a mix-effect. O'Donnell (2008) also uses the term "residual" here because i) this movement may also involve a change in scale and ii) when comparing TFP at point A with TFP at the point of maximum productivity (point E), RME is the component that remains after accounting for pure technical and scale efficiency effects.

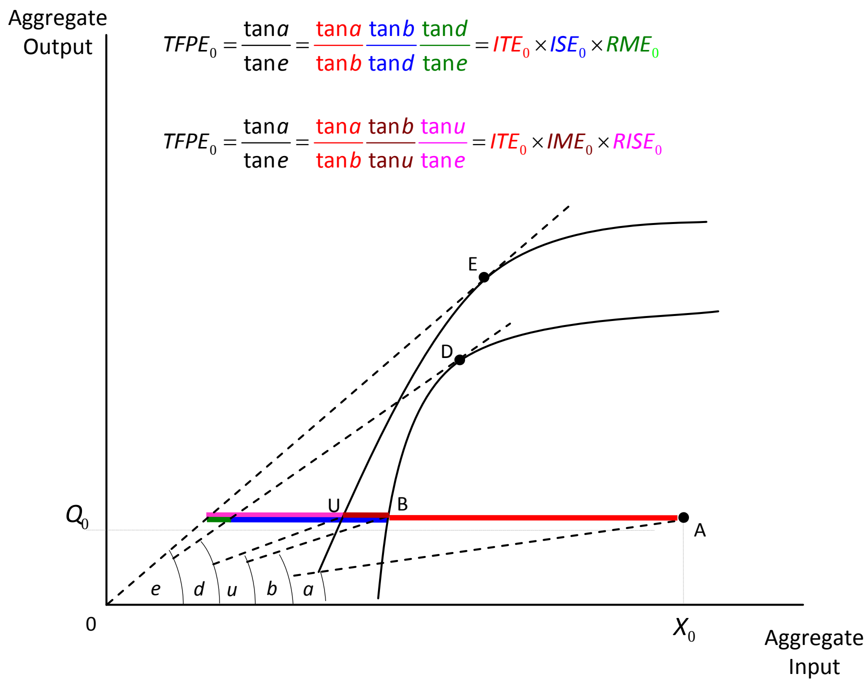

Figure 2. An Input-Oriented Decomposition of TFP Efficiency

- *Input-oriented Mix Efficiency* (IME) which measures the difference between TFP at a technically-efficient point on the mix-restricted frontier and the maximum TFP that is possible while holding the output level fixed. This measure of efficiency is represented in Figure 2 as a movement from point B to point U: $IME_0 = \tan b / \tan u$.

- *Residual Input-oriented Scale Efficiency* (RISE) which measures the difference between TFP at a technically- and mix-efficient point and TFP at the point of maximum productivity. The term scale is used because any movement around an unrestricted production frontier is a movement from one mix-efficient point to another, so any improvement in TFP is essentially a scale effect. However, the term residual is also used because, even though all the points on the unrestricted frontier are mix-efficient, they may nevertheless have different input and output mixes. Thus, what is essentially a measure of scale efficiency may contain a residual mix effect. The term residual is also appropriate in the sense that if we are interested in decomposing the difference between TFP at the observed point A and TFP at the point of maximum productivity E, then residual scale efficiency is the component that remains after we have accounted for pure technical and pure mix efficiency effects. This measure of efficiency is represented in Figure 2 as a movement from point B to point U: $RISE_0 = \tan u / \tan e$.

- *TFP Efficiency* (TFPE) which measures the difference between observed TFP and the maximum TFP possible using the available technology. This measure of efficiency is represented in Figure 2 as a movement all the way from point A to point E: $TFPE_0 = TFP_0 / TFP_0^* = \tan a / \tan e$ where TFP_0^* denotes the maximum TFP possible using the technology available in period 0.

Figure 2 illustrates two of many pathways from A to E and therefore two of many decompositions of TFP efficiency:

(3) $$TFPE_0 = \frac{TFP_0}{TFP_0^*} = ITE_0 \times ISE_0 \times RME_0$$

(4) $$TFPE_0 = \frac{TFP_0}{TFP_0^*} = ITE_0 \times IME_0 \times RISE_0$$

O'Donnell (2008) also presents several output-oriented decompositions of TFP efficiency in terms of pathways from point A to point E. Such decompositions provide a basis for an output or input-oriented decomposition of *any* multiplicatively-complete TFP index. The easiest way to see this is to rewrite (3) as $TFP_0 = TFP_0^* \times ITE_0 \times ISE_0 \times RME_0$. A similar equation holds for the firm in period *t*: $TFP_t = TFP_t^* \times ITE_t \times ISE_t \times RME_t$. It follows that

(5) $$TFP_{0t} = \frac{TFP_t}{TFP_0} = \left(\frac{TFP_t^*}{TFP_0^*}\right) \times \left(\frac{ITE_t}{ITE_0} \times \frac{ISE_t}{ISE_0} \times \frac{RME_t}{RME_0}\right)$$

The first term in parentheses on the right-hand side of equation (4) measures the difference between the maximum TFP possible using the technology available in period *t* and the maximum TFP possible using the technology available in period 0. Thus, it is a natural measure of *technical change.* The economy/industry experiences technical progress or regress as this term is greater than or less than 1. The other ratios on the right-hand side of (5) are measures of technical efficiency change, scale efficiency change, and residual mix efficiency change. O'Donnell (2008) derives the output-oriented counterparts to equations (3) and (4) and demonstrates that the input- and output-oriented measures of technical change are plausibly identical. **DPIN computes the components of (5) together with the components of the following alternative decompositions of TFP change**:

(6) $$TFP_{0t} = \frac{TFP_t}{TFP_0} = \left(\frac{TFP_t^*}{TFP_0^*}\right) \times \left(\frac{ITE_t}{ITE_0} \times \frac{IME_t}{IME_0} \times \frac{RISE_t}{RISE_0}\right)$$

(7) $$TFP_{0t} = \frac{TFP_t}{TFP_0} = \left(\frac{TFP_t^*}{TFP_0^*}\right) \times \left(\frac{OTE_t}{OTE_0} \times \frac{OSE_t}{OSE_0} \times \frac{RME_t}{RME_0}\right)$$

(8) $$TFP_{0t} = \frac{TFP_t}{TFP_0} = \left(\frac{TFP_t^*}{TFP_0^*}\right) \times \left(\frac{OTE_t}{OTE_0} \times \frac{OME_t}{OME_0} \times \frac{ROSE_t}{ROSE_0}\right)$$

where OTE, OSE, OME and ROSE are output-oriented measures of technical, scale, mix and residual scale efficiency that are analogous to the input-oriented measures.

3. THE DPIN SOFTWARE

The DPIN software uses the DEA programs developed by O'Donnell (2009) to compute and decompose the Hicks-Moorsteen TFP index. The methodology and the software can be used

- when no price data is available (i.e., when only input and output quantity data is available);
- for firms that operate in any market environment (e.g., regulated industries, perfectly competitive industries);
- for firms that have any behavioural objective (e.g., maximise profit, maximise output);
- when the technology exhibits constant or variable returns to scale; and
- to either allow for technical regress or prohibit technical regress.

Two versions of the software are available:

- a standard (**free**) version which is available at http://www.uq.edu.au/economics/cepa/dpin.htm. The standard version can analyse a balanced panel comprising **no more than 45 firms and no more than 350 observations** (i.e., firms x periods).
- a professional (commercial) version that can be obtained by contacting c.odonnell@economics.uq.edu.au. The professional version can analyse a balanced panel with no limit on the numbers of firms or observations.

The program was written[2] in C++ and compiled using the mingw compiler within the Dev-C++ integrated development environment (IDE). The program is designed to run on a Windows XP or Vista platform. Installing DPIN involves extracting the files in **DPIN.zip** into any directory. **DPIN.zip** contains the following files:

DPIN.exe	the executable file
DPIN Terms and Conditions.pdf	disclaimer
DPIN Users Guide.pdf	this document
eg1_input.csv	an example input file (see below)
eg2_input.csv	a second example input file (see below)
rice_input.csv	an input file to analyse the rice data in Coelli et al. (2005).

Running DPIN involves two steps: creating an input file and running the executable file.

Creating the DPIN Input File

DPIN input files must be created within Microsoft EXCEL and saved in a .csv (comma-delimited) format. The input file contains both commands and data. To illustrate, Figure 1 is a screenshot of the file eg1_input.csv required to run the numerical example in O'Donnell (2008). The rows before the end command are "command rows"; the row immediately after the end command is a "header row"; the remaining rows are "data rows". DPIN will read the text and data in columns A and B of every row until it encounters an end command. It will then skip the end command and the header row and start reading the data. The program is case-sensitive and will only recognise the following text strings in the command rows:

Periods	to specify the number of periods
Firms	to specify the number of firms
Inputs	to specify the number of inputs
Outputs	to specify the number of outputs
NoTechRegress	to prohibit technical regress (optional)
CRS	to impose constant returns to scale (optional)

The command rows can be listed in any order (e.g., the number of outputs can be specified first, second, or last). The program will not read the text in the header row – this row is simply used to conveniently record the names of the outputs and inputs. The output and input data must be stored in the data rows in a particular format:

- the first output variable must be stored in column D
- all output variables must be stored to the left of the input variables;
- the data must be a balanced panel;
- the data must be sorted first by period and then by firm;
- the observation, firm and period identifiers in columns A, B and C must be numeric (e.g., 1, 2, ..., not Jan, Feb ...); and
- to avoid numerical problems it is advisable to normalise all output and input variables to have unit means.

The DPIN default settings are to allow for technical regress and variable returns to scale. To prevent technical regress, insert a command row with NoTechRegress in column A. To impose constant returns to scale, insert a command row with CRS in column A. Figure 2 is a screenshot of the input file used to both prevent technical regress and impose constant returns to scale.

Running the Executable File

The DPIN executable file is DPIN.exe. Double-clicking DPIN.exe will open windows similar to those depicted in Figures 3 to 5. Figure 3 is a command window that reports program and run-time information; Figure 4 is a window used to browse and select the input file; and Figure 5 is the command window as it appears shortly after the input file eg1_input.csv has been selected. Results are written to EXCEL output files having extensions _output.csv. The

[2] The DPIN program was originally written in GAUSS by C. O'Donnell and was translated into C++ by F. Wrathmall and C. O'Donnell.

name assigned by DPIN to the output file is reported in the command window prior to exit. In the professional version of DPIN, distances are written to files having extensions _distances.csv.

Runtime Errors

In both versions of DPIN, error diagnostics are written to log files having extensions _log.txt. In log files (and the command window) the general error code for failure to solve a linear program is -1. O'Donnell (2009) explains why some linear programs may be infeasible and fail to solve. The error code for exceeding the maximum number of simplex iterations is 5. The maximum number of iterations may be reached if the linear program is degenerate. For details on degeneracy and cycling in linear programs see Winston (2004, p. 168-171).

Interpreting the DPIN Output

DPIN computes and reports measures of output- and input-oriented technical, scale and mix efficiency for every firm in every time period in the sample. It also computes and reports indexes of TFP change, technical efficiency change, and output- and input-oriented measures of technical, scale and mix efficiency change. To illustrate, Figure 6 presents the output file eg1_output.csv obtained by running DPIN on the input file eg1_input.csv. For an interpretation of these results, see O'Donnell (2009, p. 20, 21, 34, 35). Note that Column K in the output file reports *new* points of maximum productivity. If technical regress is prohibited and there is no MaxTFP value reported for a particular period then this is because the point of maximum productivity is unchanged from the previous period. In such cases, the index of technical change is 1. The missing value code used in output files is -9999. Missing values occur when a linear program cannot be solved.

Further Information

For further information, email c.odonnell@economics.uq.edu.au eg1_input.csv

	A	B	C	D	E	F	G
1	Periods	4					
2	Firms	5					
3	Inputs	2					
4	Outputs	2					
5	end						
6	Obs	Period	Firm	Y1	Y2	X1	X2
7	1	1	1	1	2	2	4
8	2	1	2	2	4	4	8
9	3	1	3	3	6	3	6
10	4	1	4	4	8	5	10
11	5	1	5	5	10	6	12
12	6	2	1	2	4	2	4
13	7	2	2	4	8	4	8
14	8	2	3	6	12	3	6
15	9	2	4	8	16	5	10
16	10	2	5	10	20	6	12
17	11	3	1	3	6	3	6
18	12	3	2	6	3	3	6
19	13	3	3	6	12	3	6
20	14	3	4	6	6	3	6
21	15	3	5	3	3	3	6
22	16	4	1	3	6	2	2
23	17	4	2	3	6	4	4
24	18	4	3	3	6	6	6
25	19	4	4	3	6	8	8
26	20	4	5	3	6	10	10
27							

Figure 1. Input file for a Model that Permits Technical Regress and Assumes VRS.

eg2_input.csv

	A	B	C	D	E	F	G
1	Periods	4					
2	Firms	5					
3	Inputs	2					
4	Outputs	2					
5	CRS						
6	NoTechRegress						
7	end						
8	Obs	Period	Firm	Y1	Y2	X1	X2
9	1	1	1	1	2	2	4
10	2	1	2	2	4	4	8
11	3	1	3	3	6	3	6
12	4	1	4	4	8	5	10
13	5	1	5	5	10	6	12
14	6	2	1	2	4	2	4
15	7	2	2	4	8	4	8
16	8	2	3	6	12	3	6
17	9	2	4	8	16	5	10
18	10	2	5	10	20	6	12
19	11	3	1	3	6	3	6
20	12	3	2	6	3	3	6
21	13	3	3	6	12	3	6
22	14	3	4	6	6	3	6
23	15	3	5	3	3	3	6
24	16	4	1	3	6	2	2
25	17	4	2	3	6	4	4
26	18	4	3	3	6	6	6
27	19	4	4	3	6	8	8
28	20	4	5	3	6	10	10
29							

Figure 2. Input file for a Model that Prohibits Technical Regress and Assumes CRS.

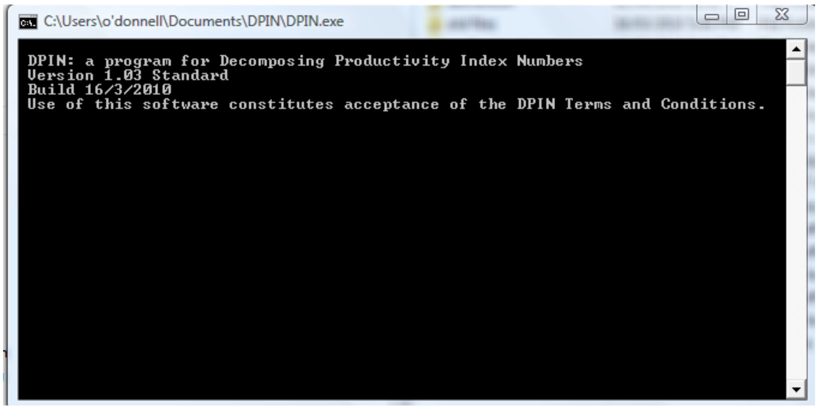

Figure 3. Initial Command Window

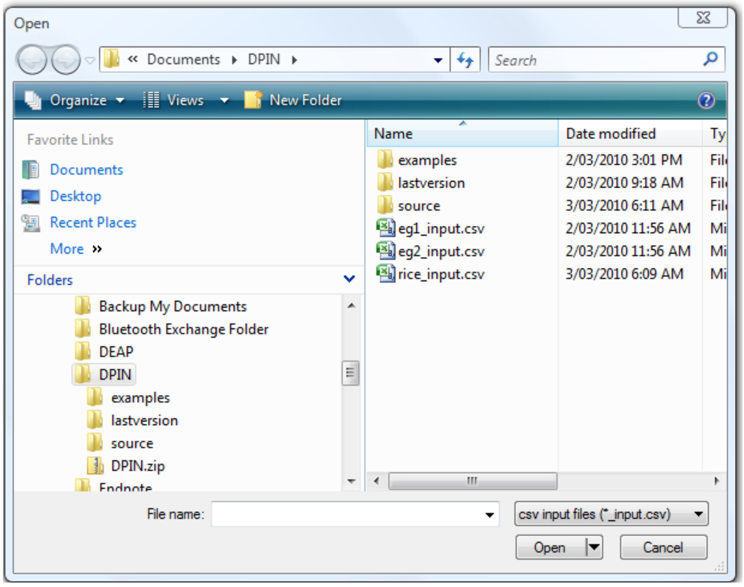

Figure 4. Input File Selection Window

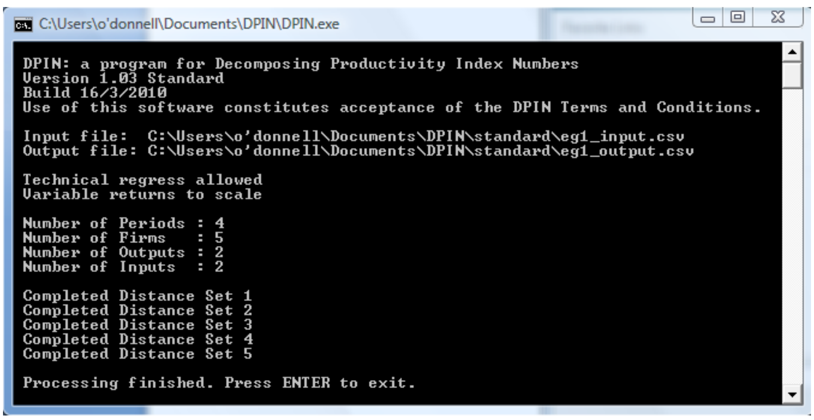

Figure 5. Command Window after Data Processing

Obs	Period	Firm		Scores								Indexes: firm n in period t relative to firm n in period t-1											
i	t	n		OTE	OSE	OME	ITE	ISE	IME	Max TFP		dTFP	dTech	dEff	dOTE	dOSE	dOME	dROSE	dITE	dISE	dIME	dRISE	dRME
1	1	1		1	0.5	1	1	0.5	1														
2	1	2		0.5455	0.9167	1	0.625	0.8	1														
3	1	3		1	1	1	1	1	1	1													
4	1	4		0.9231	0.8667	1	0.9	0.8889	1														
5	1	5		1	0.8333	1	1	0.8333	1														
6	2	1		1	0.5	1	1	0.5	1			2	2	1	1	1	1	1	1	1	1	1	1
7	2	2		0.5455	0.9167	1	0.625	0.8	1			2	2	1	1	1	1	1	1	1	1	1	1
8	2	3		1	1	1	1	1	1	2		2	2	1	1	1	1	1	1	1	1	1	1
9	2	4		0.9231	0.8667	1	0.9	0.8889	1			2	2	1	1	1	1	1	1	1	1	1	1
10	2	5		1	0.8333	1	1	0.8333	1			2	2	1	1	1	1	1	1	1	1	1	1
11	3	1		0.5	1	1	1	0.5	1			1	1	1	0.5	2	1	2	1	1	1	1	1
12	3	2		1	1	0.5	1	1	1			2	1	2	1.8333	1.0909	0.5	2.1818	1.6	1.25	1	1.25	1
13	3	3		1	1	1	1	1	1	2		1	1	1	1	1	1	1	1	1	1	1	1
14	3	4		1	1	0.6667	1	1	1			1.25	1	1.25	1.0833	1.1538	0.6667	1.7308	1.1111	1.125	1	1.125	1
15	3	5		0.5	1	0.6667	1	0.5	1			0.6	1	0.6	0.5	1.2	0.6667	1.8	1	0.6	1	0.6	1
16	4	1		1	1	1	1	1	1	2.25		2.1213	1.125	1.8856	2	1	1	0.9428	1	2	1	1.8856	0.9428
17	4	2		1	0.5	1	0.5	1	1			0.5303	1.125	0.4714	1	0.5	2	0.2357	0.5	1	1	0.9428	0.9428
18	4	3		1	0.3333	1	0.3333	1	1			0.3536	1.125	0.3143	1	0.3333	1	0.3143	0.3333	1	1	0.9428	0.9428
19	4	4		1	0.25	1	0.25	1	1			0.2652	1.125	0.2357	1	0.25	1.5	0.1571	0.25	1	1	0.9428	0.9428
20	4	5		1	0.2	1	0.2	1	1			0.4243	1.125	0.3771	2	0.2	1.5	0.1257	0.2	2	1	1.8856	0.9428

Figure 6. Output File

References

Caves, D. W., L. R. Christensen and W. E. Diewert (1982). "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity." *Econometrica* 50(6): 1393-1414.

Coelli, T. J., D. S. P. Rao, C. J. O'Donnell and G. E. Battese (2005). *An Introduction to Efficiency and Productivity Analysis*. New York, Springer.

O'Donnell, C. J. (2008). An Aggregate Quantity-Price Framework for Measuring and Decomposing Productivity and Profitability Change. *Centre for Efficiency and Productivity Analysis Working Papers*. University of Queensland.

O'Donnell, C. J. (2009). Measuring and Decomposing Agricultural Productivity and Profitability Change. *53rd Annual Conference of the Australian Agricultural and Resource Economics Society*. Cairns, Australia.

Winston, W. L. (2004). *Operations Research: Applications and Algorithms*. Belmont CA, Brooks Cole.