International Journal of Production Research
Vol. 49, No. 16, 15 August 2011, 4761–4785

A decomposition of productivity change in the semiconductor manufacturing industry

Chia-Yen Lee and Andrew L. Johnson*

Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77840, USA

(*Received 4 November 2009; final version received 23 April 2010*)

This study divides a production system into three components: production design, demand support, and operations. Efficiency is then decomposed via network data envelopment analysis and integrated into the Malmquist Productivity Index framework to develop a more detailed decomposition of productivity change. The proposed model can identify the demand effect and the identity of the root cause of technical regress. Specifically, the demand effect allows the source of technical regress to be attributed to both demand deterioration and technical regress in the production technology. An empirical study using data from 1995 to 2000 for the semiconductor manufacturing industry is presented to demonstrate and validate the proposed method. The result shows that the regress of productivity in 1997–1998 and 1999–2000 is mainly caused by demand fluctuations rather than by technical regression in production capabilities.

Keywords: productivity change; efficiency decomposition; Malmquist Productivity Index; semiconductor manufacturing

1. Introduction

In manufacturing processes, productivity analysis is a technique used to assess performance and to search for improvement alternatives. The efficient frontier can be constructed to characterise how efficiently production processes use inputs to generate outputs; given the same input resource, inefficiency is indicated by lower levels of system output. However, in practice, the decrease of actual output sometime results from insufficient demand. Demand fluctuations can bias productivity analysis. Similarly, in panel data analysis, the Malmquist Productivity Index (MPI) quantifies efficiency change and technology change over time. Technical regress is often attributed to production issues, when in reality it may be a result of demand deterioration. Thus, productivity analysis attributes changes in demand to production. The proposed model in this study can separate the demand effect and production technology effect to eliminate the bias interpretation of efficiency.

There is a limited literature discussing the effects of demand in productivity analysis. Fielding *et al.* (1985) study the performance evaluation of transportation systems. They distinguish between the production process and the consumption process, arguing that output consumption is substantially different from output production since transportation

*Corresponding author. Email: ajohnson@tamu.edu

ISSN 0020–7543 print/ISSN 1366–588X online

DOI: 10.1080/00207543.2010.497507
http://www.informaworld.com

services cannot be stored. They propose various performance indicators, specifically, service effectiveness, which is the ratio of passenger trip miles over vehicle operating miles. However, single factor productivity indicators do not represent all factors in the production system (Chen and McGinnis 2007). Lan and Lin (2005) and Yu and Lin (2008) use data envelopment analysis (DEA) and network DEA models to characterise a consumption process. However, demand and production processes characterised in transportation are different from the manufacturing industry, where manufacturers commonly depend on forecasted or contract demand based on expected sales or actual sales respectively. Longer production lead times require an 'internal' demand-supporting process; in contrast, transportation companies mainly rely on non-contract demand requested informally by customers after production, and the services must be consumed by customers immediately or they are no longer useful. We note, too, that previous studies focus on a cross-sectional analysis and do not provide estimates of productivity change over time.

Change in demand can also effect the measurement of productivity changes over time as estimated through frontier shifts indicating either technical progress or regress. Nishimizu and Page (1982) propose the first decomposition of total factor productivity change, and Färe *et al.* (1992, 1994) develop the explicit measurement of productivity change based on the MPI proposed by Caves *et al.* (1982), which uses Shephard's input distance function (Shephard 1953) to estimate inefficiency non-parametrically. The productivity change estimated via MPI can be decomposed into two sub-indices: change in efficiency and change in technology. This decomposition provides useful information in industrial application. Färe *et al.* (1992) apply the change in scale decomposition to Swedish pharmacies between 1980–1989, finding that during the latter part of the 1980s the positive productivity change is mainly due to shifts of frontier rather than changes in efficiency. In another study, 17 OECD (Organisation for Economic Cooperation and Development) countries are analysed in terms of gross domestic product, capital stock, and labour between 1979 and 1988 and an additional scale component is introduced (Färe *et al.* 1994). The results show that all of the productivity growth is chiefly due to technical change, with Japan having the highest productivity growth. Chang *et al.* (2008), who analyse performance evaluation in printed circuit board manufacturers between 2002 and 2003, find that manufacturing processes with lower efficiency and a lower MPI should be suggested for outsourcing, because they cause low capacity utilisation of expensive equipment. Several researchers develop further decomposition of productivity change, e.g. Tulkens and Vanden Eeckaut (1995), Ray and Desli (1997), Sueyoshi and Aoki (2001), Sueyoshi and Goto (2001) and Lovell (2003).

Other studies evaluate semiconductor manufacturers. Chang and Chen (2008) employ a slack-based DEA approach with two inputs (book value of tooling and cost of goods sold) and two outputs (sales revenue and average yield rate) to measure the performance of the lead frame companies at the interface between the upstream wafer and downstream printed circuit board. Their results aid the assembly/testing departments in improving supplier selection decisions, and offer managerial insights about process improvement. Lu and Hung (2010) assess the performance of vertically disintegrated firms and provide insights about the contributions of each firm to the supply chain. Their results show that efficiency can be improved by applying a consolidating strategy to achieve optimal scale and to reduce the labour force due to input congestion.

Unlike the literature above, this paper models the intermediate process by developing a decomposition that includes production facility design efficiency, sales process efficiency, and operational efficiency, while also accounting for potential frontier shifts over time. This three-phase process describes a decomposition of the black box between input and output in a production system.

First, production design efficiency measures the production capability for a given facility design. This stage assumes the facility will have enough demand for production, and will operate efficiently. The design phase has a long-term impact on production performance.

Second, the efficiency of the sales process quantifies the ability of the sales group to create enough demand to keep the facility operating at full capacity. Traditional productivity analysis assumes all deviations from the efficient frontier are attributed to inefficiency in the production system. Thus, insufficient demand may bias productivity analysis under this assumption.

Third, the operational efficiency is identified as the difference between the production level expected, given the demand, and the observed output that may be reduced by scheduling inefficiencies, machine breakdowns, inconsistent operational performance, etc. Such inefficiency in the semi-conductor industry is commonly referred to as yield loss.

This paper is organised as follows. Section 2 proposes the decomposition of the production system and explicitly quantifies the role of demand in efficiency analysis. Section 3 describes a method to estimate peak output via rolling time window and a sequential model, and then introduces a network DEA model for efficiency decomposition. Section 4 focuses on productivity change and reviews both the MPI and Shephard's distance function (Shephard 1953), while integrating demand into a decomposition of the MPI. Section 5, an empirical study of the semiconductor manufacturing industry, develops recommendations for productivity improvement based on the results of our proposed productivity change analysis. Section 6 summarises the research.

2. Production system decomposition

The production system comprises three phases, production design, demand support, and operations, and this section will describe the system decomposition. A network DEA model is proposed to model the system; thus, the necessary linking variables are defined. Figure 1 shows the decomposition of the three phases.

The first phase, production design, defines the maximal output of the production system with respect to capital investments. Inefficiency in this phase results from poor production design. The second phase describes demand support, where the sales group tries to sell enough products to keep the facility at full operation. The inefficiency in this phase results from insufficient demand, namely, production levels drop due to a lack of

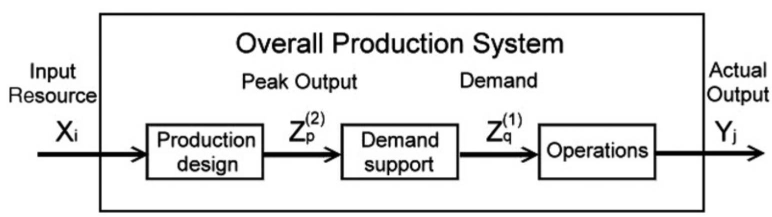

Figure 1. System process decomposition.

demand even though the production capacity is available. The third phase, operations, transforms raw materials into final goods. The inefficiency of this phase results from the poor integration of operational behaviour. In the semiconductor industry, the term 'yield' is usually employed in practice to describe the percentage of usable products resulting from the production process. In contrast, the percentage of product lost, or 'yield loss', is the result of inefficiency of operations.

This paper uses the following five metrics.

(1) *Input resources* are the items used to build up the infrastructure of the production system and support the operations of the production process.
(2) *Peak output,* the maximal output firms can achieve, characterises the 'real capability' of the production system.
(3) *Demand* is the quantity of product or output the customer is willing to consume at the current industry price. In this study the demand is estimated by product start. A *product start* is the release of raw materials to the production process. In the semiconductor manufacturing industry, product starts (wafer starts) are used to control the output level to match production and demand levels.
(4) *Actual output* is the total of final products generated.

Input resource, product start, and actual output are typically collected directly from the historical database, but peak output must be estimated (potential methods are described in Section 3.1). Using the four metrics, the efficiency of the three sub-processes (production design, demand support, and operations) can be estimated respectively.

Figure 2 illustrates the two possible scenarios that can occur between consumer and producer:

(1) demand surplus occurs where the demand for a product exceeds the supply level, or alternatively,
(2) demand shortage occurs when the demand realised is less than the supply that can be produced by the facility.

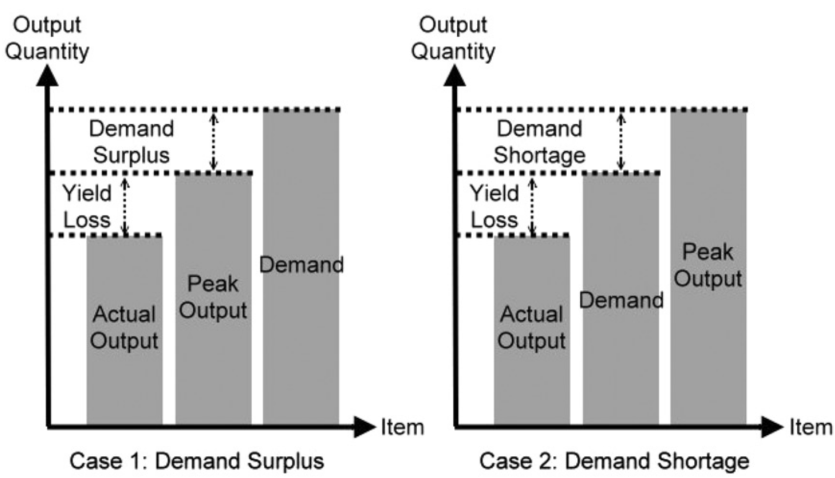

Figure 2. Scenarios of demand surplus and shortage.

In the first case, a firm may add more raw materials to the system. However, if the system was previously operating optimally, the additional materials will lead to a higher work-in-process (WIP) and increase the product cycle time (Hopp and Spearman 2001). Thus, the production system would need to extend its operating hours or outsource the additional demand. In the second case, a firm will attempt to match demand by controlling the number of product starts[1]. We will focus on the second scenario, demand quantity is less than peak output of the production system. However, demand shortage will be underestimated for a firm that matches demand to actual output.

3. Measurement of efficiency decomposition

3.1 *Peak output estimation*

To apply the network DEA model suggested for production system decomposition, it is necessary to quantify peak output. Two ways are suggested in the literature: a rolling time window analysis (Charnes *et al.* 1985) and a sequential model (Diewert 1992).

Rolling time window analysis estimates peak output via shifting time windows. It postulates no technical change within any time window. Given a certain fixed number of periods that define the time window, all observations of the production processes during that window are compared in a single analysis (Charnes *et al.* 1985). Peak output is estimated by using an output-oriented (CRS) DEA (Charnes *et al.* 1978), and a reference set constructed from only the observations of the production process under analysis within the time window. Let X_{irt} be the ith input resource of firm r in tth period, $Z^{(1)}_{qrt}$ the number of product starts for the qth product of firm r in tth period, and λ_{rt} the multiplier of firm r in tth period. θ_{rs} is the efficiency estimate of firm r in specific period s. The linear programming formulation for a specific firm is:

$$
\begin{aligned}
\text{Max } & \theta_{rs} \\
\text{s.t. } & \sum_{t \in TW} \lambda_{rt} X_{irt} \leq X_{irs}, \quad \forall i \\
& \sum_{t \in TW} \lambda_{rt} Z^{(1)}_{qrt} \geq \theta_{rs} Z^{(1)}_{qrs}, \quad \forall q \\
& \lambda_{rt} \geq 0, \quad \forall t
\end{aligned}
\tag{1}
$$

If efficiency equals 1, peak output equals the number of product starts in period s; otherwise, the peak output is equal to product start multiplied by the efficiency estimate θ_{rs}. Then, the time window is shifted to include the next period, the oldest period in the time window is dropped, and the process is repeated. Note that the reference set is constructed to analyse each individual firm relative to its own performance.

Diewert's *sequential model* constructs the production reference set by adding new observations by period. Figure 3 shows the construction procedure of a 1-input-and-2-output production reference set over time. In the first period, only one observation forms the production reference set. In the second period, when additional observations are included, the production reference set extends outward and retains the properties of monotonicity and convexity. In the third period, the frontier represents a piece-wise linear production reference set. In the fourth period, the new observation falls below the frontier, and is denoted as inefficient. Also in the fourth period, the efficiency can be estimated by

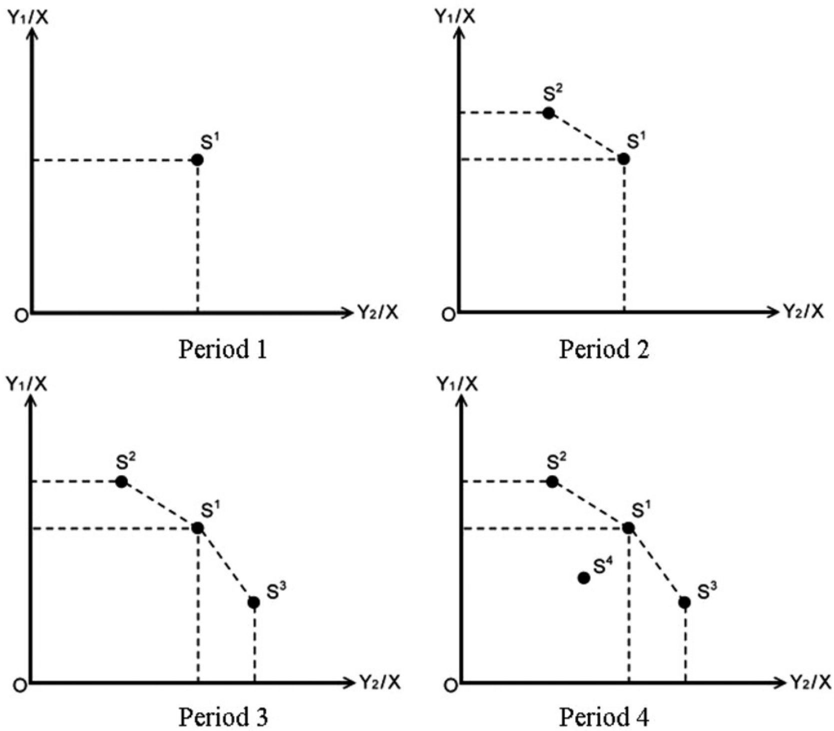

Figure 3. Diewert's sequential-type production reference set.

evaluating the current period's production relative to all of the prior periods' production, using the output-oriented CRS-DEA model to obtain peak output values.

The major differences between the rolling time window analysis and the sequential analysis are the assumptions about the production reference set. The rolling time window analysis postulates that production processes observed within a defined time window are comparable; the sequential analysis assumes that production processes can be compared to any previously observed production process. While both analyses appear in the literature, the argument for the sequential model seems justified on the basis that technology moves forward and improvement methods become available. Previous methods of operation are retained in the production possibility set regardless of the number of prior periods in which they are used. We use the sequential method to estimate peak output in our application (see Section 5).

3.2 *Efficiency measurement by network DEA*

We use a CRS rational network DEA model with series structure described in Kao (2009) for efficiency decomposition. Rational network DEA efficiency estimates will be less-or-equal to conventional network DEA (Färe and Grosskopf 1996) efficiency estimates, because the former imposes the property: if one of the sub-processes is inefficient, then the production system is inefficient. The CRS assumption allows us to compare cost-efficient production processes (Førsund and Hjalmarsson 1987).

Let X_{ikt}, $Z^{(2)}_{pkt}$, $Z^{(1)}_{qkt}$, and Y_{jkt} be the ith input resource, pth peak output, qth product start and jth actual output of kth firm in tth period respectively. v_i, $w^{(1)}_q$, $w^{(2)}_p$ and u_j are the multipliers associated with these critical variables. The system efficiency E^S_{rs} of firm r in period s is estimated with a sequential reference set using the following mathematical programming formulation:

$$
\begin{aligned}
E^S_{rs} = \text{Max } & \sum_{j\in J} u_j Y_{jrs} \\
\text{s.t. } & \sum_{i\in I} v_i X_{irs} = 1 \\
& \sum_{p\in P} w^{(2)}_p Z^{(2)}_{pkt} - \sum_{i\in I} v_i X_{ikt} \le 0, \quad \forall k, \quad \forall t = \{1,\ldots,s\} \\
& \sum_{q\in Q} w^{(1)}_q Z^{(1)}_{qkt} - \sum_{p\in P} w^{(2)}_p Z^{(2)}_{pkt} \le 0, \quad \forall k, \quad \forall t = \{1,\ldots,s\} \\
& \sum_{j\in J} u_j Y_{jkt} - \sum_{q\in Q} w^{(1)}_q Z^{(1)}_{qkt} \le 0, \quad \forall k, \quad \forall t = \{1,\ldots,s\} \\
& v_i, w^{(1)}_q, w^{(2)}_p, u_j, \ge 0, \quad \forall i,p,q,j
\end{aligned}
\tag{2}
$$

By solving this optimisation model, the optimal multipliers v^*_i, $w^{(1)*}_q$, $w^{(2)*}_p$ and u^*_j will be obtained and efficiency can be decomposed. Therefore, the efficiencies of three sub-processes of the system can be estimated by the following equations (E^P_{rs}, E^D_{rs}, and E^O_{rs} denote efficiency of production design, efficiency of demand support and efficiency of operations respectively):

$$E^P_{rs} = \left(\sum_{p\in P} w^{(2)*}_p Z^{(2)}_{prs}\right) \Big/ \left(\sum_{i\in I} v^*_i X_{irs}\right) \tag{3}$$

$$E^D_{rs} = \left(\sum_{q\in Q} w^{(1)*}_q Z^{(1)}_{qrs}\right) \Big/ \left(\sum_{p\in P} w^{(2)*}_p Z^{(2)}_{prs}\right) \tag{4}$$

$$E^O_{rs} = \left(\sum_{j\in J} u^*_j Y_{jrs}\right) \Big/ \left(\sum_{q\in Q} w^{(1)*}_q Z^{(1)}_{qrs}\right) \tag{5}$$

A property of a series-type network DEA model is the product of the components E^P_{rs}, E^D_{rs}, and E^O_{rs} equals the system efficiency,

$$E^S_{rs} = \left(\sum_{j\in J} u^*_j Y_{jrs}\right) \Big/ \left(\sum_{i\in I} v^*_i X_{irs}\right).$$

This implies the system efficiency is 1 if and only if all sub-processes are efficient.

There are other issues regarding the interpretation of the efficiency measures. Initially, a factory starting production is in a ramping-up phase (see Figure 4). During this phase, operators are training, the production system is not saturated with WIP at all stages, and peak output cannot be achieved. When the system finally enters a steady state, its performance with respect to other facilities can be estimated. In Section 5, we assume that all fabs operate under steady-state conditions; otherwise, the ramping effect will be attributed to the efficiency loss in the design, demand, and operations stages, and cause us to under-estimate the true efficiency.

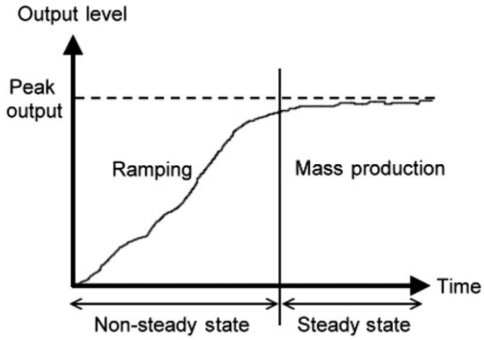

Figure 4. Production system status.

Another issue is the time delay in production. In general, for a make-to-order firm a time delay exists between realising the demand level and producing the required product. The process time to produce an individual product can be significant; in our semiconductor manufacturing case study, this production process time is usually over 50 days. Hence, production started and product completed may fall in distinct periods. A time delay arises and it is necessary to correct for this issue through data pre-processing before efficiency estimation. Namely, the time of specific product output must be shifted one product cycle to match the corresponding demand realisation. The data in our case study has been corrected to address the time delay issue (Leachman *et al.* 2007).

4. Measurement of productivity change

One method to measure the productivity change over time is MPI. Färe *et al.* (1992) define MPI as a geometric mean of two distance functions. MPI can be decomposed into a measure of change in efficiency and a measure of change in technology. This decomposition provides useful sub-indices for any study of efficiency and technical change.

4.1 *Malmquist productivity index*

Let $x^t \in R^I_+$ denote an input factor of input resource of production system at period t, and $y^t \in R^J_+$ denote an output factor of actual output of production system at period t. The input requirement set $L^t(y^t)$ is defined as:

$$L^t(y^t) = \{x^t : (x^t, y^t) \in S^t\}, \quad t \in T \tag{6}$$

where $S^t = \{(x^t, y^t) : x^t \text{ can produce } y^t\}$ is the technology set at period t. An estimate of the input requirement set $\tilde{L}^t(y^t)$ is constructed from the observations as:

$$\begin{aligned} \tilde{L}^t(y^t) = \{x^t : Y_{jt} &\le \sum_{k\in K} \lambda_k Y_{jkt}, \quad \forall j \\ X_{it} &\ge \sum_{k\in K} \lambda_k X_{ikt}, \quad \forall i \\ \lambda_k &\ge 0, \quad \forall k\}, \end{aligned} \tag{7}$$

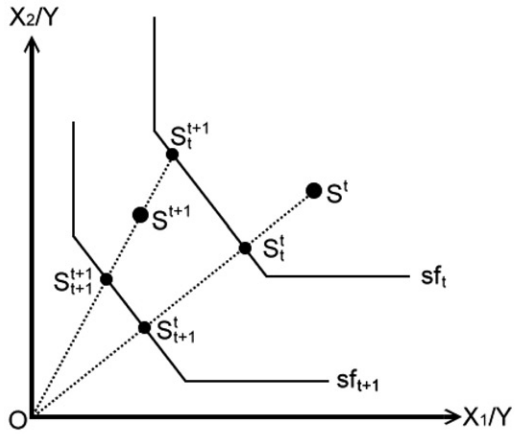

Figure 5. Input-oriented distance function and frontier shift.

where λ_k indicate the intensity variables used in the piecewise linear technology. Note that the assumption of CRS is imposed on this reference technology set as suggested by Färe *et al*. (1994). For alternative assumptions, see for example Ray and Desli (1997). Then, defining $D^t_{Input}(y^t, x^t)$ as Shephard's input-oriented distance function (Shephard 1953), the efficiency of an observation at period t can be measured relative to the reference technology at period t:

$$D^t_{Input}(y^t, x^t) = \sup\{\theta : (x^t/\theta) \in L^t(y^t)\} \tag{8}$$

Shephard's distance function (Shephard 1953) is the inverse of Farrell's measure (Farrell 1957). Figure 5 illustrates an input-oriented efficiency measure and a system frontier *sf* shifting from period t to period $t+1$, namely, from sf_t to sf_{t+1} for a two-input case. $L^t(y^t)$ forms the input requirement set, and sf_t is a piecewise linear estimate of the isoquant in period t. It also illustrates a firm shifting from S^t at period t to S^{t+1} at period $t+1$. The point S_t^{t+1} represents the system observation S^{t+1} in period $t+1$ projected to the system frontier of period t. Similar explanations apply to S_{t+1}^{t+1}, S_t^t and S_{t+1}^t. Then, the overall system efficiency of observation S^t is equal to OS_t^t/OS^t and lies between 0 and 1, the inverse of Shephard's distance function.

Estimating the MPI between period t and period $t+1$ requires us to measure the additional distance function as follows:

$$D^t_{Input}(y^{t+1}, x^{t+1}) = \sup\{\theta : (x^{t+1}/\theta) \in L^t(y^{t+1})\} \tag{9}$$

$$D^{t+1}_{Input}(y^t, x^t) = \sup\{\theta : (x^t/\theta) \in L^{t+1}(y^t)\} \tag{10}$$

$$D^{t+1}_{Input}(y^{t+1}, x^{t+1}) = \sup\{\theta : (x^{t+1}/\theta) \in L^{t+1}(y^{t+1})\} \tag{11}$$

where $D^t_{Input}(y^{t+1}, x^{t+1})$ is the cross-period distance function of an observation in period $t+1$ relative to the reference technology in period t. In Figure 5, Shephard's distance function is equal to OS^{t+1}/OS_t^{t+1} by solving the following programming formulation.

Similarly, $D^{t+1}_{Input}(y^t, x^t)$ and $D^{t+1}_{Input}(y^{t+1}, x^{t+1})$ can be defined:

$$
\begin{aligned}
&[D^t_{Input}(y^{t+1}, x^{t+1})]^{-1} = \text{Min}\ \theta \\
&\text{s.t. } Y_{jr(t+1)} \le \sum_{k\in K} \lambda_k Y_{jkt}, \quad \forall j \\
&\theta X_{ir(t+1)} \ge \sum_{k\in K} \lambda_k X_{ikt}, \quad \forall i \\
&\lambda_k \ge 0, \quad \forall k
\end{aligned} \tag{12}
$$

Färe *et al.* (1992, 1994) propose an input-oriented MPI at period t relative to period $t+1$ as:

$$
MPI^{t+1->t}_{Input}(y^{t+1}, x^{t+1}, y^t, x^t) = \left[\frac{D^t_{Input}(y^{t+1}, x^{t+1})}{D^t_{Input}(y^t, x^t)} \frac{D^{t+1}_{Input}(y^{t+1}, x^{t+1})}{D^{t+1}_{Input}(y^t, x^t)}\right]^{\frac{1}{2}} \tag{13}
$$

and this index can be decomposed into change in efficiency (CIE) and change in technology (CIT) at period $t+1$ relative to period t as:

$$
MPI^{t->t+1}_{Input}(y^{t+1}, x^{t+1}, y^t, x^t) = \frac{D^t_{Input}(y^t, x^t)}{D^{t+1}_{Input}(y^{t+1}, x^{t+1})} \left[\frac{D^{t+1}_{Input}(y^{t+1}, x^{t+1})}{D^t_{Input}(y^{t+1}, x^{t+1})} \frac{D^{t+1}_{Input}(y^t, x^t)}{D^t_{Input}(y^t, x^t)}\right]^{\frac{1}{2}} \tag{14}
$$

where the first term represents the change in efficiency from period t to period $t+1$, and the second term indicates the change in technology. Let $TSE^t = 1/D^t_{Input}(y^t, x^t)$ and $TSE^{t+1} = 1/D^{t+1}_{Input}(y^{t+1}, x^{t+1})$ as technical and scale efficiency (TSE) at period t and $t+1$, and $IEI^{t+1}_t = 1/D^t_{Input}(y^{t+1}, x^{t+1})$ and $IEI^t_{t+1} = 1/D^{t+1}_{Input}(y^t, x^t)$ as intertemporal efficiency index (IEI) at period $t+1$ relative to the reference technology at period t, and at period t relative to the reference technology at period $t+1$. Therefore, based on input-oriented measurement, the change in productivity, change in efficiency, and change in technology are each interpreted as achieving progress, no change, and regress when the values for their estimates are greater than 1, equal to 1, and less than 1.

4.2 *Efficiency decomposition of MPI*

The decomposition of efficiency proposed in Section 3.2 defines system efficiency E^S as equal to the product of production design efficiency E^P, demand support efficiency E^D, and operations efficiency E^O, $E^S = E^P \times E^D \times E^O$. Thus, we show that the MPI of the overall production system ($SMPI$) equals the MPI multiplication of production design ($PMPI$), demand support ($DMPI$), and operations ($OMPI$), namely,

$$
SMPI^{t+1}_t = PMPI^{t+1}_t \times DMPI^{t+1}_t \times OMPI^{t+1}_t. \tag{15}
$$

Below, the necessary notation and definitions are outlined to show (15) must hold. First we demonstrate that equation $SMPI^{t+1}_t = PMPI^{t+1}_t \times DOMPI^{t+1}_t$ holds, and then we show $DOMPI^{t+1}_t = DMPI^{t+1}_t \times OMPI^{t+1}_t$.

The efficiencies of demand support and operations are combined in one composite efficiency E^{DO}; thus, system efficiency is $E^S = E^P \times E^{DO}$. Since all of the distance functions used in the following sections are input-oriented measurements, we drop the subscript input for notational simplicity. Let S^t and P^t be the observations of overall production system

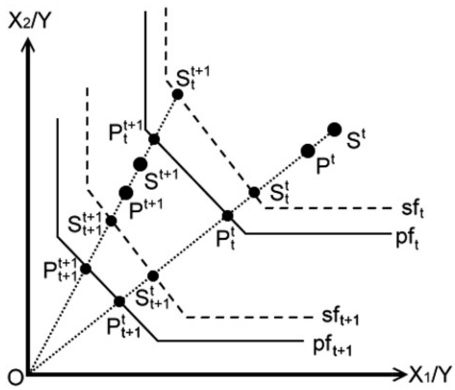

Figure 6. System and production frontier shift.

and production design, and sf_t and pf_t be the efficiency frontier of overall production system and production design at period t respectively. Then, the system and production frontier shift from period t to period $t+1$ (Figure 6). Let the notation $DS^t(y^t, x^t)$ and $DP^t(y^t, x^t)$ indicate Shephard's distance function (Shephard 1953) of overall production system and production design respectively at period t relative to the reference technology at period t.

Then, the construction of the definitions and related proofs are (also see Appendix I):

SMPI (Malmquist Productivity Index of Overall Production System)

$$\begin{aligned} SMPI_t^{t+1} &= SCIE_t^{t+1} * SCIT_t^{t+1} \\ &= \frac{DS^t(y^t, x^t)}{DS^{t+1}(y^{t+1}, x^{t+1})} \left(\frac{DS^{t+1}(y^{t+1}, x^{t+1})}{DS^t(y^{t+1}, x^{t+1})} \frac{DS^{t+1}(y^t, x^t)}{DS^t(y^t, x^t)} \right)^{1/2} \\ &= \frac{OS_{t+1}^{t+1}/OS^{t+1}}{OS_t^t/OS^t} \left(\frac{OS_t^{t+1}/OS^{t+1}}{OS_{t+1}^{t+1}/OS^{t+1}} \frac{OS_t^t/OS^t}{OS_{t+1}^t/OS^t} \right)^{1/2} \end{aligned} \tag{16}$$

PMPI (Malmquist Productivity Index of Production Design)

$$\begin{aligned} PMPI_t^{t+1} &= PCIE_t^{t+1} * PCIT_t^{t+1} \\ &= \frac{DP^t(y^t, x^t)}{DP^{t+1}(y^{t+1}, x^{t+1})} \left(\frac{DP^{t+1}(y^{t+1}, x^{t+1})}{DP^t(y^{t+1}, x^{t+1})} \frac{DP^{t+1}(y^t, x^t)}{DP^t(y^t, x^t)} \right)^{1/2} \\ &= \frac{OP_{t+1}^{t+1}/OP^{t+1}}{OP_t^t/OP^t} \left(\frac{OP_t^{t+1}/OP^{t+1}}{OP_{t+1}^{t+1}/OP^{t+1}} \frac{OP_t^t/OP^t}{OP_{t+1}^t/OP^t} \right)^{1/2} \end{aligned} \tag{17}$$

Demand and Operations Efficiency at period t

$$E^{DO} = \frac{E^S}{E^P} = \frac{DP^t(y^t, x^t)}{DS^t(y^t, x^t)} = \frac{OS_t^t/OS^t}{OP_t^t/OP^t} \tag{18}$$

Definition 4.2.1: Change in Efficiency of Demand and Operations

$$DOCIE_t^{t+1} = \frac{SCIE_t^{t+1}}{PCIE_t^{t+1}} \tag{19}$$

Definition 4.2.2: Change in Technology of Demand and Operations

$$DOCIT_t^{t+1} = \frac{SCIT_t^{t+1}}{PCIT_t^{t+1}} \tag{20}$$

Definition 4.2.3: Malmquist Productivity Index of Demand and Operations

$$DOMPI_t^{t+1} = \frac{SMPI_t^{t+1}}{PMPI_t^{t+1}} \tag{21}$$

Based on the definitions, the equation $SMPI_t^{t+1} = PMPI_t^{t+1} \times DOMPI_t^{t+1}$ is derived directly. Through a similar procedure, it can be shown that $DOMPI_t^{t+1} = DMPI_t^{t+1} \times OMPI_t^{t+1}$, where $DMPI_t^{t+1}$ and $OMPI_t^{t+1}$ are the MPI of demand support and operations respectively. Thus, the efficiency decomposition of the MPI of the overall production system is $SMPI_t^{t+1} = PMPI_t^{t+1} \times DMPI_t^{t+1} \times OMPI_t^{t+1}$.

5. Empirical study

This section will analyse the semiconductor manufacturing industry. The data set is described in Leachman *et al*. (2007). The data includes 87 records collected from 10 leading fabs which produced 200 mm wafers with 350 nm process technology in the United States, Taiwan, Japan, and Europe from 1995 to 2000. Each observation is a particular fab in a given quarter of a specific year. The data definitions of input-output factors for productivity analysis are described in Section 5.1. In Section 5.2, employing the decomposition of efficiency and applying the MPI to quantify productivity change allows a more detailed analysis of the source of inefficiency within each fab. Section 5.3 shows the efficiency difference between logic and memory products using two-stage DEA.

5.1 *Data description*

The production process in semiconductor manufacturing can be characterised by the following input resources:

- *Number of steppers* (SN) is the average number of steppers and scanners employed in the fab during a particular quarter. Steppers and scanners are exposure tools used in the lithography process to define the pattern of integrated circuit and critical dimension by depositing layers and doping region. In practice, the lithography process is typically the bottleneck of the production line, because it is the most expensive machinery in the facility.
- *Headcount* (HC) is the sum of direct and indirect headcount. Direct headcount refers to the operators and workers who operate machinery used in the production process; indirect headcount refers to the engineers, technicians and managers who support the related business activities. The amount of indirect labour is relatively stable regardless of variation in the production volume.
- *Clean-room size* (CR) indicates the size of the floor space in a clean room. A clean room controls particle dispersion and creates an uncontaminated condition for manufacturing. A general rule of thumb is that a fab's infrastructure (capital) costs are proportional to its CR. In other words, CR is a proxy for the total investment in a fab's infrastructure. The data for CR is the sum of depreciated construction cost and occupation cost per square foot during a quarter.

- *Total wafer starts* (WS) is the total number of blank silicon wafers released into the manufacturing process during a particular quarter. Wafer starts mainly depend on production capacity and demand requirements. Too many wafer starts will create high levels of WIP and extend cycle times; insufficient wafer starts cause loss of capacity and lower machine utilisation. In general, WS controls the level of production output based on demand information.
- *Actual die output* (AD) represents the amount of saleable die output actually produced by the fabrication process during a particular quarter.
- *Peak die output* (PD) is the highest output level under a given production design. This unobservable variable can be estimated via a sequential model which assumes that past performance can be used as a reference set for estimating WS.

Other factors which also affect fab operations and productivity, but which are not inputs or outputs to the production process, are often referred to as contextual variables in the productivity literature. Examples are: product mix, employment of automated material handling system (AMHS) and equipment type. In the application below the decomposition of efficiency is extended to consider the effects of product type on the productivity of a fab using a two-stage approach (Ray 1988, 1991). Product types influence fab resource allocation, and require dedicated equipment and specific changeover procedures. Different product types may have different numbers of mask layers which increase product complexity, and lead to a significant difference of efficiencies. The effects of other practices or attributes can also be analysed using this same approach.

5.2 *Productivity change analysis*

Table 1 summarises the data and the results. Note that the efficiency decomposition of one unit is derived from the production possibility set of all previous periods. In the efficiency decomposition, the overall production system efficiency can be divided into production design, demand support and operations. The system efficiency is equal to 1 only if the efficiencies of three sub-components are all equal to 1. A fab can annually investigate the efficiency variation of each component so as to make suitable remedy for productivity improvement. Using the first quarter of 1998 as an example, the system efficiency of fab 9 is 0.38. A further investigation of the components of efficiency shows this is not an issue of poor production design or operational inefficiency; rather the system inefficiency is mainly caused by insufficient demand (demand efficiency is equal to 0.44). Thus, fab 9 should focus on raising demand rather than operational changes. This clarification should lead upper management to work with sales and/or marketing to address productivity concerns.

The average productivity change of our 10 semiconductor fabs regresses in the three time periods: between 1995 and 1996, between 1997 and 1998, and between 1999 and 2000. The reason for regress is mainly the lack of demand. Table 2 shows the 10 fabs' weighted average productivity change. Specifically, between 1995 and 1996, the lower demand resulted from the introduction of 350 mm wafer technology. A small amount of demand was released for testing products, and the computer manufacturers waited for their customers' reaction as well as assessing the performance/quality of the computer chips produced. However, investors continued to provide capital to promote the production of the 350 mm product. Thus, we conclude that the demand regress is mainly due to the marketing start-up. For the time periods between 1997 and 1998 and between 1999 and 2000,

Table 1. Data information and efficiency decomposition.

Firm				Input resource			Peak output	Demand	Actual output	Efficiency decomposition			
Fab	Year	Qtr	Firm	SN	HC	CR	PD	WS	AD	System	Production	Demand	Operations
3m	1996	1	3m-1996-1	22.0	895.0	168,100	16,410	16,410	11,830	0.41	0.44	1.00	0.92
6	1996	1	6-1996-1	11.0	444.0	100,000	22,062	21,738	16,901	0.98	1.00	0.99	0.99
3m	1996	2	3m-1996-2	22.0	882.0	168,100	16,183	9639	6648	0.20	0.38	0.60	0.88
5	1996	2	5-1996-2	24.0	298.0	98,700	7332	7332	5441	0.40	0.42	1.00	0.95
6	1996	2	6-1996-2	11.0	517.0	100,000	25,248	25,248	19,770	1.00	1.00	1.00	1.00
31	1996	3	31-1996-3	22.0	855.0	168,100	2019	2019	1280	0.03	0.05	1.00	0.68
5	1996	3	5-1996-3	24.0	357.0	98,700	12,285	12,285	8809	0.32	0.41	1.00	0.77
6	1996	3	6-1996-3	11.0	566.4	100,000	25,248	21,438	15,760	0.67	1.00	0.85	0.79
7	1996	3	7-1996-3	70.0	745.0	148,789	74,409	74,409	69,075	1.00	1.00	1.00	1.00
9	1996	3	9-1996-3	25.0	478.0	89,459	30,633	30,633	24,335	0.79	0.93	1.00	0.86
10	1996	3	10-1996-3	16.0	589.0	62,431	16,953	16,953	10,843	0.53	0.77	1.00	0.69
31	1996	4	31-1996-4	22.0	835.0	168,100	2028	2028	1180	0.03	0.04	1.00	0.63
5	1996	4	5-1996-4	24.0	377.0	98,700	20,547	20,547	16,776	0.47	0.53	1.00	0.88
6	1996	4	6-1996-4	11.0	765.0	100,000	25,248	15,465	10,742	0.46	1.00	0.61	0.75
7	1996	4	7-1996-4	70.0	746.0	148,789	102,156	102,156	90,973	0.96	1.00	1.00	0.96
9	1996	4	9-1996-4	25.0	475.0	89,459	30,441	24,621	20,640	0.54	0.75	0.81	0.90
10	1996	4	10-1996-4	16.0	610.0	62,431	26,823	26,823	19,576	0.79	1.00	1.00	0.79
31	1997	1	31-1997-1	22.0	1009.0	168,100	11,400	11,400	8019	0.16	0.22	1.00	0.76
4	1997	1	4-1997-1	45.0	1091.0	72,160	43,230	43,230	37,044	0.80	0.87	1.00	0.92
6	1997	1	6-1997-1	11.0	749.0	100,000	25,248	18,267	13,394	0.56	0.99	0.72	0.79
7	1997	1	7-1997-1	70.0	742.0	148,789	101,608	99,120	87,939	0.93	1.00	0.98	0.96
9	1997	1	9-1997-1	25.0	555.0	89,459	30,633	29,370	24,067	0.58	0.69	0.96	0.88
10	1997	1	10-1997-1	16.0	648.0	65,660	35,199	35,199	24,950	0.76	1.00	1.00	0.76
31	1997	2	31-1997-2	26.0	1028.0	168,100	17,862	17,862	14,697	0.22	0.24	1.00	0.89
4	1997	2	4-1997-2	45.0	1105.0	72,160	51,894	51,894	44,778	0.93	1.00	1.00	0.93
6	1997	2	6-1997-2	15.0	569.0	100,000	25,248	24,705	19,274	0.52	0.63	0.98	0.84
7	1997	2	7-1997-2	70.0	730.0	148,789	99,965	95,991	83,977	0.90	1.00	0.96	0.94
9	1997	2	9-1997-2	25.0	711.0	89,459	32,001	32,001	26,976	0.53	0.58	1.00	0.91
10	1997	2	10-1997-2	16.0	729.0	65,660	45,693	45,693	36,989	0.87	1.00	1.00	0.87
2	1997	3	2-1997-3	3.0	209.0	46,100	5718	5718	2544	0.29	0.60	1.00	0.48
31	1997	3	31-1997-3	26.0	1026.0	168,100	18,123	18,123	15,134	0.19	0.21	1.00	0.90

6	1997	3	6-1997-3	15.0	581.0	100,000	25,248	24,564	20,374	0.46	0.53	0.97	0.89
7	1997	3	7-1997-3	70.0	712.0	148,789	97,500	92,193	79,019	0.87	0.99	0.95	0.92
9	1997	3	9-1997-3	25.7	750.0	89,459	78,258	78,258	63,251	0.87	1.00	1.00	0.87
10	1997	3	10-1997-3	16.0	849.0	65,660	45,693	40,131	32,341	0.71	0.93	0.88	0.87
2	1997	4	2-1997-4	3.3	224.0	46,100	9840	9840	6340	0.55	0.80	1.00	0.69
31	1997	4	31-1997-4	26.0	1133.0	168,100	18,123	17,811	15,516	0.16	0.18	0.98	0.94
6	1997	4	6-1997-4	15.0	608.0	100,000	25,248	22,833	19,525	0.37	0.44	0.90	0.92
7	1997	4	7-1997-4	70.0	688.0	148,789	94,214	86,253	73,765	0.83	0.99	0.92	0.92
9	1997	4	9-1997-4	27.0	783.0	89,459	97,575	97,575	78,319	0.86	1.00	1.00	0.86
10	1997	4	10-1997-4	16.0	973.0	65,660	45,693	42,732	31,950	0.59	0.78	0.94	0.81
2	1998	1	2-1998-1	4.8	216.0	46,100	9489	8886	5257	0.30	0.51	0.94	0.64
6	1998	1	6-1998-1	17.0	651.0	100,000	29,742	29,742	23,222	0.38	0.45	1.00	0.84
7	1998	1	7-1998-1	73.0	658.0	148,789	90,105	87,918	79,924	0.94	0.98	0.98	0.98
8	1998	1	8-1998-1	28.0	1022.0	80,735	4911	4911	3931	0.05	0.06	1.00	0.86
9	1998	1	9-1998-1	27.0	798.0	89,459	97,575	42,786	34,454	0.38	1.00	0.44	0.87
10	1998	1	10-1998-1	16.0	1072.0	65,660	63,423	63,423	45,455	0.77	1.00	1.00	0.77
2	1998	2	2-1998-2	6.0	258.0	46,100	12,750	12,750	6737	0.32	0.55	1.00	0.57
6	1998	2	6-1998-2	17.0	899.0	100,000	29,742	13,212	10,983	0.18	0.44	0.44	0.90
7	1998	2	7-1998-2	75.0	658.0	148,789	90,579	90,579	83,841	0.98	0.99	1.00	1.00
8	1998	2	8-1998-2	28.0	1165.0	80,735	6294	6294	4980	0.06	0.07	1.00	0.85
1	1998	3	1-1998-3	35.0	1230.0	100,000	24,168	24,168	17,201	0.17	0.22	1.00	0.76
2	1998	3	2-1998-3	6.8	400.0	46,100	16,359	16,359	10,526	0.42	0.61	1.00	0.69
6	1998	3	6-1998-3	17.0	888.0	100,000	29,742	6042	5090	0.08	0.44	0.20	0.90
7	1998	3	7-1998-3	75.0	560.0	148,789	88,710	88,710	82,605	1.00	1.00	1.00	1.00
8	1998	3	8-1998-3	28.0	1160.0	80,735	6762	6762	5061	0.06	0.08	1.00	0.80
1	1998	4	1-1998-4	35.0	1244.0	100,000	35,601	35,601	25,870	0.25	0.33	1.00	0.78
2	1998	4	2-1998-4	7.5	287.0	46,100	13,487	9276	6100	0.23	0.48	0.69	0.71
6	1998	4	6-1998-4	17.0	752.0	100,000	29,742	14,550	12,659	0.20	0.45	0.49	0.93
8	1998	4	8-1998-4	28.0	1161.0	80,735	14,103	14,103	10,145	0.12	0.16	1.00	0.77
1	1999	1	1-1999-1	35.0	1289.0	100,000	41,457	41,457	31,538	0.31	0.38	1.00	0.82
2	1999	1	2-1999-1	8.0	375.0	46,100	17,031	17,031	11,748	0.41	0.55	1.00	0.74
8	1999	1	8-1999-1	28.0	1177.0	80,735	29,445	29,445	19,826	0.24	0.33	1.00	0.72
1	1999	2	1-1999-2	35.0	1337.0	100,000	41,988	41,988	32,881	0.32	0.38	1.00	0.84
2	1999	2	2-1999-2	8.0	373.0	46,100	17,043	17,043	11,799	0.41	0.55	1.00	0.74
8	1999	2	8-1999-2	28.0	1304.0	80,735	34,620	34,620	22,520	0.27	0.39	1.00	0.70
1	1999	3	1-1999-3	35.0	1322.0	100,000	41,822	40,071	31,550	0.31	0.38	0.96	0.85

(continued)

Table 1. Continued.

Firm				Input resource			Peak output	Demand	Actual output	Efficiency decomposition			
Fab	Year	Qtr	Firm	SN	HC	CR	PD	WS	AD	System	Production	Demand	Operations
2	1999	3	2-1999-3	8.0	391.0	46,100	20,517	20,517	13,342	0.46	0.66	1.00	0.70
8	1999	3	8-1999-3	28.0	1360.0	80,735	42,024	42,024	29,182	0.36	0.48	1.00	0.75
1	1999	4	1-1999-4	35.0	1292.0	100,000	41,490	27,312	22,550	0.22	0.38	0.66	0.89
2	1999	4	2-1999-4	9.3	437.0	46,100	20,517	15,177	9194	0.28	0.58	0.74	0.65
8	1999	4	8-1999-4	29.2	1471.0	80,735	42,024	38,058	26,354	0.32	0.48	0.91	0.74
1	2000	1	1-2000-1	35.0	1276.0	100,000	41,039	19,851	15,897	0.16	0.38	0.48	0.86
2	2000	1	2-2000-1	10.0	465.0	46,100	20,517	17,049	10,397	0.29	0.54	0.83	0.65
8	2000	1	8-2000-1	31.0	1545.0	80,735	42,024	32,127	21,768	0.27	0.48	0.76	0.73
1	2000	2	1-2000-2	35.0	1280.0	100,000	41,168	15,756	13,035	0.13	0.38	0.38	0.89
2	2000	2	2-2000-2	10.0	514.0	46,100	20,517	4734	2947	0.08	0.53	0.23	0.67
8	2000	2	8-2000-2	31.0	1722.0	80,735	42,024	30,255	18,422	0.22	0.48	0.72	0.65
1	2000	3	1-2000-3	35.0	1272.0	100,000	40,910	10,290	8,497	0.08	0.37	0.25	0.89
8	2000	3	8-2000-3	31.0	1881.0	80,735	44,415	44,415	26,036	0.32	0.50	1.00	0.63

Table 2. Productivity change of semiconductor manufacturing fabs.

	MPI				CIE				CIT			
Period	System	Production	Demand	Operations	System	Production	Demand	Operations	System	Production	Demand	Operations
4Q95 –> 1Q96	1.02	0.98	0.99	1.05	1.02	0.98	0.99	1.06	1.00	1.00	1.00	1.00
1Q96 –> 2Q96	0.99	1.06	0.90	0.99	0.89	0.95	0.90	0.98	1.11	1.12	1.00	1.00
2Q96 –> 3Q96	1.03	1.16	0.90	0.95	0.72	0.99	0.90	0.80	1.41	1.18	1.00	1.19
3Q96 –> 4Q96	1.33	1.32	0.95	1.01	1.02	1.04	0.95	1.03	1.30	1.27	1.00	1.00
4Q96 –> 1Q97	1.35	1.30	1.03	1.01	1.26	1.21	1.03	1.04	1.06	1.07	1.00	1.00
1Q97 –> 2Q97	1.16	1.08	1.03	1.04	1.04	0.98	1.03	1.07	1.10	1.11	1.00	1.00
2Q97 –> 3Q97	1.35	1.40	0.97	0.99	1.13	1.17	0.97	1.00	1.15	1.15	1.00	1.00
3Q97 –> 4Q97	1.09	1.09	0.99	1.00	0.96	0.96	0.99	1.07	1.14	1.14	1.00	1.00
4Q97 –> 1Q98	1.05	1.08	0.95	1.00	1.02	1.05	0.95	0.97	1.02	1.02	1.00	1.00
1Q98 –> 2Q98	1.00	1.02	0.97	1.01	1.00	1.02	0.97	0.99	1.00	1.00	1.00	1.00
2Q98 –> 3Q98	1.10	1.08	0.97	1.03	1.02	1.02	0.97	1.04	1.08	1.06	1.00	1.00
3Q98 –> 4Q98	1.73	1.32	1.24	1.01	1.73	1.32	1.24	1.01	1.00	1.00	1.00	1.00
4Q98 –> 1Q99	1.55	1.47	1.09	1.01	1.55	1.47	1.09	1.01	1.00	1.00	1.00	1.00
1Q99 –> 2Q99	1.07	1.07	1.00	1.00	1.07	1.07	1.00	1.00	1.00	1.00	1.00	1.00
2Q99 –> 3Q99	1.12	1.12	0.98	1.02	1.12	1.12	0.98	1.00	1.00	1.00	1.00	1.00
3Q99 –> 4Q99	0.78	0.97	0.80	1.01	0.78	0.97	0.80	0.99	1.00	1.00	1.00	1.00
4Q99 –> 1Q00	0.84	0.98	0.88	0.98	0.84	0.98	0.88	0.99	1.00	1.00	1.00	1.00
1Q00 –> 2Q00	0.79	1.00	0.83	0.96	0.79	1.00	0.83	0.98	1.00	1.00	1.00	1.00
2Q00 –> 3Q00	1.23	1.03	1.25	0.97	1.23	1.03	1.25	0.98	1.00	1.00	1.00	1.00

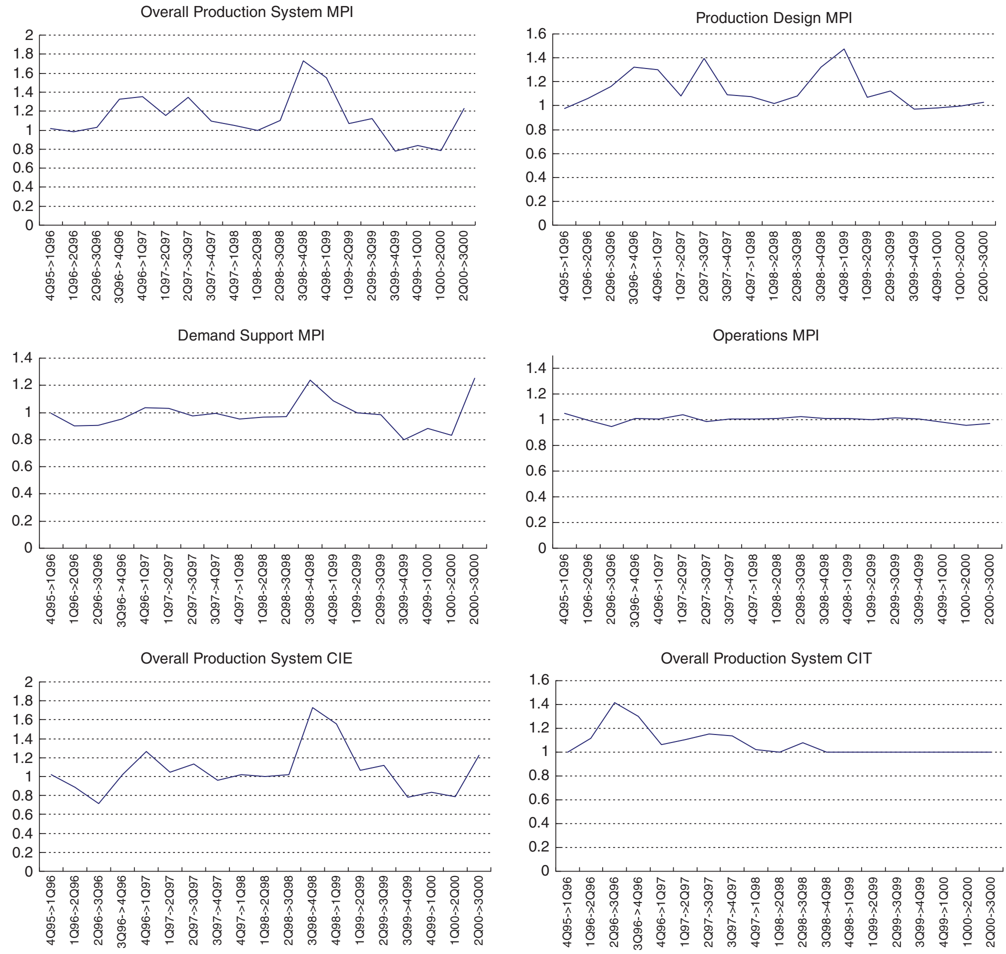

Figure 7. Trends of average productivity change in semiconductor manufacturing fabs.

the technology process matured and capacity growth was stable. Thus, we conclude the demand regress is mainly due to demand fluctuation.

For the first and second quarters of 1996, the MPI of the overall system is 0.99. The production design component is 1.06, the demand support component is 0.90 and the operational component is 0.99. Considering each component individually, a 6% improvement on average of design efficiency over the time horizon is observed, indicating the fabs have been proactive in improving their design process. Demand efficiency equal to 0.90 indicates that some fabs are failing to generate demand sufficient to keep the production facility operating efficiently. Some excess capacity is expected due to random demand fluctuations; however, there still seems to be significant room for improvement. Finally, operations efficiency equal to 0.99 indicates that most fabs are operating efficiently. Therefore, they should promote their devices and stimulate demand as the most effective way to improve productivity.[2]

The trend charts in Figure 7 show several interesting issues (also see supporting charts in Appendix II). The production design trend of MPI is usually larger than 1, indicating

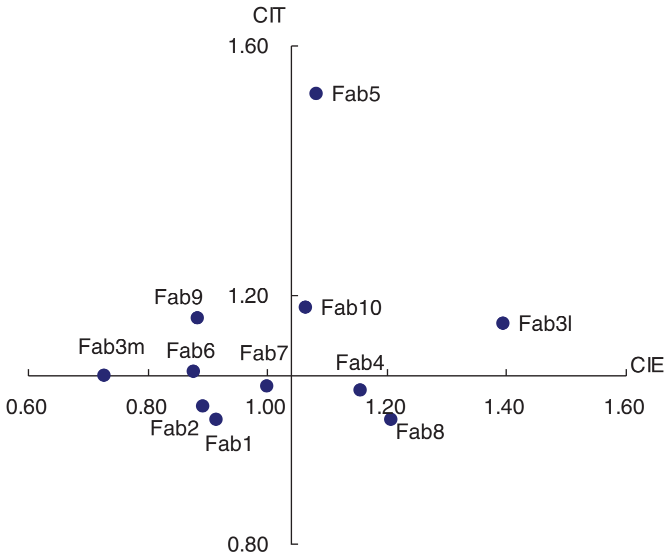

Figure 8. Fab distribution of change in efficiency and change in technology.

that the fabs are consistently improving by upgrading existing processes and equipment. The demand trend of the MPI fluctuates above and below 1, which indicates demand in the semiconductor manufacturing industry is tied to prosperity cycles. In fact, demand deteriorated from 1997 to 1998 and from 1999 to 2000; note that the second quarter is usually the weakest demand quarter. Operations productivity index is usually close to 1 with small variation, which indicates that fabs consistently perform operational processes well. In addition, the trend of the production design component is similar to that of the overall production system, and the variation of the production design component is larger than the other two components. Both indicate that the production design process is a significant sub-process and will lead to a long-term effect on productivity. The demand sub-process has a minor effect. The MPI of the overall system follows a similar pattern of CIE, but CIT tends to converge to 1, because CIE has a larger variation than CIT. For instance, a deteriorated CIE from 1999 to 2000 is mainly caused by insufficient demand; however, CIT would not represent a regress of the production frontier, since a sequential model is employed for efficiency estimation; CIT is never less than 1.

Productivity change analysis also provides benchmarking data for the 10 fabs. Figure 8 maps their CIE and CIT on a two-dimensional co-ordinate. Thus, the four quadrants highlight the strategy of productivity improvement. Using fab 6 as an example, its CIT, 1.08, is above the average, but it has a poor CIE of 0.88. Further analysis of CIE via efficiency decomposition indicates the production design is 0.94, the demand support is 0.94, and operations 0.99. Therefore, fab 6 should strive to improve its design and increase demand to catch up with other fabs. Note that CIE and CIT are not mutually independent, but have different implications for improvement strategies. CIE characterises the fab's change in efficiency and productivity, which is largely driven by process improvement, while CIT measures the frontier change of the technology with respect to a specific resource mix. Fabs can control resources for CIE improvement, but CIT can be a result of a firm's behaviour or of other firms' behaviour.

5.3 *Contextual variables*

Product type significantly affects the complexity of process in semiconductor manufacturing. As mentioned, there are two product types: memory and logic products in this industry. Nine of the fabs in our data set have a dominant product category: logic products or memory products.[3] Three hypothesis tests (Banker 1993) are commonly used to assess fab efficiency. There are two F tests (assuming inefficiency follows exponential distribution and half-normal distribution respectively) and one Kolmogorov-Smirnov test (non-parametric assumption). All three tests result in *p*-values of less than 0.01, which indicates that the distribution of inefficiency differs significantly between memory and logic products.

While Banker's tests indicate there is a difference in terms of efficiency between the different product groups, the tests do not indicate the size of the effect on efficiency. To investigate this question, we use the two-stage DEA method. Considerable controversy surrounds the exact implementation of this method; see, for example, Hoff (2007), Simar and Wilson (2007), Banker and Natarajan (2008) and McDonald (2009). Simar and Wilson (2007) argue the conventional two-stage approaches to estimate efficiency in the presence of contextual variables are invalid since none of these studies describe the underlying data-generating process (DGP). Although these studies use a variety of methods in the second stage including truncated or tobit regression to avoiding boundary problem, or ordinary least squares (OLS) ignoring the boundary problem, Simar and Wilson (2007) state a reasonable data generation process can be defined only for truncated regression. Furthermore, the authors state that in all two-stage studies the DEA efficiency estimates are serially correlated. This results in correlation among the error terms in the second stage regression and a convergence rate too slow for statistical inference on the slope estimates. However, the authors provide no proof of this claim. To address these issues, Simar and Wilson (2007) define a DGP and propose a double bootstrap procedure using truncated regression to produce bias-corrected estimates of efficiency.

Both Banker and Natarajan (2008) and McDonald (2009) argue Simar and Wilson's approach is unnecessarily complicated and based on a restrictive DGP. Banker and Natarajan prove the consistency of the two-stage method applying standard DEA followed by OLS or maximum likelihood estimation. Johnson and Kuosmanen (2009), show the same result on a considerably more general set of assumptions defining the data generation process. McDonald (2009) gives a comparison within-sample prediction performance of OLS, two-limit tobit regression, Papke-Wooldridge (PW) approach based on quasi-maximum likelihood estimation (Papke and Wooldridge 1996) and zero-inflated beta model (Cook *et al*. 2008). The comparison result shows OLS performs at least as well as the other approaches and can replace tobit as a sufficient second stage DEA. McDonald further proves that efficiency estimates are treated as descriptive measures in the second stage and generated by fractional data rather than the censoring process, thus reaching the same conclusion as Simar and Wilson that tobit regression is an inappropriate estimation procedure. The PW approach produces a result similar to OLS and is asymptotically more efficient, but requires significant programming. In contrast, OLS is an unbiased, consistent estimator, and allows hypothesis testing using White's heteroskedastic-consistent standard errors (White 1980), and is robust to heteroskedasticity in the error terms.

Based on the above reasons, OLS provides an unbiased, consistent estimator and directly illustrates the effect of the exogenous variables on efficiency estimates. In order to clarify the effect of different contexts of product on efficiency in semiconductor

Table 3. Regression model with contextual variable.

Regression	System	Production	Demand	Operations
Intercept	0.33	0.47	0.85	0.83
Slope	0.29	0.31	0.08	−0.02

manufacturing, specifically the effect of the mix of memory product and logic product on the production system, two-stage DEA as described by Banker and Natarajan (2008) is applied.

The second stage contextual variable denotes percentage of memory product.[4] If the variable is equal to 1, all products are memory products; if it is equal to 0, all products are logic products. The parameters estimated by least squares regression are shown in Table 3. The efficiency of overall production system is significantly affected by the production design component, and the memory fabs are 29% more efficient than the logic fabs, since memory products usually have fewer mask layers in the lithography process, and the layers themselves are generally a strong indicator of the complexity of the production process. Nevertheless, the statistical test for the significance of the coefficients related to the continuous variables indicating memory or logic for the analysis of demand support efficiency and operational efficiency shows that this difference is insignificant. This is perhaps because the production design is the overriding factor in determining system efficiency.

6. Conclusion

This study proposes a decomposition of efficiency to provide further insights about the sources of productivity change. An overall production system consists of three phases: production design, demand support and operations. Efficiency can be decomposed via a series-type network DEA model and MPI. The proposed model distinguishes the effect of sub-components, allowing the allocation of inefficiency to different components of the production system. In particular, the demand effect allows what was previously indicated to be technical regress to be attributed to demand deterioration rather than degraded production technology. An empirical study of productivity change from 1995 to 2000 in the semiconductor manufacturing industry is used to illustrate and validate the proposed method. In practice, the task of production design belongs to the fab design division or industrial engineering division, while the sales or marketing divisions are responsible for demand generation. Operations are the responsibility of manufacturing, process integration, or the equipment divisions. The clarification of the sources of inefficiency will allow upper-level management to allocate resources and efforts more effectively when trying to improve system performance.

Notes

1. In the case that demand exceeds production capacity, demand could be offloaded to other facilities, filled from inventories, produced using overtime or perhaps deliveries to a subset of

customers could be renegotiated to postpone the due date. This is necessary because pushing more raw materials into the system will only increase the product cycle time. In an ideal production system, the facility is designed to minimise the sum of the expected profit losses from increase costs related to excess demand and costs characterising utilisation loss due to lack of demand.

2. If non-network DEA is applied, the production system is modelled as a black box and the relationship among sub-components is not considered. Then, the efficiency estimates will be larger than those estimated by rational network DEA. Additional results using non-network DEA are available from the authors upon request.
3. The 10th fab in the data set, fab 3, switched from memory to logic products in second quarter 1996. Thus, we include observations of fab 3 prior to the second quarter in the memory group, and include later observations in the logic group.
4. Due to the data gathering process, outliers or errors in measurement are believed to be an insignificant concern. However, methods such as that described in Johnson and McGinnis (2008) can be used if these issues are a concern.

References

Banker, R.D., 1993. Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science*, 39 (10), 1265–1273.

Banker, R.D. and Natarajan, R., 2008. Evaluating contextual variables affecting productivity using data envelopment analysis. *Operations Research*, 56 (1), 48–58.

Caves, D.W., Christensen, L.R., and Diewert, W.E., 1982. The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 50 (6), 1393–1414.

Chang, D.S., Kuo, Y.C., and Chen, T.Y., 2008. Productivity measurement of the manufacturing process for outsourcing decisions: the case of a Taiwanese printed circuit board manufacturer. *International Journal of Production Research*, 46 (24), 6981–6995.

Chang, S.Y. and Chen, T.H., 2008. Performance ranking of Asian lead frame firms: a slack–based method in data envelopment analysis. *International Journal of Production Research*, 46 (14), 3875–3885.

Charnes, A., Cooper, W.W., and Rhodes, E., 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2 (6), 429–444.

Charnes, A., *et al*., 1985. A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the U.S. air forces. *Annals of Operations Research*, 2 (1), 95–112.

Chen, W.C. and McGinnis, L.F., 2007. Reconciling ratio analysis and DEA as performance assessment tools. *European Journal of Operational Research*, 178 (1), 277–291.

Cook, D.D., Kieschnick, R., and McCullough, B.D., 2008. Regression analyses of proportions in finance with self selection. *Journal of Empirical Finance*, 15 (5), 860–867.

Diewert, W.E., 1992. The measurement of productivity. *Bulletin of Economic Research*, 44 (3), 163–198.

Färe, R.S. and Grosskopf, S., 1996. *Intertemporal production frontiers: with dynamic DEA*. Boston: Kluwer Academic Publishers.

Färe, R.S., *et al*., 1992. Productivity changes in Swedish pharmacies 1980–1989: A non-parametric Malmquist approach. *Journal of Productivity Analysis*, 3 (1), 85–101.

Färe, R.S., *et al*., 1994. Productivity growth technical progress, and efficiency change in industrialised countries. *American Economic Review*, 84 (1), 66–83.

Farrell, M.J., 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A*, 120 (3), 253–290.

Fielding, G.J., Babitsky, T.T., and Brenner, M.E., 1985. Performance evaluation for bus transit. *Transportation Research Part A*, 19 (1), 73–82.

Førsund, F.R. and Hjalmarsson, L., 1987. *Analyses of industrial structure: a putty-clay approach*. Industrial Institute for Economic and Social Research, Stockholm: Almqvist and Wicksell International.

Hoff, A., 2007. Second stage DEA: Comparison of approaches for modelling the DEA score. *European Journal of Operational Research*, 181 (1), 425–435.

Hopp, W. and Spearman, M., 2001. *Factory physics*. 2nd ed. Boston, MA: Irwin McGraw-Hill.

Johnson, A.L. and McGinnis, L.F., 2008. Outlier detection in two-stage semiparametric DEA models. *European Journal of Operational Research*, 187 (2), 629–635.

Johnson, A.L. and Kuosmanen, T., 2009. How do operational conditions and practices effect productive performance? Efficient nonparametric one-Stage estimators. Working Paper. Available from: http://ssrn.com/abstract = 1485733 [3 November 2009].

Kao, C., 2009. Efficiency decomposition in network data envelopment analysis: A relational model. *European Journal of Operational Research*, 192 (3), 949–962.

Lan, L.W. and Lin, E.T.J., 2005. Measuring railway performance with adjustment of environmental effects, data noise and slacks. *Transportmetrica*, 1 (2), 161–189.

Leachman, R.C., Ding, S., and Chien, C.F., 2007. Economic efficiency analysis of wafer fabrication. *IEEE Transactions on Automation Science and Engineering*, 4 (4), 501–512.

Lovell, C.A.K., 2003. The decomposition of Malmquist productivity indexes. *Journal of Productivity Analysis*, 20 (3), 437–458.

Lu, W.M. and Hung, S.W., 2010. Assessing the performance of a vertically disintegrated chain by the DEA approach – a case study of Taiwanese semiconductor firms. *International Journal of Production Research*, 48 (4), 1155–1170.

McDonald, J., 2009. Using least squares and tobit in second stage DEA efficiency analyses. *European Journal of Operational Research*, 197 (2), 792–798.

Nishimizu, M. and Page, J.M., 1982. Total factor productivity growth, technological progress and technical efficiency: Dimensions of productivity change in Yugoslavia, 1965–78. *The Economic Journal*, 92 (368), 920–936.

Papke, L.E. and Wooldridge, J.M., 1996. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11 (6), 619–632.

Ray, S.C. and Desli, E., 1997. Productivity growth, technical progress, and efficiency change in industrialised countries: Comment. *The American Economic Review*, 87 (5), 1033–1039.

Ray, S.C., 1988. Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation. *Socio-Economic Planning Sciences*, 22 (4), 167–176.

Ray, S.C., 1991. Resource-use efficiency in public schools: A study of Connecticut data. *Management Science*, 37 (12), 1620–1628.

Simar, L. and Wilson, P.W., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136 (1), 31–64.

Shephard, R.W., 1953. *Cost and production functions*. Princeton: Princeton University Press.

Sueyoshi, T. and Aoki, S., 2001. A use of a nonparametric statistic for DEA frontier shift: The Kruskal and Wallis Rank Test. *OMEGA: International Journal of Management Science*, 29 (1), 1–18.

Sueyoshi, T. and Goto, M.S., 2001. Slack-adjusted DEA for time series analysis: performance measurement of Japanese electric power generation industry in 1984–1993. *European Journal of Operational Research*, 133 (2), 232–259.

Tulkens, H. and Vanden Eeckaut, P., 1995. Non-parametric efficiency, progress and regress measures for panel data: Methodological aspects. *European Journal of Operational Research*, 80 (3), 474–499.

White, H., 1980. A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48 (4), 817–838.

Yu, M.M. and Lin, E.T.J., 2008. Efficiency and effectiveness in railway performance using a multi-activity network DEA model. *OMEGA: The International Journal of Management Science*, 36 (6), 1005–1017.

Appendix I. Definition construction of MPI

Definition 4.2.1: Change in Efficiency of Demand and Operations

$$DOCIE_t^{t+1} = \frac{SCIE_t^{t+1}}{PCIE_t^{t+1}}$$

Proof:

$$DOCIE_t^{t+1} = \frac{\frac{DP^{t+1}(y^{t+1}, x^{t+1})}{DS^{t+1}(y^{t+1}, x^{t+1})}}{\frac{DP^t(y^t, x^t)}{DS^t(y^t, x^t)}} = \frac{\frac{OS_{t+1}^{t+1}/OS^{t+1}}{OP_{t+1}^{t+1}/OP^{t+1}}}{\frac{OS_t^t/OS^t}{OP_t^t/OP^t}}$$

$$= \frac{OS_{t+1}^{t+1} \times OS^t \times OP^{t+1} \times OP_t^t}{OS^{t+1} \times OS_t^t \times OP_{t+1}^{t+1} \times OP^t}$$

$$\frac{SCIE_t^{t+1}}{PCIE_t^{t+1}} = \frac{\frac{DS^t(y^t, x^t)}{DS^{t+1}(y^{t+1}, x^{t+1})}}{\frac{DP^t(y^t, x^t)}{DP^{t+1}(y^{t+1}, x^{t+1})}} = \frac{\frac{OS_{t+1}^{t+1}/OS^{t+1}}{OS_t^t/OS^t}}{\frac{OP_{t+1}^{t+1}/OP^{t+1}}{OP_t^t/OP^t}}$$

$$= \frac{OS_{t+1}^{t+1} \times OS^t \times OP^{t+1} \times OP_t^t}{OS^{t+1} \times OS_t^t \times OP_{t+1}^{t+1} \times OP^t}$$

$$\therefore DOCIE_t^{t+1} = \frac{SCIE_t^{t+1}}{PCIE_t^{t+1}}$$

Definition 4.2.2: Change in Technology of Demand and Operations

$$DOCIT_t^{t+1} = \frac{SCIT_t^{t+1}}{PCIT_t^{t+1}}$$

Proof:

$$DOCIT_t^{t+1} = \left(\frac{\frac{DP^t(y^{t+1}, x^{t+1})}{DS^t(y^{t+1}, x^{t+1})}}{\frac{DP^{t+1}(y^{t+1}, x^{t+1})}{DS^{t+1}(y^{t+1}, x^{t+1})}} \times \frac{\frac{DP^t(y^t, x^t)}{DS^t(y^t, x^t)}}{\frac{DP^{t+1}(y^t, x^t)}{DS^{t+1}(y^t, x^t)}} \right)^{1/2}$$

$$= \left(\frac{\frac{OS_t^{t+1}/OS^{t+1}}{OP_t^{t+1}/OP^{t+1}}}{\frac{OS_{t+1}^{t+1}/OS^{t+1}}{OP_{t+1}^{t+1}/OP^{t+1}}} \times \frac{\frac{OS_t^t/OS^t}{OP_t^t/OP^t}}{\frac{OS_{t+1}^t/OS^t}{OP_{t+1}^t/OP^t}} \right)^{1/2}$$

$$= \left(\frac{\frac{OS_t^{t+1}}{OP_t^{t+1}}}{\frac{OS_{t+1}^{t+1}}{OP_{t+1}^{t+1}}} \times \frac{\frac{OS_t^t}{OP_t^t}}{\frac{OS_{t+1}^t}{OP_{t+1}^t}} \right)^{1/2}$$

$$\frac{SCIT_t^{t+1}}{PCIT_t^{t+1}} = \frac{\left(\frac{DS^{t+1}(y^{t+1}, x^{t+1})}{DS^t(y^{t+1}, x^{t+1})} \frac{DS^{t+1}(y^t, x^t)}{DS^t(y^t, x^t)} \right)^{1/2}}{\left(\frac{DP^{t+1}(y^{t+1}, x^{t+1})}{DP^t(y^{t+1}, x^{t+1})} \frac{DP^{t+1}(y^t, x^t)}{DP^t(y^t, x^t)} \right)^{1/2}}$$

$$= \frac{\left(\frac{OS_t^{t+1}/OS^{t+1}}{OS_{t+1}^{t+1}/OS^{t+1}} \frac{OS_t^t/OS^t}{OS_{t+1}^t/OS^t} \right)^{1/2}}{\left(\frac{OP_t^{t+1}/OP^{t+1}}{OP_{t+1}^{t+1}/OP^{t+1}} \frac{OP_t^t/OP^t}{OP_{t+1}^t/OP^t} \right)^{1/2}}$$

$$= \left(\frac{\frac{OS_t^{t+1}}{OS_{t+1}^{t+1}}}{\frac{OP_t^{t+1}}{OP_{t+1}^{t+1}}} \times \frac{\frac{OS_t^t}{OS_{t+1}^t}}{\frac{OP_t^t}{OP_{t+1}^t}} \right)^{1/2} = \left(\frac{\frac{OS_t^{t+1}}{OP_t^{t+1}}}{\frac{OS_{t+1}^{t+1}}{OP_{t+1}^{t+1}}} \times \frac{\frac{OS_t^t}{OP_t^t}}{\frac{OS_{t+1}^t}{OP_{t+1}^t}} \right)^{1/2}$$

$$\therefore DOCIT_t^{t+1} = \frac{SCIT_t^{t+1}}{PCIT_t^{t+1}}$$

Definition 4.2.3: Malmquist Productivity Index of Demand and Operations

$$DOMPI_t^{t+1} = \frac{SMPI_t^{t+1}}{PMPI_t^{t+1}}$$

Proof:

$$\because DOCIE_t^{t+1} = \frac{SCIE_t^{t+1}}{PCIE_t^{t+1}} \text{ and } DOCIT_t^{t+1} = \frac{SCIT_t^{t+1}}{PCIT_t^{t+1}}$$

$$\therefore DOMPI_t^{t+1} = DOCIE_t^{t+1} * DOCIT_t^{t+1} = \frac{SCIE_t^{t+1}}{PCIE_t^{t+1}} \frac{SCIT_t^{t+1}}{PCIT_t^{t+1}} = \frac{SMPI_t^{t+1}}{PMPI_t^{t+1}}$$

Appendix II. Trends of average productivity change in semiconductor manufacturing fabs