May 2021

Essays on Capital Expenditure, Business Model, and Operating Performance in the Global Semiconductor Industry

Guangshun Qiao
Clemson University, gqiao@g.clemson.edu

Essays on Capital Expenditure, Business Model, and Operating Performance in the Global Semiconductor Industry

---

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Economics

---

by
Guangshun Qiao
May 2021

Accepted by:
Dr. Paul W. Wilson, Committee Chair
Dr. Frederick Andrew Hanssen
Dr. Matthew S. Lewis
Dr. Yichen Zhou

Abstract

This dissertation is comprised of three essays on operating performance in the global semiconductor industry, focusing on the impact of capital expenditure and business model.

In the first chapter, I apply a panel data stochastic frontier approach to investigate the impacts of different business models in operating efficiencies in the global semiconductor industry. The efficiency scores are linked with the financial ratios and specified by cumulative probit distribution function, cumulative logit distribution function, and the Gumbel function respectively after disentangling the heterogeneity by the within transformation. The estimates by the nonlinear least squares technique indicate that the asset-light fabless companies have relatively higher efficiency scores among the different operating models in the global semiconductor industry.

In the second chapter, I associate the semiparametric modified ordinary least squares approach by Simar et al. (2017) with the nonparametric shape constraint regression approach by Du et al. (2013) for performance evaluation in the semiconductor industry. Using panel data on 470 companies in the global semiconductor industry over 1999–2018, I compare technical efficiencies between the integrated device manufacturer business model and the fabless-foundry business model. The performance differences between the vertically integrated manufactures and the vertically specialized fabless and foundry firms are disentangled by the intensity of labor and capital in a very flexible format. The estimation results indicate that the capital intensive integrated device manufacturers taking the advantage of the economies of scale are operating more efficiently than the niche fabless companies in the global semiconductor industry.

In the third chapter, I use a nonparametric production frontier approach to investigate the operating efficiency differences by the impacts of business model and capital expenditure in the global semiconductor industry. By using panel data on 470 companies in the global semiconductor industry over the 20 years 1999–2018, I compare the operating efficiencies between the integrated device manufacturers and the fabless and foundry firms. Handling the impact of capital expenditure as fixed input by directional distance estimator, I find that the fabless firms are operating less efficiently on average and that vertically integrated manufacturers dominate the semiconductor industry.

Dedication

To my wife Yingna.

Acknowledgments

I would like to sincerely thank my adviser, Dr. Paul W. Wilson for his guidance throughout my study. I would also like to thank Dr. Frederick Andrew Hanssen, Dr. Matthew S. Lewis, and Dr. Yichen Zhou for serving as members of my dissertation committee and guiding me throughout.

Table of Contents

List of Tables

List of Figures

Chapter 1

Does the asset-light business model create value? A panel data stochastic frontier approach for global semiconductor industry

1.1 Introduction

Semiconductors, also referred to as integrated circuits (ICs) or chips, are crucial elements in the manufacturing of electronics over the last 70 years since the invention of transistors at Bell laboratories in 1947. The semiconductor industry is a driving force in the digital economy and is closely linked to many of the cutting-edge technologies such as advanced wireless networks, artificial intelligence, and quantum computing. During the early years of the semiconductor industry, it almost entirely involves the integrated device manufacturer (IDM) business model, that one company handles all of the production stages, including research and design (R&D), front-end wafer fabrication, and back-end assembly and test (A&T) in-house. A well-known IDM is the microprocessor manufacturer Intel, which nowadays has six wafer fabrication sites, three A&T manufacturing locations, and more than one hundred thousand employees worldwide. Due to the steadily increasing complexity of the leading-edge ICs characterized by Moore's law (e.g., see Flamm, 2017), the enormous capital expenditure (CAPEX) accompanying with proportionally increasing R&D and labor costs imposes a heavy burden even for the largest IDMs, which underlays the birth of the fabless-foundry business model in the semiconductor industry in the 1980s.

In the fabless-foundry business model, the fabless firms focus on the design and sales of chips and partner with pure-play foundries for front-end wafer fabrication as well as the third group of companies for back-end outsourced semiconductor assembly and test (OSAT). Vertical disintegration by the fabless-foundry model drastically reduces the burden of CAPEX in the semiconductor industry and brings up the prosper and flourish of the asset-light fabless firms with diversified products for various applications (e.g., see Sarma and Sun, 2017). The fabless companies, such as Qualcomm and Nvidia, direct all their resources in designing state-of-the-art chips and contract out both front-end wafer fabrication and back-end A&T so that they are risk-free in the setting up, maintaining, and upgrading of the profoundly expensive fabrication facilities. In contrast, the IDMs derive efficiency from vertical integration. In the development of bleeding-edge ICs which requires close coordination between product design and process verification, IDMs achieve performance advantages when technological developments involve complex problems and gain efficiency by the internalization of transaction costs (e.g., see Dibiaggio, 2007 and Kapoor and Adner, 2012). Hence whether the vertical integrated IDM model or the vertical specialized fabless-foundry model operates more efficiently is an empirical question in the global semiconductor industry.

Strategic management approach suggests that intra-industry performance differences can be attributed to sustainable competitive advantage (e.g., see Barney, 1991 and Mahoney, 1995). The resource-based view of competitive advantage specifies that resources are important antecedents to a firm's overall performance as well as sources of sustained competitive heterogeneity among firms (e.g., see Barney, 2001). Liou et al. (2008) and Tang and Liou (2010) suggest extending the causal relationship between competitive advantage and superior performance to a strategy-configuration performance causal series. They apply this theoretical framework to the global semiconductor industry and argue that the presence of competitive advantage of the asset-light business model can be reflected in the causal relationship between resource configuration, dynamic capability, and observable financial performance. However, though the terms competitive advantage and performance are often used interchangeably, the two constructs are acknowledged to be conceptually distinct (e.g., see Powell, 2001 and Newbert, 2008). Furthermore, the debate on a conceptually clear and unambiguous definition of competitive advantage is far from over (e.g., see O'Shannassy, 2008, and Sigalas et al., 2013).

Production frontier is another econometric approach for performance evaluation. There are rich records of efficiency estimation by production frontier in the semiconductor industry, most of which follow the data envelopment analysis (DEA) method. For example, Chu et al. (2008) use the DEA technique to measure the relative performance for global leading fabless firms, while Lu et al. (2013) and Lin et al. (2019) use the DEA model to study the semiconductor industry in the US and Taiwan respectively. Despite its flexible functional form, a main drawback of the DEA approach is the ignoring of statistical noise and accounting for all deviations from the frontier to inefficiency. In contrast, the stochastic frontier analysis (SFA) approach has the attraction of naturally including an error term in the econometric regression framework, but it also has the disadvantage of requiring ex-ante functional form for both the frontier and the inefficiency term. For instance, Kumbhakar et al. (2012) apply the SFA framework to investigate the impact of R&D activities on firm performance, but their approach mixes up firms from different industries, which hardly seems to correspond to the assumption of a consistent production function.

This paper plans to merge the advantages of both the strategic management approach and the SFA approach to investigate the impact of business model on firm-level operating efficiency in the global semiconductor industry. The panel data SFA model by Paul and Shankar (2018, 2020) is chosen for the following reasons. First, this model specifies the efficiency effects by a cumulative distribution function which eschews both the restriction of a one-sided inefficiency term and the transformation to limit the inefficiency scores in a unit interval. Hence efficiency effects can be measured by financial ratios as suggested by Tang and Liou (2010). Second, the unobserved heterogeneities are within-transformed so that it is able to estimate the firm lever efficiency scores under the production frontier for the highly globalized semiconductor industry. Third, it is a one-step approach that the frontier function and the efficiency effects are estimated simultaneously, keeping away from the measurement errors by a two-step procedure (e.g., see Schmidt, 2011).

The paper is organized as follows. Section 1.2 introduces the methodology for this study. Section 1.3 describes the data and defines the variables used for the estimation. Section 1.4 presents the empirical results of the impact of business model on performance assessment in the semiconductor industry. The last section concludes.

1.2 The Methodology

Efficiency and productivity are core concepts of economics. The SFA approach introduced by Aigner et al. (1977), Battese and Corra (1977), and Meeusen and Broeck (1977) has an appealing feature of allowing for both a one-sided inefficiency term and a two-sided statistical noise term. Schmidt and Sickles (1984), Battese and Coelli (1988), etc., extend the SFA framework to panel data with a time-invariant inefficiency term. Cornwell et al. (1990), Battese and Coelli (1992), etc., introduce models with a time-varying inefficiency term in panel data. Greene (2005) argues that these approaches treat unobserved heterogeneity as a measure of inefficiency and proposes a true fixed effects model which distinguishes between time-invariant unit-specific heterogeneity and time-varying inefficiency. However, Greene (2005) uses dummy variables to represent heterogeneity which encounters the incidental parameters problem. Wang and Ho (2010), Chen et al. (2014), Belotti and Ilardi (2018), etc., apply various maximum likelihood approaches to estimate Greene's model, all of which need extra transformation to restrict the inefficiency scores in a unit interval.

Some studies have used a two-step approach, where efficiency scores are estimated in the first step, and the estimates of the efficiency scores are regressed against a set of exogenous variables which are hypothesized to influence a firm's inefficiency in the second step. It is known that the inconsistent assumptions of the inefficiency term between the two steps generates biased estimation in such a two-step approach so that the mainstream of SFA proposes to estimate the efficiency scores and efficiency effects by a one-step procedure (e.g., see Kumbhakar et al., 1991). Battese and Coelli (1995), Kumbhakar and Wang (2005), Alvarez et al. (2006), etc., adopt different techniques to extend the one-step procedure to panel data. Recently, Parmeter et al. (2017) propose a nonparametric approach to estimate the distribution free inefficiency effects which needs additional constraints, such as the method of Du et al. (2013), to get nonnegative estimates of the inefficiency term. Paul and Shankar (2018, 2020) propose a distribution-free efficiency effect model which uses a cumulative distribution function to specify the efficiency effects and eschews the assumption of one-sided inefficiency term.

The Paul and Shankar (2020) model is an extension of Paul and Shankar (2018) which accounts for time-invariant unobserved heterogeneity and can be expressed as

$$Y_{it} = \exp\left(\alpha_i + x_{it}\beta + v_{it} + \frac{1}{\mu}\ln[H(z_{it}\gamma)]u_{it}\right), \tag{1.2.1}$$

where $i = 1, ..., N$ denotes each of the individual firms, $t = 1, ..., T_i$ denotes the observed time period of each i, α_i is the firm-specific fixed effect, v_{it} represents the random noise, and $\frac{1}{\mu}\ln[H(z_{it}\gamma)u_{it}]$ is a one-sided inefficiency term with the restriction that $0 \leq H(z_{it}\gamma) \leq 1$, and $\mu = E(u_{it})$. The equation (2.1) can be written in a logarithmic form as

$$y_{it} = \alpha_i + x_{it}\beta + \ln[H(z_{it}\gamma)] + \varepsilon_{it}, \tag{1.2.2}$$

where $y_{it} = \ln(Y_{it})$, and $\varepsilon_{it} = v_{it} + \frac{1}{\mu}\ln[H(z_{it}\gamma)](u_{it} - \mu)$. After the within transformation to eliminate the unobserved influence of α_i, we can simplify (2.2) as

$$\widetilde{y}_{it} = \widetilde{x}_{it}\beta + \ln\left(\frac{H(z_{it}\gamma)}{\left(\prod_{p=1}^{T_i} H(z_{ip}\gamma)\right)^{\frac{1}{T_i}}}\right) + \widetilde{\varepsilon}_{it} \tag{1.2.3}$$

where $\widetilde{w}_{it} = w_{it} - \frac{1}{T_i}\sum_{p=1}^{T_i} w_{ip}$ for $w \in \{y, x, \varepsilon\}$. Equation (2.3) can be estimated by the nonlinear least squares (NLS) estimator which minimizes the sum of squared errors

$$\arg\min_{\beta,\gamma} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \left(\widetilde{y}_{it} - \widetilde{x}_{it}\beta - \ln \left(\frac{H(z_{it}\gamma)}{\left(\prod_{p=1}^{T_i} H(z_{ip}\gamma)\right)^{\frac{1}{T_i}}} \right) \right)^2 \tag{1.2.4}$$

with respect to parameters β and γ. Paul and Shankar (2020) prove the consistency of the NLS estimator in (2.4) and propose to estimate the variance of the random noise v_{it} by

$$\widehat{\sigma_v}^2 = \frac{1}{L-K} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \left(\widehat{\varepsilon_{it}}^2 - \frac{[\ln(H(z_{it})\widehat{\gamma})]^2}{\widehat{\mu\delta}} \right) \tag{1.2.5}$$

where $\widehat{\mu\delta} = \sqrt{\frac{2\sum_{i=1}^{N}\sum_{t=1}^{T_i}[\ln(H(z_{it}\widehat{\gamma}))]^6}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}[\widehat{\varepsilon_{it}}\ln(H(z_{it}\widehat{\gamma}))]^3}}$, $L = \sum_{i=1}^{N} T_i$, and K is the total number of parameters β and γ.

Once the coefficient vectors β and γ are estimated, the individual fixed effects α_i can be retrieved as

$$\widehat{\alpha}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} \left(y_{it} - x_{it}\widehat{\beta} - \ln(H(z_{it}\widehat{\gamma})) \right) \tag{1.2.6}$$

and the mean technical efficiency can be derived directly as

$$\widehat{TE}_{it} = \exp\left(\ln[H(z_{it}\widehat{\gamma})]\right) = H(z_{it}\widehat{\gamma}), \tag{1.2.7}$$

which avoids the transformation to calculate the efficiency scores. The selection of the function $H(z_{it}\gamma)$ is flexible, with the only restriction that $H(z_{it})$ is in a unit interval. A cumulative distribution function such as $\Phi(z_{it}\gamma)$ or any function constrained to lie between 0 and 1, such as the Gumbel function of the form $G(z_{it}\gamma) = e^{-e^{z_{it}\gamma}}$, would be suitable for $H(z_{it})$. Another feature of using $H(z_{it}\gamma)$ to represent the technical efficiency is that it eschews the widely used assumption of a one-sided distribution of the inefficiency term in almost all the existing SFA models. Hence, it is convenient to bridge the efficiency measure $H(z_{it})$ to the operating performance, such as the financial ratios in Liou et al. (2008) and Tang and Liou (2010).

1.3 Data and Variable Specification

The data are collected from the sub-industry of semiconductors in Compustat database over the period of 20 years 1999–2018. Since the semiconductor industry is highly globalized, I combine data from both the Compustat North America database and the Compustat Global database to cover companies in the whole industry. I exclude photovoltaic producers, liquid crystal display manufactures, and light-emitting diodes manufacturers from the dataset, limiting the sample within only IC manufactures. Under such restriction, the sample includes 5136 observations from 470 unique companies in 1999–2018. Table 1.1 breaks down the sample by four kinds of business models, including fabless, IDM, foundry, and OSAT, which can naturally be grouped into three categories by the intensity of labor and capital. The first category contains fabless companies that are asset-light but labor intensive for chip design. Over half of the companies in the semiconductor industry are in the fabless model since the barriers to entry are much lower for the asset-light fabless companies than for the asset-heavy manufacturers. The second category contains foundries and OSATs, both of which focus on fabrication and depend heavily on CAPEX for the capital-intensive facility construction and equipment maintenance. The third category contains IDMs which are both labor-intensive and capital-intensive because IDMs carry out all stages of production in-house.

Identifying the inputs and outputs has always been a subject of controversy in the estimation of production frontier, without exception in the semiconductor industry. Hence I sort the most commonly used variables in the empirical papers which apply the production frontier approach for performance evaluation in the semiconductor industry and specify one output (revenue (Y)) and three inputs (labor, measured by the number of employees (X_1); cost of goods sold (X_2); and capital investment, measured by property, plant, and equipment (PP&E) (X_3)) for the production function. For the efficiency term, I follow the theory of competitive advantage, especially in case of the asset-light business model (e.g., see Liou et al., 2008 and Tang and Liou, 2010), to specify two financial ratio variables (fixed asset turnover ratio, measured by the revenue of a company divided by the value of its fixed assets (Z_1); and R&D expense to revenue ratio, measured by the percentage of sales that is allocated to R&D expenditures (Z_2)). As there are more than one hundred financial ratios in common use, Liou et al. (2008) and Tang and Liou (2010) apply a principle component analysis to identify three key factors and find that the fixed asset turnover ratio Z_1 is a key indicator of the capital management ability of a company while the R&D expense to revenue ratio Z_2 is a dedicated indicator of the knowledge management in the semiconductor industry. The other financial indicators, either overlapped or duplicated with Z_1 or Z_2, or related to the customer and supplier relationship factor, are not selected as efficiency variables in this paper. Another advantage of using the financial ratios Z_1 and Z_2 is the scale-invariant feature which matches well with the dimensionless efficiency scores. Table 1.2 gives summary statistics of the variables in 1999–2018 pooled data split by the business models. The values of X_1–X_3 and Y in Table 1.2 are in the form before the log transformation and adjusted to the 2018 US dollar by GDP deflator to set up a criterion for comparing data across different years. The distributions of Z_1 and Z_2 are skewed to the right extremely for the fabless firms, which are consistent with the asset-light feature of the fabless model in the semiconductor industry.

1.4 Estimation Results

Estimation of the model in (2.2) is straightforward by NLS after removing the individual fixed effects by the within transformation in (2.3). Table 1.3 presents the estimates of the parameters in (2.4) with three kinds of different functions for $H(z_{it}\gamma)$, including the probit cumulative distribution function, the logit cumulative distribution function, and the Gumbel function, to compare the impact of the functional form for the efficiency term $H(z_{it}\gamma)$. The estimates of output elasticities, which are represented by the coefficients of inputs in the translog production function, are positive and statistically significant and do not vary much among the three models with different forms of $H(z_{it}\gamma)$. In terms of the magnitude of elasticity, capital-investment which is represented by PP&E turns out to be a more important factor of production than labor. It is consistent with the fact that the semiconductor industry is capital-intensive more than labor-intensive by and large. In terms of the efficiency term $H(z_{it}\gamma)$, the positive sign of γ_1 implies that the higher the asset turnover ratio, the more efficient a company is at generating revenue from its assets. Similarly, the negative sign of γ_2 indicates that the efficiency of a firm decreases with the level up of R&D expenditure, implying that the severe competition and the continuous iteration of technology in the semiconductor industry make the heavily R&D spending an risky investment. Both the positive sign of γ_1 and the negative sign of γ_2 are consistent with the works of Liou (2011), Tsai et al. (2017), etc.

Figure 1.1 plots the distribution of individual fixed effects derived by (2.6). The heterogeneities among the business models are distinct and consistent in each form of $H(z_{it}\gamma)$. Figure 1.2 shows the distribution of the efficiency scores calculated by (2.7). The efficiency scores of the asset-light fabless companies have a relatively smoother distribution, while the efficiency scores of the capital intensive

IDMs, foundries, and OSATs have sharp-peaked distributions. Table 1.4 provides the pairwise Kolmogorov–Smirnov test results and Mann-Whitney test results of the distributions of the efficiency scores shown in Figure 1.2, both of which indicate that the distributions of the efficiency scores are different in all the pairwise comparisons among the business model. Furthermore, the means and standard deviations of the efficiency scores by business model over the 20 years are shown in Table 1.5 and visualized in Figure 1.3. The curve of annual mean efficiency scores of the fabless firms is conspicuously above the curves of the other business models. A plausible explanation for these results is that although investing in R&D is risky in the semiconductor industry, the fabless business model still has the attraction of lifting the heavy CAPEX burden off the small and medium enterprises' shoulders.

1.5 Summary and Conclusions

Comparing the operating efficiency between the vertical integrated IDM model and the specialized fabless-foundry model in the semiconductor industry where diversified companies are producing various products is a vexing problem. This paper applies a panel data stochastic frontier approach which has the advantage of disentangling the firm-level heterogeneity by the within transformation and estimate the efficiency scores by cumulative distribution functions. The nonlinear least squares technique in this approach eschews a priori knowledge of a one-sided inefficiency term present in almost all the existing inefficiency effects models and provides the flexibility to link the efficiency terms with the financial ratios of a firm. The estimation results indicate that the asset-light fabless companies are operating more efficiently than the firms in other operating models in the semiconductor industry. Though the vertical integrated IDMs dominate the semiconductor industry since its early days, the heavy burden of CAPEX and the law of diminishing marginal returns induce more and more companies to embrace the fabless-foundry model. Facing the high uncertainty of commercial success due to technology iteration twisting to the ups and downs in the global economic cycle, the small and medium-sized fabless companies are more flexible and adaptable to market changes in the semiconductor industry.

However, the distinction between the IDM model and the fabless-foundry model is fading away. The vertical specialized fabless-foundry model has the attraction of risk sharing and achieving high capacity utilization, so that IDMs also start to contract with other companies to manufacture some of their chips while performing all other remaining tasks internally. The complementarity and integration of the IDMs and the fabless-foundry firms can further expand the range of potential end-user applications for ICs and enable the entire semiconductor industry to thrive and prosper. This developing trend may lead to new business models in the semiconductor industry with higher operating efficiencies in the near future.

Table 1.1: Number of observations by business model

Year	All	Fabless	IDM	Foundry	OSAT
1999	125	68	38	10	9
2000	149	81	43	10	15
2001	155	83	46	10	16
2002	213	121	48	17	27
2003	241	143	49	19	30
2004	264	159	54	21	30
2005	260	162	54	17	27
2006	267	161	56	20	30
2007	269	163	52	21	33
2008	278	172	51	20	35
2009	290	180	53	21	36
2010	300	180	59	23	38
2011	298	177	60	22	39
2012	301	180	61	22	38
2013	313	183	65	24	41
2014	302	172	62	25	43
2015	288	163	59	24	42
2016	283	162	54	23	44
2017	275	156	51	23	45
2018	265	151	48	22	44
N	5,136	3,017	1,063	394	662
N_u	470	288	83	36	63

NOTE. N denotes the number of observations in the period 1999–2018. N_u denotes the number of unique firms in the period 1999–2018.

Table 1.2: Descriptive statistics

	Min	Q1	Median	Mean	Q3	Max
			Fabless			
X_1	1	113	248	803	590	35,400
X_2	1	15,465	52,182	218,222	147,582	10,210,237
X_3	5	2,748	9,412	69,806	29,689	5,627,962
Y	3	32,198	100,203	460,156	280,885	28,365,696
Z_1	0.009	4.772	9.951	27.039	21.186	1473.976
Z_2	0.000	0.118	0.191	1.084	0.305	663.200
			IDM			
X_1	28	780	2,900	8,525	8,400	107,600
X_2	49	108,201	363,525	1,230,517	1,102,348	18,226,000
X_3	298	53,482	222,571	1,524,168	917,191	4,8976,000
Y	2,679	184,205	741,897	2,932,830	2,250,155	70,848,000
Z_1	0.084	1.704	2.629	5.018	4.120	203.801
Z_2	0.000	0.058	0.120	0.132	0.181	1.663
			Foundry & OSAT			
X_1	19	439	1,577	4,114	3,931	93,891
X_2	1,123	44,567	143,472	449,317	419,758	8,841,157
X_3	88	44,868	167,886	961,037	615,005	36,542,569
Y	1,840	71,401	226,760	908,505	671,050	33,696,798
Z_1	0.039	0.794	1.218	6.792	1.924	1789.268
Z_2	0.000	0.018	0.036	0.053	0.062	0.985

NOTE. The unit of X_1 is the number of employees.
The units of X_2, X_3 and Y are thousands US$.
All the values of X_2, X_3 and Y are adjusted to the 2018 US$ by GDP deflator.

Table 1.3: Estimated stochastic frontiers and technical efficiency effects

	Efficiency Term — Logit —		Efficiency Term — Probit —		Efficiency Term — Gumbel —	
	Coef	SD	Coef	SD	Coef	SD
Frontier Function						
β_1	0.155^{***}	0.009	0.150^{***}	0.009	0.136^{***}	0.008
β_2	0.404^{***}	0.007	0.391^{***}	0.007	0.348^{***}	0.007
β_3	0.398^{***}	0.008	0.418^{***}	0.008	0.482^{***}	0.008
Efficiency effects						
γ_0	-1.469^{***}	0.035	-0.961^{***}	0.021	-0.785^{***}	0.016
γ_1	0.147^{***}	0.003	0.090^{***}	0.002	0.087^{***}	0.002
γ_2	-0.006^{***}	0.000	-0.002^{***}	0.000	-0.001^{***}	0.000
Wald stat.	$3,509^{***}$		$4,396^{***}$		$6,494^{***}$	
$\widehat{\sigma_v}^2$	0.403		0.426		0.600	
N	5,135		5,135		5,135	
N_u	470		470		470	

NOTE. The null hypothesis in the Wald test is $\gamma_0 = \gamma_1 = \gamma_2 = 0$.
N denotes the number of observations in the peroid 1999–2018.
N_u denotes the number of unique firms in the peroid 1999–2018.

Table 1.4: Tests of difference of the distributions of efficiency scores by business model

	— Logit —		— Probit —		— Gumbel —	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
			Kolmogorov–Smirnov test			
Fabless VS. IDM	0.557***	0.000	0.557***	0.000	0.557***	0.000
Fabless VS. Foundry &OSAT	0.767***	0.000	0.768***	0.000	0.768***	0.000
IDM VS. Foundry &OSAT	0.455***	0.000	0.455***	0.000	0.455***	0.000
			Mann-Whitney test			
Fabless VS. IDM	2,663,969***	0.000	2,664,607***	0.000	2,665,456***	0.000
Fabless VS. Foundry &OSAT	2,953,105***	0.000	2,954,469***	0.000	2,955,333***	0.000
IDM VS. Foundry &OSAT	247,099***	0.000	246,849***	0.000	246,711***	0.000

NOTE. H_0 in the Kolmogorov–Smirnov test is that the two distributions are equal.
H_0 in the Mann-Whitney test is that the two distributions are equal.

Table 1.5: Mean and standard deviation of efficiency scores

	Logit			Normal			Gumbel		
Year	Fabless	Foundry &OSAT	IDM	Fabless	Foundry &OSAT	IDM	Fabless	Foundry &OSAT	IDM
1999	0.456	0.246	0.290	0.436	0.227	0.271	0.364	0.167	0.212
	(0.217)	(0.068)	(0.153)	(0.221)	(0.067)	(0.155)	(0.210)	(0.063)	(0.158)
2000	0.450	0.257	0.290	0.432	0.238	0.268	0.362	0.177	0.209
	(0.232)	(0.106)	(0.147)	(0.235)	(0.104)	(0.149)	(0.222)	(0.096)	(0.151)
2001	0.440	0.237	0.271	0.420	0.218	0.252	0.353	0.158	0.193
	(0.228)	(0.090)	(0.135)	(0.230)	(0.088)	(0.136)	(0.226)	(0.082)	(0.141)
2002	0.487	0.227	0.272	0.469	0.208	0.254	0.402	0.150	0.195
	(0.263)	(0.073)	(0.132)	(0.269)	(0.071)	(0.133)	(0.268)	(0.067)	(0.138)
2003	0.539	0.240	0.281	0.522	0.221	0.262	0.453	0.161	0.202
	(0.274)	(0.096)	(0.133)	(0.281)	(0.094)	(0.134)	(0.283)	(0.087)	(0.138)
2004	0.601	0.239	0.282	0.585	0.219	0.262	0.512	0.160	0.203
	(0.270)	(0.094)	(0.129)	(0.278)	(0.092)	(0.130)	(0.279)	(0.085)	(0.133)
2005	0.597	0.231	0.301	0.582	0.212	0.282	0.510	0.153	0.221
	(0.276)	(0.089)	(0.162)	(0.284)	(0.088)	(0.163)	(0.283)	(0.080)	(0.159)
2006	0.599	0.239	0.315	0.583	0.221	0.296	0.510	0.162	0.233
	(0.271)	(0.112)	(0.176)	(0.278)	(0.115)	(0.177)	(0.277)	(0.110)	(0.171)
2007	0.590	0.235	0.296	0.574	0.217	0.276	0.499	0.158	0.213
	(0.263)	(0.108)	(0.129)	(0.272)	(0.112)	(0.128)	(0.268)	(0.107)	(0.117)
2008	0.625	0.247	0.304	0.610	0.229	0.285	0.536	0.171	0.221
	(0.270)	(0.146)	(0.145)	(0.279)	(0.150)	(0.145)	(0.279)	(0.151)	(0.132)
2009	0.603	0.225	0.298	0.589	0.206	0.278	0.520	0.148	0.215
	(0.287)	(0.059)	(0.138)	(0.296)	(0.058)	(0.138)	(0.298)	(0.054)	(0.126)
2010	0.629	0.258	0.311	0.614	0.240	0.291	0.544	0.181	0.228
	(0.279)	(0.147)	(0.128)	(0.288)	(0.150)	(0.127)	(0.292)	(0.148)	(0.116)
2011	0.584	0.252	0.298	0.570	0.234	0.278	0.503	0.176	0.216
	(0.284)	(0.146)	(0.120)	(0.293)	(0.150)	(0.120)	(0.298)	(0.151)	(0.112)
2012	0.576	0.252	0.293	0.561	0.234	0.273	0.495	0.175	0.211
	(0.281)	(0.145)	(0.113)	(0.290)	(0.148)	(0.114)	(0.298)	(0.147)	(0.104)
2013	0.591	0.240	0.300	0.576	0.222	0.280	0.510	0.162	0.218
	(0.283)	(0.103)	(0.112)	(0.292)	(0.104)	(0.112)	(0.298)	(0.096)	(0.103)
2014	0.560	0.244	0.311	0.543	0.225	0.291	0.474	0.166	0.229
	(0.274)	(0.102)	(0.139)	(0.282)	(0.104)	(0.140)	(0.285)	(0.100)	(0.134)
2015	0.544	0.238	0.310	0.527	0.219	0.291	0.458	0.160	0.229
	(0.275)	(0.086)	(0.143)	(0.283)	(0.088)	(0.144)	(0.282)	(0.080)	(0.137)
2016	0.552	0.242	0.303	0.535	0.223	0.284	0.465	0.164	0.222
	(0.273)	(0.084)	(0.136)	(0.280)	(0.085)	(0.138)	(0.282)	(0.078)	(0.130)
2017	0.546	0.244	0.300	0.529	0.225	0.280	0.460	0.167	0.217
	(0.271)	(0.098)	(0.123)	(0.278)	(0.100)	(0.123)	(0.280)	(0.102)	(0.112)
2018	0.552	0.247	0.306	0.536	0.228	0.287	0.466	0.170	0.224
	(0.280)	(0.104)	(0.158)	(0.288)	(0.106)	(0.160)	(0.287)	(0.109)	(0.147)

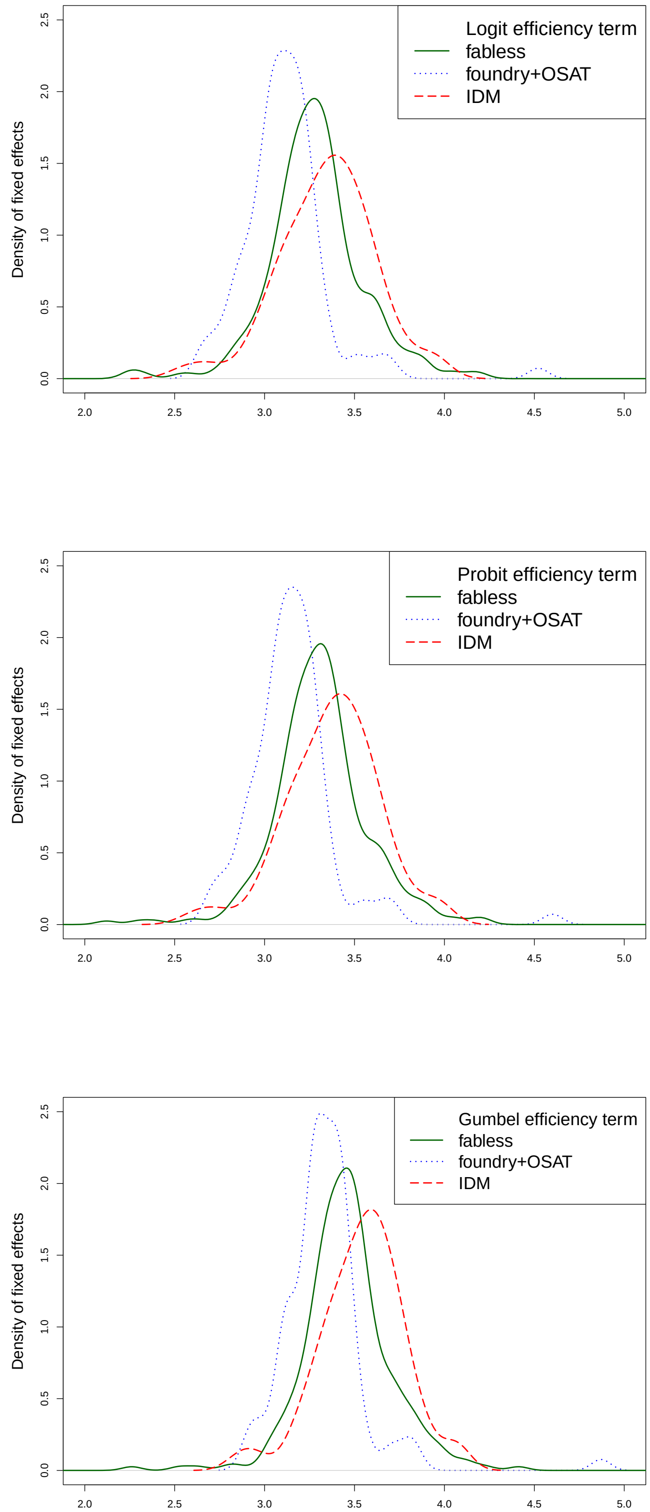

Figure 1.1: Distribution of fixed effects by business model

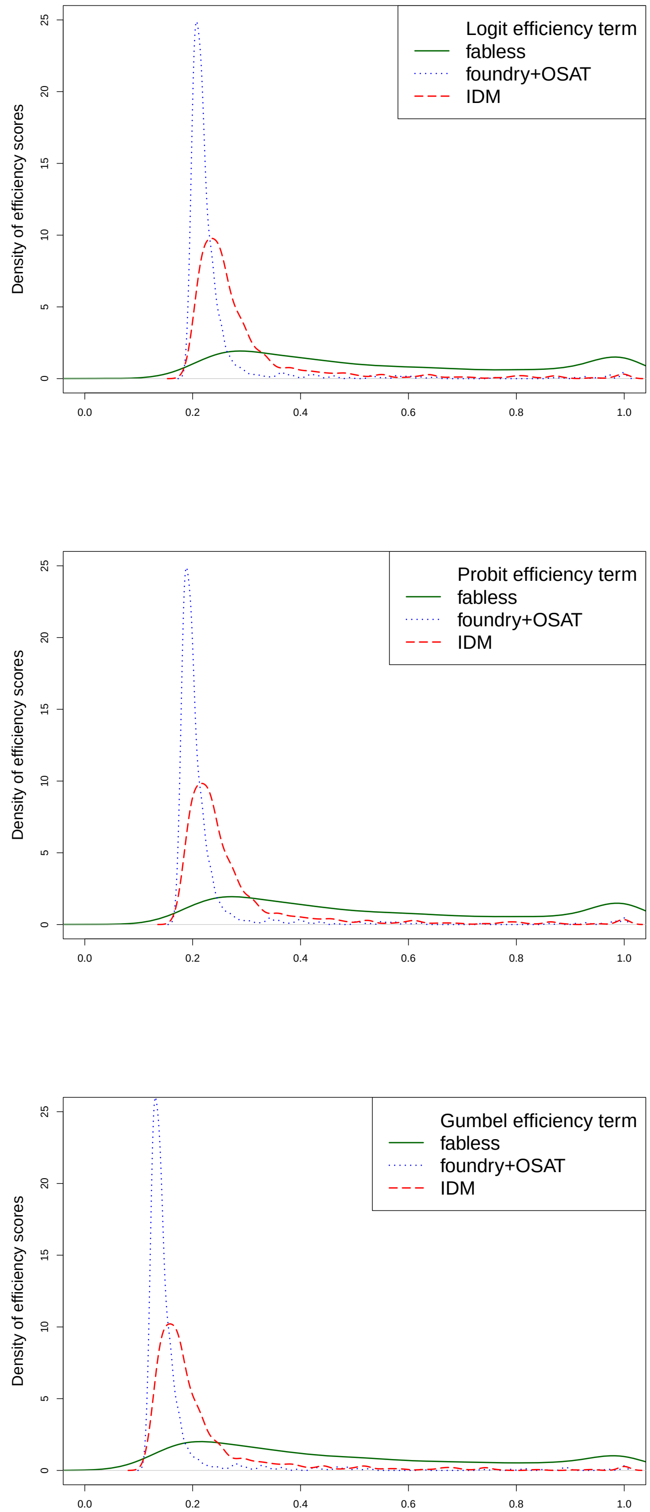

Figure 1.2: Distribution of efficiency scores by business model

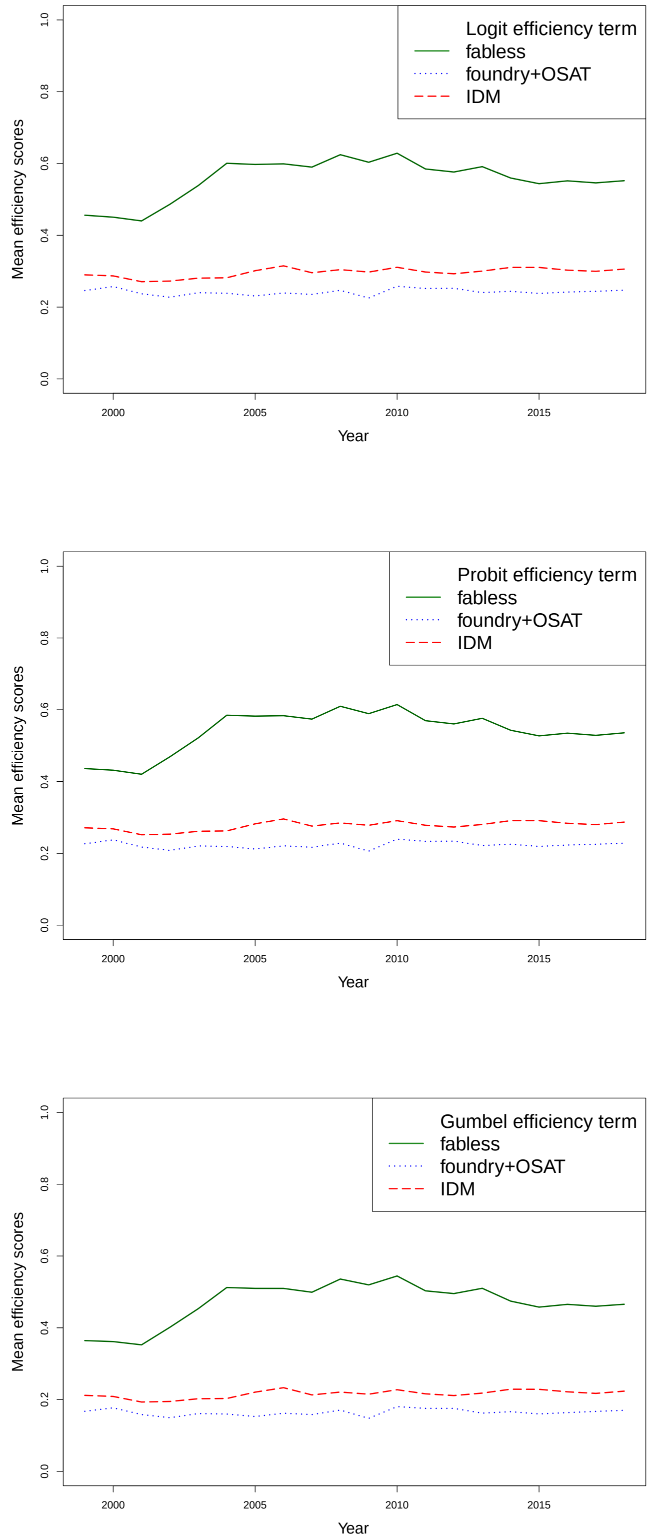

Figure 1.3: Trends of mean efficiency scores

Chapter 2

Tradeoffs between economies of scale and specialization: A constrained semiparametric least squares estimation for global semiconductor industry

2.1 Introduction

Semiconductors, also referred to as integrated circuits (ICs) or chips, are some of the most complicated products to design and produce on earth, and are arguably humanity's greatest achievement to date. The semiconductor industry, which is closely linked to many of the cutting-edge technologies such as advanced wireless networks, artificial intelligence, quantum computing, etc., is a driving force in the information age and is playing an essential role in global economic growth. Due to ever-increasing complexity of the leading-edge ICs characterized by Moore's law (e.g., see Mack, 2011), manufacturing is cost prohibitive except for all but the largest fabrications in the world. The enormous capital expenditures (CAPEX) are known to be critical of the barriers to entry in semiconductor industry, accompanying with proportionally increasing labor inputs and research and design (R&D) costs. Hence the integrated device manufacturers (IDMs), such as Intel and Texas Instruments, which handle all the production stages including R&D, front-end fabrication, and back-end assembly and test (A&T) in-house, dominate the semiconductor industry since its early days.

With the escalating complexity of ICs, the evolution of the semiconductor industry is shifting from competing on performance edge or lower price to competing on continuous technological innovation for new features. The growing demands for diversity and the raising requirements for specialization have led to the emergence of the fabless-foundry business model in semiconductor industry (e.g., see Brown et al., 2005 and Fontana and Malerba, 2010). In fabless-foundry model, the fabless companies engage solely in the design of chips, and partner with pure-play foundries for front-end fabrication and a third group of companies for back-end A&T. Since the mid-1980s, the fabless-foundry model facilites the entry of new fabless firms in the exponentially growing semiconductor industry (e.g., Balconi and Fontana, 2011). For instance, the fabless companies such as Qualcomm, Nvidia, and Semtech, direct all their funding in designing state-of-the-art chips and are risk-free in terms of the heavy CAPEX burden in setting up, maintaining, and upgrading fabrication facilities. In contrast, foundries require cleanroom and a set of profoundly expensive equipment to function, representing a financial barrier that most chip designers find impossible to surmount (e.g., see Mathews and Cho, 2000). Similarly, the major challenges for the top-notch outsourced semiconductor assembly and test (OSAT) firms are the costly CAPEX for high-end packaging solutions and the market instability. Both the foudries and OSATs seek to optimize productivity by serving many fabless companies to achieve high capacity utilization. For example, Taiwan Semiconductor Manufacturing Company (TSMC) is the first and nowaday one of the largest pure-play foundries that commit to be long term non-competitive partner with the fabless firms (e.g., see Hsieh et al., 2002).

Which business model is operating more efficiently in the semiconductor industry is a topic of wide interest. The IDM model derives efficiency from vertical integration. The huge CAPEX on semiconductor manufacturing plays the role of an economic moat for the IDMs, especially for the bleeding-edge product lines where close coordinates between product design and process development are required (e.g., see Kapoor and Adner, 2012). The predominant large IDMs, such as Intel, STMicroelectronics and Texas Instruments, take advantage of vertical integration to forge ahead with innovative and expensive technologies that are needed to iterate chips to new levels. Alternatively, the fabless-foundry model derives efficiency from delineation of tasks and specialization (e.g., see Shin et al., 2017). The risk sharing vertical specialization reduces the barriers to entry of new firms in the semiconductor inductry and expands the range of potential end-user applications for chips. Furthermore, the fabless–foundry model may shorten the cycle time from IC design to IC mass-production by using a modular design and manufacturing concept and provide stronger protection of intellectual property rights (e.g., see Sarma and Sun, 2017). Hence the cost-benefit analysis between the vertical integrated IDM model and the vertical specialized fabless-foundry model in semiconductor industry is an empirical question. This paper aims to shed light on the performance evaluation in the semiconductor industry by applying the semiparametric modified ordinary least squares (MOLS) approach, and the comparison of operating efficiency between the IDM model and the fabless-foundry model.

There are rich records of performance evaluation in the semiconductor industry. Macher and Mowery (2004), Wang and Chiu (2014), etc., compare pros and cons of the IDM model and the fabless-foundry model in semiconductor industry by using intuitive but purely qualitative analyses. Kapoor (2013), Shin et al. (2017), etc., choose the popular ordinary least squares (OLS) approach for performance assessment in the semiconductor industry, relying their estimates on linear models. Jang et al. (2016), Li et al. (2019), etc. use the nonparamtric data envelopment (DEA) approach for efficiency estimation of the semiconductor industry. Although the data driven DEA approach has an appealing feature of flexible functional form, the strong and non-testable assumption of no measurement error or random variation of the production frontier is considered the Achilles' heel in the determinstic DEA approach. Furthermore, when the studies are restricted to a small number of observations either by geographic boundaries or by business model boundaries, the measurement error are likely to be further amplified because of the slow convergence rate in the DEA approach.

An alternative approach is the stochastic frontier analysis (SFA) model, which naturally includes an error term in econometric regression. However, the standard SFA approach which is building on parametric regression techniques pays a price of assuming an ex ante functional form. Kumbhakar et al. (2012) and Bonanno (2016) use the SFA approach to investigate R&D productivity and firm efficiency, but the data they use mix up firms from different industries which hardly seem to match with the assumption of a consistent parametric functional form. The semiparametric developments, such as Kneip and Simar (1996), Kuosmanen (2008), Simar and Zelenyuk (2011), Cai et al. (2015), etc., come a long way in bridging the gap between DEA and SFA approaches. One of the latest progress in these efforts is the Simar et al.

(2017) approach, which propose a semiparametric version of the MOLS method for efficiency estimation with both flexible functional form assumption and straightforward estimation and computation convenience. This paper follows the Simar et al. (2017) approach to investigate the performance diversities among the different business models in the highly globalized semiconductor industry. The 'wrong skewness' problem in the stochastic frontier model (e.g., see Almanidis and Sickles, 2011) is solved by the Du et al. (2014) approach of shape constraint nonparametric regression.

The paper is organized as follows. Section 2.2 introduces the statistical model for this study. Section 2.3 discusses the estimation methods and the intepretation tools in this research. Section 2.4 describes the dataset of global semiconductor industry and defines the variables used for evaluation. Section 2.5 presents the empirical results of the impact of business model for performance assessment in the semiconductor industry. A last section concludes.

2.2 The Statistical Model

Efficiency and productivity are core concepts of economics. DEA and SFA are two commonly used approaches prevalent in production frontier estimation and efficiency measurement which employ quite distinct methodologies with associated strengths and weaknesses. The DEA approach which is popularized by Banker et al. (1984), Deprins et al. (1984), Färe et al. (1985), etc. has the attraction of being nonparametric, but also has the drawback of not allowing for statistical noise and attributing all deviations from the frontier to inefficiency. In contrast, the SFA approach, introduced by Aigner et al. (1977), Battese and Corra (1977), and Meeusen and Broeck (1977), has a very appealing feature of allowing for both an inefficiency term and an error term, but also has the disadvantage of requiring ex ante functional form of the frontier and the inefficiency distribution. The tradeoff between the merit of flexible functional form in nonparametric DEA model and the merit of allowing idiosyncratic error term in SFA regression model is a vexing problem. Simar and Wilson (2015) provide a guided tour for the recent works to bridge the gap between the parametric and nonparametric worlds in the field of efficiency analysis.

A general form of the SFA model is characterized by

$$\begin{aligned} Y &= m(x, z) + \varepsilon, \\ \varepsilon &= V - U, \end{aligned} \tag{2.2.1}$$

where $x \in X$, $z \in Z$, X represent production inputs, and Z reflect heterogeneous environmental conditions. The production frontier is denoted by $m(x, z)$, and the error term ε is a convolution of a one-sided inefficiency term U and a two-sided noise term V. Although the Cobb-Douglas and translog models overwhelmingly dominate the application literatures in SFA approach, a priori specified functional form in such parametric models inevitably stands to be target of criticism. On the other hand, Hall and Simar (2002) point out that a fully nonparametric model with both noise term V and inefficiency term U is unidentifiable, so that some restriction on the model is required. One compromise between the two extreme cases of a fully parametric model and a fully nonparametric model is to leave the production function $m(x, z)$ unspecified while specifying a parametric model for the stochastic part ε in (2.1). Fan et al. (1996), Kuosmanen and Kortelainen (2012) follow this semiparametric strategy but their estimation retain a homoscedasticity assumption of the stochastic terms which may import misspecification errors into the model and lead to unconvining results. Kumbhakar et al. (2007) propose a local maximum likelihood estimation (LMLE) approach which replies on a local parametric stochastic assumption on the independent random variables U and V that $U|X = x, Z = z \sim |N(0, \sigma_U^2(x, z))|$ and $V|X = x, Z = z \sim N(0, \sigma_V^2(x, z))$ to approximate the unknown functional parameters $m(x, z)$, $\sigma_U^2(x, z)$, and $\sigma_V^2(x, z)$. Simar et al. (2017) offer an alternative approach to LMLE with less assumptions and mitigated computational burdens, which is referred to as a semiparametric version of the MOLS method . This paper mainly follows the Simar et al. (2017) approach for empirical study in the semiconductor industry.

The basic idea of the MOLS approach, first introduced by Afriat (1972), Richmond (1974) and later nominated by Lovell (1993), is to use the higher order moments of the convoluted error term to construct estimator of the frontier. The MOLS approach modifies (2.1) as

$$\begin{aligned} Y &= r_1(x, z) + \varepsilon, \\ \varepsilon &= V - U + \mu_U(x, z), \end{aligned} \tag{2.2.2}$$

where U, V are independent random variables same as in (2.1), $\mu_U(x, z) = E(U|x, z)$, and $r_1(x, z)$ is defined as the average production function that $r_1(x, z) = m(x, z) - \mu_U(x, z)$. Clearly, $E(\varepsilon|x, z) = 0$ and $Var(\varepsilon|x, z) = var_U(x, z) + var_V(x, z)$ by construction. Thereafter the conditional moments of ε can be identified upon adding appropriate parametric assumptions on the moments of U and V. In case of following the popular half-normal assumption of the inefficiency term that $U|X = x, Z = z \sim |N(0, \sigma_U^2(x, z))|$, the second and third conditional moments of ε in (2.2) can be derived as

$$r_2(x, z) = E(\varepsilon^2|x, z) = \sigma_V^2(x, z) + \left(\frac{\pi - 2}{\pi}\right) \sigma_U^2(x, z) \geq 0 \tag{2.2.3}$$

and

$$r_3(x, z) = E(\varepsilon^3|x, z) = \sqrt{\frac{2}{\pi}} \left(1 - \frac{4}{\pi}\right) \sigma_U^3(x, z) \leq 0. \tag{2.2.4}$$

Note that the standard deviations of σ_U and σ_V are required to be nonnegative which set constraints for $r_2(x,z)$ and $r_3(x,z)$ in (2.3) and (2.4). A consistent estimate of the stochastic frontier $m(x,z)$ can be recovered by solving (2.3) and (2.4) for $\sigma_U(x,z)$ and $\sigma_V(x,z)$, which is the main topic of the next section.

2.3 Estimation and Interpretation

The nonparametric local polynomial least squares (LPLS) estimator, which is introduced by Fan and Gijbels (1992), can be used to estimate the conditional moments $r_1(x,z)$, $r_2(x,z)$, and $r_3(x,z)$ which are defined in Section 2.2. The estimation of $r_1(x,z)$ is straightforward because there is no constraint on $r_1(x,z)$ in (2.2). A general kernel regression estimator of $r_1(x,z)$ at any data point i can be written as linear combinations of the response Y_j that

$$\widehat{r}_1(X_i, Z_i) = \sum_{j=1}^{n} W_{j,h_1}(x,z)Y_j, \tag{2.3.1}$$

where $W_{j,h_1}(x,z)$ is a local weighting matrix and h_1 is a vector of bandwidths. This includes the Nadaraya-Watson estimator (Nadaraya, 1965 and Watson, 1964), the Priestley-Chao estimator (Priestley and Chao, 1972), the Gasser-Miiller estimator (Gasser and Miiller, 1979), and the LPLS estimator used in this paper. The corresponding estimate of ε in (2.2) can be written as

$$\widehat{\varepsilon}_i = Y_i - \widehat{r}_1(X_i, Z_i). \tag{2.3.2}$$

The estimation of $r_2(x,z)$ and $r_3(x,z)$ can also be done by using the nonparametric kernel regression estimators, but here the regressions are constrained by the restriction that both $\sigma_U(x,z)$ and $\sigma_V(x,z)$ should be nonnegative in (2.3)-(2.4).

Du et al. (2013) propose a shape constrained kernel regression method that the estimators of $r_2(x,z)$ and $r_3(x,z)$ are given by

$$\widehat{r}_2(x,z) = \sum_{i=1}^{n} W_{i,h_2}(x,z) p_{i2} \widehat{\varepsilon}_i^2 \tag{2.3.3}$$

and

$$\widehat{r}_3(x,z) = \sum_{i=1}^{n} W_{i,h_3}(x,z) p_{i3} \widehat{\varepsilon}_i^3, \tag{2.3.4}$$

where $W_{i,h_2}(x,z)$ and $W_{i,h_3}(x,z)$ are local weighting matrices similar as defined in (3.1), and the probability weights p_{i2} and p_{i3} can be solved by quadratic programming (QP) method respectively. The main idea of the QP method is to select a vector p that minimize the distance metric

$$D(p) = (p_u - p)'(p_u - p) \tag{2.3.5}$$

subject to the constraint of nonnegative $\sigma_U(x,z)$ and $\sigma_V(x,z)$, where p_u is a n-vector of 1's and p is a vector of the probability weights p_{i2} or p_{i3} in (3.3)–(3.4). R code implementing the constrained kernel regression is provided in Appendix. In practice, the requirements of nonnegative $\sigma_U(x,z)$ and $\sigma_V(x,z)$ in (2.3)–(2.4) implie the constraints of $\widehat{r}_2(x,z) \geq \left(\frac{\pi-2}{\pi}\right) \widehat{\sigma}_U^2(x,z)$ and $\widehat{r}_3(x,z) \leq 0$ in the regression of $r_2(x,z)$ and $r_3(x,z)$ in (3.3)–(3.4) respectively. Since the constraint on $r_2(x,z)$ depends on the estimates of $r_3(x,z)$, the regression of $r_3(x,z)$ should be done before the regression of $r_2(x,z)$. Plugging the constrained estimates $\widehat{r}_2(x,z)$ and $\widehat{r}_3(x,z)$ back into (2.3)–(2.4) and solving it for $\sigma_U(x,z)$ and $\sigma_V(x,z)$ we get

$$\widehat{\sigma}_U(x,z) = \left(\sqrt{\frac{\pi}{2}} \left(\frac{\pi}{\pi-4}\right) \widehat{r}_3(x,z)\right)^{\frac{1}{3}} \tag{2.3.6}$$

and

$$\widehat{\sigma}_V(x,z) = \left(\widehat{r}_2(x,z) - \left(\frac{\pi-2}{\pi}\right)\widehat{\sigma}_U^2(x,z)\right)^{\frac{1}{2}}. \tag{2.3.7}$$

Using the results in (3.6)–(3.7), we can derive consistent estimates of the inefficiency term $\mu_U(x,z)$ and the stochastic frontier $m(x,z)$ as

$$\widehat{\mu}_U(x,z) = \sqrt{\frac{\pi}{2}}\widehat{\sigma}_U(x,z) \tag{2.3.8}$$

and

$$\widehat{m}(x,z) = \widehat{r}_1(x,z) + \widehat{\mu}_U(x,z). \tag{2.3.9}$$

Elasticity is a popular measurement in economic since it has the advantage of being a unit free ratio. There is a convient way to derive the elasticity of the mean inefficiency using the third conditional moment estimate $\widehat{r}_3(x,z)$ in (3.4). The partial elasticity of the mean inefficiency $\mu_U(x,z)$ with respect to X_i, for $i = 1, 2,..., p$, and p represents the number of explanatory variables, is given by

$$\xi_{X_i} = \frac{\partial \mu_U(x,z)}{\partial X_i}\frac{X_i}{\mu_U(x,z)} = \frac{1}{3}\frac{\partial E(\varepsilon^3|x,z)}{\partial X_i}\frac{X_i}{E(\varepsilon^3|x,z)}. \tag{2.3.10}$$

A consistent nonparametric estimate of ξ_{X_i} can be obtained by replacing the true moment $E(\varepsilon^3|x,z)$ and its partial derivative $\frac{\partial E(\varepsilon^3|x,z)}{\partial X_i}$ with their nonparametric estimates in (3.10). In other words, we have

$$\widehat{\xi}_{X_i} = \frac{1}{3}\frac{\partial \widehat{r}_3(x,z)}{\partial X_i}\frac{X_i}{\widehat{r}_3(x,z)}, \tag{2.3.11}$$

where $\widehat{r}_3(x,z)$ and $\frac{\partial \widehat{r}_3(x,z)}{\partial X_i}$ can be derived from the shape constrained nonparametric regression in (3.4).

All in all, the semiparametric MOLS approach has the advantage of relaxing assumptions than its competing peers. In the estimation process through (3.1) to (3.7), the only parametric assumption is that the density of $U|x, z$ belongs to a one-parameter scale family. That is, there is only one parameter needs to be identified for the inefficiency term $U|x, z$, while the rest of the estimation process retains non-parametric. It is worth noting that the expressions in (2.3)–(2.4) are on the condition that $U|X = x, Z = z \sim |N(0, \sigma_U^2(x, z))|$. If the inefficiency term U follows other distribution, the expressions in (2.3)–(2.4) need to be revised accordingly. Besides the half-normal distribution used in this paper, the available options of the distribution of U include the exponential distribution, the gamma distribution with fixed shape parameter and the Weibull distribution with fixed shape parameter. The choice of the distribution of U may vary in different empirical questions with flexibility.

2.4 Data and Variable Specification

The data are collected from the sub-industry of semiconductors in Compustat database over the period of 20 years 1999–2018. Since the semiconductor industry is highly globalized, I combine data from both the Compustat North America database and the Compustat Global database to cover companies in the whole industry. I exclude photovoltaic producers, liquid crystal display manufactures and light-emitting diodes manufactures from the dataset, limiting the sample within only IC manufactures. Under such restriction, the sample includes 5136 observations from 470 unique companies in 1999–2018. I specify three inputs (labor, measured by the number of employees (X_1); capital investment, measured by the value of property, plant, and equipment (PP&E) (X_2); and operating expense, measured by combining the cost of goods sold, R&D expenditure, and sales & marketing expenditure (X_3)), one output (revenue (Y_1)), and two environmental variables (business model (Z_1); and year (Z_2)).

Table 2.1 breaks down the 5136 observations in global semiconductor industry in 1999–2018 by business model. Over half of the companies are in fabless model because the barriers to entry are much lower for the asset light fabless companies than for the others in the semiconductor industry. Figure 2.1 splits the yearly data of the variables by business model. The labor intensive fabless companies focusing on chip design and adoping asset light strategy have relatively low spending on PP&E. The capital intensive foundries and OSATs focusing on fabrication and depending heavily on CAPEX for facility construction and equipment maintenance have high investments in PP&E but low operating expenditures. The vertical integrated IDMs which are both labor intensive and capital intensive have high costs for all the input factors, with proportionally higher output revenues.

Table 2.2 gives summary statistics for the continuous variables in 1999–2018 pooled data. All the values in Table 2.2 are in log form because log transformation is a popular method to deal with the skewness such as the cases in Cobb-Douglas and translog models, and log transformation has the advantage of avoiding the singular matrix problem which may cause the shape constrained kernel regressions in (3.3)–(3.4) into trouble. In order to set up a criterion for comparing data from different years, the values of X_2, X_3, and Y are adjusted to 2018 US dollar by GDP deflator before log transformation. Besides the explanatory variables, it is very flexible to add environmental variables in the nonparametric MOLS approach, no matter whether the environmental variables are continuous variables or discrete variables. The environmental variable Z_1 which identifies business model is an ordinal variable with three categories. The firms fall into one of the three categories of Z_1 which are ordered by their intensity of labor and capital. The first category contains fabless companies which are labor intensive for chip design. The second category contains IDMs which are both labor intensive and capital intensive. The thrid category contains both foundries and OSATs which share the same feature of capital intensive. The environmental variable Z_2 which identifies year is also an ordinal variable because the observations in adjacent years are expected to have stronger correlations. Hence the dataset is an unbalanced panel with mixed variables. Li and Racine (2007) and Racine and Nie (2017) discuss the nonparametric regression methods for mixed data in detail.

2.5 Empirical Results

It is well known that the advantage of flexible functional form in nonparametric regression has a price of interpretation difficulty. A convenient way to interpret the estimation results in nonparametric regression is by plotting and graphing. Figures 2.2–2.3 present the estimates of σ_U and σ_V with respect to the explanatary variables. As shown in (3.8) the difference between σ_U and the conditional expectation of the inefficiency term μ_U is a constant $\frac{\pi}{2}$ for all points, so that the plots of $\widehat{\sigma}_U$ in Figure 2.2 directly depict the plots of $\widehat{\mu}_U$. The fitted lines of $\widehat{\sigma}_U$ in the top two rows of panels in Figure 2.2 show that for the whole semiconductor industry and especially for the fabless companies there are steep downward trends of the inefficiency term when the inputs are at low levels. Since the low levels of inputs probably imply that the related firms are mainly small and medium-sized enterprises (SMEs), the sharp downward trends of $\widehat{\sigma}_U$ provide evidence of economies of scale for the SMEs in semiconductor industry. However, the downward trends of $\widehat{\sigma}_U$ slow down with the expansion of company scale, which indicates that the economies of scale for the SMEs follow the law of diminishing marginal returns. Furthermore, the U-shaped curves of the fitted $\widehat{\sigma}_U$ demonstrate that if the expansion of company scale passes a turning point there are areas of diseconomies of scale for the large fabless companies. This phenomenon is rooted in the fact that more than half of the companies in the semiconductor industry are operating in fabless model and most of the fabless companies are SMEs, as shown in Table 2.1 and Figure 2.1. On the contrary, such phenomenon has not been observed either by the capital intensive foundries and OSATs or by the vertical integrated IDMs. The lower and flatter curves of fitted $\widehat{\sigma}_U$ in the last two rows of panels in Figure 2.2 indicate that companies with the feature of capital intensive are protected by the high barriers to entry and hence are operating at higher and stabler levels of efficiency.

Figure 2.3 presents the estimates of σ_V in the same format as the estimates of σ_U in Figure 2.2 in order to make comparisons. The fitted lines of $\widehat{\sigma}_V$ in the panels in Figure 2.3 illustrate downward trends of $\widehat{\sigma}_V$ for the SMEs in the whole semiconductor industry and in each of the business models. These downward trends of $\widehat{\sigma}_V$, which measure the operational risk or uncertainty, are distinct but with much larger variance comparing with the downward trends of $\widehat{\sigma}_U$ in Figure 2.2. As the expansion of company scale, which are represented by the increasing of inputs in the panels in Figure 2.3, the fitted $\widehat{\sigma}_V$ diminish gradually. The lower $\widehat{\sigma}_V$ for the large-scale companies reveal that they are facing less uncertainty comparing with the SMEs.

It can be deduced from the estimates of σ_U and σ_V in Figures 2.2–2.3 that as the semiconductor companies scaling up, they probably tend to reduce both the risk of uncertainty, which is captured by σ_V, and the risk of inefficiency, which is captured by σ_U. This finding gives corroborative evidence why there are so many merger and acquisition (M&A) in the semiconductor industry, because M&A provides a shortcut for scaling up. The only exception is for the fabless companies which are constrained largely by skilled engineers for chip design but are not heavily relying on capital investment for equipments. It is perspicuous that managing a large team of R&D engineers to work together efficiently is likely a more challenging task than managing a large-scale of precision equipments to operate in parallel efficiently. So that for the large-scale semiconductor companies, it is more likely to observe that the labor intensive fabless companies are operating in areas of diseconomies of scale, and the capital intensive foundries and OSATs, and the vertical integrated IDMs are operating in areas of economies of scale.

Figure 2.4 provides the fitted estimates of the elasticity of μ_U with respect to the explanatary variables and by difference business models. With respect to the input of labor, the elasticity ξ_{X_1} starts at a small negative value with lowest labor level, then decreases into further negative territory but finally goes back to a small negative value at the right end of range for the whole industry and the fabless firms. These U-shaped curves of ξ_{X_1} imply that there are economices of scale for the SMEs, similar as the corresponding $\widehat{\sigma}_U$ curves shown in Figure 2.2. However, for the IDMs and especially for the foundries and OSATs which are not labor-intensive, the operating efficiencies are not semsitive to the changes of labor inputs. With respect to the input of PP&E, the curves of ξ_{X_2} are also U-shaped but with a much smaller scale for the whole industry and the fabless firms, which imply that there are economies of scale for the SMEs. However, the values of ξ_{X_2} turn to positive at the right portion of range for the IDMs, foundries and OSATs, which imply there are diseconomies of scale for the enterprises above designated size. Exceptions are for the fabless companies, which are not capital intensive so that the values of ξ_{X_2} never go into the positive territory. With respect to the input of operating expenses, the elasticity ξ_{X_3} are positive with inverted U-shaped curves. Since the operating expenses are combinations of COGS, R&D expenditures, and sales & marketing expenditures, higher operating expenses are positively correlated with high-end or upgraded products. The inverted U-shaped curves of ξ_{X_3} imply that it is more risky for the SMEs than the large-scale companies to dedicate to high-end products. This phenomenon is especially distinguishing for the fabless companies, where the high sunk costs of R&D cause the niche companies to be more fragile in the rapidly evolving semiconductor industry. In general, the niche fabless firms are more elastic because of the asset-light business model, while the foundries and OSATs are inelastic because of the CAPEX constraints for fabrication and the IDMs are between the two extreme cases.

The estimates of the annual mean σ_U and σ_V in 1999–2018 are shown in Tables 2.3–2.4 and are visualized in Figure 2.5. The annual mean of σ_U are statistically significant in each year for all the business models. On the other hand, the annual mean of σ_V are statistically insignificant for any of the business models in any of the years, which is consistent with the assumption that the error term σ_V should be a randomized white noise. The annual mean of σ_U are much higher for the fabless companies than for the other business models, which is compatible with the findings in Figures 2.2–2.4. The fabless companies, most of which are SMEs, are inherently more vulnerable than the large-scale market giants in the thoroughly competitive semiconductor industry. Besides, it is undoubted that the economies of scale which is essential for new technologies and products to be popularized with affordable costs, are more likely to exist in the capital intensive business models than in the labor intensive business models, despite the cyclical fluctuations in the semiconductor industry.

2.6 Summary and Conclusions

It is an intimidating subject to compare the operational efficiency between the vertical integrated IDM model and the specialized fabless-foundry model in the semiconductor industry where technologies are everchanging. Nevertheless, the economies of scale is essential for sustainable development in the fast growing semiconductor industry, in spite of the business models. Though the genesis fabless-foundry model reduces the barriers to entry and helps small businesses prosper and flourish, CAPEX is still a causal element for companies in the semiconductor industry to be ahead of the pack. Facing the high uncertainty of commercial success due to dynamical bottlenecks and fierce competition twisting to the ups and downs in the global economic cycle, the large scale companies most of whom are capital-intensive, are operating with more steady growth and less uncertainty than the SMEs in the semiconductor industry. Since M&A is considered one of the magic tricks to take advantage of economies of scale, increasing trends toward consolidation are observed in the semiconductor industry. Consequently, the vertical integrated IDMs which have accumulated years of large-scale and comprehensive technology play the dominant role in the semiconductor industry and will continue to be at the forefront of innovation.

However, the distinction between IDM model and fabless-foundry model is fading away. In a diversified market mixing up variety demands and accelerated technology iteration, the complementarity between IDMs and fabless-foundry firms requires the coexistence of the increasing specialization at given technology node and the further concentration of the value chain. Several IDMs contract with other companies to manufacture chips while performing all other remaining tasks internally (e.g., see Li et al., 2011). This developing new trend, which is commonly called fablite in the semiconductor industry, is beyond the discussion of this paper and is left for further research.

Table 2.1: Number of observations by business model

Year	Number of Companies				
	All	Fabless	IDM	Foundry	OSAT
1999	125	68	38	10	9
2000	149	81	43	10	15
2001	155	83	46	10	16
2002	213	121	48	17	27
2003	241	143	49	19	30
2004	264	159	54	21	30
2005	260	162	54	17	27
2006	267	161	56	20	30
2007	269	163	52	21	33
2008	278	172	51	20	35
2009	290	180	53	21	36
2010	300	180	59	23	38
2011	298	177	60	22	39
2012	301	180	61	22	38
2013	313	183	65	24	41
2014	302	172	62	25	43
2015	288	163	59	24	42
2016	283	162	54	23	44
2017	275	156	51	23	45
2018	265	151	48	22	44
Obs.	5,136	3,017	1,063	394	662
Uniq. Obs.	470	288	83	36	63

NOTE. Obs. denotes the total number of observations.
Uniq. Obs. denotes the unique number of companies.

Table 2.2: Summary statistics for 1999–2018 pooled data

Variable	Min	Q1	Median	Mean	Q3	Max
Y	1.030	10.764	11.994	12.036	13.242	18.076
X_1	0.000	5..075	6.185	6.409	7.606	11.586
X_2	1.686	8.710	10.232	10.374	12.069	17.707
X_3	6.540	9.387	10.619	10.719	11.842	16.902
Obs.			5,136			
Uniq. Obs.			470			

NOTE. All the continuous variables are in log form.
The unit of X_1 is the number of employees before log transform.
The units of X_2, X_3, and Y are US$ thousand before log transform.
The values of X_2, X_3, and Y are adjusted to 2018 US$ before log transform.

Table 2.3: Estimates of the annual mean σ_U

	— Semi —		— fabless —		— IDM —		foundry+OSAT	
Year	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1999	0.367	0.136	0.470	0.085	0.230	0.068	0.272	0.046
2000	0.351	0.148	0.451	0.121	0.220	0.068	0.255	0.067
2001	0.333	0.117	0.411	0.098	0.228	0.055	0.267	0.042
2002	0.352	0.118	0.430	0.092	0.230	0.057	0.271	0.039
2003	0.338	0.109	0.401	0.092	0.230	0.052	0.261	0.037
2004	0.326	0.109	0.384	0.077	0.230	0.053	0.250	0.042
2005	0.327	0.112	0.374	0.115	0.245	0.044	0.254	0.026
2006	0.287	0.085	0.317	0.091	0.235	0.052	0.250	0.045
2007	0.287	0.081	0.309	0.093	0.249	0.036	0.256	0.046
2008	0.300	0.093	0.322	0.110	0.252	0.033	0.275	0.029
2009	0.305	0.074	0.327	0.083	0.263	0.033	0.275	0.032
2010	0.302	0.083	0.325	0.097	0.267	0.033	0.272	0.035
2011	0.323	0.095	0.353	0.111	0.281	0.035	0.278	0.038
2012	0.340	0.083	0.378	0.083	0.285	0.040	0.283	0.037
2013	0.341	0.084	0.385	0.079	0.274	0.041	0.283	0.040
2014	0.346	0.092	0.395	0.087	0.271	0.053	0.290	0.041
2015	0.361	0.102	0.420	0.089	0.271	0.065	0.295	0.041
2016	0.362	0.103	0.419	0.093	0.261	0.065	0.306	0.040
2017	0.362	0.108	0.417	0.105	0.259	0.065	0.314	0.040
2018	0.359	0.116	0.410	0.118	0.257	0.076	0.315	0.049

NOTE. Semi denotes the whole semiconductor industry.

Table 2.4: Estimates of the annual mean σ_V

	— Semi —		— fabless —		— IDM —		foundry+OSAT	
Year	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1999	0.473	0.468	0.609	0.455	0.329	0.528	0.274	0.105
2000	0.556	0.740	0.755	0.938	0.326	0.282	0.303	0.115
2001	0.511	0.449	0.673	0.555	0.318	0.133	0.331	0.103
2002	0.566	0.563	0.761	0.676	0.309	0.150	0.309	0.118
2003	0.595	0.517	0.800	0.575	0.293	0.167	0.300	0.156
2004	0.643	0.829	0.861	0.973	0.286	0.179	0.342	0.459
2005	0.636	0.828	0.857	0.980	0.277	0.167	0.267	0.118
2006	0.672	1.064	0.908	1.301	0.302	0.240	0.325	0.306
2007	0.649	1.101	0.860	1.362	0.315	0.233	0.334	0.245
2008	0.710	1.176	0.914	1.401	0.335	0.295	0.423	0.669
2009	0.792	1.439	1.067	1.759	0.356	0.337	0.330	0.229
2010	0.758	1.153	0.989	1.337	0.356	0.305	0.466	0.891
2011	0.879	1.585	1.209	1.901	0.362	0.295	0.431	0.972
2012	0.822	1.334	1.113	1.590	0.380	0.423	0.395	0.735
2013	0.945	1.925	1.378	2.408	0.394	0.522	0.274	0.115
2014	0.829	1.842	1.191	2.335	0.422	0.764	0.283	0.106
2015	0.894	1.880	1.315	2.365	0.419	0.831	0.280	0.105
2016	0.922	2.058	1.402	2.619	0.275	0.152	0.283	0.120
2017	0.955	2.697	1.476	3.494	0.276	0.142	0.270	0.151
2018	0.924	2.439	1.417	3.139	0.296	0.359	0.253	0.156

NOTE. Semi denotes the whole semiconductor industry.

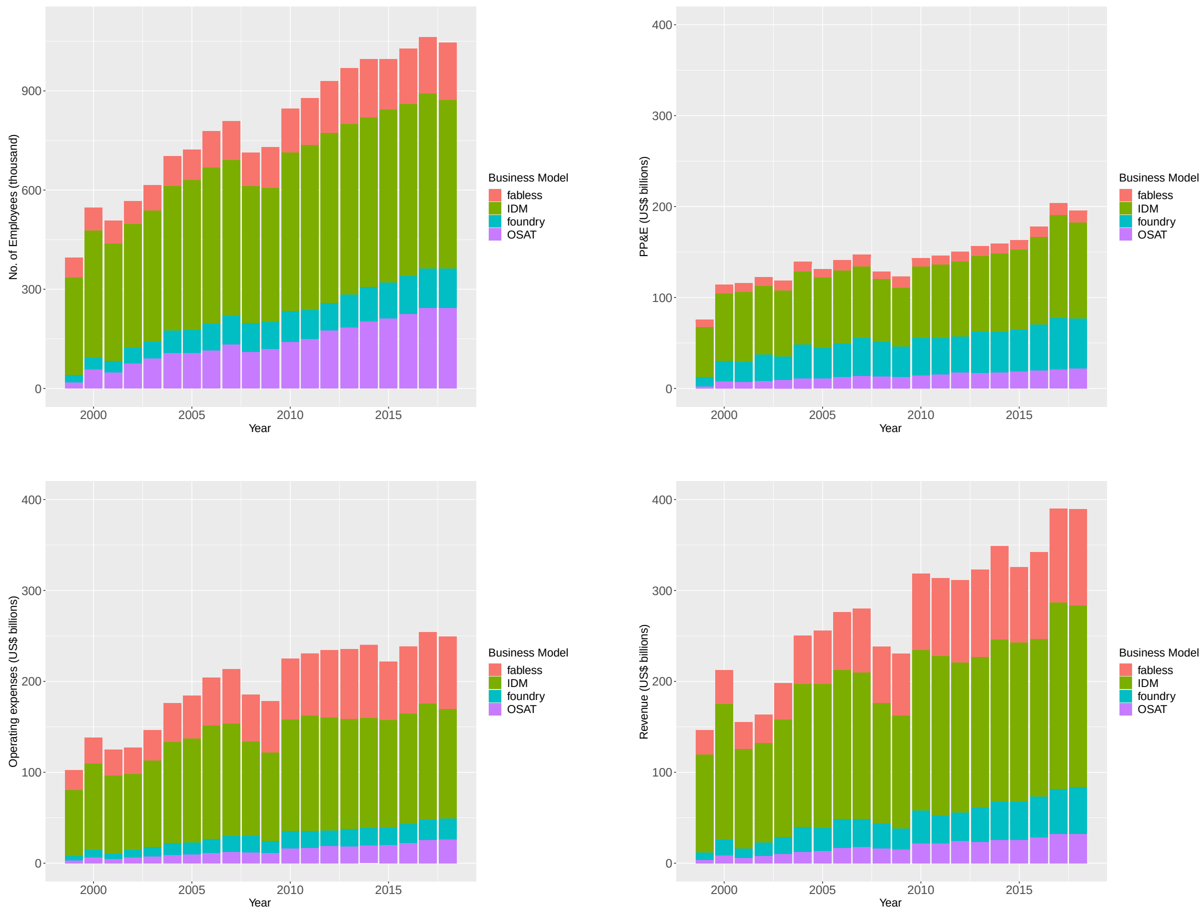

Figure 2.1: Annual data break down by business model

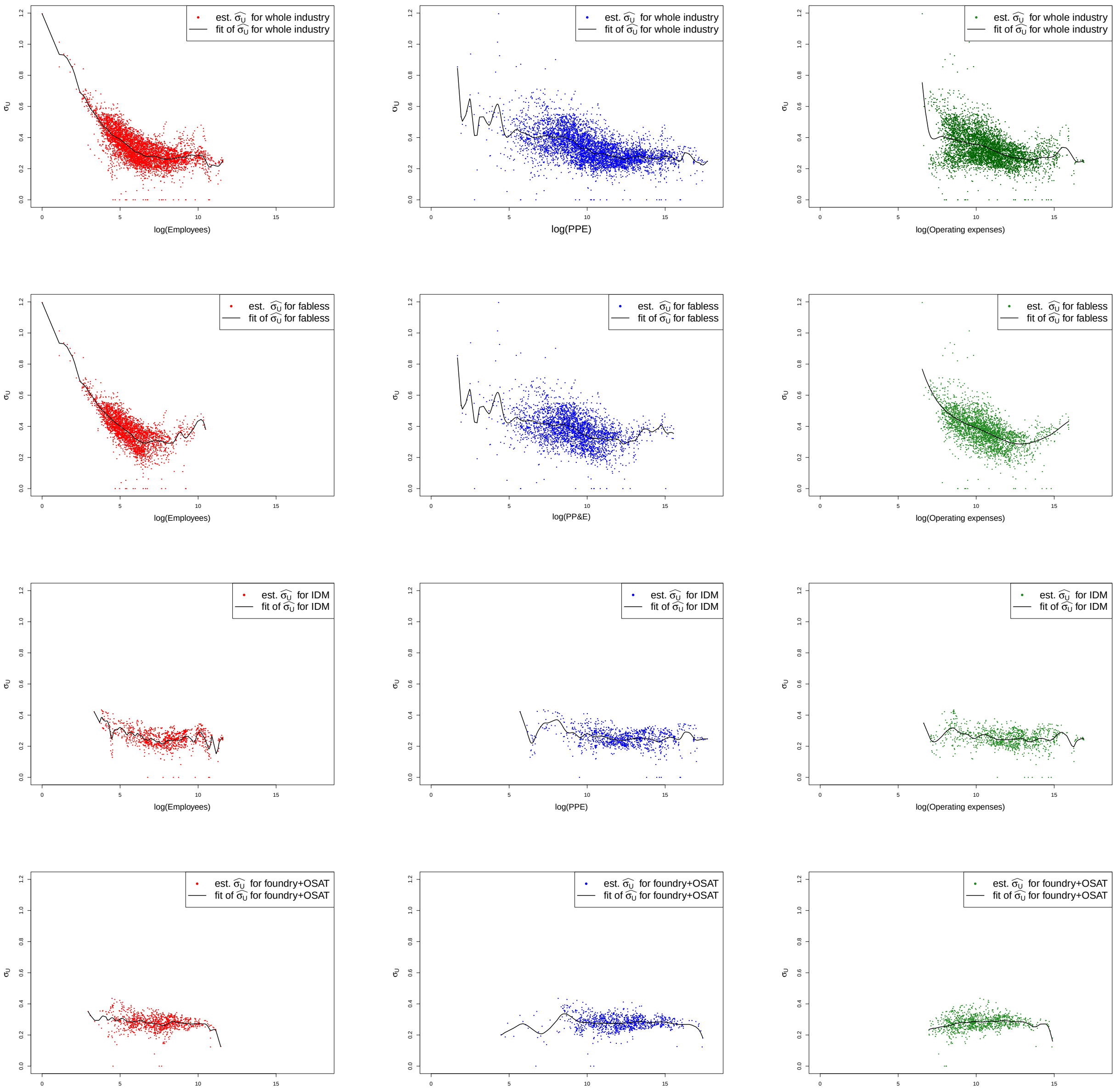

Figure 2.2: Estimates of σ_U with respect to X_i

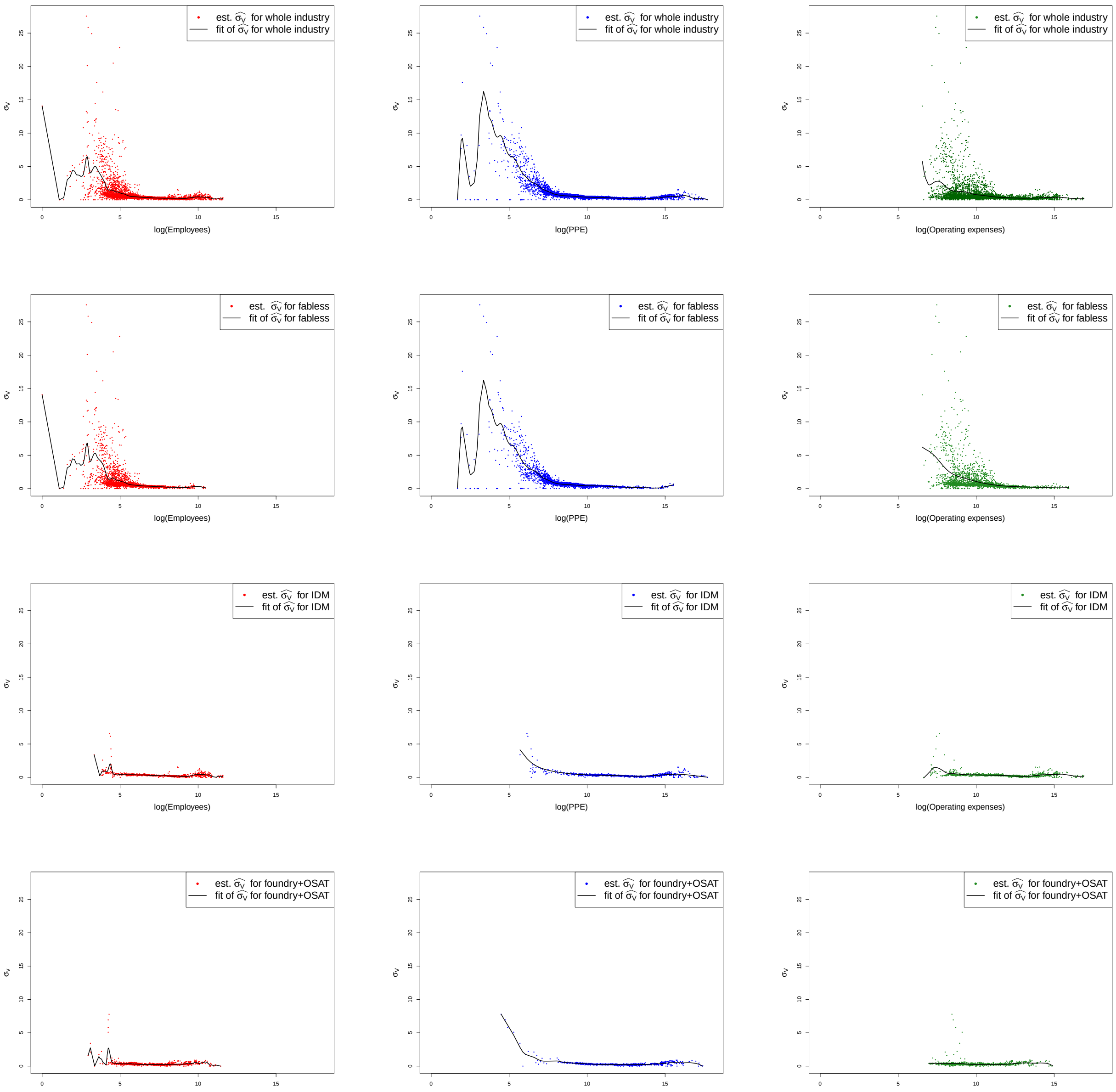

Figure 2.3: Estimates of σ_V with respect to X_i

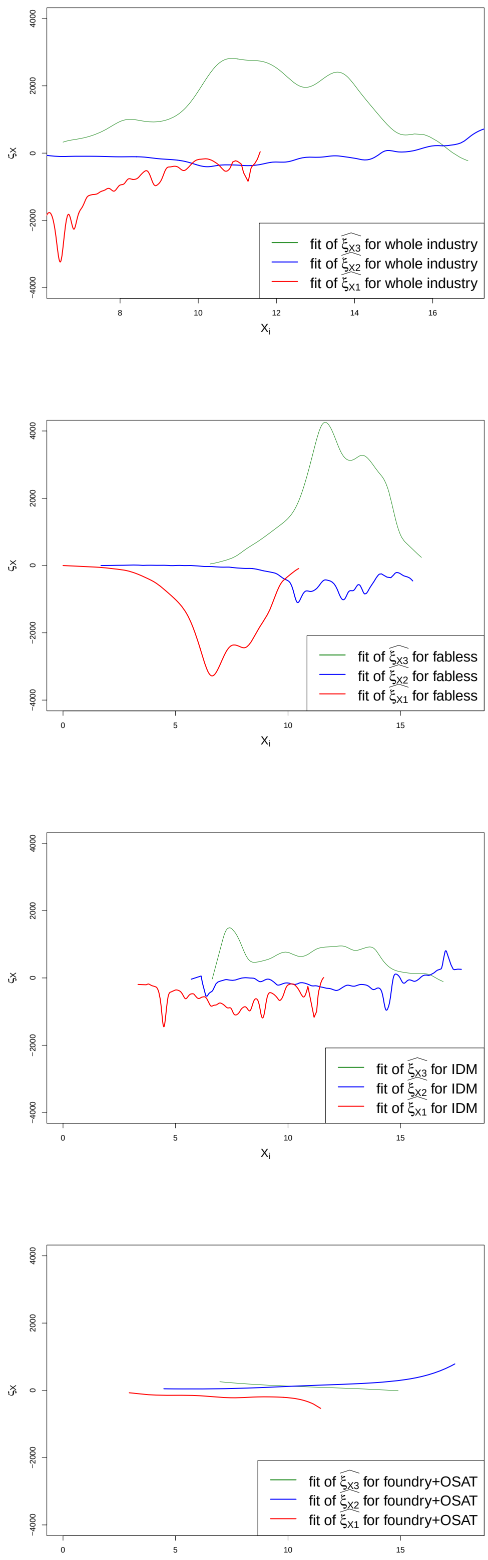

Figure 2.4: Fit of elasticities ξ_i with respect to X_i

Chapter 3

Capital Expenditure and Efficiency from Vertical Integration: A Nonparametric Frontier Estimation For Global Semiconductor Industry

3.1 Introduction

Semiconductors, also known as integrated circuits (ICs) or chips, are the brains of virtually all modern electronics. Since the release of the first commercial ICs in the 1960s, the semiconductor industry has been a driving force for growth of the electronics market. Prior to the 1980s, the semiconductor industry was dominated by integrated device manufacturers (IDMs), which perform all of the production stages, including research and design (R&D), front-end fabrication, and back-end assembly and test (A&T) in-house. As semiconductors with ever-expanding complexity approach the limits of Moore's Law (e.g., see Mack, 2011 and Flamm, 2017), the expenses of manufacturing a leading-edge chip have become prohibitive for all but a few IC suppliers. These increasing costs have kept the industry on its toes and give birth to the fabless-foundry business model in the mid-1980s. In this new business model, Fabless companies are dedicated to IC design and sales, and partner with pure-play foundries for front-end fabrication as well as a third group of companies for back-end outsourced semiconductor assembly and test (OSAT). The fabless-foundry business model has significantly changed the structure of the semiconductor industry over the last few decades.

The structural changes by specialization in the semiconductor value chain is a topic of wide interest (e.g., see Macher et al., 2007, Adner and Kapoor, 2010 and Sarma and Sun, 2017). Although the semiconductor industry is both technology-intensive and capital-intensive, much research on the topic of the structural change in the semiconductor industry emphasizes the impact of technological evolution (e.g., see Macher, 2006, Kapoor and Adner, 2012 and Hwang and Choung, 2014), while the impact of capital investments has not been discussed adequately. At the same time, vertical disintegration in the semiconductor value chain is accompanied by the trend of industry globalization (e.g., see Brown et al., 2005). Recent developments in nonparametric frontier estimation (e.g., see the survey by Simar and Wilson, 2015) provide tools to analyze the operating efficiencies in the semiconductor industry under various types of constraints such as capital investments and business model. This paper aims to use a conditional nonparametric frontier approach to shed light on disentangling the impact of capital investments and comparing the technical efficiencies between the IDMs and the vertical disintegrated fabless and foundry firms in the highly globalized semiconductor industry.

Before jumping into the details of the nonparametric frontier estimation, it is worth tracing the origin and evolution of the fabless-foundry business model in the semiconductor industry. A milestone of the vertical disintegration in the semiconductor industry is the establishment of Taiwan Semiconductor Manufacturing Company (TSMC) in 1987. Committed to be long term non-competitive partner with the fabless firms, TSMC is the first and nowaday the largest pure-play foundry worldwide that dedicates to wafer fabrication (e.g., see Hsieh et al., 2002). The drastically reduced burden of capital expenditures (CAPEX) and the reduced barriers to entry by vertical specialization ensure domination of new markets by the fabless firms. The collaboration between the asset-light fabless and the pure-play foundry also provides stronger protection of intellectual property (IP) rights when the fabless firms pass on their design blueprints to pure-play foundries, which earlier are exposed to the threats of replication and IP theft when fabless firms' ICs are manufactured by their rival IDMs (e.g., see Sarma and Sun, 2017). The entry of new fabless companies, most of which are spinoffs from industry incumbents, spur innovation and impel the diversification of products in various applications. Since the 1990s, fabless firms have substantial shares or even dominated in most of the fastest growing market segments (e.g., see Balconi and Fontana, 2011). Nevertheless, despite a trend toward vertical specialization driven by the entry of fabless firms, the vertical integrated IDMs have continued to persist and coexist with the fabless entrants in the semiconductor industry.

In the semiconductor industry, factors that determine production costs and operating efficiencies vary substantially across device types and business models. For leading-edge products such as microprocessors, manufacturing is capital intensive, with R&D and equipment expenditures rising steadily. The ever-increasing costs of building advanced fabrication facilities and the difficulties that arise from slowing development in node technology set high barriers to entry and favor the success of large IDMs, such as Intel, STMicroelectronics and Texas Instruments, which have the ability to forge ahead with innovative and expensive technologies that are needed to take ICs to the next level. For front-end fabrication and back-end A&T procedures, the major challenges are the heavy CAPEX for cleanroom and costly equipments so that both the foundries and OSATs seek to optimize productivity by serving many fabless companies to achieve high capacity utilization. In comparison, the fabless companies, many of which are niche startups, get rid of the burden in setting up, maintaining, and upgrading fabrication facilities and focus on R&D to compete with the IDM giants. Hence CAPEX are indispensable factors in the cost-benefit analysis between the IDM business model and the fabless-foundry business model in the semiconductor industry.

There has been a long-lasting debate on which business model is operating more efficiently or is more likely to dominate the semiconductor industry. On one side, Monteverde (1995) and Dibiaggio (2007) credit the efficiency of IDMs to the internalization of transaction costs. Ernst (2005), Macher (2006) and Kapoor and Adner (2012) hold the knowledge-based view that the IDMs achieve performance advantages when technological developments involve complex problems. On the other side, Li et al. (2011) show that foundries are becoming technology transferors rather than merely manufacturing capacity providers in the semiconductor industry value chain. Kapoor (2013) proposes and finds that the incumbents who persist with vertical integration increase their emphasis on systemic innovations. Besides the examples shown above that focus on analyzing the impact of technology evolution in the semiconductor industry, this paper plans to emphasize the feature of capital intensive in the semiconductor industry and focus on analyzing the impact of CAPEX and business model to the operating efficiency.

Taking advantage of a flexible functional form, data envelopment analysis (DEA) is the most popular approach for efficiency estimation. There are rich records for performance evaluation in the semiconductor industry using the DEA approach. For instance, Kozmetsky and Yue (1998) examine cost efficiency of 56 IC companies worldwide and show that US, Japanese, South Korean and Taiwanese IC companies have become the major participants in global semiconductor industry in the early 1990s. Lu and Hung (2010) compare the managerial performance efficiency of 48 leading vertically disintegrated firms in Taiwan's IC value chain and note that fabless companies perform better than foundries and OSATs. Jang et al. (2016) measure the cumulative change in R&D efficiency of 49 global leading fabless companies and note that during the period 2007-2013 the overall R&D efficiency decline slightly. Li et al. (2019) explore 64 major Chinese enterprises in the semiconductor industry and find that low levels of scale efficiency is the most significant factor limiting future improvements to innovation efficiency. One common problem of these studies, among others, such as Lu et al. (2013), Hung et al. (2014), Hsu (2015) and Tsai et al. (2017), is the slow convergence rate of the nonparametric DEA estimator accompany with the increasing numbers of input and output dimensions (e.g., see Wilson, 2018).

The issue of slow convergence rate in DEA estimation may become severe if the observations are restricted to a small number either by geographic boundary or by business model boundary. For example, the researches of Wu et al. (2006), Lu et al. (2010), and Kuo and Yang (2012) use a small number of 38-39 companies to evaluate the performance of the fabless corporations in Taiwan, while in some extreme cases, such as Hung and Lu (2008), Liu and Wang (2008), Chen and Chen (2011) and Lin et al. (2019), the studies contain only 10-25 companies, which may lead to unconvincing results. It also explains why the Free Disposal Hull (FDH) estimator, which is known as an unbiased substitution of the DEA estimator with much slower convergence rate, is not widely used in empirical articles, as slower convergence rate places greater demand for the sample size in nonparametric estimation (e.g., see Simar and Wilson, 2015). One solution to mitigate the slow convergence rate problem in nonparametric efficiency estimation is to increase the number of observations by considering the deeply globalized semiconductor industry as a whole. Furthermore, the heterogeneity by business model in the combined global semiconductor industry can be handled by the conditional efficiency estimators (e.g., see Daraio and Simar, 2007), while the heterogeneity by CAPEX can be treated specially as a fixed input variable by the directional distance estimator (e.g., see Daraio et al., 2020). This paper aims to contribute to the investigation of the impact of business model and CAPEX in the highly globalized semiconductor industry by means of a conditional nonparametric frontier approach.

The paper is organized as follows. Section 3.2 gives an overview of the nonparametric frontier framework and the available estimators for empirical works. Section 3.3 discusses the diagnostics and test statistics for choosing suitable estimator in this research. Section 3.4 describes the dataset of global semiconductor industry and defines the production function with environmental variables. Section 3.5 presents the empirical results of the effect of capital investment and business model in the semiconductor industry. Section 3.6 concludes.

3.2 The Statistical Model

The economic theory of efficiency in production can be traced to the ideas of Koopmans (1951), Debreu (1951) and Farrel (1957). Consider a production process in which p inputs are used to produce q outputs. The production set

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\} \tag{3.2.1}$$

describes the set of attainable combinations of inputs and outputs. The efficiency score of a particular production plan (x, y) is then determined by the distance from (x, y) to the efficient frontier or boundary of Ψ, which can be defined as

$$\Psi^{\partial} = \{(x, y) \in \Psi \mid (\gamma^{-1}x, \gamma y) \notin \Psi \quad \text{for all } \gamma > 1\}. \tag{3.2.2}$$

There are four kinds of commonly used efficiency measures based on different directions in which the distance is calcuated. The most widely used radial measures are the input- and output-oriented Debreu-Farrel measures. Färe et al. (1985) introduce the hyperbolic measure

$$\gamma(x, y \mid \Psi) = \inf\{\gamma \mid (\gamma x, \gamma^{-1}y) \in \Psi\} \tag{3.2.3}$$

as an alternative to selecting either an input- or output-oriented Debreu-Farrel measures, where input and output quantities are adjusted simultaneously to reach the boundary Ψ^{∂} along a hyperbolic path. Note that these three kinds of efficiency measures are all radial measures that allow for only nonnegative values of inputs and outputs.

Chambers et al. (1998) propose an additive measure of the technical efficiency which considers the feasible quantities to be added to a unit's output and simultaneously subtracted from its input and is known as the directional distance measure. The directional distance measure is given by

$$\beta(x, y \mid d_x, d_y, \Psi) = \sup\{\beta \mid (x - \beta d_x, y + \beta d_y) \in \Psi\}, \tag{3.2.4}$$

which projects the input-output vector $(x,\, y)$ onto the technology in a specified direction $(-d_x,\, d_y)$ and allows for negative values of x and y. The directional distance measure $\beta(x, y \mid d_x, d_y, \Psi)$ nests both input- and output-oriented Debreu-Farrel measures in (2.4) as special cases by setting the direction vector (d_x, d_y) as $(x,0)$ and $(0,y)$ respectively. The flexibility of the directional distance measure also comes from the choices of the directions d_x and d_y that some directions (but not all) can be set equal to zero to represent non-discretionary inputs or outputs (e.g., see Simar and Vanhems, 2012). This feature of the directional distance measure can be used to represent the impact of a kind of fixed input or output variables, such as CAPEX which can be categorized into input variables of production but are not under the direct control of the manager at least in the short run.

In real-world research problems, the attainable set Ψ is unobserved. Nonparametric methods such as FDH and DEA are developed and widely applied to estimate the unobservable production set Ψ. Using only the free disposability assumption, Deprin et al. (1984) define the FDH estimator $\widehat{\Psi}_{\mathrm{FDH}}$ as

$$\widehat{\Psi}_{\mathrm{FDH}} = \bigcup_{X_i, Y_i \in \mathcal{S}_n} \{(x, y) \in {\mathbb{R}_+}^{p+q} \mid x \geq X_i, y \leq Y_i\}, \tag{3.2.5}$$

where $\mathcal{S}_n = \{(X_i, Y_i)\}$ denote a random sample of n pairs of inputs and outputs. FDH estimators of $\widehat{\gamma}_{\mathrm{FDH}}(x, y \mid \Psi)$ and $\widehat{\beta}_{\mathrm{FDH}}(x, y \mid d_x, d_y, \Psi)$ are obtained by replacing Ψ with $\widehat{\Psi}_{\mathrm{FDH}}$ in (2.3)-(2.4) respectively. Baker et al. (1984) propose the varing-returns-to-scale DEA (VRS-DEA) estimator $\widehat{\Psi}_{\mathrm{VRS}}$, which is the convex hull of $\widehat{\Psi}_{\mathrm{FDH}}$ with the expression that

$$\widehat{\Psi}_{\mathrm{VRS}} = \{(x, y) \in {\mathbb{R}_+}^{p+q} \mid y \leq \boldsymbol{Y}\boldsymbol{\omega},\ x \geq \boldsymbol{X}\boldsymbol{\omega},\ \boldsymbol{i}_n^{'}\boldsymbol{\omega} = 1,\ \boldsymbol{\omega} \in {\mathbb{R}_+}^{p+q}\}, \tag{3.2.6}$$

where $\boldsymbol{X} = (X_1, ..., X_n)$, $\boldsymbol{Y} = (Y_1, ...Y_n)$ are $(p \times n)$ and $(q \times n)$ matrices of input and output vectors, $\boldsymbol{i}_n$ is an $(n \times 1)$ vector of ones, and $\boldsymbol{\omega}$ is a $(n \times 1)$ vector of weights.

The corresponding VRS-DEA estimators of $\widehat{\gamma}_{\mathrm{VRS}}(x, y \mid \Psi)$ and $\widehat{\beta}_{\mathrm{VRS}}(x, y \mid d_x, d_y, \Psi)$ are obtained by replacing Ψ with $\widehat{\Psi}_{\mathrm{VRS},n}$ in (2.3)-(2.4) respectively. There is another kind of DEA estimator proposed by Charnes et al. (1978), which is the convex cone of $\widehat{\Psi}_{\mathrm{FDH}}$ and is called the constant-returns-to-scale DEA (CRS-DEA) estimator. Since the CRS-DEA estimator relies on strong assumption that the returns to scale are everywhere constant, it is less widely used in empirical works. Therefore, the notation DEA is dedicated for VRS-DEA for the rest of the paper.

Beyond the combinations (X, Y) of inputs and outputs, there exist factors which are typically beyond control of the manager but may influence the production process. Denoted by $Z \in \mathbb{R}^r$, these factors are referred to as environmental factors which may reflect differences in ownership, business models, constraints of technology, regulatory, and so on. Conditions described by Z may or may not be independent of (X, Y), so that the unknown effect of Z must be estimated appropriately. Daraio and Simar (2005) propose a framework to investigate the joint behavior of (X, Y, Z) in probability terms. In detail, it defines the conditional attainable set by

$$\Psi^z = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y \text{ when } Z = z\}, \tag{3.2.7}$$

where $\Psi = \bigcup\limits_{z \in Z} \Psi^z$, so that $\Psi^z \subseteq \Psi$, for all $z \in Z$. Then the distribution of (X, Y) conditional on $Z = z$ is denoted by

$$H_{X,Y|Z}(x, y|z) = \mathrm{Prob}(X \leq x, Y \geq y \mid Z = z), \tag{3.2.8}$$

which gives the probability that a firm facing environmental conditions z will dominate the point (x, y). Given $Z = z$, the attainable set Ψ^z is the support of $H_{X,Y|Z}(x, y|z)$.

Introducing environmental factors into (2.3)-(2.4) extend the efficiency scores into their conditional counterparts. For example, the conditional hypobolic measure can be expressed as

$$\gamma(x, y \mid z) = \inf\{\gamma \mid H_{X,Y|Z}(\gamma x, \gamma^{-1} y \mid z) > 0\} \tag{3.2.9}$$

and the conditional directional distance measure can be expressed as

$$\beta(x, y \mid d_x, d_y, z) = \sup\{\beta \mid H_{X,Y|Z}(x - \beta d_x, y + \beta d_y \mid z) > 0\}. \tag{3.2.10}$$

Therefore, plugging a nonparametric estimator of $H_{X,Y|Z}(\cdot)$ from a sample $S_n = \{X_i, Y_i, Z_i \mid i = 1, ..., n\}$ into (2.9) or (2.10) can derive the estimation of the conditional efficiency scores accordingly. Such a nonparametric estimator of $H_{X,Y|Z}(\cdot)$ may be obtained by standard kernel smoothing, for example,

$$\widehat{H}_{X,Y|Z}(x, y \mid z) = \frac{\sum_i^n \boldsymbol{I}(X_i \leq x, Y_i \geq y) K(\frac{Z_i - z}{h})}{\sum_i^n K(\frac{Z_i - z}{h})}, \tag{3.2.11}$$

where $K(\cdot)$ is a kernel function with bounded support, h is a vector of bandwidths $h = (h_1, ..., h_r)$, and r is the number of environmental variables. It is well known that the selection of bandwidth h is of critical importance in kernal smoothing. Hall et al. (2004), Bădin et al. (2010, 2012), Jeong et al. (2010), and Li et al. (2013) propose the criterion of optimal bandwidth by least square cross-validation (LSCV).

There is a particular case, called separability condition in Simar and Wilson (2007), where Z has no impact on the boundaries of the Ψ^z and

$$\Psi^z = \Psi \tag{3.2.12}$$

for all $z \in Z$. Simar and Wilson (2007, 2011) emphasize that naive regression in a second-stage analysis may provide inconsistent estimation if the separability condition is violated. Daraio et al. (2018) demonstrate that standard central limit theorem results do not hold for means of nonparametric, conditional efficiency estimators and they provide new central limit theorems to construct a test of the separability condition. If the separability condition does not hold, Bădin et al. (2012, 2014) suggest a flexible nonparametric location-scale model

$$\gamma(X, Y \mid Z = z) = \mu(z) + \sigma(z)\varepsilon \tag{3.2.13}$$

in a second-stage regression, where $\mu(z)$ measures the average effect of z on the efficiency, and $\sigma(z)$ provides additional information on the dispersion of the efficiency distribution as a function of z. For empirical studies, Mastromarco and Simar (2015) use this two-step approach to gauge the effect of foreign direct investment and time on catching-up by developing countries. Cordero et al. (2017) use this two-stage approach to measure local government efficiency in Portugal. Toma (2020) uses this two-stage approach to explore the effect of size on Italian pharmaceutical firms.

3.3 Estimation and Inference

The tradeoff between FDH and DEA estimators for performance evaluation is nontrival. Simar and Wilson (2015) provide a survey of the nonparametric frontier models and summarize that under appropriate assumptions the FDH and DEA estimators converge to limiting distributions at rates $n^{\frac{1}{p+q}}$ and $n^{\frac{2}{p+q+1}}$ respectively. In either case, for a fixed sample size n, the convergence rate slows down with the increasing of dimensionality $(p + q)$, which increases the estimation error accordingly. This phenomenon is often referred to as the curse of dimensionality. Feasible approach to minimize such estimation error is either to increase the sample size n or to decrease the total dimensions of $(p+q)$. If the sample size n is limited to a small number by real world constraints, including the market scale and market scope in specific industries, geographical or political restrictions, and the high cost of data collection, dimension reduction may become an attractive solution to escape the curse of dimensionality.

Daraio and Simar (2007, pp. 148-150) propose using principal component analysis (PCA) for dimension reduction. Wilson (2018) explains why the curse of dimensionality is a serious problem in nonparametric efficiency estimation and introduces how to use PCA for a mapping $\Psi : R_+^{p+q} \longmapsto R_+^{1+1}$ for the radial measures. In detail, $(p \times n)$ matrix $\boldsymbol{X}$ and $(q \times n)$ matrix $\boldsymbol{Y}$ are transformed to $(1 \times n)$ matrices $\Lambda'_{x_1}\boldsymbol{X}$, $\Lambda'_{y_1}\boldsymbol{Y}$ by pre-multiplying the first eigenvector Λ_{x_1}, Λ_{y_1} of the moment matrices $\boldsymbol{XX'}$ and $\boldsymbol{YY'}$. Though it is not possible to give a theorem that precisely identifies situations where dimension reduction should be used, Wilson (2018) provides three diagnostics for empirical research. The first diagnostic is to compute the effective parametric sample size. Given a nonparametric estimator of n observations with the convergence rate of n^{κ} and a parametric estimator of m observations with the convergence rate of $m^{\frac{1}{2}}$, the effective parametric sample size of the nonparametric estimator can be derived as $m \approx \lfloor n^{2\kappa} \rceil$, where $\kappa = \frac{1}{p+q}$ for FDH estimator, $\kappa = \frac{2}{p+q+1}$ for DEA estimator, and $\lfloor a \rceil$ denotes the integer nearest a. Hence the criterion of judging the minimum sample size m in parametric estimation can be used as reference in nonparametric estimation.

A second diagnostic is to consider the proportion of n observations that yield efficiency scores equal to one. Since FDH estimator converges slower than DEA estimator, a robust diagnostic for the curse of dimensionality should use the FDH efficiency estimator. If more than 25%–50% of the observations yield efficiency scores equal to one, the estimation results are not convincing. A third diagnostic is to exam the ratios R_x and R_y of the largest eigenvalue of the moment matrices $\boldsymbol{X}\boldsymbol{X}'$ and $\boldsymbol{Y}\boldsymbol{Y}'$ to the corresponding sum of eigenvalues for $\boldsymbol{X}\boldsymbol{X}'$ and $\boldsymbol{Y}\boldsymbol{Y}'$ respectively. The ratios of R_x and R_y provide measures of how close the corresponding moment matrices are to rank-one. In practice, if $R_x = 0.9$, then the matrix with dimension reduction $\Lambda'_{x1}\boldsymbol{X}$ contains 90% of the independent linear information in the original matrix $\boldsymbol{X}$. In case there is evidence of excessive number of inputs or outputs provided by these diagnostics, Wilson (2018) proposes standardizing the matrices $\boldsymbol{X}$ or $\boldsymbol{Y}$ before PCA to ensure the inputs or outputs have the same scale.

After the diagnostics of dimension reduction, the choice between FDH estimator and DEA estimator can be decided by data driven hypothesis testing. Kneip et al. (2015, 2016) use new central limit theorems to construct a test of convexity for the tradeoff between $\widehat{\Psi}_{\text{FDH}}$ in (2.5) and $\widehat{\Psi}_{\text{VRS}}$ in (2.6). Note that the main difference of the test statistics by the new central limit theorems in Kneip et al. (2015, 2016) are the bias corrections construced by jackknife estimators (e.g., see Kneip et al., 2016, pp. 441-442). If the null hypothesis of convexity is rejected, the FDH estimator is the only consistent estimator. Alternatively, if the null hypothesis of convexity is not rejected, though it does not imply that the null is true, the DEA estimator may be the preferred estimator because of its faster convergence rate. Under the latter situation when the DEA estimator is preferred, the testing of return to scale (e.g., see Kneip et al., 2016, pp. 339-341) can be applied for the pros and cons between VRS-DEA and CRS-DEA estimators. However, the test of convexity proposed by Kneip et al. (2015, 2016) depends on randomly split the original sample into two independent subsamples for the calculation of the bias terms, which introduces ambiguity in practice. Simar and Wilson (2020, pp. 293-294) develop a generalized bootstrap algorithm that eliminates much of this ambiguity by repeating the random splits a large number of times, which can be used for either convexity test, return to scale test or separability test.

3.4 Data and Variable Specification

The data are collected from the Sub-Industry of Semiconductors in Compustat database. In order to treat the highly globalized semiconductor value chain as a whole, I combine data from both the Compustat North America database and the Compustat Global database to cover companies in the industry worldwide. As the semiconductor industry is famous for being a cyclical industry (e.g., see Tan and Mathews, 2010), I gather 20 years of data between 1999–2018 to cover a sufficient time period with multiple business cycles in the industry. The reason for the data to begin in 1999 is twofold. First, with 10 years of development since the inception of the fabless-foundry business model in the late-1980s, the global semiconductor value chain has been preliminarily established in the late-1990s so that there are plenty of available annual reports for the fabless and foundry firms on the open market and in Compustat database. Second, two years after the 1997 Asian financial crisis, the year 1999 is a suitable starting point to observe the trend in global semiconductor industry without massive exogenous shocks for the following years until the 2008 financial crisis. I also exclude liquid crystal display manufactures, light-emitting diodes manufactures and photovoltaic producers from the dataset, limiting the sample within only IC manufactures in a narrow sense. Hence the panel data include 5136 observations from 470 unique companies in global semiconductor industry in 1999–2018.

A side product of the flexible functional form in nonparametric frontier approach is that there is lack of theoretical fundation on the production function. Identifying the inputs and outputs has always been a subject of controversy, either in parametric or nonparametric frontier estimations, without exception in the semiconductor industry. Hence I sort the most commonly used variables in 37 empirical papers which apply the nonparametric frontier approach for performance evaluation in the semiconductor industry. Besides a few variables which are chosen for specific topics, the commonly used variables in these papers are highly concentrated into two input categories and two output categories. The first input category measures all kinds of variable inputs, including labor, raw material, R&D and sales and marketing expenditure, while the second input category measures fixed assets. Comparably, the first output category measures revenue and the second output category measures the market value of the firms. Therefore, I specify $p = 5$ inputs (labor, measured by the number of employees (X_1); COGS (X_2); R&D expenditure (X_3); sales and marketing expenditure (X_4) and fixed assets, measured by property, plant, and equipment (PP&E) (X_f)) and $q = 2$ outputs (total revenue (Y_1); and shareholders' equity, measured by common ordinary equity (CEQ) (Y_2)). Since I plan to use the directional distance estimator, I distinguish the notation of the fixed input X_f with the other four variable inputs X_1, X_2, X_3 and X_4. For the output variable Y_2, I use shareholders' equity instead of market value of a firm, because the variable of market value is suffering from missing data in Compustat database and the variable shareholders' equity is also a widely used proxy for the value of a firm.

Table 3.1 gives summary statistics for the original variables in 1999–2018 pooled data. In order to provide a uniform standard across years, all the variables except X_1 are expressed in millions of U.S. dollars and their values have been adjusted to 2018 U.S. dollar by GDP deflator. The distribution of all the variables are heavily skewed to the right, owning to the domination of several semiconductor giants in the market. In addition, I specify $r = 2$ environmental variables (business model (Z_1); and time, measured by the years 1999–2018 (Z_2)). The environmental variable Z_1

is a discrete variable, which categorize the four kinds of business models including fabless, IDM, foundry, and A&T into three groups. The first group contains fabless companies which are labor intensive for chip design, while the second group contains both foundries and OSATs which are capital intensive for fabrication, and a third group contains IDMs which are both labor intensive and capital intensive. The environmental variable Z_2 can either be treated as a continuous variable or a discrete variable, which will be discussed further in the next section.

Table 3.2 breaks down the 5136 observations by business model. It is no surprise that over half of the companies are fabless. As the barriers to entry, which relies heavily on CAPEX, is much lower for fabless than for the others, fabless companies spring up like the mushrooms in the late-1990s to the early-2000s. At the same time, the number of firms operating in other kinds of business models remain relatively stable. After the golden decade of fast growth in the semiconductor industry come to an end in the mid-2000s (e.g., see Flamm 2017), the proportions of firms in each business model are gradually fixed. Around 60% of the firms are fabless, while 20% of the firms are IDMs and the rest 20% are either front-end wafer fabs or back-end OSATs.

3.5 Empirical Results

It is well known that most nonparametric estimators suffer from the curse of dimensionality. Based on the three diagnostics introduced in Section 3.3, the necessity for dimension reduction is unambiguous. With seven dimensions (p=5 and q=2) in the original data, it is no surprise that the effective parametric sample size m for the original annual data is small, no matter using FDH or DEA estimators. A slight difference in processing PCA for the directional distance estimator is that

PCA is only on the variable inputs and outputs, but not on the fixed input X_f. Hence after dimension reduction there remain three dimensions including $\widetilde{X}$ (PCA from $X_1, ..., X_4$), $\widetilde{Y}$ (PCA from $Y_1, ..., Y_2$) and X_f. I calculate the values of the largest eigenvalue of the moment matrices of $\boldsymbol{XX'}$ and $\boldsymbol{YY'}$ to the corresponding sum of eigenvalues to be $R_x = 91.19\%$ and $R_y = 98.31\%$, indicating high correlations among the input variables $X_1, ..., X_4$ and high correlations between the output variables $Y_1, ..., Y_2$, so that dimension reduction should reduce estimation error. Thus all of the following analyses and results are based on data with dimension reduciton.

Among studies that use nonparametric frontier approach to estimate efficiency and benchmark performance of firms in the semiconductor industry, the vast majority choose DEA estimator, without comparing the pros and cons between FDH estimator and DEA estimator. The DEA estimator is probably a better choice without dimension reduction, as the slower convergence rate of FDH estimator may increase measurement error rapidly with the increasing of dimensions. However, it is worth to reevaluate the tradeoff between FDH estimator and DEA estimator with dimension reduction. The drawback of DEA estimator is imposing convexity on the production set Ψ, while FDH estimator is free of this assumption. As discussed in Section 3.3, the test of convexity versus non-convexity of the product set proposed by Kneip et al. (2016) can be applied to measure this tradeoff. The FEAR package (e.g., see Wilson, 2008) uses a bootstrap algorithm by Simar and Wilson (2020) to extend the Kneip et al. (2016) approach for the convexity test. Table 3.3 provides results of the convexity test, using the FEAR package and the choosing hyperbolic-oriented measure in (2.3). At 95% confidence level, the null hypothesis of convexity are rejected for over 80% of the 20 years annual data, except 3 years (2009, 2011 and 2012) in hyperbolic-orientation. Simar and Vanhems (2012) link the directional distance measure with the standard hyperbolic measure by a monotonic transformation, so that the results in Table 3.3 are also valid for the directional distance estimator. Hence I choose FDH estimator for the remainder of the analysis.

Daraio et al. (2020) propose a fast and efficient computation of the directional distance measures using FDH estimator. In detail, after monotonic transformation of the data that

$$X^* = \widetilde{X} \oslash d_x \quad \text{and} \quad Y^* = \widetilde{Y} \oslash d_y, \tag{3.5.1}$$

where $\oslash$ refers to Hadamard component-wise division of vectors, the FDH estimator in (2.4) can be expressed explicitly as

$$\begin{aligned} \widehat{\beta}(x, y|d_x, d_y) &= \sup\{\beta > 0 \mid \widehat{H}_{n,X^*Y^*|X_f}(x^* - \beta, y^* + \beta \mid x_f) > 0\}, \\ &= \max_{\{i|X_{f,i} \leq x_f\}} \left[\min\left\{x^* - X_i^*, Y_i^* - y^*\right\}\right], \end{aligned} \tag{3.5.2}$$

where n is the sample size and $i \in \{1, 2, ..., n\}$. It is straightforward to extend the expression in (5.2) to the conditional directional distance estimator (e.g., see Daraio et al., 2020, pp. 814) as

$$\widehat{\beta}(x, y|d_x, d_y, z) = \max_{\{i|X_{f,i} \leq x_f, |Z_i - z| \leq h\}} \left[\min\left\{x^* - X_i^*, Y_i^* - y^*\right\}\right]. \tag{3.5.3}$$

Therefore, in order to consider a discrete environmental variable such as the business model Z_1 for the estimator in (5.3), the separability condition in (2.12) needs to be examined. Similarly, in order to consider a continuous environmental variable such as Z_2 (in case Z_2 is treated as continuous) for the estimator in (5.3), the optimal bandwidth h needs to be fixed in advance.

Simar and Wilson (2020) propose a bootstrap algorithm which can be used for the separability test on the discrete environmental variable Z_1. In application to the additive directional distance measure, step [5] in Simar and Wilson (2020, pp. 293) which is originally designed for the radial measures should be revised as

$$X_i^* = X_i - \widehat{\beta}_i \times d_x + \beta_i^* \times d_x, \tag{3.5.4}$$

and

$$Y_i^* = Y_i + \widehat{\beta}_i \times d_y - \beta_i^* \times d_y, \tag{3.5.5}$$

where the directions of d_x and d_y are chosen commonly as the sample mean of X and Y. The first portion in Table 3.4 shows the separability test results with respect to the business model Z_1. Though the test statistics τ_1 and τ_2 not always give the same results, there is strong evidence to reject the separability condition in (2.12). In other words, each of the three business models in semiconductor industry has its unique production frontier for pooled data.

For the environmental variable Z_2 which represents the years 1999–2018, there is flexibility to either treat it as a discrete variable or as a continuous variable (e.g., see Mastromarco and Simar, 2015). To treat Z_2 as a discrete variable, the 20 years of 1999–2018 can be splitted into ten 2-year groups (two adjacent years as a group), five 4-year groups (four adjacent years as a group) or four 5-year groups (five adjacent years as a group). In this case the separability test with respect to Z_2 is similar to the separability test with respect to Z_1. Although Z_2 can naturally be treated as 20 individual years, it is not recommended for the directional distance measure here. Since each individual year has around 100–300 observations, the effective parametric sample size $m = n^{\frac{2}{3}}$ for the directional distance measure in each year will be a small number and hence increase the measurement error and make this approach less attrative.

Another approach is to treat Z_2 as a continuous variable and using LSCV to determine the optimal bandwidth h by minimizing

$$\frac{\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left[\mathrm{I}(\widetilde{x}_i \leq \widetilde{x}_j, x_{f,i} \leq x_{f,j}, \widetilde{y}_i \geq \widetilde{y}_j) - \frac{\frac{1}{n}\sum_{k\neq i}^{n}\mathrm{I}(\widetilde{x}_k \leq \widetilde{x}_j, x_{f,k} \leq x_{f,j}, \widetilde{y}_k \geq \widetilde{y}_j)K_h(z_i, z_k)}{\frac{1}{n-1}\sum_{k\neq i}^{n}K_h(z_i, z_k)}\right]^2}{n(n-1)}. \tag{3.5.6}$$

The optimal bandwidth is $h = 5.5$ for fabless, IDM, and pooled data, implying the smoothing window of year t is [t-5,t+5], while the optimal bandwidth is $h = 7.5$ for OSAT with the smoothing window of year t to be [t-7,t+7]. The second portion in Table 3.4 shows the separability test results with respect to the optimal time Z_2, while the third portion in Table 3.4 shows the separability test results with respect to both the business model Z_1 and the time Z_2. In any case the separability conditions are strongly rejected. Hence the efficiency scores in (5.3) are estimated with separated production frontiers per the restriction of both the conditions Z_1 and Z_2.

Table 3.5 shows the summary of the efficiency scores conditional on both the business model Z_1 and time Z_2. Whether the time Z_2 is treated as a discrete variable or a continuous variable, the distributions of the efficiency scores are skewed to the right in all kinds of business models, especially for the fabless firms. Nevertheless, on conditions that Z_2 are treated as a discrete variable, the first quartiles are either equal to zero or very close to zero, no matter how the years are grouped. Based on the second diagnostic in Wilson (2018), it is a sign that the measurement error by slow convergence rate still exist. Choosing a estimation method with larger subsample size is a feasible solution to minimize such measurement error with dimension reduction. Thus a preferred approach is to treat Z_2 as a continuous variable, rewarding faster convergence rate and more accurate estimates.

Figure 3.1 visualizes the trends of the annual mean efficiencies by business model. Either treating Z_2 as a discrete variable or as a continuous variable, the curves of the annual mean efficiencies for the fabless firms are above the curves for the other business models. This phenomenon is more visable in the bottom right panel, where Z_2 is defined as a continuous variable with more reliable estimates. As higher efficiency score infers lower technical efficiency in the directional distance measure, the curves in Figure 3.1 imply the fabless firms are operating less efficiently on average. Another interesting discovery is that the curves of different business models in the bottom right panel of Figure 3.1 tend to converge in 2008, the year of global finanical crisis. It can be interpreted that under extreme conditions the differences in operating efficiency become unconspicuous among business models. Based on the bottom right panel of Figure 3.1 which produce more accurate estimates, I use new central limit theorem (e.g., see Kneip et al., 2015, pp. 409) to derive 95% confidence interval for the annual mean efficiency curves in Figure 3.2. The variance for the fabless firms are also higher comparing with the IDMs or OSATs, implying higher risk and uncertainty for the fabless business model.

As the separability condition in (2.12) does not hold, I use a flexible nonparametric location-scale model in (2.13) for a second-stage regression. The pure efficiency defined by Bădin et al. (2012) can be derived from (2.13) and expressed as

$$\widehat{\varepsilon}(z) = \frac{\widehat{\beta}(x, y \mid z) - \widehat{\mu}(z)}{\widehat{\sigma}(z)}. \tag{3.5.7}$$

In practice, I obtain $\widehat{\mu}(z)$ by regressing $\widehat{\beta}(x, y \mid z)$ on the environmental variable z and $\widehat{\sigma}(z)$ by regressing the squared residuals of the preceding regression on z. The upper panel in Figure 3.3 illustrates the pure efficiency $\widehat{\varepsilon}(z_1, z_2)$ that cleanses efficiency scores from the influence of both the environmental factors Z_1 and Z_2, while the lower panel in Figure 3.3 illustrates the pure efficiency $\widehat{\varepsilon}(z_2)$ that cleanses efficiency scores from the influence of only the environmental factor Z_2. In the upper panel, the curves of

$\widehat{\varepsilon}(z_1, z_2)$ by different business models twist together with no clear structures, similar to white noise vibrating at small values around zero. On the contrary, in the lower panel, the curves of $\widehat{\varepsilon}(z_2)$ demonstrate clear separation by business models. Since $\widehat{\varepsilon}(z_2)$ only cleanses the influence of time, the lower panel in Figure 3.3 maintains the structure of the differences in technical efficiency by business model in Figures 3.1-3.2. Consequently, the contrast between the upper and lower panels in Figure 3.3 provide further evidence that the technical efficiencies do vary in the semiconductor by business models, and the fabless firms are operating less efficiently in the past two decades.

3.6 Summary and Conclusions

The semiconductor industry is famous for the high barriers to entry, especially in the capital intensive manufacturing portion. Taking advantage of the economic moat by huge CAPEX and economy of scale, the incumbent IDMs dominate the semiconductor industry since the onset of the industry. Nonetheless, with ever-expanding complexity of ICs and accelerated technology iterations, betting on new technologies and processes to stay ahead of the pack becomes heavy burden even for the dominating IDMs nowaday. The raising of the fabless-foundry business model diversifies the financial risks of capital investment into specified R&D, front-end wafer fabrication, and back-end A&T portions. The decentralized cooperative collaboration of the fabless-foundry alliance drastically reduces the barriers to entry into the globalized semiconductor value chain and lead to a flourishing of fabless design houses for various applications. This paper compares the operating efficiencies between the IDMs and the fabless-foundry business models to shed light on which business model will be the market trend and dominate the semiconductor industry in the long run.

Using a flexible nonparametric frontier approach, I measures the effect of business model in the semiconductor industry. Based on the capital intensive feature of the semiconductor industry, the directional distance measure is chosen to handle the constraint of CAPEX. The empirical results provide clear evidence that the fabless firms are operating less efficiently on average. Though the fabless-foundry business model encourages entrance of the fabless startups, the CAPEX barriers accompaning with technical barriers still limit the fields and applications for the fabless firms to growth and development. Taking advantage of vertical integration, the IDMs have more room to optimize the operation and lead the technology development with strategic product roadmap. The fabless-foundry business model is a complementary of the IDMs to explore a broader scope in the semiconductor industry, instead of a substitutional structure change. The IDMs will continuously dominate the semiconductor industry in the foreseeable future.

However, the distinction between IDM model and fabless-foundry model is fading away. Due to the constant and costly need to upgrade manufacturing facilities to keep up with technological advances, several IDMs contract with foundries to manufacture specific chips while performing all other remaining tasks internally. This symbiotic relationship in the semiconductor ecosystem is called the fablite business model that the complementarity between IDMs and fabless-foundry firms enhance competitiveness through increasing specialization in certain segments of the value chain. This ecosystem is together enhancing the overall competitiveness of semiconductors in capabilities, product diversities, and technological advancement. Although the so-called fablite business model is out of the scope of this paper, it is an intertesting topic that worth further studies to deduce evolutionary trends in the semiconductor industry.

Table 3.1: Summary statistics for 1999–2018 pooled data

Variable	Min	Q1	Median	Mean	Q3	Max
X_1	0.001	0.160	0.486	3.082	2.011	107.600
X_2	0.001	24.170	88.008	475.252	301.645	18,226.000
X_3	0.000	4.302	18.330	160.217	67.001	13,543.000
X_4	0.549	5.885	20.087	125.406	67.185	1,982.015
X_f	0.005	6.065	27.787	554.060	174.405	48,976.000
Y_1	0.003	47.283	161.799	1064.110	563.655	70,848.000
Y_2	0.175	44.279	151.749	1114.748	487.730	74,563.000
Obs.				5,136		
Uniq. Obs.				470		

NOTE. The unit of X_1 is thousand employees.
The units of the variables except X_1 are US$ million.
All values have been adjusted to 2018 US$ by GDP deflator.
Obs. denotes the total number of observations in 1999–2018.
Uniq. Obs. denotes the unique number of companies in 1999–2018.

Table 3.2: Number of observations by business model

	Number of Companies				
Year	All	IDM	Foundry	A&T	Fabless
1999	125	38	10	9	68
2000	149	43	10	15	81
2001	155	46	10	16	83
2002	213	48	17	27	121
2003	241	49	19	30	143
2004	264	54	21	30	159
2005	260	54	17	27	162
2006	267	56	20	30	161
2007	269	52	21	33	163
2008	278	51	20	35	172
2009	290	53	21	36	180
2010	300	59	23	38	180
2011	298	60	22	39	177
2012	301	61	22	38	180
2013	313	65	24	41	183
2014	302	62	25	43	172
2015	288	59	24	42	163
2016	283	54	23	44	162
2017	275	51	23	45	156
2018	265	48	22	44	151
Obs.	5,136	1,063	394	662	3,017
Uniq. Obs.	470	83	36	63	288

NOTE. Obs. denotes the total number of observations in 1999–2018.
Uniq. Obs. denotes the unique number of companies in 1999–2018.

Table 3.3: Results of convexity test (hyperbolic-orientation)

Year	N	Statistic	p-value
1999	125	2.222	0.011
2000	149	1.594	0.006
2001	155	1.853	0.040
2002	213	3.138	0.005
2003	241	2.901	0.001
2004	264	3.440	0.000
2005	260	3.238	0.000
2006	267	3.651	0.000
2007	269	3.915	0.003
2008	278	3.227	0.008
2009	290	2.162	0.057
2010	300	2.890	0.006
2011	298	2.102	0.088
2012	301	1.014	0.174
2013	313	1.989	0.020
2014	302	3.552	0.001
2015	288	1.452	0.041
2016	283	2.053	0.018
2017	275	4.831	0.000
2018	265	4.963	0.000

NOTE. I use 100 splits and 1000 bootstrap replications.

Table 3.4: Test of separability conditional on Z_1 and Z_2 (with dimension reduction, $p = 2$, $q = 1$, and directional distance measure)

	— τ_1 —		— τ_2 —	
	Statistic	p-value	Statistic	p-value
		— Conditional on Z_1 —		
Fabless VS. IDM	7.567	0.000	0.992	0.925
Fabless VS. OSAT	4.126	0.000	1.000	0.000
IDM VS. OSAT	3.883	0.000	0.870	0.715
		— Conditional on Z_2 —		
Pooled VS. Optimal Time	3.230	0.000	0.983	0.000
		— Conditional on Z_1, Z_2 —		
2-Year Groups	26.273	0.000	1.000	0.000
4-Year Groups	18.572	0.000	1.000	0.000
5-Year Groups	17.242	0.000	1.000	0.000
Optimal Time	19.891	0.000	1.000	0.000

NOTE. I use 10 splits and 1000 bootstrap replications.
τ_1 is the averaging of the statistics across 10 splits.
τ_2 is the Kolmogorov-Smirnov statistic obtained above.

Table 3.5: Summary statistics of the efficiency scores by directional distance estimator

Sample Set	Sample Size	Min	Q1	Median	Mean	Q3	Max
				2-year groups			
fabless	3,017	0.000	0.000	0.023	0.097	0.085	13.859
IDM	1,063	0.000	0.000	0.002	0.046	0.033	1.943
OSAT	1,056	0.000	0.000	0.000	0.041	0.029	1.167
				4-year groups			
fabless	3,017	0.000	0.005	0.036	0.129	0.107	15.171
IDM	1,063	0.000	0.000	0.009	0.067	0.061	1.939
OSAT	1,056	0.000	0.000	0.005	0.064	0.055	1.901
				5-year groups			
fabless	3,017	0.000	0.009	0.042	0.147	0.115	15.179
IDM	1,063	0.000	0.000	0.012	0.074	0.069	2.377
OSAT	1,056	0.000	0.000	0.008	0.074	0.071	2.000
				optimal time			
fabless	3,017	0.000	0.027	0.073	0.258	0.175	19.151
IDM	1,063	0.000	0.006	0.043	0.131	0.171	2.087
OSAT	1,056	0.000	0.004	0.036	0.142	0.146	2.660

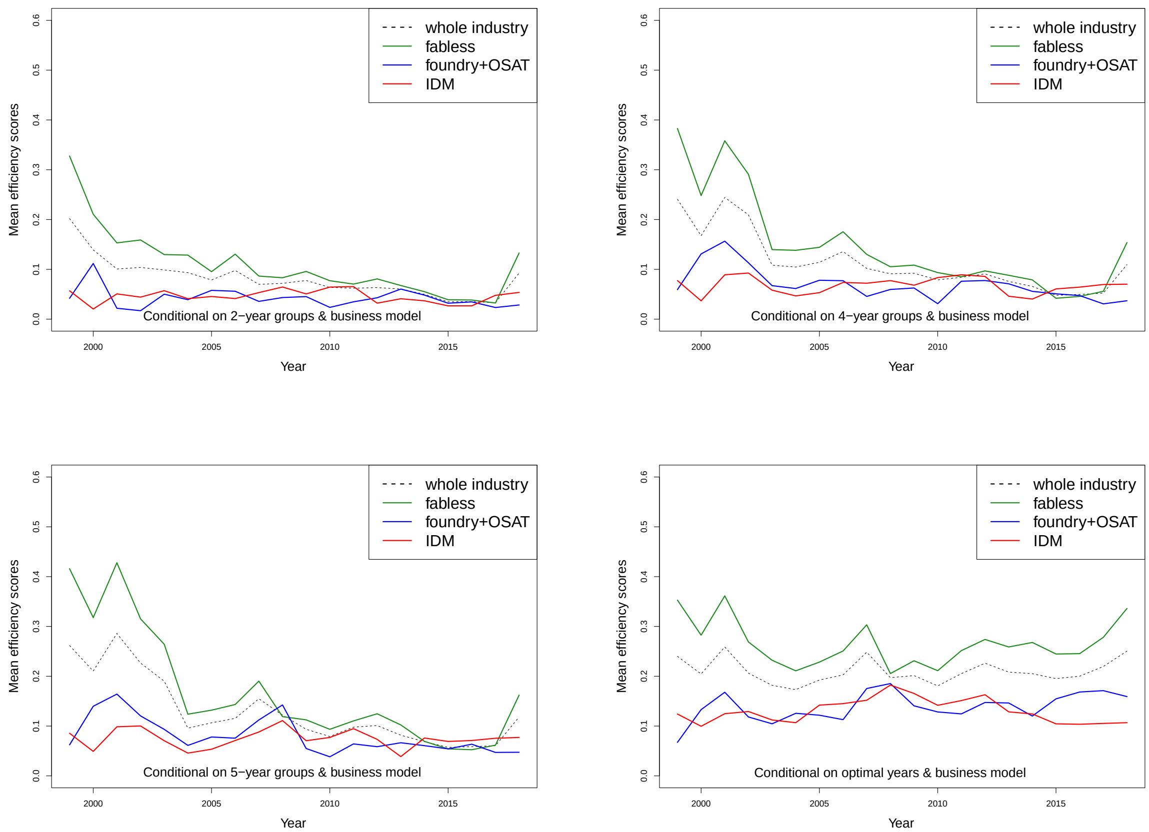

Figure 3.1: Mean β conditional on both year and business model

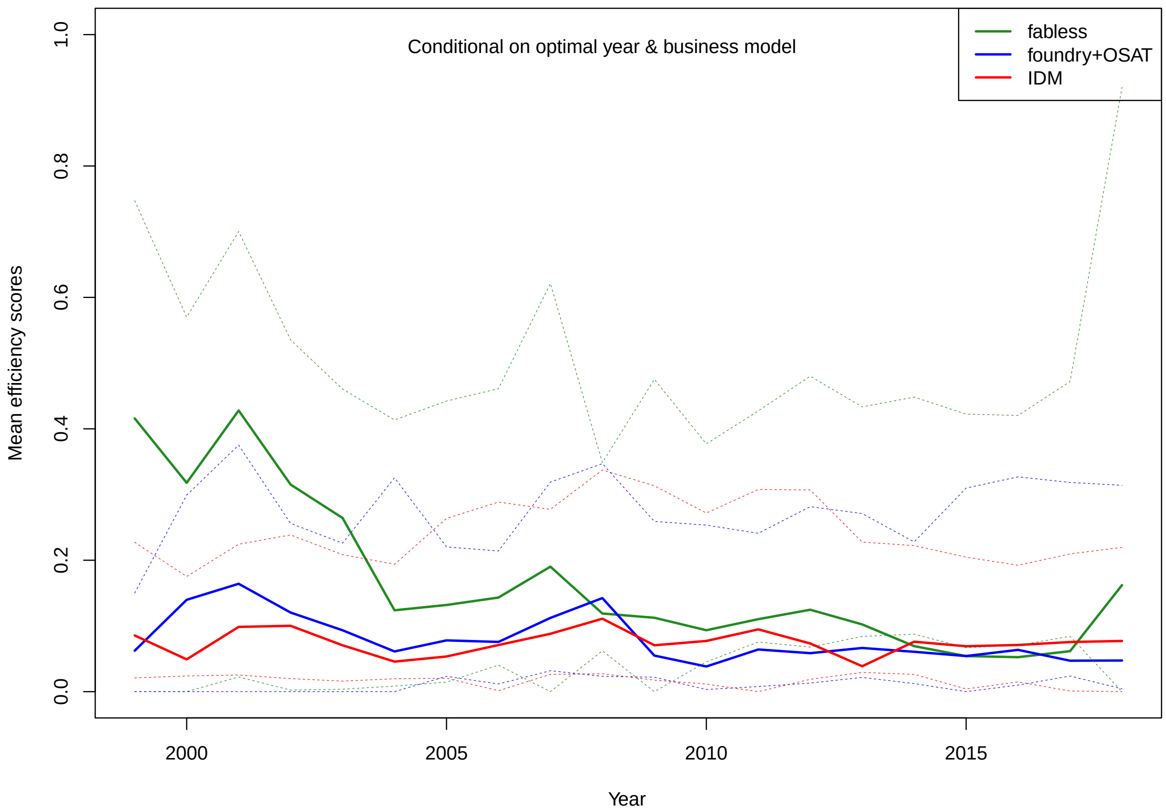

Figure 3.2: Mean efficiency conditional on optimal year and business model with 95% confidence interval

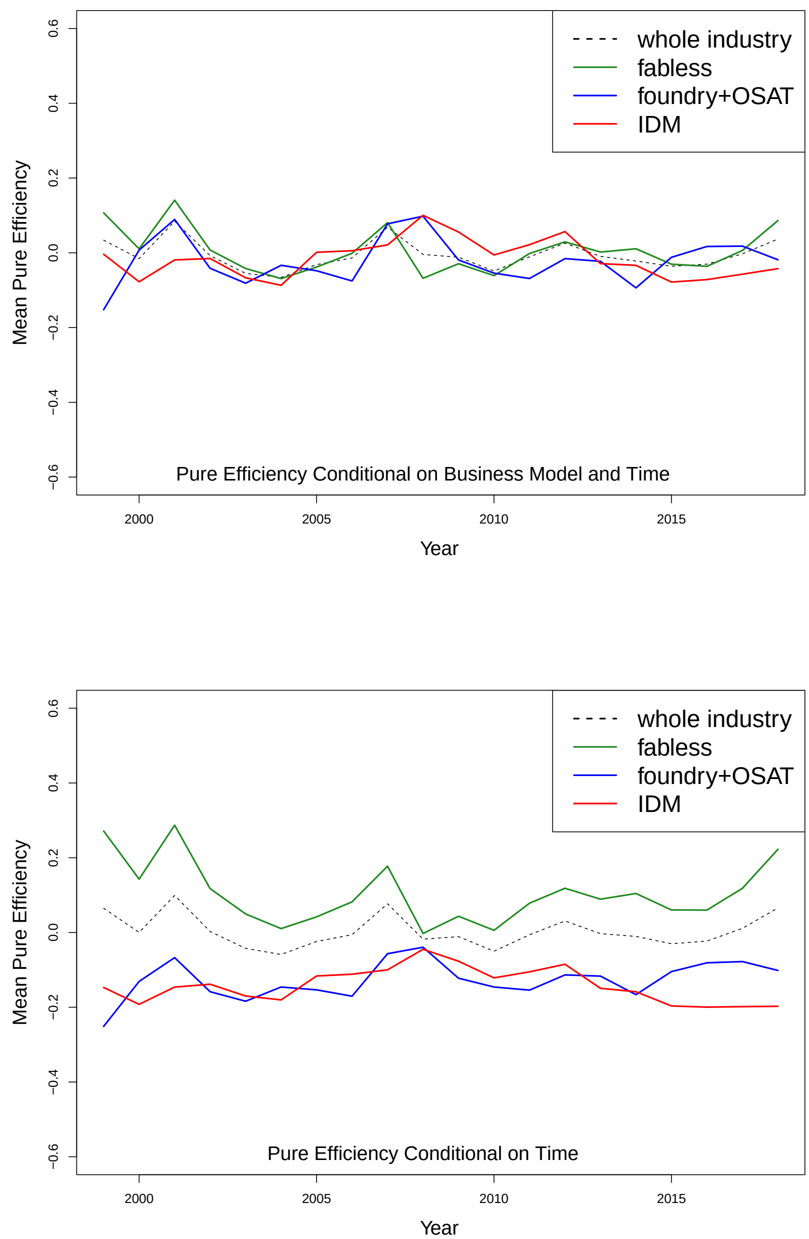

Figure 3.3: Pure efficiency

References

Adner, R., Kapoor, R., 2010, Value creation in innovation ecosystems: how the structure of technological interdependence affects firm performance in new technology generations, *Strategic Management Journal*, Vol. 31, pp. 306–333.

Afriat, S., 1972, Efficiency estimation of production functions, *International Economic Review*, Vol. 6, pp. 568–598.

Aigner, D.J., Lovell, C.A.K., Schmidt, P., 1977, Formulation and estimation of stochastic frontier models, *Journal of Econometrics*, Vol. 6, pp. 21–37.

Almanidis, P., Sickles, R.C., 2011, The skewness issue in stochastic frontiers models: fact or fiction? In I. Van Keilegom and P. W. Wilson, eds., *Exploring Research Frontiers in Contemporary Statistics and Econometrics*, Berlin: Springer-Verlag, pp. 201–228.

Bădin, L., Daraio, C., Simar, L., 2010, Optimal bandwidth selection for conditional efficiency measures: A data-driven approach, *European Journal of Operational Research*, Vol. 201, pp. 633–640.

Bădin, L., Daraio, C., Simar, L., 2012, How to measure the impact of environmental factors in a nonparametric production model, *European Journal of Operational Research*, Vol. 223, pp. 818–833.

Bădin, L., Daraio, C., Simar, L., 2014, Explaining inefficiency in nonparametric production models: the state of the art, *Annals of Operations Research*, Vol. 214, pp. 5–30.

Balconi, M., Fontana, R., 2011, Entry and innovation: an analysis of the fabless semiconductor business, *Small Business Economics*, Vol. 37, pp. 87–106.

Banker, R.D., Charnes, A., Cooper, W.W., 1984, Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, Vol. 30, pp. 1078–1092.

Battese, G.E., Corra, G.S., 1977, Estimation of a production frontier model: with application to the pastoral zone of eastern Australia, *Australian Journal of Agricultural Economics*, Vol. 21, pp. 169–179.

Bonanno, G., 2016, ICT and R&D as inputs or efficiency determinants? Analysing Italian manufacturing firms (2007–2009), *Eurasian Business Review*, Vol. 6, pp. 383–404.

Brown, C., Linden, G., Macher, J., 2005, Offshoring in the semiconductor industry: a historical perspective, *Brookings Trade Forum*, pp. 279–333.

Cai, Z., Chen, L., Fang, Y., 2015, Semiparametric estimation of partially varying-coefficient dynamic panel data models, *Econometric Reviews*, Vol. 34, pp. 694–718.

Chambers, R.G., Chung, Y.H., Färe, R., 1998, Profit, directional distance functions, and Nerlovian efficiency, *Journal of Optimization Theory and Applications*, Vol. 98, pp. 351–364.

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

Chen, Y.S., Chen, B.Y., 2011, Applying DEA, MPI, and grey model to explore the operation performance of the Taiwanese wafer fabrication industry, *Technological Forecasting & Social Change*, Vol. 78, pp. 536–546.

Chiu, C., Fang, C., Tsang, S., Chen, Y., 2018, Performance evaluation of the semiconductor industry based on a metafrontier approach, *Technological and Economic Development of Economy*, Vol. 24, pp. 825–843.

Cordero, J.M., Pedraja-Chaparro, F., Pisaflores, E.C., Polo, C., 2017, Efficiency assessment of Portuguese municipalities using a conditional nonparametric approach, *Journal of Productivity Analysis*, Vol. 48, pp. 1–24.

Daraio, C., Simar, L., 2005, Introducing environmental variables in nonparametric frontier models: a probabilistic approach, *Journal Productivity Analysis*, Vol. 24, pp. 93–121.

Daraio, C., Simar, L., 2007, *Advanced Robust and Nonparametric Methods in Efficiency Analysis*, New York: Springer Science+Business Media, LLC.

Daraio, C., Simar, L., Wilson, P.W., 2018, Central limit theorems for conditional efficiency measures and tests of the "separability" condition in nonparametric, two-stage models of production, *The Econometrics Journal*, Vol. 21, pp. 170–191.

Daraio, C., Simar, L., Wilson, P.W., 2020, Fast and efficient computation of directional distance estimators, *Annals of Operations Research*, Vol. 288, pp. 805–835.

Debreu, G., 1951, The coefficient of resource utilization, *Econometrica*, Vol. 19, pp. 273–292.

Deprins, D., Simar, L., Tulkens, H., 1984, Measuring labor-efficiency in post offices, *The Performance of Public Enterprises: Concepts and Measurement*, Amsterdam: North-Holland, pp. 243–267.

Dibiaggio, L., 2007, Design complexity, vertical disintegration and knowledge organization in the semiconductor industry, *Industrial and Corporate Change*, Vol. 16, pp. 239–267.

Du, P., Parmeter, C.F., Racine, J., 2013, Nonparametric kernel regression with multiple predictors and multiple shape constraints, *Statistica Sinica*, Vol. 23, pp. 1347–1371.

Ernst, D., 2005, Limits to modularity: reflections on recent developments in chip design, *Industry and Innovation*, Vol. 12, pp. 303–335.

Fan, J., Gijbels, I., 1992, Variable bandwidth and local linear regression smoothers, *The Annals of Statistics*, Vol. 20, pp. 2008–2036.

Fan, Y., Li, Q., Weersink, A., 1996, Semiparametric estimation of stochastic production frontier, *Journal of Business* & *Economic Statistics*, Vol. 14, pp. 460–468.

Färe, R.S., Grosskopf, S., Lovell, A.K., 1985, *The measurement of efficiency of production*, Boston: Kluwer-Nijhoff Publishing.

Farrell, M.J., 1957, The measurement of productive efficiency, *Journal of the Royal Statistical Society*, Vol. 120, pp. 253–290.

Flamm, K., 2017, Has Moore's law been repealed? An economist's perspective, *Computing in Science & Engineering*, Vol. 19, pp. 29–40.

Fontana, R., Malerba, F., 2010, Demand as a source of entry and the survival of new semiconductor firms, *Industrial and Corporate Change*, Vol. 19, pp. 1629–1654.

Gasser, T., Millier, H.G., 1979, Kernel estimation of regression functions, In *Smoothing Techniques for Curve Estimation*, pp. 23–68. Springer-Verlag, New York.

Hall, P., Racine, J.S., Li, Q., 2004, Cross-validation and the estimation of conditional probability densities, *Journal of the American Statistical Association*, Vol. 99, pp. 1015–1026.

Hall, P., Simar, L., 2002, Estimating a changepoint, boundary or frontier in the presence of observation error, *Journal of the American Statistical Association*, Vol. 97, pp. 523–534.

Hsieh, Y., Lin, N., Chiu, H., 2002, Virtual factory and relationship marketing — a case study of a Taiwan semiconductor manufacturing company, *International Journal of Information Management*, Vol. 22, pp. 109–126.

Hsu, L.C., 2015, Using a decision-making process to evaluate efficiency and operating performance for listed semiconductor companies, *Technological and Economic Development of Economy*, Vol. 21, pp. 301–331.

Hung, S.W., He, D.S., Lu, W.M., 2014, Evaluating the dynamic performances of business groups from the carry-over perspective: a case study of Taiwan's semiconductor industry, *Omega*, Vol. 46, pp. 1–10.

Hung, S.W., Lu, W.M., 2008, The comparative productivity efficiency of Taiwan's integrated circuits packaging/testing Firms, *Information Systems and Operational Reserach*, Vol. 46, pp. 189–198.

Hwang, H., Choung, J., 2014, The co-evolution of technology and institutions in the catch-up process: the case of the semiconductor industry in Korea and Taiwan, *The Journal of Development Studies*, Vol. 50, pp. 1240–1260.

Jang, H., Lee, S., Suh, E., 2016, A comparative analysis of the change in R&D efficiency: a case of R&D leaders in the technology industry, *Technology Analysis & Strategic Management*, Vol. 28, pp. 886–900.

Jeong, S., Park, B.U., Simar, L., 2010, Nonparametric conditional efficiency measures: asymptotic properties, *Annals of Operations Research*, Vol. 173, pp. 105–122.

Kapoor, R., 2013, Persistence of integration in the face of specialization: how firms navigated the winds of disintegration and shaped the architecture of the semiconductor industry, *Organization Science*, Vol. 24, pp. 1195–1213.

Kapoor, R., Adner, R., 2012, What firms make vs. what they know: how firms' production and knowledge boundaries affect competitive advantage in the face of technological change, *Organization Science*, Vol. 23, pp. 1227–1248.

Kneip, A., Simar, L., 1996, A general framework for frontier estimation with panel data, *The Journal of Productivity Analysis*, Vol. 7, pp. 187–212.

Kneip, A., Simar, L., Wilson, P.W., 2015, When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores, *Econometric Theory*, Vol. 31, pp. 394–422.

Kneip, A., Simar, L., Wilson, P.W., 2016, Testing hypotheses in nonparametric models of production, *Journal of Business and Economic Statistics*, Vol. 34, pp. 435–456.

Koopmans, T.C., 1951, An analysis of production as an efficient combination of activities. In *Activity Analysis of Production and Allocation*, Ed. T.C. Koopmans. New York: John-Wiley and Sons, Inc.; 33–97. Cowles Commission for Research in Economics, Monograph 13.

Kozmetsky, G., Yue, P., 1998, Comparative performance of global semiconductor companies, *Omega*, Vol. 26, pp. 153–175.

Kumbhakar, S.C., Ortega-Argilés, R., Potters, L., Vivarelli, M., Voigt, P., 2012, Corporate R&D and firm efficiency: evidence from Europe's top R&D investors, *Journal of Productivity Analysis*, Vol. 37, pp. 125–140.

Kumbhakar, S.C., Park, B.U., Simar, L., Tsionas, E.G., 2007, Nonparametric stochastic frontiers: a local likelihood approach, *Journal of Econometrics*, Vol. 137, pp. 1–27.

Kuo, M.H., Yang, C., 2012, Does intellectual capital matter? Assessing the profitability and marketability of IC design companies, *Quality* & *Quantity*, Vol. 46, pp. 1865–1881.

Kuosmanen, T., 2008, Representation theorem for convex nonparametric least squares, *Econometrics Journal*, Vol. 11, pp. 308–325.

Kuosmanen, T., Kortelainen, M., 2012, Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints, *Journal of Productivity Analysis*, Vol. 38, pp. 11–28.

Li, H., He, H., Shan, J., Cai, J., 2019, Innovation efficiency of semiconductor industry in China: a new framework based on generalized three-stage DEA analysis, *Socio-Economic Planning Sciences*, Vol. 66, pp. 136–148.

Li, Q., Lin, J., Racine, J.S., 2013, Optimal bandwidth selection for nonparametric conditional distribution and quantile functions, *Journal of Business* & *Economic Statistics*, Vol. 31, pp. 57–65.

Li, Q., Racine, J., 2007, *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Li, Y., Huang, M., Chen, D., 2011, Semiconductor industry value chain: characters' technology evolution, *Industrial Management* & *Data Systems*, Vol. 111, pp. 370–390.

Lin, F., Lin, S.W., Lu, W.M., 2019, Dynamic eco-efficiency evaluation of the semiconductor industry: A sustainable development perspective, *Environ Monit Assess*, Vol. 191, pp. 1–16.

Liu, F., Wang, P., 2008, DEA Malmquist productivity measure: Taiwanese semiconductor companies, *International Journal of Production Economics*, Vol. 112, pp. 367–379.

Lovell, K., 1993, Production Frontiers and Productive Efficiency, in *The Measurement of Productive Efficiency*, H. Fried, K. Lovell, and S. Schmidt, eds., Oxford University Press, Oxford, UK.

Lu, W.M., 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*, Vol. 48, pp. 1155–1170.

Lu, W.M., Wang, W.K., Lee, H.L., 2013, The relationship between corporate social responsibility and corporate performance: evidence from the US semiconductor industry, *International Journal of Production Research*, Vol. 51, pp. 5683–5695.

Macher, J.T., 2006, Technological development and the boundaries of the firm: a knowledge-based examination in semiconductor manufacturing, *Management Science*, Vol. 52, pp. 826–843.

Macher, J.T., Mowery, D.C., 2004, Vertical specialization and industry structure in high technology industries, In *Business Strategy over the Industry Lifecycle (Advances in Strategic Management, Vol. 21)*, Baum, J.A.C., McGahan, A.M., pp. 317–355, Emerald Group Publishing Limited, Bingley.

Macher, J.T., Mowery, D.C., Di Minin, A., 2007, The "non-globalization" of innovation in the semiconductor industry, *California Management Review*, Vol. 50, pp. 217–242.

Mack, C.A., 2011, Fifty Years of Moore's Law, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 24, pp. 202–207.

Mastromarco, C., Simar, L. 2015, Effect of FDI and time on catching up: new insights from a conditional nonparametric frontier analysis, *Journal of Applied Econometrics* Vol. 30, pp. 826–847.

Mathews, J.A., Cho, D., 2000, *Tiger technology: the creation of a semiconductor industry in east Asia*, Cambridge University Press.

Meeusen, W., Broeck, J.V.D., 1977, Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review*, Vol. 18, pp. 435–444.

Monteverde, K., 1995, Technical dialog as an incentive for vertical integration in the semiconductor industry, *Management Science* Vol. 41, pp. 1624–1638.

Nadaraya, E.A., 1965, On nonparametric estimates of density functions and regression curves, *Theory of Probability and Its Applications*, Vol. 10, pp. 199–203.

Priestley, M.B., Chao, M.T., 1972, Nonparametric function fitting, *Royal Statistical Society: Series B*, Vol. 34, pp. 385–392.

Racine, J., Nie, Z., 2017, *crs: Categorical Regression Splines.* R package version 0.51-31.

Richmond, J., 1974, Estimating the Efficiency of Production, *International Economic Review*, Vol. 15, pp. 515–521.

Sarma, S., Sun, L., 2017, The genesis of fabless business model: institutional entrepreneurs in an adaptive ecosystem, *Asia Pacific Journal of Management*, Vol. 34, pp. 587–617.

Shin, N., Kraemer, K., Dedrick, J., 2017, R&D and firm performance in the semiconductor industry, *Industry and Innovation*, Vol. 24, pp. 280–297.

Simar, L., Keilegom, I.V., Zelenyuk, V., 2017, Nonparametric least squares methods for stochastic frontier models, *Journal of Productivity Analysis*, Vol. 47, pp. 189–204.

Simar, L., Vanhems, A., 2012 Probabilistic characterization of directional distances and their robust versions, *Journal of Econometrics*, Vol. 166, pp. 342–354.

Simar, L., Wilson, P.W., 2007, Estimation and inference in two-stage, semi-parametric models of productive efficiency, *Journal of Econometrics*, Vol. 136, pp. 31–64.

Simar, L., Wilson, P.W., 2011, Two-stage DEA: Caveat emptor, *Journal Productivity Analysis*, Vol. 36, pp. 205–218.

Simar, L., Wilson, P.W., 2015, Statistical approaches for non-parametric frontier models: a guided tour, *International Statistical Review*, Vol. 83, pp. 77–110.

Simar, L., Wilson, P.W., 2020, Hypothesis testing in nonparametric models of production using multiple sample splits, *Journal Productivity Analysis*, Vol. 53, pp. 287–303.

Simar, L., Zelenyuk, V., 2011, Stochastic FDH/DEA estimators for frontier analysis, *Journal of Productivity Analysis*, Vol. 36, pp. 1–20.

Tan, H., Mathews, J.A., 2010, Cyclical industrial dynamics: the case of the global semiconductor industry, *Technological Forecasting & Social Change*, Vol. 77, pp. 344–353.

Toma, P., 2020, Size and productivity: a conditional approach for Italian pharmaceutical sector, *Journal Productivity Analysis*, Vol. 54, pp. 1–12.

Tsai, C.H., Wu, H.Y., Chen, I.S., Chen, J.K., Ye, R.W., 2017, Exploring benchmark corporations in the semiconductor industry based on efficiency, *Journal of High Technology Management Research*, Vol. 28, pp. 188–207.

Wang, C., Chiu, C., 2014, Competitive strategies for Taiwan's semiconductor industry in a new world economy, *Technology in Society*, Vol. 36, pp. 60–73.

Watson, G.S., 1964, Smooth regression analysis, *Sankhyā*, Vol. 83, pp. 77–110.

Wilson, P.W., 2008, FEAR 1.0: a software package for frontier efficiency analysis with R, *Socio-Economic Planning Sciences*, Vol. 42, pp. 247–254.

Wilson, P.W., 2018, Dimension reduction in nonparametric models of production, *European Journal of Operational Research*, Vol. 267, pp. 349–367.

Wu, W.Y., Tsai, H.J., Cheng, K.Y., Lai, M., 2006, Assessment of intellectual capital management in Taiwanese IC design companies: using DEA and the Malmquist productivity index, *R&D Management*, Vol. 36, pp. 531–545.