Vertical integration vs. specialization: a nonparametric conditional efficiency estimate for the global semiconductor industry

Guangshun Qiao ¹ · Zhan-ao Wang ¹

Accepted: 22 August 2021 / Published online: 13 September 2021
© The Author(s) 2021

Abstract

This paper applies a two-stage nonparametric approach to compare companies operating in different business models in the global semiconductor industry. Using panel data over 1999–2018 on 470 companies in the global semiconductor industry, we explore the operating performance of the semiconductor companies conditional on capital investment and between the integrated device manufacturers and the fabless-foundry business model. We find that vertically integrated device manufacturers are constrained heavily by capital investment. Disentangling the effects of capital investment and business model by a second-stage nonparametric regression, this paper identifies that the vertically specialized fabless-foundry business model helps to improve pure efficiency and mitigate the impact of business-cycle in the global semiconductor industry.

Keywords Semiconductor industry · Capital investment · Pure efficiency · Nonparametric frontiers · Two-stage conditional efficiency analysis

1 Introduction

Semiconductors, also known as integrated circuits or chips, are the brains of almost all modern electronics. Since the release of the first commercial chips in the 1960s, the semiconductor industry has been a driving force in the electronics market. Before the 1980s, most of the companies in the semiconductor industry operated in the integrated device manufacturer (IDM) business model, in that one company carries out all stages of production in-house, including research and design (R&D), front-end fabrication, and back-end assembly and test (A&T). As Moore's law approaches its physical limits in the semiconductor industry (e.g., see Thompson and Parthasarathy 2006), the expense of manufacturing a leading-edge chip has become prohibitive for all but a few IC suppliers. The rising pressures of developing more advanced technologies, accompanied by

exponentially increasing costs, have kept the semiconductor industry on its toes and led to the rise of the fabless-foundry business model in the mid-1980s. In the fabless-foundry model, a fabless company focuses on R&D and sales, and partners 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) operations. Vertical specialization combined with the trend toward globalization has significantly changed the structure of the semiconductor industry over the past few decades.

Whether specialization by the fabless-foundry business model increases productivity is a topic of broad interest in the semiconductor industry (e.g., see Macher 2006; Dibiaggio 2007). On the one hand, vertical disintegration notably reduces the burden of capital expenditure (CAPEX) and ensures the domination of new markets by fabless firms in the semiconductor industry (e.g., see Balconi and Fontana 2011). On the other hand, despite a trend toward specialization driven by fabless start-ups, IDMs have continued to persist and coexist with the fabless firms in the semiconductor industry (e.g., see Kapoor 2013). Much research on the topic of structural change in the semiconductor industry emphasizes the impact of technological evolution (e.g., see Kapoor and Adner 2012; Pellens and Della Malva 2018). However, few studies focus on the impact of CAPEX or discuss the effect of the business

✉ Guangshun Qiao
qiaoguangshun@wzbc.edu.cn

✉ Zhan-ao Wang
wza@wzbc.edu.cn

¹ School of Finance and Trade, Wenzhou Business College, Wenzhou 325035, China

model in the semiconductor industry adequately. This paper aims to use a two-stage nonparametric conditional approach to explore the operating performance of companies in the global semiconductor industry. The effects of capital investment and business model are handled by conditional efficiency estimators in the first stage, and the idiosyncratic efficiency, representing the Solow residual (e.g., see Mastromarco and Simar 2015) is abstracted by looking to the unexplained part of the conditional efficiencies in the second stage.

In the semiconductor industry, factors determining production costs and operating efficiencies vary substantially across device types and business models. For cutting-edge products such as microprocessors, manufacturing is capital intensive. The ever-increasing costs of building advanced fabrication facilities and the difficulties that arise from slowing development in node technology result in high barriers for entry and favor the success of large IDMs, such as Intel and Texas Instruments, which can forge ahead with innovative and expensive technologies that are needed to take chips to the next level. Comparably, for front-end fabrication and back-end A&T procedures, a significant challenge is the heavy CAPEX for cleanroom and costly equipment. The fabless-foundry model derives efficiencies from the delineation of tasks and specialization. More specifically, foundries and OSATs, both heavily invested in fixed assets, seek to optimize productivity by serving many fabless companies to achieve high capacity utilization. At the same time, the niche fabless start-ups or spinoffs, getting rid of the CAPEX burden, concentrate on R&D to compete with the IDM giants. Hence CAPEX is an indispensable factor in the performance evaluation between the IDM model and the fabless-foundry model in the capital-intensive semiconductor industry.

Data envelopment analysis (DEA) is one of the most popular approaches for performance evaluation, and there is no exception in the semiconductor industry. Among the rich records that have used DEA for efficiency estimation in the semiconductor industry, many have focused on a particular business model, while others have limited the research to a specific country or area. For example, Chu et al. (2008) measure the relative performance for global top 30 fabless firms, Chen and Chen (2011) analyze the operation performance of the foundries in Taiwan, and Lu et al. (2013) explore the relationship between corporate social responsibility and semiconductor companies' performance in the US. Though these works examine the operating efficiencies in the semiconductor industry from different angles, few have expressed a global perspective on the semiconductor industry. Furthermore, because of the slow convergence rate in the nonparametric DEA approach, the limited number of observations in these studies, either constrained by geographic boundary or a single business model, may increase

estimation error. With increasing numbers of inputs and outputs, the issue of slow convergence rate may become severe and lead to unconvincing results in the nonparametric frontier estimation (e.g., see Wilson 2018).

This paper tries to overcome the slow convergence rate in the nonparametric DEA approach in two ways. First, the data in this research cover 470 companies in the semiconductor industry from all over the world, not only helping to increase the sample base, but also helping to grasp a whole picture of the highly globalized semiconductor industry. Second, using the principal component analysis (PCA) for dimension reduction (e.g., see Daraio and Simar 2007a, pp. 148–150) further helps to reduce the estimation error. After filtering out the impacts of CAPEX and business model by the conditional nonparametric frontier framework, a flexible nonparametric location-scale model in a second stage regression identifies the idiosyncratic efficiencies in the semiconductor industry. The estimation results indicate that the vertically integrated IDM business model is constrained heavily by capital investment, and the vertically specialized fabless-foundry business model helps to improve the mean pure efficiency and mitigate the cyclical fluctuations in the global semiconductor industry.

The paper is organized as follows. Section 2 introduces the nonparametric frontier framework and discusses the diagnostics and test statistics for choosing a suitable estimator in this research. In Sect. 3, we describe the dataset and define the variables for the estimation. The empirical results of the idiosyncratic efficiency, and the impacts of capital investment and business model in the global semiconductor industry are presented in Sect. 4. The last section is the conclusion.

2 Methodology

2.1 Nonparametric production frontier

The economic theory of efficiency in production examines how firms transform their inputs into outputs. Consider a production process where a vector of inputs $x \in \mathbb{R}_+^p$ is used to produce a vector of outputs $y \in \mathbb{R}_+^q$. The production set Ψ is defined as $\Psi = \{(x, y) | x \text{ can produce } y\}$ and the efficiency score of a particular production plan (x, y) in Ψ is determined by the distance from (x, y) to the boundary of Ψ . Besides the commonly used input- and output-oriented Debreu-Farrel measures, the hyperbolic measure

$$\gamma(x, y) = \sup\{\gamma \mid (\gamma^{-1}x, \gamma y) \in \Psi\} \quad (1.1)$$

is an alternative to selecting either an input- or output-orientation. Note that the hyperbolic measure in (1.1) is the reciprocal of the hyperbolic graph measure by Färe et al.

(1985). Wilson (2011) popularizes the hyperbolic measure and derives its asymptotic properties.

In real-world research problems, the attainable set Ψ is unobserved. Nonparametric methods such as DEA and Free Disposal Hull (FDH) are widely applied to estimate the unobservable production set Ψ . Let $S_n = \{(X_i, Y_i)\}$ denote a random sample of n observations of input-output pairs. Using only the free disposability assumption, Deprins et al. (1984) define the FDH estimator of Ψ as

$$\hat{\Psi}_{\text{FDH}} = \bigcup_{(X_i, Y_i) \in S_n} \{(x, y) \mid x \geq X_i, y \leq Y_i\}. \quad (1.2)$$

The FDH estimator of $\gamma(x, y)$, denoted by $\hat{\gamma}_{\text{FDH}}(x, y)$, is obtained by replacing Ψ with $\hat{\Psi}_{\text{FDH}}$ in (1.1). Alternatively, Banker et al. (1984) propose the variable-returns-to-scale DEA (VRS-DEA) estimator of Ψ with the expression that

$$\hat{\Psi}_{\text{VRS}} = \{(x, y) \mid y \leq \sum_{i=1}^n \omega_i Y_i, x \geq \sum_{i=1}^n \omega_i X_i, \sum_{i=1}^n \omega_i = 1, \omega_i \geq 0\}, \quad (1.3)$$

where the weights ω_i and the constraints $\sum_{i=1}^n \omega_i = 1$ serve $\hat{\Psi}_{\text{VRS}}$ as the convex hull of $\hat{\Psi}_{\text{FDH}}$. The corresponding VRS-DEA estimator of $\gamma(x, y)$, denoted by $\hat{\gamma}_{\text{VRS}}(x, y)$, is obtained by replacing Ψ with $\hat{\Psi}_{\text{VRS}}$ in (1.1). Another kind of DEA estimator, called the constant-returns-to-scale DEA estimator, relies on more restrictive assumptions and is less widely used. Therefore, the notation DEA is dedicated to VRS-DEA for the rest of the paper.

2.2 Conditional efficiency scores

Beyond the combinations (X, Y) of inputs and outputs, there are factors that may influence the production process but are typically not under the manager's control in a firm. These factors, denoted by $Z \in \mathbb{R}^r$, are referred to as environmental factors that may reflect differences in ownership, business model, capital investment, etc. Cazals et al. (2002) and Daraio and Simar (2005, 2007a, 2007b) develop a framework to investigate the joint behavior of (X, Y, Z) in probability terms. Defining the conditional attainable set as $\Psi^z = \{(x, y) \mid Z = z, x \text{ can produce } y\}$, the distribution of (X, Y) conditional on $Z = z$ is denoted by

$$H_{X,Y|Z}(x, y|z) = \Pr(X \leq x, Y \geq y \mid Z = z), \quad (1.4)$$

which gives the probability that a firm facing environmental conditions z will dominate the point (x, y) . Introducing environmental factors into (1.1) extends the efficiency score of $\gamma(x, y)$ into its conditional counterpart as

$$\gamma(x, y|z) = \sup\{\gamma \mid H_{X,Y|Z}(\gamma^{-1}x, \gamma y|z) > 0\}. \quad (1.5)$$

Plugging a nonparametric estimator of $H_{X,Y|Z}(x, y|z)$ from a sample $S_n = \{(X_i, Y_i, Z_i)\}$ into (1.4) can derive the estimation of the conditional efficiency scores. Such a nonparametric estimator of $H_{X,Y|Z}(x, y|z)$ may be obtained by standard kernel smoothing, for example,

$$\hat{H}_{X,Y|Z}(x, y|z) = \frac{\sum_{i=1}^n \mathbf{I}(X_i \leq x, Y_i \geq y) K_h(Z_i - z)}{\sum_{i=1}^n K_h(Z_i - z)}, \quad (1.6)$$

where $K_h(\cdot)$ is a kernel function with bounded support, h is a vector of bandwidths that $h = (h_1, \dots, h_r)$, and r is the number of environmental variables. It is well known that the selection of bandwidth h is of critical importance in kernel smoothing. Hall et al. (2004), Bădin et al. (2010, 2012) and Li et al. (2013) propose the criterion of optimal bandwidth by least squares cross-validation (LSCV). Bădin et al. (2019) enhance the LSCV technique for bandwidth selection by a bootstrap approach.

2.3 Separability condition and second-stage regression

By construction, the conditional attainable set Ψ^z is a subset of Ψ and $\Psi = \bigcup_{z \in Z} \Psi^z$. There is a particular case, called the separability condition in Simar and Wilson (2007), where Z has no impact on the boundary of Ψ and

$$\Psi^z = \Psi \quad (1.7)$$

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 construct a separability test by new central limit theorems. Simar and Wilson (2020) enhance the separability test by a bootstrap approach.

If the separability condition in (1.7) does not hold, Bădin et al. (2012, 2014) suggest a flexible nonparametric location-scale model

$$\gamma(X, Y|Z = z) = \mu(z) + \sigma(z)\varepsilon \quad (1.8)$$

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 . The error term ε represents the unexplained part of the conditional efficiency measure and is called the idiosyncratic efficiency or pure efficiency. Mastromarco and Simar (2015) further interpret the pure efficiency as a measure of the Solow residual, which proxies for the aggregate effect of not included factors other than production inputs or conditions Z .

2.4 Tradeoff between FDH and DEA estimators

The tradeoff between the FDH and DEA estimators is nontrivial. Simar and Wilson (2015) provide a survey of the nonparametric frontier models and summarize that the FDH and DEA estimators converge to limiting distributions at rates $n^{\frac{1}{p+q}}$ and $n^{\frac{2}{p+q+1}}$ under appropriate assumptions, respectively. In either case, the convergence rate for a fixed sample size n slows down with increasing dimensionality $(p+q)$ accompanied by increasing estimation errors. This phenomenon is often referred to as the curse of dimensionality. A feasible approach to minimize the curse of dimensionality 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 industry structure, geographic restriction, and high data collection costs, dimension reduction may become an attractive solution to escape the curse of dimensionality.

2.4.1 Diagnosing the curse of dimensionality

The PCA strategy for dimension reduction amounts to a mapping from the $(p \times n)$ matrix X and the $(q \times n)$ matrix Y to $(1 \times n)$ matrices $\Lambda'X$ and $\Lambda'Y$ by pre-multiplying the first eigenvectors Λ_{x1} and Λ_{y1} of the moment matrices XX' and YY' , respectively. Though it is hard to provide theoretical guidelines of when dimension reduction should be used, Wilson (2018) proposes three diagnostics for empirical works. The first diagnostic is to compute the effective parametric sample size m . Assume a nonparametric estimator of n observations has the same level of measurement error as a parametric estimator of m observations. Equating the nonparametric convergence rate $n^{\frac{1}{p+q}}$ to the parametric convergence rate $m^{\frac{1}{2}}$, we can compute the effective parametric sample size m of the nonparametric estimator as $m \approx \lfloor n^{2\kappa} \rfloor$, where $\lfloor a \rfloor$ denotes the integer nearest a . Hence the criterion of judging the minimum sample size m in parametric estimation can be used as a comparative guideline in nonparametric estimation.

The second diagnostic is to consider the proportion of n observations that yield efficiency scores equal to one. 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 examine the ratios R_x and R_y of the largest eigenvalue of the moment matrices XX' and YY' to the corresponding sum of eigenvalues for XX' and YY' , respectively. The ratios R_x and R_y provide measures of how close the corresponding moment matrices are to rank-one, regardless of the joint distributions of inputs and outputs. For example, if $R_x = 0.9$, the matrix with dimension reduction $\Lambda'X$ contains 90% of the independent linear information in the original matrix X . In case there is evidence of excessive inputs or outputs, Wilson (2018)

proposes standardizing the matrix X or Y before PCA to ensure the inputs or outputs have the same scale.

2.4.2 Testing convexity of the attainable set

After the diagnostics of dimension reduction, the choice between the FDH estimator and the DEA estimator can be decided by data-driven hypothesis testing. Since $\hat{\Psi}_{VRS}$ in (1.3) is defined as the convex hull of $\hat{\Psi}_{FDH}$ in (1.2), Kneip et al. (2015, 2016) construct a test of convexity using new central limit theorems for the tradeoff between $\hat{\Psi}_{FDH}$ and $\hat{\Psi}_{VRS}$. In detail, the test statistic τ proposed by Kneip et al. (2016) is computed as

$$\hat{\tau} = \frac{(\hat{\mu}_{FDH,n_2} - \hat{\mu}_{VRS,n_1}) - (\hat{B}_{FDH,n_2} - \hat{B}_{VRS,n_1})}{\sqrt{\frac{\hat{\sigma}_{FDH,n_2}^2}{n_2} + \frac{\hat{\sigma}_{VRS,n_1}^2}{n_1}}} \xrightarrow{\mathcal{L}} N(0, 1), \quad (1.9)$$

where the sample S_n is randomly split into two mutually exclusive and collectively exhaustive subsets S_{n1} and S_{n2} , and $\hat{\mu}_{VRS,n_1}$, $\hat{\mu}_{FDH,n_2}$ and $\hat{\sigma}_{VRS,n_1}^2$, $\hat{\sigma}_{FDH,n_2}^2$ are the means and variances of the DEA and FDH estimators in S_{n1} and S_{n2} , respectively. The bias correction terms of $\hat{B}_{VRS,n_1}$ and $\hat{B}_{FDH,n_2}$ are computed by a generalized jackknife estimation (see Kneip et al. 2016 for details).

In practice, if the null hypothesis of convexity is rejected, the production set Ψ is nonconvex so that the FDH estimator is the only consistent estimator. Alternatively, if the null hypothesis of convexity is not rejected, the DEA estimator may be the preferred estimator because of its faster convergence rate. Under the latter situation, a further test of returns to scale can be processed for the tradeoff between the VRS-DEA and CRS-DEA estimators. However, the convexity test in (1.9) depends on randomly splitting the original sample into two independent subsamples S_{n1} and S_{n2} to calculate the bias terms, which introduces ambiguity in practice. Simar and Wilson (2020) develop a generalized bootstrap algorithm that eliminates much of this ambiguity by repeating the random splits a large number of times. The FEAR library introduced by Wilson (2008) provides convenient tools for the convexity test and the separability test.

3 Data and variable specification

The data are collected from the Sub-Industry of Semiconductors in Compustat database over 20 years ranging from 1999–2018. Since the semiconductor value chain is highly globalized, we combine data from both the Compustat North America database and the Compustat Global database to cover companies in the whole industry. We exclude photovoltaic producers, liquid crystal display

Table 1 Number of observations by business model.

Year	All	IDM	Foundry	OSAT	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
Total	5,136	1,063	394	662	3,017
Unique firms	470	83	36	63	288

manufacturers, and light-emitting diode manufacturers from the dataset, limiting the sample to only IC manufacturers. Hence, the sample includes 5,136 observations from 470 unique companies over 1999–2018. Table 1 breaks down the 5,136 observations in the global semiconductor industry by the business model from 1999 to 2018. Over half of the companies are fabless because the entry barriers, which mainly depend on CAPEX, are much lower for fabless than others. At the same time, the number of firms operating in each business model remains 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 were gradually fixed.

We specify $p = 4$ inputs (labor, measured by the number of employees (X_1); cost of goods sold (X_2); R&D expenditure (X_3); and sales & marketing expenditure (X_4)) and $q = 2$ outputs (revenue (Y_1); and shareholders' equity (Y_2)). In addition, we specify $r = 2$ environmental variables (fixed assets, measured by property, plant, and equipment (Z_1); and business model (Z_2)). The environmental variable Z_1 is a continuous variable representing the level of capital investment, which measures the fixed and tangible assets of a company that cannot quickly be sold or turned into cash in a short time. The environmental variable Z_2 is a discrete

variable, which categories the four business models, including IDM, fabless, foundry, and OSAT, into three groups. That is, the foundries and OSATs, both of which are capital intensive, are naturally combined into one group. Table 2 gives summary statistics for the original variables in 1999–2018 pooled data. Except for the discrete variable Z_2 , the distributions of the continuous variables are heavily skewed to the right, owing to several tech giants such as Intel and Qualcomm that dominate the semiconductor industry.

According to the discussion of the curse of dimensionality in Sect. 2.4.1, it is no surprise that without dimension reduction the effective parametric sample size m for either the FDH or the DEA estimator is small. For example, the effective parametric sample size m in 1999 is $125^{\frac{1}{3}} \approx 5$ for the FDH estimator and $125^{\frac{4}{3}} \approx 16$ for the DEA estimator without dimension reduction. Contrarily, the effective parametric sample size m increases to $125^1 = 125$ for the FDH estimator and $125^{\frac{4}{3}} = 625$ for the DEA estimator with dimension reduction. In the meantime, the results of eigensystem analysis in Table 3 indicate that the first principle component of the moment matrices contains most of the independent information (from the lowest ratio of 88.47% for R_x in 2014 to the highest ratio of 99.46% for R_y in 1999) in the $p = 4$ inputs and $q = 2$ outputs specified above. The evidence is clear that dimension reduction likely reduces measurement error relative to what would be obtained in the full, six-dimensional space. Therefore, all of the following results are based on data with dimension reduction.

4 Empirical results

Among studies that use nonparametric frontier approaches to estimate efficiency and benchmark performance in the semiconductor industry, the vast majority choose the DEA estimator without comparing the pros and cons between the FDH and DEA estimators. As discussed in Sect. 2, the drawback of the DEA estimator is imposing the convex assumption on the production set Ψ , while the FDH estimator is consistent regardless of the convexity of Ψ . The statistic τ_1 in Table 4 provides results of the convexity test developed by Kneip et al. (2016) and the bootstrap method developed by Simar and Wilson (2020). The efficiency scores are calculated by the hyperbolic-oriented measure defined in (1.1), and the bootstrap involves 100 splits and 1000 replications to eliminate most of the uncertainty surrounding the random splitting. At 95% confidence level, the null hypotheses of convexity are rejected for over 80% of the 20 years annual data, except for the three years of 2009, 2011, and 2012. Therefore, we use the FDH estimator for the remainder of the analysis.

Table 2 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	4.301	18.329	160.217	67.001	13,543.000
X_4	0.549	5.885	20.087	125.406	67.185	8,880.234
Y_1	0.003	47.283	161.799	1,064.110	563.655	70,848.000
Y_2	0.175	44.279	151.749	1114.748	487.729	74,563.000
Z_1	0.005	6.065	27.787	554.060	174.405	48,976.000

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.

Table 3 Eigensystem analysis by year.

Year	$R_x(\%)$	$R_y(\%)$
1999	97.28	99.46
2000	95.16	98.87
2001	96.54	99.03
2002	95.27	99.07
2003	95.06	98.84
2004	94.62	98.67
2005	95.08	98.20
2006	94.03	98.11
2007	93.55	98.23
2008	93.38	98.44
2009	93.00	97.97
2010	90.99	97.41
2011	92.68	97.32
2012	91.86	97.70
2013	91.04	98.83
2014	88.47	98.45
2015	88.52	99.09
2016	89.64	99.22
2017	89.70	98.97
2018	89.66	97.44

The statistic τ_2 in Table 4 provides results of the separability test with respect to time, using the Daraio et al. (2018) approach and the bootstrap method by Simar and Wilson (2020). Taking year as the unit of time, it is common to treat year as a discrete variable to process the separability test for each of two adjacent years. The null hypotheses are all rejected for the pairs of adjacent years, providing clear evidence that the production frontiers are different for the adjacent years. Hence each year can be treated individually in the swift-changing semiconductor industry.

Table 5 presents the results of the separability test with respect to the environmental variables Z_1 and Z_2 . The statistic τ_3 in Table 5 shows that conditional on fixed assets Z_1 ,

Table 4 Test of convexity and test of separability with respect to year.

— Test of Convexity —			—Test of Separability —		
τ_1			(w.r.t. Year) τ_2		
Year	Statistic	p -value	Year	Statistic	p -value
1999	2.222	0.011	1999–2000	8.233	0.000
2000	1.594	0.006	2000–2001	8.718	0.000
2001	1.853	0.040	2001–2002	7.425	0.008
2002	3.138	0.005	2002–2003	7.615	0.003
2003	2.901	0.001	2003–2004	8.654	0.000
2004	3.440	0.000	2004–2005	9.933	0.000
2005	3.238	0.000	2005–2006	8.561	0.004
2006	3.651	0.000	2006–2007	7.887	0.002
2007	3.915	0.003	2007–2008	8.258	0.000
2008	3.227	0.008	2008–2009	9.945	0.000
2009	2.162	0.057	2009–2010	11.178	0.000
2010	2.890	0.006	2010–2011	10.360	0.000
2011	2.102	0.088	2011–2012	9.954	0.000
2012	1.014	0.174	2012–2013	12.115	0.000
2013	1.989	0.020	2013–2014	8.786	0.002
2014	3.552	0.001	2014–2015	11.456	0.000
2015	1.452	0.041	2015–2016	10.273	0.000
2016	2.053	0.018	2016–2017	10.940	0.000
2017	4.831	0.000	2017–2018	8.351	0.001
2018	4.963	0.000	2018–1999	12.587	0.000

NOTE. Calculations of τ_1 – τ_2 include 100 splits and 1000 bootstrap replications.

over two-thirds of the years face different frontiers so that naive regressions in a second-stage analysis do not derive consistent estimation. For the separability test with respect to the discrete variable Z_2 , statistics τ_4 – τ_6 provide pairwise comparisons of the separability conditions among IDM, fabless, and foundry plus OSAT. In any case, the null hypotheses are rejected for most of the years. It is unanimous with the pros and cons analysis of the business models in Sect. 1, that the production frontiers are different and separable of the labor-intensive fabless firms, the

Table 5 Test of separability with respect to Z_1 and Z_2 .

Year	w.r.t Z_1		w.r.t Z_2					
	τ_3		τ_4		τ_5		τ_6	
	Stat	p value	Stat	p value	Stat	p value	Stat	p value
1999	0.286	0.020	0.697	0.099	5.213	0.002	3.175	0.076
2000	0.603	0.000	1.775	0.000	4.575	0.046	4.494	0.001
2001	0.070	0.353	0.883	0.054	2.949	0.238	3.658	0.034
2002	0.033	0.452	0.888	0.019	4.093	0.142	3.268	0.335
2003	0.260	0.022	0.826	0.031	3.568	0.312	4.360	0.048
2004	0.506	0.000	0.839	0.006	4.513	0.017	5.731	0.001
2005	0.572	0.000	0.913	0.017	5.043	0.006	5.282	0.000
2006	0.449	0.000	1.079	0.000	4.951	0.033	4.754	0.001
2007	0.208	0.110	0.489	0.294	4.809	0.011	4.240	0.017
2008	0.371	0.002	0.958	0.000	5.163	0.019	3.943	0.086
2009	1.022	0.000	1.177	0.000	5.711	0.001	5.539	0.004
2010	0.398	0.001	1.189	0.000	7.879	0.000	7.087	0.000
2011	0.421	0.000	1.257	0.001	5.408	0.007	5.478	0.006
2012	0.284	0.022	1.629	0.000	4.913	0.003	5.425	0.004
2013	0.505	0.000	1.540	0.000	6.745	0.000	4.950	0.045
2014	1.439	0.000	2.177	0.000	8.032	0.000	6.764	0.000
2015	0.039	0.526	1.492	0.000	6.453	0.000	5.264	0.006
2016	0.742	0.000	1.679	0.000	7.370	0.001	5.458	0.004
2017	-0.007	0.633	1.453	0.000	5.611	0.007	6.590	0.000
2018	0.027	0.519	1.617	0.000	6.247	0.000	5.871	0.002

NOTE. τ_3 denotes the test of separability w.r.t. Z_1 .

τ_4 denotes the test of separability w.r.t. Z_2 for IDM vs. fabless.

τ_5 denotes the test of separability w.r.t. Z_2 for IDM vs. foundry+OSAT.

τ_6 denotes the test of separability w.r.t. Z_2 for fabless vs. foundry+OSAT.

Calculations of τ_3 – τ_6 include 100 splits and 1000 bootstrap replications.

capital-intensive foundries and OSATs, and the IDMs which are both labor-intensive and capital-intensive, conditional on business models.

Tables 6, 7, 8 present summary statistics of the unconditional and conditional efficiency scores. The efficiency scores are measured by the hyperbolic-oriented FDH estimator and estimated under annual base as the production frontiers are different in each year. For the continuous environmental variable Z_1 , we use the LSCV proposed by Li et al. (2013) to choose the optimal bandwidth $h_{Z_1}^*$ for kernel smoothing in (1.6). By minimizing

$$CV(h) = \frac{\sum_{i=1}^n \sum_{j \neq i}^n \left[I(x_i \leq x_j, y_i \geq y_j) - \frac{\frac{1}{n} \sum_{k=1}^n I(x_k \leq x_i, y_k \geq y_j) K_h(z_i, z_k)}{\frac{1}{n-1} \sum_{k=1}^n K_h(z_i, z_k)} \right]^2}{n(n-1)}, \quad (2.1)$$

the optimal bandwidth $h_{Z_1}^*$ is equal to 922. For the discrete environmental variable Z_2 , there is no smoothing to be done. The conditional efficiency estimation under the discrete environmental variable Z_2 is simply the

unconditional efficiency estimation for a subset of the sample restricted by the discrete environmental variable Z_2 .

As shown in Table 5, the separability conditions do not hold. We use the method introduced in Bădin et al. (2012) to disentangle the impacts of the environmental variables to identify the effect on the shift of the frontier and the effect on the distribution of the inefficiencies. The effect of environmental variables on the shift of the frontier can be investigated by considering the ratio of conditional to unconditional efficiency scores. For example, based on the hyperbolic-oriented measure, such ratio can be expressed as

$$R_h(x, y|z) = \frac{\gamma(x, y|z)}{\gamma(x, y)} \quad (2.2)$$

with the boundary $R_h(x, y|z) \in (0, 1]$.

Figure 1 displays the effect of fixed assets Z_1 on the ratios of $\hat{R}_h(x, y|z_1)$ and $\hat{R}_h(x, y|z_1, z_2)$ over the 20 years from 1999–2018. Since it is not easy to visualize the three-dimensional panels in Fig. 1, we compare the mean ratios $\bar{R}_h(x, y|z_1)$ and $\bar{R}_h(x, y|z_1, z_2)$ for the different business models over time in Fig. 2. The curves of mean ratios on

Table 6 Summary statistics for unconditional efficiency scores $\hat{\gamma}(x, y)$.

Year	Min	Q1	Median	Mean	Q3	Max
1999	1.000	1.040	1.314	1.341	1.496	2.817
2000	1.000	1.156	1.352	1.502	1.715	4.001
2001	1.000	1.106	1.267	1.346	1.478	2.631
2002	1.000	1.099	1.318	1.373	1.555	2.901
2003	1.000	1.132	1.300	1.360	1.532	3.041
2004	1.000	1.125	1.267	1.365	1.516	3.490
2005	1.000	1.167	1.283	1.376	1.529	3.102
2006	1.000	1.128	1.286	1.367	1.481	3.294
2007	1.000	1.086	1.220	1.294	1.390	3.618
2008	1.000	1.064	1.213	1.281	1.441	3.519
2009	1.000	1.137	1.325	1.367	1.538	3.324
2010	1.000	1.130	1.319	1.356	1.526	2.302
2011	1.000	1.127	1.352	1.400	1.589	3.887
2012	1.000	1.137	1.343	1.408	1.593	3.243
2013	1.000	1.169	1.416	1.449	1.625	3.023
2014	1.000	1.153	1.383	1.469	1.619	3.159
2015	1.000	1.183	1.412	1.519	1.733	3.442
2016	1.000	1.152	1.362	1.459	1.667	3.298
2017	1.000	1.127	1.320	1.419	1.638	3.684
2018	1.000	1.117	1.318	1.397	1.586	3.681

Table 7 Summary statistics for conditional efficiency scores $\hat{\gamma}(x, y|z_1)$.

Year	Min	Q1	Median	Mean	Q3	Max
1999	1.000	1.006	1.282	1.317	1.466	2.817
2000	1.000	1.079	1.303	1.461	1.667	4.001
2001	1.000	1.074	1.253	1.328	1.471	2.631
2002	1.000	1.059	1.288	1.353	1.528	2.901
2003	1.000	1.112	1.281	1.347	1.526	3.041
2004	1.000	1.068	1.225	1.314	1.425	3.490
2005	1.000	1.085	1.258	1.323	1.457	3.102
2006	1.000	1.052	1.250	1.311	1.430	3.294
2007	1.000	1.054	1.195	1.270	1.369	3.618
2008	1.000	1.035	1.195	1.265	1.422	3.519
2009	1.000	1.087	1.288	1.336	1.514	3.324
2010	1.000	1.091	1.277	1.320	1.500	2.302
2011	1.000	1.086	1.330	1.379	1.585	3.887
2012	1.000	1.115	1.341	1.399	1.585	3.243
2013	1.000	1.106	1.401	1.428	1.617	3.023
2014	1.000	1.127	1.346	1.438	1.616	3.159
2015	1.000	1.141	1.372	1.487	1.706	3.442
2016	1.000	1.114	1.324	1.436	1.656	3.298
2017	1.000	1.063	1.255	1.385	1.603	3.684
2018	1.000	1.073	1.261	1.369	1.560	3.681

IDM are notably different comparing with the curves of mean ratios on fabless and on foundry plus OSAT in Fig. 2, either conditional on Z_1 or conditional on both Z_1 and Z_2 .

Table 8 Summary statistics for conditional efficiency scores $\hat{\gamma}(x, y|z_1, z_2)$.

Year	Min	Q1	Median	Mean	Q3	Max
1999	1.000	1.000	1.139	1.242	1.391	2.817
2000	1.000	1.000	1.182	1.321	1.419	4.001
2001	1.000	1.000	1.162	1.258	1.367	2.631
2002	1.000	1.006	1.219	1.300	1.465	2.901
2003	1.000	1.028	1.213	1.286	1.402	2.464
2004	1.000	1.011	1.162	1.253	1.367	3.490
2005	1.000	1.040	1.201	1.263	1.352	2.943
2006	1.000	1.017	1.190	1.262	1.341	3.083
2007	1.000	1.000	1.156	1.239	1.337	3.618
2008	1.000	1.000	1.133	1.206	1.333	2.522
2009	1.000	1.028	1.187	1.265	1.423	2.440
2010	1.000	1.017	1.190	1.251	1.430	2.220
2011	1.000	1.037	1.214	1.293	1.492	3.887
2012	1.000	1.026	1.221	1.291	1.454	2.802
2013	1.000	1.029	1.224	1.315	1.510	2.917
2014	1.000	1.037	1.213	1.303	1.459	2.787
2015	1.000	1.038	1.260	1.370	1.542	3.442
2016	1.000	1.056	1.227	1.338	1.521	2.835
2017	1.000	1.026	1.181	1.287	1.452	3.684
2018	1.000	1.011	1.188	1.280	1.447	3.225

The clues in Fig. 2 imply that the impact of capital investment shifts the production frontiers for the highly capital intensive IDM business model unambiguously. On the contrary, for fabless, foundry, and OSAT firms, capital investment has an unclear impact on the production frontiers over time.

The effects of environmental variables on the distribution of the inefficiencies can be investigated by the nonparametric location-scale model in (1.8). Bădin et al. (2012) define the pure efficiency as

$$\hat{\varepsilon}_i(t) = \frac{\hat{\gamma}_t(X_{it}, Y_{it}|Z_{it}) - \hat{\mu}(Z_{it,t})}{\hat{\sigma}(Z_{it,t})}, \quad (2.3)$$

which is the estimated resulting residual for unbalanced panel data in the second-stage regression. The pure efficiency $\hat{\varepsilon}_i(t)$ in (2.3) has cleaned off the influences of the environmental factors. Mastromarco and Simar (2015) indicate that the pure efficiency represents a new measure of the Solow residual by proxying for the aggregate effects of various factors not included in the production inputs or environmental variables. In detail, the Solow residual is calculated as

$$\widehat{\text{Solow}}_i(t) = e^{-\left(\hat{\varepsilon}_i(t) - \min_j \hat{\varepsilon}_j(t)\right)}, \quad (2.4)$$

which varies between (0,1] to help the interpretation.

Fig. 1 Effects of PP&E (Z_1) on ratios $\hat{R}_h(x, y|z_1)$ and $\hat{R}_h(x, y|z_1, z_2)$ over 1999–2018.

Fig. 2 Mean ratios $\overline{R}_h(x, y|z_1)$ and $\overline{R}_h(x, y|z_1, z_2)$ over 1999–2018.

Table 9 Mean solow residual.

Year	Filter out		Filter out	
	Z_1		Z_1, Z_2	
	Mean	SD	Mean	SD
1999	0.760	0.147	0.500	0.287
2000	0.699	0.177	0.485	0.292
2001	0.769	0.148	0.411	0.237
2002	0.764	0.155	0.326	0.214
2003	0.773	0.149	0.371	0.242
2004	0.798	0.149	0.371	0.242
2005	0.791	0.144	0.365	0.215
2006	0.796	0.151	0.429	0.251
2007	0.814	0.141	0.464	0.257
2008	0.814	0.134	0.349	0.213
2009	0.774	0.139	0.242	0.179
2010	0.777	0.129	0.204	0.154
2011	0.751	0.129	0.203	0.147
2012	0.743	0.151	0.297	0.207
2013	0.730	0.156	0.294	0.209
2014	0.726	0.162	0.279	0.179
2015	0.707	0.170	0.215	0.147
2016	0.726	0.160	0.244	0.160
2017	0.745	0.154	0.339	0.220
2018	0.751	0.151	0.319	0.214

Table 9 shows the mean Solow residual, either filtering out Z_1 or filtering out both Z_1 and Z_2 . If only filtering out the effect of Z_1 , the mean Solow residuals are relatively high and statistically significant. If filtering out the effects of both Z_1 and Z_2 , the mean Solow residuals are relatively low and statistically insignificant. Fig. 3 visualizes the changes of mean Solow residuals over time in Table 9 with confidence interval. Comparing the curve of mean Solow residuals that only filter out the impact of capital investment Z_1 with the

Fig. 3 Evolution of mean solow residual over 1999–2018

curve of mean Solow residuals that filter out the impacts of both capital investment Z_1 and business model Z_2 , the former is at a higher level with less fluctuation over time. It is demonstrated in Figure 3 that specialization by the business model not only helps to improve the pure efficiency but also helps to mitigate the impact of business cycles in the global semiconductor industry.

5 Summary and conclusions

It is well known that the barriers to entry are incredibly high in the capital-intensive semiconductor industry, especially in the manufacturing portion. Taking advantage of the economic moat by colossal capital investments and the economy of scale, the incumbent IDMs have dominated the semiconductor industry for a long time. Nonetheless, with the ever-expanding complexity of ICs and accelerated technology iterations, betting on new technologies and processes to stay ahead of the pack becomes a heavy burden even for the dominating IDMs nowadays. The raising of the fabless-foundry business model diversifies the financial risks of capital investment into specified R&D, front-end fabrication, and back-end A&T portions. The decentralized collaboration of the fabless-foundry alliance drastically reduces the barriers to entry into the globalized semiconductor value chain and leads to a flourishing of fabless design houses.

Using a flexible nonparametric conditional frontier approach, we measure the effect of business model and the constraint of capital investment in the global semiconductor industry. The empirical results are consistent with the pros and cons analysis that the entry barriers by capital investment shift the efficient frontier of the vertically integrated IDMs evidently, but have no clear effect on the efficient frontiers of the fabless-foundry models. The Solow residual estimation in a second-stage nonparametric regression

reveals that the mutually beneficial fabless-foundry partnership helps to improve the pure operating efficiency and mitigate the impact of business cycles in the global semiconductor industry.

However, the distinction between the IDM model and the fabless-foundry model is fading away. Due to the constant and costly needs 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. The complementarity between IDMs and fabless-foundry firms is similar to a symbiotic relationship in the semiconductor ecosystem, which enhances competitiveness through increasing specialization in each segment of the value chain. This ecosystem improves the overall competitiveness of the global semiconductor industry in capabilities, product diversities, and technological advancements and enables the entire semiconductor industry to thrive and prosper.

Acknowledgements We are particularly grateful to Paul W. Wilson for his advice.

Funding This study was supported by the National Social Science Fund of China (Grant No. 18ZDA093).

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bädin L, Daraio C, Simar L (2010) Optimal bandwidth selection for conditional efficiency measures: a data-driven approach. *Eur J Oper Res* 201:633–640
- Bädin L, Daraio C, Simar L (2012) How to measure the impact of environmental factors in a nonparametric production model. *Eur J Oper Res* 223:818–833
- Bädin L, Daraio C, Simar L (2014) Explaining inefficiency in nonparametric production models: the state of the art. *Ann Oper Res* 214:5–30
- Bädin L, Daraio C, Simar L (2019) A bootstrap approach for bandwidth selection in estimating conditional efficiency measures. *Eur J Oper Res* 277:784–797
- Balconi M, Fontana R (2011) Entry and innovation: an analysis of the fabless semiconductor business. *Small Bus Econ* 37:87–106
- Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manag Sci* 30:1078–1092
- Cazals C, Florens J, Simar L (2002) Nonparametric frontier estimation: a robust approach. *J Econom* 106:1–25
- Chen YS, Chen BY (2011) Applying DEA, MPI, and grey model to explore the operation performance of the Taiwanese wafer fabrication industry. *Technol Forecast Soc Change* 78:536–546
- Chu MT, Shyu JZ, Khosla R (2008) Measuring the relative performance for leading fabless firms by using data envelopment analysis. *J Intell Manuf* 19:257–272
- Daraio C, Simar L (2005) Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *J Product Anal* 24:93–121
- Daraio C., Simar, L. Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications. New York: Springer Science+Business Media, LLC. (2007a).
- Daraio C, Simar L (2007b) Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. *J Product Anal* 28:13–32
- Daraio C, Simar L, Wilson PW (2018) Central limit theorems for conditional efficiency measures and tests of the "separability" condition in nonparametric, two-stage models of production. *Econom J* 21:170–191
- Deprins, D., Simar, L., Tulkens, H. Measuring labor-efficiency in post offices. In Marchand, M., Pestieau, P., Tulkens, H., (eds.) *The performance of public enterprises: concepts and measurements*. Amsterdam: North-Holland, pp. 243–267 (1984).
- Dibiaggio L (2007) Design complexity, vertical disintegration and knowledge organization in the semiconductor industry. *Ind Corp Change* 16:239–267
- Färe, R., Grosskopf, S., Lovell, C.A.K., *The measurement of efficiency of production*. Boston: Kluwer-Nijhoff Publishing (1985).
- Flamm K (2017) Has Moore's law been repealed? An economist's perspective. *Comput Sci Eng* 19:29–40
- Hall P, Racine J, Li Q (2004) Cross-validation and the estimation of conditional probability densities. *J Am Stat Assoc* 99:1015–1026
- 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. *Organ Sci* 23:1227–1248
- 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. *Organ Sci* 24:1195–1213
- Kneip A, Simar L, Wilson PW (2015) When bias kills the variance: central limit theorems for DEA and FDH efficiency scores. *Econ Theory* 31:394–422
- Kneip A, Simar L, Wilson PW (2016) Testing hypotheses in nonparametric models of production. *J Bus Econ Stat* 34:435–456
- Li Q, Lin J, Racine J (2013) Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *J Bus Econ Stat* 31:57–65
- Lu WM, Wang WK, Lee HL (2013) The relationship between corporate social responsibility and corporate performance: evidence from the US semiconductor industry. *Int J Product Res* 51:5683–5695
- Macher JT (2006) Technological development and the boundaries of the firm: a knowledge-based examination in semiconductor manufacturing. *Manag Sci* 52:826–843

- Mastromarco C, Simar L (2015) Effect of FDI and time on catching up: new insights from a conditional nonparametric frontier analysis. *J Appl Econ* 30:826–847
- Pellens M, Della Malva A (2018) Corporate science, firm value, and vertical specialization: evidence from the semiconductor industry. *Ind Corp Change* 27:489–505
- Simar L, Wilson PW (2007) Estimation and inference in two-stage, semi-parametric models of production processes. *J Econ* 136:31–64
- Simar L, Wilson PW (2011) Two-stage DEA: caveat emptor. *J Product Anal* 36:205–218
- Simar L, Wilson PW (2015) Statistical approaches for non-parametric frontier models: a guided tour. *Int Stat Rev* 83:77–110
- Simar L, Wilson PW (2020) Hypothesis testing in nonparametric models of production using multiple sample splits. *J Product Anal* 53:287–303
- Thompson SE, Parthasarathy S (2006) Moore's law: the future of Si microelectronics. *Mater Today* 9:20–25
- Wilson PW (2008) FEAR: a software package for frontier efficiency analysis with R. *Socio-Econ Plan Sci* 42:247–254
- Wilson, P.W., Asymptotic properties of some non-parametric hyperbolic efficiency estimators. In Van Keilegom, I., Wilson, P.W., (eds.) *Exploring research frontiers in contemporary statistics and econometrics*. Berlin: Springer-Verlag, pp. 115–150 (2011).
- Wilson PW (2018) Dimension reduction in nonparametric models of production. *Eur J Oper Res* 267:349–367