Resource Allocation in Multi-divisional Multi-product Firms

BY BINLEI GONG AND ROBIN C. SICKLES*

March 27, 2018

*Binlei GONG: Department of Agricultural Economics and Management, China Academy for Rural Development (CARD), Room 1203 Qizhen Building, 866 Yuhangtang Road, Hangzhou, Zhejiang University, China 310058, (e-mail: gongbinlei@zju.edu.cn).
Robin SICKLES: Department of Economics, Rice University, Houston, TX, USA, (e-mail: rsickles@rice.edu).

Resource Allocation in Multi-divisional Multi-product Firms

This paper is concerned with specifying and estimating the productive characteristics of multidivisional multiproduct companies at the divisional level. In order to accomplish this, we augment division-level information with inputs that are imputed based on profit-maximizing allocations within each division. This study builds on work by De Loecker et al. (2016) as well as Olley and Pakes (1996), Levinsohn and Petrin (2003) and Ackerberg et al. (2015), and extends this work by lifting a key assumption that single- and multi-product/division firms have the same production technique for the same product/segment. We estimate the production function and impute input allocations simultaneously in the absence of this key assumption as well as the constant share constraint of the input portfolio. Finally, our approach is applied to estimate the division-specific productivity of firms that compete in five segments of the global oilfield market.

Key words: Multi-divisional Multi-product firm, heterogeneous technologies, productivity and performance, global oilfield market

JEL Codes: D2 L2 Q4 C5

1. Introduction

In productivity analysis, access to input and output data is essential to estimate the production function. However, detailed input and output data, especially on the input side, is often incomplete or inaccessible to researchers. An example relates to the estimation of productivity in an industry. More and more integrated companies have a multi-divisional form (MDF or M-form) and have footprints in multiple segments. The companies report total output and input data, and possibly division-level output as well, but not division-level inputs. Such information is not sufficient to reflect the company's resources invested in the targeted segment and to estimate the segment-specific production function. In this case, the input allocations across divisions/segments for each firm need to be collected or imputed to properly study the M-form firms and the segments in which they have footprints.

Such a problem also exists in modeling productivity by country/region. Many models of world productivity have been proposed in the literature, but most assume a common production function for the entire economy. However, different sectors or industries within an economy clearly utilize different technologies. Some industries are more labor-intensive, while others are more capital-intensive. Many economists overlook this economic structural heterogeneity while others choose to ignore such issues due to the lack of industry-level data.

These examples illustrate two challenges that arise as a result of unobservable input allocations, or the "black box" within multi-divisional organizations. For each multi-divisional organization, the productivity levels measured by standard production methods do not provide the difference in performance across divisions. For each segment, the production function cannot be estimated, since the inputs allocated in that specific segment are unknown for those multi-divisional firms. The focus of this paper is on these two issues, analyzing the divisional productivity of multi-divisional firms and segment-specific production functions for relevant segments.

The setup of multi-divisional firms and multi-product firms in economics and finance studies has been very similar, as most conglomerates adopt the M-form, wherein each division presents one product line that produces one output in order to produce multiple products. Many studies (Teece (1998), McEwin (2011), among them) call such a business organization a multi-divisional multi-product firm. We study such multi-divisional firms and multi-product firms and also analyze standard assumptions and conventions adopted from the literature in order that appropriate methods for modeling and estimating the technologies of such M-firms can be implemented.

It is worth noting the relationship among the three words that this research frequently uses: product, division, and segment. Suppose an energy firm has two divisions - one producing oil and the other producing natural gas; it is said that this is a multi-divisional multi-product firm. Such a company has an oil division that generates the oil product in the oil segment, as well as a natural gas division that generates the natural gas product in the natural gas segment. Therefore, the words product, division, and segment are a one-to-one match. For one product, the matching division is the department or branch of the firm that generates this product, while the matching segment refers to the market for that product.

De Loecker *et al.* (2016) point out that input allocations of multi-product firms are typically unobservable. According to their setup, these multi-product firms can be treated as multi-divisional multi-product firms. They exploit data on single-product firms to estimate the production function for each product line, assuming that the production function is the same for single- and multi-product firms that generate the same product. The production functions estimated from single-product firms allow them to back out allocations of inputs across products within a multi-product firm.

De Loecker *et al.* (2016) also note that the key assumption in their work—the production techniques for single- and multi-product firms are the same for one product line—is already implicitly employed in all previous work that pools data across single- and multi-product firms (e.g., Olley and Pakes (1996) and Levinsohn and Petrin (2003)). This is of course a testable hypothesis and we examine its validity in our analysis below. Another issue in this line of empirical study is that there may not be enough observations if there are only a few single-product firms, but many more multi-divisional firms in a product line. For example, two of the five segments in our study of the global oilfield market are dominated by multi-divisional firms. There are only seven and eight single-division firms in the completion and production segments, respectively. Another example concerns world productivity analysis, where almost all countries are multi-product/division “firms” since they all have more than one industry.

In order to implement their model, De Loecker *et al.* (2016) also assume that the share of a firm’s inputs allocated to a given product line is constant, and thus independent of the input type. For example, if a product line uses 30% of the firm’s labor, it also uses 30% of the firm’s capital. However, this assumption (assumption 4 in their paper) may not be consistent with the product-specific production function assumption (assumption 1 in their paper). For a multi-product firm, some product lines are relatively labor-intensive and require a higher labor share in the input portfolio, while other product lines are relatively capital-intensive and need a higher capital share in the input portfolio. We lift this assumption in our analysis and let the data speak to the constancy of a firm’s input shares allocated to a given product line.

Valmari (2016) also studies product-specific production function of multi-product firms. He also emphasizes the two key assumptions in De Loecker *et al.* (2016) as we mentioned above, and criticizes that the share of a profit maximizing firm’s inputs allocated to a given product line cannot be constant. However, Valmari (2016) also assumes single- and multi-product firms use similar product-specific technologies and highlights that this is the key assumption that enables De Loecker *et al.* (2016) to estimate the production functions using data on single-product firms only without simultaneously solving for the unobservable input allocations.

Thus our study makes three central contributions. First, we test the key assumption that previous studies rely on, i.e., that single- and multi-product/division firms have the same production technique for the same product/segment. Second, we estimate the production function and impute input allocations simultaneously in a setting in which the share of the input portfolio is not constrained to be constant across segments and the single- and multi-product/division firms are not assumed to have the same production technique for the same product/segment. These are two assumptions relied on by De Loecker et al. (2016). Third, we study the productivity and efficiency of a major strategic industry, the global oilfield market, also referred to as the oil and gas service industry, using methods that address these aforementioned issues. Our empirical study suggests that the key assumptions typically made in such empirical settings are not always valid and may lead to erroneous conclusions about the productivity and efficiency of firms in this key industry. Moreover, our division-level analyses provides valuable information for firms' mergers and acquisitions (M&A) decisions and are consistent with and predictive of recent major business decisions by several of the large companies in the oil and gas service industry.

To accomplish this we develop a model to explore resource allocation across divisions within a multi-divisional firm when division-level outputs, total inputs, and segment average input prices are available. The undistorted input allocations must simultaneously satisfy two systems of equations including segment-specific stochastic production functions and the maximization of the mathematical expectation of profit for each multi-divisional firm.

An iterative method is used to jointly solve these two systems and impute undistorted input allocations. Then, stochastic frontier analysis is used to estimate the segment-specific production frontiers. Since there is one frontier for each segment, this method is called the segment-specific production frontier (SSPF), while the standard stochastic frontier analysis assuming similar production frontiers across segments is called the single production frontier (SPF) approach.

This study tests the accuracy of the estimation using panel data for twenty-two Organization for Economic Co-operation and Development (OECD) countries from 2000 to 2006 where the actual input allocations are observed. The imputed undistorted input allocations are closer to the actual levels than the three competing estimations. Moreover, the iteration converges to the same point using different initial guesses.

This SSPF approach is used to explore the unobservable input allocations of multi-divisional firms in the global oilfield market, which is composed of five segments. This study finds no evidence of scale economies, as the estimated production function is constant returns to scale. However, we do find economies of scope, since the multi-divisional firms on average have higher productivity than single-divisional firms. Moreover, the single- and multi-divisional firms in the capital equipment segment have different production functions, which provide evidence that a key assumption in this literature is not always valid. Multi-divisional firms may have very different efficiency levels across segments and they prefer to invest in more efficient segments. This SSPF approach can estimate division-level and firm-level efficiency as well as average efficiency for each segment.

The remainder of the paper is structured as follows. Section 2 introduces the methodology of the SSPF approach. Section 3 presents our empirical application on the global oilfield market. Section 4 consists of the conclusions drawn. Data construction, imputation algorithms, as well as robustness checks and extended results based on the translog functional form in addition to those based on the Cobb-Douglas that we report in the main body of our paper are found in the Appendices.

2. The Model

2.1 Multi-divisional Firms and Multi-product Firms

The multi-divisional form is an organizational structure that separates a company into individual divisions based on location or products. A multi-divisional (form) firm is essentially divided into semiautonomous divisions that have their own unitary structures, and division managers are responsible for their own production and for maximizing profit division-wide. There is a central office that develops overall strategies to maximize overall company-wide profit. Thus the headquarters are in charge of overall strategic decisions, while division managers are allowed to make their own operational choices.

Multi-divisional firms, or conglomerates, have been popular in the United States since the 1960s and currently account for over 60% of the book assets and market equity of S&P 500 firms (Duchin *et al.*, 2015). This type of firm is mostly studied in corporate finance and management research. The main focus of the corporate finance and management research on multi-divisional firms has been on comparisons between the U-form (unitary form) and the M-form (multi-divisional form), the principal-agent problem between headquarters and division managers, and transfer pricing and tax minimization.

The M-form provides many advantages. First, firms can be more productive by diversifying under the assumption of neoclassical diminishing returns within the industry (Maksimovic and Phillips, 2002). Second, a large company has higher brand value and spillover effects that can be shared across divisions. Third, division managers can more efficiently handle the day-to-day operations of their own divisions. Fourth, even if some units of the firm fail, the other divisions can still be productive and profitable, which guarantees a more versatile, less risky, and more enduring organization. Fifth, diversified firms can allocate capital better than typical external market sources due to their superior inside information (Williamson, 1975). The headquarters also can expand highly productive divisions and abandon low producing divisions more effectively than the market (Stein, 1997). The M-form combines the advantages of a distinct brand and economies of scale for a large conglomerate, while maintaining the operational flexibility of a small firm.

The M-form also has disadvantages. The headquarters of multi-divisional firms may not be able to manage the various different businesses as effectively as single-division firms. In addition to coordination problems, conglomerates suffer from agency problems, wherein division managers may report incorrect information to maximize profits division-wide, rather than firm-wide. Bureaucracy and a lack of managerial focus are also frequent problems. Moreover, a lack of transparency, such as the black box of input allocations across divisions that this study discusses, also discourages investors. As a result, the stock market may value a diversified group of businesses and assets at less than the sum of its parts, which is referred to as the "conglomerate discount." Many scholars (Berger and Ofek, 1995; Lang and Stulz, 1994) find that multi-divisional firms have lower productivity levels than single-division firms.

It is important to clarify the definitions and terminology used in the literature on multi-product and multi-divisional firms. Multi-product firms produce multiple goods and face the problem of how to allocate inputs in each product line. Multi-divisional firms are those that have multiple semi-autonomous divisions and face the problem of how to allocate inputs in each division. On the one hand, multi-product firms take either a unitary form (e.g., a small farm that produces vegetables and fruits) or a multi-divisional form (e.g., General Electric operates through multiple divisions: power & water, oil & gas, aviation, healthcare, transportation, capital, and energy management). On the other hand, a multi-divisional form can be adopted by either single-product firms (e.g., an electricity company owning several power plants) or multi-product firms (e.g., the GE case mentioned above).

Since the multi-divisional form became popular in the 1960s, more and more studies focus on these organizational structures, especially for multi-product firms. Williamson (1975) theorizes the information-processing advantages that a multi-product firm can achieve by deploying a multi-divisional form. Teece (1981) concludes that large multi-product firms that adopted a multi-divisional form always perform better than those with a unitary form. Although more and more large firms are both multi-divisional and multi-product, some differences between the two were revealed in previous studies.

Earlier studies on multi-divisional firms largely have appeared in the finance and management literatures and tend to focus on the principal-agent problem (Grossman and Hart, 1983; Hölmstrom, 1979, 1982; Ross, 1973), since divisional managers are responsible for their own profit maximization, which may conflict with firm-level profit maximization. Multi-product firm studies largely are found in the literatures on empirical industrial organization, productivity and efficiency, and international trade, and center on product choices: how to concentrate on the most productive goods and drop the least productive goods in global markets (Baldwin *et al.*, 2005; Bernard *et al.*, 2011) and in national markets (Bernard *et al.*, 2010; Broda and Weinstein, 2010).

As Valmari (2016) mentioned, most of the existing studies of multi-product firms assume that all of the company's products is generated with a firm-level technology. In recent years, both multi-divisional and multi-product studies have begun to consider division-specific or product-specific production functions, rather than a unique production function at the firm level. Maksimovic and Phillips (2002) find evidence to indicate that the majority of U.S. conglomerates grow efficiently and therefore study resource allocation in the absence of an agency problem. In this paper we follow Maksimovic and Phillips (2002) and study multi-divisional firms in the absence of an agency problem. A more detailed discussion of this issue is discussed in section 2.3.

Differences between multi-divisional firms and multi-product firms are minimized when the assumptions of division (product)-specific production and the absence of an agency problem are made. The lack of an agency problem to which we refer speaks to the optimal decisions of divisional managers that are consistent with firm-level maximization. However, that every division's goal is to do their best the best may not correspond to frontier behavior. Most conglomerates have both multi-product and multi-divisional characteristics and set each product line as a division, with divisional managers making operational decisions. Although they may be attempting to profit-maximize those goals may not play out in the real world.

2.2 General Model

Suppose companies can have footprints in one or multiple segments in a "M Inputs – N Outputs/Segments – T periods" economy. We outline below methods to model such a production process and to estimate the productivity and efficiency of such enterprises, where stochastic frontier analysis[2] is used as the production method to nest the neoclassical production model that assumes fully efficient allocations. The usual approach to model such a technology, wherein one production frontier applies to all the firms, is denoted as the single production frontier (SPF, hereafter) approach, while our new model, where segment-specific production frontiers exist, is denoted as the segment-specific production frontier (SSPF, hereafter) approach.

2.2.1 Single Frontier Analysis.

The canonical stochastic frontier model for panel data, or what this study calls the SPF estimator, specifies production for an individual company i at time t as

$$Y_{it} = f(X_{it};\beta_0)\exp(\tau Z)\exp(\nu_{it})\exp(-u_{it}), \qquad (1)$$

where Y_{it} is the aggregated output of company (firm) *i* at time *t*; $X_{it} = (X_{it}^1, X_{it}^2, \ldots, X_{it}^M)$ denotes the vector of the M types of inputs; $f(X_{it};\beta_0)\cdot\exp(\tau Z)$ is the average production frontier, where $f(X_{it};\beta_0)$ is the time-invariant part of the average production function, $\beta_0 = (\beta_{01}, \beta_{02}, \ldots, \beta_{0M})$ is a vector of technical parameters to be estimated, and Z is a vector that contains a group of year dummy variables and shift the production frontier over time with corresponding coefficients τ; $\exp(\nu_{it})$ is the stochastic component that describes random shocks affecting the production process. The stochastic term ν_{it} is standard idiosyncratic noise assumed to be normally distributed with a mean of zero and a standard deviation of σ_ν while u_{it} is a one-sided stochastic term with a positive mean $\mu_{it} > 0$. Technical efficiency (TE) is defined as the ratio of observed

[2] Stochastic frontier analysis (SFA) is a method of economic modelling widely used in productivity and efficiency analysis. See Kumbhakar and Lovell (2003).

output to the expected maximum feasible output and is $TE_{it} = \exp(-u_{it})$. $TE_{it} = 1$ or $\mu_{it} = 0$ shows that the *i*-th individual allocates at the production frontier and obtains the maximum feasible output at time *t*, while $TE_{it} < 1$ or $u_{it} > 0$ provides a measure of the shortfall of the observed output from the maximum feasible output. This study uses the error components specification with time-varying efficiencies (Battese and Coelli, 1992), where $u_{it} = \exp(-\eta(t - T)) * u_i)$. To sum up, the input coefficients β_0 of the production function in Eq. (1) are time-invariant, while the firm-specific intercept term is shifted by a common time varying component. This study will test if the coefficients β_0 of the production function also change over time in the empirical analysis.

2.2.2 Segment-Specific Frontier Analysis.

Eq. (1) assumes a unique production function across segments and is not a useful representation of technologies that are segment-specific. A meta-frontier function is a widely used method to investigate firms in different groups that may not have the same technology (Battese and Rao, 2002). The meta-production function was first established by Hayami (1969) and Hayami and Ruttan (1970), which is treated as the envelope of commonly conceived neoclassical production functions (Hayami and Ruttan, 1971). This method is attractive theoretically since the producers in the same group have potential access to the same technology but there exist technology gaps across groups. It is worth noting that the technology gaps among different groups in the original meta-production studies are often caused by geographic reasons. For example, Mundlak and Hellinghausen (1982) and Lau and Yotopoulos (1989) adopt this method to compare agricultural productivity across countries.

The meta-frontier methodology is introduced in stochastic frontier analysis and becomes a function that envelopes all the frontiers of groups using different technologies (Battese *et al.*, 2004). For example, Sharma and Leung (2000) use stochastic meta-frontier model to estimate the efficiency of aquaculture farms in South Asian nations. The technology gaps between units across regions or countries are the main causes to use stochastic meta-frontier models (Huang *et al.*, 2010).

The concept of a group-specific frontier in the meta-frontier literature provides a useful paradigm that allows different segments to have different frontiers as they possess different technologies. In this paper, we assume the production frontier is segment-specific. Moreover, all the players in the same segment, including both single-division firms and the divisions of the multi-divisional firms, are assumed to have potential access to the same technology at this point. Therefore, different divisions in a multi-divisional firm that has footprints in multiple segments are under different frontiers. In a "M Inputs – N Products/Segments – T periods" economy, each equation in the system of N equations describes the production techniques for the corresponding segment. This system can be written as:

$$\begin{cases} Y_{i1t} = f_1(X_{i1t}; \beta_1) \exp(\tau_1 Z_1) \exp(\nu_{i1t}) \exp(-u_{i1t}) \\ \vdots \\ Y_{iNt} = f_N(X_{iNt}; \beta_N) \exp(\tau_N Z_N) \exp(\nu_{iNt}) \exp(-u_{iNt}) \end{cases}, \quad (2)$$

where Y_{ijt} represents the observed scalar output and X_{ijt} are vectors of unobserved inputs of firm i in segment j at time t, respectively. $f_j(X_{ijt}; \beta_j)$ is the heterogeneous production frontier for segment j, where $\beta_j = (\beta_{j1}, \beta_{j2}, \dots)$ is a vector of segment-specific technical parameters; $Z_j = (Z_{j2}, Z_{j3}, \dots, Z_{jT})$ are vectors of year dummy variables to control for production frontier changes across time and $\tau_j = (\tau_{j2}, \tau_{j3}, \dots, \tau_{jT})$ vectors the coefficients of the year dummy variables; and $u_{ijt} = (-\eta(t-T)) * u_{ij}$ is the time-variant efficiency indicator. ν_{ijt} is the noise that is considered normally distributed with a mean of zero and a standard deviation of $\sigma_{\nu j}$. The data pooled into the j-th equation in Eq. (2) depends on the validity of a key assumption. This key assumption is that single- and multi-product/division firms have the same production technique for the same product/segment). If this assumption made by De Loecker et al. (2016) is adopted, then all of the divisions for firms that are single division firms and all corresponding divisions of multi-product firms operating in the same segment (j) are included in the pooled regression model. However, we can test if this assumption is valid and if a single-division firm and the corresponding division of the multi-divisional firm that operate in the same segment (of the five segments: exploration, drilling, completion, production, capital equipment and other) have different production technologies then we can remove them from the pooled regression and allow each to have their own production regression model.

The SSPF approach can predict division-level efficiency for multi-divisional firms ($\widehat{TE}_{ijt} = e^{-u_{ijt}}$) and firm-level efficiency for single-division firms. The firm-level efficiency for the multi-divisional firm i at time t, $\widehat{TE}_{it}$, is the weighted average of its efficiency in each division.

$$\widehat{TE}_{it} = \sum_j \left[\frac{R_{ijt}}{R_{it}} \widehat{TE}_{ijt}\right],$$

where R_{it} is the firm-level revenue for firm i at time t and R_{ijt} is the division-level revenue for firm i in division j at time t.

Compared with a standard SPF, the SSPF approach considers the heterogeneity in production frontiers across segments and can derive divisional level efficiency. However, the unobserved input allocations at the divisional level need to be imputed. We next present the assumptions, the methodology, and a method to assess the accuracy of the imputations of this imputation approach.

2.3 Assumptions

Before building the model to impute input allocations across divisions, we discuss several key assumptions we must make.

Assumption 1: *Maksimovic and Phillips (2002) Agency problems are absent and thus the multi-divisional firms are profit-maximizing at the firm-level, have undistorted input allocations, and report actual and accurate price and quantity information to the headquarters.*

Each semiautonomous division has its own unitary structure and managers are responsible for their own division, which may be inconsistent with the goal of the entire company. This principal-agent problem in multi-divisional firms has been studied by many economists since Ross (1973) and Hölmstrom (1979). Hölmstrom (1982) and Grossman and Hart (1983) solve the moral hazard of the division managers by designing a compensation scheme so that the semiautonomous divisions would not deviate from the equilibrium. Maksimovic and Phillips (2002) analyze optimal firm size, growth, and resource allocation by using a profit-maximizing neoclassical model in the absence of an agency problem. They find evidence to indicate that majority of U.S. conglomerates grow efficiently and therefore study resource allocation in absence of an agency problem. We follow the ideas of Maksimovic and Phillips (2002) by assuming no limit on the inputs one company can employ, indicating that the firm-level profit maximization problem is the first best and that divisions do not have to compete for inputs. If this assumption is made then division managers are willing to truthfully report their actual operational conditions and not overstate divisional profits, anticipating that additional resources would be allocated to their division, and the upstream profit-maximizing headquarters therefore know what actual profits are. The empirical results in Maksimovic and Phillips (2002) indicate that the actual resource allocations for most conglomerates are generally consistent with the undistorted/optimal allocations of resources across divisions to maximize profit. Therefore, the undistorted input allocations for profit-maximizing multi-divisional firms can be used to impute the unobserved input allocations that companies are not required to (and thus do not) report at the division level. It is worth noting that we allow technical inefficiency in the production function and hence in the profit function, which determines the undistorted input allocations. Companies with different levels of productivity (technical efficiency) have different levels of input usage for different divisions.

Assumption 2: *Spillover effects on input prices are division-invariant functions of firm size. Firm size can be measured by total costs.*

Potential reasons for the presence of the multi-divisional form are economies of scale, as the firms become larger and more diversified, economies of scope, and spillover effects. There are of course also possibilities for diseconomies due to coordination issues and other constraints on managing a heterogeneous set of divisions. Economies of scope for multi-divisional firms would suggest that such firms enjoy higher productivity and efficiency than single-division firms, which is considered in De Loecker et al. (2016) and also allowed in this paper. In our empirical work below we do not find evidence of scale economies. However, we do find economies of scope.

Spillover effects can be another important driver of the multi-divisional form for firms in this industry. Sönmez (2013) presents a literature survey on firm survival and concludes that the survival probability of a firm increases with firm size. On the one hand, Burdett and Mortensen (1998) and Coles (2001) discuss how firms trade off higher wages against a lower quit rate when they expect a higher likelihood of survival. They also find that in equilibrium, large firms with many employees pay high wages relative to small firms with few employees that pay lower wages. On the other hand, many studies have found that larger firms pay less for capital (Botosan, 1997; Gebhardt *et al.*, 2001; Petersen and Rajan, 1994; Poshakwale and Courtis, 2005; Reverte, 2012), since large firms have a higher survival probability and lower risk. As a result, larger firms are more capital-intensive than smaller ones due to factor prices differing by firm size (Söderbom and Teal, 2004), again reinforcing our assumption that the price of capital and the price of labor are affected by firm size.

The effect of firm size on input prices is similar across divisions. Duchin *et al.* (2015) point out that the spillovers (within-firm peer effects) in compensation and capital expenditure are equally strong across related and unrelated divisions within a firm. If the effect is not equal across segments, less-benefiting divisional managers will lobby and negotiate with the headquarters. Levin (2002) presents a model where the actions of the firm toward a group of employees affect the expectations of the other employees, suggesting the importance of wage equity across divisions. Multi-divisional firms thus would pay the same premium/discount based on the segment average compensation for employees to maintain a similar quit rate across divisions. The capital prices across divisions within the same multi-divisional firm reflect the same company risk over and above the segment-specific systematic risk. Therefore, the capital prices of different divisions within the same multi-divisional firm have the same premium/discount on the basis of each segment's benchmark interest rate that depends on the systematic risk of the segment. We assume an equal effect of firm size across divisions and thus that the premium/discount on input prices due to spillover effects is equal across divisions.

Assumption 3: *Capital is chosen prior to the realization of the production, while labor and intermediate inputs (if included as an input) are chosen at the time production takes place*.

This assumption specifies the timing of input choices, which is consistent with classic productivity frameworks in Olley and Pakes (1996) and Levinsohn and Petrin (2003). In the framework of Olley and Pakes (1996), the labor used in time t is chosen when the productivity shock at time t is observed, while the capital used in time t is chosen at time $t-1$ because $K_t = (1-\delta_{t-1})K_{t-1} + I_{t-1}$, where K_t, δ_t, and I_t are capital, depreciation rate and investment, respectively, at time t. Ackerberg *et al.* (2015) describe those inputs that are chosen at the time production takes place as non-dynamic inputs. In Levinsohn and Petrin (2003), labor and intermediate inputs are both non-dynamic inputs, while capital is not. We allow labor and intermediate inputs (if included as an input) to be non-dynamic inputs (see Ackerberg, Caves, and Frazer, 2015, pg. 2421).

2.4 Imputation Algorithm

Let R_{ijt} be the division-level revenue for firm i in division j at time t; I_{it} be the information set for firm i at time t; $E[R_{ijt}|I_{it-1}]$ be the expected revenue for firm i in division j at time t given the information set at time $t-1$; X_{ijt}^k be unobserved k-th input of firm i in division j at time t for all $k < M$, and the last input, X_{ijt}^M, be the division-level capital input; p_{jt} be the average price of the output in segment j at time t; c_{jt}^k be the average price of the k-th input in segment j at time t; $r_{jj't}^k$ be the ratio of average price for the k-th input between segments j and j' at time t; $W(i,t)$ be a subset of $W = (1,2,\ldots,N)$, indicating the segments where firm i at time t has a footprint; $\beta_j = (\beta_{j1}, \beta_{j2}, \ldots)$ be a vector of technical parameters in segment j's production function; $\nu_{ijt} \sim N(0, \sigma_{\nu j}^2)$ be the white noise in segment j's production function; τ_{jt} be the coefficient of time t's dummy variable in segment j's production function; $A_{ijt} = \exp(\alpha_j + \tau_{jt} - u_{ijt})$ be the (total-factor) productivity in segment j's production function for firm i; δ_{jkl} be the coefficient of the intersection between the k-th input and l-th input in segment j's production function if the function takes Transcendental Logarithmic form; $\log(x_{ijt'})|\log(x_{ijt}) \sim N(\log(x_{ijt}) + \mu_{xj}, \sigma_{xj}^2)$ for all $x = p, A, X^k$, which means the growth rate of x is normally distributed with a mean of μ_{xj} and a standard deviation of σ_{xj}.

Theorem 1: *Under Assumptions 1-3, the undistorted input allocations in multi-divisional firms satisfy: (i)*

$$\begin{cases} \dfrac{\frac{\partial E[R_{ijt} \mid I_{it}]}{\partial X_{ijt}^{k}}}{\frac{\partial E\left[R_{ij't} \mid I_{it}\right]}{\partial X_{ij't}^{k}}} = \dfrac{c_{jt}^{k}}{c_{j't}^{k}} = r_{jj't}^{k}, \forall i, t, \forall k < M, \forall j, j' \in W(i,t) \\ \dfrac{\frac{\partial E[R_{ijt} \mid I_{it-1}]}{\partial X_{ijt}^{M}}}{\frac{\partial E\left[R_{ij't} \mid I_{it-1}\right]}{\partial X_{ij't}^{M}}} = \dfrac{c_{jt-1}^{M}}{c_{j't-1}^{M}} = r_{jj't-1}^{M}, \forall i, t, \forall j, j' \in W(i,t) \\ \sum_{j \in W(i,t)} X_{ijt}^{k} = X_{it}^{k}, \forall i, k, t \end{cases} . \quad (3)$$

(ii) When the production function has a Cobb-Douglas form, (i) derives

$$\begin{cases} X_{ijt}^{k} = \dfrac{\beta_{jk} R_{ijt}}{\sum_{j' \in W(i,t)} \beta_{j'k} R_{ij't} r_{jj't}^{k} \, exp\left(\nu_{ijt} - \nu_{ij't} + 0.5\sigma_{\nu j'}^{2} - 0.5\sigma_{\nu j}^{2}\right)} X_{it}^{k}, \ \ \forall k < M \\ X_{ijt}^{M} = \dfrac{\beta_{jM} R_{ijt}}{\sum_{j' \in W(i,t)} \beta_{j'M} R_{ij't} r_{jj't'}^{M} \left(\frac{\gamma_{j'}}{\gamma_j}\right)\left(\frac{z_{ij't}}{z_{ijt}}\right) exp\left(\nu_{ijt} - \nu_{ij't} + 0.5\sigma_{\nu j'}^{2} - 0.5\sigma_{\nu j}^{2}\right)} X_{it}^{M} \end{cases} , \quad (4)$$

where

$$z_{ijt} = \frac{p_{jt-1}}{p_{jt}} \frac{A_{ijt-1}}{A_{ijt}} \prod_{k=1,\dots,M-1} \left(\frac{X_{ijt-1}^{k}}{X_{ijt}^{k}}\right)^{\beta_{jk}} ,$$

and

$$\gamma_j = \exp\left(\mu_{pj} + \mu_{Aj} + \sum_{k=1}^{M-1} \beta_{jk} \mu_{X^k j} + \frac{\sigma_{pj}^{2} + \sigma_{Aj}^{2} + \sum_{k=1}^{M-1} \beta_{jk}^{2} \sigma_{X^k j}^{2}}{2}\right).$$

(iii) When the production function has a Transcendental Logarithmic form, (i) derives

$$\begin{cases} \dfrac{\frac{\partial E[R_{ijt} \mid I_{it}]}{\partial X_{ijt}^{k}}}{\frac{\partial E\left[R_{ij't} \mid I_{it}\right]}{\partial X_{ij't}^{k}}} = \dfrac{(\beta_{jk} + \sum_{l=1}^{M} \delta_{jkl} ln X_{ijt}^{l}) R_{ijt} X_{ij't}^{k} \, exp\left(\nu_{ij't} + 0.5\sigma_{\nu j}^{2}\right)}{(\beta_{j'k} + \sum_{l=1}^{M} \delta_{j'kl} ln X_{ij't}^{l}) R_{ij't} X_{ijt}^{k} \, exp\left(\nu_{ijt} + 0.5\sigma_{\nu j'}^{2}\right)} = r_{jj't}^{k}, \ \forall k < M \\ \dfrac{\frac{\partial E\left[R_{ijt} \middle| I_{it-1}\right]}{\partial X_{ijt}^{M}}}{\frac{\partial E\left[R_{ij't} \mid I_{it-1}\right]}{\partial X_{ij't}^{M}}} = \dfrac{z_{ijt}(\beta_{jk} + \sum_{l=1}^{M} \delta_{jkl} ln X_{ijt}^{l}) R_{ijt} X_{ij't}^{M} \exp(\nu_{ij't} + 0.5\sigma_{\nu j}^{2}) \gamma_j}{z_{ij't}\left(\beta_{j'k} + \sum_{l=1}^{M} \delta_{j'kl} ln X_{ij't}^{l}\right) R_{ij't} X_{ijt}^{M} \exp\left(\nu_{ijt} + 0.5\sigma_{\nu j'}^{2}\right) \gamma_{j'}} = r_{jj't-1}^{M} \\ \sum_{j \in W(i,t)} X_{ijt}^{k} = X_{it}^{k}, \ \ \forall i, k, t \end{cases} . \quad (5)$$

Appendix A provides a detailed imputation algorithm: Eqs. (3) - (5) are derived from the first order condition of profit maximization problem and can be solved jointly with the segment-specific production functions using an iterative method. Finally, the iterative method can impute undistorted input allocations, which can be applied in the SSPF (through Eq. (2)) to estimate the efficiency at divisional level.

We test whether this imputation method can deliver accurate estimations of input allocations using a panel of data for OECD countries where actual input allocations are observed. Details are in Appendix B and results indicate that our imputation method provides estimated input allocations that are closer to the actual allocations than the other three competing imputation methods utilized in the literature.

2.5 Endogeneity Problem

Olley and Pakes (1996) revisit the endogeneity problem inherent in production function studies due to input choices being determined by some factors that are unobserved by the econometrician, but observed by the firm (Ackerberg *et al.*, 2015). Such factors include the firm's beliefs about its productivity and efficiency. This problem is more pronounced for inputs that adjust rapidly (Marschak and Andrews, 1944), such as the oilfield market, where the decisions of the companies depend heavily on exploration and production (E&P) spending from the oil and gas firms (affected by oil price) and the business cycles. Many companies divest capital and cut headcount aggressively when oil prices fall and ramp up their use of capital and labor services when the price of oil rises. This of course leads to familiar problem of endogenous inputs in the production function, which leads to biased OLS estimates.

One of the solutions to the endogeneity problem in estimating production functions is to specify the unobserved factors as fixed effects and utilize a within transformation. In standard productivity studies utilizing linear in parameters models, such as the linear in logs Cobb-Douglas (C-D) or more flexible forms such as the Transcendental Logarithmic (T-L), the fixed-effects treatment is straightforward. However, the consistency of the fixed effects estimator rests on the assumption that unobserved factors are constant across time. An alternative advocated by Olley and Pakes (1996) is to use observed investment to control for unobserved productivity or efficiency shocks. Levinsohn and Petrin (2003) extend the idea by using intermediate inputs instead of investment as instruments. They claim the benefit is strictly data-driven, since many datasets have significant amounts of observations with zero investment, which makes investment an invalid proxy. However, as Ackerberg *et al.* (2015) note, both of the models suffer from the collinearity problems in the first step, such that the coefficients of the exogenous inputs may not be identified. Moreover, data on intermediate inputs is not available in the sample of oilfield firms we analyze in the empirical study we turn to in the next section.

The third approach, which is also the most widely used method to solve the endogeneity problem, is instrumental variables (IV) estimation. Amsler *et al.* (2016) review Two-Stage Least Square (2SLS) and applied a Corrected 2SLS (C2SLS) to solve the endogeneity problem in stochastic frontier analysis when the production function is linear in parameters, such as with the C-D production function. For C2SLS, the first step is to estimate the model by 2SLS and derive the residuals using the instruments. In the second step, these 2SLS residuals are decomposed using the maximum likelihood method, just as in classic stochastic frontier analysis. A somewhat similar two-step procedure is developed by Guan *et al.* (2009). For a T-L production function, however, Amsler *et al.* (2016) suggest that the control function method is more efficient than the C2SLS method. Moreover, they explain how to reduce the number of instrumental variables needed, while still yielding consistent estimators based on the control function method.[3]

We use the C2SLS method for the linear C-D production function and the control function method for the nonlinear T-L production; both are recommended in Amsler *et al.* (2016). The control function method can also be used to test the strict exogeneity of the inputs using t-tests for the significance of the reduced form residuals. For the selection of instrument variables, Levinsohn and Petrin (2003) mention that the potential instruments include input prices and lagged values of input use. Lagged values of inputs are valid instruments if the lag time is long enough to break the dependence between the input choices and serially correlated shocks. Blundell and Bond (2000) and Guan *et al.* (2009) both emphasize that the input levels lagged at least two periods can be valid

[3] For example, suppose two inputs, labor and capital, are both endogenous. At least five instruments are needed since all of the two inputs, their square terms, and their intersection are endogenous in the T-L production function. However, under some additional assumptions, consistent estimators can be obtained using only two control functions, not five. This point has been made by some economists, including BLUNDELL, R.W., POWELL, J.L., (2004). Endogeneity in Semiparametric Binary Response Models. *The Review of Economic Studies* **71**, 655-679., TERZA, J.V., BASU, A., RATHOUZ, P.J., (2008). Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling. *Journal of health economics* **27**, 531-543., and WOOLDRIDGE, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT press.. See detailed discussion in AMSLER, C., PROKHOROV, A., SCHMIDT, P., (2016). Endogeneity in Stochastic Frontier Models. *Journal of Econometrics* **190**, 280-288..

instruments based on annual data such as we use in our empirical study. We also use input prices as instruments.[4]

3. Empirical Study of the Oilfield Market

We demonstrate the method by using global oilfield data. SSPF approach imputes the undistorted input allocations and conducts a segment-specific frontier analysis for multi-divisional firms in the global oilfield market, where five segments exist.

3.1 The Oilfield Market

The oilfield market, or oil and gas exploration and service, is a complex market whose firms utilize processes that involve specialized technologies at each step of the oil and gas supply chain. Most oil and gas companies, even large vertically integrated companies such as Chevron and Exxon Mobil, choose to rent or buy part of the necessary equipment from oilfield services firms. Companies in the oilfield market provide the infrastructure, equipment, intellectual property, and services needed by the international oil and gas industry to explore for and extract crude oil and natural gas, and then transport it from the earth to the refinery, and eventually to the consumer. Therefore, this industry has many diverse product lines.

Firms in this industry have a total market capitalization of over $4 trillion USD, generating total revenues over $400 billion USD in 2014. The significant development of new technologies, such as hydraulic fracturing and offshore drilling, has resulted in a 10% compound annual growth rate (CAGR) from 2004 to 2014. Oil and gas companies are also paying more attention to unconventional oil and gas, offshore production, and aging reservoirs to maintain a steady supply of fossil fuel products. The higher exploration and production (E&P) spending of the oil and gas companies translates to higher revenues for the oilfield service companies.

The Oilfield Market Report (OMR) by Spears divides the oilfield industry into five segments: 1) exploration, 2) drilling, 3) completion, 4) production, and 5) capital equipment, downhole tools and offshore services (capital equipment, hereafter). OMR reports segment-level revenue for the 114 public firms in the field: 68 firms are

[4] The estimation results using lag two and lag three input quantities are robust in the empirical study.

single-division and 56 firms are multi-divisional (28 firms do business in two segments, 10 firms are active in three segments, seven firms have footprints in four segments, and only one firm covers all five segments). There are four diversified oilfield firms (the "Big Four"): Baker Hughes, Halliburton, Schlumberger and Weatherford.

3.2 Oilfield Data

In this section we discuss the outputs and the inputs used in our analysis and how they are constructed. Previous studies of this industry, such as those by Al-Obaidan and Scully (1992)) and Hartley and Medlock (2008) selected both revenue and physical products as the output measures. The use of revenue in these earlier studies instead of output quantity to estimate firm-level performance in the petroleum industry was justified on the basis of three rationales. The first was that physical output such as oil and gas produced may fail to catch the impact of subsidies (e.g., a lower domestic price) as the result of political pressure on National Oil Companies (NOCs). Second, a usual method to aggregate the multiple products (e.g., oil and gas) is to calculate their relative value at market prices. Third, revenue figures are usually easier to collect than the quantities of various products. Wolf (2009) shows the strong correlation between physical outputs and revenue in oil and gas companies in his empirical work on the performance of state vs. private oil companies during the period (1987–2006). Most of the recent studies in this industry use revenue as the output in estimating the performance of oil and gas companies (Eller *et al.*, 2011; Gong, 2018b; Hartley and Medlock III, 2013). In order to adjust the prices over time, all of these studies use oil prices as a control variable in the production function. Another method to control price over time is to deflate revenue so that the output level in different periods are comparable. Gong (2017) and Gong (2018d) use deflated revenue as the output variable in productivity analysis of oilfield companies. We follow this approach and utilize deflated revenue (based on the producer price index) as the output measure which provides a measure that is more comparable to the quantity-based production analyses utilized conducted by Gandhi *et al.* (2013) and De Loecker *et al.* (2016).

In terms of input variables, labor, measured by number of employees, is always used as an input in recent productivity analyses of the petroleum industry (Al-Obaidan and Scully, 1992; Eller *et al.*, 2011; Gong, 2018b; Hartley and Medlock III, 2013; Wolf, 2009). Assets is another important input for these companies: Al-Obaidan and Scully (1992) use total assets; Wolf (2009) use the sum of oil and gas reserves as an extra input in addition to total asset; Eller *et al.* (2011), Ike and Lee (2014) and Ohene-Asare *et al.* (2017) drop total assets and separate oil reserves and gas reserves as capital input variables; Hartley and Medlock III (2013) and Gong (2018b) keep oil reserves and gas reserves, and add refining capacity as another important capital variable, as refining capacity is the key assets in downstream activities. No related literature uses intermediate inputs, as it is often unavailable, which is also the case in our dataset. Different from the petroleum industry, the oilfield market consists of oilfield service companies, which provide services to the petroleum exploration and production industry but do not typically produce petroleum themselves. For the oil and gas companies, reserves are an important input as well as asset and these companies pay E&P (exploration and production) spending to "purchase" this input. The service firms exchange services such as directional drilling and hydraulic fracturing for the E&P spending from the oil and gas companies. Therefore, reserves are not an input for service companies, most of which do not own reserves, let alone refining capacity, which is the downstream asset the petroleum industry that has no bearing on oilfield companies. Similar to many industries, oilfield companies use many different types of equipment and assets in the production process. A typical solution is to sum them up to one input – capital, using the Perpetual Inventory Method (PIM). The PIM interprets a firm's capital stock as an inventory of investments, which already takes various equipment and assets into consideration.

To summarize, our study uses deflated revenue as the output, the number of employees and capital as the inputs as others have done in this literature and due to data availability. Although (Gong, 2017) and Gong (2018d) also consider some characteristics of the M-form in oilfield companies, their estimations of productivity and efficiency stay at the firm-level. This paper, however, estimates segment-level production function and efficiency, which provides much more information within those multi-divisional companies. Different production functions are specified for the five segments and this leads to a "2 Inputs – 5 Products/Segments – 18 Periods" technology that each firm in the industry utilizes. Division-level revenue data from 1997 to 2014 for each of the 114 public firms are collected from the three waves of the OMR (2000, 2011, and 2015) dataset. Appendix C introduces this report, the method employed to combine the three waves of data, and the detailed segmentation of the oilfield market. The Bureau of Labor Statistics publishes the Producer Price Index (PPI) by North American Industry Classification System (NAICS) division. The average output price index for each of the five segments, as well as the overall oilfield market, can thus be calculated. The output price indices deflate the revenue, which is used as the measure of segment specific output.

Data on annual overall revenue, the number of employees, and total capital for the 114 public firms during the sample period is collected from Thomson ONE, Bloomberg, and FactSet. Capital services are assumed to be proportional to accounting capital, which is the sum of equity and long-term debt, adjusted based on the approach of Berlemann and Wesselhöft (2014) using PIM. Appendix D explains in more detail how the variables were constructed. The overall revenue of a firm is not always equal to the total revenue in the oilfield market as reported by the OMR. In some cases, the former may be larger because the company has some business outside the oilfield market. On the other hand, the former could be smaller, as the OMR adds the acquired firm's revenue to the mother firm's revenue even in the years before acquisition. The input proportionality assumption suggested by Foster *et al.* (2008) is used to adjust the labor and capital used in the oilfield market.

Input and output price data are constructed using conventional methods in the literature. The service price of capital is the sum of the depreciation rate and interest rate. The depreciation rate is calculated using the depreciation and capital data from Thomson ONE, Bloomberg, and FactSet. Using the firm-level beta[5], the risk-free rate, and the expected market return from the same database, the firm-level interest rate can be estimated with a capital asset pricing model (CAPM). For labor price, many international firms have compensation cost information, but North American firms have no such information published. The Bureau of Labor Statistics has average compensation per employee figures for each NAICS division in the Labor Productivity and Cost (LPC) Database. The labor prices of those firms without wage information are set equal to the

[5] In finance, the beta of an investment or a company is a measure of the risk arising from exposure to general market movements as opposed to idiosyncratic factors. The market portfolio of all investable assets has a beta of unity.

corresponding NAICS division average. Average labor and capital prices are constructed for each segment.

Table 1 summarizes the input and output quantities in the oilfield market and in each of the five segments. Firm-level labor and capital are observed, but segment-level labor and capital are not. The average revenues in drilling and completion are significantly higher than those of other segments. The exploration and drilling segments have a high average wage while the capital equipment segment has a low average wage. Moreover, the capital price in exploration is relatively high when compared to the other segments.

[Insert TABLE 1 Here]

3.3 Empirical Results

Our imputation of the input allocations assumes that the production function is either a C-D or T-L functional form. . The stopping criterion in the imputation algorithm ($c=1*10^{-6}$) was attained after eight iterations in the C-D case and after eleven iterations in the T-L case. We report below results for the C-D case while Appendix E provides the corresponding results for the T-L case. Overall, the findings are quite comparable for the two functional forms.

3.3.1 Estimations of Production Functions

Table 2 presents the estimation results of the SPF approach in Eq. (1), which reflect the production frontier at firm-level for the oilfield market.

[Insert TABLE 2 Here]

The first column of Table 2 reports the MLE estimation (assuming inputs are exogenous) and the second column presents the C2SLS estimation (assuming inputs may not be exogenous) of the total oilfield industry's production frontier. The method discussed by Amsler *et al.* (2016) is followed to test the exogeneity of labor and capital. The capital is assumed to be exogenous, while labor is allowed to be endogenous, which is consistent with the literature (Levinsohn and Petrin, 2003; Olley and Pakes, 1996). Griliches and Mairesse (1998) remark that fixed effect estimators have frequently led researchers to find point estimates for the capital coefficient that are very low and often not significantly different from zero. This problem appears to have been addressed, as the coefficient on capital in the C2SLS model (0.23 with a t-value of 19.2) is both statistically and economically more significant than the estimation without considering endogeneity (0.13 with a t-value of 8.6). Moreover, the sum of the output elasticities of the labor and capital factor inputs is insignificantly different from one in the C2SLS estimation, indicating a technology that does not display economies or diseconomies of scale when endogeneity is addressed.

The estimators in Table 2 make it possible to draw 3D images of the production frontiers. Figure 1 plots the production frontiers for the oilfield market. The left is the MLE estimation and the right is the C2SLS estimation. These 3D images visualize the production-labor-capital relation. It is obvious that the elasticity of capital in the C2SLS estimation is larger than that in the MLE estimation.

[Insert FIGURE1 Here]

The C2SLS estimation of the production function also derives firm-level technical efficiency and total factor productivity. Table 3 compares the productivity of single-division and multi-divisional firms. The first column regresses productivity on a dummy variable of multi-divisional firms and a group of year dummy variables, while the second column regresses productivity on number of divisions each company entered and a group of year dummy variables. The results show that multi-divisional firms on average have higher productivity than single-division firms, indicating the existence of economies of scope in this market.

[Insert TABLE 3 Here]

Table 4 presents the estimation results of the SSPF approach in Eq. (2), which reflects the segment-specific production frontiers for the oilfield market. This study first uses the

SSPF approach to impute input allocations at the divisional level and estimate segment-specific production functions, assuming that both single- and multi-divisional firms have the same production techniques within a segment. Then, the validity of this assumption is tested for each regression and it is found that the production techniques for single- and multi-divisional firms are significantly different in the capital equipment segment, which indicates that the key assumption previous studies rely heavily on is not always valid. Therefore, the observation from single-division firms cannot solely be used to estimate the production function that multi-divisional firms also follow.

[Insert TABLE 4 Here]

Since the production techniques are different for single- and multi-divisional firms in the capital equipment segment, this study drops the observations of single-division firms in that segment from the production function and reruns the iterations to reach the undistorted input allocations. Columns 1 - 4 of Table 4 list the estimated production frontiers for each of the first four segments, and Columns 5 and 6 show the estimated production frontiers in the capital equipment segment for single-and multi-divisional firms, respectively.

The production frontiers are different across segments. Compared with the drilling, completion, and capital equipment segments, the exploration and production segments are relatively more capital-intensive. This finding implies that the constant share in the input portfolio across segments assumed in De Loecker *et al.* (2016) is not a valid assumption.

Tables 2 and 4 also present the average efficiency for the oilfield market and each segment, respectively. The top and bottom 5% of the estimated efficiencies are dropped to eliminate possible outliers. After controlling the endogeneity problem, the average efficiency for the oilfield market as a whole is 0.541. For segment-level average efficiency, the exploration segment (0.719) and the multi-divisional firms in the capital equipment segment (0.673) are the highest, followed by the production (0.659) and drilling (0.633) segments, while the completion segment (0.606) and the single-division firms in the capital equipment segment (0.610) are the lowest.

This study tests if the coefficients of inputs in the production frontier are time-invariant by adding the intersections between each year dummy variable and each input into the production equations. All the coefficients of the intersections are insignificantly different from zero, which supports $f(X_{it};\beta_0)$ in Eq. (1) is time-invariant as modeled. Results for each year are not reported here and are available on request.

3.3.2 Investment Decision and Efficiency

In the oilfield market, mergers, acquisitions, investment and divestment are common ways to change the share of revenue by segment for multi-divisional firms. These companies take such strategic actions frequently in response to market volatility and to compete with peers. Compared with firms only generating revenue in one segment, multi-divisional firms can improve their overall efficiencies not only by improving division-level efficiency, but also by transferring more resources and business from less efficient segments to more efficient ones.

Suppose the revenues of a firm in segments *I* and *II* were half-and-half, while its efficiency in the two segments was 0.1 and 0.7, respectively. After a decade, this firm improved efficiency by 0.1 in each segment. During the same period, this firm divested in segment *I* and invested in segment *II*, so that the share of revenue by segment changed from 1:1 to 1:3 in segments *I* and *II*▷ Therefore, the aggregate efficiency for this firm changed from 0.4 to 0.65 in ten years. The improved efficiency in each segment contributes to an increase of 0.1, while the change in business portfolio contributes to an increase of 0.15 in overall efficiency. Table 5 checks if the multi-divisional firms in the oilfield market invested more in more efficient divisions than in less efficient divisions. For each multi-divisional firm, the change is calculated in the share of revenue by segment between the first year and the last year in the dataset. Two-divisional firms, on average, transfer 7.5% of their business from the low-efficiency division to the high-efficiency division. This transfer from the low- to high-efficiency division is not that significant for firms that are present in more than two segments. Overall, multi-divisional firms transferred about 4% of their business from low- to high-efficiency divisions.

[Insert TABLE 5 Here]

3.3.3 "Big Four" Comparison and the HAL/BHI Merger

Thus far, this section presents the average efficiency at the industry and segment levels. This subsection compares the firm-level and division-level efficiency of the "Big Four," namely Schlumberger (SLB), Halliburton (HAL), Baker Hughes (BHI), and Weatherford (WFT).

Figure 2 shows the efficiency of the four diversified oilfield services companies at the firm level over time. Baker Hughes had the highest efficiency level, followed by Schlumberger and Halliburton, which are very close to the industry average efficiency, while Weatherford is below industry average in terms of firm-level efficiency.[6]

[Insert FIGURE1 Here]

Looking at the divisional level, Table 6 provides detailed efficiency levels by division for the four giants in 2014. The efficiency levels within the companies across segments were very different. Schlumberger, the biggest company in the oilfield industry, had footprints in all five segments. This firm has very high efficiency in the exploration segment and no significant weakness in terms of division-level efficiency. Halliburton, the second largest firm in the field, has the highest efficiency in the exploration segment, but the lowest efficiency in the production segment among the "Big Four." Baker Hughes, which does business in only three segments, had the highest efficiency in the production segment. Compared with its peers, Weatherford is the smallest diversified oilfield company and is not as efficient as the other three companies in most segments.

In the fourth quarter of 2014, Halliburton (number two in the field) spent $35 billion USD to acquire Baker Hughes (number three in the field). The comparison of efficiency at the divisional level explains the incentive of this merger. Comparing the four segments that both Schlumberger and Halliburton entered,[7] Halliburton is slightly more efficient

[6] Although not listed in this study, the estimated efficiencies for these four companies are much higher compared with the industry average when overlooking the endogeneity problem. This is because the MLE method predicts a much lower returns to scale (see Column 1 in Table 2) and hence overestimates the efficiency of big companies. For the same reason, the MLE approach predicts that Schlumberger and Halliburton are more efficient than relatively small Baker Hughes.

[7] The capital equipment segment that only Schlumberger entered can be ignored in this comparison because it only generates

than Schlumberger in the exploration (0.95 vs. 0.90), drilling (0.70 vs. 0.68), and completion (0.62 vs. 0.61) segments, but significantly less efficient in the production (0.61 vs. 0.73) segment. Baker Hughes, however, is very efficient in the production segment (0.91). Moreover, the efficiencies in Baker Hughes' drilling and completion segments are also slightly higher than that of Halliburton and Schlumberger. Therefore, Halliburton can improve the efficiency in its production segment dramatically, while maintaining the efficiency in other segments after the merger. The new company can be as competitive as Schlumberger in every major segment.

On the other hand, massive divestment must happen after the wedding of Halliburton and Baker Hughes in order to be approved by the Department of Justice. This is a good opportunity for Weatherford to buy some good assets and improve efficiency. Weatherford should first buy drilling assets and then completion assets but not production assets.

[Insert TABLE 6 Here]

4. Conclusion

This paper aims to estimate the distribution of inputs across segments within a firm since such information is necessary for division-level productivity and efficiency analysis but in many cases remains unobserved. Solving a system of segment-specific production functions and a system of firm-level profit maximization problem simultaneously, this paper builds a model to impute the undistorted input allocations. This approach can test if some assumptions that previous studies rely on are valid and derive a result even in the absence of those assumptions.

The accuracy of the undistorted input allocations is tested using panel data from OECD countries where the actual division-level inputs are available. The undistorted estimation outperforms the three competing imputation methods. In the empirical study, the input allocations for all the multi-divisional firms in the global oilfield market are estimated, which make it possible to perform reduced-form stochastic frontier analysis for each of

0.5% of Schlumberger's total revenue.

the five segments. This SSPF approach predicts division-level and firm-level efficiency as well as the average efficiency for each of the five segments. Evidence is also found that the assumptions used by previous works are not always valid.

Acknowledgments: We thank Dr. Tang Xun, Dr. Gustavo Grullon, Dr.Steven Chu, Dr. Roger Myerson, Dr. Hulya Eraslan, Dr. Peter Hartley, Dr. Kenneth Medlock, and Dr. Marc Santugini. We are grateful to Spears and Associates, Inc. for providing their data, Weatherford International for giving us the opportunity to closely observe firms in the oilfield market, and Baker Institute for supporting our research on energy economics.

Reference

ACKERBERG, D.A., CAVES, K., FRAZER, G., (2015). Identification Properties of Recent Production Function Estimators. *Econometrica* **83**, 2411-2451.

AL-OBAIDAN, A.M., SCULLY, G.W., (1992). Efficiency Differences between Private and State-Owned Enterprises in the International Petroleum Industry. *Applied Economics* **24**, 237-246.

ALLEN, C., HALL, S.G., 1997. Macroeconomic Modelling in a Changing World: Towards a Common Approach. John Wiley & Son Ltd.

AMSLER, C., PROKHOROV, A., SCHMIDT, P., (2016). Endogeneity in Stochastic Frontier Models. *Journal of Econometrics* **190**, 280-288.

BALDWIN, J.R., CAVES, R.E., GU, W., (2005). Responses to Trade Liberalization: Changes in Product Diversification in Foreign-and Domestic-Controlled Plants. *Economic Analysis (EA) Research Paper*.

BATTESE, G.E., COELLI, T.J., 1992. Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India. Springer.

BATTESE, G.E., RAO, D.P., (2002). Technology Gap, Efficiency, and a Stochastic Metafrontier Function. *International Journal of Business and Economics* **1**, 87-93.

BATTESE, G.E., RAO, D.P., O'DONNELL, C.J., (2004). A Metafrontier Production Function for Estimation of Technical Efficiencies and Technology Gaps for Firms Operating under Different Technologies. *Journal of Productivity Analysis* **21**, 91-103.

BERGER, P.G., OFEK, E., (1995). Diversification's Effect on Firm Value. *Journal of Financial Economics* **37**, 39-65.

BERLEMANN, M., WESSELH FT, J.-E., (2014). Estimating Aggregate Capital Stocks Using the Perpetual Inventory Method–a Survey of Previous Implementations and New Empirical Evidence for 103 Countries. *Review of Economics/Jahrbuch für Wirtschaftswissenschaften* **65**, 1-34.

BERNARD, A.B., REDDING, S.J., SCHOTT, P.K., (2010). Multiple-Product Firms and Product Switching. *American Economic Review* **100**, 70-97.

BERNARD, A.B., REDDING, S.J., SCHOTT, P.K., (2011). Multiproduct Firms and Trade Liberalization. *The Quarterly Journal of Economics* **126**, 1271-1318.

BLUNDELL, R., BOND, S., (2000). Gmm Estimation with Persistent Panel Data: An Application to Production Functions. *Econometric reviews* **19**, 321-340.

BLUNDELL, R.W., POWELL, J.L., (2004). Endogeneity in Semiparametric Binary Response Models. *The Review of Economic Studies* **71**, 655-679.

BOTOSAN, C.A., (1997). Disclosure Level and the Cost of Equity Capital. *The Accounting Review* **72**, 323-349.

BRODA, C., WEINSTEIN, D.E., (2010). Product Creation and Destruction: Evidence and Price Implications. *American Economic Review* **100**, 691-723.

BURDETT, K., MORTENSEN, D.T., (1998). Wage Differentials, Employer Size, and Unemployment. *International economic review* **39**, 257-273.

COELLI, T.J., RAO, D.S.P., O'DONNELL, C.J., BATTESE, G.E., 2005. An Introduction to Efficiency and Productivity Analysis. Springer.

COLES, M.G., (2001). Equilibrium Wage Dispersion, Firm Size, and Growth. *Review of Economic Dynamics* **4**, 159-187.

DE LA FUENTE, A., DOM NECH, R., (2006). Human Capital in Growth Regressions: How Much Difference Does Data Quality Make? *Journal of the European Economic Association* **4**, 1-36.

DE LOECKER, J., GOLDBERG, P.K., KHANDELWAL, A.K., PAVCNIK, N., (2016). Prices, Markups, and Trade Reform. *Econometrica* **84**, 445-510.

DUCHIN, R., GOLDBERG, A., SOSYURA, D., (2015). Spillovers inside Conglomerates: Incentives and Capital. *Stanford University Graduate School of Business Research Paper* No. 14-22.

ELLER, S.L., HARTLEY, P.R., MEDLOCK, K.B., (2011). Empirical Evidence on the Operational Efficiency of National Oil Companies. *Empirical Economics* **40**, 623-643.

FOSTER, L., HALTIWANGER, J., SYVERSON, C., (2008). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *The American Economic Review* **98**, 394-425.

GANDHI, A., NAVARRO, S., RIVERS, D., (2013). On the Identification of Production Functions: How Heterogeneous Is Productivity? *Mimeo, University of Wisconsin-Madison*.

GEBHARDT, W.R., LEE, C.M.C., SWAMINATHAN, B., (2001). Toward an Implied Cost of Capital. *Journal of Accounting Research* **39**, 135-176.

GONG, B., (2017). Multi-Dimensional Interactions in the Oilfield Market: A Jackknife Model Averaging Approach of Spatial Productivity Analysis (Forthcoming). *Energy Economics*.

GONG, B., (2018a). Agricultural Reforms and Production in China Changes in Provincial Production Function and Productivity in 1978–2015. *Journal of Development Economics* **132**, 18-31.

GONG, B., (2018b). Different Behaviors in Natural Gas Production between National and Private Oil Companies: Economics-Driven or Environment-Driven? . *Energy Policy* **114**, 145-152.

GONG, B., (2018c). The Impact of Public Expenditure and International Trade on Agricultural Productivity in China (Forthcoming). *Emerging Markets Finance and Trade*.

GONG, B., (2018d). The Shale Technical Revolution -- Cheer or Fear? Impact Analysis on Efficiency in the Global Oilfield Service Market. *Energy Policy* **112**, 162-172.

GRILICHES, Z., MAIRESSE, J., 1998. Production Functions: The Search for Identification, *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*. Cambridge University Press, p. 169.

GROSSMAN, S.J., HART, O.D., (1983). An Analysis of the Principal-Agent Problem. *Econometrica* **51**, 7-45.

GUAN, Z., KUMBHAKAR, S.C., MYERS, R.J., LANSINK, A.O., (2009). Measuring Excess Capital Capacity in Agricultural Production. *American Journal of Agricultural Economics* **91**, 765-776.

HOLMSTROM, B., (1979). Moral Hazard and Observability. *The Bell Journal of Economics* **10**, 74-91.

HOLMSTROM, B., (1982). Moral Hazard in Teams. *The Bell Journal of Economics* **13**, 324-340.

HARTLEY, P., MEDLOCK, K.B., (2008). A Model of the Operation and Development of a National Oil Company. *Energy Economics* **30**, 2459-2485.

HARTLEY, P.R., MEDLOCK III, K.B., (2013). Changes in the Operational Efficiency of National Oil Companies. *The Energy Journal* **34**, 27-57.

HAYAMI, Y., (1969). Sources of Agricultural Productivity Gap among Selected Countries. *American Journal of Agricultural Economics* **51**, 564-575.

HAYAMI, Y., RUTTAN, V.W., (1970). Agricultural Productivity Differences among Countries. *The American Economic Review* **60**, 895-911.

HAYAMI, Y., RUTTAN, V.W., 1971. Agricultural Development: An International Perspective. Baltimore, Md/London: The Johns Hopkins Press.

HUANG, Y.-J., CHEN, K.-H., YANG, C.-H., (2010). Cost Efficiency and Optimal Scale of Electricity Distribution Firms in Taiwan: An Application of Metafrontier Analysis. *Energy Economics* **32**, 15-23.

IKE, C.B., LEE, H., (2014). Measurement of the Efficiency and Productivity of National Oil Companies and Its Determinants. *Geosystem Engineering* **17**, 1-10.

KAMPS, C., (2006). New Estimates of Government Net Capital Stocks for 22 Oecd Countries, 1960-2001. *IMF Staff Papers* **53**, 120-150.

KLACEK, J., VOŠVRDA, M., SCHLOSSER, Š., (2007). Kle Translog Production Function and Total Factor Productivity. *Statistika* **87**, 4.

KOOP, G., OSIEWALSKI, J., STEEL, M., 1995. Measuring the Sources of Output Growth in a Panel of Countries. Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

KOOP, G., OSIEWALSKI, J., STEEL, M.F.J., (1999). The Components of Output Growth: A Stochastic Frontier Analysis. *Oxford Bulletin of Economics and Statistics* **61**, 455-487.

KUMBHAKAR, S.C., LOVELL, C.A.K., 2003. Stochastic Frontier Analysis. Cambridge University Press.

LANG, L.H., STULZ, R.M., (1994). Tobin's Q, Corporate Diversification, and Firm Performance. *Journal of Political Economy* **102**, 1248-1280.

LAU, L.J., YOTOPOULOS, P.A., (1989). The Meta-Production Function Approach to Technological Change in World Agriculture. *Journal of Development Economics* **31**, 241-269.

LEVIN, J., (2002). Multilateral Contracting and the Employment Relationship. *Quarterly Journal of economics*, 1075-1103.

LEVINSOHN, J., PETRIN, A., (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of Economic Studies* **70**, 317-341.

MAKSIMOVIC, V., PHILLIPS, G., (2002). Do Conglomerate Firms Allocate Resources Inefficiently across Industries? Theory and Evidence. *The Journal of Finance* **57**, 721-767.

MARSCHAK, J., ANDREWS, W.H., (1944). Random Simultaneous Equations and the Theory of Production. *Econometrica, Journal of the Econometric Society*, 143-205.

MASTROMARCO, C., GHOSH, S., (2009). Foreign Capital, Human Capital, and Efficiency: A Stochastic Frontier Analysis for Developing Countries. *World Development* **37**, 489-502.

MCEWIN, R.I., 2011. Intellectual Property, Competition Law and Economics in Asia. Bloomsbury Publishing.

MUNDLAK, Y., HELLINGHAUSEN, R., (1982). The Intercountry Agricultural Production Function: Another View. *American Journal of Agricultural Economics* **64**, 664-672.

OHENE-ASARE, K., TURKSON, C., AFFUL-DADZIE, A., (2017). Multinational Operation, Ownership and Efficiency Differences in the International Oil Industry. *Energy Economics* **68**, 303-312.

OLLEY, G.S., PAKES, A., (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica* **64**, 1263-1297.

PETERSEN, M.A., RAJAN, R.G., (1994). The Benefits of Lending Relationships: Evidence from Small Business Data. *The Journal of Finance* **49**, 3-37.

POSHAKWALE, S., COURTIS, J.K., (2005). Disclosure Level and Cost of Equity Capital: Evidence from the Banking Industry. *Managerial and Decision Economics* **26**, 431-444.

PUIG-JUNOY, J., (2001). Technical Inefficiency and Public Capital in U.S. States: A Stochastic Frontier Approach. *Journal of Regional Science* **41**, 75-96.

REVERTE, C., (2012). The Impact of Better Corporate Social Responsibility Disclosure on the Cost of Equity Capital. *Corporate Social Responsibility and Environmental Management* **19**, 253-272.

ROSS, S.A., (1973). The Economic Theory of Agency: The Principal's Problem. *The American Economic Review* **63**, 134-139.

SODERBOM, M., TEAL, F., (2004). Size and Efficiency in African Manufacturing Firms: Evidence from Firm-Level Panel Data. *Journal of Development Economics* **73**, 369-394.

SONMEZ, A., (2013). Firm Entry, Survival, and Exit. *Academic Journal of Interdisciplinary Studies* **2**, 160.

SHARMA, K.R., LEUNG, P., (2000). Technical Efficiency of Carp Pond Culture in South Asia: An Application of a Stochastic Meta - Production Frontier Model. *Aquaculture Economics & Management* **4**, 169-189.

STEIN, J.C., (1997). Internal Capital Markets and the Competition for Corporate Resources. *Journal of Finance*, 111-133.

TEECE, D.J., (1981). Internal Organization and Economic Performance: An Empirical Analysis of the Profitability of Principal Firms. *The Journal of Industrial Economics* **30**, 173-199.

TEECE, D.J., 1998. Economic Performance and the Theory of the Firm. Edward Elgar Publishing.

TERZA, J.V., BASU, A., RATHOUZ, P.J., (2008). Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling. *Journal of health economics* **27**, 531-543.

VALMARI, N., 2016. Estimating Production Functions of Multiproduct Firms. Research Institute of the Finnish Economy.

WEI, Z., HAO, R., (2011). The Role of Human Capital in China's Total Factor Productivity Growth: A Cross-Province Analysis. *The Developing Economies* **49**, 1-35.

WILLIAMSON, O.E., 1975. Markets and Hierarchies : Analysis and Antitrust Implications ; a Study in the Economics of Internal Organization, 1. print. ed. New York, NY : Free Press.

WOLF, C., (2009). Does Ownership Matter? The Performance and Efficiency of State Oil Vs. Private Oil (1987–2006). *Energy Policy* **37**, 2642-2652.

WOOLDRIDGE, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT press.

ZELLNER, A., KMENTA, J., DREZE, J., (1966). Specification and Estimation of Cobb-Douglas Production Function Models. *Econometrica: Journal of the Econometric Society*, 784-795.

For Online Publication

Supplement to Resource Allocation in Multi-divisional Multi-product Firms: Examining the Divisional Productivity of Energy Companies

By Binlei Gong and Robin C. Sickles

Appendix A Imputation Method

This appendix presents the imputation method including how to derive Theorem 1 under Assumptions 1-3. A simple example is introduced first to illustrate how to impute undistorted input allocations in multi-divisional firms when they maximize profits. Then, the general model is given under a "*M* Inputs – *N* Products/Segments – *T* periods" economy. Finally, parametric production function specification is discussed.

A.1 A Simple Example of the Model

A.1.1 Setup

This model assumes that firms use two inputs: labor and capital. All firms can operate in a maximum of two segments following segment-specific production functions. Suppose that a total of six companies is observed: 1) firms *A* and *B* do business only in segment *I*; 2) firms *C* and *D* produce only in segment*II*; and 3) firms *E* and *F* have footprints in both segments. Therefore, firm *E*(*F*) is a multi-divisional firm with division *E1*(*F1*) in segment *I* and division *E2*(*F2*) in segment *II*. As a result, segment *I* has four competitors (firms *A*, *B*, *E1*, and *F1*), and segment *II* also has four competitors (firms *C*, *D*, *E2*, and *F2*).

A.1.2 First System of Equations: Segment-specific Production Functions

The first system of equations consists of two segment-specific production functions, one for each segment. The production functions are assumed to be stochastic and have a Cobb-Douglas form.

$$\begin{cases} \ln Y_{i1t} = \alpha_{i1t} + \beta_{1L}\text{In}(L_{i1t}) + \beta_{1K}\text{In}(K_{i1t}) + v_{i1t}\ ,\ \forall i,t \\ \ln Y_{i2t} = \alpha_{i2t} + \beta_{2L}\text{In}(L_{i2t}) + \beta_{2K}\text{In}(K_{i2t}) + v_{i2t}\ ,\ \forall i,t \end{cases} , \qquad \text{(A-1)}$$

where Y_{ijt}, L_{ijt}, and K_{ijt} are production, labor, and capital, respectively, for firm *i* in segment/division *j* at time *t*; the first equation is the production function frontier of segment *I* and the second equation is the production function frontier of segment II. β_{mn} is the coefficient of input *n* at segment m. *v* is the noise that is considered normally distributed with a mean of zero and a standard deviation of σ_v. The productivity term $\alpha_{ijt} = \alpha_{jt} - u_{ijt}$ has different characteristics under different production methods. In the stochastic frontier analysis, the technical efficiency term u_{ijt} has bounded support (positive) and the exponential of $-u_{ijt}$ indicates the efficiency level of firm *i* in segment/division *j* at time *t* (TE_{ijt}=$e^{-u_{ijt}}$). In the Olley-Pakes framework, the technical efficiency term is a first-order Markov process with unbounded support. All the coefficients β_{mn} can be estimated if the input allocations for firms *E* and *F* are given.

It is worth noting the observations in each of the two regressions in Eq. (A-1). If single-division and multi-divisional firms are assumed to have the same production technique within a segment, then the first equation includes *A, B, E1,* and *F1* (i.e. $\forall i$ = *A, B, E*1*, F*1), while the second equation includes *C, D, E2,* and *F2* (i.e. $\forall j$ = *C, D, E*2*, F*2). If single-division and multi-divisional firms are assumed to have different production techniques within a segment, then the first equation includes only *E1* and *F1* (i.e. $\forall i$ = *E*1*, F*1), while the second equation includes only *E2* and *F2* (i.e. $\forall j$ = *E*2*, F*2). In other words, single-division firms are removed from the regression if the latter assumption is selected.

A.1.3 Second System of Equations: Profit Maximization

The second system of equations is the profit maximization problem for multi-divisional firms *E* and *F*. Under Assumption 1, multi-divisional firms maximize the net-present-value of discounted future profits by maximizing firm-level profits period-by-period, because firms' decisions don't have any intertemporal consequences. Therefore, the division managers are willing to report their true profit functions in each period so that the headquarters can better predict firm-level profit function and value function. Under Assumption 3, this study follows the framework in Olley and Pakes (1996) so that the labor used in time *t* is chosen when the productivity shock at time *t* is observed, while the capital used in time *t* is chosen at time *t*-1.

Since this study develops stochastic production functions, the profits are also stochastic, rather than deterministic. Therefore, profit-maximizing firms maximize the mathematical expectation of profit (Zellner *et al.*, 1966). In this model, firm *i* first decides capital allocations to maximize the expected firm-level profit of time *t* knowing the production function and price information of time *t*-1 and then decides labor allocations to maximize the expected firm-level profit of time *t* knowing the production function and price information of time *t*.

$$\begin{cases}\max_{K_{ijt}} \mathrm{E}[\pi_{it}|I_{it-1}] = \max_{K_{ijt}} \sum_{j=1,2}\left[p_{jt-1}A_{ijt-1}L_{ijt-1}^{\beta_{jL}}K_{ijt}^{\beta_{jK}}e^{\frac{\sigma_{vj}^2}{2}}\gamma_j - c_{ijt-1}^{L}\sigma_j - c_{ijt-1}^{K}\varphi_j\right] \\ \max_{L_{ijt}} \mathrm{E}[\pi_{it}|I_{it}] = \max_{L_{ijt}} \sum_{j=1,2}\left[p_{jt}A_{ijt}L_{ijt}^{\beta_{jL}}K_{ijt}^{\beta_{jK}}e^{\frac{\sigma_{vj}^2}{2}} - c_{ijt}^{L} - c_{ijt}^{K}\right]\end{cases}, \tag{A-2}$$

where

$$c_{ijt}^{L} = h_{i1t}\omega_{jt}L_{ijt} \quad and \quad c_{ijt}^{K} = h_{i2t}\rho_{jt}K_{ijt} \ ,$$

$$\gamma_j = \exp(\mu_{pj} + \mu_{Aj} + \beta_{jL}\mu_{Lj} + \frac{\sigma_{pj}^2+\sigma_{Aj}^2+\sigma_{Lj}^2\beta_{jL}^2}{2}),$$

$$\sigma_j = \exp\left(\mu_{\omega j} + \mu_{Lj} + \frac{\sigma_{\omega j}^2+\sigma_{Lj}^2}{2}\right) \ and \ \ \varphi_j = \exp(\mu_{pj} + \frac{\sigma_{pj}^2}{2}),$$

$$h_{ist} = h_{ist}(\omega_{1t}L_{i1t}+\omega_{2t}L_{i2t} + \rho_{1t}K_{i1t} + \rho_{2t}K_{i2t} \quad \forall s = 1,2 \ .$$

π is the profit, and p is the output price; A_{ijt} equals $\exp(\alpha_{ijt})$ in Eq. (A-1) and indicates the productivity; p_{jt}, ω_{jt} and ρ_{jt} are the segment average price of output, labor and capital in segment *j* at time *t*, respectively; h_1 and h_2 are the spillover effect on labor and capital price that follows Assumption 2, respectively. Firms predict $\log(x_{ijt})|\log(x_{ijt-1}){\sim}N\big(\log\big(x_{ijt-1}\big) + \mu_{xj},\ \sigma_{xj}^2\big)$, where $x = p, A, L, \omega, \rho$ and all the μ and σ are known. Therefore, $\gamma_{j}, \sigma_{j}, \varphi_{j}$, and σ_{vj}^2 are known by companies. They are constant across firms and time, but vary across segments. Eq. (A-2) applies to firm *i*=*E*, *F*.

Setting the first order condition of Eq. (A-2) to zero, the result is:

$$\begin{cases}\beta_{1K}p_{1t-1}A_{i1t-1}L_{i1t-1}^{\beta_{1L}}K_{i1t}^{\beta_{1K-1}}\exp\left(\frac{\sigma_{v1}^2}{2}\right)\gamma_1 - \rho_{1t-1}c_{1t-1} = 0 \\ \beta_{2K}p_{2t-1}A_{i2t-1}L_{i2t-1}^{\beta_{2L}}K_{i2t}^{\beta_{2K-1}}\exp\left(\frac{\sigma_{v2}^2}{2}\right)\gamma_2 - \rho_{2t-1}c_{1t-1} = 0\end{cases}, \tag{A-3}$$

and

$$\begin{cases}\beta_{1L}p_{1t}A_{i1t}L_{i1t}^{\beta_{1L-1}}K_{i1t}^{\beta_{1K}}\exp\left(\frac{\sigma_{v1}^2}{2}\right) - \omega_{1t}c_{2t} = 0 \\ \beta_{2L}p_{2t}A_{i2t}L_{i2t}^{\beta_{2L-1}}K_{i2t}^{\beta_{2K}}\exp\left(\frac{\sigma_{v2}^2}{2}\right) - \omega_{2t}c_{2t} = 0\end{cases}, \tag{A-4}$$

where

$$c_{1t} = h'_{i1t}(\omega_{1t}L_{i1t}\sigma_1 + \omega_{2t}L_{i2t}\sigma_2) + h_{i2t}\varphi_2 + h'_{i2t}(\rho_{1t}K_{i1t}\varphi_1 + \rho_{2t}K_{i2t}\varphi_2),$$

$$c_{2t} = h'_{i1t}(\omega_{1t}L_{i1t} + \omega_{2t}L_{i2t}) + h_{i1t} + h'_{i2t}(\rho_{1t}K_{i1t} + \rho_{2t}K_{i2t}).$$

This study first solves the "perfectly variable" input labor in Eq. (A-4). From Eq. (A-1), the following is known:

$$R_{ijt} = p_{jt}A_{ijt}L_{ijt}^{\beta_{jL}}K_{ijt}^{\beta_{jK}}\exp(v_{ijt})$$

$$\Leftrightarrow p_{jt}A_{ijt}L_{ijt}^{\beta_{1L-1}}K_{ijt}^{\beta_{1K}} = R_{ijt}/\left[L_{ijt}exp(v_{ijt})\right] \quad \text{(A-5)}$$

$$\Leftrightarrow p_{jt}A_{ijt}L_{ijt}^{\beta_{1L}}K_{ijt}^{\beta_{1K-1}} = R_{ijt}/\left[K_{ijt}exp(v_{ijt})\right]. \quad \text{(A-6)}$$

Plug Eq. (A-5) into Eq. (A-4), and the result is

$$\begin{cases} \dfrac{\beta_{1L}\exp\left(\frac{\sigma_{v1}^2}{2}\right)R_{i1t}}{[L_{i1t}\exp(v_{i1t})]} = \omega_{1t}c_{2t} \\ \dfrac{\beta_{2L}\exp\left(\frac{\sigma_{v2}^2}{2}\right)R_{i2t}}{[L_{i2t}\exp(v_{i2t})]} = \omega_{2t}c_{2t} \end{cases}.$$

Then, this study divides the first by the second equation. Adding the observed firm-level labor input equation, the result is

$$\begin{cases} \dfrac{\beta_{1L}R_{i1t}L_{i2t}\exp(v_{i2t} + 0.5\sigma_{v1}^2)}{\beta_{2L}R_{i2t}L_{i1t}\exp(v_{i1t} + 0.5\sigma_{v2}^2)} = \dfrac{\omega_{1t}}{\omega_{2t}} \\ L_{i1t} + L_{i2t} = L_{it} \end{cases}$$

$$\Leftrightarrow \begin{cases} L_{i1t} = \dfrac{\beta_{1L}R_{i1t}\omega_{2t}}{\beta_{1L}R_{i1t}\omega_{2t} + \beta_{2L}R_{i2t}\omega_{1t}exp(v_{i1t} - v_{i2t} + 0.5\sigma_{v2}^2 - 0.5\sigma_{v1}^2)}L_{it} \\ L_{i2t} = \dfrac{\beta_{2L}R_{i2t}\omega_{1t}}{\beta_{1L}R_{i1t}\omega_{2t}\,exp(v_{i2t} - v_{i1t} + 0.5\sigma_{v1}^2 - 0.5\sigma_{v2}^2) + \beta_{2L}R_{i2t}\omega_{1t}}L_{it} \end{cases}. \quad \text{(A-7)}$$

In Eq. (A-7), segment actual revenues (R) and average segment labor price (ω) are observed. The rest of the variables, including the production parameters (β) and the noises (v), can be estimated in Eq. (A-1).

After solving the labor allocations, this study solves the capital inputs. Eq. (A-3) can be written as

$$\begin{cases} \beta_{1K}\dfrac{p_{1t-1}}{p_{it}}\dfrac{A_{i1t-1}}{A_{i1t}}\left(\dfrac{L_{i1t-1}}{L_{i1t}}\right)^{\beta_{1L}}\left[p_{1t}A_{i1t}L_{i1t}^{\beta_{1L}}K_{i1t}^{\beta_{1K}-1}\right]\exp\left(\dfrac{\sigma_{v1}^2}{2}\right)\gamma_1 = \rho_{1t-1}c_{1t-1} \\ \beta_{2K}\dfrac{p_{2t-1}}{p_{2t}}\dfrac{A_{i2t-1}}{A_{i2t}}\left(\dfrac{L_{i2t-1}}{L_{i2t}}\right)^{\beta_{2L}}\left[p_{2t}A_{i2t}L_{i2t}^{\beta_{2L}}K_{i2t}^{\beta_{2K}-1}\right]\exp\left(\dfrac{\sigma_{v2}^2}{2}\right)\gamma_2 = \rho_{2t-1}c_{1t-1} \end{cases}. \quad \text{(A-8)}$$

Plugging Eq. (A-6) into Eq. (A-8):

$$\begin{cases}\beta_{1K} z_{i1t}\left[R_{i1t}/[K_{i1t}\exp(v_{i1t})]\right]\exp\left(\frac{\sigma_{v1}^2}{2}\right)\gamma_1 = \rho_{1t-1}c_{1t-1} \\ \beta_{2K} z_{i2t}\left[R_{i2t}/[K_{i2t}\exp(v_{i2t})]\right]\exp\left(\frac{\sigma_{v2}^2}{2}\right)\gamma_2 = \rho_{2t-1}c_{1t-1}\end{cases},$$

where

$$z_{ijt} = \frac{p_{jt-1}}{p_{jt}}\frac{A_{ijt-1}}{A_{ijt}}\left(\frac{L_{ijt-1}}{L_{ijt}}\right)^{\beta_{jL}}.$$

Similar to the transformation of labor, this study divides the first equation by the second. Adding the observed firm-level capital input equation, the result is

$$\begin{cases}\dfrac{\beta_{1K} z_{i1t} R_{i1t} K_{i2t}\exp(v_{i2t} + 0.5\sigma_{v1}^2)\gamma_1}{\beta_{2K} z_{i2t} R_{i2t} K_{i1t}\exp(v_{i1t} + 0.5\sigma_{v2}^2)\gamma_2} = \dfrac{\rho_{1t-1}}{\rho_{2t-1}} \\ K_{i1t} + K_{i2t} = K_{it}\end{cases}$$

$$\Leftrightarrow \begin{cases} K_{i1t} = \dfrac{\beta_{1K}R_{i1t}\rho_{2t-1}}{\beta_{1K}R_{i1t}\rho_{2t-1} + \beta_{2K}R_{i2t}\rho_{1t-1}\left(\frac{\gamma_2}{\gamma_1}\right)\left(\frac{z_{i2t}}{z_{i1t}}\right)\exp(v_{i1t} - v_{i2t} + 0.5\sigma_{v2}^2 - 0.5\sigma_{v1}^2)} K_{it} \\ K_{i2t} = \dfrac{\beta_{2K}R_{i2t}\rho_{1t-1}}{\beta_{1K}R_{i1t}\rho_{2t-1}\left(\frac{\gamma_2}{\gamma_1}\right)\left(\frac{z_{i1t}}{z_{i2t}}\right)\exp(v_{i2t} - v_{i1t} + 0.5\sigma_{v1}^2 - 0.5\sigma_{v2}^2) + \beta_{2K}R_{i2t}\rho_{1t-1}} K_{it}\end{cases}. \qquad \text{(A-9)}$$

In Eq. (A-9), segment actual revenues (R), average segment capital price (ρ), and average segment labor price (ω) are observed. If Eq. (A-1) can be estimated, other variables, including the production parameters (β), the productivity shock (A), the noises (v) and their variances (σ_v^2), as well as z and γ, are all known. Then the capital allocations in Eq. (A-9) can be solved.

Eqs. (A-7) and (A-9) consist of the solutions of the second system of equations, which can be combined into Eq. (A-10)

$$\begin{cases} L_{i1t} = \dfrac{\beta_{1L}R_{i1t}\omega_{2t}}{\beta_{1L}R_{i1t}\omega_{2t} + \beta_{2L}R_{i2t}\omega_{1t}\exp(v_{i1t} - v_{i2t} + 0.5\sigma_{v2}^2 - 0.5\sigma_{v1}^2)} L_{it} \\ L_{i2t} = \dfrac{\beta_{2L}R_{i2t}\omega_{1t}}{\beta_{1L}R_{i1t}\omega_{2t}\exp(v_{i2t} - v_{i1t} + 0.5\sigma_{v1}^2 - 0.5\sigma_{v2}^2) + \beta_{2L}R_{i2t}\omega_{1t}} L_{it} \\ K_{i1t} = \dfrac{\beta_{1K}R_{i1t}\rho_{2t-1}}{\beta_{1K}R_{i1t}\rho_{2t-1} + \beta_{2K}R_{i2t}\rho_{1t-1}\left(\frac{\gamma_2}{\gamma_1}\right)\left(\frac{z_{i2t}}{z_{i1t}}\right)\exp(v_{i1t} - v_{i2t} + 0.5\sigma_{v2}^2 - 0.5\sigma_{v1}^2)} K_{it} \\ K_{i2t} = \dfrac{\beta_{2K}R_{i2t}\rho_{1t-1}}{\beta_{1K}R_{i1t}\rho_{2t-1}\left(\frac{\gamma_2}{\gamma_1}\right)\left(\frac{z_{i1t}}{z_{i2t}}\right)\exp(v_{i2t} - v_{i1t} + 0.5\sigma_{v1}^2 - 0.5\sigma_{v2}^2) + \beta_{2K}R_{i2t}\rho_{1t-1}} K_{it}\end{cases}, \qquad \text{(A-10)}$$

where

$$z_{ijt} = \frac{p_{jt-1}}{p_{jt}}\frac{A_{ijt-1}}{A_{ijt}}\left(\frac{L_{ijt-1}}{L_{ijt}}\right)^{\beta_{jL}},$$

$$\gamma_j = \exp\left(\mu_{pj} + \mu_{Aj} + \beta_{jL}\mu_{Lj} + \frac{\sigma_{pj}^2 + \sigma_{Aj}^2 + \sigma_{Lj}^2\beta_{jL}^2}{2}\right).$$

This system says that for each input, the ratio of expected marginal revenue equals the input price ratio across divisions. In other words, the cost for an input to generate additional expected revenue is equal across divisions at the undistorted allocations; hence, this study calls it the "equal (expected) marginal revenue per cost" condition. If the first system of equations in Eq. (A-1) is estimated, the input allocations for both firms E and F can be imputed in the second system of equations in Eq. (A-10).

A.1.4 Iterative Algorithm

These two systems of Eqs. (A-1) and (A-10) need to be jointly solved to derive division-level inputs and segment-specific production functions. This study implements an iterative method through the following procedures.

Step 1: For each input, preset a reasonable guess of the initial allocations. This study assumes input proportionality, which is recommended by Foster *et al.* (2008) and utilized as the first guess of division-level inputs.[8]

Step 2: Estimate the segment-specific production functions in Eq. (A-1) using the initial guess of division-level inputs in Step 1.

Step 3: Update the division-level input allocations in Eq. (A-10) using the estimated segment-specific production functions derived in Step 2.

Step 4: Repeat Steps 2 and 3 until a given stationary threshold of the change in coefficients is achieved by successive iterations.

Finally, the estimated division-level inputs are the undistorted allocations of the unobserved inputs, and the estimated production functions present how inputs are converted into output in each segment.

After the input allocations are imputed from the iterations, this study tests if the key assumption that single-division and multi-divisional firms use the same production techniques within a segment is correct. For each segment, this study creates a dummy variable of multi-divisional firms and then regresses the output on inputs, the dummy, and the intersections between each input and the dummy variable. If any coefficient of the intersections and the dummy variable is statistically significant, then the key assumption that most previous works (De Loecker *et al.*, 2016; Levinsohn and Petrin, 2003; Olley and Pakes, 1996) rely on is invalid. On

[8] In practice, other methods will also be used to set the initial allocations to see if the iteration converges at the same undistorted allocations. The methods and results are discussed in Appendix B.

the other hand, evidence is achieved to support this key assumption if all the coefficients of the intersections are insignificant.

This study recommends imputing input allocations assuming the key assumption is valid first so that both observations of single-division and multi-divisional firms can be used in the first system of equations. This is especially important with a small sample size. If the test then shows the assumption to be invalid, the iterations can be rerun after removing the observations of single-division firms in the first system of equations.

The next two subsections generalize the model, allowing for more inputs, segments, and multi-divisional firms, as well as different production functional forms. The algorithm is analogous to the simple model above and can derive Theorem 1. Those who are not interested in the details can skip the rest of Appendix A.

A.2 General Model

A.2.1 Setup

As an extension of the simplified model, the general model considers a generalized production function and analyzes a "M Inputs – N Products/Segments – T periods" economy. Moreover, firms can have footprints in one or multiple segments and enter or exit segments over time.

A.2.2 Single Frontier Analysis

The canonical stochastic frontier analysis for panel data, or what this study calls the SPF estimator, runs the regression for individual (country, region, company, etc.) i at time t[9]:

$$Y_{it} = f(X_{it};\beta_0)\exp(\tau Z)\exp(v_{it})\exp(-u_{it}), \qquad \text{(A-11)}$$

where Y_{it} is the total output of individual i at time t; X_{it}= (X_{it}^1, X_{it}^2, ..., X_{it}^M) vectors the M types of inputs; $f(X_{it};\beta_0)\cdot\exp(\tau Z)$ is the production frontier over time, where $f(X_{it};\beta_0)$ is the time-invariant part of the production function,[10] β_0 = (β_{01}, β_{02}, ..., β_{0M}) is a vector of technical parameters to be estimated. Z vectors a group of year dummy variables, controls the production frontier change across time and τ vectors the coefficients of time; $\exp(v_{it})$ is the stochastic component that describes random shocks affecting the production process, where v_{it}

[9] Eq. (A-11) is the same as Eq. (1).

[10] In the simple example for this study, this production function has a Cobb-Douglas form.

is assumed to be normally distributed with a mean of zero and a standard deviation of σ_v; and $TE_{it} = \exp(-u_{it})$ denotes the time-variant technical efficiency defined as the ratio of observed output to maximum feasible output $TE_{it} = 1$ or $u_{it} = 0$ shows that the *i*-th individual allocates at the production frontier and obtains the maximum feasible output at time *t*, while $TE_{it} < 1$ or $u_{it} > 0$ provides a measure of the shortfall of the observed output from the maximum feasible output. This study uses the error components specification with time-varying efficiencies (Battese and Coelli, 1992), where $u_{it} = \exp(-\eta(t - T)) \cdot u_i$. To sum up, the coefficients of the production function in Eq. (A-11) are time-invariant, while the firm-specific intercept term is shifted by a common time varying component. This study will test if the coefficients of the production function also change over time in the empirical analysis.

A.2.3 Segment-Specific Frontier Analysis

As has been discussed, Eq. (A-11) assumes a unique production function across segments and can only derive firm-level efficiency without segment-specific production function concern. The SSPF analysis in a "*M* Inputs – *N* Products/Segments – *T* periods" economy imputes the input allocations by solving two systems of equations simultaneously.

Each equation in the first system of *N* equations (A-12) describes the production techniques for the corresponding segment[11].

$$\begin{cases} Y_{i1t} = f_1(X_{i1t};\beta_1)\exp(\tau_1 Z_1)\exp(v_{i1t})\exp(-\mu_{i1t}) \\ \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \cdots \\ Y_{iNt} = f_N(X_{iNt};\beta_N)\exp(\tau_N Z_N)\exp(v_{iNt})\exp(-\mu_{iNt}) \end{cases}, \quad \text{(A-12)}$$

where Y_{ijt} represents the observed scalar output and X_{ijt} vectors the unobserved inputs of individual *i* in segment *j* at time *t*, respectively. $f_j(X_{ijt};\beta_j)$ is the heterogeneous production frontier for segment *j*, where β_j = (β_{j1}, β_{j2}, ...) is a vector of segment-specific technical parameters; Z_j = (Z_{j2}, Z_{j3}, ..., Z_{jT}) vectors a group of year dummy variables to control the production frontier change across time and τ_j = (τ_{j2}, τ_{j3}, ..., τ_{jT}) vectors the coefficients of the year dummy variables; and $\mu_{ijt} = \exp(-\eta(t - T)) \cdot \mu_{ij}$ is the time-variant efficiency indicator. v_{ijt} is the noise that is considered normally distributed with a mean of zero and a standard deviation of σ_{vj}. The data pooled into the *j*-th equation in Eq. (A-12) depends on the

[11] Eq. (A-12) is the same as Eq. (2).

validity of the key assumption. If assuming that single-division firms and multi-divisional firms have the same production techniques, all the players in segment *j* are included in the regression. Otherwise, the single-division firms are removed from the regression.

The second system of equations solves the profit maximization problem of multi-divisional firms. For firm *i* at time *t*, the mathematical expectation of the profit function is

$$\begin{cases} \max_{X_{ijt}^M} \mathrm{E}[\pi_{it}|I_{it-1}] = \max_{X_{ijt}^M} \sum_{j\in W(i,t)} (\mathrm{E}\left[p_{jt}Y_{ijt}|I_{it-1}\right] - \mathrm{E}\left[\sum_{k=1}^{M} h_{ikt}c_{jt}^k X_{ijt}^k|I_{it-1}\right]) \\ \max_{X_{ijt}^k} \mathrm{E}[\pi_{it}|I_{it}] = \max_{X_{ijt}^k} \sum_{j\in W(i,t)} (\mathrm{E}\left[p_{jt}Y_{ijt}|I_{it}\right] - \mathrm{E}\left[\sum_{k=1}^{M} h_{ikt}c_{jt}^k X_{ijt}^k|I_{it}\right]), \forall k < M \end{cases}, \quad \text{(A-13)}$$

where

$$h_{ikt} = h_{ikt}(\sum_{k=1}^{M} c_{jt}^k X_{ijt}^k), \forall k = 1,2, \dots, M$$

and

$$Y_{ijt} = f_j\left(X_{ijt};\ \beta_j\right) exp\left(\tau_j Z_j\right) exp\left(v_{ijt}\right) ex\,p\left(-u_{ijt}\right), \forall j \in W(i,t), \qquad \text{(A-14)}$$

where Y_{ijt} and X_{ijt}^k represent the observed scalar output and the unobserved *k*-th input of individual *i* in segment *j* at time *t*, respectively. The last input, X_{ijt}^M, is the division-level capital input. h_{ikt} is the spillover effects on the *k*-th input for firm *i* at time *t*. c_{jt}^k is the average price of the *k*-th input in segment *j* at time *t*. Similar to the simple example in Subsection A.1.1.3, the general model also assumes that firms predict $\log(x_{ijt'})|\log(x_{ijt}) \sim N\left(\log\left(x_{ijt}\right) + \mu_{xj}, \sigma_{xj}^2\right)$ where $x = p, A, X^k, \omega, \rho$ and all the μ and σ are known. $W(i,t)$ is a subset of $W = (1,2, \dots, N)$, indicating the segments in which individual *i* at time *t* has a footprint. When estimating region or country productivity, $W(i,t)$ is usually similar to W since most countries have production in every sector or segment. A counterexample is Singapore, which has a negligible agriculture industry. For company productivity analysis, $W(i,t)$ is smaller than W in most cases, as only a few integrated companies do business in every segment. Take this study's empirical study as an example; only one out of 114 firms does business in all five segments of the oilfield market.

Using the same strategy in Eq. (A-2), this study first solves all the "perfectly variable" inputs and then solves the capital input. By solving the first order condition of Eq. (A-13), the "equal (expected) marginal revenue per cost" condition can be derived. Adding the observed firm-level inputs, the first part of Theorem 1 (Eq. (3)) is derived

$$\begin{cases} \dfrac{\frac{\partial E[R_{ijt} \mid I_{it}]}{\partial X_{ijt}^k}}{\frac{\partial E\left[R_{ij't} \mid I_{it}\right]}{\partial X_{ij't}^k}} = \dfrac{c_{jt}^k}{c_{j't}^k} = r_{jj't}^k, \forall i, t, \forall k < M, \forall j, j' \in W(i,t) \\ \dfrac{\frac{\partial E[R_{ijt} \mid I_{it-1}]}{\partial X_{ijt}^M}}{\frac{\partial E\left[R_{ij't} \mid I_{it-1}\right]}{\partial X_{ij't}^M}} = \dfrac{c_{jt-1}^M}{c_{j't-1}^M} = r_{jj't-1}^M, \forall i, t, \forall j, j' \in W(i,t) \\ \sum_{j \in W(i,t)} X_{ijt}^k = X_{it}^k, \forall i, k, t \end{cases}, \quad \text{(A-15)}$$

where c_{jt}^k is the average price of the *k*-th input in segment *j* at time *t*; $r_{jj't}^k$ represents the ratio of average price for the *k*-th input between segments *j* and *j'* at time *t*.

Suppose $N(i,t)$ is the number of elements (i.e., the number of segments entered by firm *i* at time *t*) in set $W(i,t)$, and the system of Eq. (A-15) has $\sum_{i,t} N(i,t) \cdot M$ equations with $\sum_{i,t} N(i,t) \cdot M$ unknown division-level inputs (i.e. the black box of input allocations). If the first system of equations in Eq. (A-12) can be estimated, then system (A-15) is just-identified and the input allocations can be imputed.

This study looks for the undistorted allocations of segment-level inputs (X_{ijt}^k) that jointly solves the system of Eq. (A-12) and the system of Eq. (A-15). In order to find the undistorted input allocations and the segment-specific production functions that satisfy Eq. (A-12) and Eq. (A-15) simultaneously, this study implements the same iterative method discussed in Subsection A.1.4.

A.3 Parametric Production Function Specification

Since the stochastic frontier analysis is a parametric approach, the mathematical form for the production function needs to be specified. This study discusses the solution of the undistorted input allocations when the production function $f_j(X_{ijt}; \beta_j)$ has the most widely used forms: the Cobb-Douglas and Transcendental Logarithmic forms.

A.3.1 Cobb-Douglas Production Function

Given a C-D form, the canonical model in Eq. (A-11) has the form:

$$Y_{it} = \exp(\alpha) \prod_{k=1}^{2} (X_{it}^k)^{\beta_{0k}} \exp(\tau Z) \exp(v_{it}) \exp(-u_{it})$$

$$\Leftrightarrow \ln Y_{it} = \alpha + \sum_{k=1}^{2} \beta_{0k} \ln(X_{it}^{k}) + \tau Z + v_{it} - u_{it}. \quad \text{(A-16)}$$

The segment-specific production function in Eq. (A-12) becomes:

$$\begin{cases} \ln Y_{i1t} = \alpha_1 + \sum_{k=1}^{M} \beta_{1k} \ln\left(X_{i1t}^{k}\right) + \tau_1 Z_1 + v_{i1t} - u_{i1t} \\ \quad \cdots \\ \ln Y_{iNt} = \alpha_N + \sum_{k=1}^{M} \beta_{Nk} \ln\left(X_{iNt}^{k}\right) + \tau_N Z_N + v_{iNt} - u_{iNt} \end{cases}. \quad \text{(A-17)}$$

The “equal marginal revenue per cost” constraint is then:

$$\begin{cases} \dfrac{\frac{\partial E[R_{ijt} \mid I_{it}]}{\partial X_{ijt}^{k}}}{\frac{\partial E\left[R_{ij't} \mid I_{it}\right]}{\partial X_{ij't}^{k}}} = \dfrac{\beta_{jk} R_{ijt} X_{ij't}^{k} exp(v_{ij't} + 0.5\sigma_{vj}^{2})}{\beta_{j'k} R_{ij't} X_{ijt}^{k} exp(v_{ijt} + 0.5\sigma_{vj'}^{2})} = r_{jj't}^{k}, \forall i, t, \forall k < M, \forall j, j' \in W(i,t) \\ \dfrac{\frac{\partial E[R_{ijt} \mid I_{it-1}]}{\partial X_{ijt}^{M}}}{\frac{\partial E\left[R_{ij't} \mid I_{it-1}\right]}{\partial X_{ij't}^{M}}} = \dfrac{\beta_{jk} z_{ijt} R_{ijt} X_{ij't}^{M} exp(v_{ij't} + 0.5\sigma_{vj}^{2}) \gamma_j}{\beta_{j'k} z_{ij't} R_{ij't} X_{ijt}^{M} exp(v_{ijt} + 0.5\sigma_{vj'}^{2}) \gamma_{j'}} = r_{jj't-1}^{M}, \forall i, t, \forall j, j' \in W(i,t) \end{cases}.$$

where

$$z_{ijt} = \frac{p_{jt-1}}{p_{jt}} \frac{A_{ijt-1}}{A_{ijt}} \prod_{k=1,\dots,M-1} \left(\frac{X_{ijt-1}^{k}}{X_{ijt}^{k}}\right)^{\beta_{jk}},$$

$$A_{ijt} = \exp(\alpha_j + \tau_{jt} - u_{ijt}) \ .$$

Given that and the observed firm-level inputs, it is easy to solve the system of equations:

$$\begin{cases} X_{ijt}^{k} = \dfrac{\beta_{jk} R_{ijt}}{\sum_{j' \in W(i,t)} \beta_{j'k} R_{ij't} r_{jj't}^{k} exp(v_{ijt} - v_{ij't} + 0.5\sigma_{vj'}^{2} - 0.5\sigma_{vj}^{2})} X_{it}^{k}, \forall k < M \\ X_{ijt}^{M} = \dfrac{\beta_{jM} R_{ijt}}{\sum_{j' \in W(i,t)} \beta_{j'M} R_{ij't} r_{jj't}^{M} \left(\frac{\gamma_{j'}}{\gamma_j}\right) (\frac{z_{ij't}}{z_{ijt}}) \exp(v_{ijt} - v_{ij't} + 0.5\sigma_{vj'}^{2} - 0.5\sigma_{vj}^{2})} X_{it}^{M}, \end{cases} \quad \text{(A-18)}$$

where

$$\gamma_j = \exp(\mu_{pj} + \mu_{Aj} + \sum_{k=1}^{M-1} \beta_{jk} \mu_{X^k j} + \frac{\sigma_{pj}^{2} + \sigma_{Aj}^{2} + \sum_{k=1}^{M-1} \beta_{jk}^{2} \sigma_{X^k j}^{2}}{2}).$$

Eq. (A-18) is the second part of Theorem 1 (Eq. (4)). The solution exists when all the parameters β_{jk} are positive. Finally, the iteration method is used to solve the undistorted input allocations and technical parameters jointly from Eqs. (A-17) and (A-18).

A.3.2 Transcendental Logarithmic Production Function

Given a T-L form, the production function in Eq. (A-11) becomes:

$$\ln Y_{it} = \alpha + \sum_{k=1}^{M} \beta_{0k} \ln\left(X_{it}^{k}\right) + \frac{1}{2}\sum_{k=1}^{M}\sum_{l=1}^{M} \delta_{0kl}\, lnX_{it}^{k} lnX_{it}^{l} + \tau Z + v_{it} - u_{it} \ ,$$

where

$$\delta_{0kl} = \delta_{0lk}, \forall k, l \in (1,2,\dots,M).$$

Similarly, the segment-level production function in Eq. (A-12) becomes:

$$\begin{cases} \ln Y_{i1t} = \alpha_1 + \sum_{k=1}^{M} \beta_{1k} \ln\left(X_{i1t}^{k}\right) + \frac{1}{2}\sum_{k=1}^{M}\sum_{l=1}^{M} \delta_{1kl}\, lnX_{i1t}^{k} lnX_{i1t}^{l} + \tau_1 Z_1 + v_{i1t} - u_{i1t} \\ \dots \\ \ln Y_{iNt} = \alpha_N + \sum_{k=1}^{M} \beta_{Nk} \ln\left(X_{iNt}^{k}\right) + \frac{1}{2}\sum_{k=1}^{M}\sum_{l=1}^{M} \delta_{Nkl}\, lnX_{iNt}^{k} lnX_{iNt}^{l} + \tau_N Z_N + v_{iNt} - u_{iNt} \end{cases},$$

where

$$\delta_{jkl} = \delta_{jlk}, \forall k, l \in (1,\dots,M), \forall j = 1,\dots,N \ .$$

The “equal marginal revenue per cost” constraint is then:

$$\begin{cases} \dfrac{\frac{\partial E[R_{ijt} \mid I_{it}]}{\partial X_{ijt}^{k}}}{\frac{\partial E[R_{ij't} \mid I_{it}]}{\partial X_{ij't}^{k}}} = \dfrac{(\beta_{jk} + \sum_{l=1}^{M} \delta_{jkl} lnX_{ijt}^{l}) R_{ijt} X_{ij't}^{k} exp(v_{ij't} + 0.5\sigma_{vj}^{2})}{(\beta_{j'k} + \sum_{l=1}^{M} \delta_{j'kl} lnX_{ij't}^{l}) R_{ij't} X_{ijt}^{k} exp(v_{ijt} + 0.5\sigma_{vj'}^{2})} = r_{jj't}^{k}, \forall k < M \\ \dfrac{\frac{\partial E[R_{ijt} \mid I_{it-1}]}{\partial X_{ijt}^{M}}}{\frac{\partial E[R_{ij't} \mid I_{it-1}]}{\partial X_{ij't}^{M}}} = \dfrac{z_{ijt}(\beta_{jk} + \sum_{l=1}^{M} \delta_{jkl} lnX_{ijt}^{l}) R_{ijt} X_{ij't}^{k} exp(v_{ij't} + 0.5\sigma_{vj}^{2})\gamma_j}{z_{ij't}(\beta_{j'k} + \sum_{l=1}^{M} \delta_{j'kl} lnX_{ij't}^{l}) R_{ij't} X_{ijt}^{k} exp(v_{ijt} + 0.5\sigma_{vj'}^{2})\gamma_{j'}} = r_{jj't-1}^{M} \end{cases}, \quad \text{(A-19)}$$

where

$$z_{ijt} = \frac{p_{jt-1}}{p_{jt}} \frac{A_{ijt-1}}{A_{ijt}} \prod_{k=1,\dots,M-1} \left(\frac{X_{ijt-1}^{k}}{X_{ijt}^{k}}\right)^{\beta_{jk}} \prod_{k,l=1,\dots,M} \frac{exp(\frac{1}{2}\delta_{1kl} lnX_{i1t-1}^{k} lnX_{i1t-1}^{l})}{exp(\frac{1}{2}\delta_{1kl} lnX_{i1t}^{k} lnX_{i1t}^{l})}$$

$$A_{ijt} = \exp(\alpha_j + \tau_{jt} - u_{ijt}) \quad and \quad k + l < 2M$$

$$\gamma_j = \exp\{\mu_{pj} + \mu_{Aj} + \sum_{k=1}^{M-1} \beta_{jk}\mu_{X^k j} + \frac{\sigma_{pj}^{2} + \sigma_{Aj}^{2} + \sum_{k=1}^{M-1} \beta_{jk}^{2}\sigma_{X^k j}^{2}}{2} +$$

$$\sum_{k,l=1}^{M} [\left(\beta_{jk}\mu_{X^k j} + \frac{\beta_{jk}^{2}\sigma_{X^k j}^{2}}{2}\right)\left(\beta_{jl}\mu_{X^l j} + \frac{\beta_{jl}^{2}\sigma_{X^l j}^{2}}{2}\right)]\}.$$

Compared with Eq. (A-17) in the C-D case, Eq. (A-19) in the T-L case adds the $\sum_{l=1}^{M} \delta_{jkl}\, lnX_{ijt}^{l}$ part, which makes the latter case much more complicated to solve. Interestingly, this ratio itself is a C-D function.

In the T-L case, the system of equations in Eq. (A-15) becomes

$$\begin{cases} \frac{\frac{\partial E[R_{ijt} \mid I_{it}]}{\partial X_{ijt}^{k}}}{\frac{\partial E\left[R_{ij't} \mid I_{it}\right]}{\partial X_{ij't}^{k}}} = \frac{(\beta_{jk}+\sum_{l=1}^{M} \delta_{jkl} lnX_{ijt}^{l})R_{ijt}X_{ij't}^{k} exp(v_{ij't}+0.5\sigma_{vj}^{2})}{(\beta_{j'k}+\sum_{l=1}^{M} \delta_{j'kl} lnX_{ij't}^{l})R_{ij't}X_{ijt}^{k} exp(v_{ijt}+0.5\sigma_{vj'}^{2})} = r_{jj't}^{k}, \forall k < M \\ \frac{\frac{\partial E[R_{ijt} \mid I_{it-1}]}{\partial X_{ijt}^{M}}}{\frac{\partial E\left[R_{ij't} \mid I_{it-1}\right]}{\partial X_{ij't}^{M}}} = \frac{z_{ijt}(\beta_{jk}+\sum_{l=1}^{M} \delta_{jkl} lnX_{ijt}^{l})R_{ijt}X_{ij't}^{k} exp(v_{ij't}+0.5\sigma_{vj}^{2})\gamma_{j}}{z_{ij't}(\beta_{j'k}+\sum_{l=1}^{M} \delta_{j'kl} lnX_{ij't}^{l})R_{ij't}X_{ijt}^{k} exp(v_{ijt}+0.5\sigma_{vj'}^{2})\gamma_{j'}} = r_{jj't-1}^{M} \\ \sum_{j \in W(i,t)} X_{ijt}^{k} = X_{it}^{k}, \forall i,k,t \end{cases} . \quad \text{(A-20)}$$

Eq. (20) is the third part of Theorem 1 (Eq. (5)). This system of equations is just-identified, where the number of unknown variables equals the number of equations. However, whether a solution exists depends on the initial inputs, i.e., β_{jk}, δ_{jkl}, Y_{ijt}, X_{it}^{k}, $r_{jj't}^{k}$, M, and $N(i,t)$. The T-L production function is a flexible functional form and a generalization of the C-D production function (Allen and Hall, 1997). Compared with the C-D form, the T-L form does not need the perfect substitution between inputs restriction or the linear input-output restriction (Klacek *et al.*, 2007). However, an easy solution cannot be obtained for the T-L form in Eq. (A-20) as was done for the C-D form solution in Eq. (A-18).

Appendix B Accuracy Test of the Estimation

B.1 OECD Data

This appendix uses panel data for OECD countries to test whether the imputation method estimates accurate input allocation. Many studies treat countries or regions as firms and use a production frontier approach to estimate the productivity and efficiency of a sector or the whole economy. Koop *et al.* (1999) decompose output change into technical, efficiency and input changes for seventeen OECD countries by using stochastic frontier methods. Mastromarco and Ghosh (2009) use stochastic frontier analysis to derive the total factor productivity for fifty-seven developing countries. Koop *et al.* (1995) measure the source of growth in four regions, consisting of East Asia, Africa, Latin America, and West. Other studies focus on cross-state or cross-province rather than cross-country productivity analysis, such as the thirty provinces of China (Gong, 2018a; Wei and Hao, 2011) and the forty-eight contiguous US states (Puig-Junoy, 2001). However, all of these studies utilize one production function for the entire economy without considering differences across segments. With regard to segment-specific production functions, the SSPF method can impute segment-level input allocations in each country.

This study uses panel data for twenty-two OECD countries[12] from 2000 to 2006 where labor and capital are the inputs and the value-added is the output. The dataset is mainly collected from the Structural Analysis (STAN) Database, the Annual National Accounts (ANA) Database, and the Monthly Monetary and Financial Statistics (MEI) Database, all of which are in the OECD iLibrary[13].

Based on the sector classification of OECD data, the total economy of a country can be divided into primary, secondary, and tertiary sectors and further classified into nine industries:

I) the primary sector is segmented into 1) agriculture, hunting, forestry and fishing, and 2) mining and quarrying;

II) the secondary sector is divided into three categories, consisting of 3) manufacturing, 4) electricity, gas and water supply, and 5) construction; and

[12] These twenty-two OECD countries are Australia, Austria, Belgium, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Iceland, Ireland, Italy, Japan, the Netherlands, New Zealand, Poland, the Slovak Republic, Spain, Sweden, the United Kingdom, and the United States.

[13] http://www.oecd-ilibrary.org/

III) the tertiary sector includes 6) wholesale, retail trade, restaurants and hotels, 7) transport, storage and communications, 8) finance, insurance, real estate and business services, and 9) community, social, and personal services.

Our dataset has segment-level output, input allocations, and price data for each country. The output is value added in US dollars by segment and country. For the labor input, total employment and compensation for employees are collected. The former is a total quantity and the latter is divided by the former, thus providing labor price information. For the capital input, net capital stock/net fixed assets, consumption of fixed assets, and interest rate are available. The consumption of fixed assets divided by the net capital stock/net fixed asset is the depreciation rate. The price of capital, or the user cost of capital, is the sum of the depreciation rate and the interest rate. Table B-1 provides summary statistics for the inputs and outputs by segment.

[Insert TABLE B-1 Here]

A 3 Segments model is set up for OECD countries, where the output is value added (Y_{it}), the two inputs are labor ($X_{it}^1 = L_{it}$) and capital ($X_{it}^2 = K_{it}$), and the three segments are the primary, secondary, and tertiary sectors. In order to check the robustness of the results, this study further classifies the economy using nine industries and imputes the undistorted input allocations for a 9 Segments model.

Since all estimations of the technical parameters (β_{jk}) are positive, the system of equations in Eq. (A-12) has a solution for all the observations in the C-D case. When T-L production functions are assumed, however, less than 10% of the observations have solutions in Eq. (A-14) and can be refined. Therefore, only the accuracy of the estimation when production functions have C-D forms is tested.

B.2 Testing Method

"Equal (expected) marginal revenue per cost" estimation. The undistorted allocations of segment-level inputs $\hat{X}_{ijt}^k$ for the OECD countries imputed from the SSPF approach are called the "equal (expected) marginal product per cost" estimation,[14] which satisfies

[14] γ_j is removed because empirical data shows that the difference between segments is negligible.

(B-1)
$$\begin{cases} \frac{\beta_{jk} R_{ijt} X_{ij't}^{k} \exp(v_{ij't}+0.5\sigma_j^2)}{\beta_{j'k} R_{ij't} X_{ijt}^{k} \exp(v_{ijt}+0.5\sigma_{j'}^2)} = r_{jj't}^{k}, \forall k < M, \forall j, j' \in W(i,t) \\ \frac{z_{ijt} \beta_{jk} R_{ijt} \mathrm{X}_{ij't}^{\mathrm{M}} \exp(v_{ij't}+0.5\sigma_j^2)}{z_{ij't} \beta_{j'k} R_{ij't} X_{ijt}^{M} \exp(v_{ijt}+0.5\sigma_{j'}^2)} = r_{jj't-1}^{M}, \forall i, t, \forall j, j' \in W(i,t) \end{cases},$$

where

$$z_{ijt} = \frac{p_{jt-1}}{p_{jt}} \frac{A_{ijt-1}}{A_{ijt}} \prod_{k=1,\dots,M-1} \left(\frac{X_{ijt-1}^{k}}{X_{ijt}^{k}} \right)^{\beta_{jk}} .$$

This section tests whether method 1 (Eq. (B-1)) imputes allocations that are closer to the actual allocations than those based on the other three methods (Eqs. (B-2)-(B-4)). The three alternative methods for generating the missing allocations are presented below.

"Equal revenue per input" estimation. If both the differences in price and production function across segments are ignored in Eq. (B-1), then the segment-level input allocations are in proportion to the actual revenue across segments:

(B-2)
$$\frac{\frac{R_{ijt}}{X_{ijt}^{k}}}{\frac{R_{ij't}}{X_{ij't}^{k}}} = 1, \ \forall k = 1,2, \forall j, j' \in W(i,t) \ .$$

The input proportionality assumption is widely used, as it requires the least amount of information. For example, Foster *et al.* (2008) allocate inputs based on products' revenue shares. This study regards this estimation as the benchmark estimation since it is a reasonable first estimate; it is used in the first step of the iterative method. In business, financial analysts use revenue per employee and revenue per capital as efficiency ratios. This "Equal revenue per input" estimation assumes that these efficiency ratios are equal across segments within a company.

"Equal (expected) marginal revenue per input" estimation. In order to test whether considering the difference in price across segments in Eq. (B-1) improves the accuracy of the estimation, this study also estimates the input allocations under the condition

(B-3)
$$\begin{cases} \frac{\beta_{jk} R_{ijt} X_{ij't}^{k} exp(v_{ij't}+0.5\sigma_j^2)}{\beta_{j'k} R_{ij't} X_{ijt}^{k} exp(v_{ijt}+0.5\sigma_{j'}^2)} = 1, \forall i, t, \forall k < M, \forall j, j' \in W(i,t) \\ \frac{z_{ijt} \beta_{jk} R_{ijt} X_{ij't}^{M} exp(v_{ij't}+0.5\sigma_j^2)}{z_{ij't} \beta_{j'k} R_{ij't} X_{ijt}^{M} exp(v_{ijt}+0.5\sigma_{j'}^2)} = 1, \forall i, t, \forall j, j' \in W(i,t) \end{cases},$$

where the differences in market price of an input across divisions are ignored. This is called the "equal (expected) marginal revenue per input" approach.

“Equal revenue per cost” estimation. Similar to the benchmark estimation in Eq. (B2), this method considers average revenue rather than expected marginal revenue. However, this approach also takes input price information into consideration.

(B-4) $$\frac{\frac{R_{ijt}}{x_{ijt}^{k}}}{\frac{R_{ij't}}{x_{ij't}^{k}}} = r_{ijj't}^{k}, \forall k = 1,2, \forall j, j' \in W(i,t) \ .$$

The estimation assumes that the average revenue per input is proportional to the price of the same input across segments. This approach guarantees that the average costs of the value-added are equal across segments and is therefore denoted as the “equal revenue per cost” estimation.

To sum up, only the undistorted estimation (“equal expected marginal revenue per cost”) considers both segment-specific input prices and production functions. The benchmark estimation “equal revenue per input” is the easiest to derive, as it requires no additional information. The “equal revenue per cost” estimation takes input price into account on the basis of the benchmark estimation and is likely to be close to the actual level. Both of these estimations ignore the segment-specific technical parameters, but are more computationally friendly than other methods, since no segment-specific production regressions (stochastic frontier analysis in this case) and iterations are involved. The “equal marginal revenue per input” estimation, on the other hand, considers segment-specific production function but ignores the heterogeneity in input prices across segments. This estimation can therefore substitute the undistorted estimation if segment-level input prices are unobserved. Based on the amount of information used, this study predicts that the undistorted estimation is the most accurate and that the benchmark estimation is the least accurate. There is no measurable evidence about the relative accuracy between the “equal revenue per cost” and “equal marginal revenue per input” estimations. The advantage of the former is its lower computational burden, while the advantage of the latter is the fewer amounts of data (segment-level inputs) needed.

This study uses the mean square error (MSE) and the mean absolute error (MAE) methods for the accuracy test. Suppose the actual segment-level value of the *k*-th input is X_{ijt}^{k}, while the values for the undistorted estimation (“equal expected marginal revenue per cost”), “equal revenue per input” estimation, “equal revenue per cost” estimation, and “equal marginal revenue per input” estimation are $\hat{X}_{ijt}^{k}(1)$, $\hat{X}_{ijt}^{k}(2)$, $\hat{X}_{ijt}^{k}(3)$, and $\hat{X}_{ijt}^{k}(4)$, respectively. Then, the mean square error of the input allocations can be calculated using

$$MSE_{(p)} = \frac{\sum_k \sum_i \sum_j \sum_t \left[\frac{\hat{X}_{ijt}^k(p) - X_{ijt}^k}{X_{ijt}^k}\right]^2}{\sum_k \sum_i \sum_j \sum_t [\mathrm{I}(X_{ijt}^k \neq 0)]}.$$

Similarly, the mean absolute error of the input allocations can be calculated using

$$MAE_{(p)} = \frac{\sum_k \sum_i \sum_j \sum_t |(\hat{X}_{ijt}^k(p) - X_{ijt}^k)/X_{ijt}^k|}{\sum_k \sum_i \sum_j \sum_t [\mathrm{I}(X_{ijt}^k \neq 0)]},$$

where $I(\cdot)$ is the indicator function that takes on a value of one if the argument is correct and a value of zero otherwise. The numerator of the MSE (MAE) is the sum of the square error (absolute error) of the segment-level estimation of the inputs to the actual level. The denominator is the number of nonzero segment-level values for the inputs. Finally, this study calculates and compares MSE and MAE to check the accuracy of those estimations, where a lower error or deviation implies a more accurate estimation.

In the iterative method to estimate the undistorted allocations, the benchmark estimator in Eq. (B-2) is set as the initial guess. The other two competing methods in Eqs. (B-3) and (B-4) can also be alternative initial guesses of the iteration. This study checks if the iteration converges to the same allocations when given various initial allocations, which implies the robustness of the SSPF estimation.

B.3 Test Results

This study uses stopping criterion c=1*10^{-6} for the iteration process. When the mean error of the production parameters in two consecutive iterations is smaller than this criterion, it is believed that a stationary threshold has been achieved and that the estimated segment-level input locations satisfy the assumption that the ratio of expected marginal products equals the price ratio of an input across segments for each country.

Table B-2 presents the error results of the 3 Segments model and the 9 Segments model. The first model takes four iterations to pass the criterion and the second model takes thirty iterations to achieve a stationary condition.

[Insert TABLE B-2 Here]

The first four columns of Table B-2 show the MSE and MAE for the undistorted estimation and the other three estimations, respectively, for comparison. The fifth column shows the MSE and MAE ratios of the undistorted estimation and the benchmark estimation, while the sixth column shows the MSE and MAE ratios of the undistorted estimation and the most accurate estimation among the three competing methods.

Compared with the benchmark "equal revenue per input" estimate, the undistorted estimation significantly decreases the mean square error and hence improves the accuracy. In the 3 Segments model, the MSE of the undistorted estimation is 16% of that of the benchmark estimation. A significant improvement in estimating the input allocations is also achieved in the 9 Segments model, where the MSE of the undistorted estimation is 21% of that of the benchmark estimation. Moreover, the MAE of the undistorted estimation is around 60% of that of the benchmark estimation in both the 3 Segments model and the 9 Segments model, which also indicates that the undistorted estimation is more accurate than the benchmark input proportionality estimation used in Foster *et al.* (2008)

The numbers in the last column are all smaller than one, which implies that the undistorted estimation is more accurate than all three competing estimations. Among the three estimations for comparison, the "equal revenue per cost" estimation that uses input price information is the most accurate one, while the other two are very close in their inaccuracy.

Moreover, this study uses all three competing methods (Eqs. (B-2)-(B-4)) to impute the input allocations and set each of these three allocations as the initial guess in the iteration, respectively. In Eq. (A-18), it can be found that the variation of input allocations only depends on the variation of the production functions, because all the revenue and prices in the equation are observed and fixed. Therefore, if the three initial allocations can derive similar production functions, these initial allocations can also impute similar undistorted input allocations.

Tables B-3 and B-4 present the estimated labor coefficient and capital coefficient in each segment for the 3 Segments model and the 9 Segments model, respectively. All the coefficients derived by the three different initial allocations are consistent, indicating that the iterative method converges to the same undistorted input allocations using those different initial guesses.

[Insert TABLES B-3 AND B-4 Here]

Appendix C: OMR Data Introduction and Adjustment

This study uses data from the Oilfield Market Report (OMR) by Spears & Associates. This report details the global oilfield equipment and service markets associated with five macro-segments: exploration, drilling, completion, production, and capital equipment. Spears & Associates began tracking the oilfield market in 1996 and publish its OMR annually. Each year, the report not only releases new data for the current year, but also updates previously published data. Most numbers in the OMR are estimates developed by Spears through five sources: public company reports (about 100 firms), published information, interviews (about 2,000 discussions), trade shows, and site visits.

There are several advantages of using the OMR dataset. Firstly, this report brings estimations under the same criteria. Different firms have different segmentations, so direct use of their revenue declarations by product line from their financial reports is not wise. Secondly, this dataset is widely used by most firms and clients in the field. Thirdly, Spears has investigated the numbers through many sources to confirm its estimations in the past twenty years. Lastly, the OMR is updated each year, which alters any incorrect numbers according to the newest information.

In this study, three versions of the OMR (2000, 2011, and 2015) are used to collect firm-level data from 1997 to 2014, which is denoted as OMR1997-2014. OMR2000 includes firm-level revenue by segment from 1997 to 2000, OMR2011 includes firm-level revenue by segment from 1999 to 2011, and OMR2015 includes firm-level revenue by segment from 2005 to 2014. Since different waves of data have different market divisions, this study uses the market segmentation of OMR2015 and adjusts the other two datasets to acquire statistically comparable numbers.

The revision in OMR2000 consists of 1) the “Mud Logging” segment being renamed as the “Surface Data Logging” segment; 2) the “Field Processing Equipment” segment being removed from the market; 3) the “Offshore O&M Services/Contracting” segment being added to the “Offshore Contract Drilling” segment; and 4) the “Production Logging” segment being added to the “Wireline Logging” segment. Moreover, the “Casing & Cementation Products” segment in both OMR2000 and OMR2011 is added to the “Completion Equipment & Services” segment. Finally, the “Pressure Pumping Service” segment in both datasets is divided into the “Cementing” and “Hydraulic Fracturing” segments.

The OMR1997–2014 contains share and size analysis for 32 micro-market segments within the 5 macro-segments from approximately 600 companies working around the world. OMR1997–2014 gives detailed revenue by segment for 275 companies, 114 of which are public firms that publish complete financial information annually. The other 300 smaller companies have been added to “Others” in the report. The detailed segmentation is as follows:

I) Exploration segment includes 1) Geophysical Equipment & Services;

II) Drilling segment includes 2) Cementing, 3) Casing & Tubing Services, 4) Directional Drilling Services, 5) Drill Bits, 6) Drilling & Completion Fluids, 7) Inspection & Coating, 8) Land Contract Drilling, 9) Logging-While-Drilling, 10) Offshore Contract Drilling, 11) Oil Country Tubular Goods, 12) Solids Control & Waste Management, 13) Surface Data Logging;

III) Completion segment includes 14) Completion Eqpt & Services, 15) Coiled Tubing Services, 16) Hydraulic Fracturing, 17) Productions Testing, 18) Rental & Fishing Services, 19) Subsea Equipment, 20) Surface Equipment, 21) Wireline Logging;

IV) Production segment includes 22) Artificial Lift, 23) Contract Compression Services, 24) Floating Production Services, 25) Specialty Chemicals, 26) Well Servicing; and

V) Capital Equipment, Downhole Tools & Offshore Services segment includes 27) Downhole Drilling Tools, 28) Petroleum Aviation, 29) Offshore Construction Services, 30) Rig Equipment, 31) Supply Vessels, and 32) Unit Manufacturing.

Appendix D: Estimating Capital Stocks Using Perpetual Inventory Method

The perpetual inventory method (PIM) is the most widely employed approach to estimate capital stocks in many statistical offices (Gong, 2018c). Berlemann and Wesselhöft (2014)review the three PIM approaches used most frequently in the literature, consisting of the steady state approach, the disequilibrium approach, and the synthetic time series approach. After comparing the advantages and disadvantages of those three methods, they are able to combine them into a unified approach in order to prevent the drawbacks of the various methods. Their approach follows the procedure proposed by de la Fuente and Doménech (2006).

The PIM interprets a firm's capital stock as an inventory of investments. The aggregate capital stock falls in the depreciation rate per period. Therefore, the capital stock in period t is a weight sum of the history of the capital stock investment:

$$K_t = \sum_{i=0}^{\infty}(1-\delta)^i I_{t-(i+1)} \ .$$

However, a complete time series of past investments from day one is not available for many companies. Thomson ONE, Bloomberg, and FactSet only cover the recent portion of investment history. Suppose the investment can only be tracked back to period t_1, then the current capital stock can be estimated by using

(D-1) $$K_t = (1-\delta)^{t-t0}K_{t0} + \sum_{i=0}^{t-1}(1-\delta)^i I_{t-(i+1)}.$$

Therefore, the information needed to calculate capital stock includes a time series of investment $I_{t-(i+1)}$, the rate of depreciate δ, and the initial capital stock K_{t0}. Firstly, de la Fuente and Doménech (2006) propose smoothing the time-series investment data since the economies are on their adjustment path towards equilibrium rather than staying in a steady state most of the time. Hence, this study smooths the observed capital expenditure (investment) using a regression $I_{it} = \alpha_i + \beta_1 t + \epsilon$ for each firm. Secondly, this study follows the lead of Kamps (2006) and uses time-varying depreciation schemes, which seems to be the most plausible variant. The time-variant smooth depreciation rate can be estimated as the fitted value of the regression $\delta_t = \alpha + \beta_2 t + \epsilon$. This study collects a given firm's annual depreciation and total capital data to calculate the depreciation rate in accounting and use this information to run the regression. Finally, the initial capital stock at time t_0 can be calculated from the investment I_{t_1}, the long-term investment growth rate g_I, and the estimated depreciation rate δ: $K_{t0} \approx$

$I_{t_1}/(\,g_I + \delta_{t_1})$, where the growth rate g_I is β_1 and the investment I_{t_1} is the fitted value in the same regression. Similar to the method used in Berlemann and Wesselhöft (2014), this study assumes that all the years before t_1 without desegregated data have the same constant depreciation rate as year t_1. But for all the recent years that we have investment data, the depreciation rate is time variant. Therefore, Eq. (D-1) becomes:

$$K_t = \prod_{i=t1}^{t}(1-\delta_i)\frac{I_{t_1}}{g_I+\delta_{t1}} + \sum_{i=0}^{t-1}\prod_{j=t-(i+1)}^{t-1}(\,1-\delta_j)\,I_{t-(i+1)}\,.$$

In our empirical study, t is 2014 for most companies that are still active while t_1 presents the first year of investment data and varies across firms.

Appendix E: Robustness Results Assuming a T-L Function

In this paper, the functional form of the production function is assumed to be Cobb-Douglas (C-D). This appendix gives the estimation results if Transcendental Logarithmic (T-L) is the formation of the production function. In the T-L model, no solution for the system of equations in Eq. (A-20) exists for about 16% of the observations. To sum up, the estimated results in the T-L case are consistent with the ones in the C-D case, supporting the robustness of our approach. However, C-D model guarantees all the input allocations can be updated as long as the technical parameters are positive.

[Insert TABLE E-1 Here]

[Insert FIGURE E-1 Here]

[Insert TABLES E-2 AND E-3 Here]

TABLE 1

Oilfield Market Summary Statistics

	Value Added	Output Price Index	Labor	Price of Labor	Capital	Price of Capital
Unit	$\$1 \cdot 10^9$	--	$1 \cdot 10^3$	$\$1,000$	$\$1 \cdot 10^9$	%
Oilfield Industry	1.7	2.0	6.1	76.0	1.4	20.5
I) Exploration	0.6	1.7	--	80.1	--	25.8
II) Drilling	1.4	2.4	--	79.8	--	20.9
III) Completion	1.2	2.1	--	75.3	--	20.6
IV) Production	0.5	2.1	--	76.1	--	19.7
V) Capital Equipment	0.8	1.9	--	68.3	--	19.2

TABLE 2

Estimate of Cobb-Douglas Production Function at Firm-Level

	(1)	(2)
	(w/o IV)	(w/IV)
lnL	.579***	.726***
	(.016)	(.013)
lnK	.134***	.226***
	(.016)	(.011)
Time dummy	Yes	Yes
Intercept	1.62***	-1.28***
	(.170)	(.095)
# of firms	114	113
# of obs.	1525	1298
Avg. Eff	.458	.541

Notes: Significant at: *5, * *1 and * * * 0.1 percent; Standard error in parentheses.

TABLE 3

Productivity between Single-division and Multi-divisional Firms

	(1)	(2)
Multi-divisional Firms	.171 ***	--
	(.029)	--
Number of Divisions	--	.016***
	--	(.005)
Time dummy	Yes	Yes
Intercept	1.868***	.582***
	(.078)	(.027)

Notes: Significant at: *5, * *1 and * * * 0.1 percent; Standard error in parentheses.

TABLE 4

Estimate of Segment-specific Cobb-Douglas Production Function

	(1)	(2)	(3)	(4)	(5)	(6)
	Exploration	Drilling	Completion	Production	Capital Equipment	
	(w/IV）	(w/IV)	(w/IV)	(w/IV)	(w/IV)	(w/IV)
	Both	Both	Both	Both	Single-division	Multi-division
lnL	.647***	.804***	.944***	.619***	.664***	.816***
	(.037)	(.022)	(.025)	(.027)	(.032)	(.032)
lnK	.283***	.185***	.072***	.309***	.187***	.173***
	(.03)	(.019)	(.024)	(.023)	(.027)	(.036)
Time dummy	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	-.721***	-1.7***	-2.24***	-.137	-.672***	-2.17***
	(.269)	(.136)	(.164)	(.192)	(.228)	(.222)
# of firms	18	53	44	28	22	22
# of obs.	218	606	493	325	257	285
Avg. Eff	.719	.633	.606	.659	.610	.673

Notes: Significant at: *5, * *1 and * * * 0.1 percent; Standard error in parentheses.

TABLE 5

Change in Share of Revenue by Segment for Multi-divisional Firms

Segments Entered	Firm #	Lowest Efficient Division Share Change	Highest Efficient Division Share Change
2	28	-7.5 %	7.5 %
3	10	-5.6 %	1.1 %
4&5	8	-0.4 %	0.9 %
Total	46	-3.2 %	4.3 %

TABLE 6

Efficiency Levels of the "Big Four" by Segment in 2014

Segments	Schlumberger	Halliburton	Baker Hughes	Weatherford	Oilfield Average
Exploration	0.90	0.95	--	--	0.71
Drilling	0.68	0.70	0.72	0.53	0.61
Completion	0.61	0.62	0.64	0.55	0.61
Production	0.73	0.61	0.91	0.71	0.63
Capital Equipment	0.63	--	--	0.50	0.61

TABLE B-1

OECD Data Summary Statistics

	Value Added	Labor	Price of Labor	Capital	Price of Capital
Unit	$\$1 \cdot 10^{12}$	$1 \cdot 10^{6}$	$\$1{,}000$	$\$1 \cdot 10^{12}$	%
Total Economy	1.196	18.572	22.69	3.295	7.95
I) Primary Sector	0.030	0.729	12.26	0.125	9.56
Agriculture, Hunting, Forestry & Fishing	0.018	0.668	10.58	0.064	9.78
Mining & Quarrying	0.013	0.064	37.22	0.064	9.83
II) Secondary Sector	0.281	4.447	25.83	0.459	10.03
Manufacturing	0.190	2.865	28.27	0.254	11.26
Electricity, Gas & Water Supply	0.025	0.165	37.79	0.140	7.97
Construction	0.066	1.417	19.89	0.068	10.88
III) Tertiary Sector	0.885	13.396	22.47	2.730	7.43
Trade, Restaurants & Hotels	0.195	3.948	15.64	0.182	10.02
Transport, Storage & Communications	0.087	1.074	36.16	0.279	9.72
Finance, Insurance, Real Estate & Biz Svs	0.323	2.483	27.32	1.607	6.72
Community, Social & Personal Services	0.284	5.890	23.73	0.728	7.98

TABLE B-2

Accuracy Test Results

Model	Error Type	(1) $error(1)$	(2) $error(2)$	(3) $error(3)$	(4) $error(4)$	(5) $\frac{error(1)}{error(2)}$	(6) $\frac{error(1)}{min_{p\neq 1} error(p)}$
3Segments Model	MSE	0.160	0.985	0.377	0.989	16%	42%
	MAE	0.242	0.373	0.340	0.354	65%	71%
9Segments Model	MSE	1.532	7.341	2.327	7.617	21%	66%
	MAE	0.612	1.093	0.657	1.093	56%	93%

Notes: $error(1)$ is the error of the undistorted “equal marginal revenue per cost” estimation; $error(2)$ is the error of the “equal revenue per input” estimation; $error(3)$ is the error of the “equal revenue per cost” estimation; and $error(4)$ is the error of the “equal marginal revenue per input” estimation.

TABLE B-3

Iteration Results by Different Initial Guess in 3 Segments Model

		(1)	(2)	(3)
I) Primary Sector	Labor	0.467	0.467	0.467
	Capital	0.385	0.385	0.385
II) Secondary Sector	Labor	0.701	0.702	0.702
	Capital	0.287	0.286	0.286
III) Tertiary Sector	Labor	0.644	0.644	0.644
	Capital	0.302	0.302	0.302

Notes: (1) is the estimation using "equal revenue per input" derived allocation as the initial guess; (2) is the estimation using "equal revenue per cost" derived allocation as the initial guess; and (3) is the estimation using "equal marginal revenue per input" derived allocation as the initial guess.

TABLE B-4

Iteration Results by Different Initial Guess in 9 Segments Model

		(1)	(2)	(3)
1) Agriculture, Hunting, Forestry & Fishing	Labor	0.407	0.408	0.408
	Capital	0.379	0.379	0.377
2) Mining & Quarrying	Labor	0.483	0.483	0.483
	Capital	0.343	0.343	0.343
3) Manufacturing	Labor	0.673	0.672	0.673
	Capital	0.319	0.319	0.319
4) Electricity, Gas & Water Supply	Labor	0.456	0.456	0.456
	Capital	0.364	0.364	0.364
5) Construction	Labor	0.625	0.625	0.625
	Capital	0.291	0.291	0.291
6) Trade, Restaurants & Hotels	Labor	0.670	0.670	0.670
	Capital	0.222	0.222	0.222
7) Transport, Storage & Communications	Labor	0.633	0.633	0.633
	Capital	0.203	0.203	0.203
8) Finance, Insurance, Real Estate & Biz Svs	Labor	0.602	0.602	0.602
	Capital	0.311	0.311	0.311
9) Community, Social & Personal Services	Labor	0.660	0.660	0.660
	Capital	0.328	0.328	0.328

Notes: (1) is the estimation using “equal revenue per input” derived allocation as the initial guess; (2) is the estimation using “equal revenue per cost” derived allocation as the initial guess;. and (3) is the estimation using “equal marginal revenue per input” derived allocation as the initial guess.

TABLE E-1

Change in Share of Revenue by Segment for Multi-divisional Firms (T-L Model)

Segments Entered	Firm #	Lowest Efficient Division Share Change	Highest Efficient Division Share Change
2	28	-13.5 %	13.5 %
3	10	-4.9%	2.2 %
4&5	8	1.6%	-1.1%
Total	46	-8.9%	8.0%

TABLE E-2

Estimate of Transcendental Logarithmic Production Function

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Entire Industry		Exploration	Drilling	Completion	Production	Capital Equipment	
	(w/o IV)	(w/IV)	(w/IV）	(w/IV)	(w/IV)	(w/IV)	(w/IV)	(w/IV)
	Both	Both	Both	Both	Both	Both	Single-division	Multi-division
lnL	.368***	.074	-.321***	-1.04***	-1.07***	-.698***	-.891***	-.990***
	(.065)	(.063)	(.120)	(.098)	(.099)	(.089)	(.180)	(.243)
lnK	.291***	.524***	.153	1.04***	1.32***	.314***	2.06***	.749***
	(.04)	(.043)	(.174)	(.079)	(.130)	(.106)	(.200)	(.135)
ln LlnL	-.034***	-.080***	-.049**	-.173***	-.229***	-.092***	-.154***	-.117***
	(.007)	(.006)	(.023)	(.014)	(.018)	(.014)	(.023)	(.025)
lnKlnK	-.013***	-.026***	.034	-.030**	-.075***	-.003	.063*	.044***
	(.005)	(.006)	(.026)	(.013)	(.025)	(.012)	(.035)	(.013)
lnLlnK	.045***	.086***	-.062	.152***	.273***	.007	.139***	.014
	(.01)	(.01)	(.046)	(.025)	(.037)	(.022)	(.052)	(.031)
Control *lnK*	--	.151***	.513***	.060	.213**	.128	-.228	.045
	--	(.044)	(.131)	(.061)	(.093)	(.081)	(.147)	(.095)
Time dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	1.59***	.080	2.67***	2.91***	1.99***	3.77***	.242	3.83***
# of firms	(.202)	(.219)	(.523)	(.328)	(.309)	(.284)	(.762)	(.685)
# of obs.	114	113	18	53	44	28	22	22
Avg. Eff	1525	1298	218	606	493	325	257	285

Notes: Significant at: *5, * *1 and * * * 0.1 percent; Standard error in parentheses.

TABLE E-3

Efficiency Levels of the "Big Four" by Segment in 2014 (T-L Model)

Segments	Schlumberger	Halliburton	Baker Hughes	Weatherford	Oilfield Average
Exploration	0.66	0.89	--	--	0.67
Drilling	0.74	0.73	0.84	0.66	0.72
Completion	0.57	0.46	0.62	0.42	0.50
Production	0.71	0.67	0.73	0.62	0.61
Capital Equipment	0.46	--	--	0.40	0.55

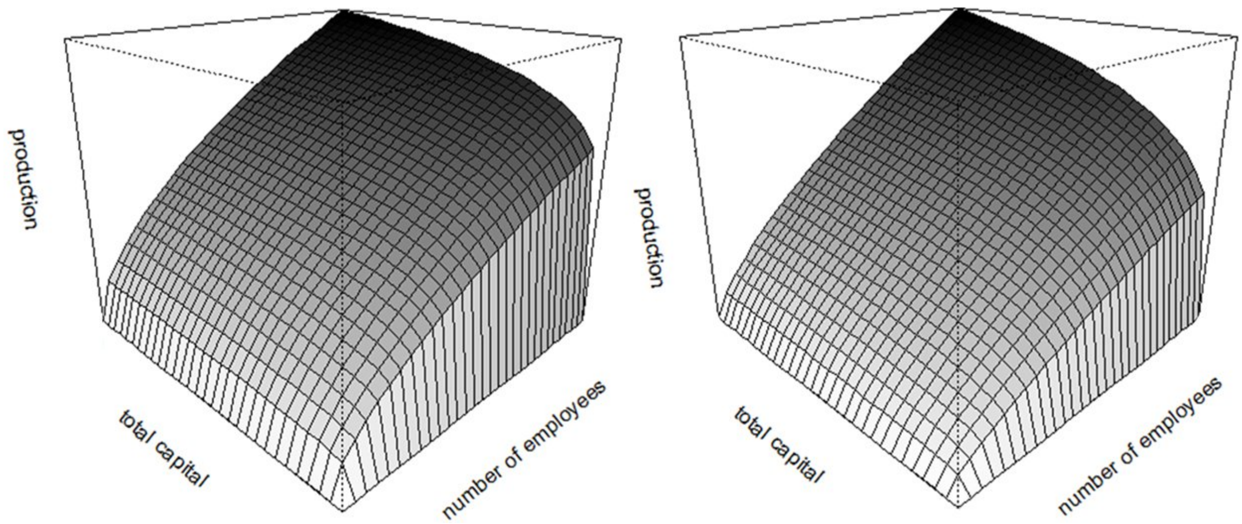

FIGURE 1

The Estimated Cobb-Douglas Production Function

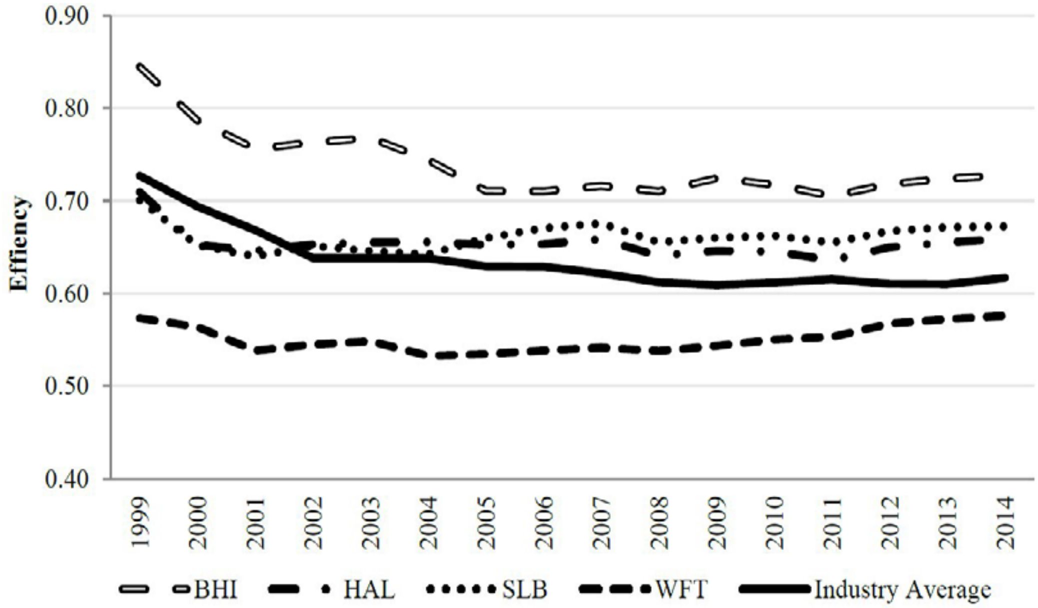

FIGURE 2

Efficiency Level for the "Big Four" and the Industry Average

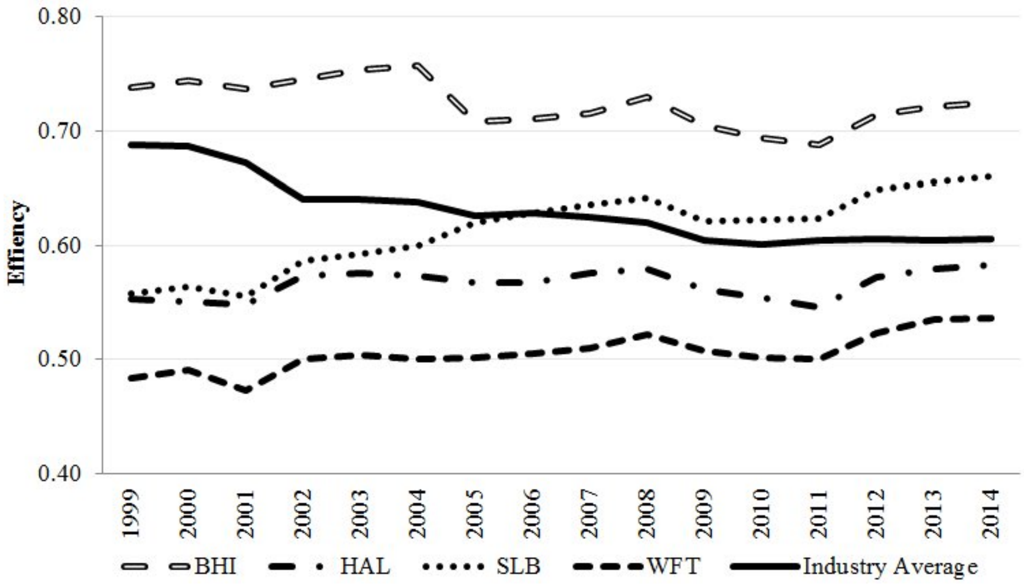

FIGURE E-1

Efficiency Level for BHI, HAL, SLB, WFT, and Industry Average (T-L)