Introduction

In this brief introduction we discuss motivations for the study of productivity, when optimizing behavior is not always practiced by the firm or productive unit under observation, that mostly come from important literatures that we will not cover in depth in our book. We discuss several sets of motivating factors and introduce them under the generic headings of management practices, behavioral economics, and X-efficiency. We purposely introduce the arguments in these literatures without reference to the more recent work in DEA and stochastic frontiers that make up the methodological coverage in the remaining chapters of the book. We do so in order to highlight the broad consensus that has existed since at least the early twentieth century for the persistent and transitory existence of suboptimal behaviors coming from these relatively distinct but very related literatures in management, psychology, sociology, and from classical economics.

Management Practices and Inefficiency in Production

Inefficiency in production, and the approaches to address it that we detail in the chapters to follow, can be linked to the sometimes overlooked but quite important literature on management practices, variations in which are what we almost always interpret as changes in the level of a firm's operating efficiency. Empirical studies have shown large differences in productivity across both firms and countries (Lieberman et al., 1990; Foster et al., 2008; Hsieh and Klenow, 2009; Hall and Jones, 1999), and one of the clear determinants of such differences can be attributed to management practices, as pointed out by Glaister (2014). In a study of microdata from 45 developing countries, Nallari and Bayraktar (2010) found that research and development, capacity utilization, and adoption of foreign technology were clearly related to productivity differences among the micro-units, and these are all determined by the decisions of management, which for all intents and purposes is an unobserved latent variable. In the classic work on variations in management practices, Bloom et al. (2012) find an $R^2 = 0.81$ in their regressions of gross domestic product (GDP) per capita on management practices

across 17 countries. Bloom et al. (2013) found similar variations in productivity explained by management practices within India. The exact formal mechanisms that map management skills and practices into a particular firm's productivity advantages is frankly not well understood, and empirical research on the topic has provided little in the way of a consensus on how such a formal linkage can be modeled. We often think of such a linkage in terms of a reduced form model that dispenses with a formal structure for the mediating factors and view the latent management input as being a function of a number of observable factors. There appears to be a consensus on the main factors that influence the effectiveness of management practices and it would appear that innovation is one of them. However, it is often difficult to distinguish between the innovation that is typically identified with technical changes, and the innovations instituted by management practices. Incremental innovations in management practices are ubiquitous throughout market economies (Dodgson and Gann, 2010, p. 15). Such adjustments to management practices may be informal and based on trial by error, involving the marketing of new models of existing products and services, changes in organizational structures and processes, coordination, routine supervision, and many other factors. These changes are incremental and it is their accumulation that triggers measurable productivity impacts (Rosenberg, 1982, p. 62–70; Mokyr, 1985, p. 28; McCloskey, 1985, p. 66; Blaug, 1999, p. 110). Moreover, there would appear to be a growing consensus that small to medium size enterprises are at a disadvantage in providing the resources needed to systematically impact productivity via R&D expenditures vis-à-vis larger enterprises (Haltiwanger et al., 1999; Shane, 2009; Van Praag and Versloot, 2007).

Prices also may drive financial performance just as much as productivity does. As pointed out by Lovell (2013), as well as Grifell-Tatjé and Lovell (2013, 2014, 2015), managers can get lucky and manage at the right time (minerals in Australia during 1995–2005) or the wrong time (minerals in Australia during 2006–2013). However, management practices also can be mitigating factors in response to such demand shocks that are out of the control of the firm. Bloom and Van Reenen (2007) studied 732 firms in the USA, the UK, France, and Germany and found that productivity gaps could be explained by 18 key management practices. Among the most important management practices were (1) shop floor operations (2) monitoring (3) targets (4) incentives. Why do management practices differ? Is it is due to product market competition? Is it due to ownership? What are the consequences of management practices on total factor productivity (TFP), profitability, sales growth, survival, Tobin's Q? In Brea-Solís et al. (2015) they provide counterexamples to the notion that productivity growth is necessary in order for a firm to be successful, based in part on Walmart's remarkable corporate history. This is also echoed in the volume by Grifell-Tatjé and Lovell (2015), which is devoted in part to the topic of business strategy, an expression apparently coined by Drucker (1954). Business strategy is of course connected to productivity but has important differences.

Do entrepreneurs engage in ventures that create value and how is this value captured as profit? Value in this context is the wedge between customers' willingness to pay and suppliers' willingness to sell (Brandenburger and Stuart, 1996). Creating value and capturing it as profit is the function of the business model. Creating value is an outcome of the dynamic interplay of productivity, innovation, and business practices. Value creation may in some cases be facilitated by finance and financial engineering, but these financial enterprises do not, in and of themselves, tend to create value.[1]

Although financial services have a role in intermediation, their role as a casino does raise doubt about their influence on aggregate economic growth. Does the focus on financial services and not on value creation impact income inequality, which it is argued by many puts a damper on productivity growth? In the USA there has been a disconnect since the 1970s between TFP growth and wage growth. This has been pointed out by a number of scholars. Sickles et al. (2016) have assessed counterfactuals based on relinking productivity growth and wage compensation during this period and have provided evidence that income inequality in the USA would be more comparable to present day Sweden than the countries with which it shares its current ranking (e.g., Iran and Bulgaria) had productivity growth been distributed to wage compensation as it had been in the post WWII years through the early 1970s. Income inequality is both a determining factor of productivity growth and a factor determined by it and demands more attention by scholars to measure its impact and disentangle the causal links between productivity growth and income inequality (Jones, 2015).

As creating value and capturing it as profit is the function of the business model, value must be created. The business model is developed to lay out the "the logic of the firm, the way it operates, and how it creates and captures value for its stakeholders" (Brandenburger and Stuart, 1996). Brea-Solís et al. (2015) point out the characterization of business models by Casadesus-Masanell and Ricart (2007, 2010, 2011) models by Casadesus-Masanell and Ricart (2007, 2010, 2011) and Casadesus-Masanell and Zhu (2010). They argue that the business model is grounded in the choices that are under the control of management, such as assets, policies, the governance of those policies and assets, and the consequences of these choices, which cannot be controlled by management. Brea-Solís et al. (2015) use this description of the business model and management's role in executing it in a very insightful case study of the sources of profit variations over time over the tenure of three of Walmart's CEO's: Sam Walton (1970–1988), David Glass (1998–2000), and Lee Scott (2000–2008).

[1] Recall the *Economist* article in 1999 heralding Houston as the City of the Twenty-First Century. The Jewel in the Crown was Enron. Enron was a finance company that sold itself as something very different. At the end of the day, when it was liquidated several years later, it was clear that its assets were often somewhat dubious titles to titles of titles to assets. There was no value added.

How does management assess the success of these practices? It typically does so by recognizing that, although innovation is the key aspect of a production process that management can impact, it is innovation in the workplace, or in-process innovation, that is impacted most by effective management. How is this measured?

In our brief discussion of management practices and their linkages to productivity we also can reconsider the debate that occurred during the 1970s between Harvey Leibenstein with his view of X-inefficiency (Leibenstein, 1966) and George Stigler (1976) with his view of internal inefficiency.[2] Stigler's rather condescending comment that "We may sympathize with Leibenstein's desire to associate his X-efficiency with economic behavior, but this shotgun marriage is not fertile" is still a view harbored by many neoclassical economists who stay true to assumptions that market forces result in only short-run divergences between first-order conditions for optimizing behavior and thus can be ignored. Indeed, in his Presidential Address to the American Economic Association, Zvi Griliches (1994), who was then Director of the National Bureau of Economic Research's (NBER) Productivity Program and of course a pioneer in productivity research, commented in regard to NBER's Productivity Program: "Harvey Leibenstein's (1966) ideas about X-efficiency, or more correctly X-inefficiency, did not get much of a sympathetic ear from us." In his address Griliches, foreshadowing the work by Bloom (current Co-Director of the NBER's Productivity and Growth Program) and his colleagues, noted that X-inefficiency gains importance at times when the firm finds itself in disequilibrium or faces unexploited profit opportunities and develops innovative ways to close the gap between its current state and the "frontier." [3]

X-inefficiency, and which management practices directly influence it, is important not when the firm finds itself in "disequilibrium" but also when it finds itself in equilibrium; that is, all of the time. Equilibrium is a factor that posits conditions under which optimizing behaviors should trend, much like the equilibrium concepts of strong attractors and cointegrating long-run and short-run relationships that are represented in error-correction models (Engle and Granger, 1987). Such issues are addressed later in this book in our chapter on dynamics and convergence from a stochastic and deterministic point of view. It is important to remember that management practices as drivers of innovation and efficiency change are important not only when the firm is in equilibrium or disequilibrium but also when the firm faces opportunities and,

[2] Much of this discussion is based on Lovell (2013).

[3] It is hard to overstate the influence that the late Zvi Griliches and his Harvard colleague Dale Jorgenson, whose current World KLEMS initiative is refocusing researchers on sector specific productivity issues through the world, have had on the study of productivity. Nerlove (2001), who himself contributed so much to duality theory and to the panel data literature whose approaches are utilized in Chapters 11 and 12, has a particularly insightful treatise on Griliches' particular contributions in his thoughtful and detailed discussion of Griliches' life and contributions to economics.

if unwilling to take advantage of opportunities, may incur the opportunity cost of not undertaking a particular strategy that may generate net value and hence profits.

There is indirect evidence for the presence and for concern by the management of their firm's inefficiency. The existence of a substantial consulting industry devoted to benchmarking analysis (that is, efficiency analysis) would suggest that the notion of inefficiency and its measurement using the techniques discussed in this book have passed the market test. It is clear that businesses and regulators use formal models to measure productivity and focus their attention on mechanisms to incentivize its diminution. Moreover, it is unlikely that anyone involved in a public/government service has not tried to find inefficiency and remove it.

Why are these measures taken by enterprises and their leadership? It is of course to improve performance, that is, to improve productivity. It is why the management consultant exists and is highly paid, it is the reason for cost cutting, rightsizing, etc. Although there is a semantic difference, and a difference that can be developed more fully at a theoretical level, the question of whether or not the presence of inefficiency results in the firm's reoptimizing or trimming the fat, there is little difference in fact. The firm is not doing as well as it could with the resources at its disposal relative to other firms with which it competes.

The evidence for the need for benchmarking/efficiency analyses is also direct. The Journal of Productivity Analysis (Springer) is a case in point. So is the work by Bloom and Van Reenen (2007), Bloom et al. Bloom et al. (2012, 2013). As Bloom et al. (2013) conclude: "improved management practice is one of the most effective ways for a firm to outperform its peers."

Behavioral Economics and Inefficiency in Production

As we discussed in the preface, behavioral economists have provided motivation for allowing nonoptimal allocations in production to be a testable restriction rather than an *a priori* assumption. That firms often operate inefficiently and in markets where inefficiencies can persist, is not a new concept nor a new empirical or theoretical insight. A treatment of production theory and productivity with optimal decisions a testable restriction is consistent with the recent Nobel Prize in Economic Sciences winner Robert Thaler's views on human nature (recalling that managers are the decision-makers on the supply side of the market just as consumers are decision makers on the demand side of the market). Decision making by inefficient firms was studied exhaustively by many economists and sociologists in the 1940s and 1950s. Such scholars as Harvey Leibenstein, George Katona, and Herbert Simon laid the foundation for the more recent cohort of behavioral economists and experimental economists such as George Akerlof, Daniel Kahneman, Richard Thaler, Vernon Smith, and Amos Tversky (Frantz et al., 2017). Efficiency and rationality are the cornerstones of modern neoclassical economics, but these cornerstones have always been fragile and in need of repair. John Maynard Keynes was clear in his views

on the subject, devoting the first page of his General Theory of Employment, Interest, and Money (1936) to the fact that the term "General" was meant to question *status quo* beliefs by the classical economists of his time who stifled intellectual developments (Frantz et al., 2017, p. 92). Similar points about "irrational" behavior were made by George Katona (1951) with his psychological economics and by Herbert Simon (1957) with his selective rationality. Such inefficient firm decision-making was treated extensively by Leibenstein and others, as documented in the excellent Routledge Handbook of Behavioral Economics (Frantz et al., 2017). Frantz (1997) notes a revealing and poignant observation made by Leibenstein in his discussion of Lionel Robbins' (1932) book, An Essay on the Nature and Significance of Economic Science in the forward to his X-Efficiency: Theory, Evidence and Applications. In Robbins's book economics is defined as the study of the efficient allocation of scarce resources in their alternative uses. Leibenstein's comment is:

> What got lost. . . was the businessman's idea and the engineer's idea of efficiency, which signify how well or poorly people and machines are working. Once allocative efficiency is combined with the maximization-of-utility or profits postulate there is no longer any room for the businessman's and the engineer's concept of efficiency. Thus, the idea disappeared that suboptimal operations by the firm and inside the firm are possible. . . Businessmen, engineers, and psychologists are aware of suboptimal behavior, but standard economic theory somehow does not easily or readily lend itself to the possibility of suboptimal operations.

Leibenstein's major work, X-efficiency theory, was based on an interdisciplinary perspective in efficiency analysis that brings together psychology, management and engineering studies. We also adopt this approach in our book by integrating the fields of operations research, engineering, management science (a field, it has been argued, that was begun at Carnegie Mellon University by Herbert Simon and one of the seminal contributors to the field of DEA discussed in this book, W. W. Cooper), statistics, and economics. Although the classical concept of an idealized rational optimizing agent has scientific merit, allowing for such behavior to be testable instead of assumed *a priori* leads to a research perspective that we hope will broaden the field of productivity analysis and result in a deeper and more meaningful theoretical and empirical body of knowledge that can be brought to bear on its study.

X-Efficiency in Production: Perspectives on Efficiency by Harvey Leibenstein

Before Harvey Leibenstein introduced X-Efficiency (XE) theory in 1966, economists had been studying efficiency, allocative efficiency specifically, based on the assumption that agents are rational and thus are utility maximizers. It was believed that firms could achieve full efficiency in a competitive market in which firms set prices equal to marginal costs. XE theory challenged the well-accepted assumption that firms try to minimize costs in

their production. Repeated observations of irrational behavior of firms and individuals have established the foundation of XE theory. The neoclassical theory is one of the cases XE theory considers when firms produce optimally in accordance with their production and cost functions. However, XE theory also incorporates irrationality in firms' behavior and allows for institutional environments that may lead to X-Inefficiency (XIE).

Sources of XIE are prevalent in economics. The agency problem is an important source for XIE. When managers pursue risky investments, or devote resources based on personal benefits rather than on the basis of increasing firms' profits – in the short run and in the long run – XIE occurs even if the firms are allocatively efficient. Market power also gives rise to inefficiency. Since the agency problem is more likely to exist when a firm has market power, monopolistic firms tend to have incentives for non-cost-minimizing behavior. In general, XIE is expected to be higher when government regulation, state ownership, or monopoly exists in the market. Leibenstein (1966, 1975, 1987) argued that inefficiency is persistent in production due to such agency problems in a firm, as well as due to information asymmetry and to monitoring required by regulatory oversight.

XE theory is in line with the ideas in productivity and efficiency literature that we explore in our book wherein total factor productivity growth is decomposed into technological change (innovation) and technical efficiency change (catch-up). XE has been incorporated into technical efficiency in hundreds of empirical studies on firm/industry efficiency. Empirical evidence from various industries, e.g., health care, telecommunications, airlines, and education, are consistent with the implications of XE theory (Frantz, 1997, 2007). XIE has been empirically estimated to be much more significant than allocative inefficiency. Studies (Kwan, 2006; Jiang et al., 2009; Fu and Heffernan, 2009) utilizing stochastic frontier analysis on the Chinese banking system, where periods of government deregulation had a different impact on banks' efficiency levels, estimate an average 25 percent XIE level from 1990 to 2015. Studies (Yao et al., 2008; Rezvanian et al., 2011) have also shown that Chinese banks' XE increased in periods after admission to WTO. Other studies on efficiency of financial institutions in various countries have shown consistent estimates with XE theory. US financial institutions, for example, have an average 20 percent XIE level from 1980s to 1990s, and they tend to have higher XE levels in periods after the banking deregulation (Bauer and Hancock, 1993; Mester, 1993; DeYoung, 1998).

As the literature on XE has flourished since Leibenstein introduced the theory, so has criticism. The most forcefully presented leisure-effort argument, for example, argues that the product of a firm also includes "non-traded" output which can be the health and leisure of employees. A decrease in employees' effort to pursue more leisure changes output rather than inefficiency, and a rational agent would not choose leisure if its cost outweighed its value. Such an argument that XIE does not exist and all inefficiencies are allocative is based on the assumption of complete information in the market for, and prices of, leisure by employees, which would appear to be unrealistic. Different schools

of criticisms are essentially tautological in the way that they disprove XE theory by assertion but they can't prove the existence of maximizing behavior. As Frantz (1997) raises the question in his book, "what becomes of the word maximize if non-maximize is not possible? Is the concept of efficiency important if the possibility of inefficiency is ruled out *a priori*?" The importance of efficiency remains as long as economics remains important.

Modeling Inefficiency in Production – A Road Map on How We Synthesize Different Perspectives on the Sources of Productive Inefficiency

The term "frontier firm" is ubiquitous in economics and business. In the relatively recent OECD Report Frontier Firms, Technology Diffusion and Public Policy: Micro Evidence from OECD Countries (Andrews et al., 2015) the term is used 336 times. In the recent *Economist* review of Haskel and Westlake (2018) and the book itself, the term is also used throughout as a key concept in various discussions. It is clear from the brief set of studies we have discussed in this introduction, many of which have not originated in the relatively recent literature on productivity and efficiency on which the remainder of the book is focused, that economists and business analysts have great interest and a large number of ways to approach the modeling and explanations of such production outcomes. It is the purpose of our book to provide the economics and business community with broad methodologies within an operational paradigm to identify such frontier firms, measure deviations from such benchmarks, analyze their sources, and to understand how other firms can improve their performance.

What we attempt to do in the remaining chapters is to provide a sound theoretical grounding in how such outcomes of the production process can be squared with models of optimizing behavior and how suboptimal and optimal production decisions can be measured, estimated, and interpreted. Most of the theoretical foundations are treated in Chapters 1–5. Chapters 6–10 discuss how productivity can be measured empirically, with Chapter 6 focusing on functional forms for estimating the technology and Chapters 7–10 focusing on various methods to estimate production outcomes, technical change, and efficiency levels and changes in them using linear programming methods from the field of operations research and statistics. Chapters 11–16 discuss a number of econometrically motivated topics, beginning with the stochastic production frontier paradigm and ending with methods to build consensus estimates from competing models (model averaging) and to impose regularity constraints on econometrically specified nonparametric representations of technology or its many dual forms. Our final Chapter 17 speaks to the main sources of data that productivity researchers have and are using, and focuses on the World KLEMS initiative and many recent studies using it that have appeared in the literature. We discuss a number of other data sets briefly and also introduce the various linear programming and econometrics software programs that are publicly available to estimate productivity and efficiency.