

1

INTRODUCTION AND PRELIMINARIES

1.1 A Nontechnical Overview

Economies rely on a rich set of different inputs to produce goods and services. The composition of the pool of productive resources varies dramatically across countries. Some countries have a lot of workers with little education and few highly educated workers, while others have an abundance of workers with many years of schooling. Some countries have a copious supply of natural resources, relative to their stock of productive machinery and structures, while others are resource-poor but have plenty of equipment. Some countries have a lot of people and little capital; others have high capital-labor ratios.

Do these big differences in the composition of the bundle of factors of production lead to correspondingly large differences in the way productive activities are organized? Do productive units in a country tailor their mode of production, choice of technology, and management practices to the particular patterns of scarcity and abundance of inputs in their country? And if so, how? Do they choose production methods which make the most of the abundant factors—giving up, so to speak, on those that are relatively limited in their countries? Or do they rather focus on maximizing the contribution of the inputs in limited supply, to make up for their scarcity?

Similar observations, leading to similar questions, can be made for a given country over time. In the space of years and decades, the relative supply of skills, equipment and structure, natural resources, and even workers belonging to different demographic

groups can change substantially. Do modes of production change and adapt in response to these changes? If so, does the adaptation compound the changes, by increasing the reliance of production units on the inputs that are becoming more abundant; or does it lean against the wind, by strengthening the ability to contribute to output of the factors of production that are becoming relatively scarce?

This book reports on my attempts to provide some answers to these questions. It shows that the mode of production—or the “production technology” in the words I use throughout the book—does vary systematically across countries, depending on their endowments of different factors of production. The production technology also changes over time, as factor supplies change. As to whether this adaptation favors the abundant or the scarce inputs, the answer turns out to be “it depends.” When the factors becoming scarce are not that different, in terms of the role they play in the production process, from those becoming abundant, the production technology adapts to maximize the impact of the abundant factors. On the other hand, when the scarce factors are special and are difficult to replace using the abundant factors, the production technology mutates to bolster the contribution of the scarce inputs.

Admittedly, none of the above sounds particularly surprising. What is perhaps more surprising is that these patterns of technology adaptation are completely ignored by the vast majority of economists’ thinking on how the organization of production varies across countries and (to a slightly lesser extent) over time. The dominant view, instead, is that some countries just have “better” technology than others, regardless of differences in factor endowment. Therefore, regardless of factor endowments, countries with “poor” technology should strive to copy as best they can the modes of organization they observe in countries with “good” technology. Similarly, over time technology mostly just gets “better,” lifting the productivity of all factors equally. This pervasive, and rather boring, view of technology is

incompatible with the evidence presented in this book: technology differences over space and time are much more interesting than most economists make them out to be!

1.2 A Slightly More Technical Overview

Economists characterize the relationship between a country's productive resources and its GDP by means of the *aggregate production function*. The aggregate production function can be used to answer two types of questions: (i) If country *A* has $x\%$ more of a given input (say, labor) than country *B*, by how much will country *A*'s GDP exceed country *B*'s (everything else being constant)? (ii) If country *A* experiences an $x\%$ increase in a given input between years t and $t + 1$, by how much will its GDP increase between the two years?

In empirical applications, economists have long noticed that production functions are not stable. Namely, the mapping from inputs onto outputs changes both across countries and over time. It is customary to refer to this instability as *technology differences* (across countries) and *technical change* (over time).

A long tradition of studies attempt to quantify the *pace* of technical change. This endeavor is usually referred to as *growth accounting*. Growth-accounting exercises usually find technical change to be very important in driving changes in GDP over time. There is also a more recent, but now well-established, strand on quantifying the *magnitude* of technology differences across countries. These *development accounting* studies tend to find that technology differences are very important in determining differences in GDP. Both these sets of findings have profoundly influenced the way economists think about economic performance in the long run.

While these traditions have been effective at quantifying the extent of technology differences and technical change, they have arguably been less successful at characterizing their nature. In the

vast majority of cross-country empirical applications technology is assumed to be *factor neutral*. Roughly speaking (and I will of course be more precise below) factor neutrality implies that if the efficiency with which a country uses one input is $x\%$ higher than the same input's efficiency in another, then the efficiency of all other inputs is also $x\%$ higher. Discussions of growth-accounting exercises are sometimes more nuanced and more aware of the possibility of technical change that is not factor neutral. Still, the methodology delivers a single quantitative measure of the pace of technical change, and is unsuited to characterize its nature. For this reason, growth-accounting results are still almost universally interpreted as if technical change was factor neutral.

This factor-neutral representation of technology is incompatible with various facts about factor prices across countries and time. As I will explain, if that view was correct, skill premia in (skill-) poor countries would be much higher, compared to premia in (skill-) rich countries, than they are; and skill premia couldn't have grown over time the way they have. Also, the capital share in income would vary much more across countries and over time than it does in the data. Changes in the experience premium over time are also inconsistent with neutral technical change.

This book uses these and other observations to show that technology differences and technical change are *factor biased*: they change not only the overall efficiency with which a country exploits its bundle of productive inputs, but also the relative efficiency with which different factors contribute to production. In fact, in some cases evidence shows that as the efficiency with which one input increases, the efficiency of another *decreases*. Allowing factor-biased technical change helps explain, among other things, why skill premia are very similar in poor and rich countries; why they have been growing over time; why the experience premium has not declined in response to the maturing of the baby-boom generation; and why the capital share in income is fairly constant both across countries and over time.

More specifically, I show that in richer countries the efficiency with which skilled labor is used relative to unskilled labor is

greater than in poorer countries; similarly, the efficiency with which reproducible capital (equipment and structure) is used relative to natural capital (mineral deposits, land, timber, etc.) is higher in rich countries; also, when comparing the efficiency of an overall bundle of labor inputs (appropriately combining skilled and unskilled labor) and an overall capital input (which combines reproducible and natural capital), rich countries use labor relatively more efficiently. Furthermore, the absolute efficiency with which physical capital is used appears to be not lower, and may even be higher, in poor countries.

Over time, I document (like others before me) an increase in the relative efficiency of skilled labor. I also find an increase in the relative efficiency of older workers relative to younger ones (holding skills constant). Finally, in an echo of the corresponding result in the cross section, the efficiency with which physical capital is used has been declining over time.

I interpret these findings by means of a simple theoretical model of endogenous technological choice. In the model, firms choose from a menu of technologies (production functions). The key consideration turns out to be the degree of substitutability among factors. When two factors of production are highly substitutable, firms choose technologies that maximize the efficiency of the cheaper factor (at the expense of the efficiency of the more expensive factor). Instead, when two factors are poor substitutes, firms choose to maximize the efficiency of the expensive factor.¹

To see how this framework sheds light on the empirical findings, consider skilled and unskilled labor. Rich countries have larger relative supplies of skilled labor, and hence skilled labor is relatively cheap there. Since skilled and unskilled labor are pretty good substitutes, firms in rich countries seek to make

¹ In this book I use the phrase “technology choice” to indicate a choice of parameters for the production function, i.e. a choice about the mapping between inputs and outputs. This is different from an alternative usage, encountered particularly but not exclusively in the classic trade literature, where technology choice refers to a choice of a particular input (e.g. capital-labor) ratio.

the most of skilled labor, and end up picking technologies that imply a high relative efficiency of skilled labor compared to poor countries. Rich countries also have larger relative supplies of physical capital (broadly construed to include natural and reproducible capital) compared to labor (broadly construed to account for the larger proportion of skilled workers). But labor and capital are (thought to be) poor substitutes, so in this case rich countries choose technologies that bolster the relative efficiency of capital. The other empirical patterns uncovered can be interpreted along similar lines.

The factor-neutrality approach implies a Manichean view where some countries “get it right” and others “get it wrong.” Countries either make the most of their skilled and unskilled labor, reproducible and natural capital, or they fail to use any of these efficiently. One implication is that poor countries should strive to reproduce rich countries’ technological choices, regardless of their factor endowments and other determinants of optimal technology choice. The nonneutrality findings in this book, and in the research on which this book is based, point to a more nuanced picture. To be sure, firms in poor countries lag far behind the technology frontier to which rich-country firms have access. But technology transfer and adoption should be selective and tailored to local conditions.

1.3 Aggregate Production Functions

The central—indeed the only—analytical tool used in this book is the aggregate production function. The aggregate production function is a mapping from a country’s input *quantities* to a country’s *output*, and we express it as

$$Y_{ct} = F_{ct}(X_{1ct}, X_{2ct}, \dots), \quad (1.1)$$

where Y_{ct} is aggregate output in country c in year t , X_{jct} is the quantity of input j used in production, and F_{ct} is the mapping in question. Note that the mapping carries subscripts c and t ,

indicating that the aggregate production function is *country and time specific*.²

The empirical counterpart of output Y_{ct} is gross domestic product (GDP). More specifically, when we are concerned with cross-country comparisons, we focus on GDP at purchasing power parity (PPP GDP). PPP GDP adds up the quantities produced of all final goods and services using a common set of prices (PPPs) as weights. When making comparisons over time, constant-price series must be used.

This book is about *how* F_{ct} varies across countries and over time. The *factor-neutral* case can be represented as

$$F_{ct}(X_{1ct}, X_{2ct}, \dots) = A_{ct} \tilde{G}(X_{1ct}, X_{2ct}, \dots). \quad (1.2)$$

In this special case, the production functions F differ by, and only by, the multiplicative term A . There already is a lot of literature documenting that A contributes substantially to changes in GDP over time and very substantially to cross-country differences in GDP.³

Factor neutrality is a natural first step in investigating cross-country technology differences as well as technical change, but a glance at equation (1.2) clearly shows that it is highly restrictive. The book focuses on the following conceptual generalization:

$$F_{ct}(X_{1ct}, X_{2ct}, \dots) = G(A_{1ct} X_{1ct}, A_{2ct} X_{2ct}, \dots). \quad (1.3)$$

In (1.3) the technology parameter A_{jct} augments factor j . A country (year) may have a relatively high value of one of the A_{jct} s

² I will not list here the conditions under which an economy admits an aggregate production function, but I do remind the reader that this list is exceptionally stringent. I am not aware of systematic attempts to assess the extent to which one can still count on aggregate production functions “approximately” holding when such assumptions are violated, as they certainly are. But everything in this book is subject to the (very big) caveat that something like an approximate production function still holds despite the certainty of violation of the theoretical underpinnings.

³ A brief overview of growth accounting (which studies changes in A_{ct} over time) with references to the classic contributions can be found in Caselli (2008a). A brief overview of development accounting (across countries) is provided in Caselli (2008b) and a detailed one in Caselli (2005).

without having a proportionally high value of another. In other words, technology differences need not be factor neutral—though neutrality is admitted as a possible special case. The book is about asking whether—and if so how—the A_{jct} s vary across countries and over time.⁴

To this end, we must begin by identifying the list of relevant factors of production. I focus on four broad aggregates: unskilled labor, skilled labor, physical reproducible capital, and natural capital. The breakdown of the main factors of production into labor and capital is almost as old as economics, and the breakdown of labor into skilled and unskilled is also well established. The importance of accounting for reproducible and physical capital has recently been emphasized by Caselli and Feyrer (2007).

We must also specify a functional form for G . The book applies methods originally developed by Caselli and Coleman (2002, 2006) and Caselli (2005) which allow for identification of the A_{jct} s when the production function features constant elasticities of substitution (CES). Accordingly, in most of the book, I work with the following specification:

$$Y_{ct} = [(A_{Kct}K_{ct})^\sigma + (A_{Lct}L_{ct})^\sigma]^{1/\sigma}, \quad (1.4)$$

$$K_{ct} = [(A_{Nct}N_{ct})^\eta + (A_{Mct}M_{ct})^\eta]^{1/\eta}, \quad (1.5)$$

$$L_{ct} = [(A_{Uct}U_{ct})^\rho + (A_{Sct}S_{ct})^\rho]^{1/\rho}. \quad (1.6)$$

Hence, the production process is represented by a sequence of nested CES aggregators. Beginning from the bottom, unskilled labor U and skilled labor S are combined into an aggregate labor input L with elasticity of substitution $1/(1 - \rho)$. Similarly, natural capital N and reproducible capital M (M for machine) are combined into the aggregate K , with elasticity of substitution $1/(1 - \eta)$. Finally, labor and capital are aggregated with elasticity

⁴ Clearly (1.3) remains restrictive in that it only admits technology differences of the factor-augmenting kind—there are no c or t subscripts to the function G .

$1/(1 - \sigma)$ to produce output. Technology differences are captured by differences in the factor-augmenting terms A_{Uct} , A_{Sct} , A_{Nct} , A_{Mct} , A_{Kct} , and A_{Lct} , which are the object of this study.⁵

The advantage of the nested-CES structure is twofold. First, it keeps the number of parameters (other than the augmentation factors A) to a minimum, i.e. the three elasticities of substitution. Second, as we will see, it allows for breaking up the problem of identifying the relative efficiency of any two factors into stages, i.e. first between skilled and unskilled labor, then between reproducible and natural capital, and only then between labor and capital. Admittedly other nestings are in principle possible, and little in the literature offers guidance on the most appropriate one. I have chosen the one in (1.4)–(1.6), as it is the most consistent with traditions emphasizing the distinction between skilled and unskilled labor, and labor and capital. Perhaps more important, the existence of these traditions provides (some) information on the plausible values of the corresponding elasticities of substitution.

With a slight modification (discussed below in section 1.8), the CES aggregates in (1.4)–(1.6) nest the Cobb-Douglas case as a special case. Macroeconomists often use the Cobb-Douglas assumption, particularly for (1.4), on the ground that the capital share is constant in the United States. The historical trendlessness of the capital share in the United States, however, can of course be replicated by CES models with the “right” time series behavior of the effective supplies of capital and labor (i.e. $A_K K$ and $A_L L$). Furthermore, there is clear evidence of substantial fluctuations in the capital shares of many countries other than the United States, and even there in recent years [e.g. Oberfield and Raval (2012), Elsy et al. (2013), Neiman and Karabarbounis (2014)].

⁵ In chapter 6 I add a further level of nesting “under” equation (1.6), where U and S are further broken down by the amount of experience.

1.4 Factor Bias

It is useful to establish a terminology to characterize particular patterns of variation of technology across countries and over time. To do so, I build on the terminology that was developed to characterize technical change over time, and extend it to the cross-country context.

Consider again an aggregator of the form

$$X = [(A_1 X_1)^\zeta + (A_2 X_2)^\zeta]^{1/\zeta}. \quad (1.7)$$

In the time series, it is customary to say that technical change is *factor- i augmenting* if A_i increases over time. Furthermore, technical change is said to be *biased toward factor i* if $(A_i/A_j)^\zeta$ increases over time.⁶

To see the rationale for the definition of factor bias note that

$$\frac{MP_i}{MP_j} \propto \left(\frac{A_i}{A_j}\right)^\zeta \left(\frac{X_i}{X_j}\right)^{\zeta-1},$$

where MP_i (MP_j) is the marginal product of factor i (j). Hence, technical change is biased toward factor i if it increases the relative marginal productivity of factor i when relative factor quantities are held constant. In recent years the idea of factor bias in technical change has played a prominent role in attempts to explain changes in the wage structure [e.g. Katz and Murphy (1992), Acemoglu (1998, 2002), Autor, Katz, and Krueger (1998), Katz and Autor (1999), Caselli (1999), Goldin and Katz (2008)].

In a cross section of countries, similar definitions are possible if we replace time with a suitable criterion to order observations. The natural criterion is income per worker. Hence, we will say that technology differences across countries are *factor- i augmenting*

⁶ The definitions of factor augmenting, neutral, and biased technical change go back to Hicks (1939).

if A_i is higher in countries with higher GDP. Furthermore, technology differences across countries are biased toward factor i if $(A_i/A_j)^\zeta$ is higher in countries with higher GDP.⁷

1.5 Alternative Representation

It is immediate that an alternative representation for an aggregator of the form (1.7) is

$$X = \Omega_1 [(X_1)^\zeta + \Omega (X_2)^\zeta]^{1/\zeta}, \quad (1.8)$$

where the mapping is

$$\begin{aligned} \Omega_1 &= A_1 \\ \Omega_2 &= \left(\frac{A_2}{A_1}\right)^\zeta. \end{aligned} \quad (1.9)$$

In words, we can work with aggregators that are specified in terms of the augmentation coefficients of both inputs or in terms of one augmentation coefficient and one factor-bias coefficient. In the book, I will exploit this representational equivalence extensively.

1.6 Plan for the Book

The book is divided into three parts.

Part I is the “across-countries” part. In chapters 2 and 3 I will use the specification in (1.8) and (1.9) for equations (1.6) and, respectively, (1.5), to identify the *factor bias* (if any) in labor and capital aggregation. In other words, in these chapters I (drop time subscripts and) estimate A_{Sc}/A_{Uc} and, respectively, A_{Mc}/A_{Nc} , and characterize how they vary across countries—particularly as a function of GDP. While these chapters produce estimates of the ratios A_{Sc}/A_{Uc} and A_{Mc}/A_{Nc} , they do not pin down the absolute

⁷ For a precedent on replacing the time index with a country’s ranking in the world income distribution see Hall and Jones (1996).

levels of A_{Uc} and A_{Nc} . As mentioned, I find that both $(A_{Sc}/A_{Uc})^\rho$ and $(A_{Mc}/A_{Nc})^\eta$ are positively correlated with income per worker.

In chapter 4 I turn to equation (1.4), which I keep in its original form. Substituting from equations (1.6) and (1.5), in their alternative form, we have

$$Y_c = [(A_{Kc}A_{Nc}\tilde{K}_c)^\sigma + (A_{Lc}A_{Uc}\tilde{L}_c)^\sigma]^{1/\sigma}, \quad (1.10)$$

where

$$\tilde{K}_c = \left[(N_c)^\eta + \left(\frac{A_{Mc}}{A_{Nc}} M_c \right)^\eta \right]^{1/\eta}, \quad (1.11)$$

$$\tilde{L}_c = \left[(U_c)^\rho + \left(\frac{A_{Sc}}{A_{Uc}} S_c \right)^\rho \right]^{1/\rho}. \quad (1.12)$$

These substitutions reveal that, in a system of nested CES functions, it is not possible to separately identify the augmentation coefficient of all inputs at all levels of the nesting. Accordingly, chapter 4 focuses on estimating the augmentation coefficients

$$\begin{aligned} \tilde{A}_{Kc} &= A_{Kc}A_{Nc}, \\ \tilde{A}_{Lc} &= A_{Lc}A_{Uc}. \end{aligned}$$

We can think of these coefficients as augmentation coefficients for “natural capital equivalents” \tilde{K} , i.e. the capital input expressed in efficiency units of natural capital, and “unskilled-labor equivalents” \tilde{L} , or the labor input in efficiency units of unskilled labor. My finding is that \tilde{A}_{Lc} is increasing in income per worker, while \tilde{A}_{Kc} is either unrelated or perhaps even slightly decreasing in income per worker.

Part III is the “over time” part. In chapter 6 I extend the definition of the aggregate labor input in (1.6) to further break down the skilled and unskilled labor aggregates by experience. This results in an additional layer of CES nesting. I then show that the efficiency of experienced skilled (unskilled) workers increases over time in the United States relative to the efficiency of inexperienced skilled (unskilled) workers. I also look at the evolution over time of the relative efficiency of skilled workers to unskilled workers and confirm the skilled-biased technical change (SBTC) result.

Chapter 7 extends the time series analysis to a panel of Organization for Economic Cooperation and Development (OECD) countries and investigates both skill bias in technical change and the evolution of the efficiency of labor relative to capital. The analysis confirms that SBTC is a global phenomenon. More originally, I find that in almost all OECD countries \tilde{A}_{Kct} has been declining over time.

Between the empirical parts I and III, in part II I pause for a theoretical interlude. I present a model of endogenous technological choice and use it to interpret the results of part I. At the end of part III I also return to the theoretical model to interpret that part's results.

1.7 Relation to Previous Work

All of the empirical results presented in this book are previously unpublished, in the sense that, at a minimum, they are obtained with data that have been updated with the most recent available sources. In most cases, however, I also extend previous work in various conceptual and methodological directions.

The analysis of skilled bias across countries in chapter 2 is based on Caselli and Coleman (2006). The data used in that paper refer to the year 1985 and covers a cross section of 52 countries. Here I report updated results on two cross sections: 1995 (66 countries) and 2005 (34 countries). I also improve very substantially on the methodology to construct the skilled and unskilled labor aggregates and to estimate the skill premium, which is a key input in backing out relative efficiencies.

The analysis of the relative efficiency of reproducible and natural capital in chapter 3 is novel to this book, though it is inspired by my work with Jim Feyrer [Caselli and Feyrer (2007)], which shows the importance of accounting for natural capital in estimating aggregate returns to capital across countries.

Chapter 4, which investigates how the efficiency of capital and labor (both broadly construed) varies across countries, updates the corresponding analysis in section 7 of Caselli (2005). There I

looked at 96 countries in 1996. Here I present estimates for 1995 (but with revised data) and 2005. I also measure both the capital and the labor aggregates differently. In particular, I include natural capital in the former, and allow for imperfect substitution between skilled and unskilled labor in the latter. I also present extensions not present in Caselli (2005) in which the health status of the population and cognitive skills are allowed to contribute to differences across countries in the labor endowment.

In chapter 6, the study of experience bias in the United States is novel to this book, though it is heavily indebted to the original investigation of this theme in Katz and Murphy (1992). So is the study of skill bias, which, however, is methodologically closer to Caselli and Coleman (2002). The exercise on the OECD panel in chapter 7 is novel to this book.

The two-factor theoretical model of endogenous technology choice in part II is from Caselli and Coleman (2006). The extension to four factors is novel to this book.

1.8 A Note on “Share Parameters”

Before starting, a quick note to reassure readers who find my representation of aggregators of the form (1.7) unfamiliar. It would indeed be more rigorous to write

$$X = [\omega(\tilde{A}_1 X_1)^\zeta + (1 - \omega)(\tilde{A}_2 X_2)^\zeta]^{1/\zeta}, \quad (1.13)$$

where ω is customarily referred to as the “share parameter.” This specification is more accurate because it allows retrieval of the Cobb-Douglas specification as the limiting case when $\zeta \rightarrow 0$. In this limit, ω and $1 - \omega$ are indeed the factor shares (hence the terminology).

The factor shares are omitted here exclusively for ease of notation. The reader should simply keep in mind that any estimate of A_1 (A_2) presented in the book is really an estimate of $\omega^{1/\zeta} \tilde{A}_1$ ($(1 - \omega)^{1/\zeta} \tilde{A}_2$), as is easily verified by comparison of (1.7) with (1.13).