The Substitution Structure of Production

De-Xing Guan*

October 9, 2023

Abstract

Economic growth and development is concerned with how and why the level and the growth rate of per capita income might change over time and across countries. Based on the endogenous growth theories of Lucas and Romer, we incorporate Coase's transaction costs into our model. Marshall's principle of substitution and Smith's idea about the make-or-buy decision are emphasized. Our model shows that to explain economic growth and development, we should better understand the substitution structure of production in the first place, and Shannon's information theory will play an important role in the determination and transmission of useful information, which is the cornerstone of economic growth and development.

Keywords: Substitution Structure, Transaction Cost, Growth and Development
JEL classification: O11, O30, O43

* Department of Economics, National Taipei University.

I. Introduction

By the substitution structure of production, I mean the structure of substitution between productive factors in the process of production. This substitution can happen at the intensive margin as well as at the extensive one. In the process of economic development, for example, free trade means that there is more substitution in goods and productive factors between trading partners, or more substitution at the extensive margin. On the contrary, protectionism would have less substitution between goods and factors. In this case of economic development, substitution at the intensive margin means substitution between domestic goods and factors. In this paper we will show that both intensive and extensive substitutions are necessary for a country to move from the system of agriculture to that of commerce.

Technology, property rights, institutions, culture, ideas are all important factors for a country to march into capitalism. But all of these factors are results of a common factor: the principle of substitution first proposed by Marshall (italics original):

As far as the knowledge and business enterprise of the producers reach…the sum of the supply prices of those factors which are used is, as a rule, less than the sum of the supply prices of any other set of factors which could be substituted for them; and whenever it appears to the producers that this is not the case, they will, as a rule, set to work to substitute the less expensive method…We may call this…*The principle of substitution*. The applications of this principle extend over almost every field of economic inquiry.[1]

The principle of substitution is actually concerned with the substitution at multiple margins, as Marshall said in the following paragraph:

Each man's actions are influenced by his special opportunities and resources, as well as by his temperament and his associations: but each, taking account of his own means, will push the investment of capital in his business in each several direction until what appears in his judgment to be the outer limit, or margin, of profitableness is reached…The margin of profitableness…is not to be regarded as a mere point on any one fixed line of possible investment; but as a boundary line of irregular shape cutting one after another every possible line of investment. This principle of substitution is closely connected with, and is indeed partly based on, that tendency to a diminishing rate of return from any excessive application of resources or of energies in any given direction.[2]

1 Marshall (1920, p. 284).
2 Marshall (1920, p. 296).

When firms and households produce goods and services, they are minimizing costs, given their objectives: profits for firms and utility for households. But there are many margins for them to choose factors of production to minimize costs. The substitution between factors of production is not at a single margin but at multiple ones. If there are not transaction costs, then, according to Coase (1988), all margins would shrink to a single one since there are no costs moving factors across different margins. This means that the substitution structure of production is determined by transaction costs. But, on the other hand, the ease of substitution within or across margins determines the magnitude of the transaction cost. The easier to substitute one factor for another, the lower the cost of moving factors within or across margins. Substitution structure and transaction costs are therefore two sides of the same coin: substitution structure determines the costs of transacting factors within or across margins, and transaction costs tell firms and households how easy they can produce through substituting one factor for another.

Following Lucas (2002) and Parente and Prescott (2005), a classical model means a model with production but without utility. In this sense the first classical model on economic development might be Lewis (1954). In this paper we use Lewis model as a benchmark to characterize the relationship between economic development and the substitution structure of production. In the tradition of Becker (1993) and by using the household production function, households can be viewed as producers. Both firms and households are therefore assumed to minimize costs subject to the production technology. The novelty here is that the full cost of either firms or households would include both prime cost and transaction cost which was emphasized by Coase (1988) and Ben-Porath (1980).

Section II is concerned with the so-called *Smith-Coase framework*, which is the workhorse of this paper. Section III briefly discusses the connection between the information theory and economic growth. Section IV is mainly about the substitution structure of production and economic development, and the role played by transaction costs. Section V concludes. The proof of Proposition 1 will be put in the Appendix.

II. The Smith-Coase Framework

Suppose that there is an aggregate production function: $Y = G(A, K, L, N)$, where Y is output, K is capital, L is labor, N is land or natural resource, and A is for idea. All factors of production are assumed to be private goods, that is excludable and rival, except that idea is an excludable but nonrival good, as argued by Romer (1990) and Jones (2011). Following Stokey (2001) and Parente and Prescott (2005), the supply of land is assumed fixed and can be normalized to unity such that the aggregate production function can be redefined as $Y = G(A, K, L, 1) \equiv H(A, K, L)$.[3]

The aggregate production function would be derived from a cost minimization problem, where the business firm makes a decision of staying with the original technology or switching to a new one, and the problem for this firm is to choose a less costly way to produce the same amount of final output. We call the original way of producing goods the intensive margin, and the new way the extensive margin, following the terminology often used in such field as labor economics. When the firm chooses the extensive margin, this margin would become a new intensive one because the firm would stay with it for at least a while. Then the firm would face another round of choice between this new intensive margin and a newer extensive one. And the process will not stop until the firm would no more change its positions.

The cost minimization problem of producers can be described as a two-stage problem. At the first stage those who would like to sell the final good, say coffee, in the market should learn how to accumulate their expertise in making coffee. Then these professional coffee makers use labor and capital to produce the coffee at the second stage. Now consumers have more choices because the cost of using coffee market has been reduced by the café and the specialized coffee makers.

To introduce Coasian transaction cost into our model let us assume that some efforts X are necessary in using markets to produce goods. These efforts include, for examples, searching for information, bargaining and negotiating, enforcing the contracts, and measuring the quality and quantity of goods. Without loss of generality, assume that the efforts of using markets are linearly related to the expertise of professional coffee makers, or assume that $A_E = \mu X$, where A_E represents the idea or expertise a typical professional coffee maker in the market would have in making coffee, and $\mu > 0$ is a variable representing the efficiency of using efforts to produce the expertise. A larger value of μ implies that professional coffee makers have better expertise such as more information, better knowledge and know-how, better skills in making coffee, and so on.

---

[3] In a more general case I am working on, following Stokey (2001) and Hansen and Prescott (2002), the supply of land is still fixed and normalized to unity, but the share of the rent of land would not be included as a part of capital share, as used in the present paper. In the Malthusian era, we can do this only when the rent of land was rent certain as adopted in England, where the rent was a lump-sum and would have no marginal effects on functional income distribution.

Let the price of efforts be P_X, that is, the cost of a unit of efforts in terms of the final good. Note that $1/\mu$ is the cost of producing a unit of expertise in terms of efforts, so P_X/μ is the cost of producing a unit of expertise in terms of the final good, which we define as marginal transaction cost (C^T) of producing market-made goods, or $C^T = P_X/\mu$. The efforts of using markets are factors of production and therefore intermediate goods of producing final goods. They are produced by other factors of production such as labor and capital. Assume that this production function is Cobb-Douglas: $X = K_E{}^{\beta} L_E{}^{1-\beta}$, such that we have $A_E = \mu X = \mu K_E{}^{\beta} L_E{}^{1-\beta}$, where L_E, K_E are labor and capital devoted to the accumulation of expertise, respectively, $0 < \beta < 1$.

Firms which want to enter into the extensive margin (say, to start a café in the city) have two options: producing the market-made goods by themselves or buying them from other firms in the market. If they choose the former they become sellers of the good, and they become buyers if they choose the latter. According to Coase (1988), transaction costs are the costs involved in using institutions such as markets, firms, and the law. When there are no transaction costs the equilibrium condition would require that the price of expertise be equal to the discounted sum of profits or net cash flow the expertise will generate, as described in Romer (1990). But when there are transaction costs the equilibrium arbitrage condition would require that

$$(1)\quad C^T + P_E = \frac{\pi_1}{1+r} + \frac{\pi_2}{(1+r)^2} + \ldots + \frac{\pi_n}{(1+r)^n}$$

where P_E is the price of A_E, π_i is the flow of profits generated by the expertise (such as license) in the *i*th period, and *n* is the duration of the expertise, the license, or the café, $i = 1,2,\ldots,n$. This means that after *n* periods either the license or the café will be out of date. Equation (1) indicates that the sum of the discounted profits or net cash flow of acquiring new expertise is equal to the full cost of doing so. And the full cost includes not only the cost of acquiring the expertise itself, but also the transaction cost of protecting and enforcing the property rights of it.[4]

[4] This could also be considered as a special case of the well-known *Coase Conjecture* (Coase (1972)). Actually I believe that Romer knew this problem, but what he had done was to assume it away! In p. S82 of his 1990 paper he said that “It is also easier to assume that the firm that buys a design…rents its durables instead of selling them outright…this shows that there are market mechanisms that avoid the usual durable-goods-monopoly problem.”

After acquiring the expertise (or license) people have to provide some efforts for protecting and enforcing their property rights. The price of doing this is P_X, as discussed above, and the full cost would be $P_E A_E + P_X X = FA_E$, where F is the unit full cost of the expertise. When full cost is greater than net cash flow, people would have less incentives to learn new skill in making coffee; otherwise they would like to learn more. In equilibrium the full cost must be equal to the net cash flow of learning the expertise. Note that $P_X = C^T \mu$ and $A_E = \mu X$, so $P_X X = C^T A_E$. This implies that $FA_E - C^T A_E = P_E A_E$, or simply $F = C^T + P_E$. In equilibrium the full cost *F* is obviously the full price of the expertise. In terms of Smith and Coase, total cost can be divided into two parts: prime cost and transaction cost. $P_E A_E$ and $C^T A_E$ are the prime cost and the transaction cost of producing the expertise, respectively. Because $P_E A_E = P_E \mu X = P_E (P_X / C^T) X$ and $C^T A_E = P_X X$, the transaction cost to total cost ratio is $1/(1+(P_E / C^T)) = C^T / (C^T + P_E)$ and the prime cost to total cost ratio is $(P_E / C^T)/(1+(P_E / C^T)) = P_E / (C^T + P_E)$, as expected.

Now we consider consumer's problem. Assume that consumers face a Smithian make-or-buy decision: to make the good by themselves or to buy it in the market. The purpose of consumers is assumed to get the good they want in the least costly way. According to the principle of comparative advantage, sellers in the market are usually better at producing goods than buyers. Because using markets is costly, buyers should pay transaction costs such that sellers are willing to bring goods to the market. The cost minimization problem of consumers can be described as follows:

(2) $C = Y \min\{\min(F, \gamma w^{1-\alpha} r^{\alpha}), (1-\gamma) w^{1-\alpha} r^{\alpha}\}$

where *C* is total cost of producing the final good, say *Y* cups of coffee, $w^{1-\alpha} r^{\alpha}$ is the unit cost of labor and capital in making coffee, where $0 < \alpha < 1$, and γ is the fraction of labor and capital devoted to the production of the final good at the extensive margin (in the market) such that $1-\gamma$ is the other fraction devoted to the intensive margin (at home), $0 < \gamma < 1/2$.[5] Professional coffee makers use the

[5] The expertise of professional coffee makers could be their knowledge concerning coffee, their skills

expertise together with their labor and capital to produce the final product. Equation (2) indicates that all the three factors of production: expertise, labor, and capital are necessary to make coffee in the market, but only labor and capital are required to make coffee at home. People can either produce coffee for themselves, or buy it in the market. They just choose the least costly way to have a cup of coffee.

In equilibrium the cost of making or buying coffee would be the same. Because there are three inputs: professional coffee maker's expertise, labor, and capital, the total cost function can be written as $C = FA_E + wL_N + rK_N$, where *w* is wage rate, *r* is rental price of capital, and L_N, K_N are labor and capital in producing coffee, the final output.[6] We assume that both labor and capital markets are competitive but the market for expertise is not. The first minimization problem inside the curly bracket of equation (2) requires that both expertise and labor/capital are necessary in making coffee in the market, that is, to solve the sub-problem:

(3) $C' = Y' \min(F, \gamma w^{1-\alpha} r^{\alpha})$

where $C' = FA_E + \gamma(wL_N + rK_N)$ is total cost of production in the market, and Y' is total output in the market, respectively. The solution to (3) is $C' = FY' = \gamma w^{1-\alpha} r^{\alpha} Y'$. This means that $FY' = C' = FA_E + \gamma(wL_N + rK_N)$, or $Y' = A_E + \gamma(wL_N + rK_N)/F$. Since the unit cost function is assumed to be Cobb-Douglas, an implication of this result is that $(1-\alpha)(wL_N + rK_N) = wL_N$. Putting these equations together will have

(4) $Y' = A_E + (wL_N)/[(1-\alpha)(w^{1-\alpha} r^{\alpha})]$

By Shephard's Lemma: $L_N = \partial C'/\partial w = \gamma(1-\alpha) w^{-\alpha} r^{\alpha} Y'$ and

$K_N = \partial C'/\partial r = \gamma \alpha w^{1-\alpha} r^{\alpha-1} Y'$, so $K_N / L_N = \alpha w/[(1-\alpha)r]$. Inserting this into (4) and rearranging terms would have

(5) $Y' = A_E + A_N' K_N^{\alpha} L_N^{1-\alpha}$

in making coffee, or any other know-how which ordinary people could not easily obtain. Because café needs both experts and ordinary workers, the fraction of labor/capital making coffee at home should not be less than one half. Otherwise no one will go to the café because the coffee is too expensive there.

[6] As will be shown later, the expertise is in turn produced by both labor and capital, and the aggregate production function will be a weighted average of the outputs produced by people at home and those in the market, with the weights being the fractions of labor and capital allocated to these two kinds of production.

where $A_N' = [(1-\alpha)/\alpha]^{\alpha}/[(1-\alpha)]$.

The solution for the second cost minimization problem outside the curly bracket of equation (2) requires that the total cost of making and buying coffee would be the same in equilibrium, so we have

$$(6)\quad (F + \gamma w^{1-\alpha} r^{\alpha})Y = (1-\gamma) w^{1-\alpha} r^{\alpha} Y$$

The solution to equation (6) is equivalent to that of the following redefined problem:

$$(7)\quad C = Y \min\{F, (1-2\gamma) w^{1-\alpha} r^{\alpha}\}$$

Now the market product and the home product are aggregated such that total cost is $C = FA_E + wL_N + rK_N$. The solution to equation (7) is $C = FY = (1-2\gamma) w^{1-\alpha} r^{\alpha} Y$. This means that $FY = C = FA_E + wL_N + rK_N$, or $Y = A_E + (wL_N + rK_N)/F$. Again by Shephard's Lemma, we can get the following equation:

$$(8)\quad Y = A_E + A_N K_N{}^{\alpha} L_N{}^{1-\alpha}$$

where $A_N = A_N'/(1-2\gamma)$, $0 < \gamma < 1/2$. Equation (8) is the aggregate production function based on the assumption that economic agents are following the least cost principle to solve the Smithian make-or-buy problem.

In this paper we assume that $\beta > \alpha$. This means that the marginal productivity of per capita capital at the extensive margin (acquiring expertise) is greater than that at the intensive one (no-expertise efforts), or that the production function of goods made in the market has larger marginal product than that made at home. Otherwise there are no consumers who would buy goods in the market if their qualities or convenience are the same. These two margins are illustrated in Figure 2. Combining $A_E = \mu X$ with equation (8) gives rise to the following aggregate production function:

$$(9)\quad Y = \mu K_E{}^{\beta} L_E{}^{1-\beta} + A_N K_N{}^{\alpha} L_N{}^{1-\alpha}$$

Before the completion of the model, we first explore the relation between the aggregate production function and the market equilibrium of final goods. First, when there are no transaction costs ($C^T = 0$), $P_E = F$, and this is the standard arbitrage equilibrium condition: at the margin, the cost of buying the good is equal to the discounted sum of profits (or monopoly rent) generated by selling this good in the market. But when $C^T \to 0$, $\mu = P_X / C^T \to \infty$, so $X = A_E / \mu \to 0$: no efforts will be devoted to using the market. This contradicts the fact that using markets is costly in the real world. The second aspect is that when a firm would like to buy goods in the market place it must pay the costs involved in using the market. If it does pay the full price, that is, prime costs plus transaction costs, then its demand for the good becomes Smith's *effectual demand*; otherwise, it is an *absolute demand*.[7] Obviously here the effectual demand is represented by the full price $P_E + C^T$ such that without paying for transaction costs, the firm's demand would become absolute and it will not be realized in the market. The firm must pay not only the prime cost but the transaction cost to bring the good to the market. The firm would buy nothing if it only pays for the fixed cost. Another implication of equation (6) is that, for any goods to be effectively brought to the market, marginal benefits (rent) must exceed marginal costs (transaction cost) of doing so, or $F > C^T$. If the benefit fails to be larger than the cost, no new goods would be created. In the extreme case that $C^T \to \infty$, it is too costly for the firm to start a new business, such that there are no new goods to be produced at all. Mathematically, $A_E = P_X X / C^T \to 0$ as $C^T \to \infty$.

To close this model we need market-clearing conditions for both labor and capital. Assume that there is a θ fraction of people who would like to learn the expertise, where $0 < \theta < 1$, and the remaining $1-\theta$ has two choices: γ fraction of it would choose to work at the extensive margin (in the market), while $1-\gamma$ of it would work at the intensive margin (at home). For simplicity, we also assume that the proportions of capital employed at these two margins are the same as those of labor. Again nothing important would be changed if this assumption were relaxed. The labor and capital markets clear if $L_E + L_N = L$ and $K_E + K_N = K$, where L, K are the aggregate supply of labor and capital, respectively. When all markets clear, equation (9) would become

(10) $$Y = \theta\mu K^\beta L^{1-\beta} + (1-\theta)[\gamma/(1-2\gamma)]A_N' K^\alpha L^{1-\alpha} + (1-\theta)[(1-\gamma)/(1-2\gamma)]A_N' K^\alpha L^{1-\alpha}$$

$\theta\mu K^\beta L^{1-\beta}$ is the fraction of skilled labor and capital in accumulating the expertise. $(1-\theta)A_N K^\alpha L^{1-\alpha}$ can be decomposed into two parts: $(1-\theta)[\gamma/(1-2\gamma)]A_N' K^\alpha L^{1-\alpha}$ and $(1-\theta)[(1-\gamma)/(1-2\gamma)]A_N' K^\alpha L^{1-\alpha}$. The first part is the fraction of unskilled labor and

7 For the distinction between effectual demand and absolute demand, see Smith (1776, Bk. I, Ch. VII).

capital devoted to making coffee in the market, and the second part is the fraction devoted to making coffee at home. Equation (10) then characterizes the aggregate production possibility frontier. It is a weighted average of the production functions at both intensive and extensive margins.

All of these results can be illustrated by Figure 1. In a world without transaction costs, no market-made goods would be produced because using the market is not costless. This implies that $A_E = 0$, and the point *B* in Figure 1 will shrink to the origin immediately. In a world with positive transaction costs there are two situations. First, if transaction costs are no less than the rent the firm might earn from its production of the new good, that is, if $F \leq C^T$, then obviously no goods will be produced. The point *B* in Figure 1 will again shrink to the origin. Second, if $F > C^T$, then the new good will be produced, and in equilibrium, $F - C^T = P_E > 0$, a positive price which is necessary for A_E to exist.

Transaction costs, therefore, act as thresholds to the introduction of new ideas or new goods into the economy. When transaction costs are lower because of better legal system, more information, less unnecessary lawsuits, less political conflicts, among others, point *B* in Figure 1 will move rightward to point *B'*, and the intersection point of the two production functions (point *A*) will move upward along the production curve at the extensive margin to another newer extensive margin (to point *A'* in equilibrium). This is because now the firm would have better expertise due to the reduction of transaction costs. This process will go on and on if more transaction costs are reduced and therefore better institutions are established. The long-run aggregate production possibility frontier will be the upper envelope of the production functions at various margins. There is always another better extensive margin out there for people to pursue if they can find a better way to get to it.

In the long run the model economy will grow along the balanced growth path (BGP). In particular per capita output and per capita capital will grow at the same rate at the BGP, or

$$(11) \quad \frac{\dot{y}}{y} = \frac{\dot{k}}{k} = g$$

where $\dot{y}$, $\dot{k}$ are the time derivatives of per capita output y and per capita capital

k, respectively. The common growth rate at the BGP can be calculated through equation (9). Simple calculation results in the following equation:

$$(12)\quad \frac{\dot{y}}{y} = \eta\left(\frac{\dot{\mu}}{\mu} + \beta\frac{\dot{k}}{k}\right) + (1-\eta)\alpha\frac{\dot{k}}{k}$$

where $\eta = \theta\mu k^{\beta} / y$ and $1-\eta = (1-\theta)A_N k^{\alpha} / y$, $0 < \eta < 1$. Using equation (10), we can calculate the BGP growth rate as

$$(13)\quad g = \frac{\eta g_{\mu}}{1-\alpha+(\alpha-\beta)\eta},$$

where $g_{\mu} = \frac{\dot{\mu}}{\mu} = \frac{\dot{P}_X}{P_X} - \frac{\dot{C}^T}{C^T}$. From equation (13) we can find what the factors that determine the long-run growth rate might be. First, larger capital shares in both margins, that is larger α and β, would have higher growth rates. This is a standard result in almost all growth models. Second, a lower growth rate of transaction costs would result in a higher growth rate of μ and therefore a higher growth rate of total product. The decrease in the growth of transaction costs would have growth effects. This implies that any country that has a better institution would grow faster. If the decrease in transaction costs is in their level, not in growth rates, then the result is still the same. This is because a smaller C^T means a larger μ and hence a larger η, and this would most of the time imply a higher growth rate of total product. A simple mathematics can show this:

$$\frac{\partial g}{\partial \eta} = \frac{(1-\alpha)g_{\mu}}{[1-\alpha+(\alpha-\beta)\eta]^2} > 0$$

if and only if $g_{\mu} > 0$, since we have assumed $0 < \alpha < 1$. Unless the growth rate of μ is negative a larger η would imply a larger g. Finally, a larger θ has a similar effect as a smaller C^T because both of them imply a larger η, the extensive-margin share of total output.

III. Information Theory and Economic Growth

Endogenous growth theory began with a conversation between Lucas and Romer, when the former gave an advice to the latter that "Well, why don't you use external increasing returns?"[8] What Lucas had in mind was Marshall's theory of increasing returns, which was meant by requiring the marginal product of capital (MPK) to be an increasing function of the capital which would have positive external effects on MPK. Nevertheless, Marshall had a clear but different definition of increasing returns, as he said that (italics original)

> while the part which nature plays in production shows a tendency to diminishing return, the part which man plays shows a tendency to increasing return. The *law of increasing return* may be worded this:--- An increase of labour and capital leads generally to improved organization, which increases the efficiency of the work of labour and capital.[9]

Marshall's increasing return emphasizes that MPK is not increasing with capital, but with organization, the fourth factor of production of Marshall.[10] Because improving organization is costly, this means that institutions and transaction costs are important if we want to deal with the problem of external increasing returns.

Romer (1990) initiated the second generation of endogenous growth theory which emphasizes that ideas are excludable but nonrival factors of production. This theory departed from the first generation models of Romer (1986) and Lucas (1988), which were based on the assumption of perfect competition and external increasing returns of the physical or human capital. Romer (1990) assumed monopolistic competition in the production of ideas, and there is a fixed R&D cost needed to be incurred for the firm to create a new idea. But Lucas did not totally agree with Romer, as he said that

> Are these ideas the achievements of a few geniuses, Newton, Beethoven… as external to the activities of ordinary people? Are they the product of a specialized research sector, engaged in the invention of patent-protected processes over which they have monopoly rights?... neither seems to me central. What is central, I believe, is the fact that the industrial revolution involved the emergence (or rapid expansion) of a *class* of educated people… who spend entire careers exchanging ideas, solving work-related problems, generating new knowledge.[11]

---

8 Warsh (2006), p. 208.
9 Marshall (1920), p. 265.
10 Coase (1988), p. 35.
11 Lucas (2009), p. 1.

Obviously, R&D is important in explaining the creation of new ideas, and is central to Romer, but this is not the case for Lucas. His response was to initiate a new theory in which people exchange ideas and create new knowledge. The basis of this third generation endogenous growth model was what Kortum (1997) called the *technology frontier*. When people meet randomly, they would exchange ideas from one another, and the resulting knowledge of everyone is the maximum of all knowledge involved. This means that exchanging ideas would widen both individual's knowledge and the technology frontier of the society.

Mathematically, the core of the third generation endogenous growth model is the Boltzmann equation in statistical mechanics. This equation appeared in Lucas (2009, 2018), Lucas and Moll (2014), and many other applications of this framework. But, as we have shown above, the transmission of information, and hence ideas, is not free.[12] People needs time and efforts to search, transfer, and absorb information coming from others. The existence of the cost of using information implies that there are noises in the process of the transmission of information.

Sims (2003) first noticed the importance of applying Shannon's information theory to the study of economic problems.[13] He called his work about the information cost a *rational inattention* problem. The cost can be either physical or psychological, and in the latter case it is indeed the *cognitive cost*,[14] as in the literature of psychology and behavioral economics.[15]

The information cost in the literature of rational inattention is what Shannon (1948) called *equivocation* in communication theory. Equivocation is the measure of noises or *average ambiguity*. Perfect information is impossible because there is always noise. The purpose of economic agents is therefore not to pursue perfect information, but the maximum useful information or *mutual information*.[16] Theories based on Boltzmann

or Gibbs distribution are usually assuming a noiseless world without considering the ambiguity, or uncertainty in general. Economists often use variance as the measure of uncertainty, but it is at most a measure of risk. This is because uncertainty, as noted by Knight (1921), is unmeasurable. All moments of probability distribution are important

---

[12] Romer (2015) had a similar opinion about the transmission of information.
[13] Actually Sims (1998) had an appendix titled "A Brief Introduction to Information Theory," but he did not mention Shannon there.
[14] Thaler and Sunstein (2009, p.8).
[15] For a review of the theory of rational inattention, see Maćkowiak, Matějka, and Wiederholt (2023).
[16] This is standard in communication literature, see Cover and Thomas (2006). What Shannon (1948) used was the *rate of transmission*, and Kelly (1956) called it the *information rate*.

in the determination of uncertainty, not just the first two moments of it.

IV. Substitution, Transaction Cost, and Economic Development

If transaction costs are so important for the economic performance of a country, then we might ask what are the factors underlying these costs? As discussed in the last Section, there are at least three kinds of such costs, namely, costs of searching for information, bargaining and negotiating, and enforcing the contracts. But there is a tautological problem here. Take the information cost as an example. The creation and transmission of information is costly, and if any information is available without incurring any cost, it must be useless or just common knowledge. The same argument can be applied to the discussion of transaction costs. When people buy cars, they search for the cheapest one with the quality satisfying them. So why do they search? This is because their information is imperfect. But why is the information imperfect? This is because searching for useful information is costly. And why is useful information costly? This is because most of the information is not common and therefore imperfect. We get back to the starting point of the argument: a tautology. If transaction costs are defined as any costs involved in using institutions, such as markets, firms and the law, there is still a tautology. For example, why using markets is costly? This is because to discover the price needs costs. And why is that? The answer is that buyers and sellers have to search for information, bargain with each other, and enforce the contracts they have agreed with. All of these activities are costly. But why are they costly? In reality we must know it is true, but in theory the answer would be because the information is imperfect, the bargaining power is asymmetric, or the contract is incomplete. All of these explanations are certainly true, but they are still tautological. Admittedly, any theory in some sense is inevitably tautological. In this Section we want to propose another explanation of transaction costs, though the reader might argue that it is still tautological. Anyway, it is just another explanation.

1. Substitution and Transaction Cost

First of all let us reconsider what perfect competition really means. Usually it is a situation where there are at least perfect information, homogeneous goods, and free entry and exit. Information is perfect only if the cost of creating and transmitting information is zero. Free entry and exit means that the cost of entry and exit is zero.

Both conditions indicate that perfect competition is a situation where there are no transaction costs, but the case of homogeneous goods is not easy to explain in this way.[17] To avoid tautology and to reconcile the condition of homogeneous goods with other criteria of perfect competition, we use Proposition 1 to organize our thoughts:

Proposition 1: Perfect competition is a situation where there are no transaction costs. In a world with positive transaction costs, it is impossible for all markets to be perfect competition. The smaller the transactions costs, the larger the elasticity of substitution between factors of production at different margins such that markets will be more, but never be perfectly, competitive.

The first sentence of Proposition 1 was actually proposed by George Stigler. Coase has clearly described this: "Stigler states the Coase Theorem in the following words: "… under perfect competition private and social costs will be equal." Since, with zero transaction costs, as Stigler also points out, monopolies would be induced to "act like competitors," it is perhaps enough to say that, with zero transaction costs, private and social costs will be equal."[18] The market of idea in our model acts an example to see if Stigler's statement is right. The marginal private benefit of selling an idea is P_E, but the marginal social benefit generated by the idea is F, which is the sum of discounted future profits. On the other hand, $P_E + C^T$ is the marginal private cost of buying this idea, and P_E is the marginal social cost.[19] Since in the idea market equilibrium, $P_E + C^T = F$. It is obvious that we have $P_E = F$ if $C^T = 0$, or private benefits will be equal to social benefits if there are no transaction costs. Similarly, we have $P_E + C^T = P_E$ if $C^T = 0$, or private costs will be equal to social costs if there are no transaction costs. Both Stigler and Coase were right.

Now let us consider the rest of the Proposition. Without loss of generality, we use

[17] That goods are homogeneous reflects the fact that either there is only one good or the cost of searching for the quality and quantity of goods is zero such that people, for example, can always pick out the same good from different stores without incurring any information cost. This might be confused with monopolistic competition where there is product differentiation without transaction costs.

[18] Coase (1988, p. 158).

[19] Coase (1988, p. 158) has defined social and private costs as follows: "Social cost represents the greatest value that factors of production would yield in an alternative use. Producers…are not concerned with social cost and will only undertake an activity if the value of the product of the factors employed is greater than their private cost (the amount these factors would *earn* in their best alternative employment)." In the idea market there are positive externalities from nonrival ideas. The social benefit is therefore greater than the private one. In other words the social cost is *less* than the private one, as indicated in our example.

the growth model in this paper to do this work. Suppose that a firm chooses between extensive and intensive margins. If all factors of production are perfect substitutes between these two margins, then the solution to this choice problem is quite simple: it is indifferent between them. Either margin will produce the same output at the same costs. In such situation these two margins are actually reduced to a single one. The boundary between them just disappears.

If we can meaningfully separate the extensive margin from the intensive one, then it must be that some goods in some margins are not perfect substitutes, so that there are transaction costs in switching across boundaries of different margins. That some goods at some margins are imperfect substitutes and that there are positive transaction costs are on the different sides of the same coin! Margins can be interpreted as a production technology, a market, a good, a different time, an idea, a method or rule to rearrange factors of production, or a legal system. They can be goods. They can also be institutions. The substitution of different goods at different margins is always imperfect because there are many government restrictions, imperfect information, incomplete contracts, barriers to entry, and so on. This imperfect substitution reflects the transaction costs incurred by factors of production when people want to move them across different margins to minimize their costs of production. The worse of the institutions a country might have, the greater transaction costs there would be, and the less competitive the market is in that country. This is why many economists, such as North (1981), North, Wallis, and Weingast (2013), and Acemoglu and Robinson (2012), have tried to figure out what is the role that institutions might play in the analysis as well as in the process of economic growth and development.

In our model transaction costs are related to the ease of substitution at extensive and intensive margins or, in general, the ease of substitution at multiple margins. We use Morishima elasticity of substitution (MES) to measure the ease of substitution between margins.[20] The total cost function in our model can be described as

(14) $C = Y\{\theta P_X(1/C^T)w^{1-\beta}r^{\beta} + (1-\theta)w^{1-\alpha}r^{\alpha}\}$

[20] MES was first proposed by Michio Morishima in 1967. It has been considered as a better measure of the ease of substitution than the usual Hicks-Allen elasticity of substitution when there are more than two factors of production. The original MES assumed that the output is fixed. This might be inadequate in a growth model. Fortunately, Blackorby, Primont, and Russell (2007) proved that the net MES (with fixed output) is equal to gross MES (with changing output) if the production function is homothetic. Because the aggregate production function in our model is homothetic, these two definitions of MES are equivalent. We use net MES in this Section because it is easier to calculate.

The first term inside the curly bracket in equation (14) is the unit cost of producing goods at the extensive margin, and the second one is that at the intensive margin. MES is defined as

$$(15)\quad M_{ij} = \frac{P_i C_{ij}}{C_j} - \frac{P_i C_{ii}}{C_i}$$

where subscript *i* of the cost function indicates partial derivative with respect to the price of the *i*th productive factors. For simplicity, we only use one example to illustrate the economic implications of the MES. The case we choose is the elasticity of substitution between efforts *X* and labor *L*. The first factor only appears at the extensive margin but the latter at both two margins. This elasticity shows some important aspects of the ease of substitution between these two margins. The MES between efforts and labor is

$$(16)\quad M_{12} = \frac{P_1 C_{12}}{C_2} - \frac{P_1 C_{11}}{C_1} = \frac{\theta P_X (1/C^T)(1-\beta) w^{-\beta} r^{\beta}}{\theta P_X (1/C^T)(1-\beta) w^{-\beta} r^{\beta} + (1-\theta)(1-\alpha) w^{-\alpha} r^{\alpha}}$$

There are two aspects to see the relationship between transaction costs and MES. First, note that because $C^T = P_X / \mu$ and $\mu > 0$, if $C^T \to 0$, then $P_X \to 0$. This means that any efforts to delimit property rights are free of charge. This in turn means that no rights would be protected and therefore no R&D would be undertaken: $A_E \to 0$. The extensive margin would simply disappear, and there is only one (intensive) margin left. And because in our model there are no other margins, this reduces to the case of perfect competition. Another way to think about this aspect is to notice that if $C^T \to 0$, then $M_{12} \to 1$, and this means that the efforts enter into total cost function in a Cobb-Douglas way, the same as labor and capital. When all factors of production can be grouped in a Cobb-Douglas cost function, where the elasticity of substitution between any two of them is unity, the aggregate production function is also Cobb-Douglas. And this implies that the market of final good is competitive. Because now all markets of productive factors (including effort market) are also competitive, all markets in this model are competitive. On the other hand, from equation (16), if $C^T \to \infty$ then $M_{12} \to 0$. When transaction costs are restrictively high, no one could substitute any factors of production for those at different margins, and the efficiency of production and markets would greatly be reduced.

The second aspect of the relationship between transaction costs and MES can be illustrated below. A smaller C^T would induce a larger elasticity of substitution, or

$$(17)\quad \frac{\partial M_{12}}{\partial C^T} = \frac{-\theta P_X (1/C^T)^2 (1-\beta)(1-\theta)(1-\alpha) w^{-(\alpha+\beta)} r^{\alpha+\beta}}{\{\theta P_X (1/C^T)(1-\beta) w^{-\beta} r^{\beta} + (1-\theta)(1-\alpha) w^{-\alpha} r^{\alpha}\}^2} < 0$$

From equation (16) it is clear that if transaction costs decline, the elasticity of substitution between productive factors would increase, and it becomes easier for factors, goods, ideas, and all of the possible rearrangements of these resources to move between margins. This, together with the above result, confirms Proposition 1. Because the MES is not symmetric it is better to see if the counterpart of the above elasticity of substitution still has the same property as equation (16) has had. A similar calculation shows

$$(18)\quad M_{21} = \frac{P_2 C_{21}}{C_1} - \frac{P_2 C_{22}}{C_2} = 1 - \beta + \frac{\beta\theta P_X (1/C^T)(1-\beta) w^{-\beta} r^{\beta} + \alpha(1-\theta)(1-\alpha) w^{-\alpha} r^{\alpha}}{\theta P_X (1/C^T)(1-\beta) w^{-\beta} r^{\beta} + (1-\theta)(1-\alpha) w^{-\alpha} r^{\alpha}}$$

The same argument also applies here for the case of zero transaction costs. In particular, if $C^T \to 0$, then $M_{21} \to 1$, the same result as in the case of equation (16). The economic explanation is also the same, which is omitted here. Now take a look at the partial differentiation of M_{21} with respect to C^T:

$$(19)\quad \frac{\partial M_{21}}{\partial C^T} = \frac{-\theta P_X (1/C^T)^2 (1-\beta)(1-\theta)(1-\alpha)(\beta-\alpha) w^{-(\alpha+\beta)} r^{\alpha+\beta}}{\{\theta P_X (1/C^T)(1-\beta) w^{-\beta} r^{\beta} + (1-\theta)(1-\alpha) w^{-\alpha} r^{\alpha}\}^2} < 0, \text{ if } \beta > \alpha$$

The condition $\beta > \alpha$ is usually satisfied because the marginal productivity of capital at the extensive margin is usually larger than that at the intensive margin. Without loss of generality, we make this assumption. Smaller transaction costs again induce larger elasticities of substitution and, accordingly, more competitive markets. Equation (19) therefore further confirms Proposition 1. There are nine MES for the case of three productive factors. We will discuss the rest of these MES in the Appendix. All main results in this paper are unchanged.

2. Substitution and Economic Development

We have shown that the substitution structure is important for the explanation of economic development. Class struggle was in effect concerned with the internal substitution structure. Marx only saw the dark side of class struggle, but Smith had looked at both bright and dark sides of it. On the other hand, foreign trade has been a good example of the external substitution structure. We can use these two substitution structures and the Lewis model to characterize the three different stages of economic development, which we might call the stages of Malthus, from Malthus to Solow, and of Solow, respectively.

2.1. The stage of Malthus (stagnation): $\gamma = 0$, $\theta = 0$

In this stage there was no per capita output growth, there were no capitalists, the economic system was mainly agricultural, and there was no rural-urban migration. This stage corresponds to the case: $\gamma = 0$ and $\theta = 0$ in our theoretical model. When these two parameters are zero, the BGP growth rate $g = 0$, and there are no migration from the country to the city. Only landlords (including the king) would hire farmers, and the demand curve for the labor would in general not shift to the right, as shown by the far left demand curve in Figure 2. Because almost all of the rent was collected by landlords, farmers could not accumulate capital and landlords did not have any incentive to do so. The dearth of investment opportunities was the reality for most countries in the feudal society.[21] The period of stagnation might last for thousands of years since both the internal and the external substitution structures are needed for a nation to have free burghers and rural-urban migration.

2.2. The stage from Malthus to Solow (transition): $0 < \gamma < 1/2$, $\theta = 0$

In this stage there was still no per capita output growth, there were some free burghers or bourgeois, the economic system was still mainly agricultural but with some handcrafts working in the city, and there was some rural-urban migration but no ideas were produced in a commercial way. This stage corresponds to the case:

---

[21] Koo (2018) has discussed the relationship between the dearth of investment opportunities and economic development.

$0 < \gamma < 1/2$ and $\theta = 0$ in our theoretical model. When $\theta = 0$, the BGP growth rate $g = 0$. There are some migration from the country to the city because $0 < \gamma < 1/2$. Both landlords (including the king) and guilds would hire farmers such that the demand curve for the labor would shift to the right, but not enough to let real wage begin to increase.[22] This can be shown in Figure 2 by the shift of demand curves to the Lewis turning point. Most of the rent was collected by landlords, but now bourgeois could accumulate some capital. This capital is not large enough to generate sustained growth in the subsistence level of consumption and hence in the real wage. In terms of the growth theory, the level of per capita income might be increasing in this second stage, but there was still no long-run growth in per capita output.

2.3. The stage of Solow (growth): $0 < \gamma < 1/2$, $0 < \theta < 1$

In this stage there was sustained growth in per capita output, and free burghers or bourgeois were so many that the economic system gradually became commercial. Rural-urban migration was more popular, and capitalists began to produce ideas in a commercial way. This stage corresponds to the case: $0 < \gamma < 1/2$ and $0 < \theta < 1$ in our theoretical model. When $0 < \theta < 1$, the BGP growth rate $g > 0$. The migration from the country to the city was increasing. Capitalists became the main employers of farmers such that the demand curve for the labor would not only shift to the right, but would pass the Lewis turning point. Technology progress made sustained growth in real wage possible. In terms of the growth theory, both the level and the growth rate of per capita income might be increasing in this third stage.

Stages of economic development might not be only three. Koo (2018) proposed a fourth stage. But the number of development stages is not the point we would like to make in this paper. Different economists would certainly have different opinion about the stages of economic development. But the three stages described above have been the basic ones. History and many empirical studies have provided evidences about their relevance. Now we can use Proposition 2 to summarize the above results.

Proposition 2: The three basic stages of economic development can be characterized by using parameters in our theoretical model: (1) the stage of Malthus: $\gamma = 0$, $\theta = 0$, (2) the stage from Malthus to Solow: $0 < \gamma < 1/2$, $\theta = 0$, and (3) The stage of Solow: $0 < \gamma < 1/2$, $0 < \theta < 1$.

In terms of growth theory, γ is concerned mainly with the level effect, and θ

[22] This is consistent with the big-push theory of Murphy, Shleifer, and Vishny (1989).

with the growth effect. In terms of development theory, γ is concerned primarily with rural-urban migration, and θ with sustained growth by using and producing ideas. And in both theories the substitution structure of production and the corresponding transaction costs are important for explaining the performance of the growth and development of a nation.

V. Conclusions

In the opening chapter of the *Wealth of Nations*, Adam Smith said: “It is the great multiplication of the production of all the different arts, in consequence of the division of labour, which occasions, in a well-governed society, that universal opulence which extends itself to the lowest ranks of the people.”[23] Then he continued in the second chapter: “The division of labour, from which so many advantages are derived, is not originally the effect of any human wisdom… It is the necessary, though very slow and gradual, consequence of a certain propensity in human nature…; the propensity to truck, barter, and exchange one thing for another.”[24] And finally he wrote in the third chapter: “As it is the power of exchanging that gives occasion to the division of labour, so the extent of this division must always be limited by the extent of that power, or, in other words, by the extent of the market.”[25] From these passages it is clear that the logic of Smith has been that the extent of market causes or determines the extent of division of labor, and this in turn determines the production and thus opulence of the people. This is the great idea of Smith.

But one might ask a deeper question: what are the factors that determine the extent of market? There are many answers but Coase proposed the following heuristic one: “… without the establishment of this initial delimitation of rights there can be no market transactions...”[26] In other words, the prelude of market transactions, according to Coase, is the delimitation of rights. This is the great idea of Coase. It tells us that the market cannot function by itself alone. It is an institution, and using institutions is not costless. To delimit rights would incur transaction costs, so if there were no transaction costs, then there were no rights delimited unless the delimitation of them is costless. So if we want to understand the sources and processes of economic growth, then we must first find out what are the relevant transaction costs that would determine the extent of the market.

[23] Smith (1789; 1994, p. 12).
[24] Smith (1789; 1994, p. 14).
[25] Smith (1789; 1994, p. 19).
[26] Coase (1988, p. 104).

This paper has tried to build a theoretical model to incorporate transaction costs explicitly into the growth and development theory. We find that lower transaction costs would induce better institutions and therefore more rapid economic growth. We also find that it is easier to substitute factors employed at one margin for those employed at another, if transaction costs are lowered. Easy substitution of the productive factors between margins would result in more competitive markets that foster economic growth.

We also find that the stages of economic development can be characterized by the substitution structure of production. Those three basic development stages include stagnation (Malthus), transition (from Malthus to Solow), and growth (Solow). Based on the classical Lewis (1954) model, we have incorporated transaction costs into the endogenous growth theory of Romer (1990) to describe these development stages. More empirical studies are needed to see if this model or, more precisely, if the Smith-Coase framework could match the data of growth and development, and explain the facts we observe in real life.

Appendix

In the case of three factors of production there are nine MES, namely, M_{11}, M_{22}, M_{33}, M_{12}, M_{21}, M_{13}, M_{31}, M_{23}, and M_{32}, where $M_{11} = M_{22} = M_{33} = 0$, by definition. Let

(A1) $$M_{ij} = \frac{P_i C_{ij}}{C_j} - \frac{P_i C_{ii}}{C_i} = \varepsilon_{ji} - \varepsilon_{ii}$$

Then if we know these ε_{ij}, we get MES. We list all ε_{ij} as follows:

(A2) $$\varepsilon_{11} = 0$$

(A3) $$\varepsilon_{22} = \frac{-\beta\theta P_X (1/C^T)(1-\beta) w^{-\beta} r^{\beta} - \alpha(1-\theta)(1-\alpha) w^{-\alpha} r^{\alpha}}{\theta P_X (1/C^T)(1-\beta) w^{-\beta} r^{\beta} + (1-\theta)(1-\alpha) w^{-\alpha} r^{\alpha}}$$

(A4) $$\varepsilon_{33} = \frac{-\beta\theta P_X (1/C^T)(1-\beta)w^{1-\beta}r^{\beta-1} - \alpha(1-\theta)(1-\alpha)w^{1-\alpha}r^{\alpha-1}}{\beta\theta P_X (1/C^T)w^{1-\beta}r^{\beta-1} + \alpha(1-\theta)w^{1-\alpha}r^{\alpha-1}}$$

(A5) $$\varepsilon_{12} = 1-\beta$$

(A6) $$\varepsilon_{21} = \frac{\theta P_X (1/C^T)(1-\beta)w^{-\beta}r^{\beta}}{\theta P_X (1/C^T)(1-\beta)w^{-\beta}r^{\beta} + (1-\theta)(1-\alpha)w^{-\alpha}r^{\alpha}}$$

(A7) $$\varepsilon_{13} = \beta$$

(A8) $$\varepsilon_{31} = \frac{\beta\theta P_X (1/C^T)w^{1-\beta}r^{\beta-1}}{\beta\theta P_X (1/C^T)w^{1-\beta}r^{\beta-1} + \alpha(1-\theta)w^{1-\alpha}r^{\alpha-1}}$$

(A9) $$\varepsilon_{23} = \frac{\beta\theta P_X (1/C^T)(1-\beta)w^{-\beta}r^{\beta} + \alpha(1-\theta)(1-\alpha)w^{-\alpha}r^{\alpha}}{\theta P_X (1/C^T)(1-\beta)w^{-\beta}r^{\beta} + (1-\theta)(1-\alpha)w^{-\alpha}r^{\alpha}}$$

(A10) $$\varepsilon_{32} = \frac{\beta\theta P_X (1/C^T)(1-\beta)w^{1-\beta}r^{\beta-1} + \alpha(1-\theta)(1-\alpha)w^{1-\alpha}r^{\alpha-1}}{\beta\theta P_X (1/C^T)w^{1-\beta}r^{\beta-1} + \alpha(1-\theta)w^{1-\alpha}r^{\alpha-1}}$$

From equations (A2) to (A10) it is easy to calculate the MES, and we leave this for the interested reader. Simple calculation will reach the conclusion that $\partial M_{ij} / \partial C^T < 0$, and $M_{ij} \to 1$ as $C^T \to 0$, $\forall i \neq j$. All of these results confirm the Proposition 1 in this paper.

References

Acemoglu, Daron, “Distorted Innovation: Does the Market Get the Direction of Technology Right?” *AEA Papers and Proceedings*, 2023, 1-28.

--- and James A. Robinson, *Why Nations Fail*, Crown Business, 2012.

Becker, Gary S., *A Treatise on the Family*, Harvard University Press, 1993.

Ben-Porath, Yoram, “The F-Connection: Families, Friends, and Firms and the Organization of Exchange,” *Population and Development Review*, 1980, 1-30.

Blackorby, Charles, Daniel Primont, and R. Robert Russell, “The Morishima Gross Elasticity of Substitution,” *Journal of Productivity Analysis*, 2007, 203-208.

Coase, Ronald H., “Durability and Monopoly,” *Journal of Law and Economics*, 1972, 143-149.

--- *The Firm, the Market, and the Law*, University of Chicago Press, 1988.

Cover, Thomas A., and Joy A. Thomas, *Elements of Information Theory*, 2nd edition, Wiley, 2006.

Deane, Phyllis, *The First Industrial Revolution*, Cambridge University Press, 1965.

Ferguson, Niall, *Civilization: The West and the Rest*, The Penguin Press, 2011.

Galinsky, Adam, and Maurice Schweitzer, *Friend and Foe*, Crown Business, 2015.

Galor, Oded, and David N. Weil, “Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond,” *American Economic Review*, 2000, 806-828.

Hansen, Gary D., and Edward C. Prescott, “Malthus to Solow,” *American Economic Review*, 2002, 1205-1217.

Huang, Ray, *China: A Macro History*, revised edition, M. E. Sharpe, 1997.

Jedwab, Remi, Noel D. Johnson, and Mark Koyama, “The Economic Impact of the Black Death,” *Journal of Economic Literature*, 2022, 132-178.

Jones, Charles I., “Intermediate Goods and Weak Links in the Theory of Economic Development,” *American Economic Journal: Macroeconomics*, 2011, 1-28.

Kelly, John L., Jr., “A New Interpretation of Information Rate,” *Bell System Technical Journal*, 1956, 917-926.

Knight, Frank H., *Risk, Uncertainty, and Profit*, Hart, Schaffner, and Marx, 1921.

Koo, Richard C., *The Other Half of Macroeconomics and the Fate of Globalization*, John Wiley & Sons, 2018.

Kortum, Samuel S., “Research, Patenting, and Technological Change,” *Econometrica*, 1997, 1389-1419.

Lewis, W. Arthur, “Economic Development with Unlimited Supplies of Labour,” *The Manchester School*, 1954, 139-191.

Lucas, Robert E., Jr., “On the Mechanics of Economic Development,” *Journal of Monetary Economics*, 1988, 3-42.

--- “The Industrial Revolution: Past and Future,” in *Lectures on Economic Growth*, Harvard University Press, 2002, 109-188.

--- “Ideas and Growth,” *Economica*, 2009, 1-19.

--- “What Was the Industrial Revolution?” *Journal of Human Capital*, 2018, 182-203.

--- and Benjamin Moll, “Knowledge Growth and the Allocation of Time,” *Journal of Political Economy*, 2014, 1-51.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, “Rational Inattention: A Review,” *Journal of Economic Literature*, 2023, 226-273.

Malthus, Thomas Robert, *An Essay on the Principle of Population*, 1798; Oxford University Press, 1993.

Marshall, Alfred, *Principles of Economics*, 8th edition, Macmillan, 1920.

Marx, Karl, and Friedrich Engels, *The Communist Manifesto*, 1848; Oxford University Press, 1992.

McCloskey, Deirdre N., *Bourgeois Inequality*, University of Chicago Press, 2016.

Mokyr, Joel, *A Culture of Growth*, Princeton University Press, 2017.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny, “Industrialization and the Big Push,” *Journal of Political Economy*, 1989, 1003-1026.

North, Douglass C., *Structure and Change in Economic History*, W. W. Norton, 1981.

---, John Joseph Wallis, and Barry R. Weingast, *Violence and Social Orders*, paperback edition, Cambridge University Press, 2013.

Parente, Stephen, and Edward C. Prescott, “A Unified Theory of the Evolution of International Income Levels,” in *Handbook of Economic Growth*, Vol. 1B, edited by Philippe Aghion and Steven N. Durlauf, Elsevier B. V., 2005, 1371-1416.

Romer, Paul M., “Increasing Returns and Long-Run Growth,” *Journal of Political Economy*, 1986, 1002-1037.

--- “Endogenous Technological Change,” *Journal of Political Economy*, 1990, S71-S102.

--- “Mathiness in the Theory of Economic Growth,” *American Economic Review: Papers & Proceedings*, 2015, 89–93.

Shannon, Claude E., “A Mathematical Theory of Communication,” *Bell System Technical Journal*, 1948, 379-423 and 623-656.

Sims, Christopher A., “Stickiness,” *Carnegie-Rochester Conference Series on Public Policy*, 1998, 317-356.

--- “Implications of Rational Inattention,” *Journal of Monetary Economics*, 2003, 665-690.

Smith, Adam, *An Inquiry into the Nature and Causes of the Wealth of Nations*, 5th edition, 1789; Modern Library, 1994.

Stokey, Nancy L., "A Quantitative Model of the British Industrial Revolution, 1780-1850," *Carnegie-Rochester Conference Series on Public Policy*, 2001, 55-109.

Thaler, Richard H., and Cass R. Sunstein, *Nudge*, revised and expanded edition, Penguin, 2009.

Warsh, David, *Knowledge and the Wealth of Nations*, Norton, 2006.

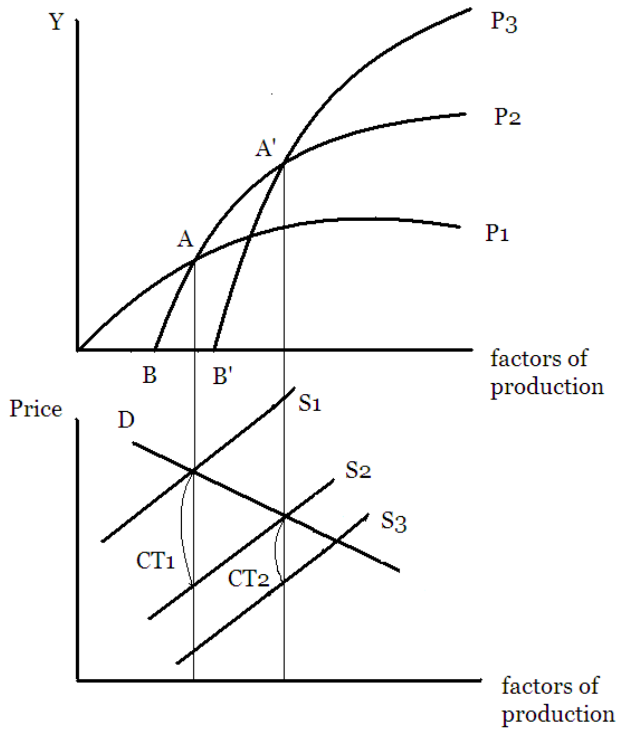
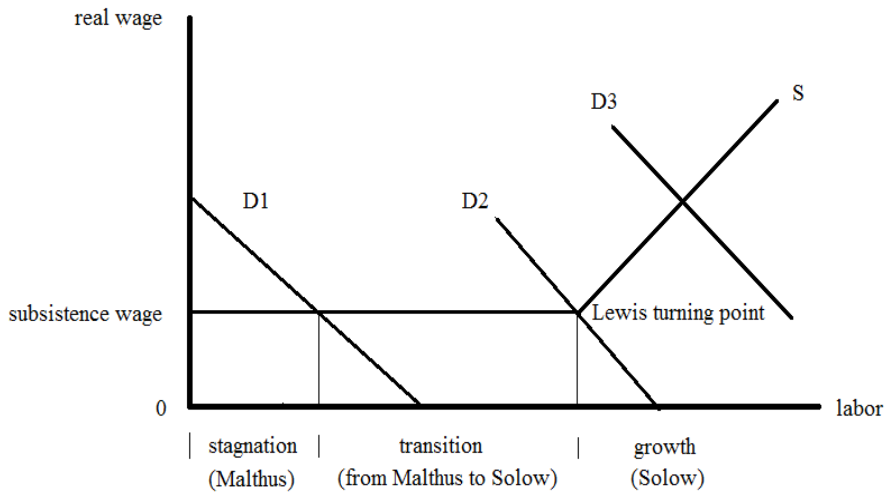


Figure 1: The Substitution Structure of Production


real wage
D3
S
D1
D2
subsistence wage
Lewis turning point
0
labor
stagnation
(Malthus)
transition
(from Malthus to Solow)
growth
(Solow)