CHAPTER ONE

Production and Distribution of Income in a Market Economy

THE AIM OF THIS BOOK is to study the implications of economic interactions between heterogeneous individuals, both for macroeconomic outcomes and for the evolution of the income and wealth distribution. As these interactions are extremely complex, we organize our analysis around several key simplifications.

First, we will assume throughout that there are *two factors of production*: an "accumulated" factor and a "non-accumulated" factor. We will frequently refer to the former as "capital" and to the latter as "labor." As we discuss below, however, the important point is that the economy's (as well as the households') endowment with the former is endogenously determined by savings choices, whereas the economy's endowment with the latter is exogenously given.

Second, we will assume throughout that *all individuals have the same attitude toward savings*, i.e., *that any two individuals would behave identically if their economic circumstances were identical*. This is not to say that heterogeneity in preferences between present and future consumption is unimportant in reality. Allowing for systematic differences across individuals along this dimension, however, would tend to yield tautological results: one might, for example, find that the poor are and remain poor due to their low propensities to save. It is much more insightful to highlight other sources and effects of large differences in incomes across individuals: we will highlight the role of macroeconomic phenomena (such as capital accumulation and associated changes in factor prices, market imperfections, and economic policies) for the dynamics of the distribution of income and wealth and their feedback to the long-run process of economic development. Heterogeneous propensities to save are clearly of some importance in reality, but will not induce a systematic bias in our results if they are random and unrelated to economic circumstances.

Third, in many of our derivations we will assume that only *one good* is produced in the economy and can be used for either consumption or investment. Investment then coincides with forgone consumption, to be understood broadly as leisure choices are subsumed in consumption choices. The single-good assumption is adopted throughout part 1 (with the exception of the appendix to chapter 4) and part 2. In part 3, we relax it and

consider the interrelation between distribution and growth when there are many goods and when the structure or consumption differs between rich and poor consumers.

As a further general principle, we will apply standard tools of modern macroeconomic analysis, formulating all models in formally precise and consistent terms. Even as we strive to take individual heterogeneity into account when studying macroeconomic phenomena, we will often find it useful to refer to situations where some or all of the implications of heterogeneity are eliminated by appropriate, carefully discussed assumptions, so that a representative agent perspective is appropriate for some or all aspects of the analysis. Specifying and carefully discussing deviations from these assumptions will make it possible to highlight clearly problems of heterogeneity and distribution, as well as their interaction with macroeconomic phenomena.

This first chapter sets the stage for our analysis. We introduce notation and set out basic relationships both at the level of the family and at the aggregate, making the important distinction between accumulated and non-accumulated income sources. Then, we analyze the relationship between distribution and the efficiency of production in a "neoclassical" setting of perfect and complete markets. Firms maximize profits and take prices as given, all factors of production are mobile, there is complete information, and all economic interactions are appropriately accounted for by prices (there are no externalities). In that setting we discuss in some detail the conditions under which macroeconomic aggregates do not depend on income distribution and on technological heterogeneity, so that production and accumulation can be studied as if they were generated by decisions of "representative" consumers and producers. As is often the case in economics, the model's assumptions are quite stringent, so we discuss briefly conceptual problems arising when certain tractability conditions are not met. In particular, if factors of production cannot be reallocated, aggregation becomes very problematic unless stringent conditions are met regarding the character of technological heterogeneity. This qualifies, but certainly does not eliminate, the usefulness of stylized models as a benchmark when assessing the practical relevance of deviations from the neoclassical assumptions.

1.1 ACCOUNTING

Consider an economy with many households endowed with two types of production factors: accumulated and non-accumulated. By definition, *accumulated* factors are inputs whose dynamics are determined by microeconomic savings decisions. At the aggregate level, these decisions affect

both the distribution of accumulated factors across individuals and the dynamics of macroeconomic accumulation. In contrast, *non-accumulated* factors are, by definition, production factors that evolve exogenously (or, for simplicity, remain constant) in the aggregate. We will frequently refer to the accumulated factor as "capital" and to the non-accumulated factor as "labor." However, the simple capital/labor distinction may be misleading. For instance, an individual's human capital is essential for the efficiency of its "labor" but clearly affected by an individual's savings choices. In contrast, incomes from real estate ("land") as well as non-contestable monopolies are often counted as part of capital income but are, according to our definition, part of non-accumulated factors' rewards.

While here we take the evolution of non-accumulated factors as given, it is important to note that, in reality, the economy's supply with these factors is subject to households' supply choices. Here we abstract from the endogeneity of the supply of their non-accumulated factors and from endogenous fertility behavior. We subsume labor/leisure choices under the consumption choice.

A family or household i is endowed with $k(i)$ units of an accumulated factor and $l(i)$ units of a non-accumulated factor. In general, households differ in endowments k and l. Moreover, factor rewards may also differ between households, hence $r = r(i)$ and $w = w(i)$. However, when there are perfect factor markets, all households get the same returns and r and w no longer depend on individual endowment levels but are determined by their aggregate counterparts.

The models reviewed below can be organized around a simple accounting framework. The income flow y accruing to a family also depends on endowments k and l and equals

$$y(i) = w(i) \cdot l(i) + r(i) \cdot k(i).$$

The dynamic budget constraint, at the household level, is given by

$$\Delta k(i) = y(i) - c(i), \text{ or } \Delta k(i) = r(i)k(i) + w(i)l(i) - c(i), \qquad (1.1)$$

where $c(i)$ denotes the consumption flow of a household who owns accumulated factor k and non-accumulated factor l in the current period. The change in the family's stock of the accumulated factor, denoted $\Delta k(i)$, coincides with forgone consumption (income not consumed). Income $y(i)$ is measured *net of depreciation* of the accumulated factor, and $r(i)$ is the *net return* of this factor. Consumption c, income y, and savings Δk are, in general, heterogeneous across individuals. This heterogeneity may be due

to two sources: households own different baskets of factors $(k(i), l(i))$, and they may earn different rewards $r(i)$ and/or $w(i)$.

There are two important assumptions implicit in the above formulation. The first is that there is only one consumption good, and the second is that consumption is convertible one to one into the accumulated factor. We will stick to these assumptions throughout most of parts 1 and 2 of this book. In part 3 we will relax the first assumption: we will study conditions under which differentiating output by different consumption purposes becomes relevant for distribution and growth. In appendix 4.6 we will address the latter assumption. There a model with two sectors is presented where the accumulated factors and consumption goods are produced with different technologies.

Any of the variables on the right-hand sides of the expressions in (1.1) may be given a time index, and may be random in models with uncertainty. In (1.1), $\Delta k(i) \equiv k_{t+1}(i) - k_t(i)$ is the increment of the individual family's wealth over a discrete time period. In continuous time, the same accounting relationship would read

$$\dot{k} = y - c = rk + wl - c, \tag{1.2}$$

where $\dot{k}(t) \equiv dk(t)/dt = \lim_{\Delta t \to 0} \left[\left(k(t + \Delta t) - k(t) \right) / \Delta t \right]$ is the rate of change per unit time of the family's wealth.

The advantage of a continuous-time formulation is that it frequently yields simple analytic solutions, and it is not necessary to specify whether stocks are measured at the beginning or the end of the period. The advantage of discrete time models is that empirical aspects and the role of uncertainty are discussed more easily in a discrete-time framework. We will use the continuous-time formulation in some chapters, the discrete-time formulation in others.

Aggregating across individuals leaves us with the macroeconomic counterparts of income, consumption, and the capital stock. We allow the distribution to be of discrete or continuous nature. In the former case, $p(i)$ denotes the population share of group i, with n different groups in the population, we have $\sum_{i=1}^{n} p(i) = 1$. If distribution is continuous, $p(i)$ denotes the density, and with a population distributed over the interval $[0, 1]$ we have $\int_0^1 p(i)di = 1$. For the sake of compact notation we use the Stjelties integral, which encompasses both the discrete and the continuous case. The measure $P(\cdot)$, where $\int_{\mathcal{N}} dP(i) = 1$, assigns weights to subsets of $\mathcal{N}$, the set of individuals in the aggregate economy of interest. To gain more intuition with the weight function $P(\cdot)$ consider the special case where $\mathcal{N}$ has n elements (of equal population size). Then, the weight function $P(i) = 1/n$ defines Y as the arithmetic mean of individual income levels $y(i)$.

With continuous distribution, the relative size or weight $P(A)$ of a set $A \subset \mathcal{N}$ of individuals is arbitrarily small, and conveniently lets the idiosyncratic uncertainty introduced in chapter 8 average to zero in the aggregate.

We use the convention to write uppercase letters for the aggregate counterpart of the corresponding lowercase letter. Hence aggregate income is denoted by Y and equals

$$Y \equiv \int_{\mathcal{N}} y(i) dP(i), \tag{1.3}$$

where $\mathcal{N}$ denotes the set of families. For the most part, we take $\mathcal{N}$ as fixed. However, when we want to study issues like population growth, finite lives, or immigration, we will allow $\mathcal{N}$ to be variable over time.

Recall that heterogeneity of the non-accumulated income flow wl may be accounted for by differences in w and/or l across individuals. We take l as exogenously given. Hence we sum up and get

$$L \equiv \int_{\mathcal{N}} l(i) dP(i), \tag{1.4}$$

where L denotes the amount of non-accumulated factors available to the aggregate economy.

Recall from (1.1) that we assumed the relative price of c and Δk to be unitary. This allows us to aggregate families' endowments with the accumulated factor. The aggregate stock of the accumulated factor K is measured in terms of forgone consumption

$$K \equiv \int_{\mathcal{N}} k(i) dP(i). \tag{1.5}$$

The definitions in (1.3), (1.4), and (1.5) readily yield a standard aggregate counterpart of the individual accumulation equation (1.1):

$$\begin{aligned} \Delta K &= \int_{\mathcal{N}} \Delta k(i) dP(i) = \int_{\mathcal{N}} (y(i) - c(i)) \, dP(i) \\ &= Y - C = RK + WL - C. \end{aligned} \tag{1.6}$$

Corresponding to its individual counterpart we define $Y = RK + WL$, where R and W denote the aggregate rate of return on the accumulated and non-accumulated factor, respectively. The definition directly implies that R and W are weighted (by factor ownership) averages of their heterogeneous microeconomic counterparts,

$$R = \int_{\mathcal{N}} r(i) \frac{k(i)}{K} dP(i), \qquad W = \int_{\mathcal{N}} w(i) \frac{l(i)}{L} dP(i). \tag{1.7}$$

Interestingly, the economic interpretation of these aggregate factor prices is not straightforward in a world where inequality plays a role. In the models discussed in part 1, all units of each factor are rewarded at the same rate. In this case $r(i) = R$ and $w(i) = W$, which denotes an economy-wide interest rate and wage rate (or land rent), respectively. In the more complex models of part 2, however, unit factor incomes may be heterogeneous across individuals. This introduces interesting channels of interaction between distribution and macroeconomic dynamics. At the same time, such heterogeneity also makes it difficult to give an economic interpretation to aggregate factor supplies and remuneration rates.

Finally, note that the individual-level budget constraint (1.1) features net income flows, and so does (1.6). Hence, the aggregate Y flow is obtained subtracting capital depreciation, say δK, from every period's gross output flow, say $\tilde{Y}$, and (1.6) may equivalently be written

$$\Delta K = \tilde{Y} - \delta K - C.$$

In order to economize on notation and obtain cleaner typographical expressions, from now on we abstain from making explicit the indexing of (lowercase) individual-level variables. A convention we adopt throughout the book is the use of *lowercase letters* to denote variables relating to *individuals* and *capital letters* for variables relating to the *aggregate economy*.

Before proceeding it is important to note that we use the term "inequality" as a relative concept. More inequality can therefore be characterized by a shift in the Lorenz curve, which clearly is measured in relative terms. For example, the Lorenz curve for income depicts the relative share of total income of the poorest x percent of the population where the population percentages are on the horizontal axis. Obviously, we could also be interested in absolute differences in income. However, most of our discussions will not depend on details of such definitions. The interested reader is referred to Cowell (2000).

1.2 The Neoclassical Theory of Distribution

Let production take place in firms that rent factors of production from households, and use these factors in (possibly heterogeneous) production functions. (Now lowercase letters refer to a particular firm rather than a household.) A firm produces $y = f(k, l)$ units of output, takes as given the (possibly heterogeneous) rental prices r and w of the factors it employs,

and maximizes profits as in

$$\max_{k,l} \left(f(k,l) - rk - wl \right). \tag{1.8}$$

If technology is convex, i.e., $f(\cdot,\cdot)$ is a concave function, the first-order conditions

$$\frac{\partial f(k,l)}{\partial k} = r, \quad \frac{\partial f(k,l)}{\partial l} = w \tag{1.9}$$

are necessary and sufficient for solution of the problem (1.8). Note that $f(\cdot,\cdot)$, r, and w may, in general, be different by firms.

Now assume that there are perfect factor markets. If factors can be costlessly relocated between production units, then, in equilibrium, the same factor must be rewarded at the same rate, irrespective of the particular firm where it is employed. Otherwise, arbitrage opportunities would exist, and reallocation meant to exploit them would eliminate all marginal productivity differentials.

It is easy but instructive to show that an equilibrium where, for all firms, $w = W$ and $r = R$ maximizes the aggregate production flow obtained from a given stock of the two factors. Formally the equilibrium allocation solves the problem

$$F(K,L) \equiv \max_{\{l(j),k(j)\}} \int_{\mathcal{F}} f^{(j)}(k(j), l(j)) \, dQ(j) \tag{1.10}$$

$$\text{s.t.} \quad \int_{\mathcal{F}} l(j) \, dQ(j) \leq L, \quad \int_{\mathcal{F}} k(j) \, dQ(j) \leq K,$$

where j indexes firms, $\mathcal{F}$ denotes the set of all firms, j is a firm index, and $Q(j)$ is the distribution function of firms. The first-order conditions of (1.10) are necessary and sufficient due to the same concavity assumptions that make (1.9) optimal at the firm level.

$$\frac{\partial f^{(j)}(k(j), l(j))}{\partial l(j)} = \lambda_L \quad \text{if } l(j) > 0 \tag{1.11}$$

$$\frac{\partial f^{(j)}(k(j), l(j))}{\partial k(j)} = \lambda_K \quad \text{if } k(j) > 0.$$

The optimality conditions (1.11) say that marginal products across firms have to be equalized whenever this factor is employed at firm j in positive amounts. This condition is exactly met by the firms' optimality conditions (1.9), because $r(j) = R$ and $w(j) = W$ for all j holds in equilibrium. Then, the factors' unit incomes coincide with the shadow prices λ_L and λ_K of

the two aggregate constraints in (1.10),

$$w = W = \lambda_L = \frac{\partial F(K, L)}{\partial L}, \qquad r = R = \lambda_K = \frac{\partial F(K, L)}{\partial K}, \qquad (1.12)$$

and (1.10) defines an aggregate production function $F(\cdot, \cdot)$ as the maximum aggregate production obtainable from any given set of factors.

Hence we can state a central result: if markets are *perfect*, all factors are *mobile*, and firms choose inputs to *maximize profits*, aggregate production is at its efficient frontier. Under our assumptions of a single output good, efficiency means that aggregate output reaches its maximum level. Under neoclassical conditions, it is possible to abstract from distributional issues and technological heterogeneity. The allocation of resources and the distribution of income among factors of production can be viewed as if they were generated by decisions of representative consumers and producers. The *distribution across families* of production factors has no effect on productive efficiency, since factors can be reallocated across firms so as to equalize marginal products. Clearly, the initial distribution of endowments with factors of production does matter for the size distribution of income across families. The *distribution of technological knowledge* across firms plays no role for the existence of a well-defined aggregate production function for a similar reason. The mobility of production factors equalizes their marginal product across production units, hence the effect on aggregate output of increasing the aggregate stock of a factor by one unit is well defined. Aggregate output can thus be represented as a function of the aggregate stock of production factors. Clearly, the functional form of the aggregate production function $F(\cdot, \cdot)$ does reflect the heterogeneity of technologies, and the size distribution of firms will mirror the technological differences: firms with a better production technology will produce at a larger scale. In cases where no misunderstandings are possible, we will not explicitly index firms in what follows.

1.2.1 *Returns to Scale*

When all individual production functions have constant returns to scale, so does the aggregate production function. In that case, aggregate factor-income flows coincide with total net output by Euler's theorem:

$$F(K, L) = \frac{\partial F(K, L)}{\partial L} L + \frac{\partial F(K, L)}{\partial K} K = WL + RK \qquad (1.13)$$

The irrelevance of distribution and technological heterogeneity for the macroeconomic equilibrium does not hinge upon the assumption of constant returns: decreasing returns to scale at the firm level can be accom-

modated by including any fixed factors in the list of (potentially) variable factors. The rents accruing to these fixed factors are part of aggregate income. Obviously, the presence of decreasing returns in production with respect to k and l leaves the above central result unchanged. Marginal products of k and l are still equalized across production units. Similarly, factor-ownership inequality does not affect aggregate output, and a well-defined aggregate production function $F(K, L)$ exists despite technological heterogeneity across firms.[1]

Equation (1.13) states how income is distributed to the factors of production. According to (1.12), factors are paid their marginal product. In this neoclassical setting, each factor is paid according to its contribution to output. Equation (1.13) shows further that perfect factor markets and a competitive reward of factors can only exist if returns to scale are non-increasing. Were the technology to exhibit increasing returns to scale (*non-convexities*), the factor rewards $(\partial F(K, L)/\partial L) L + (\partial F(K, L)/\partial K) K$ would more than exhaust the total value of production. Consequently, at least one factor has to be paid less than its marginal product, implying that the respective market is not competitive. In other words, the neoclassical analysis has to rule out increasing returns.[2]

1.2.2 *Mobility of Production Factors*

The above discussion suggests that the mobility of production factors is crucial. It is therefore interesting to ask what happens if one factor is immobile. Consider, for instance, the case where the non-accumulated factor is firm-specific: a firm's production may involve use of a peculiar natural resource, or of its owner's unique entrepreneurial skills, and may therefore increase less than proportionately to employment of factors that are potentially or actually mobile across firms in the economy considered. It turns out that, when technologies are homogeneous across production units, factor-price equalization is still ensured. Since the marginal products of the mobile factor must be equal, the homogeneity of technologies implies that all firms produce with the same factor intensity.

---

[1]In general, an aggregate of the (immobile) fixed factor does not exist. While aggregate production function depends only on the stock of K and L, it will depend on the distribution of the fixed factors across production units just like the functional form $F(\cdot, \cdot)$ depends on the distribution of technologies.

[2]Note that, by the accounting conventions of section 1.1, both firm-level and aggregate production functions are defined net of capital depreciation. This has no implications for this argument: if the gross production function is concave and has constant returns to scale, so does net production as long as, as is commonly assumed, a fixed portion of capital in use depreciates within each period.

We also note that, if the non-accumulated factor is immobile and technologies are homogeneous, the distribution of production is determined by the distribution of l. The following exercise proves this claim formally.[3]

EXERCISE 1 Assume each firm is endowed with a fixed amount l of labor. Instead, k is mobile. All firms use the same CRS technology: $y = F(k, l)$. (Note this implies that the production function for a firm is the same as the aggregate production function.) Show that the reward of the immobile factor w is equal across firms and that the firm output is proportional to the endowment l of the immobile factor.

1.2.3 Heterogeneous Technologies and Immobile Factors

In the general case, with heterogeneous technologies and immobile factors, serious aggregation problems arise. As shown by Fisher (1969) and Felipe and Fisher (2001), aggregation is only possible under very restrictive assumptions on technological heterogeneity. Translated into our context, Fisher's aggregation result states that an aggregate production function exists *if and only if* technological heterogeneity is restricted to augmenting differences in the immobile factor. This means that if technological heterogeneity takes the form

$$f(k, l) = F(k, b\tilde{l})$$

there exists a well-defined measure for the aggregate stock of the immobile non-accumulated factor and aggregate output can be represented as $F(K, L)$. Of course, the appropriate aggregate measure of the immobile factor is then $L = \int b(j)\tilde{l}(j)dQ(j)$, and coincides with definition (1.4) if the (exogenously given) immobile factor is sensibly measured in efficiency units.

The following exercises show that mobility of some factors may suffice to ensure factor-price equalization if all firms have the same technology, and that some technologies remain unused if different firms have access to different technologies and factors are mobile.

---

[3]Obviously, when both factors are immobile no interaction takes place. There exists a collection of family firms that produce and consume in isolation, which differ not only in their ownership of productive factors, but also in the incomes earned by each unit of their factors. There is no macroeconomic equilibrium in such a situation: each family firm constitutes its own "macroeconomy."

EXERCISE 2 Discuss factor rewards and equilibrium allocation across two firms with production function

$$f^{[1]}(k,l) = A_1 k^\alpha l^\beta + A_2 k$$
$$f^{[2]}(k,l) = B_1 k^\gamma l^\delta + B_2 k$$

For what values of the parameters are these functions strictly concave? Suppose there is a total amount K of factor k, mobile across the two firms: is its employment positive at both firms if $A_2 = B_2$ and if $A_1 = B_1 = 0$? If l is immobile, are there parameter configurations such that its marginal productivity is equalized by mobility of k only?

EXERCISE 3 (a) For what parameter values are returns to scale constant in the functional forms proposed in exercise 2? (b) Discuss the form of the relevant aggregate production functions when $A_2 = B_2$. (Hint: Determine first whether both firms produce in equilibrium or not.)

The macro models of distribution reviewed in later chapters give up the neoclassical framework and study systematically deviations from these assumptions. The literature reviewed in chapters 4 and 6 studies models with increasing returns and treats distribution as exogenously given, and discusses the consequences of distribution for growth. Models in part 2 in which capital market imperfections play a central role typically feature technological heterogeneity and immobile factors of production ("human capital") in which aggregation conditions are clearly not satisfied. Models in part 3 study the consequences of distribution for macroeconomic outcomes when there are imperfections in product markets and the distribution of income among factors of production is affected by the heterogeneity in the families' initial endowments.