1 Introduction

1.1 The definition of economics

In *An Essay on the Nature and Significance of Economic Science* (1984) Lionel Robbins (1898–1984) defines economics as "the science which studies human behavior as a relationship between given ends and scarce means which have alternative uses." This famous "all-encompassing" definition of economics is still used to define the subject today, according to *The Concise Encyclopedia of Economics* (http://www.econlib.org/library/CEE. html). The underlying idea is that absent scarcity, all needs could be satisfied, no choices would have to be made, and, therefore, no economic problem would be present. But which resources are scarce? Is air scarce? If not, maybe clean air is scarce? If we stick to the traditional definition of economics, these questions must be answered prior to any economic analysis: in some mysterious way, all the scarce resources are known. In my opinion, however, the enumeration of scarce resources should be included in the definition of economics. I therefore modify Robbins' definition by omitting the adjective "scarce." In short, I define economics as the study of the allocation of resources among alternative ends – or, more precisely, the study of the allocation of resources to production units for commodities and the distribution of the latter to the population. Some resources may be scarce, others may not be. Scarcity will be signaled by a price. If resources are not scarce, they will have a zero price.

It is not necessary to be very specific at this stage as regards the concepts of "production units," "commodities," "distribution," and "households." The essence of economics is merely that something is maximized. Production units, or firms, maximize profits and households maximize their levels of income or well-being. The objectives can be fulfilled only to limited extents because of resource constraints. Maybe air is not scarce, but there is certainly only a limited stock of it, however large. The limited availability of some resources will act as a bottleneck in the furthering of the objectives. The economic problem can thus be summarized as the maximization of some objective subject to constraints. It is crucial to understand the principles of *constrained maximization*, and to relate them to the basic economic concept of a price, but first we must quickly review some elementary principles of mathematics.

1.2 Mathematical preliminaries

The two main streams of elementary mathematics are calculus and matrix algebra. *Calculus* is about functions, particularly of real numbers, and the manipulations that can be done with them, such as taking derivatives or integrals. *Matrix algebra* extends operations such as addition and multiplication to higher dimensions. It is handy for the extension of calculus to functions of several variables.

By definition, a *function*, f, maps *every* element, x, of one set (the domain) to *precisely one* element of a second set (the range), $f(x)$. The standard case is where both the domain and the range are the set of real numbers. Examples are given by (1)–(3) and counterexamples by (4) and (5):

(1)	$f(x) = x$	the identity function
(2)	$f(x) = cx + d$	a linear function
(3)	$f(x) = x^n$	a power function
(4)	$f(x) = \sqrt{x}$	the square root
(5)	$f(x) = \pm\sqrt{x}$	the solution to $y^2 = x$

Counterexample (4) is not a function from the real numbers to the real numbers, because it does not take every element to another one. However, by restricting the domain to the non-negative numbers, it becomes a function. Counterexample (5) is not a function from the real numbers to the real numbers, because it does not take elements to precisely one other. By restricting the value to either the non-negative or the non-positive one, counterexample (5) becomes a function. The *inverse* of function f is the function f^{-1} defined by $f^{-1}(y) = x$ with $f(x) = y$. The inverse of function 1 is $f^{-1}(y) = y$. The inverse of function 2 is $f^{-1}(y) = (y - d)/c$. The inverse of function 3 is $f^{-1}(y) = x^{1/n}$ for n odd. For n even, say 2, case (4) would be the candidate solution, but it is not a function.

Let x be input and $f(x)$ output. Then *average* product is $f(x)/x$. The *marginal* product is the rate at which output increases:

$$\frac{f(x + \Delta x) - f(x)}{\Delta x}, \ \Delta x \to 0 \tag{1.1}$$

Expression (1.1) is called the *derivative* of f in x and is denoted $f'(x)$. The symbol $\to$ means "tends to." The derivative of x^n is nx^{n-1}:

$$(x^n)' = nx^{n-1} \tag{1.2}$$

Let me illustrate the rule for $n = 2$: By definition (1.1) the derivative of x^2 is

$$\frac{(x + \Delta x)^2 - x^2}{\Delta x} = 2x + \Delta x$$

with $\Delta x \to 0$, hence $2x$. If Δx does not tend to zero but is small, the quotient

$$\frac{(x + \Delta x)^2 - x^2}{\Delta x}$$

is approximately equal to the derivative, as we overlook the residual term, Δx. In general:

$$\frac{f(x+\Delta x)-f(x)}{\Delta x} \approx f'(x), \ \Delta x \quad \text{small} \tag{1.3}$$

The approximate equality (1.3) is called the first-order approximation. Other handy rules of differentiation are the *sum*, *product* and *chain rules*:

$$(f+g)' = f' + g' \tag{1.4}$$

$$(fg)' = f'g + fg' \tag{1.5}$$

$$f[g(x)] \text{ has the derivative } f'[g(x)]g'(x) \tag{1.6}$$

The proof of the sum rule (1.4) is trivial. The proof of the product rule needs a little work. By definition,

$$\begin{aligned}(fg)'(x) &= \frac{f(x+\Delta x)g(x+\Delta x)-f(x)g(x)}{\Delta x} \\ &= g(x+\Delta x)\frac{f(x+\Delta x)-f(x)}{\Delta x} + f(x)\frac{g(x+\Delta x)-g(x)}{\Delta x}\end{aligned} \tag{1.7}$$

with $\Delta x \to 0$, so that (1.7) proves the rule (1.5). Finally, the proof of the chain rule (1.6) is straightforward. The derivative of $f[g(x)]$ is

$$\frac{f[g(x+\Delta x)]-f[g(x)]}{\Delta x} = \frac{f[g(x+\Delta x)]-f[g(x)]}{g(x+\Delta x)-g(x)} \cdot \frac{g(x+\Delta x)-g(x)}{\Delta x} \tag{1.8}$$

with $\Delta x \to 0$. Substituting $y = g(x)$ and $\Delta y = g(x+\Delta x) - g(x)$, the first factor on the right-hand side of (1.8) reads

$$\frac{f(y+\Delta y)-f(y)}{\Delta y}$$

Since Δy tends to zero as $\Delta x \to 0$, the proof of the chain rule is complete.

Taking the derivative of a product function, one obtains the marginal product function. Now the reverse operation from differentiation is taking the *integral* – or, briefly, integration. Hence by integrating the marginal products one retrieves the underlying production function. The symbol for an integral is $\int$. For example, by (1.2), $\int nx^{n-1}dx = x^n$. Since the derivative of a constant is zero, one may add this to the integral. So, strictly speaking, $\int nx^{n-1}dx = x^n + c$, where c is any constant number.

Integrating the marginal products between a and b, one obtains the total output that comes with an increase of input from a to b: $\int_a^b f'(x)dx = f(b) - f(a)$. For example, $\int_0^1 nx^{n-1}dx = (1^n + c) - (0^n + c) = 1$. Hence $\int_0^1 x^{n-1}dx = 1/n$.

Let us model the production of a single output from *two* inputs. Then x in $f(x)$ is a list of two numbers or a *vector*, with components x_1 and x_2. The marginal product of the first input is the partial derivative of $f(x_1, x_2)$ with respect to x_1. By definition, this is the ordinary derivative of the function of x_1 keeping x_2 fixed. It is denoted f_1'. The row vector of partial derivatives is denoted $f' = (f_1' \quad f_2')$. For example, if the production function is $f(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}$, then the marginal products are given by $(\alpha x_1^{\alpha-1} x_2^{1-\alpha} \quad (1-\alpha) x_1^{\alpha} x_2^{-\alpha})$. If both inputs are rewarded according to their marginal products, the total cost is

$$\alpha x_1^{\alpha-1} x_2^{1-\alpha} x_1 + (1-\alpha) x_1^{\alpha} x_2^{-\alpha} x_2.$$ [1]

It is also possible to model *multiple outputs*. The two inputs may produce two outputs, each with its own production function. The vector of outputs is denoted

$$\begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix}$$

We may now list for each output the row vector of marginal products:

$$\begin{pmatrix} \dfrac{\partial f_1(x_1, x_2)}{\partial x_1} & \dfrac{\partial f_1(x_1, x_2)}{\partial x_2} \\ \dfrac{\partial f_2(x_1, x_2)}{\partial x_1} & \dfrac{\partial f_2(x_1, x_2)}{\partial x_2} \end{pmatrix}$$

This table is a 2×2-dimensional *matrix*. The first index indicates the row (output, in this case), the second index the column (input, in this case). The (i, j)th element of the matrix represents the marginal i-product of input j.

The numbers of inputs and outputs need not match. In fact, we have dealt with the case of one output and two inputs, where we had a row vector of marginal products $f' = (f_1' \quad f_2')$. This is a 1×2 matrix. In general an $m \times k$-dimensional matrix B has m rows and k columns. The element in row i and column j is denoted b_{ij}. $b_{i\bullet}$ denotes row i and $b_{\bullet j}$ denotes column j. Notice that the dimension of any row of matrix B is k, which is the number of columns. Similarly, the dimension of any column is m, the number of rows.

1.3 Constrained maximization

An objective function ascribes values to the various magnitudes of all the variables of an economy. If the variables are $x_1, \ldots, x_n$ (representing the activity levels of the production units, for example), then the outcome (national income, for example), will be some real number $f(x_1, \ldots, x_n)$, or $f(x)$ for short, where f is the objective function. Formally, an objective function f maps the n-dimensional variable space to the one-dimensional space of the real numbers, that is $f : \mathbb{R}^n \to \mathbb{R}$. It is important to distinguish the objective function, f, and the values it may take, $f(x)$. The latter merely measure the performance of the economy for given magnitudes of all the underlying variables, while the former denotes the relationship between performance and the underlying variables. In other words, function f summarizes the structure of the economy. There may be many constraints. With each level of the variables of an economy $x = (x_1, \ldots, x_n)$, we may associate labor requirements – say, $g_1(x)$ – and other resource requirements – say, $g_i(x)$ – where resource i is any input

[1] This expression happens to be equal to $f(x_1, x_2)$, a finding that reflects the constant returns to scale property of f.

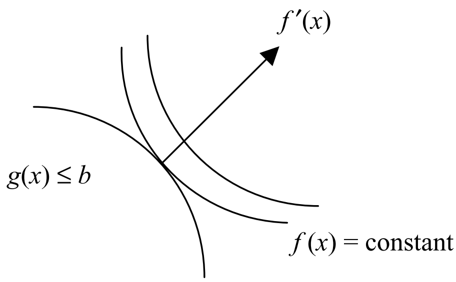

Figure 1.1 The feasible region of constraint function g and two isoquants and the derivative of objective function f

that must be present before production takes place, such as mineral resources, equipment, etc. Let the number of resources be m. Then the requirements are $g_1(x), \ldots, g_m(x)$ and the resource constraints can be written by $g_1(x) \leq b_1, \ldots, g_m(x) \leq b_m$, where the right-hand sides are the available quantities of the resources. The inequalities may be summarized by:

$$g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix} \leq \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = b \tag{1.9}$$

In constraint (1.9) g is the constraint function and b is the bound. Function g associates with every n-dimensional list of variables, x, m requirements, that is a point in m-dimensional space. Formally, we write $g : \mathbb{R}^n \to \mathbb{R}^m$. Constrained maximization is the problem:

$$\max_x f(x) : g(x) \leq b \tag{1.10}$$

The colon in program (1.10) stands for the phrase "*subject to*." The program can be depicted graphically in the variable space, particularly when there are two variables ($n = 2$) and only one constraint ($m = 1$). The set of points that fulfills the constraint, (1.9), is the *feasible* region. The objective function can be represented by so-called *isoquants*, which connect points x of equal value, $f(x)$. Perpendicular to these isoquants are the vectors of steepest ascent which are given by the partial derivatives:

$$f'(x) = \left(\frac{\partial f}{\partial x_1}(x) \cdots \frac{\partial f}{\partial x_n}(x) \right) \tag{1.11}$$

For example, if the isoquant is given by $3x_1 + x_2 = 6$, which is a steep line with horizontal intercept $x_1 = 2$ and vertical intercept $x_2 = 6$, then the vector perpendicular to the isoquant is (3 1). For a non-linear example see figure 1.1.

The objective function f takes a maximum value on the feasible region where the isoquant is tangent to the boundary. Since the boundary is an isoquant of the constraint function, g, an equivalent condition is that the vectors of steepest ascent point in the same direction:

$$f'(x) = \lambda g'(x), \lambda \geq 0 \tag{1.12}$$

In (1.12) proportionality constant λ cannot be negative, for then a movement in the direction $f'(x)$ would go into the feasible region and constitute an improvement, contradicting the assumed maximization. Note also that the above condition covers the case where the constraint is *not* binding. Then maximization merely requires that the objective function is flat: $f'(x) = 0$. This is covered by a *zero* λ in (1.12).

First-order condition (1.12) of constrained maximization holds in the general case where $g : \mathbb{R}^n \to \mathbb{R}^m$. In short, the derivatives of the objective function are proportional to those of the constraint function and the proportions are non-negative. The following matrix defines the derivative of g:[2]

$$g' = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n} \end{pmatrix} \tag{1.13}$$

The proportionality constants, one for each constraint, are listed in row vector $\lambda = (\lambda_1 \ldots \lambda_m)$ and the product of λ and matrix g' is defined in the usual way by a row vector (of the same dimension as f'):[3]

$$\begin{aligned} \lambda g' &= (\lambda_1 \cdots \lambda_m) \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n} \end{pmatrix} \\ &= \left(\lambda_1 \frac{\partial g_1}{\partial x_1} + \cdots + \lambda_m \frac{\partial g_m}{\partial x_1} \ \cdots \ \lambda_1 \frac{\partial g_1}{\partial x_n} + \cdots + \lambda_m \frac{\partial g_m}{\partial x_n} \right) \end{aligned} \tag{1.14}$$

Mathematicians call the proportionality constants λ in (1.12) *Lagrange multipliers*. Again, when a constraint is *not* binding, the Lagrange multiplier is *zero*:

$$g_i(x) < b_i \Rightarrow \lambda_i = 0 \tag{1.15}$$

Because of inequality (1.9), implication (1.15) may be written as:

$$\lambda_i [b_i - g_i(x)] = 0 \tag{1.16}$$

Using the fact that a sum of non-negative terms is zero if and only if every term is zero, the system of all equations (1.16) is equivalent to the single equation:

$$\sum_{i=1}^{m} \lambda_i [b_i - g_i(x)] = 0 \tag{1.17}$$

Invoking the notation of the product of row vector λ and a matrix, (1.14), (1.17) simply

[2] This notation is consistent with that of the partial derivatives of a real-valued function (such as f), as the case $m = 1$ shows.
[3] In (1.14), the first component is the product of λ and the first column of matrix g', etc. A precise treatment of matrix multipication is postpond to chapter 2.

reads:

$$\lambda[b - g(x)] = 0 \tag{1.18}$$

Equation (1.18) is a brief reflection of the condition that a constraint is binding or has a zero Lagrange multiplier.

The first-order conditions (1.12) and the so-called *complementary slackness* conditions (1.18), are a concise mathematical statement of the solution to the constrained maximization problem, (1.10). The proportionality constants between the objective function derivatives and the constraint function derivatives (the Lagrange multipliers) have an economic interpretation, which we shall establish below. As a matter of fact, we shall show that the Lagrange multipliers measure the marginal productivities of the constraining entities. By definition, a marginal productivity is the amount by which the objective value goes up when an additional unit is available. So consider the situation in which one unit is added to the bound of the ith constraint. The new bound is $b + e_i$, where e_i is the ith unit vector:

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \\ \\ \\ \leftarrow \text{ place } i \\ \\ \\ \\ \end{matrix} \tag{1.19}$$

In (1.19), the ith entry is one and all others are zero. Let x^* be the new optimum, reserving unstarred x for the old optimum (bounded by b). Making first-order approximations (1.3) to the increase of both the objective and the constraint function values and substituting (1.12) and the new bound we get:

$$\begin{aligned} f(x^*) - f(x) &\approx f'(x)(x^* - x) = \lambda g'(x)(x^* - x) \\ &\approx \lambda[g(x^*) - g(x)] = \lambda[g(x^*) - b] \leq \lambda e_i = \lambda_i \end{aligned} \tag{1.20}$$

Inequality (1.20) indicates that the marginal productivity of the ith constraining entity does not exceed λ. If $\lambda = 0$, this "increase" in the value is attained trivially by $x^* = x$. If $\lambda > 0$, the increase in the value is actually attained by the solution x^* to the equation defined by (1.20) with a *binding* inequality. In either case, the value of the objective function goes up by an amount of λ when one unit is added to the ith bound. The derivation will be presented rigorously in the context of linear objective and constraint functions in chapter 4.

1.4 Linear analysis

This section is a quick introduction to material that will be explained in detail in subsequent chapters. Readers who do not know matrices should proceed directly to chapter 2.

If an economy features constant returns to scale and the objective is to maximize the value of the net product, then the constraints and the objective function are linear. Problem (1.10) turns out as:

$$\max_x ax : Cx \leq b \tag{1.21}$$

The constrained maximization problem (1.21) is called a *linear program*. The first-order conditions (1.12) turn out as:

$$a = \lambda C, \lambda \geq 0 \tag{1.22}$$

Finally, the complementary slackness conditions (1.18) turn out as:

$$\lambda(b - Cx) = 0 \tag{1.23}$$

Multiplying (1.22) by solution x and substituting (1.23), we derive the important result:

$$ax = \lambda b \tag{1.24}$$

Equation (1.24) imputes the optimal value to the bounds. Each binding unit gets a value of λ_i. The result confirms that the marginal productivities of the bounds (given by vector b) are the components of row vector λ. There is a neat way to characterize these Lagrange multipliers. Consider *any* row vector μ fulfilling condition (1.22):

$$a = \mu C, \mu \geq 0 \tag{1.25}$$

Then we have, using the inequality in (1.21), the equality in (1.25), and (1.24):

$$\mu b \geq \mu Cx = ax = \lambda b \tag{1.26}$$

According to (1.24) the inequality is binding for λ. In other words, λ minimizes the left-hand side of (1.26). In other words, the Lagrange multipliers solve:

$$\min_{\lambda \geq 0} \lambda b : \lambda C = a \tag{1.27}$$

Minimization problem (1.27) is the dual program associated with the original maximization problem or primal program (1.21). Notice that the values of the primal and dual programs are equal according to (1.24). If a so-called *shadow price* of λ_i is assigned to the entity of constraint i, then the value of the ith bound is $\lambda_i b_i$ and the total value of bound b exhausts the value of the objective function. Since the shadow prices are equal to the marginal productivities, a competitive mechanism can bring them about. This approach is borne out in the following example.

In traditional input-output analysis, variable x lists the gross outputs of the sectors of an economy. Assuming constant returns to scale and fixed input proportions, sector 1 requires $a_{11}x_1, \ldots, a_{n1}x_1$ units of the various sectors as inputs in its production of x_1 units of output. Demand for the product of sector 1 amounts to $a_{11}x_1$ by sector 1 itself, $a_{12}x_2$ by sector 2, $\ldots$, $a_{1n}x_n$ by sector n, and y_1 final demand by the non-producing sectors of the economy, such as the households. Organize these demand coefficients in a row vector:

$$a_{1\bullet} = (a_{11} \cdots a_{1n}) \tag{1.28}$$

The condition that total demand for the product of sector 1 is bounded by supply can be written succinctly as follows:

$$a_{1\bullet}x + y_1 \leq x_1 \tag{1.29}$$

Organize the different row vectors in a matrix A:

$$A = \begin{pmatrix} a_{1\bullet} \\ \vdots \\ a_{n\bullet} \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \tag{1.30}$$

Then constraint (1.29) is the first component of the following inequality:

$$Ax + y \leq x \tag{1.31}$$

Let the economy maximize the value of the net output, $\underline{p}y$, on world markets. Here $\underline{p}$ is a given row vector of *world prices*. If the net output y does not agree with household demand, it is traded for other commodities. The maximization of the value of net output yields the greatest purchasing power in world markets, which is clearly in the interest of the domestic households. Since the value of net output is constrained by commodity balance (1.31), the factor balances, and a non-negativity constraint, we face the program:

$$\max_{x,y} \underline{p}y : Ax + y \leq x, kx \leq M, lx \leq N, x \geq 0 \tag{1.32}$$

In program (1.32) row vector k lists the amount of capital required per unit of output in each sector, M is the available stock of capital, and l and N are the corresponding labor statistics. Introduce matrix notation for the objective function and constraint coefficients, respectively:

$$a = (0 \quad \underline{p}), C = \begin{pmatrix} A - I & I \\ k & 0 \\ l & 0 \\ -I & 0 \end{pmatrix} \tag{1.33}$$

Then program (1.32) reads

$$\max a \begin{pmatrix} x \\ y \end{pmatrix} : C \begin{pmatrix} x \\ y \end{pmatrix} \leq \begin{pmatrix} 0 \\ M \\ N \\ 0 \end{pmatrix} \tag{1.34}$$

In program (1.34) multiplication of the first row of coefficients matrix C of (1.33) with the stacked vector $\binom{x}{y}$ reproduces the first inequality, (1.31). Multiplication of the other rows of matrix C with the vector of variables reproduces the further inequalities in program (1.32).

Denote the shadow prices associated with the material constraints, the capital and labor constraints, and the non-negativity conditions by:

$$\lambda = (p \quad r \quad w \quad \sigma) \tag{1.35}$$

The notation (1.35) suggests commodity price, rental rate of capital, wage rate, and slack, as will be explained shortly. The shadow prices are determined by the first-order condition (1.22) – or, substituting specifications (1.33) and (1.35),

$$(p \quad r \quad w \quad \sigma)\begin{pmatrix} A-I & I \\ k & 0 \\ l & 0 \\ -I & 0 \end{pmatrix} = (0 \quad \underline{p}) \tag{1.36}$$

The first component of (1.36) is the product of the row vector and the first column of the matrix – or, after a slight rearrangement of terms:

$$p = pA + rk + wl - \sigma \tag{1.37}$$

The second component of (1.36) reads:

$$p = \underline{p} \tag{1.38}$$

By (1.38) the shadow prices of the materials are simply *equal* to the world prices. And by (1.37) the prices are equal to the sum of the material costs of the inputs, the capital costs, the labor costs, and the slack. If the slack is positive, costs exceed price. However, since the slack is a Lagrange multiplier, the underlying constraint is binding by the complementary slackness conditions (1.23). But since this is a non-negativity constraint, it means that the output of such a sector is zero. Thus, unprofitable sectors are inactive. Conversely, if sectors are active, the non-negativity constraint is not binding, the associated slack variable is zero, and, therefore, price equals cost by (1.37). Thus, the shadow prices, particularly of capital and labor, make the active sectors break even while rendering the inactive sectors unprofitable. Profit maximizing entrepreneurs would target the right sectors. Moreover, since the shadow prices are minimal, yielding negative or zero profits, a process of free entry can bring them about. The competitive market mechanism is a device for the *optimal allocation of resources*.

The value of final demand, $\underline{p}y$, accrues to the resources in proportion to their marginal productivities, rM for capital and wN for labor. Thus, if resources are rewarded according their shadow prices, the value of the net output of the economy is exhausted. This equality of costs and revenues reflects the constant returns to scale. A precise derivation is by the application of the equality of the primal and dual solution values, (1.24):

$$(0 \quad \underline{p})\begin{pmatrix} x \\ y \end{pmatrix} = (p \quad r \quad w \quad \sigma)\begin{pmatrix} 0 \\ M \\ N \\ 0 \end{pmatrix} \tag{1.39}$$

Equation (1.39) is the well-known macroeconomic identity of the national product and the national income:

$$\underline{p}y = rM + wN \tag{1.40}$$

In (1.40), national income comprises no profit under constant returns to scale and competition. If resources are paid according to their marginal productivities, income matches the