Final version, March 2001

NON-CONSTANT RETURNS, PARETO OPTIMALITY AND COMPETITIVE EQUILIBRIUM

David Laibman

Program in Economics, The Graduate School, City University of New York, 365 Fifth Avenue, New York NY 10016, USA

Competitive equilibrium is not Pareto optimal if returns to scale are not constant, except in special and accidental circumstances. This result is demonstrated using a classical production model; it holds quite generally, and independently of all other sources of Pareto inefficiency, such as externalities, imperfect information and quantity constraints. It establishes a general and ubiquitous basis for critique of the "invisible hand" ideology, which still dominates both the textbooks and wider reaches of social thought.

The author would like to thank Vivian Walsh and anonymous referees for their comments on earlier drafts of this paper. The final version is, of course, the author's responsibility alone.

1. The Problem Stated

The object of this paper is to show that, when returns to scale are not constant, competitive equilibrium is not Pareto optimal, except in special and accidental cases.

In today's post—Arrow-Debreu environment of quantity-constrained equilibria, imperfect and asymmetric information, etc., this may not seem to matter much: if an economy is not thought likely to reach equilibrium, the Pareto properties of that equilibrium are of little interest.[1] In the textbooks, however, diminishing returns and concave production possibility curves are central to the ideological thrust of mainstream economics. Pareto optimality also serves in an ideological capacity, protecting competitive equilibrium from the ravages of nihilistic realism: in a hypothetical world of perfect information, costless mobility of resources, absence of externalities, etc., the invisible hand lives on. For this reason it is important to examine the implications of non-constant (especially decreasing) returns for the Pareto

[1] Foundation texts for neoclassical general equilibrium theory are Arrow & Hahn (1971); Debreu (1959). A recent survey that focuses on the post—Arrow-Debreu developments is Mukherji (1990).

optimality of competitive equilibrium when the usual perturbing conditions are absent.

I will do this using a classical model of production and distribution. Deriving a production possibility curve from a linear production model under decreasing returns is itself novel, and challenges the claim that classical—linear production theory has difficulty handling non-constant returns to scale. This procedure also lets us distinguish the "macro" production possibility curve reflecting the effects on production coefficients of large-scale shifting of resources from the "micro" production possibility curve for the individual producer, whose own perceptible activity shifts have negligible impact on the macro coefficients.

I begin with a subcase of the classical model: one in which all active agents are both producers and traders, performing labor with productive resources which they also own. There is no separation here into labor and capital markets, and the producer-traders maximize, and therefore bring about the competitive equalization of, a single variable – the net income per unit of labor expended, or net income ratio, η. I then argue that the result extends readily to the full classical model, with separate labor and capital markets and simultaneous equalization of a wage rate and a profit rate, and by implication to the neoclassical model with fixed endowments and production functions.

2. The Model, in n and 2 Sectors

Consider a multi-sector economy, without technical change.[2] There are no capital stocks (stocks of produced inputs). Non-produced inputs other than labor are assumed to be present; their role in production will be expressed indirectly, however, through their effect on labor and material input flows. The overall scale of production is determined by the aggregate quantity of homogeneous labor, which closes the system. The production technology can then be described in well-known input-output fashion by a row vector of homogeneous per-unit labor inputs, $\mathbf{l}$, and a square matrix of flow per-unit input coefficients, $\mathbf{A}$. Net and gross outputs are, respectively, the column vectors $\mathbf{Y}$ and $\mathbf{X}$; these are related by $\mathbf{Y} = (\mathbf{I} - \mathbf{A})\mathbf{X}$, where $\mathbf{I}$ is the identity matrix. We will also need, for analytical purposes, the vertically integrated unit labor input vector

$$\lambda = \lambda\mathbf{A} + \mathbf{l} = \mathbf{l}(\mathbf{I} - \mathbf{A})^{-1}, \qquad (1)$$

from which, with L representing the total labor force,

$$\lambda\mathbf{Y} = \mathbf{l}(\mathbf{I} - \mathbf{A})^{-1}(\mathbf{I} - \mathbf{A})\mathbf{X} = \mathbf{l}\mathbf{X} = L. \qquad (2)$$

---

[2] A good source for models in this tradition is Kurz & Salvadori (1995); see also Kurz & Salvadori (1998). The seminal foundation text is Sraffa (1960).

$\boldsymbol{\lambda}$**Y** = *L* is the equation of the net transformation surface, the simplex expressing the tradeoff among all of the net outputs of the

$$Y_i = \frac{L}{\lambda_i} - \sum_{j=1}^{n} \frac{\lambda_j}{\lambda_i} Y_j \qquad j \neq i \qquad (2')$$

economy, which can also be written as

From (2') the maximum net output of any commodity, corresponding to

$$Y_i^{\max} = \frac{L}{\lambda_i}.$$

zero net outputs for all other commodities, is seen to be

It should be noticed that (2') reveals the <u>marginal rate of transformation</u> (mrt) between any two commodities to be

$$\frac{\partial Y_i}{\partial Y_j} = -\frac{\lambda_j}{\lambda_i}. \qquad (3)$$

This system can be interpreted in two different ways. The first, and most obvious, is to assume constant returns to scale in all sectors of the economy, in which case the elements of **A** and **l** are invariant to changes in the allocation of the total available labor, *L*, among sectors. (2) and (2') then express the literal truth about the range of possible choices of net output combinations available. Alternatively, however, one may regard (2) and (2') as representing <u>hypothetical extrapolations</u> from a particular set of outputs and the associated activity scales in the various sectors; they are a logical deduction from a singular reality, and by no means preclude the possibility that any actual shift of labor and allied resources among sectors will cause some or all of the input coefficients to change.

Such change results from the existence of non-produced and non-homogeneous inputs; these enter the model indirectly by altering the labor input and produced input coefficients, as scale changes. Decreasing returns may then be set in motion by large-scale, aggregative shifting of resources among sectors. Shifting by an individual producer, by contrast, is likely to have an unmeasurable impact on the technical coefficients, which are therefore constant within the range of variation that is relevant for decision making by individual agents.

An example may help illustrate the distinction between the macro and micro perspectives on the effect of scale changes on the technical coefficients. In a hypothetical agricultural economy, economy-wide shifting between crop farming and dairy farming may cause input coefficients to change in the presence of fixed qualities of land that favor one or the other agricultural activity. The individual farmer, by contrast, may make a decision to move from crops to dairy, perhaps selling crop land and purchasing pasturage of equal value for the purpose, without causing any change in the crop/dairy tradeoff (the marginal rate of transformation). The individual producer’s action is so minute that it results in no perceptible change in relative prices (of either land or goods), and no perceptible change in productivities. As in the case of the perfectly competitive firm, whose output decisions move the economy along the market demand curve to such an imperceptible degree that there can be no perceived effect on the price, the farmer's effect on the aggregate balance of resource use is negligible.

In the simple producer–trader subcase, competitive arbitrage results in the formation of a uniform net income ratio, η. The price equation is

$$\mathbf{p} = \mathbf{pA} + \eta\mathbf{l} = \eta\mathbf{l}(\mathbf{I} - \mathbf{A})^{-1} = \eta\boldsymbol{\lambda}, \tag{4}$$

where η is the equalized net income ratio, the outcome of the process $\eta_i \to \eta$. The equilibrium price vector for this economy is therefore revealed to be proportional to the gross labor input coefficients, $\boldsymbol{\lambda}$.[3]

We may quickly write down the structure of the better-known "capitalist" subcase, in which an equalized wage rate and an equal-

[3] The value of used-up inputs is replaced before net income is calculated, so that differences in ratios of material inputs to labor (however measured) do not "pull" prices away from those that equalize net income ratios across sectors. Differences in startup costs due to divergent material/labor ratios constitute a one-time element that may affect decisions of new producers choosing a sector to enter, but these should have vanishing significance for the price structure in general. In a model with significant fixed capital, however, differing material/labor ratios might well complicate the equilibrium price vector, even in this world of small producers for whom ownership is inseparable from their own labor activity.

ized profit rate take shape. Here price adjustment follows the form

$$\boldsymbol{\pi} \rightarrow w\mathbf{l}\,[\mathbf{I} - (1 + r)\mathbf{A}]^{-1}, \qquad (5)$$

where $\boldsymbol{\pi}$ is the vector of equilibrium prices for this case, from the price equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{A}(1 + r) + w\mathbf{l}$.

We begin with the producer—trader subcase, as explained above, prior to generalizing the results. As a final preliminary,we will

$$\frac{\partial \lambda_i}{\partial l_j} > 0, \ \frac{\partial \lambda_i}{\partial a_{jk}} > 0. \qquad (6)$$

need the property (presented without proof here) that
This states simply that the vertically integrated labor coefficients, λ, are monotonic increasing functions of all of the input coefficients.

In a two-good version of this economy, the hypothetical net

$$Y_2 = \frac{L}{\lambda_2} - \frac{\lambda_1}{\lambda_2} Y_1. \qquad (7)$$

transformation curve (2') reduces to

In the producer—trader model, we have, in competitive equilibrium, prices proportional to λ_i (as we have seen above). Since the goal is to examine the properties of equilibrium, I assume from here on that convergence is complete, and that the λ_i represent, for this case, the appropriate price concept. Clearly, the agents in this model will evaluate the market price ratio in relation to

λ_1/λ_2, the marginal rate of transformation, which therefore emerges as the long-run equilibrium price ratio.

It may be worthwhile to establish this result as explicitly as

$$\lambda = (\lambda_1 \;\; \lambda_2) \qquad \mathrm{l} = (\mathrm{l}_1 \;\; \mathrm{l}_2) \qquad \mathrm{A} = \begin{bmatrix} \mathrm{a}_{11} & \mathrm{a}_{12} \\ \mathrm{a}_{21} & \mathrm{a}_{22} \end{bmatrix}$$

possible. For the two-good case we have

$$\frac{\lambda_1}{\lambda_2} = \frac{l_1(1 - a_{22}) + l_2\, a_{21}}{l_2(1 - a_{11}) + l_1\, a_{12}}.$$

from which, using (1),

Producers move between the two sectors by comparing net income

$$(\eta_1, \eta_2) = \left\{ \frac{p_1(1 - a_{11}) - p_2\, a_{21}}{l_1}, \frac{p_2(1 - a_{22}) - p_1\, a_{12}}{l_2} \right\}.$$

ratios. This amounts to choosing the maximum of

It is easy to show that $\frac{p_1}{p_2} \qquad \frac{\lambda_1}{\lambda_2} \Rightarrow \eta_1 \qquad \eta_2$; producers will therefore migrate in one direction or the other until $\frac{p_1}{p_2} = \frac{\lambda_1}{\lambda_2}$, which is therefore established as the long-run competitive equilibrium condition for this economy. The ratio $\frac{\lambda_1}{\lambda_2}$ is, as we have seen, the marginal rate of transformation <u>seen by the individual producing agent</u>, since it is a condensed expression of the production coefficients defined by the established scale of the sectors of the economy. The individual producer is both powerless to alter, and uninterested in altering, these coefficients, and certainly cannot know or anticipate changes in them.

Our next task is to derive the economy-wide net transformation curve, under the assumption of decreasing returns to scale in both sectors: this is the implicit role of the non-produced inputs in the model.[4] All or some of the input coefficients in a sector are thus assumed to vary directly with the scale of activity in that

[4] The distinction being drawn here is essentially the same as that between (dis)economies internal to the firm and those external to the firm. These have a long history in 20th century discussion, but this has usually been in the context of increasing returns to scale and with regard to the problem of the long-term viability of competitive firm size, a problem that plays no role in my argument (see, e.g., Young, 1928). There is, so far as I am aware, no prior discussion of the relation between non-constant returns and Pareto optimality, despite the large literatures on the two concepts taken separately.

sector, as measured by the proportion of the sector's value added,

$$\mu = \frac{\lambda_1 Y_1}{L} \quad 1 - \mu = \frac{\lambda_2 Y_2}{L}$$

$\lambda_i Y_i$, in total value added L. Defining we can use the monotonic property (6) to express decreasing returns as affecting the gross labor coefficients λ_i: an increase (decrease) in scale in sector j, for example, forces some or all of the input flow coefficients in that sector up (down), which in turn alters λ_j in the same direction. For simplicity (and I believe without loss of generality), the relation between the scale parameter μ and the λ_i is assumed to be linear:

$$\lambda_1 = a + b\mu \tag{8}$$

$$\lambda_2 = c - d\mu \qquad (c > d) \tag{9}$$

Relations (8) and (9) are the central innovation proposed here; they bring non-constancy of returns into the "linear" classical model. The restriction $c > d$ assures that λ_2 is positive at its minimum value, when all labor is allocated to sector 1 ($\mu = 1$). From (8) we have $Y_1^{max} = \frac{L}{\lambda_1(\mu=1)} = \frac{L}{a+b}$; from (9), $Y_2^{max} = \frac{L}{\lambda_2(\mu=0)} = \frac{L}{c}$.

It should be noted that $\lambda_1 Y_1 + \lambda_2 Y_2 = L$, and that $L_1 + L_2 = L$, but that, except at one particular and analytically uninteresting allocation of labor, $\lambda_i Y_i \neq L_i$. $\lambda_i Y_i$ goes to 0 as Y_i goes to 0, which occurs when L_i is still positive (sector i must run at a minimum level adequate to ensure replacement of good *i* used up in both sectors).

$$\mu = \frac{a Y_1}{L - b Y_1} \qquad 1 - \mu = \frac{L - (a+b) Y_1}{L - b Y_1} \qquad (10)$$

Combining the definition of μ with (8), we find

$$\lambda_2 = c - \frac{ad\, Y_1}{L - b Y_1} = \frac{cL - (ad + bc) Y_1}{L - b Y_1}.$$

And from (9) and the definition of μ,

$$Y_2 = L \; \frac{L - (a+b) Y_1}{cL - (ad + bc) Y_1}. \qquad (11)$$

Putting this into $Y_2 = \frac{(1-\mu) L}{\lambda_2}$, we finally obtain

(11) is the equation of the <u>aggregate net transformation curve</u>, or production possibility curve, incorporating the (decreasing-returns) effect on scale when economy-wide shifts of labor (changes in μ) occur. It should be noticed that $Y_1 = 0 \_ Y_2 = Y_2^{max} = L/c,$ and $Y_2 = 0 \_ Y_1 = \; Y_1^{max} = L/(a+b),$ as required.

The “macro” aggregate marginal rate of transformation, MRT, is

$$\mathrm{MRT} = \frac{\partial Y_2}{\partial Y_1} = - \frac{a\,(c-d)\, L^2}{[cL - (ad+bc)\, Y_1]^2}, \qquad (12)$$

found as

which is < 0 owing to the restriction $c > d$; this also guarantees that the denominators in (11) and (12) are positive for all $Y_1 \in (0, Y_1^{max})$. The second derivative, $\frac{\partial^2 Y_2}{\partial Y_1^2}$, is also < 0, as expected.

Now the “micro” marginal rate of transformation, mrt, is eas-

$$\text{mrt} = -\frac{\lambda_1}{\lambda_2} = -\frac{a + b\mu}{c - d\mu};$$

$$= -\frac{a L}{cL - (ad + bc)\, Y_1}. \qquad (13)$$

ily derived:

From (12) and (13) the relation between the two marginal rates of

$$\text{MRT} = \text{mrt}\left[\frac{(c - d)\, L}{c\, L - (ad + bc)\, Y_1}\right]. \qquad (14)$$

transformation is seen to be

The important point here is that <u>the two marginal rates of transformation are in general different</u>. In fact, MRT = mrt if and only if the term in square brackets in (14) = 1, which implies $Y_1 = \frac{d\,L}{ad + bc}$, a particular value of Y_1 falling between 0 and Y_1^{max}. Comparison of MRT and mrt at different compositions of output is straightforward. In the following table, they are presented in absolute value (<u>i.e.</u>, the negative sign is ignored for simplicity):

Y_1	0	$dL/(ad+bc)$	Y_1^{max}
mrt	$\frac{a}{c}$	$\frac{a}{c-d}$	$\frac{a+b}{c-d}$
MRT	$\frac{a}{c}\ \frac{c-d}{c}$	$\frac{a}{c-d}$	$\frac{a+b}{c-d}\ \frac{a+b}{a}$

The information obtained thus far is presented visually in Figure 1. The top panel shows the production possibility curve. The straight lines have slopes determined at their points of tangency or intersection with the curve; the labels attached to them are the economic interpretation to be given to their slopes. In the bottom panel, the MRT and mrt curves are again drawn in absolute value, for ease of presentation.

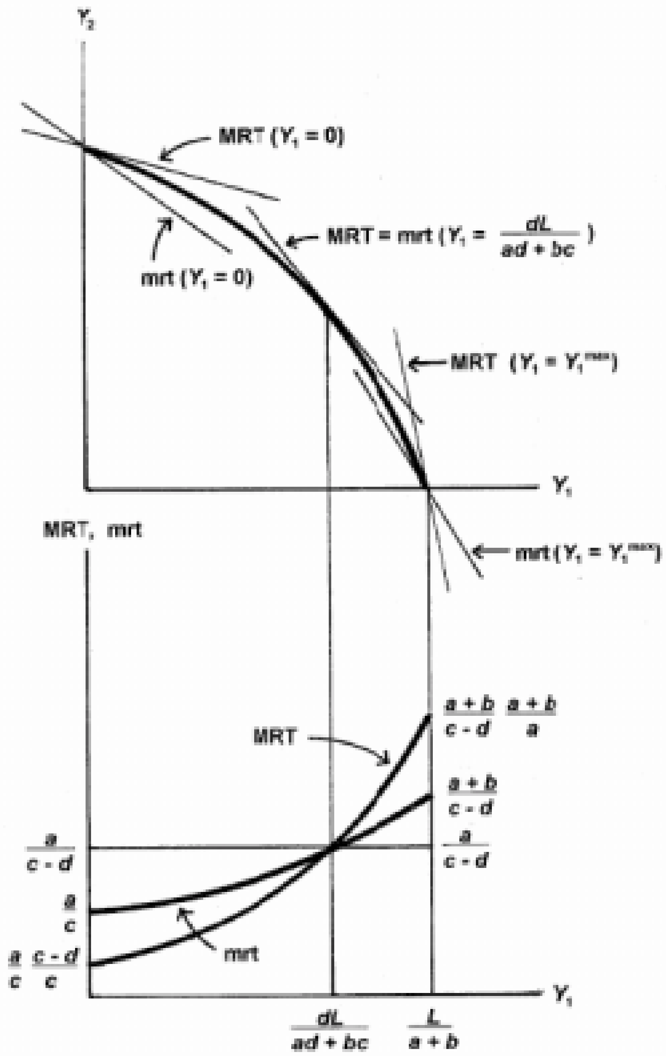

Fig. 1. The two marginal rates of transformation.

The major implication of this construction should now be clear. Individual producers follow mrt, not MRT: their own resource shifts are infinitesimally small in comparison to the macro scale of the economy, and the decreasing returns that operate at that scale are invisible to them. With the economy at some position along the macro transformation curve, the operational marginal rate of transformation is given simply by $-\frac{\lambda_1}{\lambda_2}$, and is constant for the reasonably small ranges of shifting perceivable by any individual producer. This is nothing but a general equilibrium version of the usual competitive assumption: individual firms confidently (and correctly) assume that their own actions are too insignificant to affect either prices or scale conditions at a magnitude to alter the technical coefficients. Rational action, then, equates the marginal rate of substitution in consumption, MRS, to mrt, and, by implication, therefore, MRS ≠ MRT —— unless the parameters underlying the optimal allocation happen by accident to place the economy at $Y_1 = \; dL/(ad+bc)$.

This outcome can be formalized using a standard Cobb-Douglas

$$U = Y_1^{\alpha} Y_2^{\beta}, \qquad (15)$$

utility function:[5]

from which the marginal rate of substitution (once again ignoring

$$\mathrm{MRS} = \frac{U_1}{U_2} = \frac{\alpha\, Y_2}{\beta\, Y_1} = \frac{\alpha}{\beta} \; \frac{L}{Y_1} \; \frac{L - (a+b)\, Y_1}{c\, L - (ad+bc)\, Y_1} \qquad (16)$$

the negative sign for convenience) is

This expression has a vertical asymptote at $Y_1 = 0$, and becomes 0 at $Y_1 = Y_1^{\max}$.

[5] This of course falls short of a truly rigorous presentation in terms of individual utility functions, and falls prey to all of the well-known paradoxes associated with the notion of a social utility function and community indifference curves. It is sufficient to make the point about the welfare loss from mrt ≠ MRT.

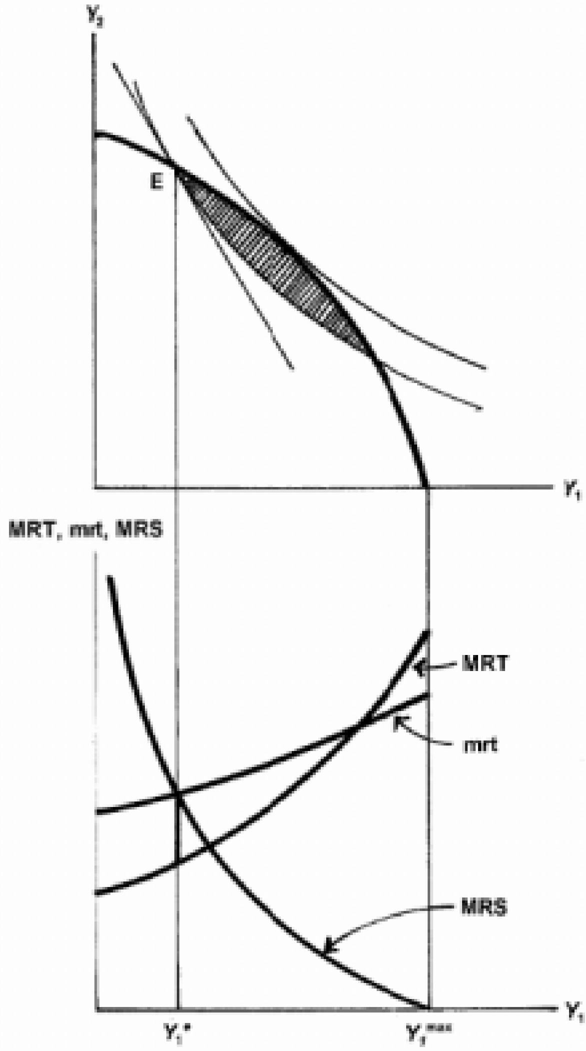

Fig. 2. The welfare loss in competitive equilibrium.

The relation among all three ratios MRS, mrt and MRT is shown in Figure 2. The competitive equilibrium of the economy is at point E in the top panel. This equilibrium is fully determined by both technical conditions and demand. The important point is that, since mrt = MRS at E, that point represents a stable equilibrium: neither producers nor consumers have an incentive to shift. The

MRS and mrt curves clearly have a single intersection, guaranteeing uniqueness of equilibrium, at least in this simple setting. The Pareto inefficiency appears as the heavy line segment in the bottom panel, showing (in this case) MRT < MRS at the equilibrium allocation; and as the shaded area in the top panel, measuring the welfare loss.

3. Concluding and Summarizing

The conclusion is unmistakable: non-constant returns by themselves imply that long-run competitive equilibrium is not Pareto optimal. This conclusion actually follows simply from the observation that individual producers cannot simulate the effects of a combined (macro) shift in resources, and therefore have no way of knowing **MRT** (the vector of marginal rates of transformation for the *n*-good case). Moreover, even if they had the econometric capability to estimate it, they would not want to do so; their momentary profit opportunities would impel them to act according to **p** **mrt** (where **mrt** is the vector of micro marginal rates of transformation). **MRT** is not a factor in producer optimal calculations.

Does this result survive translation to the capitalist case, in which labor markets and capital markets have been separated and a distinct rate of profit formed? First, it is interesting to observe that in a competitive regime of profit-rate maximization/equalization, the relevant rate of transformation is $\frac{\pi_i}{\pi_j}, \neq \frac{\lambda_i}{\lambda_j}$. Capital owners in this case are oblivious to the actual marginal rates of transformation in the economy —— **mrt** as well as **MRT**. (To see this, remember that with **p** = **mrt**, in general profit rates differ among sectors.) Forces other than the actual technical possibilities of transformation are driving price from the supply side.

With profit-rate equalization defining the special rate of transformation $\boldsymbol{\pi}$, there is every reason to expect that **MRS** = $\boldsymbol{\pi} \neq$ **MRT** in competitive equilibrium. The welfare loss persists. By extension, in a more sophisticated dynamic world, all of the widely recognized problems for the welfare theorems come into play on top of the welfare loss identified in the static case, but there is no reason to believe that that loss is somehow eliminated or superseded. The point remains general.

As a final step toward complete generality, we may briefly consider the world of fixed endowments. Imagine a large number of producers, each with a unique transformation curve and vector $\mathbf{mrt}_i$, expressing the perceived transformation opportunities in production; imagine also, for the moment, that these vectors are the relevant ones for optimal calculation. Once we recognize that $\mathbf{mrt}_i = f_i(\boldsymbol{\mu})$, -- that individual production possibilities depend on aggregate tradeoffs embodying decreasing returns and in principle unknown and unknowable to the individual agents -- we still have $\mathbf{mrt}_i \neq \mathbf{MRT}_i$, and a large collection of welfare loss areas such as the one depicted in the top panel of Figure 2. This Pareto inefficiency is more visible in the “classical” world, but it exists also in the neoclassical fixed-endowments models.

It should, perhaps, be noted that the deviation of **mrt** from **MRT** has nothing to do with specific externalities in production, such as the fabled orchardists and bee-keepers or the polluters of rivers with industrial waste. These external effects and other market imperfections constitute further sources of divergence between perceived and actual tradeoffs, and failure to achieve Pareto optimal outcomes. The argument of this paper is that, even absent all such specific external effects and information failures, the very atomistic, “parametric” condition of the individual producer entails failure to perceive and act upon the full social tradeoffs in production.

One may see this failure as yet another externality, in the sense that the effect of an individual producer’s actions on technical coefficients is external to his/her calculations and behavior. Alternatively, it may be considered an information problem: by definition, the individual does not have knowledge of the social **MRT** and would not act on that knowledge if it were available. Semantics should not detain us. If my point is thought to involve externalities and/or imperfect information, it is only in a sense that these qualities are ubiquitous and inherent in <u>all</u> competitive production, as long as returns to scale are not constant.

I make no effort to estimate the relative importance of the welfare loss arising from **mrt** ≠ **MRT**, and in fact doubt whether any attempt at estimation could be meaningful. The role of the link between Pareto optimality and competitive equilibrium is, as implied at the outset, more ideological than practical. I am quite prepared to agree that the result presented in this paper has no practical significance -- as long as the same conclusion is drawn concerning the invisible hand thinking that still dominates the textbooks, despite the guarded and nihilistic conclusions of the post—Arrow-Debreu theorists. This doctrine has the advantage of incumbency, with a deep and enduring impact on social thought. Its refutation is therefore a useful project, despite the apparent simplicity and abstractness of the argument. If atomistic competition is <u>not</u> socially efficient, even on the austere, static terrain on which it performs best, then speculation about economic possibilities is protected from succumbing to formulaic negativism, and new ways of thinking about institutional arrangements seem both possible and worth pursuing.

REFERENCES

Arrow, K. & Hahn, F. H. (1971) General Competitive Analysis (San Francisco, Holden Day).

Debreu, G. (1959) The Theory of Value: An Axiomatic Analysis of Economic Equilibrium (New York, Wiley).

Kurz, H. D. & Salvadori, N. (1995) Theory of Production: A Long-Period Analysis (Cambridge: Cambridge University Press).

Kurz, H. D. & Salvadori, N. (Eds) (1998) The Elgar Companion to Classical Economics (two volumes) (Cheltenham, UK and Northampton, Massachusetts, USA, Edward Elgar).

Mukherji, A. (1990) Walrasian and Non-Walrasian Equilibria: An Introduction to General Equilibrium Analysis (Oxford: Clarendon Press).

Sraffa, P. (1960) Production of Commodities by Means of Commodities (Cambridge, England, Cambridge University Press).

Young, A. A. (1928) Increasing returns and economic progress, Economic Journal, XXXVIII:152 (December), pp. 527-542.