Constant Returns to Scale: Can the Neoclassical Economy Exist?

Alam, M. Shahid

Northeastern University

16 March 2013
Constant Returns to Scale: Can the Neoclassical Economy Exist?

Abstract

Constant returns to scale (CRS) is one of the cornerstones of the competitive general equilibrium paradigm of neoclassical economics. This note argues that the equilibrium solutions of this paradigm are not compatible with CRS. CRS implies that all producers (whatever their scale of production) can produce goods at the same unit costs; and this makes self-production a feasible alternative to market production. In the event, an infinite number of equilibria become possible with a mix of markets and self-production. If labor is the only factor of production, self-production becomes the only option: and the market economy ceases to exist.

M. Shahid Alam*
Professor of Economics
Northeastern University

* The author wishes to acknowledge comments on the paper from Salim Rashid, Professor Emeritus at University of Illinois, Urbana-Champaign, and James Dana, Professor of Economics and Strategy, Northeastern University. The usual disclaimer applies: only I am responsible for the contents of this paper.
All too often the core competitive paradigm of neoclassical economics has been criticized for its many unrealistic assumptions. Many of its critics have also developed new approaches to the study of economics based on more realistic assumptions: among these one may mention the old and new institutional economics, behavioral economics, the economics of limited information, evolutionary economics, and the applications of game theory to the study of strategic behavior. Regardless of these critiques and the presence of alternative approaches to the study of the economy - the neoclassical paradigm could not be dislodged from its position of eminence in economics. Kenneth Arrow (1994: 451) thinks that this is because “competitive general equilibrium theory is still the only coherent account of the entire economy.” Others are of the opinion that the success of neoclassical economics is assured by the kind of economic system it supports: the capitalism of unfettered markets.¹

Can the competitive paradigm be shown to be flawed on grounds of internal consistency: not because of its unrealistic assumptions? It is unarguably the case that neoclassical economics would be hamstrung without the assumption of constant returns to scale in production. Allyn Young (1928) and later Nicholas Kaldor (1967 and 1972) have shown that equilibrium economics becomes irrelevant in the presence of increasing returns to scale. But neoclassical economics is in trouble even with the assumption of constant returns to scale: only, this is a problem that has gone unnoticed. At least since Leon Walras, economists have constructed mathematical models to establish the conditions under which a market economy - with fixed resources, tastes and

¹ Joseph Stiglitz (2001) agrees: in his Nobel Price lecture, he says, “One might ask, how can we explain the persistence of the paradigm for so long? Partly, it must be because, in spite of its deficiencies, it did provide insights into many economic phenomena. … But one cannot ignore the possibility that the survival of the [neoclassical] paradigm was partly because the belief in that paradigm, and the policy prescriptions, has served certain interests.” In a similar vein, James Tobin (1985: 30-31) writes, “In positive as well as normative theory, neoclassical economics was in a much better position than classical economics to respond to the Marxist challenge.”
technology - will attain an equilibrium that also fulfils the conditions for the efficient allocation of resources. Constant returns to scale - or the more general assumption of non-increasing returns to scale - is the cornerstone of all these models. Given this assumption’s centrality, it is surprising that no one has asked if the general equilibrium economy is logically compatible with this assumption. Our concern here is not with the existence - or uniqueness and stability - of solutions to the system of equations that define the neoclassical economy. Instead, we ask a simpler and more basic question. Can a neoclassical economy exist in the presence of constant returns to scale?

A short note might suffice to answer this question, but lest it become too short we will start with the mercantilists and Adam Smith. This digression will provide some historical perspective on how constant returns to scale became one of the cornerstones of neoclassical theory. A little history cannot hurt even in a discipline that does not take too kindly to history.

1. Classical Economics

A central purpose of Adam Smith in writing the Wealth of Nations was to discredit the dirigiste policies of the mercantilist and make the case for laissez faire. It is therefore ironic that he should open his treatise on political economy with a central insight from mercantilist writings regarding the power of increasing returns to scale in manufacturing.²

“The greatest improvement,” writes Adam Smith, “in the productive powers of labor, and the greater part of the skill, dexterity, and judgment with which it is anywhere directed, or applied, seem to have been the effects of the division of labor.” Division of labor contributes to labor productivity in three ways: it augments worker skills, saves time lost in moving from one task to another, and facilitates the increasing use of machines. The use of machinery is perhaps the most important of these three sources of gains in productivity that flow from division of labor: it “enable[s] one man to do the work of many.” In pin manufacturing, Adam Smith tells us, the average output of each man—working as part of a team of ten workers—is 4,800 pins per day; working independently “they certainly could not each of them have made twenty [pins], perhaps not one pin in a day.” The returns to scale in pin-making then are quite extraordinary: division of labor increased the output of each worker by a multiple of 240. Moreover, these gains were attained without the use of any external sources of energy.

Adam Smith makes it clear that the magnitude of productivity gains from division of labor in pin making is not an outlier. “In every other art and manufacture,” he writes, “the effects of the division of labor are similar to what they are in this trifling one; though, in many of them, the labor can neither be so much subdivided, nor reduced to so great a simplicity of operation. The division of labor, however, so far as it can be introduced, occasions, in every art, a proportionate increase in the productive powers of labor (emphasis added).” Agriculture is the exception to these extraordinary productivity gains: the sequential nature of many of its operations limit the scope of division of labor in this sector.

Although he does not elaborate upon it, the dynamic implications of the division of labor are clear to Adam Smith. “As it is the power of exchanging,” he writes in chapter three of his treatise, “that gives occasion to the division of labor, so the extent of this division must always be limited by the extent of

---

4 Smith (1776/1975): 5-6.
that power, or, in other words, by the extent of the market." Clearly, this establishes a two-way feedback between the division of labor and the size of markets: of necessity, this connection must create a virtuous circle of growth. The mercantilists understood this dynamic, at the center of whose thinking stood 'art and manufacture,' and hence they were willing to move heaven and earth to advance manufactures and externally directed commerce. This was the chief engine of growth.

But this is not where Adam Smith wanted to go: straight into the territory of the mercantilists. He was going in the wrong direction.

Having started on the wrong foot, Adam Smith hastily gave up further talk of division of labor. This was not the foundation on which he could build his critique of dirigisme; he would have to change his focus. And this he did in a hurry. After spending the first three chapters of his magnum opus discussing foreign trade, size of markets, productive powers and the division of labor, in chapter four he turned his attention to price theory and allocation of resources in free markets. The division of labor would receive only a few passing mentions in the rest of the Wealth of Nations.

---

5 Smith (1776/1975): 17. Allyn Young (1928: 529) has written that this was "one of the most illuminating and fruitful generalizations which can be found anywhere in the whole literature of economics."

6 "In manufacturing," writes an Italian writer of the early seventeenth century, "production can be multiplied not merely twofold but a hundredfold, and at a proportionately lower cost (emphasis added)." Antonio Serra, A short treatise on the wealth and poverty of nations, translated by Jonathan Hunt and edited by Sophus A. Reinert (London: Anthem Press, 1613/2011):121.

7 In the words of Nicholas Kaldor (1972/1989, 378) - in the middle of chapter four of the Wealth of nations - Adam Smith "suddenly gets fascinated by the distinction between money price, real price and exchange value, and from then on, hey presto, his interest gets bogged down in the question of how values and prices are determined for products and factors of production."

8 Smith (1776/1975): 64, 84,243-44,415-416, 659, and 707. In only two places (415-16 and 706), Adam Smith again briefly touches upon the connection between markets size and division of labor.
Adam Smith sought the superiority of markets in their allocative function. Motivated only by regard for their own interests, and making decisions directed by market prices, buyers and sellers vary the amounts they buy and sell until each market converges upon its "central price to which the prices of all commodities are continually gravitating."\(^9\) The concept of a central price - alternatively 'natural price' - was rooted in constant returns to scale.\(^10\) John Hicks (1989: 10) writes that in spite of what Adam Smith has to say about scale economies in his theory of growth, “we still find that in his value theory, his cost of production value theory, he does not get away from CRS [constant returns to scale].” In the presence of increasing returns to scale, output and price changes in one market would be transmitted to other markets through forward and backward linkages making the concept of an equilibrium and a natural price irrelevant. The die was cast: the ‘invisible hand’ would have to lean on the crutch of constant returns to scale.

Adam Smith offered at least three arguments in favor of free trade. It augmented markets, thus giving impetus to growth via division of labor; it allowed capital to flow into the most productive channels based on a country’s absolute advantage; it gave vent to the country’s surplus. The first argument was flawed since it reinforced existing advantages in manufactures and commerce; therefore it would disadvantage countries that entered free trade with a handicap in these two sectors. The validity of the second argument was problematic because it would work only if capital or labor was completely free to cross borders.\(^11\) The third argument offered free trade as a remedy for the surpluses created by shifts in trade away from a country’s exports; but greater exposure to trade might worsen the problem of surpluses. In other words, the Wealth of Nations, failed to make a strong case for free trade. This deficiency had to be remedied.

\(^11\) Alam (2000: 50-52) has written about the first two contradictions.
It is believed that David Ricardo (1817/1996: 93-95) brilliantly rose to the occasion with his theory of comparative advantage. But he succeeded only by changing the central question concerning the advantages of international trade. For four centuries, the mercantilists had framed their arguments for protectionism within a dynamic framework: they argued that a lagging country was unlikely to improve its chances of growth or preserve its sovereignty under free trade. Ricardo framed his question within a purely static framework: he ruled out changes in technology and tastes, and labor too is available in fixed quantities and is immobile between countries. He makes another crucial assumption: production is subject to constant returns to scale. Within this static framework, free trade for all countries is the best policy regardless of where their comparative advantage lies in potato chips or computer chips.

The classical economists’ shift towards static analysis was an ideological necessity. Britain had outgrown the mercantilist policies that had elevated it from the ranks of a backward country in the early fifteenth century to become the world’s leading economy by the mid-eighteenth century. In an essay on the rich country-poor country debate among Scottish thinkers in the eighteenth-century, J. M. Low writes that this was “the central point of the Wealth of Nations.” “Smith and [Josiah] Tucker,” he continues, “were agreed that Britain was already a rich country and hence that there really was no need for the government to intervene to safeguard advantages which we were in no danger of losing.” Unlike Adam Smith, however, Josiah Tucker had the candor to concede that what was good policy for Britain was not good for poor countries. It would be judicious for them to take protectionist measures to raise productivity in their manufacturing activities.

2. Neoclassical Economists

Once the classical economists had chosen to demonstrate the superiority of free markets within the static framework of allocative efficiency, the path was
clear for the marginalist revolution of the 1870s with its focus on methodological individualism and marginalist analysis.\textsuperscript{12}

In order to demonstrate the efficiency of markets in equilibrium, economic theory would have to show that the decisions of self-serving buyers and sellers led to simultaneous equilibria in all markets that also produced the best allocation of resources.\textsuperscript{13} Leon Walras formally launched this quest in his \textit{Elements of Pure Economics}, published in two parts in 1874 and 1877. Crucially, his general equilibrium analysis assumed constant returns to scale in production: also, capital, labor and land always entered his production function with fixed coefficients. Employing more rigorous mathematics, Kenneth Arrow (1951), Gérard Debreu (1959) and Gérard Debreu (1954) demonstrated the existence of a solution to their system of equations that they claimed is unique; they also established the two fundamental theorems of welfare economics. Their system was built on the assumption of non-increasing returns to scale.

The neoclassical production function with smooth factor substitution and constant returns to scale underpins at least two other ventures in neoclassical economics. Although the concept of marginal product of labor had been around since David Ricardo’s theory of distribution, the marginal productivity theory of factor prices was developed in the 1890s. While earlier writers

\textsuperscript{12} The Austrian economists did not obsess about equilibrium. Unlike the neoclassical economists who assume that economic agents possess perfect information and their decisions produce efficient market equilibria, the Austrians view individuals as possessing only local information about matters that affect that concern them. Nevertheless, they alone have the best chance of making the right decisions in the face of changing tastes, technology and prices.

\textsuperscript{13} Other factors too may have propelled the timing of this shift or the mathematization that accompanied it. In part, at least, it would appear that the marginalist shift away from the labor theory of value may have been a response to the revolutionary connotations that Karl Marx gave to the labor theory of value. Philip Mirowski (1990: chapter 7) has written about the physics envy that pushed some economists in the late nineteenth century to adopt marginalist analysis and mathematical formulae.
had spoken of the law of diminishing returns in the context of labor applied to land, in 1888 John Bates Clark developed a nearly full-blown theory of factor prices that was rooted in the law of diminishing returns. He claimed that this law applies to the marginal return to any homogenous factor when it is combined with fixed quantities of other factors, provided the technology of production remains fixed. However this theory faced a problem: it would not work if payments to products did not fully exhaust the total product. Philip Wicksteed (1894) pointed out that product exhaustion would occur only in the presence of constant returns to scale. Nearly five decades later, George Stigler (1941: 49) added another twist: the law of diminishing returns itself may or may not hold in the presence of increasing returns to scale. In other words, without constant returns to scale the neoclassical theory of production would have to abandoned.

3. Can the Neoclassical Economy Exist?

If constant returns to scale is the cornerstone of the neoclassical economy, it is necessary to ask if this assumption is compatible with the existence of the neoclassical economy. John Hicks has (1989) devoted an entire paper to the question of constant returns to scale, where examines this assumption from many different aspects; but he does not mention the concerns raised in this note. Allyn Young (1928) and Nicholas Kaldor (1972), two trenchant critics of this assumption, also do not refer to any internal inconsistency in the competitive paradigm.

Generations of neoclassical economists have appropriated high-powered mathematics to construct and complete the competitive paradigm. They claim to have discovered the exact conditions that allow for the existence of a

---

14 Indeed, John Hicks (1989: 16) offers these closing thoughts in his paper: "My general conclusion is that CRS, if used with proper precaution in the approximate contexts, can be a help; but if misused, or if it is applied in the wrong context, it can indeed, as Kaldor thought, be a hindrance."
unique and stable equilibrium in their decentralized economy; they also claim that this equilibrium is efficient. Perhaps it is tautologically true that this equilibrium cannot exist if we were to modify any of its key assumptions; no neoclassical economist has made the claim that the competitive paradigm is compatible with a variety of alternative assumptions. Thus Allyn Young (1928), Nicholas Kaldor (1972) and others have demonstrated the irrelevance of equilibrium economics in the presence of increasing returns to scale. This note offers a different critique that questions the internal consistency of the competitive paradigm. It will argue that the competitive model is logically flawed: the equilibrium it identifies cannot exist under its own assumptions. In particular, this equilibrium is not compatible with one of the assumptions it makes about production technology: that it is characterized by constant returns to scale. This demonstration does not depend on abstruse mathematics.

We use two different approaches to establish this result. The first takes the neoclassical paradigm as it is and shows that the equilibrium it establishes is not compatible with constant returns to scale. It also identifies a limiting situation - within the parameters of this model - in which a natural economy can replace the competitive market economy: where the natural economy is one in which no exchanges take place. Alternatively, if we assume with Ricardo that labor is the only factor of production we discover that a natural will necessarily replace the competitive market economy.

In the presence of constant returns to scale, the unit costs of all goods are fixed regardless of their scale of output. This constancy of unit costs implies that no advantage can accrue to any economic agent from specializing in the production of any good. In other words, each person can supply himself with the goods that he needs as cheaply as anyone else however small the quantities of goods that he needs for his own consumption. This line of thinking establishes a presumption that a market economy may not emerge in the presence of constant returns to scale: except when a person’s endowment of factors does not permit him to produce his preferred consumption bundle.
First, consider the competitive economy as it is. Suppose $x_{1i}^*$ and $x_{2i}^*$ denote individual $i$’s optimal consumption bundles in a competitive economy that produces two goods, $x_1$ and $x_2$. In the competitive paradigm the individual $i$ attains his optimal consumption bundle by selling all his factor endowments on the market and then using the proceeds to buy his optimal consumption bundle. It has escaped the attention of neoclassical economists that an individual may walk an alternative path for attaining his optimal consumption bundle. Under constant unit costs an individual may choose to avoid the market and instead self-produce his consumption bundle. This would be quite straightforward if his endowment of factors turned out to be identical to the bundle of factor services that has gone into the production of $x_{1i}^*$ and $x_{2i}^*$. If the factor endowments of all individuals coincides their factor endowments and factor requirements (for the consumption bundle), production for the market would become unnecessary. Individuals would be indifferent in this economy between self-producing their consumption bundles or buying them on the market. In this event, they would prefer self-production over buying on the market. It might still be possible to preserve the market economy: but then we would have to assume that all individuals prefer market-acquired consumption bundles over self-produced ones. On the other hand, the existence of a market economy could be ruled out if we introduce transaction costs into the model.

It is much more likely, however, that an individual may not possess exactly the factor endowment that is required to produce his optimal consumption bundle. An individual can overcome this mismatch between the two bundles via different equivalent exchanges: trading goods, trading factors, or some combination of the two. In the absence of transaction costs - another assumption of the neoclassical model - individuals will be indifferent between the three types of trades. An individual could use his factor endowment to produce goods and meet the deficit between this and his optimal consumption bundle via trading on the markets. Alternatively, he may use some part of his factor endowment for self-production and trade both goods and factors to
attain his desired consumption bundle. As a result, an infinite number of options exist for each individual to attain his consumption objective. If all individuals chose to trade all of their factors this and this alone would reproduce the competitive equilibrium of neoclassical economics. But this is a limiting case: one equilibrium amongst an infinite number of possible equilibria. In all other scenarios, with at least some individuals engaging in some amount of self-production, the economy would combine markets and self-production to satisfy the demands of consumers.

If labor is the only factor of production, it is easy to show that the presence of constant returns to scale will rule out all market exchanges. Since labor is the only factor of production, each individual can produce any bundle of goods so long as this satisfies his labor constraint. If \( x_1 \) and \( x_2 \) are two goods in this economy, and \( a_{l1} \) and \( a_{l2} \) are the corresponding labor coefficients, the individual’s production function is given by

\[
(1) \quad x_1.a_{l1} + x_2.a_{l2} = L,
\]

where \( L \) is his endowment of labor. Equation (1) also represents the individual’s budget constraint. In other words, every individual can self-produce any point he chooses on his budget constraint because the budget constraint is also his production function. Since production and consumption bundles in this economy are identical for all individuals, there is no need for trade. Trade would arise only if production technology - the labor coefficients for different goods - varied across individuals. But the competitive economy rules this out. Hence, the neoclassical market economy cannot exist.

It does seem odd that while economists have been trying so hard to establish the existence of equilibrium in a decentralized economy - and later sought to establish conditions under which this equilibrium would also be unique and stable - it did not occur to them to ask whether there is some way in which constant returns to scale to undermine their economy. We have shown that such an economy cannot exist if labor is the only factor of production. In such an economy there would be no exchange since there was no advantage to
be obtained from specialization. If production required the use of more than one factor, except in the limiting case, this could produce a mismatch between an individual’s factor endowment and his factor requirements. This would force the necessity of trading factors, goods or both. In turn, this creates the possibility of an infinite number of equilibria depending on how much of their factor endowment individuals decide to use for self-production. Apart from one limiting case in which all individuals trade all their factors on the market, a natural economy would co-exist with a market economy under the assumptions of the competitive paradigm.

This is not the first occasion that neoclassical economics has been - so to speak - hoist with its own petard. Once neoclassical economists had established that a perfectly competitive economy is also Pareto-optimal, it nearly ensured that the competitive economy would be the gold standard against which all real-world economies would be judged. Any departures from the perfectly competitive economy would be deemed a market failure. Government interventions then could only be justified if they were aimed at correcting market failures. In the case of externalities, positive or negative, economists proposed that they be addressed by a combination of taxes and subsidies. This logic was used by economists to recommend piecemeal correction of market failures; they ignored the general-equilibrium impact of correcting any one market failure by implicitly assuming away all other market failures. Lipsey and Lancaster (1956-57) were troubled by this piecemeal correction of market failures. As a result, they employed a general equilibrium framework to examine the impact of removing one market failure on other market failures. In what has come to known as the theory of the second-best, they demonstrated that if there are \( n \) market failures in any economy, the correction of all but one of these failures is no guarantee that this will bring the economy closer to Pareto-optimality. In the presence of multiple market failures, some failures may be attenuating the distortionary effects of others. As a
result, removing some of these failures may worsen the distortionary effects of the remaining market failures.

The theory of the second-best should have had a sobering impact on the neoclassical focus on allocative efficiency. Needless to say, the neoclassical economists ignored the second-best theory; when correcting one market failure they simply assume that no others exist. Economists continue to propose seriatim correction of market failures as if unaware of the cautionary tale contained in the second-best theory. This is the standard procedure in the general equilibrium demonstrations of free trade: they assume that there are no distortions in the economy other than the wedge between domestic and international prices introduced by tariffs or quotas on imports or exports. Remove these trade-related distortions: and this will move the economy towards a Pareto-efficient outcome.\(^\text{15}\)

4. Some Concluding Observations

How did neoclassical economists miss this flaw in their core model? Perhaps the answer is to be found in the imperatives of the ideology that the core of mainstream economics has been tasked to defend: the advantages of free market capitalism.

It was Adam Smith who first took up this challenge. With the marginalist revolution of the late nineteenth century, this became the obsessive quest of neoclassical economists: as it were, their holy grail. Adam Smith and David Ricardo had already shown the way: the first by framing the efficiency of markets in a static framework, the second by doing the same for the theory of

\(^\text{15}\) Indeed, a team of distinguished economists spent a great deal of time developing an elaborate taxonomy of distortions and assigning appropriate policies for each of these distortions. They developed a general rule: the best corrective policy is one that attacks the distortions at its source. Policies that correct distortions at their source were ranked as first-best, others as second-best. See Jagdish N. Bhagwati (1971).
trade. In addition, both understood that they needed constant returns to scale if they were to employ a static framework. Once the neoclassical economists had settled on Vilfred Pareto’s definition of an efficient allocation of resources, it appeared to them that their objective was at hand. All that they needed to do was mathematize economics. Physics and engineering became the new model for how to do economics. However, even with mathematical tools, the goal of establishing the Pareto-optimality remained elusive. Nevertheless, this quest could not be abandoned: the more elusive the goal of Pareto-optimality appeared, the greater has been the willingness of economists to suspend their disbelief - to turn their eyes away from the real world. The neoclassical economists believe that they finally reached their Valhalla with the work of Kenneth Arrow and Gerard Debreu; although some critics argue that this is doubtful.\footnote{Frank Ackerman and Alejandro Nadal (2004).}

It is in this intellectual atmosphere - ideological while strenuously claiming to be scientific - that we must seek to understand the failure of neoclassical economics to work out those implications of constant returns to scale that are developed in this note. They could not look too minutely at an assumption that has been so integral to their enterprise. Could not Adam Smith and his neoclassical followers develop a defense of free markets on less restrictive assumptions? The Austrian economists claim that they have done so: and they did it without mathematics too.

Why then do the Austrians take a distant second place to neoclassical economics in the economics discipline? A pregnant question whose answer may have been supplied by Philip Mirowski. In the late nineteenth century, as it was forced to change its methodology under the challenge of the Marxist doctrine, mainstream economics was invaded by engineers and physicists determined to import metaphors from these disciplines into economics. After this, the triumph of mathematics seemed inevitable. Paul Samuelson may have clinched this trend with his \textit{Foundations of Economic Analysis} published in
1947. Mathematics aided in the professionalization of economics: it outfitted economists with the ‘tools’ and ‘models’ that made them look like physicists and engineers. The fate of economics was sealed.

On this subject - mathematization and professionalization of economics - we can let Axel Leijonhufvud have the last word:

The dominant feature, which makes status relations among the Econ of unique interest to the serious student, is the way that status is tied to the manufacture of certain types of implements, called “modls.” The status of the adult male is determined by his skill at making the “modl” of his “field.” The facts (a) that the Econ are highly status-motivated, (b) that status is only to be achieved by making "modls," and (c) that most of these “modls” seem to be of little or no practical use, probably accounts for the backwardness and abject cultural poverty of the tribe. Both the tight linkage between status in the tribe and modl-making and the trend toward making modls more for ceremonial than for practical purposes appear, moreover, to be fairly recent developments, something which has led many observers to express pessimism for the viability of the Econ culture.
References


