UNIVERSITY of GUELPH

DEPARTMENT OF ECONOMICS AND FINANCE

Lectures on John Maynard Keynes' *General Theory of Employment, Interest and Money* (3):

Chapter 3, "The Principle of Effective Demand"

Brian S. Ferguson

CME

Lectures on John Maynard Keynes' *General Theory of Employment, Interest and Money* (3):

Chapter 3, "The Principle of Effective Demand"

Brian S. Ferguson
Department of Economics
University of Guelph
Guelph, Ontario, Canada N1G 2W1
brianfer@uoguelph.ca

July 2013

Abstract

In Chapter 3 of the *General Theory* Keynes sketches out what he calls the essence of the General Theory of Employment. He introduces the Keynesian expenditure-based model, his aggregate demand function and also his aggregate supply function, a concept which spawned much debate among Post-Keynesian economists but which was, for a long time, virtually ignored in mainstream macroeconomics. He sets out the Savings = Investment version of Say's Law and outlines how an economy can settle into an equilibrium at less than full employment.

Lectures on John Maynard Keynes' *General Theory* (3):

Chapter 3, "The Principle of Effective Demand"

Introduction: Definitions, User Cost and Value Added:

Chapter 3 is where Keynes starts to spell out the structure of his model of the macro economy. It is also in this chapter that he establishes firmly his rather annoying practice of introducing a concept with a bare mention and saying that a proper definition will have to wait for a later chapter, so he starts the chapter with, as he puts it, "a few terms which will be defined precisely later".

The first definition he introduces is <u>factor cost</u>, by which he means the payments which entrepreneurs make to the factors of production which they employ, but not including the payments they make to other entrepreneurs. Thus factor cost is made up primarily, but by no means entirely, of labour cost.

The next term he defines is <u>user cost</u> of employment, which covers "the amounts which [the entrepreneur] pays out to other entrepreneurs for what he has to purchase from them together with the sacrifice which he incurs by employing the equipment instead of leaving it idle". Payments to other entrepreneurs are for things like intermediate inputs, and also for capital. In fact, capital as the term is used here is not just fixed capital, or machinery, but includes any manufactured inputs so it includes working capital[1]. In addition, the definition user cost includes

---

[1] In his *Treatise on Money* (1930), Volume I Book III Ch. 9, section (iii) on The Classification of Capital, Keynes defines three classes of capital: Fixed Capital, which consisted of "Goods in use, which are only capable of giving up gradually their full yield of use or enjoyment", Working Capital, which included "Goods in process, i.e. in course of preparation by cultivation or manufacture for use or in consumption, or in transport, or with merchants, dealers and retailers, or awaiting the rotation of the seasons" and Liquid Capital, meaning "Goods in stock, but which are yielding nothing but are capable of being used or consumed at any time". He goes on to say "In the case of unfinished goods there is sometimes ambiguity as to whether raw materials are best regarded as liquid or in process. In Vol, ii chap. 28 we shall complete our definition to the effect that normal stocks required for efficient business are part of Working Capital and therefore in process, whilst surplus stocks are to be regarded as liquid. Thus the unfinished goods existing at any time consist partly of Working Capital and partly of Liquid Capital." Note

that bit about the sacrifice which the entrepreneur incurs as a result of using his equipment: Keynes is getting at an item of opportunity cost here, although, as he says in a footnote, we have to wait for Chapter 6 before we will be given a precise definition of user cost. Then we have:

> The excess of the value of the resulting output over the sum of its factor cost and its user cost is the profit or, it we shall call it, the *income* of the entrepreneur. The factor cost is, of course, the same thing, looked at from the point of view of the entrepreneur, as what the factors of production regard as their income. Thus the factor cost and the entrepreneur's profit make up, between them, what we shall define as the *total income* resulting from the employment given by the entrepreneur.

Note that the total income of the enterprise doesn't include user cost – he gets to the reason for this in a moment. The total income generated by employing a certain amount of labour – i.e. the sum of factor cost plus profit – he also labels the proceeds of the employment. Then he labels as the aggregate supply price associated with a certain level of employment "the expectation of proceeds which will just make it worth the while of the entrepreneurs to give that employment".

Keynes uses "proceeds" of an enterprise in a way which comes across to us as odd: it doesn't refer to the revenue of an enterprise, rather it refers to the incomes of all of the people who are paid directly by that enterprise – the sum of factor costs and profit. It doesn't include the incomes of people who work at other firms and earn their incomes by selling intermediate inputs to our firm – their revenue comes from our firm's user cost, which includes the amounts which our entrepreneur pays out to other entrepreneurs. The same issue comes up a few lines later when Keynes refers to the firm's aiming to maximize the excess of proceeds over factor cost. It sounds as if this should refer to factor cost plus user cost, until you remember that proceeds is defined as factor cost plus profit. Saying that the firm aims to maximize the excess of proceeds over factor cost is just another way of saying that it aims to maximize profit.

In his discussion here of his notion of the aggregate supply price, Keynes makes the point that his concept differs in two ways from the Marshallian concept of the supply price for a particular commodity. The first is simply that the aggregate supply price is an aggregate figure whereas the micro supply price function for a commodity shows the price per unit which would just be sufficient to bring forth successive units of that commodity from a firm. Although he does refer

---

that Keynes' definition of Fixed Capital as including goods which give up their enjoyment gradually would put consumer durables in this category.

to the aggregate supply price of the output of a given amount of employment, Keynes' focus in his definition is on the level employment, not the level of output, so what he is looking at is the aggregate proceeds which would just be sufficient to support a particular level of aggregate employment.

The second relates to what is included in supply price. In micro theory, the supply curve for a competitive firm is its marginal cost curve above the minimum point on its average variable cost curve. Because this curve shows the amount per unit which, if received as revenue, would just cover the addition to total cost of producing each successive unit of output, it gives us the supply price schedule for that commodity. In principle, then, we would expect that finding the aggregate supply price of a certain level of output (and hence of the level of employment required to produce that level of output) would involve taking the area under the marginal cost curve up to the target level of output for each firm and adding up across firms. This isn't the aggregate which Keynes has in mind though, because, as he points out in footnote 2 on page 24, neither his definition of the proceeds of any given level of employment nor of the supply price of that level of employment include the user cost of that level of employment.

User cost, remember, he defined as including both "the amounts which [the entrepreneur] pays out to other entrepreneurs for what he has to purchase from them together with the sacrifice which he incurs by employing the equipment instead of leaving it idle". Since inputs bought from other entrepreneurs will include raw materials, which will be variable inputs and hence contribute to variable and marginal costs, the exclusion of user cost can't be some sort of fixed cost exclusion. Rather, the exclusion of user costs both from proceeds and from aggregate supply price ties in with Keynes' concerns about measurement of aggregate economic concepts. His aim here is to come up with a measure of aggregate income which is both meaningful and sensibly measurable. It's clear from footnote 2 on page 24 that his primary concern here is with double counting. The revenue of our firm (in the sense of price per unit times quantity sold, not to be confused with Keynes' proceeds) has to include an amount which will cover that part of its total costs which it pays out to other firms for intermediate inputs and other kinds of capital. Payments by our firm become income for the second firm, and if we simply added up the revenue of the two firms, we'd be double counting part of the first firm's user costs. By

excluding user cost, Keynes is aiming to use the value added method of measuring national income.

He says in footnote 2 on pg 24:

> If the term is to be interpreted gross of user cost, they can only be overcome by making special assumptions relating to the integration of entrepreneurs in groups according as they produce consumption-goods or capital-goods which are obscure and complicated in themselves and do not correspond to the facts. If, however, aggregate supply price is defined as above *net* of user cost, the difficulties do not arise. The reader is advised, however, to await the fuller discussion in Chapter 6 and its appendix.

The way we handle aggregating revenue defined as including user cost, when we are calculating GDP, is by defining GDP as the value of all new, final goods and services produced in the economy in a given time period. The use of the term "final" in the definition of the expenditure-based measure of GDP is our way of "making special assumptions relating to the integration of entrepreneurs in groups according as they produce consumption-goods or capital-goods". Keynes' preference is to avoid the issue of double counting by totalling the value added at each stage of production, which approach we usually mention, but pass over pretty quickly, in introductory macroeconomics courses. Assuming that the national income accounts have dealt with obscurities and complications, so that our measures do correspond reasonably well with the facts, the two approaches should give the same total value for GDP. Keynes' slightly odd (at least to us) terminology is a reflection of his concern that the *General Theory* should be of practical use to policy makers, and not turn into a purely theoretical exercise.

His aggregate supply price, then, is the aggregate level of proceeds which would just be sufficient to induce firms to produce the level of output which would require them to employ a certain number of workers. His use of "proceeds" here means that he's looking at an aggregate which has been calculated by the value added method, eliminating double counting problems. Given the equivalence between the value added and the expenditure on final goods approaches to calculating GDP, we don't actually have to give much attention to the details of the way he uses "proceeds" either with regard to the aggregate supply price of the aggregate demand price (although we do have to remember that by price he means GDP): Keynes' exposition here was

written in the context of a period in which national income accounting was just starting to be tackled in a serious way[2]. We can simply think of this in terms of GDP.

Aggregate Demand and Aggregate Supply:

Given all of this, Keynes then goes on to define aggregate supply and demand functions:

> Let Z be the aggregate supply price of the output from employing N men, the relationship between Z and N being written $Z = \phi(N)$, which can be called the *aggregate supply function*. Similarly, let D be the proceeds which entrepreneurs expect to receive from the employment of N men, the relationship between D and N being written $D = f(N)$, which can be called the *aggregate demand function*.

Note that Keynes' aggregate supply and demand functions are written in terms of GDP (aggregate expenditure on final goods and services, as defined above) and the level of employment, not expenditure and income. While this isn't the way we write them it makes perfect sense given that Keynes was writing during the Depression and given that his concern was with ways to stimulate British employment. It also matches up with his discussion of what is in essence the dynamic stability of a Keynesian equilibrium, which he gets to in a couple of pages.

Figure 1 below shows the *General Theory* AD diagram as it is usually drawn[3], with employment on the horizontal axis and aggregate demand (D) and aggregate supply (Z) on the vertical. Aggregate demand corresponds to aggregate expenditure as we typically draw the Keynesian cross diagram, with two differences. One is the obvious one, that in Figure 1 we have N instead of Y on the horizontal axis. The other is that Keynes' D curve is typically drawn concave, whereas the aggregate expenditure line in the Keynesian Cross diagram is typically a straight line. There are two reasons for this change in shape: one is that, even if expenditure is linear in income, in a short run model (as the *GT* model is), moving to the right along the horizontal axis involves increasing employment while holding capital constant. Diminishing marginal

---

[2] See, for example, J. M. Keynes (1940): "The Concept of National Income: A Supplementary Note" *The Economic Journal* 50(197), March, 60-65

[3] We say "usually", since Keynes didn't draw it. One of the frustrations encountered in reading the *General Theory* is that Keynes didn't use diagrams – there is only one in the whole of the book, in Chapter 14.

productivity implies that equal increases in N, under those circumstances, will translate into smaller and smaller increases in income, and therefore smaller increases in D. The other is that Keynes assumed that, as a community's income increased, its marginal propensity to consume decreased, so that even if we drew the Keynesian Cross aggregate expenditure curve, it would be slightly concave, looking like the D curve in Figure 1.

There has been a certain amount of controversy in the literature around Keynes' D curve, much of it related to the question of the relation between the D curve and the value of the marginal product of labour curve at the level of individual firms and with Keynes' reference to "the proceeds which entrepreneurs expect to receive" from different levels of employment. If we are talking about a single firm, the increase in revenue which it expects to receive from employing another worker is simply the value of the marginal product of labour at that point, and the firm will decide to hire the worker if the VMPL exceeds the wage. We have to make allowance for the fact that Keynes is dealing with proceeds rather than revenue and the fact that he is working at the aggregate level, but it does not seem unreasonable to suggest that he was thinking of entrepreneurs collectively and defining $D = f(N)$ in terms of expectations as to how expenditure on final goods and services would change as aggregate employment changed. The D curve is a representation of how entrepreneurs collectively view the state of the economy – whether there is likely to be a widespread feeling among them that it might be worth hiring more workers and producing more output.

Keynes' Aggregate Supply Function has been the subject of considerable controversy in the literature. Much of that controversy comes from questions about what Keynes meant in a footnote in Chapter 6 of *GT*; we will deal with that issue when we reach that point in the *GT*. Another part deals with the definition of the supply price, and how it relates to the conventional supply price, given that Keynes has made a point of saying that it is not the same thing. Again, it is important to remember that Keynes is working with his concept of proceeds, so the aggregate

supply price can't simply be an aggregation of the variable cost curves of individual firms since that would result in double counting[4].

The general idea of the Aggregate Supply Price curve, or Aggregate Supply Function, $Z = \phi(N)$ is that we should be able to determine the level of proceeds which firms would have to receive in the aggregate if they were to be going to offer various aggregate levels of employment. Apart from the need to remember that we are working in aggregate terms, and not in per worker or per unit of output terms, this is fairly straightforward.

Converting this to measurable terms is trickier. We cannot simply regard it as the aggregate value, across all firms, of the product of each firm's Marshallian supply price for the level of output produced by various levels of labour inputs and the values of those outputs themselves, (i.e. a simple aggregate of $p_{Si}Q(N_i)$ where p_{Si} is the Marshallian supply price for output level $Q(N_i)$) since each firm's supply price curve has to cover its costs of production, and those costs will, for most firms, include amounts paid to other entrepreneurs, often for variable inputs. Aggregating areas under individual firms marginal cost curves, taking account of the production relation between Q and N to put N on the horizontal axis, would result in double counting. We could aggregate across all final goods and services, to give us a required GDP figure to compare with the expected GDP figure which is associated with the aggregate demand curve, or we could use the value added approach, as Keynes does. Again this is conceptually reasonable although not easy to put into practice – it would effectively involve finding the Marshallian supply price curve for each firm, subtracting from it the amounts paid to other firms (since those will appear in the aggregation when we do the same thing for the producers of intermediate inputs), multiplying by output to get a total proceeds measure, matching output to the level of employment used to produce it, and summing across firms.

The need to match the aggregate supply price to a level of aggregate employment rather than to a level of aggregate output must also be kept in mind when interpreting the Keynesian aggregate supply function: normally, when we draw AD/AS diagrams today, we have a measure of aggregate output on the horizontal axis and a measure of price per unit of aggregate output on the

---

[4] The issue of the relation between Keynes' aggregate supply price and the traditional supply price schedule for a single commodity will reappear in relation to some material at the end of Chapter 4 of the *General Theory*, so, in the manner of Keynes, we will leave the answer incomplete at this point.

vertical, and we interpret the AS curve as an aggregate supply price curve in a sense more closely analogous to the Marshallian supply price curve for an individual product, although still in terms of GDP.

In Figure 2 below we have drawn a Keynesian aggregate supply function with employment on the horizontal axis. The vertical line at NF represents the full employment level of employment, because in Chapter 3, Keynes defines full employment as "a situation in which aggregate employment is inelastic in response to an increase in the effective demand for its output". He argues that this definition is equivalent to the definition he used in the previous chapter, where he defined full employment as the maximum quantity of employment which is compatible with a given real wage. The labour market is at full employment if it is at a real wage which happens to lie on both the demand and supply curves[5]. Keynes, remember, did not reject the classical labour demand and supply curves, rather he argued that there was no dynamic process inherent to the labour market which would cause the real wage to converge on the equilibrium level. That doesn't preclude the possibility that a change in the real wage will lead to a change in the full employment level of employment – i.e. we do not have to accept Pigou's backward-L labour supply curve to use the Keynesian model, but we would have to allow for the possibility that the NF point in Figure 2 below would shift to the right as the real wage rose.

We have drawn the aggregate supply function as linear in employment – Keynes discusses its shape in a later chapter of the *General Theory* – but this should not be confused with the proposition that, were we to draw it with aggregate output on the horizontal axis, it would still be linear. As Keynes defines it, the aggregate supply function shows the aggregate level of proceeds which would just be sufficient to support a particular level of employment. There is no particular reason why that total should not rise linearly with the level of employment, if we assume the wage rate is constant. On the other hand, diminishing marginal productivity means that aggregate output rises more slowly than employment does, so the required proceeds, being calculated on the basis of employment levels, would rise faster than output, meaning that the aggregate supply function would be positively sloped and convex when plotted against output. If

---

[5] Holding the real wage constant – given – at the level associated with the intersection of the labour demand and supply curves and increasing the demand for output, which increases the demand for labour, we will find that at that unchanged real wage employment does not increase (there will be an excess demand for labour) so aggregate employment is inelastic in response to a demand for output.

we were to define an aggregate supply price in per unit of output terms, by dividing the Keynesian aggregate supply price (proceeds) by the aggregate level of output associated with each level of employment, we would find that this AS curve was positively sloped, as in modern textbook AD/AS diagrams.

It is not impossible for the aggregate supply function to be convex when drawn against employment – at lower levels of employment it will generally be possible to increase employment without having to raise the wage, since there will be a pool of involuntarily unemployed labour to draw from. As we approach full employment, however defined, firms which wish to expand output will have to bid a certain amount of labour away from other firms and those other firms will have to respond with higher wages if they do not want to have to contract, so every firm's cost curves will shift upwards. Even after we have converted to proceeds terms, this means that the aggregate supply price of labour will increase at an increasing rate as employment increases, and we will have a diagram like that in Figure 3 below.

The primary use of the Aggregate Demand-Aggregate Supply diagram is as a way of setting out the question of whether increases in aggregate employment are expected to yield sufficient additional GDP-type expenditure to cover the costs of the added production resulting from the increased employment. If we are at a level of N at which D is greater than Z, firms will have an incentive to increase N because the increase in aggregate demand will exceed the increase in the aggregate supply price associated with the increased employment. At the level of the individual firm, this presumably means that the VMPL exceeds the marginal cost of increasing the level of labour employed. We say marginal cost of labour instead of wage because for a firm to increase its output will presumably require its buying more intermediate inputs from a second firm, and the first firm's spending on intermediate inputs will generate the proceeds expected by the second firm. When we are aggregating the gaps between marginal revenue and marginal employment cost across firms, though, we don't run into the conceptual problems associated with aggregating revenue across firms. Keynes' D-Z gap basically reflects an economy-wide expectation of an excess of the value of the MPL over the marginal cost of employing the additional labour, where the marginal cost of the additional output which that labour will produce (holding, as always, the capital stock fixed) has been included in the marginal cost of employing additional labour.

Another point at issue in the literature deals with Keynes' interpretation of the point of intersection between the D and Z curves. He refers to this value of D at this point as the level of effective demand. We would be more inclined to regard the whole of the D curve as effective demand, in the sense that it is the aggregate demand which would appear at different levels of aggregate employment, and to refer to the D-Z intersection point as the Keynesian equilibrium point. The main point at issue, though, relates to the fact that Keynes refers to this point as the point at which entrepreneurs' expectation of profit will be maximized. The argument often made is that if the D and Z curves are total revenue and total cost curves, profit will be maximized where the gap between them is greatest, meaning at the point where their slopes are equal (MR = MC), not the point at which they are equal. The problem with this argument is that they are not in fact total revenue and cost curves. The D curve is expected aggregate expenditure on final goods and services, which will be distributed across all firms as the producers of final commodities pay for the capital and intermediate goods which they use, and the Z curve shows the level of aggregate value added type expenditure which would just be sufficient to support different aggregate levels of employment of labour. In Keynes' terminology, a gap between D and Z at any given level of N means that firms in the aggregate expect that if they expand employment, the additional proceeds which they will receive will more than pay for the additional labour and output. For all the difference between Keynes' aggregate supply price and the Marshallian per unit supply price for an individual commodity, his Z curve is still a supply price curve in the sense that it shows how much firms must be offered – in this case, in aggregate and allowing for Keynes' definition of proceeds – if they are to be willing to produce a certain level of output and hire a certain amount of labour. It will still be true that no individual firm will be willing to expand output unless the marginal revenue from the additional output exceeds the marginal cost, in the usual sense[6].

---

[6] Marshall's discussion of the dynamics of market adjustment when price was not at the equilibrium level focused on quantity responses on the part of firms:

> When therefore the amount produced (in a unit of time) is such that the demand price is greater than the supply price, then sellers receive more than is sufficient to make it worth their while to bring goods to market to that amount; and there is at work an active force tending to increase the amount brought forward for sale. On the other hand, when the amount produced is such that the demand price is less than the supply price, sellers receive less than is sufficient to make it worth their while to bring goods to market on that scale; so that those who were just on the margin of doubt as to whether to go on producing are decided not to do so, and there is an active force at work tending to diminish the amount brought forward for sale. When the demand

Although the Keynesian model is very much focused on aggregate demand, aggregate supply is not neglected in the development of the *General Theory*. It has been argued, though, that in the early days of the Keynesian revolution, when Keynes' model was starting to be taught in undergraduate economics courses, there was an excessive focus on aggregate demand to the point that aggregate supply was treated almost as an afterthought. Harcourt[7] has argued that had proper attention been paid to teaching Keynesian aggregate supply along with aggregate demand, the policy missteps of the stagflationary period could have been avoided.

Before proceeding further into Chapter 3, it is worth going back, briefly, to the third of the footnotes on page 24:

> An entrepreneur, who has to reach a practical decision as to his scale of production, does not, of course, entertain a single undoubting expectation of what the sale-proceeds of a given output will be, but several hypothetical expectations held with varying degrees of probability and definiteness. By his expectation of proceeds I mean, therefore, that expectation of proceeds which, if it were held with certainty, would lead to the same behaviour as does the bundle of vague and more various possibilities which actually makes up his state of expectation when he reaches his decision.

Expectations play a key role in the model of the *General Theory*, but so too does Keynes' skepticism about the formal mathematics of probability and expectations. We will leave a

---

price is equal to the supply price, the amount produced has no tendency either to be increased or to be diminished; it is in equilibrium.
Alfred Marshall (1920): *Principles of Economics* 8th edition, London, Macmillan, Book V, Chapter III. Marshall's discussion did not mean that prices never adjusted – as the quantity offered for sale increased, the supply price increased and the demand price (representing consumers' willingness to pay per unit for the increased quantity) fell. The first step in the Marshallian process of adjustment to disequilibrium, though, involved a quantity response on the part of firms; if firms could sell their output for a higher price per unit than the minimum price per unit which (based on their costs of production) they would be willing to accept, they would increase the amount they produced and offered for sale. The increased amount offered to the market would drive down the price they could get per unit, effectively sliding along the demand curve. This process would stop when the demand price and the supply price were equal.

[7] G. C. Harcourt (1992): "Is Keynes Dead" *History of Economics Review* 18, summer, 1-9. Harcourt is referring in particular to the discussion of aggregate supply in chapter 32 of Lorie Tarshis's introductory economics text: Lorie Tarshis (1947) *The Elements of Economics: An Introduction to the Theory of Price and Employment*, pub. Houghton Mifflin Company. We should note that Tarshis draws the Aggregate Supply Curve with Quantity on the horizontal axis and price per unit on the vertical, as we tend to draw it today, rather than with employment on the horizontal and Keynes' aggregate supply price on the vertical.

discussion of Keynes' views on probability theory until we reach the chapter on long term expectations, and here just note a couple of points.

One is that Keynes is using the term "expectation" a bit more loosely than we would today. Instead of treating it as referring to the mathematical expectation of a probability distribution, he uses it to refer to the possible outcomes in that distribution. The other is that, in this chapter, Keynes is working in certainty equivalent terms – instead of calculating a mathematical expectation, he simplifies matters by treating firms as holding their certainty equivalent with certainty. His handling of expectations becomes much sharper later in the *General Theory*.

Say's Law:

At this point in *General Theory*, Keynes turns his attention back to Say's Law, framing it in terms of his AD/AS diagram[8]. What we make of his discussion depends on whether we accept his characterization of Say as having said that "supply creates its own demand". We have already seen that Say did argue that if there was insufficient demand for products in general, it was because not enough income was being generated to support additional demand, and that the reason not enough income was being generated was that some individuals weren't producing, or were choosing not to produce very much. Keynes' interpretation of this seems to go roughly as follows: In terms of Figure 5 below, suppose the economy is at equilibrium at E1, which we assume is below full employment. Some firms hire in order to increase their output. The increased income which their additional production generates causes the aggregate demand curve to shift up to give a new equilibrium at E2. They hire again and the equilibrium moves to E3 and so on up – the fact that demand rises with supply, in the sense of the D curve rising whenever firms move out along the Z curve means that expected proceeds rise by the full amount of the increase in the aggregate supply price associated with increased employment and output, so firms are rewarded for increasing output with increased revenue. Since expected proceeds continues to rise as the aggregate supply price rises, their expansion of output will be profitable. Competition will induce them to continue to expand output until full employment is reached and

---

[8] He will come back to Say's Law one more time at the end of Chapter 3.

no additional output can be produced. Empirically we wouldn't be able to identify the aggregate demand function, since we will always be at a D-Z intersection. This is the basis for Keynes' argument, later in this chapter, that the classical economists were hopelessly optimistic in their expectation that the system would always tend to move towards full employment, but as we've already noted, there is doubt as to whether anyone but Keynes (and depending on how you read him, possibly Say) really believed in this version of Say's Law. There is, however, a somewhat more plausible story which we can tell about Say's Law, but to tell that we need to run through the next part of Chapter 3, which is Keynes' outline of his economic model.

The essence of the General Theory of Employment:

In Part II of Chapter 3, Keynes summarizes the model which he intends to set out in the following chapters, "even though it may not be fully intelligible" at this stage in the book. He then says:

> In this summary we shall assume that the money-wage and other factor costs are constant per unit of labour employed. But this simplification, with which we shall dispense later, is introduced solely to facilitate the exposition. The essential character of the argument is precisely the same whether or not money-wages, etc., are liable to change.

This statement: that constant money wages are a simplifying assumption and not a crucial one, is one which many authors have cited but which still hasn't dispelled the view that the *General Theory* is at heart a sticky-wage model. Perhaps that is because slow wage adjustment seems to satisfy the statement that wage constancy will be dispensed with, or perhaps it is simply that Keynes' later exposition of a flexible wage model strikes people as unconvincing. The material in Chapter 2 and in the 1939 *Economic Journal* paper with regard to the cyclicality of real wages shows that Keynes wasn't being too dogmatic about wage rigidity, and his simplifying assumption really only applies to the summary model which he sets out in this chapter.

Turning to Keynes' summary, then, we find the following:

> When employment increases, aggregate real income is increased. The psychology of the community is such that when aggregate real income is increased aggregate consumption is increased, but not by so much as income.

This is really an unfortunate statement. Keynes frequently[9] refers to the tendency for consumption to increase by less than income as a fundamental psychological law, which creates the impression that he was some kind of 1930s behavioural economist and that his theory of consumer behavior was not rooted in utility maximization theory but rather in his assumption about humans obeying some kind of ill defined psychological drive, or habit. As we will see when we reach the chapters on the propensity to consume this is not correct, but his repeated references to this fundamental psychological tendency have made it part of the conventional wisdom about Keynesian economics.

Having set himself up this way, Keynes goes on to discuss the implications of the marginal propensity to consume being less than one. In terms of Figure 6 below, his story runs as follows. Assume the economy is in equilibrium at N1, with D1 = Z1. Now suppose that firms in the consumer goods sector decide to increase employment from N1 to N2. This way of starting the process – with consumer-goods firms hiring more labour despite being at equilibrium – is important for a couple of reasons. At one level it helps us see why Keynes thought in terms of a diagram with N, rather than Y, on the horizontal axis, at least at this stage in the *GT*. On another level, Keynes generally treats firms as playing a very fundamental decision making role in the macro-economy. In terms of Figure 6, then, consumer-goods producers expand employment and move the economy from N1 to N2 on the horizontal axis. The aggregate supply price associated with this new level of employment is found at Z2 on the aggregate supply function, but because the marginal propensity to consume is less than one, consumer spending increases along the D curve, not the Z curve, and settles at D2, which is below Z2.

This is where Keynes' assumption that the initial increase in employment was in the consumer goods sector comes in useful as a simplifying device. When the consumer goods firms increased (total) employment by N2-N1, they spent an amount equal to the increase in the aggregate supply price which was associated with that increase in N, as shown by the Z curve. We might think of them as drawing on their financial reserves, or as borrowing, to do the hiring, but by the

---

[9] See Chapter 8, for example.

definition of the aggregate supply price actual spending will have moved to Z2. All of the increase in Z is incurred by the consumer goods sector, since that was where the increase in employment occurred. The increase in D is spending on consumer goods, since that is what consumers are assumed to spend their income on. Because the newly hired workers will save part of their newly acquired incomes, while spending on consumer goods does increase, it does not increase by enough to cover the costs of employing the additional labour. The upshot is that while the consumer goods sector which created the new jobs will get some of their costs back in the form of increased spending on consumer goods, they won't get it all back and will lose money on the increased employment. The result is that they will cut back and lay the workers off again.

The only way, then, that spending will rise by enough to support the new hires, given that the marginal propensity to consume is less than one, is if investment spending increases by enough to fill - or swallow - the gap Z2-D2. Since that gap is equal to the saving by the newly hired workers, the saving is available to finance that increased amount of investment.

Savings, Investment and Say's Law:

This brings us back to Say's Law, in particular to the issue of the direction of causality between savings and investment. Both the Keynesian and the classical models assume that S = I; the issue between them is whether an increase in S drives an increase in I. Say certainly assumed this, and the classicals generally did, although the mechanism they had in mind was quite different, relying not on direct investment by savers but on changes to the interest rate bringing investment into equilibrium with (increased) savings.

Whatever the mechanism, if increases in S are automatically invested, so that I goes up by the same amount as S has increased then, in terms of Figure 7 below, when N increases from N1 to N2, increased consumer spending causes D to increase from D1 to D2 leaving Z2–D2 as increased savings. Firms in the investment sector borrow the increased savings and spend it on investment, causing the D curve to shift up by exactly the amount of the savings, Z2-D2. The result is that aggregate demand increases by exactly the amount needed to cover the remainder of

the increase in Z and the firms whose increased hiring set the process off will receive increased proceeds sufficient to cover their increased aggregate supply price. The economy will settle in at a new equilibrium point, with employment level N2.

So long as there are unemployed workers, consumer goods firms can repeat this process, and, with causality running from S to I (as Keynes says it does in the classical model, and certainly in Say's model) will do so until full employment is reached. In Keynes' model, since increases in S do not cause increases in I, the experiment will fail and the system will revert to its original equilibrium at N1. As Keynes puts it:

> Thus — except on the special assumptions of the classical theory according to which there is some force in operation which, when employment increases, always causes D_2 to increase sufficiently to fill the widening gap between Z and D_1 — the economic system may find itself in stable equilibrium with N at a level below full employment, namely at the level given by the intersection of the aggregate demand function with the aggregate supply function.

One question which often arises in connection with the model in the *General Theory* is whether Keynes was asserting that the economy could remain stuck at a level of employment below full employment or whether he was arguing that it will tend to return to full employment, although the speed of adjustment will be slow. In Chapter 3 he quite explicitly says that it can settle in at a less-than-full employment equilibrium. It might be argued that this is a consequence of downward wage rigidity, but it must be borne in mind that when Keynes assumed fixed wages he did so only as a simplifying assumption, and that he claimed that the model also worked in the case of flexible wages.

In Keynes' more formal summary of his model, he defines some notation which will reappear through the *General Theory*. Aggregate Demand has already been labeled D, but now is subdivided into spending on consumption goods, D_1 and spending on investment, D_2. (We should note that these are expected values – as we saw above, expectations are key elements on Keynes' model.)

Keynes assumes that the economy is at equilibrium, though not necessarily at full employment equilibrium. Equilibrium is defined as the level of employment at which $D_1 + D_2 = \phi(N)$, where $\phi(N)$ is the aggregate supply function. This equilibrium cannot be at a level of employment

which is greater than the full employment level from classical theory, since "the real wage cannot be less than the marginal disutility of labour". This statement is a reminder that Keynes still has in his model a labour supply curve which reflects labour's utility maximizing income-leisure trade-off problem; what he has rejected is the mechanism by which the wage moves to the level at which the value of the marginal product of labour curve (labour demand curve) cuts this labour supply curve, hence there is no reason to expect the equilibrium level of employment to be equal to the classical full employment level. The actual equilibrium satisfies $D_1 + D_2 = \phi(N)$, and D_1 can be written as a function of N, $\chi(N)$, so we have $\chi(N) + D_2 = \phi(N)$ or, Keynes' preferred way of looking at it for purposes of exposition of the model, $\phi(N) - \chi(N) = D_2$, which says that at equilibrium the gap between the aggregate supply price and spending on consumer goods (both written as functions of N) must be filled by investment spending. If this condition does not hold, the system cannot be at equilibrium. Keynes emphasizes this point by saying:

> (5) Hence the volume of employment in equilibrium depends on (i) the aggregate supply function, ϕ, (ii) the propensity to consume, χ, and (iii) the volume of investment, D_2. This is the essence of the General Theory of Employment.

It is worth noting that Keynes here includes the aggregate supply function as a key part of his model. For a long time, that part of the model got short shrift in the way Keynesian economics was taught.

A couple more elements in Keynes' summary of his model are worth quoting here:

> For every value of N there is a corresponding marginal productivity of labour in the wage-goods industries; and it is this which determines the real wage. (5) is, therefore, subject to the condition that N cannot *exceed* the value which reduces the real wage to equality with the marginal disutility of labour. This means that not all changes in D are compatible with our temporary assumption that money-wages are constant. Thus it will be essential to a full statement of our theory to dispense with this assumption.

And further, on the matter of real wages:

> Thus the volume of employment is not determined by the marginal disutility of labour measured in terms of real wages, except in so far as the supply of labour available at a given real wage sets a *maximum* level to employment. The propensity to consume and the rate of new investment determine between them the volume of employment, and the volume of employment is uniquely related to a given level of real wages — not the other way round. If the propensity to consume and the rate of new investment result in a deficient effective demand, the actual level of employment will fall short of the supply

> of labour potentially available at the existing real wage, and the equilibrium real wage
> will be *greater* than the marginal disutility of the equilibrium level of employment.

Again, what he is saying here is that while the real wage/employment point will be on the MPPL curve, it is the level of employment which determines the real wage and not vice versa. And again, as we have noted before, this means that in both the Keynesian and Classical models, increases in employment (in the short run) will be associated with reductions in the real wage. What differentiates the models is the direction of causality, and that, empirically, is not easy to test.

This leads us to one more statement about unemployment equilibria:

> This analysis supplies us with an explanation of the paradox of poverty in the midst of plenty. For the mere existence of an insufficiency of effective demand may, and often will, bring the increase of employment to a standstill *before* a level of full employment has been reached. The insufficiency of effective demand will inhibit the process of production in spite of the fact that the marginal product of labour still exceeds in value the marginal disutility of employment.

Keynes goes on to make a few observations about the marginal propensity to consume (a term which he introduces here but doesn't define until Chapter 10). One we've already mentioned; he assumes that the mpc will be lower in a rich community than in a poor one, which implies that a rich community will have to fill a bigger gap between D and Z than will a poor one, if it wants to remain rich. If it fails to fill that gap, "the working of the principle of effective demand will compel it to reduce its actual output, until, in spite of its potential wealth, it has become so poor that its surplus over its consumption is sufficiently diminished to correspond to the weakness of the inducement to invest." He also says, in an observation which will play an important role later in the *General Theory*:

> Not only is the marginal propensity to consume weaker in a wealthy community, but, owing to its accumulation of capital being already larger, the opportunities for further investment are less attractive unless the rate of interest falls at a sufficiently rapid rate; which brings us to the theory of the rate of interest and to the reasons why it does not automatically fall to the appropriate level, which will occupy Book IV.

This remark about the diminished opportunities for investment in rich communities will play a significant role in the policy arguments Keynes makes in later chapters.

The remainder of Chapter 3 is a reprise and elaboration on Keynes' complaint about the neglect of the concept of aggregate demand in Classical economic theory: "It could only live on furtively, below the surface, in the underworlds of Karl Marx, Silvio Gesell or Major Douglas."[10]

And finally, this section contains another of Keynes' better known (among economists, at least) aphorisms: "For, since Malthus was unable to explain clearly (apart from an appeal to the facts of common observation) how and why effective demand could be deficient or excessive, he failed to furnish an alternative construction; and Ricardo conquered England as completely as the Holy Inquisition conquered Spain."

---

[10] Major C.H. Douglas was the theorist behind the Social Credit movement, some of whose policies were almost implemented by the Social Credit party government in the Canadian province of Alberta in the 1930s, only to be blocked by a ruling of the Supreme Court of Canada. Silvio Gesell was a proponent of stamped money, meaning currency which lost value the longer you held it, the idea being that this would prompt people to spend it rather than save it – an idea similar to the idea of deliberately generating inflation to stimulate spending and discourage saving. Despite Keynes having associated them here, Major Douglas did not approve of Gesell's ideas.
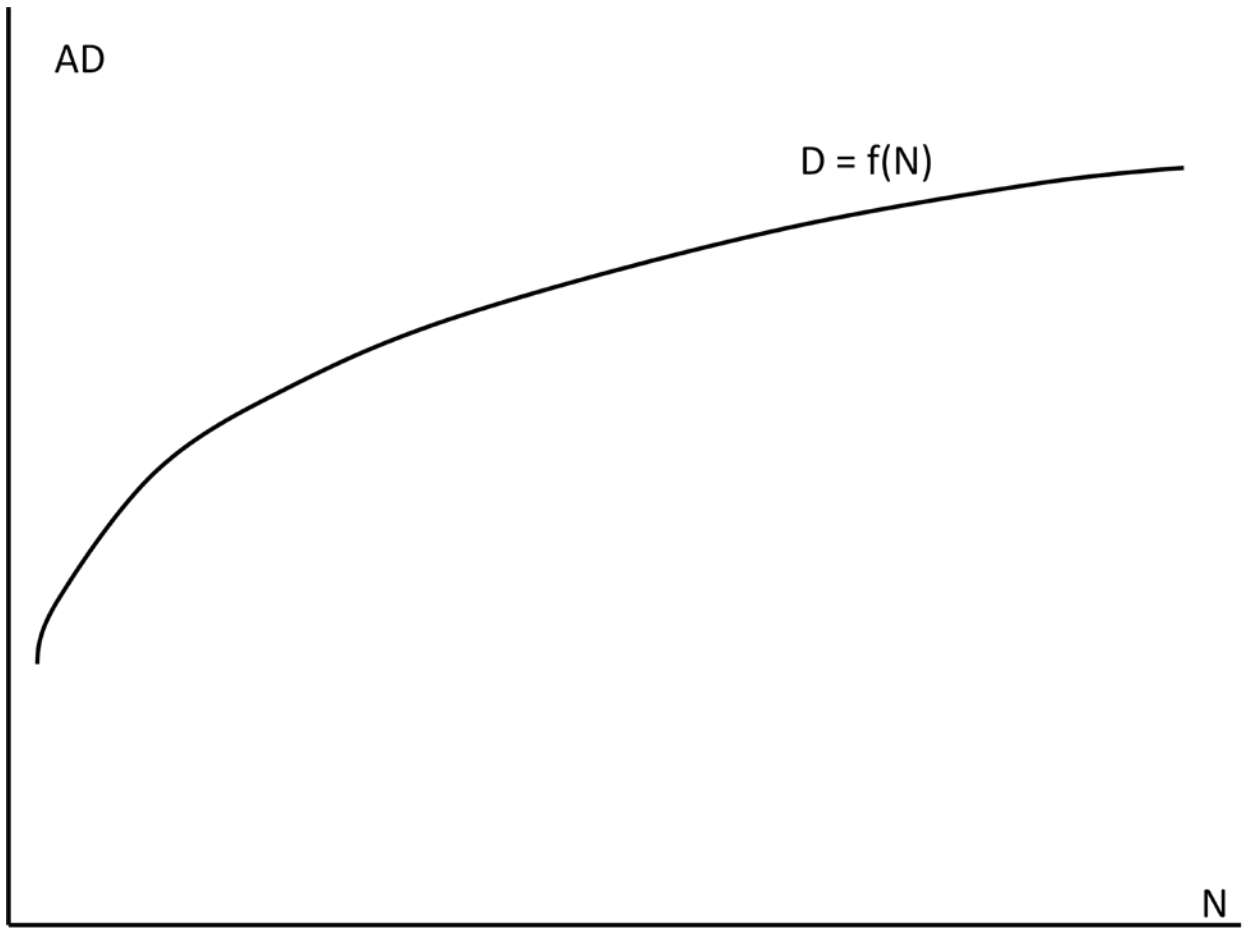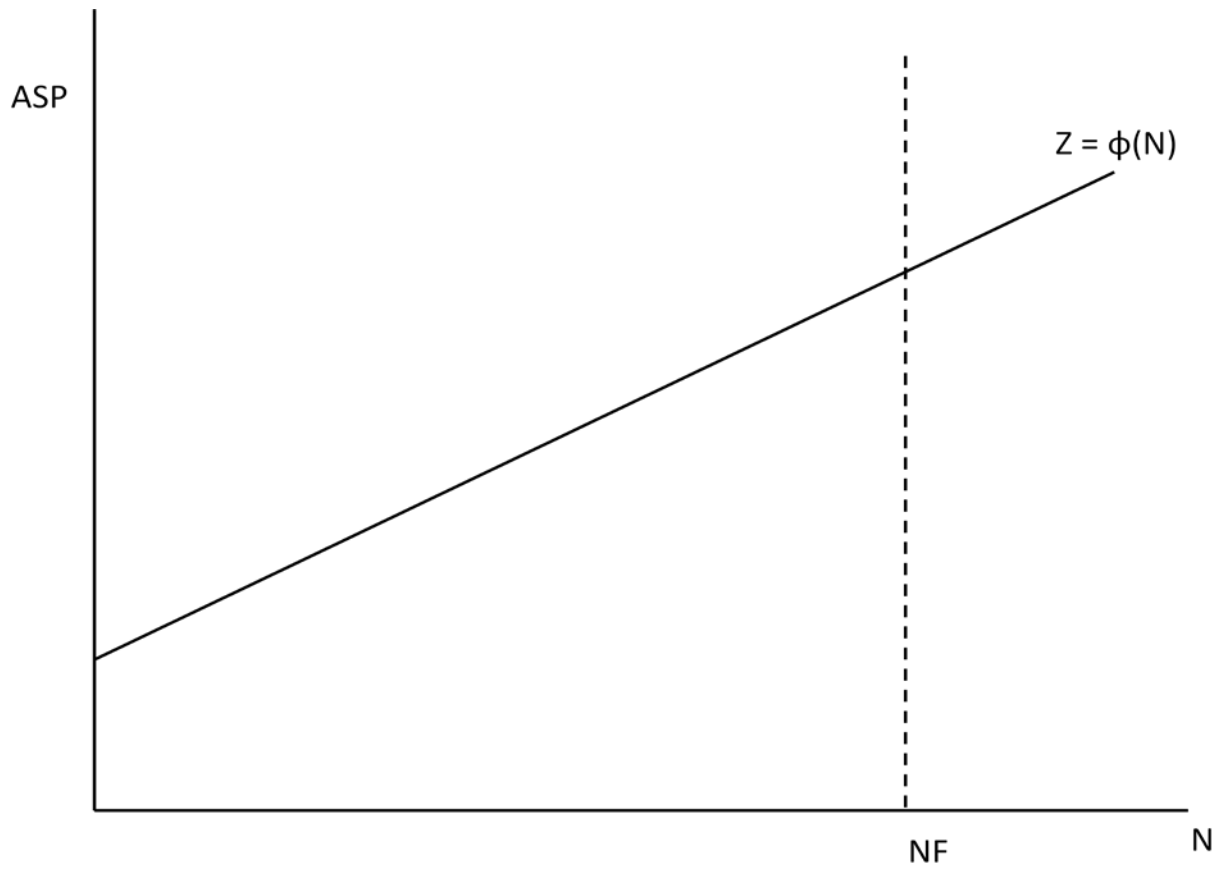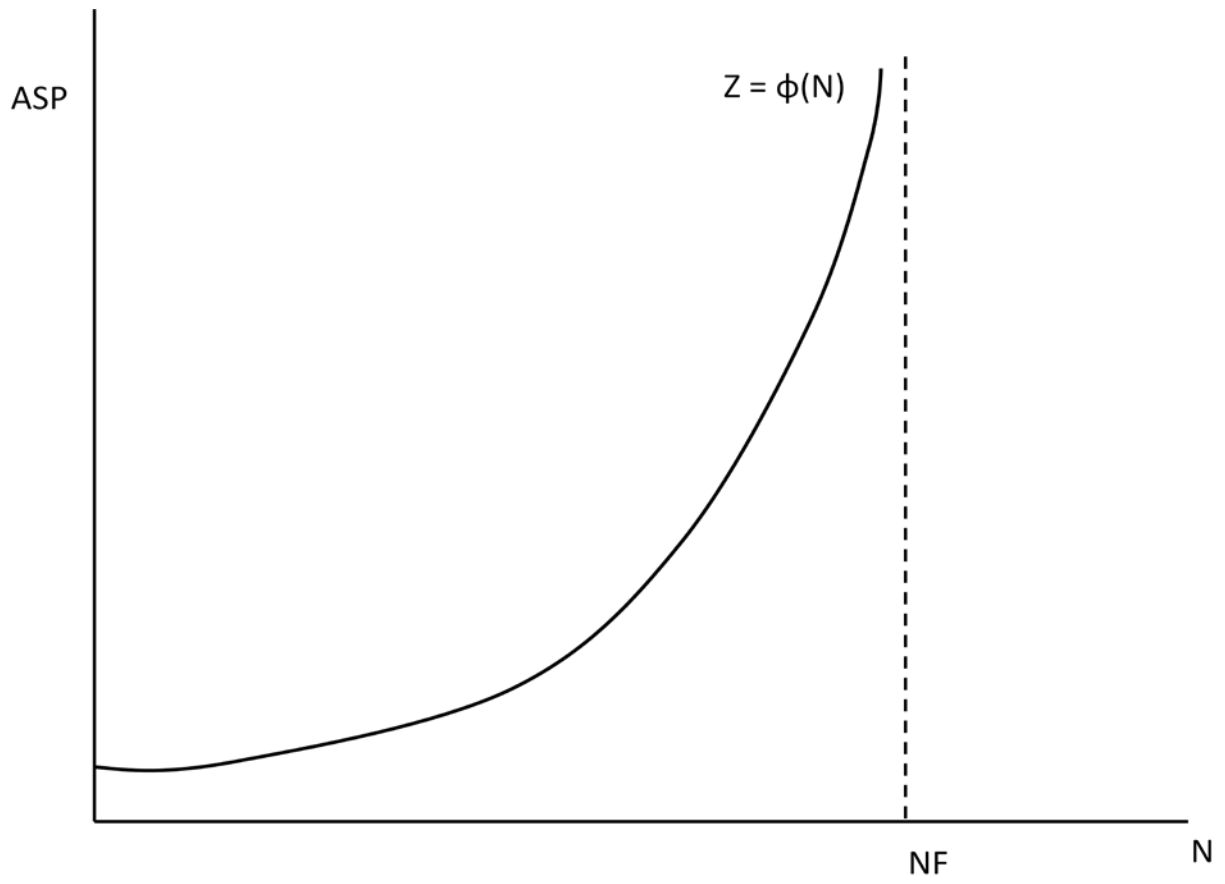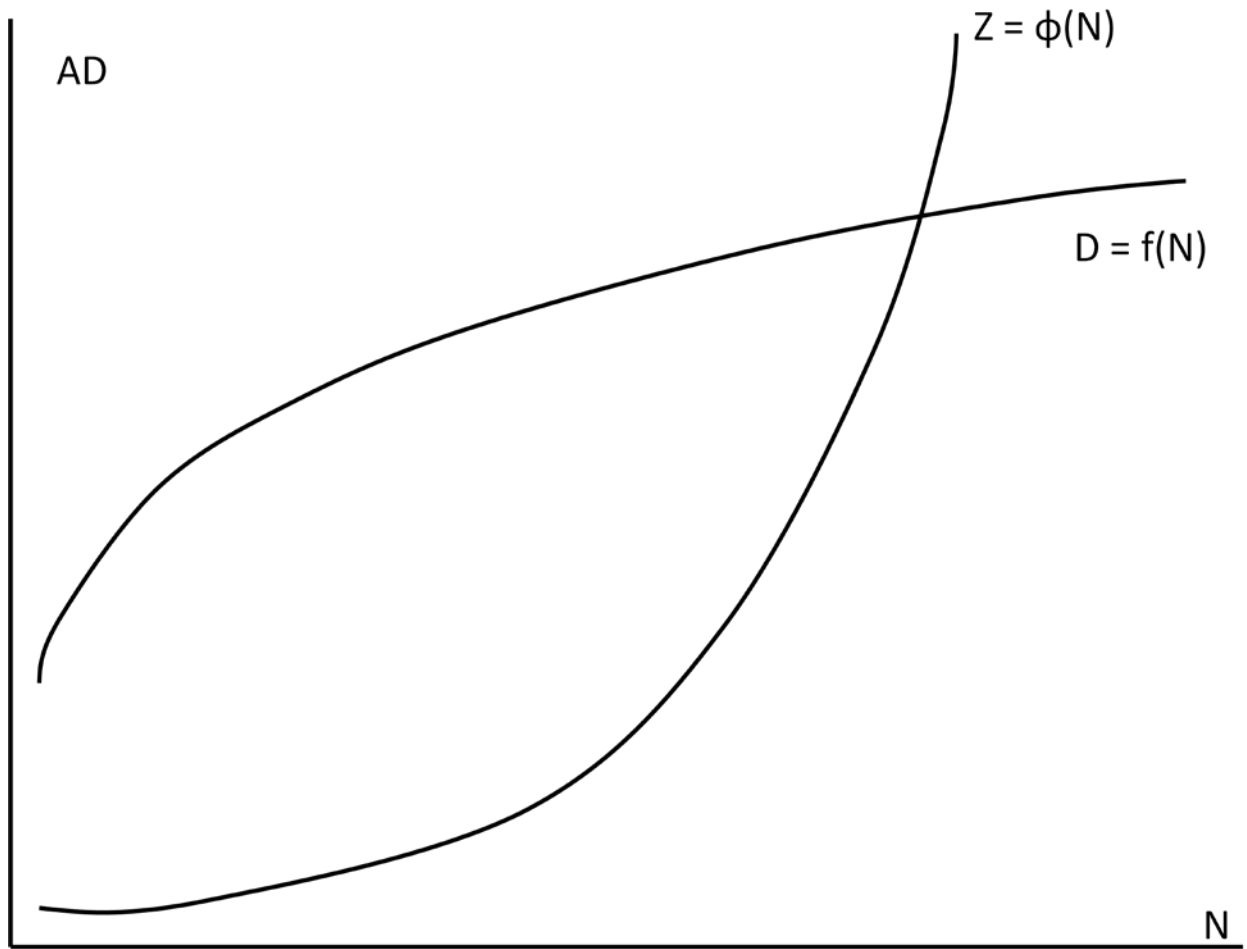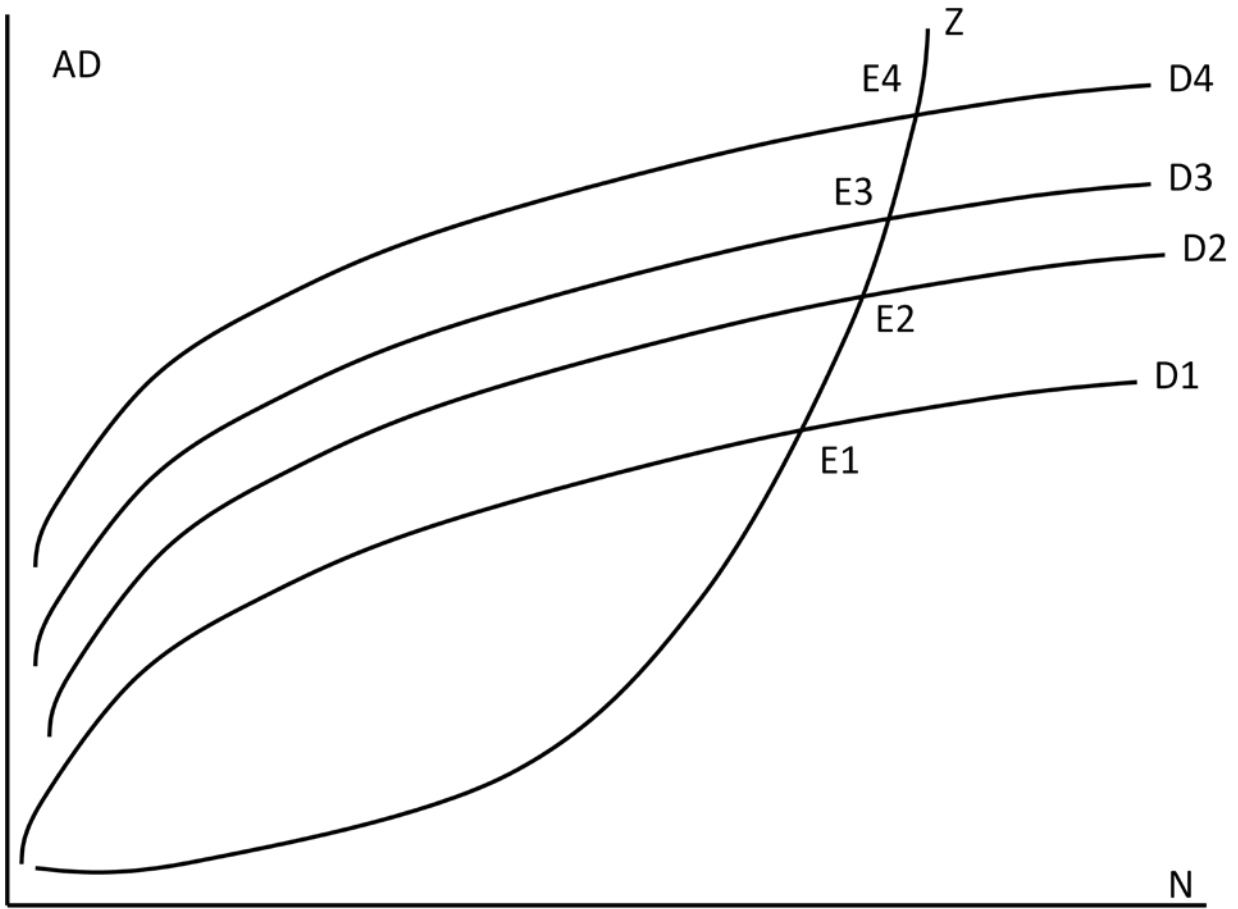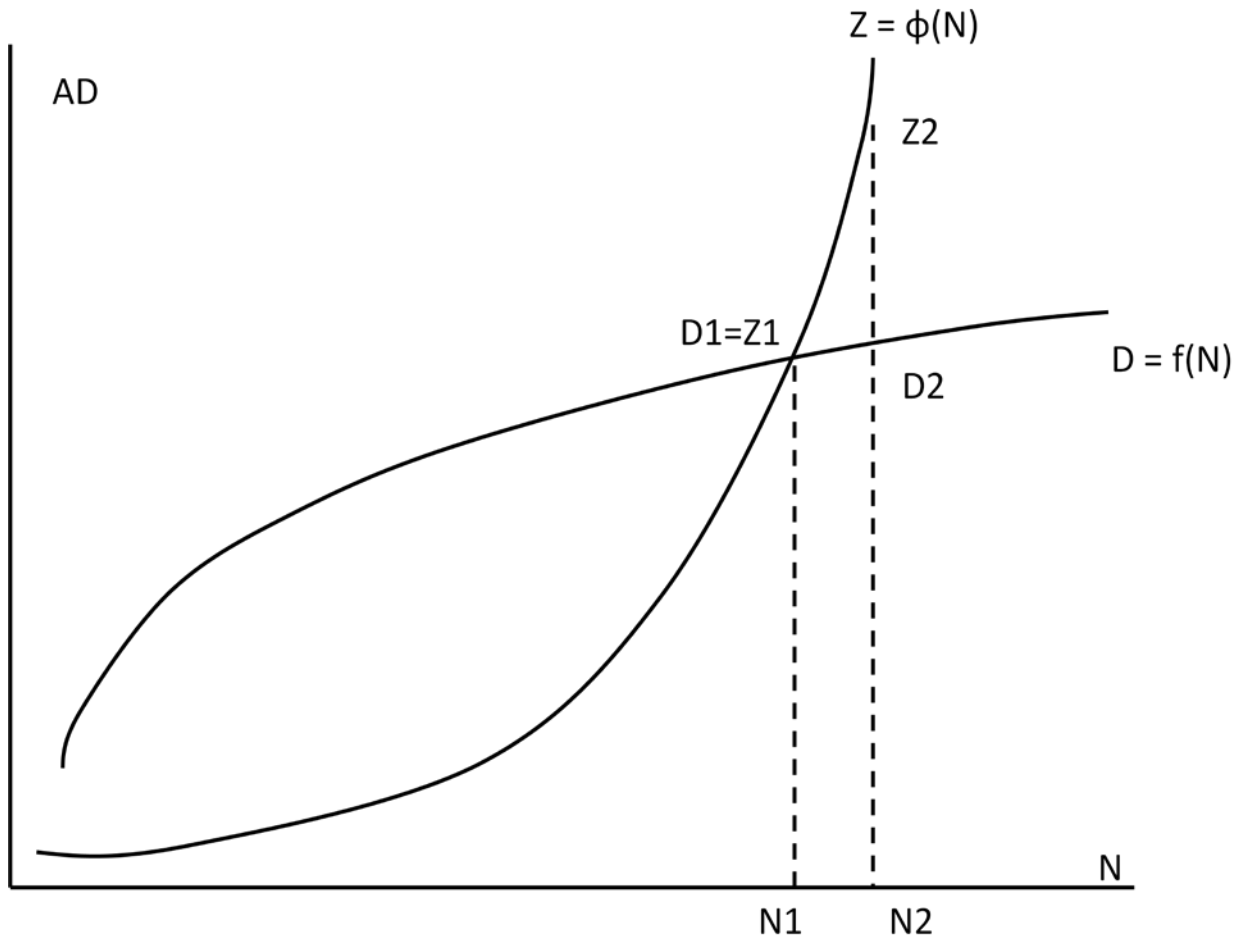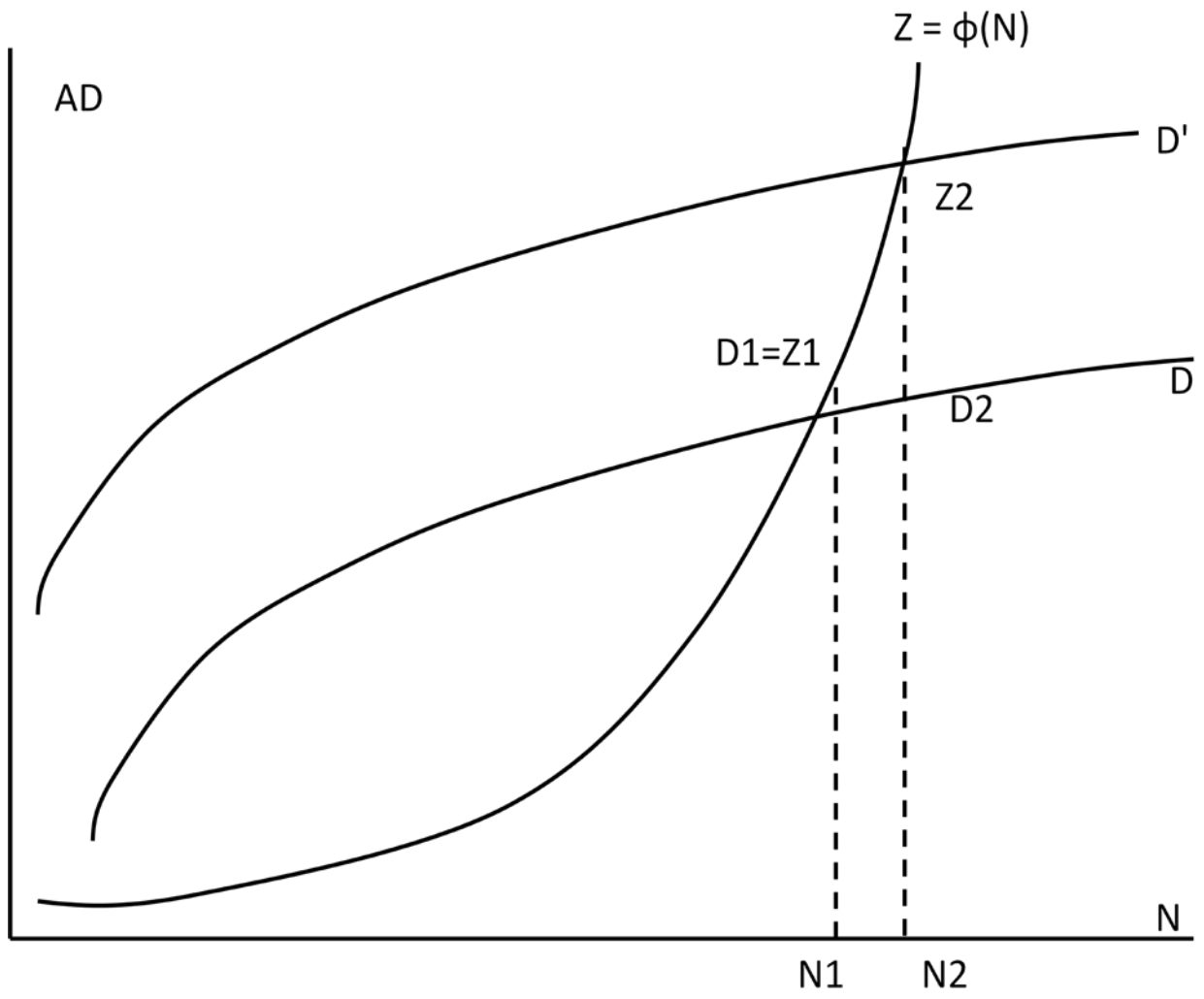
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7