



DEPARTMENT OF ECONOMICS AND FINANCE

DISCUSSION PAPER 2013-09

Lectures on John Maynard Keynes' *General Theory of Employment, Interest and Money* (4):

Chapter 4, "The Choice of Units"

Chapter 5: "Expectations as Determining Output and Employment"

Brian S. Ferguson

JULY 2013



College of Management and Economics | Guelph Ontario | Canada | N1G 2W1
www.uoguelph.ca/economics

Lectures on John Maynard Keynes' *General Theory of Employment, Interest and Money* (4):

Chapter 4, "The Choice of Units"

Chapter 5, "Expectations as Determining Output and Employment"

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

July 2013

Abstract

In Chapter 4 of the *General Theory*, Keynes discusses the units of measurement he will be using in the remainder of the book, in particular his reason for measuring in nominal rather than real terms, objection to aggregate measures of real output and physical capital stock, and his concept of wage units, which is a source of difficulty in following bits of the later exposition. Chapter 5 introduces expectations and discusses the role of short run expectations in determining the behavior of firms and of economic aggregates.

JEL Codes: B10, B12, B13, B22, B31, E12, N12, N14

Keywords: Keynes, General Theory, Keynesian Economics, Classical Economics, Wage Units, Labour Units, Short Run Expectations, Long Run Expectations, Economic Dynamics

Lectures on John Maynard Keynes' *General Theory* (4):

Chapter 4, "The Choice of Units"

Chapter 5, "Expectations as Determining Output and Employment"

Chapter 4: Units of Measurement: Introduction:

Chapter 4 of the *General Theory* is entitled "The Choice of Units", and is the first of three chapters in which "...we shall be occupied with an attempt to clear up certain perplexities which have no peculiar or exclusive relevance to the problems which it is our special purpose to examine.". This would seem to suggest that it is a chapter which can be skipped over, but as it happens the definitions and approach introduced here are key to the development of Keynes' theoretical model. It is in this chapter that he makes clear why he chooses to work in the terms in which he does work, and defines things like wage-units and labour-units; for most readers trying to keep track of what it means to say that a variable is measured in terms of wage units is one of the more frustrating and distracting aspects of reading the *General Theory*.

Keynes begins Chapter 4 with a brief discussion of the national dividend, a concept which Pigou introduced as a measure of aggregate output. According to Keynes, the national dividend refers to real output, not output measured in money terms, and to net rather than gross real output, or "the net addition, that is to say, to the resources of the community available for consumption or for retention as capital stock, due to the economic activities and sacrifices of the current period, after allowing for the wastage of the stock of real capital existing at the commencement of the period.". Since consumption, as Keynes uses the term, includes what we would now separate into C, G, and net exports, this is a definition which is consistent with the measure of net national or net domestic product in real terms. Keynes objects to the notion that the real output of the various industries in the economy can be aggregated into some sort of meaningful real aggregate output – as he puts it, "the community's output of goods and services is a non-homogeneous complex" and he takes the view that any aggregate of these very disparate industry-level real outputs makes no logical sense and has no sensible meaning.

He has a similar objection to the idea of an aggregate real capital stock – capital is too non-homogeneous for a single real aggregate value to have any useful interpretation. With capital, too, there is the problem of obsolescence, in that new capital may be more productive than old capital, and old capital may suffer from obsolescence simply because a new, more productive type of capital has come onto the market. In his calculation of the national dividend Pigou, according to Keynes, includes a deduction for normal obsolescence, but when you're working in real terms that amounts to making a deduction from physical capital and so reducing the measured amount of physical capital when there has been no actual reduction in the real quantity of capital in any industry. Obsolescence is an issue which arises in growth models when we do not want to treat technological change as a simple multiplicative factor applied to all capital equally. One way of getting at the issue is through a model with different vintages of capital, but then we run into the issue of how to construct a sensible aggregate of the different vintages of capital when the introduction of a new vintage doesn't by itself reduce the physical quantity of, or the amount of output produced by, older vintage equipment. Keynes' problem here is not with the notion of capital heterogeneity, nor with the idea of vintages, it's with the notion that you can create a logically consistent, and useful, single aggregate measure of the total physical capital stock in the economy. The issue of how to measure the aggregate capital stock became one of the battlegrounds of the Cambridge Capital debates of the 1960s, with Joan Robinson and the Cambridge UK critics of neoclassical capital theory lined up against Paul Samuelson and the MIT (Cambridge Mass.) defenders of neoclassical theory. Keynes does not go into fine detail in the *General Theory*, rather he asserts that he doesn't believe that the concepts of aggregate real output and of an aggregate real capital stock make any logical sense, and he moves on to look for other ways of measuring aggregate economic activity¹.

¹ In his review of Keynes' *Treatise on Money*, Hayek criticizes the assumption which Keynes made in that book to the effect that there exists a meaningful aggregate measure of capital: "The question is whether any increase of the value of the existing capital is to be considered [as an addition to the capital stock] – in this case, of course, such an addition could be brought about without any new production of such goods – or whether only additions to the physical quantities of capital goods are to be counted as such an addition – a method of computation which clearly becomes impossible when the old capital goods are not replaced by goods of exactly the same kind, but when a transition to more capitalistic methods brings it about that other goods are produced in place of those used up in production." F. A. von Hayek (1931): "Reflections on the Pure Theory of Money of Mr. J. M. Keynes" *Economica* 33, August, 270-295, quote from pg. 281. Keynes replied in his response to Hayek's review that "...Dr. Hayek also criticizes the conception of "quantity of capital" as being invalid on the grounds that the different types of specific goods constituting capital are not always identical, and when non-identical are non-commensurable. But this is simply the same problem as that of the conception of "price-level" and the associated conception of

We typically measure aggregate real output by taking aggregate nominal output (or income) and deflating by a price index. Keynes considers this approach but concludes that

...the well-known, but unavoidable, element of vagueness which admittedly attends the concept of the general price-level makes this term very unsatisfactory for the purposes of a causal analysis, which ought to be exact.

Again, it is not that Keynes rejects the idea of a price index for aggregate output, rather that he doubts that it is possible to construct one which will work well enough for his purposes². The *General Theory* was intended to be a guide for practical policy making, so it was important to Keynes that the theory set out in it only use concepts which were logical, measurable, and which could be measured precisely enough for policy makers to be able to rely on them. It was not that he rejected the construction of aggregate price indices completely: they could be useful in their place, but their place wasn't policy making:

The fact that two incommensurable collections of miscellaneous objects cannot in themselves provide the material for a quantitative analysis need not, of course, prevent us from making approximate statistical comparisons, depending on some broad element of judgment rather than of strict calculation, which may possess significance and validity within certain limits. But the proper place for such things as net real output and the general level of prices lies within the field of historical and statistical description, and their purpose should be to satisfy historical or social curiosity, a purpose for which perfect precision — such as our causal analysis requires, whether or not our knowledge of the actual values of the relevant quantities is complete or exact — is neither usual nor necessary.

And he goes on to another well-known Keynesism:

To say that net output to-day is greater, but the price-level lower, than ten years ago or one year ago, is a proposition of a similar character to the statement that Queen Victoria was a better queen but not a happier woman than Queen Elizabeth — a proposition not without meaning and not without interest, but unsuitable as material for the differential calculus. Our precision will be a mock precision if we try to use such partly vague and non-quantitative concepts as the basis of a quantitative analysis.

real-wages when the complex of goods refers to changes in its make-up. This I have discussed at great length in Book II of my *Treatise*, and it arises of course in all types of monetary theory alike." J. M. Keynes (1931): "The Pure Theory of Money: A Reply to Dr. Hayek" *Economica* 34, November, 387-397, quote from pg. 397. Apparently by the time he wrote the *General Theory*, Keynes had changed his mind on these matters.

² One of Keynes' earliest major pieces of work was his 1909 Adam Smith Prize essay on Index Numbers. Reproduced in *The Collected Writings of John Maynard Keynes*, Vol. XI "Economic Articles and Correspondence, Academic"

In addition to arguing that things like aggregate real output, real capital or a general price index are unmeasurable, or at least, not measurable with sufficient accuracy for his purposes, he argued that they were not crucial to the development of a model of the macroeconomy:

Nevertheless these difficulties are rightly regarded as “conundrums.” They are “purely theoretical” in the sense that they never perplex, or indeed enter in any way into, business decisions and have no relevance to the causal sequence of economic events, which are clear-cut and determinate in spite of the quantitative indeterminacy of these concepts. It is natural, therefore, to conclude that they not only lack precision but are unnecessary. Obviously our quantitative analysis must be expressed without using any quantitatively vague expressions. And, indeed, as soon as one makes the attempt, it becomes clear, as I hope to show, that one can get on much better without them.

The reference to these real measures not entering into business decisions is significant, since Keynes’ aim is to build a macro model which has firm micro foundations. Arguably, though, he uses his focus on the Marshallian short run as a way of avoiding certain logical problems. In the short run the firm’s capital stock is fixed, so the business decision comes down to deciding what quantity of variable inputs to apply to the fixed capital stock – how hard to work the capital stock. This means that he can treat each industry’s the capital stock as given, and although much of the *General Theory* deals with investment decisions the focus on the short run means that Keynes doesn’t have to deal with the implications of investment spending for changes in the total quantity of capital equipment in the economy, nor with how to incorporate that increased quantity into a model which is built on and expressed in terms of aggregate measures. He recognizes that aggregates can make sense at the level of the industry: “In the case of an individual firm or industry producing a homogeneous product we can speak legitimately, if we wish, of increases or decreases of output.” where the industry output is an aggregate of the homogeneous output of the individual firms in the industry, but:

...when we are aggregating the activities of all firms, we cannot speak accurately except in terms of quantities of employment applied to a given equipment. The concepts of output as a whole and its price-level are not required in this context, since we have no need of an absolute measure of current aggregate output, such as would enable us to compare its amount with the amount which would result from the association of a different capital equipment with a different quantity of employment.

Note that when he speaks of current aggregate output here he means real output – output defined as an aggregate of the disparate physical products made by the whole range of industries in the

economy. All of this leads him to two decisions which it is important to understand if we are to follow the exposition of the *General Theory*. One is that, while Keynes often refers to real measures – labour’s real income, for example – for modeling purposes he works in nominal terms. Again, this is not because he rejects the real measures as logical constructs at the microeconomic level, it is because he rejects the idea that the micro concepts, sensible though they are, can be aggregated up to sensible macro concepts. The second is that he will work in the short run and in the absence of a sensible measure of aggregate output take the level of employment as his measure of aggregate real economic activity. Holding the capital stock constant, if the level of employment increases then the aggregate level of real output being produced must have increased, even if we cannot measure that increased output sensibly. In the short run, with the capital stock held fixed, changes in the level of employment will be associated with changes in the level of real output in the same direction, even though there may not be a stable proportional relation between them – i.e. even though we may not be able to write a sensible aggregate production function with aggregate real output as a mathematically stable function of aggregate employment. The level of aggregate employment is an index of the level of aggregate real output, albeit a noisy one.

Wage Units and Labour Units

This, of course, raises the question of whether there exists a sensible aggregate measure of employment or labour, if there doesn’t exist either a sensible measure of aggregate capital or of aggregate real output. The short answer is that Keynes thinks that there does exist such an aggregate.

He starts from the assumption that labour is paid the value of its marginal product, and interprets this to mean that when different types of labour are paid different amounts per hour the differences in their wages can be taken to reflect differences in the amount of what we would now term effective, or efficiency adjusted, labour which they supply per hour. In other words, he assumes that all types of labour can be defined in terms of the number of units of standard labour to which they are equivalent. This clearly requires that we define a concept of standard labour

units, but this is not so unusual an assumption – when we are looking at the effects of changes in the age distribution of the labour force, for example, it is common to characterize the different ages of labour in terms of their relative productivities, and then define a labour aggregate which is a weighted sum of the different age groups, where the weights reflect their relative productivities. We might, for example, estimate a Cobb-Douglas production function with specific age groups of labour as inputs, and use the relative exponents on each age group as the basis for the relative weight to be applied to each member of the labour force, with one age group – the middle labour force age group, for example – serving as the reference case, in the sense that each individual in that age group would be given a weight of 1. Younger, less experienced workers will typically be weighted as less than one full labour equivalent. When we take that approach, it is possible that as the age distribution of our labour force changes, becoming more experienced, the number of equivalent, or effective, units of labour could change even though the number of discrete human beings in the labour force didn't. We do the same thing, implicitly, when we estimate the difference in productivity among workers who have accumulated different amounts of specific human capital in the form of specialized education or training.

Adam Smith adopted a similar device in *Wealth of Nations* when he was writing about making comparisons between income across place and time. Smith assumed that there existed a certain base type of labour – in his case probably unskilled agricultural labour – which yielded the same amount of disutility per hour in all times and at all places. He then assumed that the wage paid to that type of labour, regardless of the currency units in which it was paid or the number of those units which the base-case worker received, represented the supply price of that labour – i.e. that it just compensated for the disutility of an hour of the base-case labour. This then meant that you could compare prices and incomes across countries by calculating the number of base case labour hours which it would take to buy different commodities. In essence Smith was proposing a purchasing power parity approach to comparing currencies, but taking base labour as the common unit rather than looking at the cost of a common bundle of goods³. The difference

³ Actually, Smith comes even closer to the modern purchasing power parity approach to calculating exchange rates when he says that "Equal quantities of labour will at distant times be purchased more nearly with equal quantities of corn, the subsistence of the labourer, than with equal quantities of gold and silver, or perhaps of any other commodity. Equal quantities of corn, therefore, will, at distant times, be more nearly of the same real value, or

between Smith's and Keynes' approaches is that Smith assumes that the wage of a unit of basic labour is on the labour supply curve, while Keynes assumes that it is not on the supply curve but rather that it is on the labour demand curve. Thus, even though Keynes does not believe that labour markets clear, he is Marshallian enough to assume that firms hire along the VMPL curve and that differences in the hourly wage of very different types of labour reflect differences in the productivity of those types of labour. He therefore takes the wage of a unit of the basic labour type – perhaps unskilled manual labour – as his basis for weighting other types of labour and says that:

.....in so far as different grades and kinds of labour and salaried assistance enjoy a more or less fixed relative remuneration, the quantity of employment can be sufficiently defined for our purpose by taking an hour's employment of ordinary labour as our unit and weighting an hour's employment of special labour in proportion to its remuneration; i.e. an hour of special labour remunerated at double ordinary rates will count as two units.

He then defines an hour of basic labour as what he calls the labour-unit, and defines the hourly money wage of this type of labour as the wage-unit. "Thus, if E is the wages (and salaries) bill, W the wage-unit, and N the quantity of employment, $E = N \cdot W$."

It is important to notice what is being calculated here. Keynes is taking the wage-bill in money terms, E , as given and is taking W , the wage-unit, the wage of an hour of basic labour, as given. Then when he writes $E = N \cdot W$ he is actually calculating $N = E/W$. This means that, as in our age-distribution example above, there could be more hours of N than there are actual hours of unweighted labour reported by firms. Keynes goes on, in a footnote, to say:

If X stands for any quantity measured in terms of money, it will often be convenient to write X_w for the same quantity measured in terms of the wage-unit.

So if we let Y be national income measured in money terms, we would write Y_w for national income measured in wage units. But $Y_w = Y/W$, so Y_w represents the number of units of "W" that Y contains, and since W is the wage of an hour of basic labour, Y_w is the number of hours of basic labour which could be bought with a nominal national income of Y . This means that when Keynes talks about measuring national income, say, in wage units, he is talking about measuring

enable the possessor to purchase or command more nearly the same quantity of the labour of other people."
Wealth of Nations Book 1 Ch V.

it in terms of the number of hours of basic labour which it could buy. This almost sounds as if Keynes is adopting a labour theory of value, but he isn't – it is a matter of tackling the problem of measuring macroeconomic aggregates. As Keynes puts it:

It is my belief that much unnecessary perplexity can be avoided if we limit ourselves strictly to the two units, money and labour, when we are dealing with the behaviour of the economic system as a whole; reserving the use of units of particular outputs and equipments to the occasions when we are analysing the output of individual firms or industries in isolation; and the use of vague concepts, such as the quantity of output as a whole, the quantity of capital equipment as a whole and the general level of prices, to the occasions when we are attempting some historical comparison which is within certain (perhaps fairly wide) limits avowedly unprecise and approximate.

Keynes may believe that he's avoiding unnecessary perplexity, but unless we keep in mind what the wage-unit means, bits of the *General Theory* become very heavy slogging.

Further:

It follows that we shall measure changes in current output by reference to the number of hours of labour paid for (whether to satisfy consumers or to produce fresh capital equipment) on the existing capital equipment, hours of skilled labour being weighted in proportion to their remuneration. We have no need of a quantitative comparison between this output and the output which would result from associating a different set of workers with a different capital equipment.

In the context of this excerpt, it is worth remembering what we do when we convert nominal to real GDP today. We take one of the measures of nominal national income and divide it by a price index. If we are deflating by the CPI, the price index is a measure of the price of a basket of consumption goods (relative to their base year price) so real GDP calculated using the CPI is a measure of the quantity of consumer goods which could be bought with national income. We use the price index of a bundle of consumer goods to convert a nominal variable to a real (in terms of purchasing power) one; Keynes, who didn't have much confidence in the precision of price indices, converts to purchasing power in terms of basic labour.

It's also worth noting that a lot of the time we do not differentiate explicitly between different sub-types of labour when we are doing macroeconomic analysis. When we calculate nominal GDP per capita we generally do not standardize the labour units, mainly because, as a reading of the *General Theory* helps convince us, doing so would not necessarily make the story any clearer

(especially when we are writing newspaper articles). It is also not unusual to take the total wage bill and divide it by the number of workers to get an average wage figure – in terms of $E = N \cdot W$, we are collecting data on E and N and calculating $W = E/N$.

Keynesian and Marshallian Supply:

Keynes goes on, in the last, brief, section of Chapter 4, to write:

It is easily shown that the conditions of supply, such as are usually expressed in terms of the supply curve, and the elasticity of supply relating output to price, can be handled in terms of our two chosen units by means of the aggregate supply function, without reference to quantities of output, whether we are concerned with a particular firm or industry or with economic activity as a whole.

What he is doing in this section is relating his aggregate supply price notions, his definitions of key concepts and his approach to aggregation to notions which are familiar from Marshallian microeconomics. He defines $Z_r = \phi_r(N_r)$ as the aggregate supply function for industry r (which is assumed to be producing a homogeneous output). He is deliberately using the same notation here, apart from the addition of an industry subscript, as he did when he defined $Z = \phi(N)$ to be the aggregate supply function, where Z was the aggregate supply price, so when we are working at the industry level Z_r is the proceeds net of user cost which would be just sufficient to induce the industry to employ N_r hours of labour. He wants to relate his concepts to the Marshallian supply price which, since double counting isn't an issue when we are talking about a single industry producing a single, homogeneous output, has to be high enough to cover the user cost in addition to the elements in Z_r . Thus the Marshallian supply price has to be just sufficient to cover $\phi_r(N_r) + U_r(N_r)$ and, since it is the Marshallian industry-level supply price, it has to be calculated on a per unit of output basis, something which Keynes does not believe can be done at the aggregate level but with which he has no problem when we are talking about the supply price of a single, homogeneous product. He defines the (physical) output of industry r as O_r , and is quite happy to write a production function for O_r , $O_r = \psi_r(N_r)$. Note that the production function, $\psi_r(N_r)$, is written as a function of N_r alone, not N_r and K_r because Keynes is working in the short run and so holding capital fixed (and because he would presumably balk defining an aggregate

measure of capital even at the level of the individual firm). Then, letting p stand for the Marshallian price, he writes

$$p = \frac{Z_r + U_r(N_r)}{O_r} = \frac{\varphi_r(N_r) + U_r(N_r)}{\psi_r(N_r)}$$

where $U_r(N_r)$ is the user cost of employing N_r units of labour in industry r in the production of good r . We could, if we wanted, write $p = p_r(N_r)$. We can calculate Z_r for each industry (and note that)

$$Z_r = p\psi_r(N_r) - U_r(N_r)$$

and aggregate across industries to get Z , since Z is in money terms, and we can aggregate across N_r given the way Keynes has defined N , but we can't sensibly aggregate O_r values. Keynes' point here is to show that there is a logical link between his notion of the aggregate supply price and the Marshallian supply price for an individual commodity.

Interestingly, in the first printing of *General Theory*, Keynes wrote:

It is easily shown that the conditions of supply, such as are usually expressed in terms of the supply curve, and the elasticity of supply relating output to price, can be handled in terms of our two chosen units by means of the aggregate supply function, without reference to quantities of output, whether we are concerned with a particular firm or industry or with economic activity as a whole. For the aggregate supply function for a given firm (and similarly for a given industry or for industry as a whole) is given by

$$Z_r = \phi_r(N_r),$$

where Z_r is the return the expectation of which will induce a level of employment N_r . If, therefore, the relation between employment and output is such that an employment N_r results in an output O_r , where $O_r = \psi_r(N_r)$, it follows that

$$p = Z_r/O_r = \phi_r(N_r)/\psi_r(N_r)$$

is the ordinary supply curve.

A former student of his, Hugh Townshend, who was working at the British Post Office⁴ wrote to Keynes to point out that he was at the very least using notation in a confusing manner, since in this original version he referred to Z_r as return, not proceeds, and where the logic of the equations meant that this definition of Z_r had to include user cost whereas Z , aggregate proceeds, did not. That was confusing to the reader, to put it mildly. Keynes changed the wording of this section in the second printing of the *General Theory* to reflect Townshend's comments.

Chapter 5: Expectations; Introduction:

Chapter 5 is another short chapter focused on defining terms, but as in the case of Chapter 4 is not one which can be skipped, given the key role played by expectations in the structure of the *General Theory*.

Actually, one of the most important observations which Keynes makes about expectations in this context is not made in the *General Theory* but in the notes he prepared for his 1937 lectures⁵, where he says

Entrepreneurs have to endeavour to forecast demand. They do not, as a rule, make wildly wrong forecasts of the equilibrium position. But, as the matter is very complex, they do not get it just right; and they endeavour to approximate the true position by a process of trial and error. Contracting where they find they are overshooting their market, expanding where the opposite occurs.

In Chapter 5 Keynes defines two sorts of expectations, long term expectations and short term expectations. Long term expectations get a chapter of their own later on, so they are treated very briefly in this chapter. Basically, short term expectations are expectations about the state of the market in the short run, and since this is the Marshallian short run that means a period during which the firm's capital stock is fixed. Short term expectations deal with issues such as what

⁴ Their correspondence is included in Volume XXIX of the *Collected Writings of John Maynard Keynes*.

⁵ Published in *The Collected Writings of John Maynard Keynes, Vol XIV, The General Theory and After, Part II, Defence and Development* pp 179-183

prices or demand will be in the near future, and determine the amount of output which the firm will produce using its existing, unchanged capital stock. Long term expectations, on the other hand, come into play when the firm is deciding how much capital to invest in to be used in producing output at some point in the future. A firm's current capital stock, then, was determined by the long run expectations it had held at some point in the past, while the intensity with which it runs that capital stock today is determined by its short run expectations about the state of the market in the near future.

In discussing short run expectations, Keynes says:

The *actually realised* results of the production and sale of output will only be relevant to employment in so far as they cause a modification of subsequent expectations.

The best way to make sense of this statement is to think in terms of a commodity for which there is a definite period of production, in the sense of a well-defined interval between when the firm begins a production run and when it places the output on the market. The clearest example of this sort of production process is probably in the case of agricultural products for which there is a definite planting period and, some known time later, a definite harvesting period. This is the example which underlies what is known as the cobweb model, one of the most basic of dynamic models in economics.

In a cobweb model, demand in the current period (period t) depends on the current price of the product and a set of other variables – income, say – giving a demand function of the form

$$D_t = \alpha_0 - \alpha_1 P_t + \alpha_2 Y_t$$

While supply depends on some actual current variables, which we will call X_t but because of the time structure of this production process Supply today depends on the expectation which was formed yesterday about what the price of the output would be today: $E_{t-1}P_t$. There is no requirement that $E_{t-1}P_t = P_t$, although there is nothing to forbid it. The supply curve, then, is:

$$S_t = \beta_0 + \beta_1 E_{t-1}P_t + \beta X_t$$

In the simplest cobweb model, the realized price in period t is determined by the intersection of demand and supply at time t , but while the quantity demanded in period t depends on the current price, supply in period t is exogenous, at a level determined by $E_{t-1}P_t$. The dynamics of the market, then, depend on how expectations are formed. In the case of perfectly myopic expectations, $E_{t-1}P_t = P_{t-1}$ and we get classic cobweb dynamics – when P_{t-1} is high, firms set large production runs in motion, resulting in the period t supply being large and P_t being low, which in turn leads firms to cut back on their production runs, which will drive the price up in period $t+1$. Stability of the dynamic process then depends on the relative elasticities of demand and supply⁶. Keynes' remarks about the role of actually realized values basically come down to saying that if we write the supply function as

$$S_t = \beta_0 + \beta_1 P_t + \beta X_t$$

We are mis-specifying the relationship. In the simple cobweb example this mis-specification could lead to our believing we have found a negatively sloped supply curve.

While the simple cobweb is the clearest example of what Keynes means when he says that short term output depends on expectations about certain variables rather than on the actually realized values of those variables, as the quote about firms generally getting the expectations right indicates he doesn't have quite such a stark case in mind. Still, he is concerned with the way changes in expectations about some variables will lead to changes in the evolution of the actual values of other variables over time – i.e. about macroeconomic dynamics. As he puts it, “Now, in general a *change* in expectations (whether short-term or long-term) will only produce its full effect on employment over a considerable period.”

Dynamic Economic Modelling:

Much of Chapter 5, in fact, deals with economic dynamics. To see this, define Y as the dependent variable, X as the explanatory variable and the relation between them as $Y(X)$. Then consider the following definition:

⁶ A version of the cobweb model is often used to explain the dynamic behaviours of incomes of certain types of labour. When earnings in a particular field are high, students enroll in educational programs aimed at allowing them to work in those fields, which means that, when their cohort graduates they are likely to find competition pushing their incomes down below what they were expecting.

If we suppose a state of expectation to continue for a sufficient length of time for the effect on employment to have worked itself out so completely that there is, broadly speaking, no piece of employment going on which would not have taken place if the new state of expectation had always existed, the steady level of employment thus attained may be called the long-period employment corresponding to that state of expectation. It follows that, although expectation may change so frequently that the actual level of employment has never had time to reach the long-period employment corresponding to the existing state of expectation, nevertheless every state of expectation has its definite corresponding level of long-period employment.

This is a description of what we would now refer to as the long run relation⁷ between Y and X , which we can write as $Y^*(X)$. Keynes is dealing here with the way Y changes in response to actual and expected changes in X . In comparative static analysis, when X changes, Y will jump immediately to its new long run level. In dynamic analysis, which Keynes is clearly treating here as the more realistic case, changes in X are associated with immediate changes in Y^* , by definition, but Y itself adjusts with a lag, following a trajectory which may converge smoothly on the new value of Y^* as in Figure 1 below or which may involve overshooting as in Figure 2. In each of these figures we have plotted Y on the vertical axis and time on the horizontal, and we assume that X changes in a manner which causes Y^* to increase from $Y^*(X_1)$ to $Y^*(X_2)$. This corresponds to, although as we shall see in a moment is not precisely the same as, what Keynes calls a change in long-term expectations.

Keynes goes on to say that:

An uninterrupted process of transition, such as the above, to a new long-period position can be complicated in detail. But the actual course of events is more complicated still. For the state of expectation is liable to constant change, a new expectation being superimposed long before the previous change has fully worked itself out; so that the economic machine is occupied at any given time with a number of overlapping activities, the existence of which is due to various past states of expectation.

which would correspond to a figure like Figure 3 below where each change in the value of X would lead to a different Y^* value. If the adjustment of Y to Y^* were instantaneous, as in the comparative statics case, there would be no problem, but on the assumption that Y adjusts towards Y^* with a lag, we will observe changes in Y even when there are no changes in X , and on the usual assumptions about the differential equations which characterize dynamic economic models, the slope of the trajectory for Y on our time series diagram will vary depending on how

⁷ And sometimes as the equilibrium relation between Y and X .

close Y is to the current Y^* value. It is worth remembering, too, that in general we do not know the $Y^*(X)$ function but have to estimate it, so in the case of Figure 3 we would observe three discrete values of X and a continuous trajectory of values of Y , (so the time series plot of what we know would be as in Figure 4 below, where the time intervals during which X takes on specific values are indicated) and have to use that information to estimate both the $Y^*(X)$ function and the equation explaining the dynamic adjustment process which Y follows⁸. In econometric terms⁹, this typically involves estimating either a general dynamic equation of the form

$$(1) \quad Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_t + \alpha_3 X_{t-1} + \varepsilon_t$$

where the long run relation between Y and X is found on the assumption that X has been constant long enough for Y to have settled down, giving $X_t = X_{t-1} = X$ and $Y_t = Y_{t-1} = Y$ and

$$(2) \quad Y = \frac{\alpha_0}{[1 - \alpha_1]} + \frac{[\alpha_2 + \alpha_3]}{[1 - \alpha_1]} X$$

or the counterpart of the general dynamic form, an error correction model (ECM), which can be derived from the general dynamic form (1) first by subtracting Y_{t-1} from both sides, giving

$$(3) \quad Y_t - Y_{t-1} = \alpha_0 + \alpha_1 Y_{t-1} - Y_{t-1} + \alpha_2 X_t + \alpha_3 X_{t-1} + \varepsilon_t \\ = \alpha_0 + [\alpha_1 - 1] Y_{t-1} + \alpha_2 X_t + \alpha_3 X_{t-1} + \varepsilon_t$$

and then adding and subtracting $\alpha_2 X_{t-1}$ on the right hand side, giving

$$(4) \quad Y_t - Y_{t-1} = \alpha_0 + [\alpha_1 - 1] Y_{t-1} + \alpha_2 X_t - \alpha_2 X_{t-1} + \alpha_3 X_{t-1} + \alpha_2 X_{t-1} + \varepsilon_t$$

⁸ We need to estimate both the long run $Y^*(X)$ relation and the equation of motion for Y since we need to be able to separate the elements of the time series plot of Y due to each. Going after one part alone will be econometrically less efficient and may well lead to bias in our estimated results.

⁹ It is well known that Keynes was very skeptical about econometrics. On the other hand, many of the issues of econometric theory and methodology which he raised in his debate with Tinbergen have been tackled since their time.

$$= \alpha_0 + [\alpha_1 - 1]Y_{t-1} + \alpha_2 [X_t - X_{t-1}] + [\alpha_3 + \alpha_2] X_{t-1} + \varepsilon_t$$

and then, using Δ to represent first differences, giving

$$(5) \quad \Delta Y_t = \alpha_0 + \alpha_2 \Delta X_t + [\alpha_1 - 1]Y_{t-1} + [\alpha_3 + \alpha_2] X_{t-1} + \varepsilon_t$$

In the notation of equation (5), the Δ terms represent short run effects (so α_2 is the short run effect of a change in X on Y) and we find the long run version of the relation by setting the Δ terms to zero (reflecting the assumption that X , and therefore Y , have settled down in the long run) and re-write (5) so that we have the same form as in equation (2) above.

The reason we said above that our figures do not correspond exactly to the cases which Keynes is discussing is that, as they are drawn, the dynamic adjustment process by which the actual value of Y moves from its old long run value at $Y^*(X_1)$ to its new long run value does not begin until the change in X has actually occurred. In fact, Keynes assumes that Y begins to change when entrepreneurs' expectations about X change. Looking at Figure 5 below, we have reproduced Figure 1 but have also added a trajectory along which the actual value of Y begins to change as soon as entrepreneurs become convinced that X will change from X_1 to X_2 , so that in our notation it is still the case that the long run values of Y can be written $Y^*(X_1)$ and $Y^*(X_2)$, but now the actual values of Y must be written $Y(EX_1)$ and $Y(EX_2)$. The same could be done for the case shown in Figure 3. The transition could begin as soon as entrepreneurs are confident that there is going to be a long run change in X , not just a short run blip, to which they might respond with small changes in Y but which, being a blip, will quickly be undone, with X returning to its original value (this is the sort of effect which could show up in the short run part of the ECM). The actual process of transitioning from $Y^*(X_1)$ to $Y^*(X_2)$ (which are the true long run values of Y conditional on X : when we write $Y(EX)$ we are referring to the actual, observed values of Y) will then depend on the balance of two factors – the cost of adjustment and the cost of being out of the long run relation at both ends of the process. Here again we need to consider Keynes warning about not simply replacing EX by X_t in the expression for Y : if we estimate Y as though entrepreneurs are responding to the actual, rather than the expected, value

of X, then depending on which part of the time series of actual values of Y happens to be in our sample, we could well conclude that the relation between Y and X is dynamically unstable, since we would estimate it as Y moving away from the actual $Y^*(X)$ instead of as Y moving towards the value which $Y^*(X)$ is expected to move to at a certain instant (the timing of which is also a matter of expectation) in the future.

As we noted above, expectations, and in particular long-term expectations, play a key role in the development of the *General Theory*. This raises the question of how Keynes conceived of expectations as being formed. His references (to which we shall come later) to animal spirits have often been taken to suggest that there is something non-structural about Keynesian expectations; that they are in a sense irrational. Keynes does not exclude the possibility of irrational exuberance, again as we shall see later, but the expectations being discussed in Chapter 5 are much more rational.

The Formation of Expectations:

Again, since long-term expectations are going to be given a chapter of their own, most of what Keynes has to say here about how expectations are formed relates to short-term expectations. His key take is that actual production decisions depend on expectations in two ways – past long-term expectations are embodied in the current capital stock, and current short-term expectations determine how hard that capital stock is to be worked in the short run. Short term expectations can fluctuate, but:

It would be too complicated to work out the expectations *de novo* whenever a productive process was being started; and it would, moreover, be a waste of time since a large part of the circumstances usually continue substantially unchanged from one day to the next. Accordingly it is sensible for producers to base their expectations on the assumption that the most recently realised results will continue, except in so far as there are definite reasons for expecting a change.

Thus unless there is reason to believe a fundamental change is underway, the most sensible rule about short term expectations is $E_t X_{t+1} = X_t$, where t now represents a very short calendar

interval¹⁰. Note that this only makes sense if the expectations formation mechanism is such that, as Keynes noted in his 1937 lecture, entrepreneurs generally get it pretty much right. Short term expectations can be tweaked, or adapted, on the basis of actually observed outcomes, so we can justify using the adaptive process:

$$(6) \quad E_t X_{t+1} = E_{t-1} X_t + \delta [X_t - E_{t-1} X_t]$$

where δ will generally be a small positive value. Here if the expectation formed in period $t-1$ about period t 's X proves to have been correct, expectations will remain unchanged into the future, and since $E_{t-1} X_t$ turned out to be equal to X_t , we are back to $E_t X_{t+1} = X_t$.

Long term expectations, on the other hand, cannot be tweaked on the basis of realized short term outcomes, by their very nature. Long term expectations, therefore, must depend on the entrepreneur's understanding of the markets within which he is operating, along with the discipline imposed by the fact that any entrepreneur who consistently makes mistakes in forming long-term expectations will find himself continually operating with non-optimal stocks of fixed capital and will wind up going bankrupt. Keynes' assumption that entrepreneurs basically get their expectations right, and his argument that they base their expectations on recent results unless there are definite reasons for believing that conditions will change means that they are forming their long term expectations on the basis of a model of how the bits of the economy which affect them work, and getting those expectations basically right, which can be read as meaning that Keynes is assuming that long-term expectations are essentially what are now referred to as rational expectations.

This argument might sound odd, since the Rational Expectations revolution of the 1970s and 1980s is generally taken as part of the counter-revolution against the Keynesian revolution, but it must be understood that the non-Keynesian implications of many of the rational expectations models arose not from the nature of the expectations process but from the fact that those models assumed a classical, market clearing world. The basic premise, that expectations are formed rationally by taking what is essentially a correct model of the economy and using all currently available information to project future values of variables which enter current behavioural

¹⁰ Clearly the cobweb model to which we referred earlier doesn't fall into the category of short-period models as Keynes is thinking of them here.

equations as expectations rather than as current values seems fully consistent with the approach Keynes takes in the *General Theory*.¹¹

¹¹ From the applied econometrics perspective, the argument that entrepreneurs get expectations correct, within a relatively small margin of error suggests that we could treat realized future values of expectational variables as being equal to the true expectations plus an error term which has a zero mean, and employ econometric techniques suitable for the case where we have a variable which is measured with error. This still leaves the empirical issue of the timing of the response of current Y to expected X.

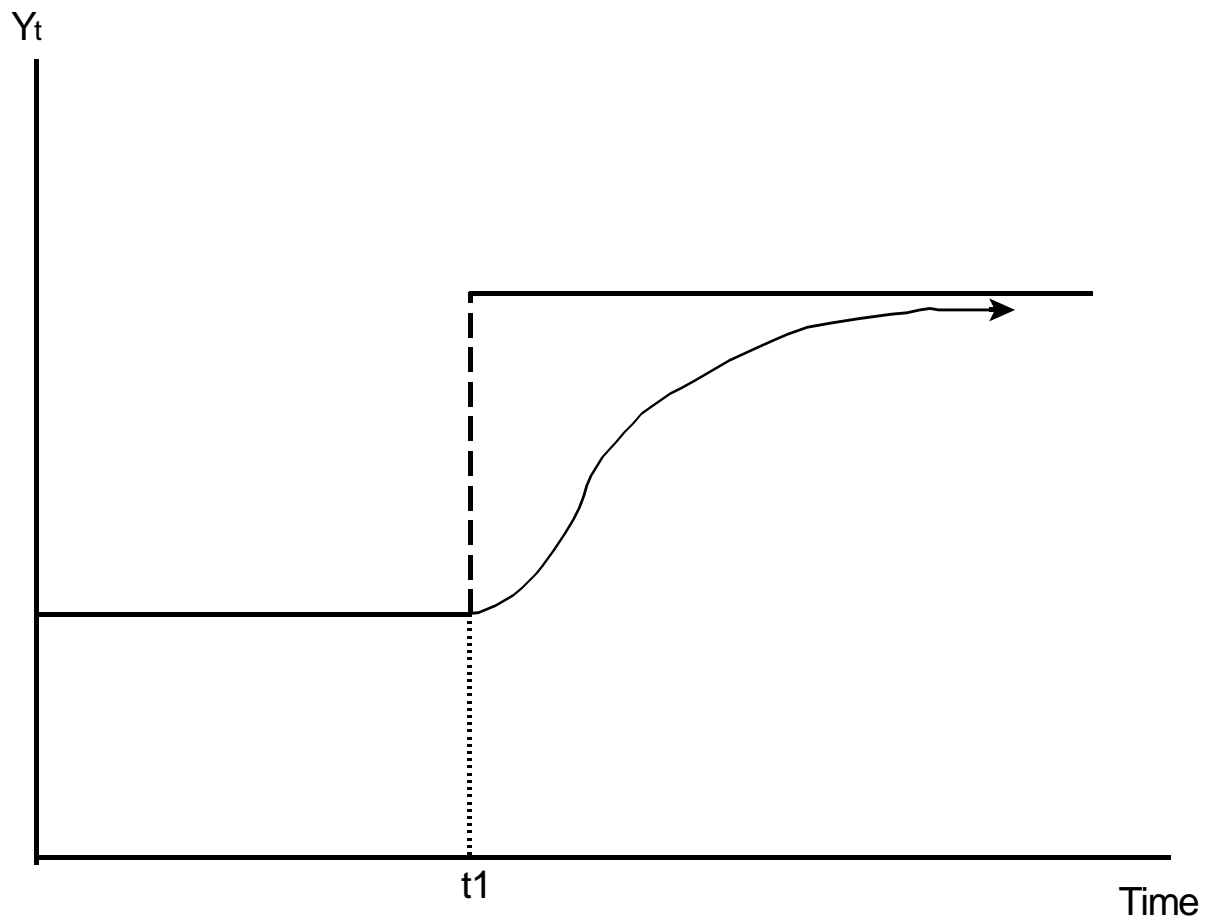


Figure 1

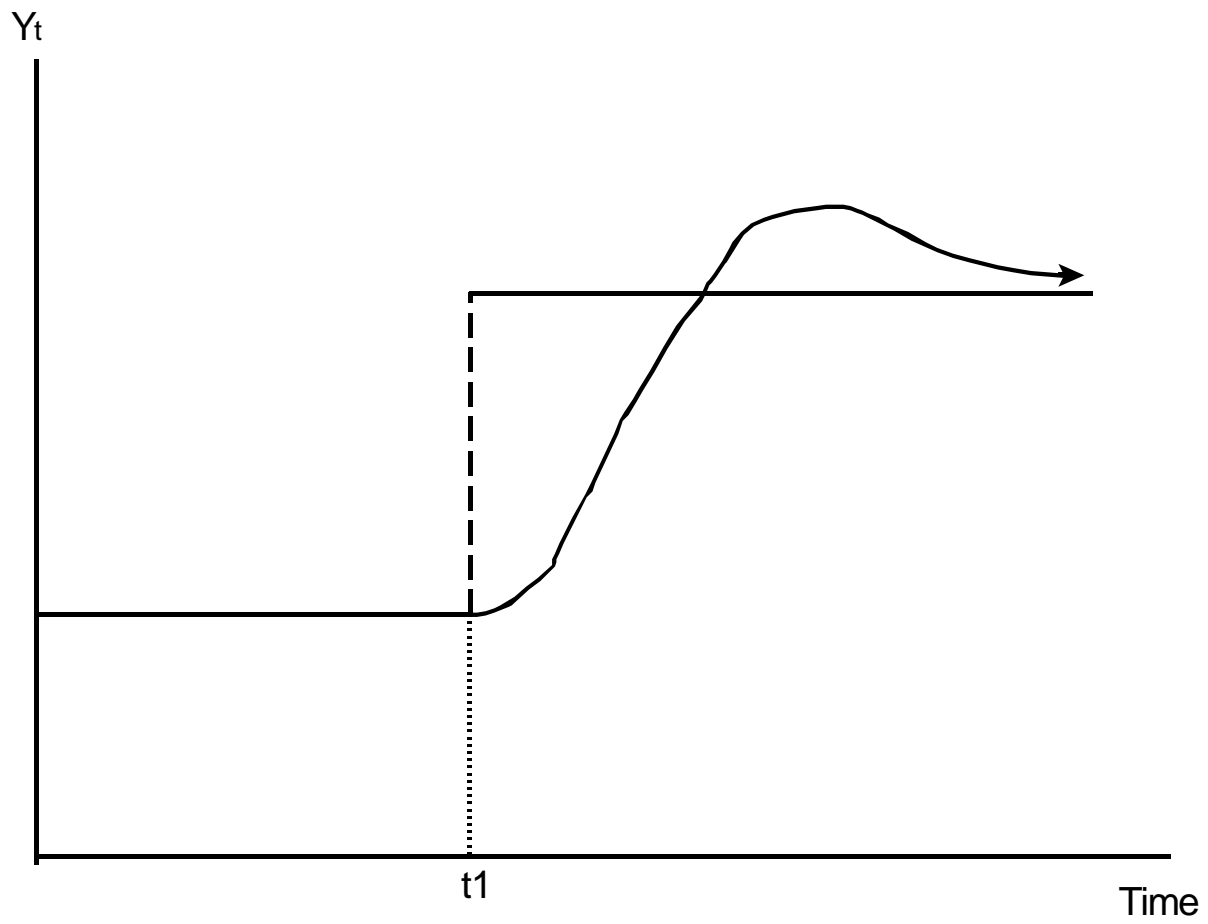


Figure 2

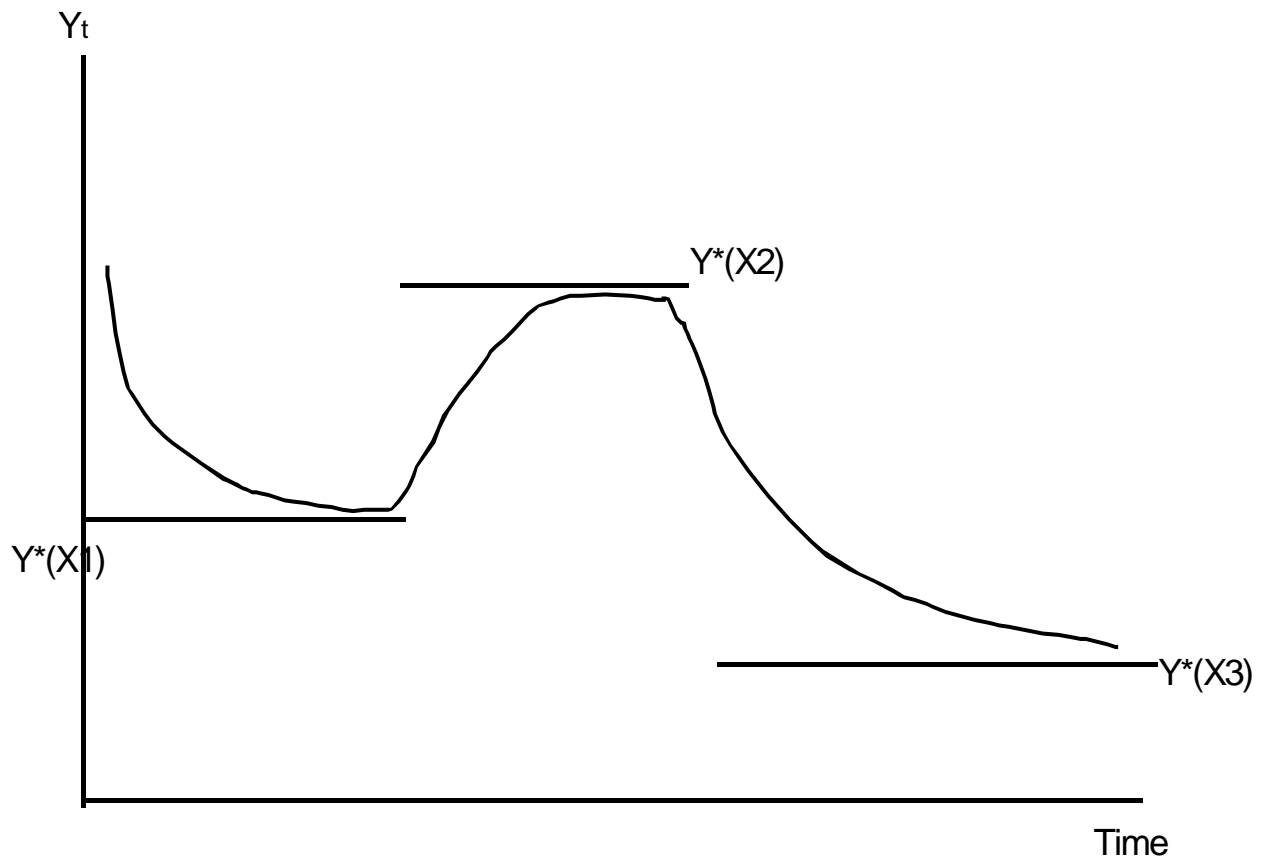


Figure 3

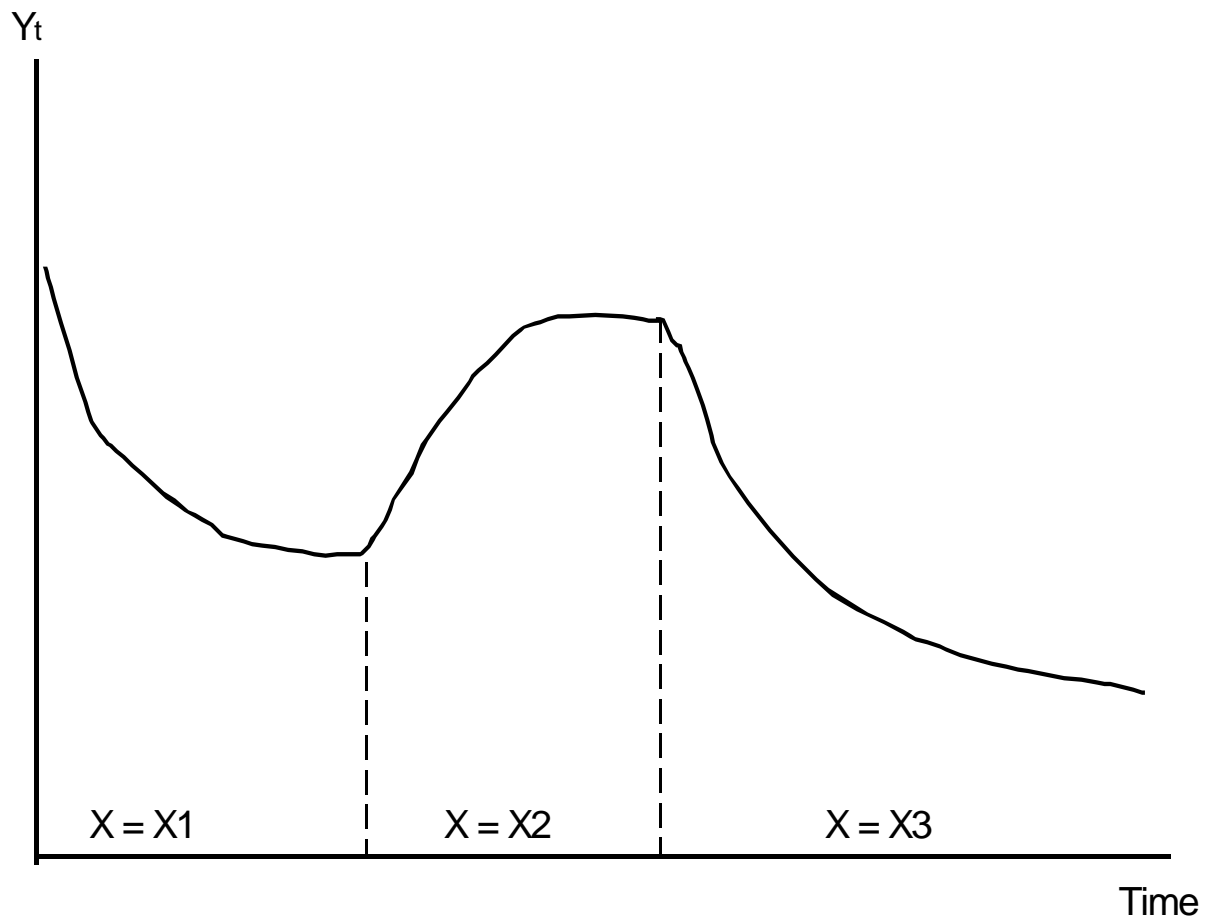


Figure 4

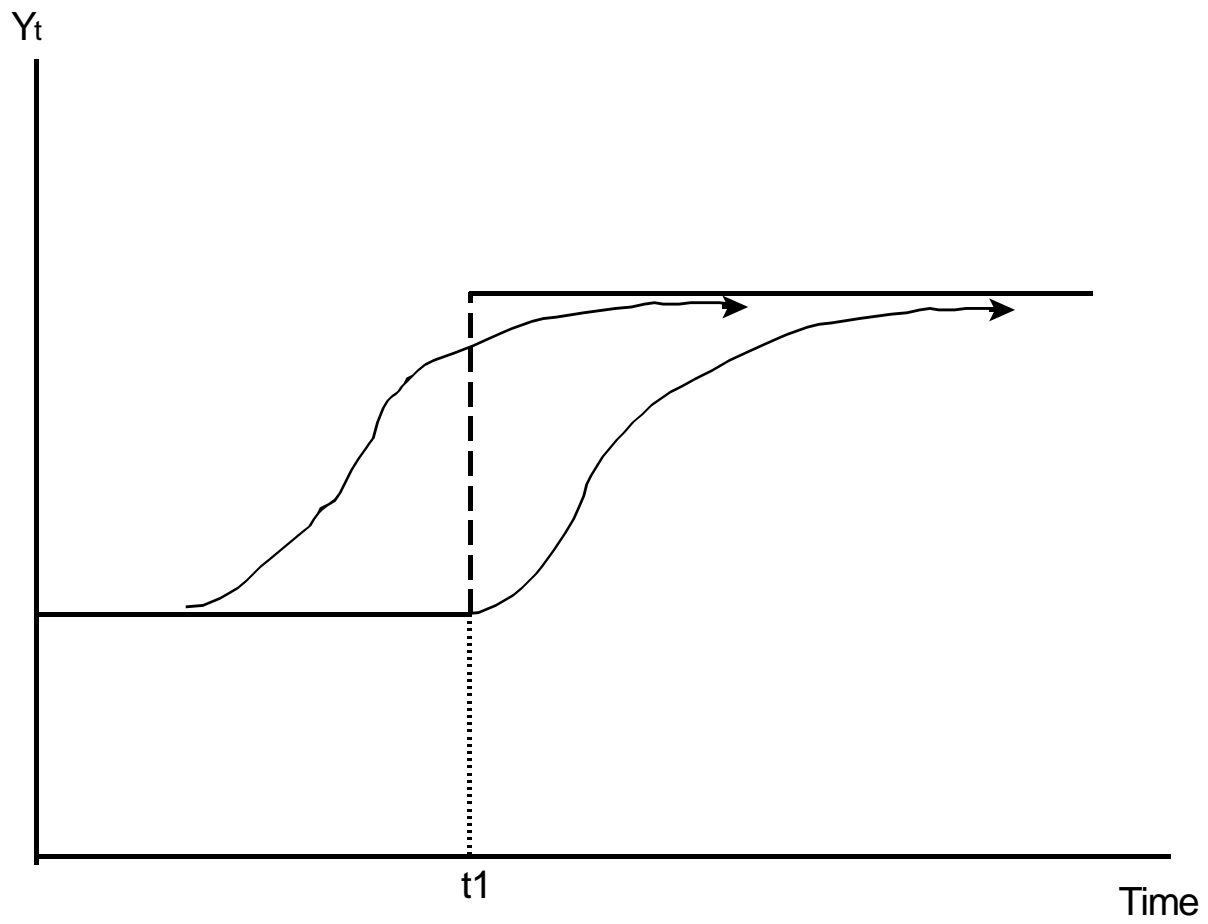


Figure 5