13
Price expectations and the behaviour of the price level

R.M. Solow
Massachusetts Institute of Technology

Chapter 13. *Mathematical Models in Economics*. MOL Bacharach, MAH Dempster, and JL Enos, eds. University of Oxford (1990).

13

Price expectations and the behaviour of the price level

R.M. Solow
Massachusetts Institute of Technology

Abstract

This chapter reproduces the third of three lectures given at the University of Manchester in March 1969 and published by the Manchester University Press in 1970 under the title *Price Expectations and the Behaviour of the Price Level*. It is appropriately included here because the lectures were actually written in Oxford while I was Eastman Professor and Fellow of Balliol College in 1968-69. They represented an early attempt to understand critically what later came to be called 'the expectations-augmented Phillips curve'. The first two lectures reported on some tentative empirical work for the U.S. and U.K. which seemed to show that the full pass-through of 'expected' price inflation into actual inflation, and thus the vertical character of the long-run Phillips curve – however convincing for the longest run – was apparently undetectable at the business-cycle time scale. The third lecture offered a macro-model which might make sense of the apparent paradox. It is a model in which neither the labour market nor the goods market is supposed to clear in the short-to-medium run. The later literature managed to transform the Phillips curve from a description of disequilibrium dynamics into a specification of the 'supply of labour' or even of 'aggregate supply'. That still seems to me to have been a mistake.

The notion that there exist trade-off relations, connecting the rates of change of prices and money wages with real characteristics of the economy, has certain paradoxical implications that have sometimes been noticed. One of them – the apparent failure of the expectations hypothesis over usefully long periods of time – has already been discussed. Another apparent paradox depends on some of the characteristics of the empirically fitted trade-off surfaces.

It is well known that statistical Phillips surfaces – which are trade-off equations with rate of change of the money wage as the dependent variable—usually or always show a strong back-effect from price changes, perhaps with a short lag. Thus an equation to explain w will have p as an explanatory variable,[1] just as my equations for p have w or ulc on the right-hand side. Typically—maybe even universally—the coefficient of p is less than 1, usually around $\frac{1}{2}$ or even less. This is by itself paradoxical, because it smacks of severe money illusion. It implies that the process of wage determination does not compensate the worker fully for price inflation (and would over-compensate him for price deflation). If the rate of price inflation were to rise by one percentage point and stay there, all other things equal, money wage rates would inflate by only half a point a year faster than they had been doing. Therefore, real wages would rise half a point slower than they had been rising before. If there were no change in the rate at which output per worker were rising, the share of wages in total output would fall steadily. If this story were really so, we would certainly have noticed it.

On the other hand, the price trade-off equations I have been describing all have the rate of change of the money wage as an explanatory variable (or the rate of change of unit labour cost, which comes to the same thing with a given rate of productivity increase). In all of them, w appears with a coefficient less than one, even in the permanent trade-off equation with fully adapted expectations. The coefficient may be in the neighbourhood of the share of wages in output, but that is not very important in the present context.

This seems to lead to the opposite paradox. Imagine an initial situation with unit labour cost constant (i.e. $ulc = 0$) and prices constant. Now let the wage begin to rise at 1 per cent a year, everything else unchanged. The price level will begin to rise, and rise faster as expectations adapt; but eventually, according to the equations, p will stabilize at a value equal to the coefficient of w or ulc in the permanent trade-off equation, therefore less than 1 per cent a year. It appears, therefore, that the real wage will continue to rise. With an unchanged rate of productivity gain, the share of wages will rise steadily.

From the point of view of the wage trade-off, it appears that faster inflation cheats the wage earner; from the point of view of the price trade-off, it appears that faster inflation cheats the recipient of nonwage income. Yet I imagine most of us suspect that, except in the short run, faster inflation cheats neither group systematically (except the recipients of fixed monetary incomes, who can be neglected for this purpose). What can be happening?

---

[1] These lower-case symbols represent rates of change.

One possible escape from this paradox presents itself immediately: take account of both the wage equation and the price equation. This has to be part of the answer, because it is obviously right to take account of both equations if they both in fact hold. But it is not the whole answer, as one can easily see.

Suppose we think of an eonomy with stationary productivity, so $w = ulc$; it is easy to make the appropriate modification when output per man is actually rising. Take the simplest sort of permanent trade-off equations:

$$p = a + bw$$

$$w = A + Bp,$$

where a and A are not necessarily constants, but depend on the real configuration of the economy in ways we need not specify now. According to the empirical generalization I mentioned earlier, both b and B are between zero and one. Now for given values of a and A, these are two linear equations in the two unknowns w and p. They can always be solved for w and p because their determinant is $1 - bB$, which is positive. The unique solution is:

$$p = \frac{a + bA}{1 - bB}$$

$$w = \frac{A + Ba}{1 - Bb},$$

which are both positive if a and A are positive; all combinations of signs are possible. This is the only possible state of steady wage and price inflation corresponding to any particular real configuration of the economy.

Remember that I have assumed productivity to be stationary. In conditions of steady inflation, one would expect the price level and the money wage to be rising at the same rate. Then the real wage and distributive shares would be constant or else would be changing slowly in a direction and at a pace governed by real factors, not by the mere fact of regular inflation.

But the little model tells us that w will be bigger or smaller than p according as $A + Ba$ is greater or smaller than $a + bA$. Indeed, if V is the real wage and v its rate of change,

$$v = w - p = \frac{A(1 - b) - a(1 - B)}{1 - bB},$$

and the real wage rises or falls according as the numerator of that fraction is positive or negative. The behaviour of the real wage in regular inflation appears to depend on the parameters of both trade-off equations, and on the particular real configuration in which the model economy is established. For v to be zero, we must have $A(1-b) = a(1-B)$ and that would be a coincidence.

So taking account of both trade-off equations gets us somewhere; the rates of wage and price inflation corresponding to any real state of the economy are determined.

The significance of having b and B both less than one is that then the 'equilibrium' rates of inflation are stable. If the product bB were bigger than one, a chance increase of w would induce a rise in p, and that in turn would induce a further rise in w bigger than the initial impulse. The result must be hyperinflation or hyperdeflation, depending on the direction of the original impulse. When bB is less than one, there is leakage, as in a multiplier system with a marginal propensity to spend less than one. An initial impulse sets off a sequence of echoes, but they diminish in size fast enough so that they cumulate only to determine finite rates of price and wage inflation.

The 'incidence' of inflation depends, as I have said, on all the parameters of both trade-offs, and on the real situation as well. Notice, for instance, that if $b < 1$ and $B = 1$ the stability condition is satisfied, but the steady-state w is necessarily greater than the steady-state p (provided that A is positive, i.e. that the labour market is tight enough to cause the money wage to rise even if the price level were constant). Thus the real wage is always rising in regular inflation if there is 'money illusion' in the price equation but not in the wage equation. And, of course, if $b = 1$ and $B < 1$, so all the money illusion is in the wage equation, then regular inflation cheats the wage earner and the real wage falls. In the case that seems to correspond to most regression results, we are left with no presumption at all, but only the unsatisfactory implication that regular inflation will cheat one side or the other permanently, except in the improbable case that all the parameters fall out just right. I call this implication unsatisfactory because it does not correspond to common observation. We expect inflation to generate distributional shifts in the short run, but hardly continuing shifts if the inflation becomes regular.

(This line of argument sounds like the strict expectations hypothesis again. Quite so; the analogy is pretty close. The argument that 'rationality requires $b = B = 1$, and therefore in fact it must be true that $b = B = 1$' operates just like the strict expectations hypothesis. Inserted into the trade-off equations, it requires that $p = a+w = w-A$, and therefore that $a + A = 0$. Since a and A depend on real factors, we are back to the Friedman notion of a 'natural rate of unemployment' by a slightly different route.)

So far I have discussed the real wage as if it were nothing more than the casual outcome of a tug-of-war between the money wage and the price level. It is that too, no doubt, and the whole discussion of trade-off equations is intended to elucidate that side of the matter. But presumably the real wage figures also as a non-trivial part of the functioning of the real economy. And since real factors appear in the trade-off equations, the anlysis so far has been incomplete. In terms of the simplified model I have been using, it is not enough to say that a and A depend on real factors and, together with the trade-off parameters, determine w, p, and therefore $v = w - p$. It is not enough because the real factors, on which a and A depend, may have to change precisely because the real wage is changing. A complete theory must take that dependence into acount.

A complete theory would be a self-contained macroeconomics; I am in no position to offer that. For the particular point at issue, I hope I can get away with a considerably simplified version of a model analysed in the *Quarterly Journal of Economics* (November 1968) by Joseph Stiglitz and myself.

To begin with, suppose we agree to represent the 'real factors' in both trade-off equations simply by the current level of real output, Y. Obviously, I don't believe that will really do, nor do I expect anyone else to believe it; it is merely an expository simplification. Holding to linear relations, I can write the slightly extended trade-off equation as:

$$p = a + bw + cY$$

$$w = A + Bp + cY.$$

Here c and C are positive coefficients; $-a/c$ and $-A/C$ are, respectively, the level of real output that would stabilize the price level if the money wage were constant, and the money wage if the price level were constant. (I am still imagining productivity to be constant; otherwise replace w by $w + r$, where r is the rate of change of labour requirements per unit of real output.)

As before, these two equations can be solved for p and w because $bB < 1$. One gets

$$p = \frac{a + bA}{1 - bB} + \frac{c + bC}{1 - bB} Y$$

$$w = \frac{A + Ba}{1 - Bb} + \frac{c + bC}{1 - bB} Y$$

and therefore,

$$v = w - p = (1 - bB)^{-1} \times$$
$$\{[A(1-b) - a(1-B)] + [C(1-b) - c(1-B)]Y\}.$$

This is no more than we had before, except that the dependence on Y is made explicit. Evidently there is one value of Y, which I shall call Y^*, that makes $v = 0$, i.e. one level of real output that causes the money wage and the price level to rise at the same rate, so that the real wage is constant.

It is important to know whether v is an increasing or decreasing function of Y; that is, whether a still higher level of output than Y^*, a tighter economy, would make the real wage rise or fall. A higher level of output makes the money wage rise faster; it also makes the price level rise faster. The real wage will rise if wage inflation accelerates more than price inflation; in the opposite case, a higher level of output will cause the real wage to decline or to rise more slowly, even though money wages and prices both rise more rapidly.

From the equation, v is an increasing function of Y if $C(1-b)$ exceeds $c(1-B)$, a decreasing function of Y if the inequality is reversed. I think it is a fair reading of most statistical work in this field that b is probably slightly bigger than B but that C tends to be bigger than c. That judgment translates into the statement that the back-effect of wages on prices is perhaps slightly larger than the reverse effect of prices on wages; and that the pressure of demand drives the rate of inflation more strogly through the labour market than through the market for goods and services. If this is a correct description of the state of affairs, it is hard to extract a clear presumption about the relation of v to Y.

Guesses have been made, of course. Professor Kaldor, for example, holds that wages are more sticky than prices. That is to say, when Y exceeds Y^* margins increase and the real wage falls; when Y is less than Y^*, wage rates tend to be better maintained while margins weaken, so the real wage rises. In this view, v is a decreasing function of Y. (Professor Kaldor identifies Y^* with 'full employment', but that definition, if that is what it is, appears to have no particular merit.)

One difficulty with my brief summary of empirical findings is that it refers to a pair of linear trade-off equations, whereas much empirical work, though not all, suggests that both equations may be nonlinear. Many statistical Phillips-curves either assume or conclude that the relation between w and the unemployment rate is hyperbolic; that would be roughly equivalent to make w an increasing, convex function

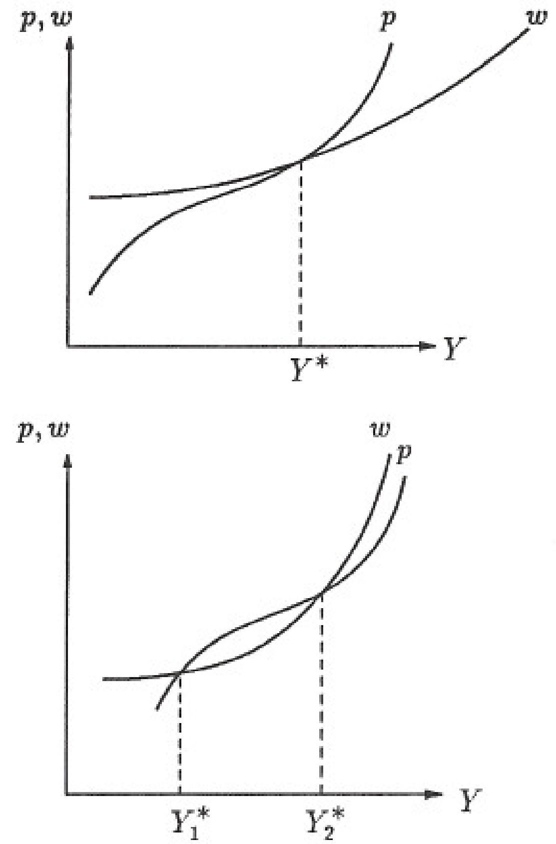

Figure 1

of Y. On the other hand, my own tentative version of the price trade-off equation makes p a nonlinear function of the rate of capacity utilization, and therefore of Y in the short run.

I have drawn two possible configurations in Fig. 1. The upper picture illustrates the Kaldor configuration. To the right of Y^*, p exceeds w, so the real wage falls; to the left of Y^*, w exceeds p. Thus v is a decreasing function of Y. In the lower version, there are actually two levels of Y at which w and p inflate or deflate at the same rate, so that V is constant. The smaller one is qualitatively like the upper picture. The larger stationary point is just the opposite; near it, the rate of change of the real wage is an increasing function of Y. Obviously, then, in this nonlinear case the relation between v and Y is not monotone, but U-shaped. Still other patterns are possible, including some in which the two curves never intersect; but that does not sound economically interesting when one remembers that the 'true' trade-off relations, if there are any, must include several determining variables, some of which will affect the labour market more strongly and some the market for goods.

The only safe course is to retain both possibilities and to consider separately the case where v rises with Y and the case where it falls.

I turn now to the other side of the argument, the relation between Y and V, between the level of output and the real wage (not its rate of change). In the dynamic trade-off relation we have been considering, the causation runs from Y to v. For simplicity, I shall formulate the other relation in static terms; the causation runs generally from V to Y. The real wage is regarded as a determinant of the level of output.

The real wage rate functions macroeconomically in two ways: as a source of income and as an item of cost. Correspondingly, the size of the real wage operates on the level of aggregate output both from the demand side and from the supply side. So the character of its influence may depend on whether demand factors or supply factors predominate.

The normal presumption is that aggregate demand in real terms is an increasing function of the real wage rate. For a given amount of output and employment, a higher real wage necessarily redistributes income from other forms toward wages. On the presumption that the marginal propensity to consume wage income is very high, the result will be an increase in real consumption demand. One ought to offset against this upward shift in the consumption function the possibility that the higher real wage reduces the prospective profitability of investment. If it does, the investment schedule may fall at each level of output. The net effect on the combined aggregate demand schedule is moot in principle, but one supposes that the consumption effect will usually outweigh the investment effect. In that case, aggregate demand is an increasing function of the real wage. In circumstances where actual current real output is determined on the demand side, current output and employment will be higher the higher the real wage.

I would not expect this relation to be very strong (by which I mean I would not expect the elasticity of real aggregate demand with respect to the real wage rate to be very large). The simplest sort of Keynesian model of income determination makes real consumption a function of real income and real investment more or less exogenous. In that case, aggregate demand is independent of the real wage. No one would justify so crude a model of income determination these days. But if it is even a fair first approximation, the real wage effect must belong to the second approximation, or even perhaps the third. Not much is known in fact about differences among the marginal propensities to spend different sorts of incomes, primarily because there are no official data on disposable income by income-type. In the absence of evidence, the safe course is probably to regard aggregate demand as a gently increasing function of the real wage.

Aggregate supply presumably goes the other way. In the short run, with the economy's stock of capital goods given, the aggregate supply of output for given money wage and price level is the volume of real output industry as a whole is willing to produce. For a given money wage, aggregate supply will be higher the higher the commodity price level. The wage bill is the largest element of prime cost; at a given money wage, a higher price level permits profitable operation of older, less productive, higher cost, plant and equipment. If the price level were lower or the money wage higher, some marginal capacity would be unable to earn the required quasi-rents and would be shifted into idleness. There would be a corresponding reduction in employment. One supposes, therefore, that the aggregate supply of real output is an increasing function of P/W, the price level in wage-units, or a decreasing function of $W/P = V$, the real wage.

It follows that whenever there is excess demand, so that actual output is determined on the supply side, current output and employment will be lower the higher the real wage.

There have not been many—perhaps not any—attempts to estimate aggregate supply functions econometrically. In principle, an estimate of the aggregate production function plus an assumption about the average 'degree of monopoly' will determine an aggregate supply curve. There is a difficulty of interpretation, however. Most attempts to estimate aggregate production functions seem to be getting at long-period relations. In the present context, we want a short-run curve relating fluctuations in output and the corresponding variations in employment, with variations in effective demand the prime mover. One doesn't know if long-run production functions contain the right sort of information.

There can be no harm in an example. Suppose that $Y = N^h$ in the short run, where N is employment and h is between zero and one. (If the traditional Cobb–Douglas parameters hold for short-run fluctuations of output and employment, then h is about $\frac{3}{4}$.) Then marginal cost is W/hN^{h-1} and we can set $P = mW/hN^{h-1}$, where m is the mark-up of price over marginal cost, which I shall suppose insensitive to small variations in Y It now follows that the aggregate supply function is $Y = (mV/h)^{h/h-1}$. For instance, if $h = \frac{3}{4}$, Y is proportional to V^{-3}; if $h = 0.6$, Y is proportional to $V^{-3/2}$. It seems plausible that the proper short-run value of h might be smaller than the proper long-run value, but nothing much depends on that.

At very low real wage rates, one would suppose that aggregate demand considerations predominate. There is plenty of spare capacity and unemployment. Real output would be higher if there were a market for it at the going level of prices. In this state of affairs, a higher real wage will expand aggregate demand. It will also reduce margins, but since the limit to output is on the demand side, output will expand. If aggregate demand is only a very slowly increasing function of the real wage rate, then this effect will be small.

At very high real wage rates, there is more likely to be excess demand and aggregate supply considerations will predominate. Real output would be higher if it could be profitably produced at the going wage-price configuration; by definition there is an unsatisfied market at the going price. In this state of affairs, a higher real wage corresponds to a lower level of output; marginal capacity will be priced out of operation. (Of course, when there is excess demand both wages and prices will be rising rapidly; but when the level of output is supply-determined, it will respond to the ratio of P to W. There are undoubtedly lags and frictions in this mechanism, but I shall ignore them. They get a more careful treatment in the article referred to earlier.)

The situation is represented in Fig. 2, with the real wage V measured horizontally and real output Y measured vertically. The curve we have just been discussing has an inverted-V shape; but we reserve the possibility that the rising branch is relatively flat. On this Fig. 2 superimpose a horizontal line at $Y = Y^*$, the level of output at which the trade-off relations generate equal rates of increase of W and P, and so a constant V. There are now two main cases according as v is an increasing or decreasing function of Y. Special cases arise when the horizontal at Y^* intersects only one branch of the $Y-V$ curve, or lies wholly above it. These are also illustrated in Fig. 2.

The diagram works as follows. The momentary state of the economy is summarized by two variables, V and Y. By assumption, V and Y always lie on the humped curve that relates the level of output to the real wage rate. The direction in which V is changing comes from the combined trade-off equations. In case (a) v is an increasing function of Y, so that V is decreasing (v is negative) below Y^* and increasing (v is positive) above Y^*. In all cases, Y moves as it must move to stay on the curve. The arrows show what happens.

In part (a) of the diagram there are two possible situations with constant real wage; that is to say, there are two points on the $Y-V$ curve where $Y = Y^*$ and $v = 0$. But the left-hand one is unstable. The arrows indicate that, if the economy is disturbed from that point, it will move further away, either into a situation of falling real wage and falling output or into a stage of rising real wage which leads to

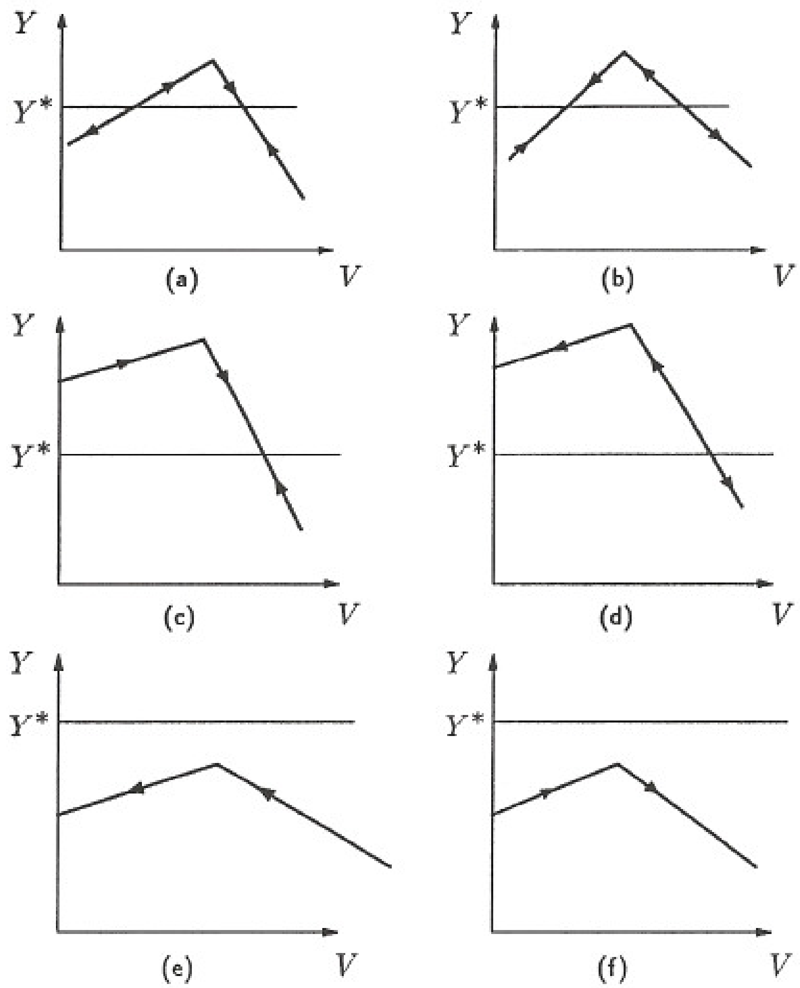

Figure 2

the right-hand constant-V point. This one is stable; if the economy is disturbed slightly in either direction, it will return to its initial configuration. Part (b) of the diagram is similar except that the left-hand constant-V situation is the stable one.

Parts (c) and (d) of the diagram illustrate cases in which the left-hand excess-supply equilibria disappear and only the exacess-demand equilibria remain. In (c) there is a stable configuration with constant real wage; in case (d) there is not. In principle, there are symmetrical cases in which only the excess-supply equilibrium remains; but those cases are unlikely if aggregate demand is only weakly dependent on the real wage. Parts (e) and (f) illustrate situations with no equilibrium point.

The interesting situations are those illustrated in (a) and (b). They show how a closed model can dissipate the paradox with which I introduced this chapter. In part (a) of the diagram there is a stable equilibrium (at least in the short run) with excess demand for output. Almost certainly there is inflation; the price level and money wage are rising. But they are rising at the same rate so that the real wage is not changing. The distribution of national income is also constant. The parameters of the trade-off equations determine the actual rate of inflation, and they help to determine the equilibrium level of real output and the real wage. But whatever the parameters of the trade-off equations, regular inflation can cheat neither wage nor nonwage income systematically.

In the situation of part (b) the short-run equilibrium is demand-limited; there is excess supply in commodity markets and presumably unemployment in the labour market. The diagram doesn't say whether the price level is deflating or inflating. That can be found out from either trade-off equation. Such situations do in fact seem to arise, with the price level actually rising despite the existence of what any reasonable man would call excess capacity and unemployment. That observation was the starting point for this discussion.[2]

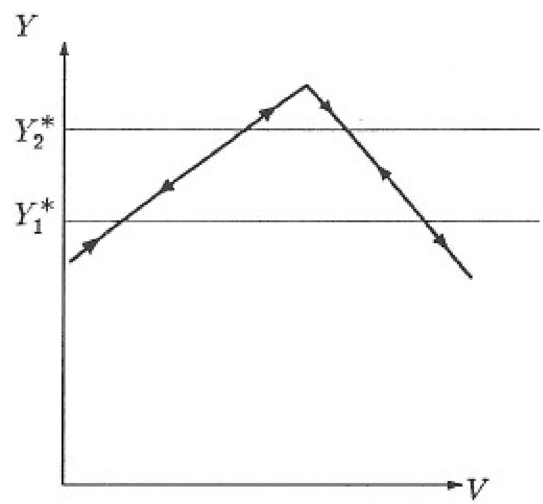

Figure 3

I will conclude with one further possibility. Earlier, in Fig. 1, I mentioned that in plausible circumstances there might be two levels

[2] Fig. 2(b) has one unacceptable implication as it stands. Suppose the aggregate demand schedule were to rise (by virtue of government spending, say). The upward-sloping part of the $Y - V$ curve would shift upward. The new short-run equilibrium position would have a lower real wage, but the *same* level of real output Y^*. The implausible result, that a shift in aggregate demand has no effect on real output in an underemployment equilibrium, is an implication of the assumption that $v = 0$ implies $Y = Y^*$. If the relation between v and Y involves other variables as well, in particular V, then the implausible implication disappears. I have made v depend only on Y^* simply as an expository simplification.

of income at which the real wage would stabilize. That leads to Fig. 3, which contains four possible equilibrium points, two of which are stable and two unstable. In such a case, depending on where it 'begins', the economy may tend in the short run either to an excess-demand equilibrium or to one with excess supply. The interesting possibility emerges that the economy might be jolted out of an underemployment equilibrium and transferred to a new 'initial position' from which it might find its way to an inflationary excess-demand equilibrium, or *vice versa*.