

2 The origins and further development of the natural rate of unemployment

Edmund Phelps

The idea of the natural rate of unemployment, more generally the equilibrium path of the unemployment rate, challenged the main tenets of Keynesian thought: that doctrine held that the unemployment rate was a function of effective demand, and demand management could aim for the unemployment rate deemed best. In the new Keynesian version, fiscal policy was to be governed by neoclassical considerations while monetary policy was to engineer a point on the unemployment–inflation trade-off – on the Phillips curve.

Modelling of the natural rate idea led to two propositions. One of these was a conclusion from the model sketched in my 1968 paper (Phelps, 1968, part 2). Management of monetary demand cannot engineer an arbitrary unemployment rate other than the natural level without sooner or later generating a continuing disequilibrium manifested by rising inflation or mounting deflation – then collapse.¹ Maintaining the unemployment rate below the natural level, for example, would ultimately or immediately pull the actual inflation rate above the expected rate, which would drive up the expected rate; but each such rise would push the actual rate up by as much, leaving the disequilibrium undiminished.

The other proposition was implied by my 1967 paper (Phelps, 1967) and rather a similar thesis was forcefully argued by Milton Friedman in his 1968 paper (Friedman, 1968). Monetary policy can make a permanent difference only to nominal variables: a policy to generate a finite increase or decrease in the inflation rate will generate only a transient dip of the actual unemployment rate relative to the path it would otherwise have taken. In particular, the actual unemployment rate, though occasionally hit by such shocks, is constantly homing in on the natural rate. This last part – equilibration – makes this the stronger proposition of the two, as we could believe the former without having much faith in the homing in.

Some of these terms need defining, since some authors use them in other senses. 'Equilibrium' here means, as theorists from Marshall to Hayek meant, that expectations of wages, etc. are borne out, absent a new and

unforeseen shock; 'disequilibrium' means a gap between expectations and outcomes. Under the maintained assumptions, each equilibrium path of the unemployment rate (indexed by its starting point) converges to the path of the natural rate – conceived as a path independent of monetary policy, at least approximately. The current natural rate may be defined as the current stationary rate: the level of the unemployment rate such that it would be unchanging for the moment as least.² For analytical convenience the natural rate was taken to be a constant, hence equal to the equilibrium stationary state level.³ The first proposition assumes that in the stated circumstances there is a tendency for expectations to narrow the disequilibrium unless policymakers keep people off balance. The second proposition assumes that the tendency to equilibrium is general. Then the actual unemployment rate converges to one of the equilibrium 'glide paths', all of which converge to the natural rate path.

Strikingly, this message that economists took away from the natural rate discussions of the late 1960s – the neutrality of money and inflation and the homing in – was less a theory of employment determination than a set of *axioms* that we might require of such a model independently of what substantive building blocks the model is made from.⁴ Indeed, the concept of the natural rate, as I have on occasion acknowledged, had existed for decades prior to the arrival on the scene of Milton Friedman and myself. The postulate that inflation was neutral for the equilibrium path of output, employment and some other 'real' variables was introduced by Abba Lerner in the 1940s and by William Fellner (a great teacher of mine) in the 1950s.⁵ Perhaps my 1967 paper and Friedman's were more emphatic and explicit about homing in to the natural rate (on which I was more cautious in my 1968 paper).

There was a tendency among quite a few scholars, myself included, to forget that my 1968 paper on equilibrium unemployment sketched a *substantive model* of the determination of the *size* of the natural unemployment rate and the course of the equilibrium unemployment rate path which leads to it. The 1968 paper by Friedman also had a substantive side, though it sketched departures of the labour force (to which employment was equal) from its equilibrium path: unemployment does not appear.

The present commentary on the natural rate begins by looking back: to what my model of equilibrium unemployment was, and how that model arose. The second section argues that the full implications of this model and kindred sequels – all featuring what I prefer to call 'incentive wages' – have not, until very recently at any rate, been grasped. The discussions of 1968 hit upon one implication of central consequence for Keynesian thought, the scope of monetary policy. But, I will maintain, there are other ramifications of this kind of model of equilibrium unemployment – implications for fiscal

policy and other non-monetary impulses – that are, if judged empirically plausible, no less devastating for Keynesian doctrine.

Background to the original model

It might be thought that the late 1960s modelling of the Phillips curve and, as a byproduct, the natural rate grew out of great agitation and yearning for light. In fact, the Keynes–Phillips orthodoxy was sailing on smooth waters, the object of much congratulation, rather like the liner *Titanic* prior to its collision with the fateful iceberg. In the present case the iceberg was the neutrality axiom of Lerner and Fellner. What made a collision of the Phillips curve with neutrality inevitable was that sooner or later someone – or some *two* – would provide one or more micro–macro models of the Phillips curve, and with that development the difficulty of maintaining that inflation was non-neutral, which the Keynes–Phillips position implies, would be exposed.

My efforts at a theoretical understanding of the Phillips curve began in earnest over the summer of 1966 in the Sidgwick Avenue building at Cambridge, and my first few months at the University of Pennsylvania in the autumn. In the preceding winter I had written a paper on optimal inflation/unemployment control, published in 1967, in which an expectations-augmented quasi-Phillips curve was written down:

$$p - p_{-1} = \phi(u) + p^e - p_{-1}, \quad (1)$$

equivalently

$$p = \phi(u) + p^e, \quad (1')$$

where p is the money price level being set, p^e is what it is expected to be, and u is the unemployment rate. But there was nothing about the micro-economics of the function ϕ . Furthermore, the money wage level was implicitly the passive partner of the price level rather than the other way around, as Phillips and most practitioners supposed. A microeconomic understanding of the relation between inflation and unemployment did not yet exist.

With the benefit of hindsight the puzzles I was struggling with can be reduced to a few basic problems: How can there be involuntary unemployment, particularly in conditions of equilibrium in the expectational sense? How could the unemployment rate remain, however briefly, below its natural level? In such a infra-natural state, what is the process by which nominal wages go on spiralling upward? How might one introduce into this model the Lerner–Fellner acceleration hypothesis that as long as monetary policy, say, kept the unemployment rate below its natural level, the rate of

increase of the average wage would steadily increase? I had only a foggy notion at best of the answers to any of these questions. However, I did have the sense that the way to the answers was somehow to lay out a model – not a complete system of differential equations but nonetheless a serviceable description of a highly stylised hypothetical economy.

There were bits of labour economics that I started from with each new attempt at a model. I had read a little of Dunlop and Slichter, the Harvard labour economists, Paish, the LSE economist, and Wallich, my colleague over several years at Yale. From them I took away the impression that when the economy is pressured, at least for a time, into operation at a level in excess of its equilibrium steady state level, the low unemployment rate poses various inconveniences for firms, which try in turn to cope by setting higher wage rates. I also had a more recent memory of the dynamics of employment arising from employee turnover behaviour as it was modelled by Richard Lipsey in an otherwise econometric paper of his on wage inflation and employment (1960). Yet these insights, however necessary, were missing something fundamental, it seemed to me. They did not put us into the mind of the firm, or its personnel manager. Man is a thinking, expectant being! What was needed was a model of a sequence: the firm's expectations, its subsequent actions and those of the others, the discovery of the others' actions, the formation of new expectations, and so forth.

I had also read the (1964) paper on wages and employment, replete with econometric estimates, by Sargan of LSE. This paper postulated a required nominal wage level that is an increasing function of the employment rate (hence decreasing in the unemployment rate), given expectations of the price level. I took from this paper the rather important point that the rate of increase of nominal wages is a function not just of the *level* of unemployment but also the *change* of employment. It also encouraged my impression that when firms plan to increase employment they offer an increased wage simultaneously; the wage is not completely described as a feedback response to discovery of changes in the market wage and the total unemployment rate. On the other hand, at the embryonic stage of my thinking then, this paper was a distraction and an unnecessary complication. For weeks, I focused exclusively upon expectations of the price level by the personnel manager and his employees rather than their expectations of what the general money wage level was going to be.

By the time I was settled into the University of Pennsylvania I had a 'story' about labour market equilibrium and wage dynamics – to use the phrase I finally used as a title. The unemployment rate might move to so low a level that, to moderate the associated quit rate, every firm wants to offer its employees a better *real* wage as an inducement not to quit with such readiness; but as all firms pass along the implied money wage increase, the

price level increases in proportion (beyond what it was going to do anyway), an increase that is unexpected; to keep the unemployment rate down, there must be a succession of such wage increases and hence continually unexpected inflation – greater than whatever rate was expected. In such a scenario, the unemployment rate was below equilibrium; the steady state equilibrium must be one with a larger rate.

One day, though, it struck me that something was terribly wrong with that story. Suppose that each wage increase is accompanied by a proportional increase of productivity, so that the price level remains unchanged and the real wage is increased. Then it would be implied by the original story that the reduced unemployment rate was consistent with equilibrium. With each advance of the real wage and productivity, the equilibrium volume of unemployment would be reduced again, with unemployment vanishing in the limit. Furthermore, as I went on to see, it is presumably the effect of its increased wage scale on the expected *relative* wage that the firm counts on to induce a reduction of the quit rate. Employees might not respond as assumed to a higher money wage if they could observe the same increase in money wage rates at all the other firms.

The model was then reconstructed: the quit rate is a decreasing function of the firm's relative wage. For simplicity, only the relative wage and the unemployment rate determine the quit rate, not the real wage. In the revised version, if the unemployment rate is driven to a sufficiently low level, every firm raises its wage in the expectation of achieving an increase in its relative wage in order to induce a moderation of its quit rate; but as all firms try to outpay one another the result can only be disappointment – a disequilibrium in which expectations of the money wage at other firms are found to be too low. Equilibrium in the labour market thus requires a large unemployment rate – large enough to dissuade the representative firms from attempting an unrepresentative outcome. The resulting Phillips curve was

$$w - w_{-1} = \phi(u) + w^e - w_{-1}, \quad (2)$$

where w denotes the money wage level. The equilibrium steady state unemployment rate, which makes $\phi(\cdot)$ equal to zero, is a positive number. If monetary policy keeps on yanking up firms' nominal demand prices in order to induce firms to go on employing beyond the steady state rate, firms will pass along each round of wage increase in proportionally higher prices; money wages will continue to go up, round after round, always in excess of what firms expect them to go up by.

A number of features of this model stood out. As already noted, the invariance of labour market equilibrium to whatever inflation rate was expected was a sensational aspect. This was not because there was intense

substantive interest on the part of economists in whether a steady inflation of, say, 6% per year, might make for tighter labour markets than 5%. As I suggested at the start of this chapter, the fascination lay in the implication that Keynesian aggregate demand management – through monetary policy, at least – could not achieve an arbitrarily chosen unemployment rate within some admissible and reasonable range. Keynesian forces could only make transient departures from the gravitational pull of the natural rate.

A second feature was that the unemployment existing at the natural rate, and indeed virtually everywhere on any equilibrium path, was involuntary – not just in Keynes' sense of the term but in the everyday sense that the unemployed could not get a job by offering their labour for less than the going wage. As far as I can recall, this point was not well understood by me at the time of writing, nor for some time after.⁶ But eventually it became clear to me why the model implied that an unemployed worker could not obtain a job that way: if the firm were to accept such a worker at a lower wage though that worker did not apparently differ from employed workers with regard to the likelihood of quitting, the firm would have to assume that the worker's quit rate would be higher as a result; but that trade-off would be sub-inoptimal for the firm to accept since it had already calculated the optimum on the wage-quitting opportunity locus.

Another feature – an 'optional extra' – of the model was the property that, starting from unemployment in excess of the natural level, the equilibrium path would approach the natural rate only gradually. The argument was simply that firms will not jump their employment rolls to the natural level since they face rising marginal cost of imparting firm-specific training, or induction programmes, to new recruits. Whether a specified unemployment rate today will generate unexpected inflation thus depends on the rate yesterday. The augmented Phillips curve became

$$w - w_{-1} = \phi(u, u_{-1}) + w^e - w_{-1}. \quad (2')$$

Hence there was an equilibrium path of the unemployment – a path along which the expected wage is always matched by the actual wage, hence a path given by $\phi(u, u_{-1}) = 0$ – that approaches the natural rate only asymptotically. This was the notion of 'persistence'. In contrast, the idea of 'hysteresis', as used in my 1972 book, at any rate, referred to the effect of unemployment history on the natural rate, either a permanent effect or a long-lasting one.

In another respect, however, the spirit of the model was uncomfortably remote from the behaviour of aggregate data: the setting evoked by the paper suggested that firms would raise their wages – up or down – immediately in response to a shock, then learn what the general wage was

doing once the data for that week or month were finally collected, processed and released. But in reality, though the preponderance of firms may change their wages by discrete amounts, instantaneously or when it suits them, the weekly time series of the average wage rate appears to be virtually continuous – without large increases no matter how large and widely perceived the shock. This modelling problem was met in a formal way by the apparently artificial device of imposing on the individual firm a gradual response of its wage; implicitly there was some lag in the reporting of the daily or weekly wage. In the last section of the text, however, there was a verbal discussion of the importance of moving to a more complex version of the model that would allow for overlapping wage rates that are revised in a non-synchronous and regular manner, as from a uniform distribution. In an altered version of that paper prepared for a January 1969 conference on which the (1970) conference volume was based, a version that suffered from effects to make it closer to the companion paper by Dale Mortensen, I sketched the algebra of the wage rate leap frogging by such firms for the special case in which unemployment is held at the natural rate. But the major work remained for the 1970s in New Keynesian papers by John Taylor, Guillermo Calvo and myself.⁷

As a result of synchronous wage-setting and either fixed mark-up pricing or an auction market for goods, the model implied that if expectations were rational, instead of adaptive as I had supposed, a jump in demand would generate what could be regarded as a disequilibrium disturbance, resting on forecast errors by the firms about the upcoming wage level, only for as long as it took for the wage data (or other proxy date) to come in. Once the data were in, wage rates would jump the remaining distance needed to put the economy again on an equilibrium track. This was the thrust of Robert Lucas' (1972) model, based on my parable of imperfectly communicating 'islands'. Fortunately, by the early 1970s, Friedman's wonderful example of the peso problem arrived to give one sort of reply: if, in an economy that was at the fixed natural rate, the central bank went on a money-creating spree having always the same probability of ending with the present 'period', then employment will be above natural as long as the spree continues, as wage-setters will not have treated its continuation as a certainty; employment will drop below its natural level in the first period after the rampage is over, regaining the natural level gradually. So there *can* be booms and slumps, not just weekly vibrations. The other reply, which was developed both independently and collaboratively by Roman Frydman and myself in the 1980s was the argument that expectations cannot be presumed to be rational in a world in which the correct model (or theory, as I tend to say) is a source of continuing disagreement. Then your rational forecast is not my rational forecast, unless we are fellow travellers down the

same model's road. Then the economy can take a non-equilibrium path for a long time, even indefinitely. (Even in the 1968 paper, I might add, a certain ambiguity arises over whether the economy assuredly approaches the natural rate path – hence the distinction between the concept of the natural rate and the natural rate hypothesis.⁸)

There was another missing feature – trade unions. In the economics of those days, the fashion had been to suppose that the necessity for so high an unemployment rate in order to contain inflation was ultimately a consequence of the presence of labour unions, which grew aggressive *whenever the labour market became tight*. I took a different view, and in a paper where so little had been eliminated in the interest of simplicity and tractability, I felt it was necessary to draw the line somewhere. Labour unions were banished. Since then, there has been a considerable amount of modelling of unions. It has become clear what unions can add to the story. But a rough idea of the quantitative importance of that addition is not yet in hand.

One last comment. The above account glosses over the confusions and ambiguities that were present in my mind over the months of writing the paper and for months afterwards. The worst of these was the unfortunate remark early in the exposition that it would be necessary to think of the labour force as heterogeneous in order to make sense of positive unemployment as an equilibrium phenomenon. In fact, though, heterogeneity of workers plays no role in the defensive upward push of the product wage in response to the turnover problem that is the driving force behind the creation of unemployment in the model; it is the heterogeneity of firms in the minds of workers that underlies labour turnover. Worker heterogeneity merely plays a supporting role as one of the contributors to the marginal cost of hiring; it is needed to underpin firms' concern about the quit rate only if marginal hiring costs do not include any firm-specific on-the-job training costs, which ultimately I came to depend on more. The next-worst confusion was created by the property of rising marginal hiring costs on which the existence of continuous equilibrium paths depended; it led to a variable vacancy rate, which played a role alongside the unemployment rate as a determinant of the rate of wage inflation. Some readers mistook the resulting model as generalisation of an excess demand model of wage dynamics (in which excess demand was measured by the excess of vacancies over unemployment).

If I have gone on rather lengthily about this now rather old piece of work, it is because it will be useful to have a good idea of what that model of the natural rate was for grasping the simple propositions of the next section on the determinants of the natural rate and hence the possible causes of its long-term swings and shifts.

Further developments

Natural rate models, to repeat, are so constructed that, if they contain nominal variables at all, the equilibrium path from every initial condition is invariant to 'money' and to the purely monetary effects of disturbances generally. The presence of non-synchronous wage- (or price) -setting introduces nominal variables into equilibrium near-term behaviour, but it does not alter the two main propositions: that keeping the unemployment rate bounded away from the natural rate path must ultimately result in inflation or deflation rising to a problematic level, and that a small change in monetary policy will make only a transient difference for the path of the unemployment rate. The riskiness of this doctrine warrants comment. There is another large aspect of the subject, however, with at least as urgent a claim on our attention.

The question arises: what *are* the key determinants of the size of the natural rate – money not being one, by construction – and in what direction do they pull? My 1968 paper observed that faster growth of the labour force would produce a higher natural rate through the rising marginal cost of training. There was also the suggestion that faster steady growth, if it entailed a higher average rate of layoff in the economy, would also produce a higher natural rate. But there was little more than that. It is striking that this paper and all the succeeding papers in the next months and years did not explore the 'comparative statics' of the natural rate. In time this omission in the theory was elevated to the status of a normal property – that the natural rate was a constant, as if a force of nature, a magical thing that is not itself caused.

This constancy, which served only to cover up the yawning gap in our understanding of the natural rate, may lie behind the suspicion with which natural rate doctrine has come to be regarded in some quarters in the past dozen or so years. Natural rate models can hardly be satisfactory if the natural rate, in any suitably general view, has been pulled to and fro by real shocks of the evolution of real conditions while our *models* of the natural rate continue to treat it as a parameter rather than a variable.

There has been a crying need, then, for a *general equilibrium view* of the natural rate – defined now in the general sense of the *current* equilibrium stationary rate, given the *current* capital stock and any other state variables. (It is the unemployment rate that, if it were the actual rate at the moment, would make the current rate of change of the associated equilibrium unemployment rate path equal to zero.) In such a framework, the equilibrium unemployment path from a given initial point is still driven by a natural rate, but the latter is a *variable* of the system – an endogenously moving target that the equilibrium path constantly pursues.

If we view the natural rate path as endogenous, pushed like other economic variables by non-monetary forces, and take on board the rest of natural rate doctrine – actual unemployment tending soon to equilibrium, and all equilibrium paths approaching the natural path – we arrive at a new paradigm: *a non-monetary equilibrium theory of unemployment movements* – an endogenously moving-natural rate theory of movements in the *actual* rate of unemployment.

Over the past several years I have managed to develop to a rudimentary, working stage a general equilibrium theory of this kind in the form of a family of intertemporal micro-macro models. Each one revolves around a distinct kind of asset acquired by the firm that is of importance for its hiring decisions: the trained employee, the customer and fixed capital equipment. Collectively these models provide a 'structuralist story' of how the equilibrium unemployment path is determined and thus disturbed by changing parameters and conditions. The evidence so far encourages me to believe that this theory does rather well at explaining the *shifts* and *long swings* of the unemployment rate in the postwar experience of the Western industrial countries. A detailed exposition of this approach, with a statistical and historical investigation of its explanatory power, is presented in my recent monograph (Phelps, 1994). It is possible, though, to give a glimpse of what the approach has to offer by taking up two themes regarding the closed economy.

One of these themes concerns the effect of shocks to the technology. Consider a one-time shock to the level of labour augmentation appearing in all production functions. In the admirable rendering of the natural rate by Steven Salop (1979), the implication in this regard went unnoticed. Such a productivity shock had no effect on the natural rate path and, accordingly, on the equilibrium path from the initially given unemployment rate; the shock was followed instantly by a neutralising increase of the real wage at each firm, hence no change in the quit rate, and no reason for firms to speed or slow their hiring of labour, the wage having offset the increased productivity of employees. This was a comforting result from one point of view since it meant that secular progress did not have the counterfactual implication of an ever-decreasing unemployment rate; but it likewise implied that there was no near-term disturbance to the employment rate either. This result was a consequence of the wage-wage view of labour turnover behaviour of my 1968 paper on which Salop built. In the later, utility-theoretic formulation by Carl Shapiro and Joseph Stiglitz (1984), the contrast was complete. Such a labour-augmenting shock, in driving up the economy-wide real wage, decreased the propensity to quit, causing the supply wage to shift up less than the demand wage

in the employment–wage plane, with the result that the employment rate is increased and the wage rises proportionately less than productivity. This finding was a consequence of the utility function chosen and the background setting.

The models I have developed take an additional step in introducing the income from wealth in the quitting or shirking function contained by the model. Nevertheless, a one-time productivity shock has no effect on the natural rate, nor on the equilibrium path from the given initial condition, in the two models where there is no slow-moving state variable (other than, as in one model, the stock of employed workers, hence the employment rate itself). In the model where there is such a state variable in the form of the stock of fixed capital, however, there is an effect on the natural rate: if the capital stock could increase in proportion to productivity, to which level it will tend gradually as saving and investment jump up, there would be no effect; the wage rate at the firm, in the economy as a whole, and the employees' non-wage income would all increase in proportion to productivity; as a result, the demand wage and the supply wage (or incentive wage) would increase equally, putting the unemployment rate at its preshock level. But since the capital stock cannot increase, there is a near-term effect. It tends to be contractionary through an important channel since the implied drop in capital relative to the new-found level of labour augmentation operates to raise the real rate of interest and lower the real price of the capital good; the latter effect in turn operates to move the capital goods sector down its (new) supply curve and, since that is the labour-intensive sector, thus to reduce the demand wage relative to the new level of labour augmentation.⁹

The effect of an unanticipated positive increase in the *rate* of labour-augmenting progress expected to prevail over the indefinite future is another story. It might be thought that the effect would be exhilarating for the economy, precipitating new doses of investment. But in fact the prospect of new levels of real income in the future propel an expansion of consumption, while not providing the technical means for increased supply of output in the present, with the result that the real price of assets has to seek a lower level, with real interest rates accordingly pushed up. In one model the effect is a drop in the capital goods output supplied; in both cases employment is contracted. Thus it may happen that the growth rate of an economy is the enemy of its employment rate.

The emerging general equilibrium theory makes demand shocks as much as supply shocks the great movers and shakers of the economy's equilibrium path. In the version constructed here, the theory sees shifts in profitability and thrift as prime sources of disturbances. Adjustments of

domestic assets and of wealth operate to amplify or ultimately to tend to correct the early effect on unemployment.

Here the theory is an echo of pre-Keynesian doctrine in sounding the theme of slump through 'undersaving': public debt and other fiscal stimulus to consumer spending are seen as contractionary. Yet the results have in common with Keynesian doctrine the theme of slump through 'under-investment': in particular, government armaments purchases (and in all but special cases manpower buildups too), as occur in wartime, and more generally any government spending on goods produced by the capital goods sector of the economy are implied to be expansionary – without any reference to the liquidity of a money economy, which is crucial to the Keynesian analysis.

The pre-Keynesian part follows from a very simple mechanism. An increase of consumer demand, whether the response to a spontaneous increase of the rate of pure time preference or the artefact of a recent increase of public debt, creates an excess of consumption demand over consumption supply, with the result that real asset prices have to drop and real interest rates rise in order to eliminate that excess demand. The repercussion of these changes in financial prices is an induced decrease in *the amount of investment of the various types that firms are willing to undertake*. The effect in turn is a decline in the equilibrium path of the employment rate.

The Keynesian part arises from the property that an increased demand by the government for the capital good serves to pull up the relative price of the good whose production is the more labour-intensive and thus to pull up the demand wage in the aggregate employment–wage plane. As an empirical matter, however, it is not clear that this effect is generally strong; it may be confined to a small sub-set of capital goods.

Introducing the interactions of open economies adds further twists to the story. It is found to be theoretically possible that a consumption-demand stimulus in an open economy, if it small enough, will have an expansionary effect at home – a result more Keynesian than that obtained by some Keynesian models – while having a contractionary effect abroad (in proportion to its size). This is the same 'locomotive' in reverse, or 'crowding out' at a distance, previously found in the part-monetary models of Fitoussi and Phelps (1988). The argument is that the fiscal largesse of the country on its own citizenry drives up the domestic real interest rate and thus the world real interest rate, which entails a real exchange rate depreciation abroad, the effect of which is to push up the natural unemployment rate abroad.

The empirical sections of the monograph suggest that this theory of equilibrium unemployment succeeds to an important degree in shedding light on the contrasts between the long period of extraordinarily high

economic activity from the early postwar years to the end of the 1960s and the two nearly global slumps in the 1970s and 1980s. (The view of the 1990s it offers is not far from the mark either.) Of course, other non-classical representatives of the equilibrium approach to unemployment also exist: the insider–outsider models (especially their non-monetary versions), and the still embryonic models based on modern finance notions of credit rationing to firms and households. So it is not suggested that the particular models of the natural rate that I have developed are the only vehicles for a structuralist approach to slumps and booms.

Some reservations

The thrust of these reflections is clear. The natural rate idea remains a powerful concept for macroeconomic analysis. The endogenisation of the natural rate in the work of Salop, Stiglitz and myself – and the parallel work of a less micro founded sort by Kouri, Bruno and Sachs, Fitoussi and Phelps, and Newell and Symonds, to mention only a few – has freed the natural rate to do a great deal more work than it was called upon to do earlier.

Yet it would be remiss of me not to draw back for a critical look at the edifice that has been going up. How much of the macroeconomic world can be captured by a theory that emphasises the natural rate, downplaying monetary factors and channels and downplaying expectational disequilibrium, and that bases the natural rate on incentive wage considerations – quitting, shirking, and adverse selection?

The claim that money and inflation are neutral deserves more qualification than it usually receives. The monetary aspects of the economy do not possess the sort of linearity that we imply in our usual models. In the neighbourhood of the normal range of the economy, money may be neutral ‘in the small’. A 1% increase or decrease of the money supply will make little difference. A change of 15 or 25 per cent, however, may make a large difference. And in the neighbourhood of such a disturbed range, even a small change of the money supply may make a large difference. It is remarkable how firmly and how early Irving Fisher got a hold of this point, and how little attention was paid to it by Keynes and most Keynesians.

Fisher’s point may be important for understanding the various forces behind the high unemployment of recent years. It is plausible that the slump of the 1990s is not altogether a holdover of the lengthy 1980s’ slump in the economies but is, rather, *in part* to be explained by the huge drop in inflation over the 1980s, which has left households and corporations saddled with larger levels of real indebtedness than they anticipated when they took on much of this debt in previous years. By the same logic, it must

be that some of the 1980s' slump – particularly the high unemployment rates still lingering toward the end of the decade – is also to be explained by this phenomenon of 'debt deflation', in Fisher's term. Very likely the bulging levels of real debt overhang had a sobering effect on households and firms contemplating their customary accumulations of real assets and financing them by the customary mix of debt and equity. If this hypothesis is true and quantitatively important, it means, as I see it, that the natural rate hypothesis is not as good an approximation of reality as it first seemed to us in the 1960s. However, no purpose would be served by conceding this small but valuable piece of territory to the resurgent monetary forces until the econometric scouts have confirmed that the insurgents have solid control of it. At the time of writing, the relevant econometric tests have not yet been performed. Scraps of circumstantial evidence and supporting observations are useful but not conclusive.

The reliability of convergence to the natural rate path is, in my view, no less serious an issue. It is one thing to suggest a broad tendency toward equilibration (not in any exact way, of course) over a range of circumstances and histories, and quite another to close the door to the intuition, ventured by Keynes and a few others, that there are apt to be episodes in which the economy lingers away from equilibrium or moves back and forth past the correct expectations to have. Research by Roman Frydman and myself, some of it collaborative, has sought to identify some conditions that may block convergence to equilibrium. On the other hand, if my recent econometric work implementing the variable natural rate theory (mentioned on p. 26 above) is basically right, there is – on average, at any rate – a tendency for the unemployment rate to approach the natural rate path. A question very much left open, however, is whether, following a shock, there is a systematic tendency for a stampede, culminating in overshooting – a phenomenon perhaps traceable to expectational errors or perhaps to a desire of each enterprise manager not to be seen having taken fewer precautions than other managers even if the situation of the enterprise does not call for them.

Finally, is the labour market really a matter of incentive pay and nothing else? Of course not. Students should understand that seeing *N* papers in a row without a mention of the word 'union', say, or 'insiders' and 'outsiders', should not be taken to indicate that the author doubts the importance of those considerations. The more convenient model is always preferred, and as between equally convenient models, the one in which the writer has a vested interest is selected.

The insider–outsider theory greatly enriches the dynamics of employment, and it has interesting implications for the real wage response to shocks. Some models by Andrew Oswald and by Assar Lindbeck and

Dennis Snower (1989, for example) are among those that have had an impact in the 1980s. It does not seem that these models are contrary by their very nature to the natural rate concept. There are insider–outsider models in which the unemployment rate gravitates back to the level that could be regarded as the natural rate. Indeed, there are elements of the 1960s' models, such as the marginal costs of hiring, that have fed directly into insider–outsider models. A careful recognition of such costs tends to generate a natural rate *zone* instead of a point, since the employment level to which a firm will downsize its workforce when starting with an excess of employees is possibly greater than the level to which it will increase its workforce when starting with a deficiency.¹⁰ One could go further with the contention of similarity, arguing that these models offer another dimension – the extortion dimension – of incentive pay: wages are set high enough that the workforce will not sabotage production, refuse to work with new recruits, and so forth.

Lastly, trade unions. To Europeans it must seem a sort of wilfulness to have left them out. As I commented earlier, unions had been overemphasised in the preceding years and, as a reaction, I took the position that unions were inessential for explaining why equilibrium unemployment was substantial. Since then, of course, a number of models of unemployment centred on trade unions have appeared. Models of the insider–outsider type in which unions figure have already been mentioned. With the (1991) monograph by Richard Layard and his colleagues has come another model of wage-setting and unemployment based on bargaining between unions and imperfectly competitive firms – a sort of Nash–Zeuthen model focusing on unemployment. Yet these union models do not incorporate considerations of incentive pay, just as incentive models do not have unions.

We are stuck with the fact that every tractable model can offer only a slice of reality. I feel fortunate to have seized on a slice that, despite so many years, has gone on yielding rich insights into the determination of unemployment.

Notes

- 1 The process of rising inflation would cause the rate finally to hit a barrier such that sellers of money would find no takers, money having ceased to play any useful role. The process of rising deflation would cause the rate finally to reach a level such that sellers of real wealth claims would be unable to find any takers, money having come to dominate all non-monetary assets such as equities and consumer durables.
- 2 This definition of the natural rate seems to be very convenient in econometric work. In a purely theoretical paper it might be more natural (no pun intended)

- to define the natural rate as the equilibrium steady state path.
- 3 Clearly the thrust of the doctrine would survive if the natural rate became a *path* instead of a constant, even an endogenous path provided it is invariant to monetary policy.
 - 4 Dahrendorf would say that the proposition belongs to critical rather than to theoretical economics; see Dahrendorf (1993).
 - 5 These papers were widely known. Lerner's, in the *Review of Economics and Statistics* (1949), was the stimulus to Friedman's famous complaint that, contrary to what Lerner had implied, inflation had *some* real allocative maleffects, namely the time-consuming efforts of people to economise on cash balances.
 - 6 Among those who zeroed in on this aspect of the model were Dale Mortenson and some of his colleagues at Northwestern, Donald Gordon, then at UBC, and Arthur Okun at Brookings.
 - 7 The first wave of new Keynesian models, developed at Columbia in the latter half of the 1970s, is discussed in Phelps (1991).
 - 8 I could not be certain that the distinction was really there were it not that I remember some readers having called attention to it, William Nordhaus for one.
 - 9 Another effect works the other way, as the deficiency of capital relative to the long run means that non-wage income is decreased relative to the wage, which tends to decrease the propensity to shirk in the model and thus to reduce the supply wage (or incentive wage). As a consequence of this contrary tendency, there is no theoretically unambiguous result regarding the unemployment effect of a permanent shock to the level of labour augmentation and of a change in the initial capital stock.
 - 10 Appendix A to the 1970 version of my paper has some equations bearing on the dynamics of employment under the explicit presence of rising marginal hiring cost. Without doubt that discussion is somewhat confused, having been written in late 1966, well before the significance and real meaning of the model I was working on had become more or less fully clear to me.

References

- Dahrendorf, R., 1993. *Essays in the Theory of Society*, Stanford, CA: Stanford University Press
- Fellner, W. J., 1959. 'Demand Inflation, Cost Inflation, and Collective Bargaining', in P.D. Bradley (ed.), *The Public Stake in Union Power*, Charlottesville: University of Virginia Press
- Fitoussi, J.-P. and Phelps, E.S., 1988. *The Slump in Europe*, Oxford: Basil Blackwell
- Friedman, M., 1968. 'The Role of Monetary Policy', *American Economic Review*, 58(1) (March), 1-17
- Layard, R., Nickell, S.J. and Jackman, R., 1991. *Unemployment: Macroeconomic Performance and the Labour Market*, Oxford: Oxford University Press
- Lerner, A. P., 1948, 'The Inflationary Process - Some Theoretical Aspects', *Review of Economics and Statistics*, 31 (August).

- Lindbeck, A. and Snower, D.J., 1989. *The Insider–Outsider Theory of Employment and Unemployment*, Cambridge, MA: MIT Press
- Lipsey, R.G., 1960. 'The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1862–1957: a Further Analysis', *Economica*, 27 (February), 1–31, reprinted in AEA Series (1966), *Readings in Business Cycles*, 456–87
- Lucas, R.E., 1972. 'Expectations and the neutrality of money', *Journal of Economic Theory*, 4 (April), 103–24
- Phelps, E.S., 1967. 'Phillips Curves, Expectations of Inflation and Optimal Unemployment Over Time', *Economica*, 34(3) (August), 254–81
1968. 'Money Wage Dynamics and Labor Market Equilibrium', *Journal of Political Economy*, 76 (August), part 2, 678–711
1970. 'Money Wage Dynamics and Labor Market Equilibrium', in E.S. Phelps *et al.*, *Microeconomic Foundations of Employment and Inflation Theory*, New York: Norton and London: Macmillan (1971)
1972. *Inflation Policy and Unemployment Theory*, London: Macmillan
1991. *Seven Schools of Macroeconomic Thought*, Oxford: Oxford University Press
1994. *Structural Slumps: the Modern Equilibrium Theory of Unemployment, Interest and Assets*, Cambridge, MA: Harvard University Press
- Salop, S.C., 1979. 'A Model of the Natural Rate of Unemployment', *American Economic Review*, 69 (May), 117–25
- Sargan, D., 1964. 'Wages and Prices in the UK: a Study in Econometric Methodology', in P.E. Hart, G. Mills and J.K. Whitaker (eds), *Econometric Analysis for National Economic Planning*, London: Butterworth
- Shapiro, C. and Stiglitz, J.E., 1984. 'Equilibrium Unemployment as a Worker Discipline Device', *American Economic Review*, 74 (June), 433–44