

Part 4: Short term fluctuations

Chapters

14 Real Business Cycles.....	205
15 (New) Keynesian Theories of Fluctuations: A Primer.....	219
16 Unemployment.....	243



Filipe Campante,
Federico Sturzenegger
and Andrés Velasco

Advanced Macro- economics

LSE Press

Suggested citation: Campante, Filipe; Sturzenegger, Federico; and Velasco, Andrés. (2021) 'Part 4: Short term fluctuations', in Filipe Campante, Federico Sturzenegger and Andrés Velasco, *Advanced Macro-Economics: An Easy Guide*. London: LSE Press, 2021. Chapters 14–16. <https://doi.org/10.31389/lsepress.ame.4>



Filipe Campante, Federico Sturzenegger and Andrés Velasco 2021. Published under a Creative Commons Attribution NonCommercial licence allowing users to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license does not allow for commercial use.

Short Term Fluctuations

Real business cycles

So far we have mostly talked about long-term dynamics, the process of capital accumulation, inter-generational issues, etc. However, a lot of macroeconomics focuses on the short term – the departures from the long-run trend that we’ve been mostly concerned about. This is, of course, particularly evident in recession times! Some of the biggest questions in macroeconomics revolve around this: how can we understand and influence the short-run, cyclical evolution of the economy? What can we (or should we) do about recessions?

These are obviously important questions, and they are very much at the heart of the development of macroeconomics as a discipline, as we discussed in the first chapter of the book. In fact, business cycles is where the distinction between macroeconomic schools of thought became more evident – giving credence to the idea that economists never agree with each other. Many of your policy recommendations will derive from which view of the world you have.

Essentially, one school is ingrained in the Keynesian perspective where there is scope for intervening on the cycle and that doing so is welfare-improving. Its modern version is the New Keynesian approach originated in the 1980s in response to the empirical and methodological challenges from the 1970s. The second approach is quite skeptical about what policy can or should do, as it views the cycle as the result of optimal adjustments to real shocks. Its modern version was born, also in the 1980s, with the so-called Real Business Cycle (RBC) framework, which argued that a perfectly competitive economy, with no distortions or aggregate imbalances of the Keynesian type, but subject to productivity shocks, could largely replicate the business-cycle frequency data for real-world economies.

Recent years have seen a great deal of methodological convergence, with both views adopting, to a large extent, the so-called dynamic stochastic general equilibrium (DSGE) framework that essentially implements the NGM with whatever exogenous shocks and market imperfections that you may feel are relevant. Because this model can be specified to work as a perfectly competitive distortion-free economy, or as one with more Keynesian-type characteristics, it has become the new workhorse of macroeconomics. This has allowed for a more unified conversation in recent decades.

In light of that, and because we have covered much of the ground when we studied the NGM, we will start by describing the RBC framework, which started the trend that turned the NGM into the workhorse of modern macroeconomics. This framework, from a theory standpoint, is, conceptually, a simple extension of the NGM to a context with stochastic shocks.

How to cite this book chapter:

Campante, F., Sturzenegger, F. and Velasco, A. 2021. *Advanced Macroeconomics: An Easy Guide*.

Ch. 14. ‘Real business cycles’, pp. 205–218. London: LSE Press.

DOI: <https://doi.org/10.31389/lsepress.ame.n> License: CC-BY-NC 4.0.

Yet we will see that it has a sharp message. In particular, it delivers the ultimate anti-Keynesian statement: to the extent that business cycles are optimal responses to productivity shocks, policy interventions are worse than ineffective; they are bad, because they deviate the economy from its intertemporal optimum. From this benchmark, we will then turn to the Keynesian theories.

14.1 | The basic RBC model

The basic RBC model is simply the NGM (in discrete time) with two additions: stochastic shocks to productivity (to generate fluctuations in output) and a labour supply choice (to generate fluctuations in employment). Fluctuations come from optimal individual responses to the stochastic shocks.

The basic RBC model, first introduced by Kydland and Prescott (1982), is built around a typical NGM framework of intertemporal maximisation. There are three differences with what we've seen so far in the book. First, we introduce uncertainty in the form of exogenous productivity shocks, without which (as we've seen) no fluctuations emerge. Second, we also introduce a choice of how much labour will be supplied – in other words, there is a labour-leisure choice. This is what will enable us to say something about fluctuations in employment. Finally, RBC models typically use discrete time. This is so because the objective is to compare the simulated data from the model with that of the real data, which is always discrete, and also because the models quickly become too complicated for analytical solutions. One has to resort to numerical methods of solution, and computers can more easily handle discrete data.

The consumer's problem works as follows:

$$\text{Max } E \left[\sum_t \left(\frac{1}{1 + \rho} \right)^t ((1 - \phi) u(c_t) + \phi v(h_t)) \right], \quad (14.1)$$

subject to the household budget constraint in which individuals own the capital stock and labour endowment, and rent those out to the firms,

$$k_{t+1} = l_t w_t + (1 + r_t) k_t - c_t, \quad (14.2)$$

the production function,

$$f(k_t, l_t, z_t) = z_t k_t^\alpha l_t^{1-\alpha}, \quad (14.3)$$

the labour endowment equation,

$$h_t + l_t = 1, \quad (14.4)$$

and a productivity shock process

$$z_{t+1} = \varphi z_t + \varepsilon_{t+1}. \quad (14.5)$$

c_t is consumption, h_t indicates leisure, r_t is the rate of return on capital (net of depreciation), k_t is the capital stock, l_t is the amount of labour devoted to market activities.¹ Finally, z_t is a productivity parameter which is subject to random shocks ε_t . The rest are parameters which should be relatively self-explanatory.

14.1.1 | The importance of labour supply

As we've pointed out, one of the RBC literature's main departures from the standard NGM is the presence of a labour supply choice. This is crucial to generate fluctuations in employment, which are a pervasive feature of actual business cycles. Let us consider this choice within the context of the basic model. With log utility, the consumer's objective function can be thought of as:

$$E \left[\sum_t \left(\frac{1}{1 + \rho} \right)^t [(1 - \phi) \log(c_t) + \phi \log(h_t)] \right].$$

Notice that the household has two control variables, consumption and leisure. We have seen before the solution to the problem of optimal intertemporal allocation of consumption; it is the familiar Euler equation:

$$u'(c_t) = \frac{1 + r_{t+1}}{1 + \rho} E [u'(c_{t+1})]. \quad (14.6)$$

Leaving aside uncertainty, for the moment, and using the log assumption, we can rewrite this as:

$$c_{t+1} = \frac{1 + r_{t+1}}{1 + \rho} c_t. \quad (14.7)$$

The labour-leisure choice, in contrast, is static; it takes place in each period with no implication for the next period. The FOC equates the marginal benefit of additional consumption to the marginal cost of lost leisure:

$$w_t(1 - \phi)u'(c_t) = \phi v'(h_t). \quad (14.8)$$

Again using the log utility assumption, we get

$$w_t(1 - \phi)h_t = \phi c_t. \quad (14.9)$$

To simplify things further, assume for the moment that r is exogenous – think of a small open economy. In this setup, we can use (14.9) into (14.7) to obtain

$$\frac{h_{t+1}}{h_t} = \frac{1 + r}{1 + \rho} \frac{w_t}{w_{t+1}} \Rightarrow \frac{1 - l_{t+1}}{1 - l_t} = \frac{1 + r}{1 + \rho} \frac{w_t}{w_{t+1}}. \quad (14.10)$$

This means that leisure will be relatively high in those periods when the wage is relatively low; in other words, a higher wage increases the supply of labour. We can also see the impact of the interest rate: a high interest rate will lead to a higher supply of labour. The intuition is that it is worth working more in the present the higher the interest rate, as it provides a higher return in terms of future leisure. These responses of the labour supply, driven by intertemporal substitution in labour and leisure, are at the very heart of the fluctuations in employment in RBC models.

The long-run labour supply

Let's pause for a minute to explore a bit deeper the shape of this labour supply curve. Consider the case when wages and income are constant, a case that would be akin to analysing the effect of permanent shocks to these variables. Let's see the form of the labour supply in this case.

Since consumption is constant at income level

$$c_t = w(1 - h_t), \quad (14.11)$$

substituting this into (14.8) we obtain

$$(1 - \phi) u'(w(1 - h_t)) = \frac{\phi v'(h_t)}{w}. \quad (14.12)$$

Using the log specification for consumption and allowing for $\phi = \frac{2}{3}$ allows us to simplify

$$\left(\frac{1}{3}\right) \frac{1}{w(1 - h_t)} = \frac{2}{3} \frac{1}{wh_t}, \quad (14.13)$$

$$\frac{1}{(1 - h_t)} = \frac{2}{h_t} \Rightarrow h_t = \frac{2}{3}. \quad (14.14)$$

This is a strong result that says that leisure is independent of the wage level. You may think this is too strong but, surprisingly, it fits the data very well, at least when thinking about labour supply in the very, very long run.² It does seem that for humans income and substitution effects just cancel out (or maybe you prefer a more Leontieff setup in which people just can't work more than eight hours per day, or where the disutility of labour beyond eight hours a day becomes unbearable if repeated every day).

Does this mean that labour supply does not move at all? Not really. The above was derived under the assumption of the constancy of the wage. This is akin to assuming that any change in wages is permanent, which induces a very large response in the shadow value of consumption that works to offset the labour supply effect of the change in wages (totally cancelling it in the log case). But if the wage moves for a very short period of time we can assume the shadow value of consumption to remain constant, and then changes in the wage will elicit a labour supply response. Thus, while the long-run elasticity of labour supply may be zero, it is positive in the short run.

The basic mechanics

In its essence, the RBC story goes as follows: consider a positive productivity shock that hits the economy, making it more productive. As a result of that shock, wages (i.e. MPL) and interest rates (i.e. MPK) go up, and individuals want to work more as a result. Because of that, output goes up. It follows that the elasticity of labour supply (and the closely related elasticity of intertemporal substitution) are crucial parameters for RBC models. One can only obtain large fluctuations in employment, as needed to match the data, if this elasticity is sufficiently high. What is the elasticity of labour supply in this basic model? Consider the case when $\frac{(1+r)}{(1+\rho)} = 1$, in which consumption is a constant. We can read (14.8) as implying a labour supply curve (a relation between l_t and w_t):

$$\phi v'(1 - l_t) = \lambda w_t, \quad (14.15)$$

where λ is the (constant) marginal utility of consumption. Let's assume a slightly more general, functional form for the utility of leisure:

$$v(h) = \frac{\sigma}{1 - \sigma} h^{\frac{\sigma-1}{\sigma}}, \quad (14.16)$$

plugging this in (14.15) gives

$$\phi h_t^{-\frac{1}{\sigma}} = \lambda w_t \quad (14.17)$$

or

$$h_t = \left(\frac{\lambda w_t}{\phi} \right)^{-\sigma}, \quad (14.18)$$

which can be used to compute the labour supply:

$$l_t = 1 - \left(\frac{\lambda w_t}{\phi} \right)^{-\sigma}. \quad (14.19)$$

This equation has a labour supply elasticity in the short run equal to

$$\frac{dl}{dw} \frac{w}{l} = \varepsilon_{l,w} = \frac{\sigma \left(\frac{\lambda w_t}{\phi} \right)^{-\sigma-1} \left(\frac{\lambda w_t}{\phi} \right)}{1 - \left(\frac{\lambda w_t}{\phi} \right)^{-\sigma}} = \frac{\sigma \left(\frac{\lambda w_t}{\phi} \right)^{-\sigma}}{1 - \left(\frac{\lambda w_t}{\phi} \right)^{-\sigma}} = \frac{\sigma h_t}{l_t}. \quad (14.20)$$

If we assume that $\sigma = 1$ (logarithmic utility in leisure), and that ϕ and λ are such that $\frac{h}{l} = 2$ (think about an 8-hour workday), this gives you $\varepsilon_{l,w} = 2$. This doesn't seem to be enough to replicate the employment fluctuations observed in the data. On the other hand, it seems to be quite high if compared to micro data on the elasticity of labour supply. Do you think a decrease of 10% in real wages (because of inflation, for instance) would lead people to work 20% fewer hours?

14.1.2 | The indivisible labour solution

The RBC model thus delivers an elasticity of labour supply that is much higher than what micro evidence suggests, posing a challenge when it comes to matching real-world fluctuations in employment. One proposed solution for the conundrum is to incorporate the fact that labour decisions are often indivisible. This means that people may not make adjustments so much on the intensive margin of how many hours to work in your job, but more often on the extensive margin of whether to work at all. This implies that the aggregate elasticity is large even when the individual elasticity is small.

Hansen (1985) models that by assuming that there are fixed costs of going to work. This can actually make labour supply very responsive for a range of wage levels. The decision variables are both days of work: $d \leq \bar{d}$, and, then, the hours of work each day: n . We assume there is a fixed commuting cost in terms of utility κ , which you pay if you decide to work on that day, regardless of how many hours you work (this would be a sort of commuting time).

The objective function is now

$$\text{Max} E \left[\sum_t \left(\frac{1}{1+\rho} \right)^t [u(c_t) - d_t v(n_t) - \kappa_t d_t] \right], \quad (14.21)$$

where we leave aside the term ϕ to simplify notation, and abuse notation to have $v(\cdot)$ be a function of hours worked, rather than leisure, entering negatively in the utility function. The budget constraint is affected in that now wage income is equal to $w_t d_t n_t$.

It is easy to see that we have the same FOCs, (14.7) – which is unchanged because the terms in consumption in both maximand and budget constraint are still the same –, and (14.8) – because the term in n_t is multiplied by d_t in both maximand and budget constraint, so that d_t cancels out. What changes is that now we have an extra FOC with respect to d_t :

$$[v(n_t) + k_t] \geq u'(c_t) w_t n_t. \quad (14.22)$$

Assume $\frac{(1+r)}{(1+\rho)} = 1$, so that c_t is constant. Then (14.8) simplifies to

$$v'(n_t) = \lambda w_t \implies n^*(w), \quad (14.23)$$

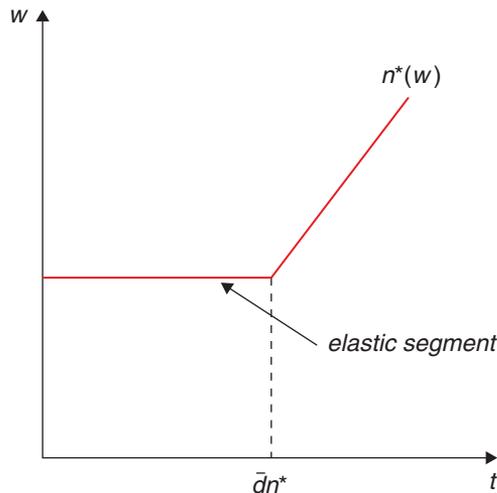
which gives the optimal amount of hours worked (when the agent decides to work). Then (14.22) simplifies to

$$v(n^*) + k_t \geq \lambda w_t n^*. \quad (14.24)$$

If $v(n^*) + k > \lambda w n^*$, then $d = 0$, otherwise $d = \bar{d}$. This gives rise to a labour supply as shown in Figure 14.1

The important point is that this labour supply curve is infinitely elastic at a certain wage. The intuition is that on the margin at which people decide whether to work at all or not, the labour supply will be very sensitive to changes in wages.³

Figure 14.1 The Hansen labour supply



14.2 | RBC model at work

RBC models typically cannot be solved analytically, and require numerical methods. We discuss an example of a calibration approach to assess the success of the model in describing actual economic fluctuations.

Having discussed the basic intuition behind generating output and employment fluctuations from real shocks to the economy, let us now talk a little bit more generally about how RBC models are usually handled. The main challenge is that even simple specifications are impossible to solve analytically, so the alternative is to use numerical methods.

How is this done? In a nutshell, the strategy is to solve for the FOCs of the model which, in addition to the equations determining the nature of the stochastic shocks, will describe the dynamic path of the variables of interest. (This will often imply a step in which the FOCs are log-linearised around the balanced growth path, since it is easiest to analyse the properties of a linear system.) We then need to provide numbers for the computer to work with – this is done by calibrating the parameters. (Remember what we have discussed of calibration when talking about growth empirics – this approach was pretty much pioneered by the RBC literature.) Because the model is calibrated on the basis of parameters that are brought from “outside the model”, the procedure provides somewhat of an independent test of the relevance of the model.

With this in hand, the model is simulated, typically by considering how the variables of interest responds after being exposed to a stochastic draw of exogenous productivity shocks. The results are then compared to the data. In both cases, the simulated and the real data, we work with the business cycle component, i.e. detrending the data. This is usually done using the Hodrick-Prescott filter, which is a statistical procedure to filter out the trend component of a time series. What output of the model is then compared to the data? Mostly second moments: variances and covariances of the variables of interest. A model is considered successful if it matches lots of those empirical moments.

14.2.1 | Calibration: An example

Let us consider the basic RBC model, and the calibration proposed by Prescott (1986) which is the actual kick-off of this approach and where Prescott tackles the issue of assigning parameters to the coefficients of the model. For example, at the time, he took as good a capital share of $\alpha = 0.36$.⁴ To estimate the production function, he starts with a Cobb-Douglas specification we’ve used repeatedly

$$f(k) = k^\alpha. \quad (14.25)$$

Remember that the interest rate has to equal the marginal product of capital,

$$f'(k) = \alpha k^{\alpha-1}, \quad (14.26)$$

which means that we have an equation for the return on capital:

$$r = \alpha \frac{Y}{K} - \delta. \quad (14.27)$$

Now let’s put numbers to this. What is a reasonable rate of depreciation? Let’s use (14.27) itself to figure it out. If we assume that the rate of depreciation is 10% per year (14.27) becomes

$$0.04 = 0.36 \frac{Y}{K} - 0.10 \quad (14.28)$$

$$0.14 = 0.36 \frac{Y}{K} \quad (14.29)$$

$$\frac{0.36}{0.14} = \frac{K}{Y} = 2.6. \quad (14.30)$$

This value for the capital output ratio is considered reasonable, so the 10% rate of depreciation seems to be a reasonable guess.

How about the discount factor? It is assumed equal to the interest rate. (This is not as restrictive as it may seem, but we can skip that for now.) This implies a yearly discount rate of about 4% (the real interest rate), so that $\frac{1}{1+\rho} = 0.96$ (again, per year).

As for the elasticity of intertemporal substitution, he argues that $\sigma = 1$ is a good approximation, and uses the share of leisure equal to $2/3$, as we had anticipated (this gives a labour allocation of half, which is reasonable if we consider that possible working hours are 16 per day).

Finally, the productivity shock process is derived from Solow-residual-type estimations (as discussed in Chapter 6 when we talked about growth accounting), which, in the case of the U.S. at the time, yielded:

$$z_{t+1} = 0.9 * z_t + \varepsilon_{t+1}. \quad (14.31)$$

This is a highly persistent process, in which shocks have very long-lasting effects. The calibration for the standard deviation of the disturbance ε is 0.763.

So, endowed with all these parameters, we can pour them into the specification and run the model over time – in fact, multiple times, with different random draws for the productivity shock. This will give a time series for the economy in the theoretical model. We will now see how the properties of this economy compare to those of the real economy.

14.2.2 | Does it work?

Let's start with some basic results taken directly from Prescott's paper. Figure 14.2 shows log U.S. GDP and its trend. The trend is computed as a Hodrick-Prescott filter (think of this as a smoothed, but not fixed, line tracing the data). It is not a great way to compute the business cycle (particularly at the edges of the data set), but one that has become quite popular. Once the trend is computed, the cycle is easily estimated as the difference between the two and is showing in figure 14.3.

Figure 14.3 also shows the variation over the cycle in hours worked. As you can see, there is a large positive correlation between the two.

Real business cycle papers will typically include a table with the properties of the economy, understood as the volatility of the variables and their cross-correlation over time. Table 14.1 and 14.2 show this from Prescott's original paper for both the real data and the calibrated model.

As you can see, things work surprisingly well in the sense that most characteristics of the economy match. The volatility of output and the relative volatility of consumption and investment appear to be the optimal response to the supply shocks. The only caveat is that hours do not seem to move as much as in the data. This is why Prescott implemented Hansen's extension. Figure 14.4 shows how labour and output move in the Hansen economy (they seem to match better the pattern in Figure 14.1).

The Appendix to this chapter (at the end of the book) will walk you through an actual example so that you learn to numerically solve and simulate an RBC-style model yourself!

Figure 14.2 The U.S. output

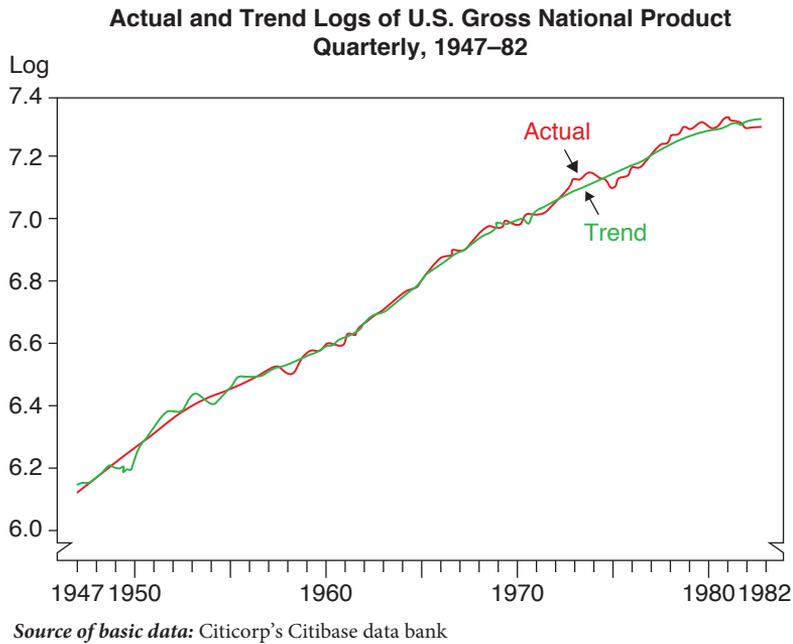


Figure 14.3 The cycle in the U.S.

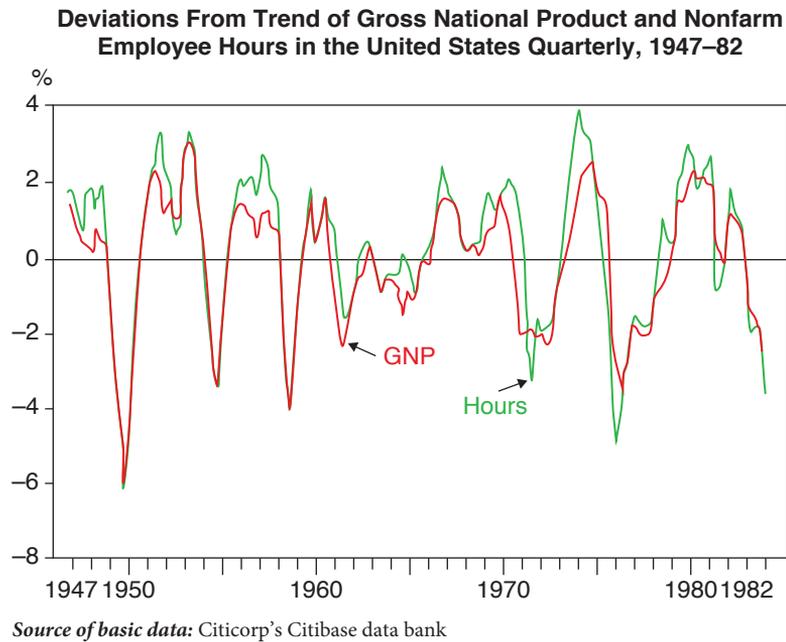


Table 14.1 The data for the U.S. cycle, from Prescott (1986)

Cyclical Behavior of the U.S. Economy Deviations From Trend of Key Variables. 1954:1–1982:4				
Variable x	Standard Deviation	Cross Correlation of GNP With		
		$x(t - 1)$	$x(t)$	$x(t + 1)$
Gross National Product	1.8%	.82	1.00	.82
Personal Consumption Expenditures				
Services	.6	.66	.72	.61
Nondurable Goods	1.2	.71	.76	.59
Fixed investment Expenditures	5.3	.78	.89	.78
Nonresidential Investment	5.2	.54	.79	.86
Structures	4.6	.42	.62	.70
Equipment	6.0	.56	.82	.87
Capital Stocks				
Total Nonfarm Inventories	1.7	.15	4.8	.68
Nonresidential Structures	.4	−.20	−.03	.16
Nonresidential Equipment	1.0	.03	.23	.41
Labor Input				
Nonfarm Hours	1.7	.57	.85	.89
Average Weekly Hours in Mfg.	1.0	.76	.85	.61
Productivity (GNP/Hours)	1.0	.51	.34	−.04

Source of basic data: Citicorp's Citibase data bank

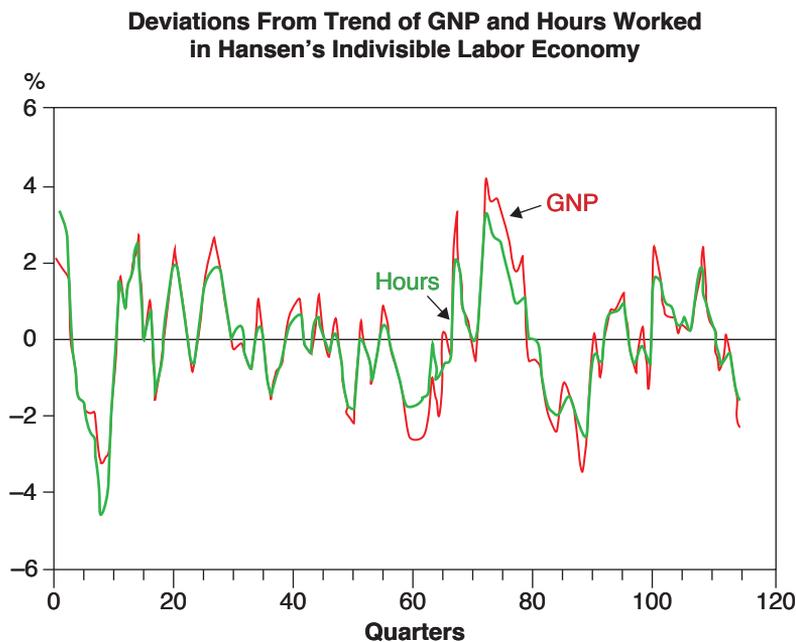
Table 14.2 The variables in the Prescott model, from Prescott (1986)

Cyclical Behavior of the Kydland-Prescott Economy*				
Variable x	Standard Deviation	Cross Correlation of GNP With		
		$x(t - 1)$	$x(t)$	$x(t + 1)$
Gross National Product	1.79%	.60	1.00	.60
	(.13)	(.07)	(—)	(.07)
Consumption	.45	.47	.85	.71
	(.05)	(.05)	(.02)	(.04)
Investment	5.49	.52	.88	.78
	(.41)	(.09)	(.03)	(.03)
Inventory Stock	2.20	.14	.60	.52
	(.37)	(.14)	(.08)	(.05)

Capital Stock	.47 (.07)	-.05 (.07)	.02 (.06)	.25 (.07)
Hours	1.23 (.09)	.52 (.09)	.95 (.01)	.55 (.06)
Productivity (GNP/Hours)	.71 (.06)	.62 (.05)	.86 (.02)	.56 (.10)
Real Interest Rate (Annual)	.22 (.03)	.65 (.07)	.60 (.20)	.36 (.15)

*These are the means of 20 simulations, each of which was 116 periods long. The numbers parentheses are standard errors.
 Source: Kydland and Prescott 1984

Figure 14.4 The correlation of output and hours in the Hansen model



Source: Gray D. Hansen, Department of Economics, University of California, Santa Barbara

14.3 | Assessing the RBC contribution

The RBC approach led to a methodological revolution in macroeconomics; all macro models from then on have been expected to be framed as a dynamic stochastic general-equilibrium model with fully optimising agents and rational expectations. Whether or not you buy it as an explanation for business cycle fluctuations in general, and the associated critique of policy interventions, the approach can be useful in understanding at least some aspects of actual fluctuations.

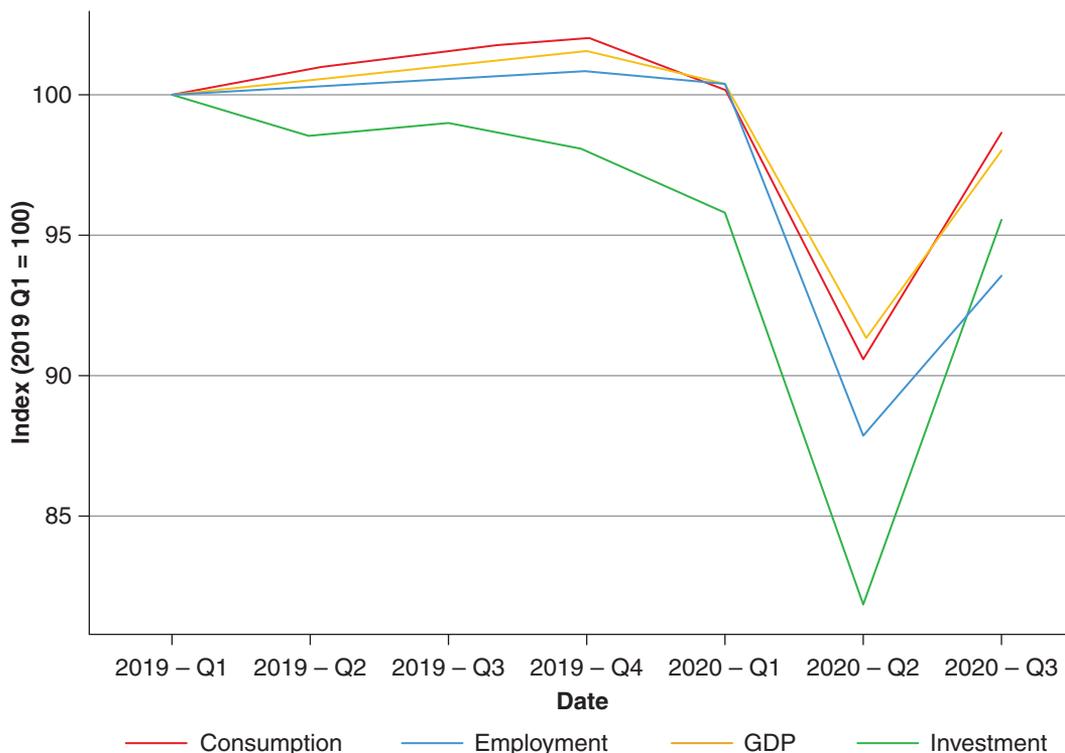
Prescott (1986) summarises the claim in favour of the RBC model: “Economic theory implies that, given the nature of the shocks to technology and people’s willingness and ability to intertemporally and intratemporally substitute, the economy will display fluctuations like those the U.S. economy displays.” His claim is that his model economy matches remarkably well the actual data and, to the extent that it doesn’t, it’s probably because the real-world measurement does not really capture what the theory says is important – hence the title of ‘Theory Ahead of Business Cycle Measurement’.

The startling policy implications of these findings are highlighted as follows: “Costly efforts at stabilization are likely to be counterproductive. Economic fluctuations are optimal responses to uncertainty in the rate of technological change.” In other words, business cycle policy is not only useless, but harmful. One should focus on the determinants of the average rate of technological change.

The macro literature has vigorously pursued and refined the path opened by Prescott. Lots of different changes have been considered to the original specification, such as different sources for the shocks (for instance, government spending) or the inclusion of a number of market distortions (e.g. taxation). On the other hand, many objections have been raised to the basic message of RBCs. Do we really see such huge shifts in technology on a quarterly basis? Or, is the Solow residual capturing something else? (Remember, it is the measure of our ignorance...) Do we really believe that fluctuations are driven by people’s willingness to intertemporally reallocate labour? If these are optimal, why do they feel so painful? How about the role of monetary policy, for which the RBC model has no role? Finally, it seems that the features of the fluctuations that are obtained are very close to the nature of the stochastic process that is assumed for the shocks – how much of an explanation is that?

Kydland and Prescott eventually received the Nobel Prize in Economics in 2004, partly for this contribution. (We will return to other contributions by the pair when we discuss monetary policy.) More importantly, the approach led to two developments. First, it generated a fierce counterattack to Keynesianism. The rational expectations revolution had stated that Keynesian policy was ineffective; Kydland and Prescott said it was wrong and harmful. Second, by validating the model, this calibrated stochastic version of the NGM became the workhorse of macroeconomics, so that the RBC approach won the methodological contest. In macro these days, people are pretty much expected to produce a DSGE model with fully optimising agents (with rational expectations) that is amenable to the discipline of calibration. Even the folks who believe in Keynesian-style business cycles are compelled to frame them in such models, though including some price rigidity, monetary disturbance, and so on, as we will see soon.

In addition, even if you believe that the Keynesian approach is a better description of business cycles in general, it may still be the case that a simple RBC framework can explain some important economic episodes. For instance, take a look at Figure 14.5, which depicts GDP, employment, consumption, and investment data for the U.S. over 2019–20. The sharp drop we see is, of course, the economic response to the Covid-19 pandemic, which fits well the supply shock paradigm. The response looks a lot like the kind of logic we have seen in this chapter: a shock to productivity – in this case, the threat posed by a virus – radically changes the intertemporal tradeoff, and leads to a postponement of labour supply and, consequently, to drops in employment and output. Notice that consumption in this case is not smoother than output even though investment is the most volatile variable, as the model predicts. Why is consumption more volatile in this context? Eichenbaum et al. (2020) provide an explanation. They make the realistic assumption that during the Covid-19 pandemic, people attached a risk of contagion to the act of consuming (consumption means going out to a restaurant, shopping mall, etc.) and, therefore, reduced consumption more than they would have done if only adjusting for the change in intertemporal wealth. The example illustrates that some specific changes in the setup may be required on occasion to adjust empirical observations.⁵

Figure 14.5 Trajectories of macro variables in response to Covid-19

This particular example illustrates that real shocks actually exist, and shows, more broadly, that you do not have to be an RBC true believer to accept that the logic they illuminate has practical applications.

14.4 | What have we learned?

The RBC approach to business cycle fluctuations is conceptually very straightforward; take the basic NGM model, add productivity shocks (and a labour-supply choice), and you will get business cycle fluctuations. It highlights the importance of intertemporal substitution and labour supply elasticities as important potential driving factors behind these fluctuations, and can provide a useful lens with which to understand real-world fluctuations, at the very least, in some circumstances (as illustrated by the case of the Covid-19 pandemic).

The approach also has a very sharp message in terms of policy: you should not pursue counter-cyclical policy. If fluctuations are simply the optimal response of a distortion-free economy to real shocks, policy would only add noise to the process, and make adjustments harder. As we will see, the Keynesian approach has a very different, more policy-friendly message. The contraposition of these two traditions – and particularly the role they assign to policy intervention – is very much at the heart of macroeconomic policy debates.

But we also learned that, underpinning this policy divergence, is a substantial degree of methodological convergence. All of mainstream modern macroeconomics, to a first approximation, speaks

the language that was first introduced by the RBC approach – that of dynamic, stochastic general equilibrium (DSGE) models.

14.5 | What next?

Readers who are interested in the RBC approach can go to McCandless (2008), *The ABCs of RBCs*, which, as the title indicates, provides a simple and practical introduction to solving RBC models. The paper by Prescott (1986) is also worth reading, as it provides a veritable manifesto of the original approach.

Those who want to dig deeper into the type of recursive methods that have become ubiquitous in macroeconomics, to solve models in the methodological tradition inaugurated by the RBC approach, should look into Ljungqvist and Sargent (2018). You will actually see many of the themes that we discuss in this book, but presented at a whole other level of formal rigor.

Notes

¹ You will often see l used to refer to leisure and n to labour, but we are going to stick with l for labour, for consistency. Think about h as standing for holidays.

² One way to think about this is asking yourself the following question: how many times higher are real wages today than, say, 300 years ago? And how many more hours do we work?

³ Here is a weird little prediction from this model: note that consumption is constant regardless of the employment decision. This means that unemployed and employed workers have the same consumption. But, since work generates disutility, that means that unemployed workers are better off! For a discussion on the state-of-the-art of this debate, see Chetty et al. (2011).

⁴ Remember that Parente and Prescott (2002) argue for $\alpha = .66$, but this was later...

⁵ An alternative story is provided by Sturzenegger (2020). In his specification, *utility* changes and people, due to lockdowns, require fewer goods to obtain the same utility. The result is a sharp fall in optimal consumption as in Eichenbaum et al. (2020).

References

- Chetty, R., Guren, A., Manoli, D., & Weber, A. (2011). Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins. *American Economic Review*, 101(3), 471–75.
- Eichenbaum, M. S., Rebelo, S., & Trabandt, M. (2020). *The macroeconomics of epidemics* (tech. rep.). National Bureau of Economic Research.
- Hansen, G. D. (1985). Indivisible labor and the business cycle. *Journal of Monetary Economics*, 16(3), 309–327.
- Kydland, F. E. & Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica*, 1345–1370.
- Ljungqvist, L. & Sargent, T. J. (2018). *Recursive macroeconomic theory*. MIT Press.
- McCandless, G. (2008). *The ABCs of RBCs*. Cambridge, Massachusetts, London: Harvard.
- Parente, S. L. & Prescott, E. C. (2002). *Barriers to riches*. MIT Press.
- Prescott, E. C. (1986). Theory ahead of business-cycle measurement. *Carnegie-Rochester Conference Series on Public Policy*, 25, 11–44.
- Sturzenegger, F. (2020). Should we hibernate in a lockdown? *Economics Bulletin*, 40(3), 2023–2033.

(New) Keynesian theories of fluctuations: A primer

Keynesian thinking starts from a different viewpoint, at least compared with that of the RBC approach, regarding the functioning of markets. In this perspective, output and employment fluctuations indicate that labour markets, good markets, or both, are not working, leading to unnecessary unemployment. The idea is that, at least in some circumstances, the economy is demand-constrained (rather than supply-constrained), so that the challenge is to increase expenditure. If that could be done, then supply will respond automatically. (This is Keynes's Principle of Effective Demand.¹) As a result, Keynesian models focus on aggregate demand management as opposed to supply-side policies. Later on in the book we will discuss specifically the role of fiscal and monetary policy in aggregate demand, but in this chapter we need to understand the framework under which this aggregate demand management matters.

Of course there is a lot of controversy among economists as to how is it possible that a situation where markets fail to clear may persist over time. Why is there unemployment? Can unemployment be involuntary? If it is involuntary, why don't people offer to work for less? Why are prices rigid? Why can't firms adjust their prices? How essential is price fixing in comparison with distortions on the labour market? And, in this setup, do consumers satisfy their intertemporal budget constraints?

These are difficult questions that have led to a large amount of literature trying to develop models with Keynesian features in a microfounded equilibrium framework with rational expectations. This line of work that has been dubbed New Keynesianism, emerged as a reaction to the challenge posed by the New Classical approach. Over time, and as New Classical thinking evolved into the RBC approach, the literature coalesced around the so-called DSGE models – with the New Keynesian literature building on these models while adding to them one or many market imperfections.

In any event, this is a very broad expanse of literature that we will not be able to review extensively here. We will thus focus on three steps. First, we will revisit the standard IS-LM model. This model captures the essence (or so most economists think) of the Keynesian approach, by imposing the assumption of price rigidities, which gives rise to the possibility of aggregate demand management. This simple approach, however, begs the question of what could explain those rigidities. Our second step therefore will be to provide a brief discussion of possible microfoundations for them. There may be many reasons for why a nominal price adjustment is incomplete: long-term client relationships, staggered price adjustment, long-term contracts, asymmetric information, menu costs, etc. Not all of

How to cite this book chapter:

Campante, F., Sturzenegger, F. and Velasco, A. 2021. *Advanced Macroeconomics: An Easy Guide*.

Ch. 15. '(New) Keynesian theories of fluctuations: A primer', pp. 219–242. London: LSE Press.

DOI: <https://doi.org/10.31389/lsepress.ame.o> License: CC-BY-NC 4.0.

these lead to aggregate price rigidities, but we will not get into these details here. We will instead focus on a model where asymmetric information is the reason for incomplete nominal adjustments. This microfounded model highlights, in particular, the role of expectations in determining the reach of aggregate demand management.² Last but not least, we will see how these microfoundations combine to give rise to the modern New Keynesian DSGE models, which reinterpret the Keynesian insights in a rather different guise, and constitute the basis of most of macroeconomic policy analysis these days. (We include at the end of the book an appendix that takes you through the first steps needed so that you can run a DSGE model of your own!)

With at least some analytical framework that makes sense of the Keynesian paradigm and its modern interpretation, in later chapters we will discuss the mechanisms and policy levers for demand management, with an emphasis on monetary policy and fiscal policy.

15.1 | Keynesianism 101: IS-LM

We revisit the basic version of the Keynesian model that should be familiar from undergraduate macroeconomics: the IS-LM model.

In 1937, J.R. Hicks³ provided a theoretical framework that can be used to analyse the *General Theory* Keynes had published the previous year. Keynes's book had been relatively hard to crack, so the profession embraced Hicks's simple representation that later became known as the IS-LM, model and went on to populate intermediate macro textbooks ever since (Hicks won the Nobel Prize in Economics in 1972 for this work). While much maligned in many quarters (particularly because of its static nature and lack of microfoundations), this simple model (and its open-economy cousin, the Mundell-Fleming model) is still very much in the heads of policy makers.

The model is a general equilibrium framework encompassing three markets: goods, money and bonds, though only two are usually described as the third will clear automatically if the other two do (remember Walras Law from micro!). It is standard to represent the model in terms of interest rates and output, and to look at the equilibrium in the money and goods market. The corresponding equations are:

A money market equilibrium locus called the *LM* curve:

$$\frac{M}{P} = L \left(\overset{(-)}{i}, \overset{(+)}{Y} \right), \quad (15.1)$$

and a goods market equilibrium called the *IS* curve:

$$Y = A \left(\overset{(-)}{r}, \underbrace{\overset{(+)}{Y}}_{<1}, \overset{(+)}{\text{Fiscal}}, \overset{(+)}{RER} \right) \quad (15.2)$$

where *Fiscal* stands for government expenditures and *RER* for the real exchange rate, or, alternatively,

$$Y = A \left(\overset{(-)}{r}, \overset{(+)}{\text{Fiscal}}, \overset{(+)}{RER} \right). \quad (15.3)$$

Finally, a relationship between nominal and real interest rates:

$$r = i - \pi^e. \quad (15.4)$$

15.1.1 | Classical version of the IS-LM model

In the classical version of the model, all prices are flexible, and so are real wages. Thus the labour market clears fixing the amount of labour as in Figure 15.1.

With full employment of labor and capital, output is determined by the supply constraint and becomes an exogenous variable, which we will indicate with a bar:

$$\bar{Y} = F(K, \bar{L}). \quad (15.5)$$

The *IS*, can then be used to determine r so that savings equals investment ($S = I$, thus the name of the curve). The nominal interest rate is just the real rate plus exogenous inflation expectations (equivalent to the expected growth rate of prices). With Y and i fixed, then the *LM* determines the price level P given a stock of nominal money:

$$P = \frac{\bar{M}}{L(i, \bar{Y})}, \quad (15.6)$$

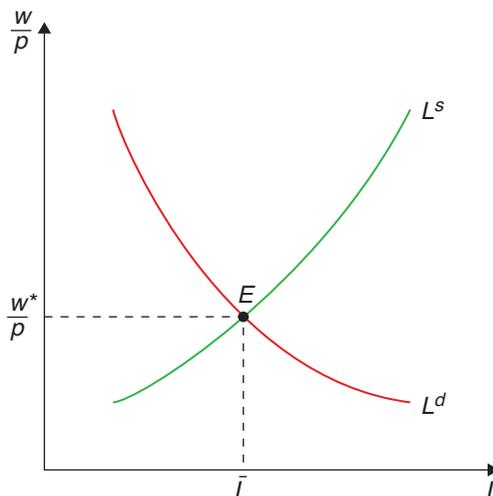
which is an alternative way of writing the quantity equation of money:

$$MV = PY. \quad (15.7)$$

In short, the structure of the model is such that the labor market determines the real wage and output. The *IS* determines the real and nominal interest rate, and the money market determines the price level.

This is typically interpreted as a description of the long run, the situation to which the economy gravitates at any given moment. The idea is that prices eventually adjust so that supply ends up determining output. That is why we ignored aggregate demand fluctuations when discussing long-run growth. There we concentrated on the evolution of the supply capacity. In the classical version of the model (or in the long run) that supply capacity determines what is produced.

Figure 15.1 The classical model



15.1.2 | The Keynesian version of the IS-LM model

However, as Keynes famously quipped, in the long run, we will all be dead. In the short run, the Keynesian assumption is that prices are fixed or rigid, and do not move to equate supply and demand:

$$P = \bar{P}, \quad (15.8)$$

so now the *IS* and *LM* curves jointly determine Y and i as in Figure 15.2.

Notice that Y is determined without referring to the labour market, so the level of labour demand may generate involuntary unemployment.

It is typical in this model to think of the effects of monetary and fiscal policy by shifting the *IS* and *LM* curves, and you have seen many such examples in your intermediate macro courses. (If you don't quite remember it, you may want to get a quick refresher from any undergraduate macro textbook you prefer.) We will later show how we can think more carefully about these policies, in a dynamic, microfounded context.

15.1.3 | An interpretation: The Fed

Is the model useful? Yes, because policy makers use it. For example, when the Fed talks about expanding or contracting the economy it clearly has a Keynesian framework in mind. It is true that the Fed does not typically operate on the money stock, but one way of thinking about how the Fed behaves is to think of it as determining the interest rate and then adjusting the money supply to the chosen rate (money becomes somewhat endogenous to the interest rate). In our model, i becomes exogenous and M endogenous as in Figure 15.3:

$$\frac{M}{\bar{P}} = L(\bar{i}, Y) \quad (15.9)$$

$$Y = A(\bar{i} - \pi^e, \text{Fiscal}, \dots). \quad (15.10)$$

Figure 15.2 The IS-LM model

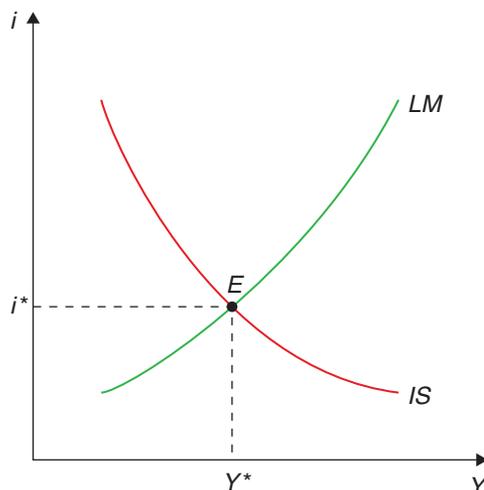
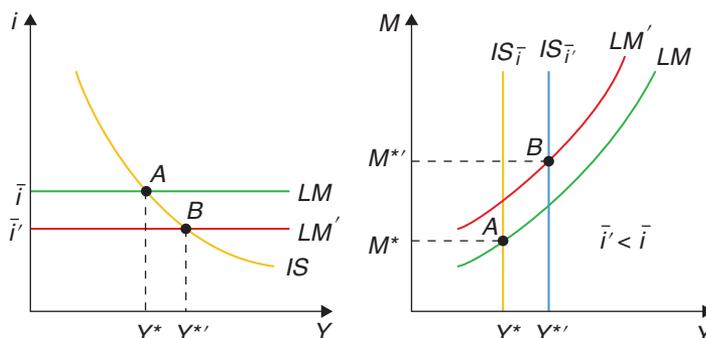


Figure 15.3 The IS-LM model with an exogenous interest rate

We can represent this using the same Y, i space, but the LM curve is now horizontal as the Fed sets the (nominal) interest rate. Alternatively, we can think about it in the Y, M space, since M is the new endogenous variable. Here we would have the same old LM curve, but now the IS curve becomes vertical in the Y, M space. Both represent the same idea: if the Fed wants to expand output, it reduces the interest rate, and this requires an expansion in the quantity of money.

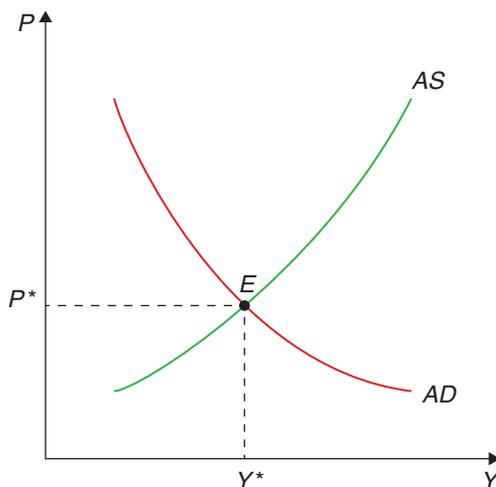
As a side note, you may have heard of the possibility of a liquidity trap, or alternatively, that monetary policy may hit the zero interest lower bound. What does this mean? We can think about it as a situation in which interest rates are so low that the demand for money is infinitely elastic to the interest rate. In other words, because nominal interest rates cannot be negative (after all, the nominal return on cash is set at zero), when they reach a very low point an increase in the supply of money will be hoarded as cash, as opposed to leading to a greater demand for goods and services. In that case, the IS-LM framework tells us that (conventional) monetary policy is ineffective. Simply put, interest rates cannot be pushed below zero!

This opens up a series of policy debates. There are two big questions that are associated with this: 1) Is monetary policy really ineffective in such a point? It is true that interest rate policy has lost its effectiveness by hitting the zero boundary, but that doesn't necessarily mean that the demand for money is infinitely elastic. The Fed can still pump money into the economy (what came to be known as quantitative easing) by purchasing government (and increasingly private) bonds, and this might still have an effect (maybe through expectations). 2) In this scenario, can fiscal policy be effective? These are debates we'll come back to in full force in our discussions of fiscal and monetary policy.

15.1.4 | From IS-LM to AS-AD

Another way to understand the assumption on price rigidity in generating a role for aggregate demand management is to go from the IS-LM representation to one in which P is one of the endogenous variables. The LM curve implies that an increase in prices leads to a decrease in the supply of real money balances, which shifts LM to the left. Since IS is not affected, that means that a higher P leads to a lower level of output Y . This is the aggregate demand (AD) curve in Figure 15.4.

An increase in aggregate demand (through monetary or fiscal policy) will shift the AD curve to the right. The effect that this will have on equilibrium output will depend on the effect of this on prices, which in turn depends on the aggregate supply (AS) of goods and services. The classical case is one in

Figure 15.4 AS-AD model

which this supply is independent of the price level – a vertical AS curve. This is the case where prices are fully flexible. The Keynesian case we considered, in contrast, is one in which AS is horizontal (P is fixed), and hence the shift in AD corresponds fully to an increase in output. This is an economy that is not supply constrained. In the intermediate case, where prices adjust but not completely, AS is positively sloped, and shifts in aggregate demand will have at least a partial effect on output.

The positively-sloped AS curve is the mirror image of the *Phillips curve* – the empirical observation of a tradeoff between output/unemployment and prices/inflation. We assume you are familiar with the concept from your intermediate macro courses, and we will get back to that when we discuss the modern New Keynesian approach.

15.2 | Microfoundations of incomplete nominal adjustment

We go over a possible explanation for the incomplete adjustment of prices or, more broadly, for why the AS curve may be upward-sloping. We study the Lucas model of imperfect information, which illustrates how we can solve models with rational expectations.

We have now reestablished the idea that, if prices do not adjust automatically, aggregate demand management can affect output and employment. The big question is, what lies behind their failure to adjust? Or, to put it in terms of the AS-AD framework, we need to understand why the AS curve is positively sloped, and not vertical. Old Keynesian arguments were built on things such as backward-looking (adaptive) expectations, money illusion, and the like. This, in turn, rubbed more classical-minded economists the wrong way. How can rational individuals behave in such a way? Their discomfort gained traction when the Phillips curve tradeoff seemed to break down empirically in the late 1960s and early 1970s. The model is also useful from a methodological point of view: it shows how a rational expectations model is solved. In other words, it shows how we use the model to compute the expectations that are an essential piece of the model itself. Let's see!

15.2.1 | The Lucas island model

The challenge to Keynesian orthodoxy, and hence the initial push from which modern Keynesian theories were built, took the shape of the pioneering model by Lucas (1973) – part of his Nobel-winning contribution. He derived a positively-sloped AS curve in a model founded at the individual level, and where individuals had rational expectations. This model also more explicit the role of expectations in constraining aggregate demand policy. The key idea was that of imperfect information: individuals can observe quite accurately the prices of the goods they produce or consume most often, but they cannot really observe the aggregate price level. This means that, when confronted with a higher demand for the good they produce, they are not quite sure whether that reflects an increase in its relative price – a case in which they should respond by increasing their output – or simply a general increase in prices – a case in which they should not respond with quantities, but just adjust prices. We will see that rational expectations implies that individuals should split the difference and attribute at least part of the increase to relative prices. (How much so will depend on how often general price increases occur.) This yields the celebrated *Lucas supply curve*, a positively-sloped supply curve in which output increases when the price increases *in excess of its expected level*.

The model is one with many agents (Lucas's original specification places each person on a different island, which is why the model is often referred to as the Lucas island model). Each agent is a consumer-producer that every period sees a certain level of demand. The basic question is to figure out if an increase in demand is an increase in real demand, which requires an increase in production levels, or if it is simply an increase in nominal demand, to which the optimal response is just an increase in prices. The tension between these two alternatives is what will give power to the model. In order to solve the model we will start with a specification with perfect information and, once this benchmark case is solved, we will move to the case of asymmetric information, which is where all the interesting action is.

15.2.2 | The model with perfect information

The representative producer of good i has production function

$$Q_i = L_i, \quad (15.11)$$

so that her feasible consumption is

$$c_i = \frac{P_i Q_i}{P}. \quad (15.12)$$

Utility depends (positively) on consumption and (negatively) on labour effort. Let's assume the specification

$$u_i = c_i - \frac{1}{\gamma} L_i^\gamma \quad \gamma > 1. \quad (15.13)$$

If P is known (perfect information), the problem is easy; the agent has to maximize her utility (15.13) with respect to her supply of the good (which is, at the same time, her supply of labour). Replacing (15.11) and (15.12) in (15.13) gives

$$u_i = \frac{P_i L_i}{P} - \frac{1}{\gamma} L_i^\gamma. \quad (15.14)$$

The first order condition for L is

$$\frac{P_i}{P} - L_i^{\gamma-1} = 0, \quad (15.15)$$

which can be written as a labour supply curve

$$L_i = \left(\frac{P_i}{P} \right)^{\frac{1}{\gamma-1}}, \quad (15.16)$$

or, if expressed in logs (denoted in lower case letters), as

$$l_i = \left(\frac{1}{\gamma-1} \right) (p_i - p). \quad (15.17)$$

As expected, supply (production) increases with the relative price of the good.

Next, we need to think about demand for every good i , and about aggregate demand. The former takes a very simple form. It can be derived from basic utility but no need to do so here as the form is very intuitive. Demand depends on income, relative prices, and a good-specific taste shock – in log format it can be written as

$$q_i = y + z_i - \eta (p_i - p) \quad \eta > 0, \quad (15.18)$$

where y is average income and p is the average price level. The taste shock z_i is assumed to affect relative tastes, hence it averages to zero across all goods. It is also assumed to be normally distributed, for reasons that will soon be clear, with variance v_z .

How about aggregate demand? We will assume that there is an aggregate demand shifter, a policy variable we can control, which in this case will be m . It can be anything that shifts the AD curve within the AS-AD framework developed above, but to fix ideas we can think about monetary policy. To introduce, it consider a money demand function in log form:

$$y = m - p. \quad (15.19)$$

We assume that m is also normally distributed, with mean $E(m)$ and variance v_m .

Equilibrium

To find the equilibrium we make demand equal to supply for each good. This is a model with market clearing and where all variables, particularly p , are known.

$$\left(\frac{1}{\gamma-1} \right) (p_i - p) = y + z_i - \eta (p_i - p), \quad (15.20)$$

from which we obtain the individual price

$$p_i = \frac{(\gamma-1)}{1 + \eta\gamma - \eta} (y + z_i) + p, \quad (15.21)$$

and from which we can obtain the average price. Averaging (15.21) we get

$$p = \frac{(\gamma-1)}{(1 + \eta\gamma - \eta)} y + p \quad (15.22)$$

which implies

$$y = 0.$$

You may find it strange, but it is not: remember that output is defined in logs. Replacing the solution for output in (15.19) we get that

$$p = m, \quad (15.23)$$

i.e. that prices respond fully to monetary shocks. In other words, the world with perfect information is a typical classical version with flexible prices and where aggregate demand management has no effect on real variables and full impact on nominal variables.

15.2.3 | Lucas' supply curve

When there is imperfect information, each producer observes the price of her own good, p_i , but cannot observe perfectly what happens to other prices. She will have to make her best guess as to whether a change in her price represents an increase in relative prices, or just a general increase in the price level. In other words, labour supply will have to be determined on the basis of expectations. Because we assume rational expectations, these will be determined by the mathematical expectation that is consistent with the model – in other words, individuals know the model and form their expectations rationally based on this knowledge.

Denote relative prices as $r_i = (p_i - p)$, then the analog to (15.17) is now⁴

$$l_i = \left(\frac{1}{\gamma - 1} \right) E(r_i | p_i). \quad (15.24)$$

It so happens that if the distribution of the shocks z_i and m is jointly normal, then so will be r_i , p_i , and p . Since r_i and p_i are jointly normally distributed, a result from statistics tells us that the conditional expectation is a linear function

$$E(r_i | p_i) = \alpha + \beta p_i. \quad (15.25)$$

More specifically, in this case, we have what is called a signal extraction problem, in which one variable of interest (r_i) is observed with noise. What you observe (p_i) is the sum of the signal you're interested in (r_i), plus noise you don't really care about (p). It turns out that, with the assumption of normality, the solution to this problem is

$$E(r_i | p_i) = \frac{v_r}{v_r + v_p} (p_i - E(p)), \quad (15.26)$$

where v_r and v_p are the variances of relative price and general price level, respectively. (They are a complicated function of v_z and v_m .) This expression is very intuitive; if most of the variance comes from the signal, your best guess is that a change in p_i indicates a change in relative prices. Substituting in (15.24) yields

$$l_i = \left(\frac{1}{\gamma - 1} \right) \frac{v_r}{v_r + v_p} (p_i - E(p)). \quad (15.27)$$

Aggregating over all the individual supply curves, and defining

$$b = \frac{1}{\gamma - 1} \frac{v_r}{v_r + v_p} \quad (15.28)$$

we have that

$$y = b (p - E(p)), \quad (15.29)$$

which is actually a Phillips curve, as you know from basic macro courses.

This became known as the *Lucas supply curve*. Note that this is a positively-sloped supply curve, in which output increases when the price increases *in excess of its expected level*. Why is it so? Because when facing such an increase, imperfectly informed producers rationally attribute some of that to an increase in relative prices. It also says that labour and output respond more to price changes if the relative relevance of nominal shocks is smaller. Why is this? Because the smaller the incidence of nominal shocks, the more certain is the producer that any price shock she faces is a change in real demand.

Solving the model

We know from the AS-AD framework that, with a positively-sloped supply curve, aggregate demand shocks affect equilibrium output. How do we see that in the context of this model? Plugging (15.29) into the aggregate demand equation (15.19) yields

$$y = b(p - E(p)) = m - p, \quad (15.30)$$

that can be used to solve for the aggregate price level and income:

$$p = \frac{m}{1+b} + \frac{b}{1+b}E(p), \quad (15.31)$$

$$y = \frac{bm}{1+b} - \frac{b}{1+b}E(p). \quad (15.32)$$

Now, rational expectations means that individuals will figure this out in setting their own expectations. In other words, we can take the expectations of (15.31) to obtain:⁵

$$E(p) = \frac{1}{1+b}E(m) + \frac{b}{1+b}E(p), \quad (15.33)$$

which implies, in turn, that

$$E(p) = E(m). \quad (15.34)$$

Using this and the fact that $m = E(m) + m - E(m)$ we have that

$$p = E(m) + \frac{1}{1+b}(m - E(m)), \quad (15.35)$$

$$y = \frac{b}{1+b}(m - E(m)). \quad (15.36)$$

In short, the model predicts that changes in aggregate demand (e.g. monetary policy) will have an effect on output, but only to the extent that they are unexpected. This is a very powerful conclusion in the sense that systematic policy will eventually lose its effects; people will figure it out, and come to expect it. When they do, they'll change their behaviour accordingly, and you won't be able to exploit it. This is at the heart of the famous *Lucas critique*: as the policy maker acts, the aggregate supply curve will change as a result of that, and you can't think of them as stable relationships independent of policy.

As we can see, the imperfect information approach highlights the role of expectations in determining the effectiveness of macro policy. This insight is very general, and lies behind a lot of modern policy making: inflation targeting as a way of coordinating expectations, the problem of time inconsistency, etc. In fact, we will soon see that this insight is very much underscored by the modern New Keynesian approach.

15.3 | Imperfect competition and nominal and real rigidities

We show that, with imperfect competition and nominal rigidities, there is a role for aggregate demand policy. Imperfect competition means that firms can set prices, and that output can deviate from the social optimum. Nominal rigidities mean that prices fail to adjust automatically. The two combined mean that output can be increased (in the short run), and that doing so can be desirable. We discuss how real rigidities amplify the impact of nominal rigidities.

The Lucas model was seen at the time as a major strike against Keynesian policy thinking. After all, while it illustrates how we can obtain a positively-sloped AS curve from a fully microfounded approach with rational agents, it also fails to provide a justification for systematic macro policy. The New Keynesian tradition emerged essentially as an attempt to reconcile rational expectations and the possibility and desirability of systematic policy.

Consider the desirability: in the Lucas model (or the RBC approach that essentially came out of that tradition), business cycles are the result of optimal responses by individuals to disturbances that occur in this economy, and aggregate demand policy can only introduce noise. If the market equilibrium is socially optimal, then any fluctuation due to aggregate demand shocks is a departure from the optimum, and thus undesirable. The New Keynesian view departs from that by casting *imperfect competition* in a central role. The key to justifying policy intervention is to consider the possibility that the market-determined level of output is suboptimal, and imperfect competition yields exactly that. In addition, this is consistent with the general impression that recessions are bad and booms are good.

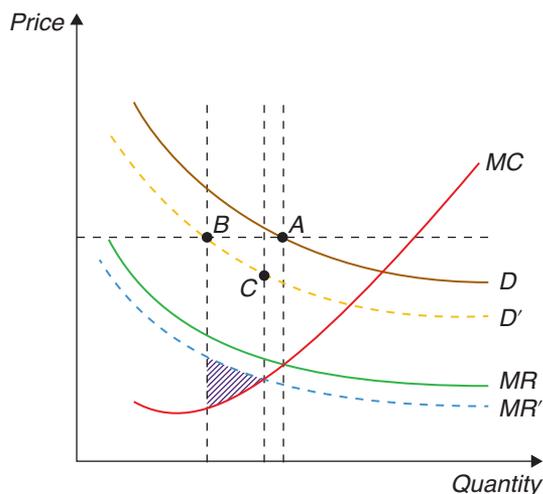
Besides the issue of desirability, we have argued that the Lucas model also implies that systematic policy is powerless; rational agents with rational expectations figure it out, and start adjusting prices accordingly. The second essential foundation of New Keynesian thinking is thus the existence and importance of barriers to price adjustment. Note that this is also related to imperfect competition since price adjustment can only matter if firms are price-setters, which requires some monopoly power. It is not enough to have imperfect competition to have these rigidities, however, as monopolists will also want to adjust prices rather than output in response to nominal shocks.

We thus have to understand how barriers, that are most likely rather small at the micro level, and which have become known in the literature by the catch-all term *menu costs*, can still have large macroeconomic effects. Do we really think that in the real world the costs of adjusting prices are large enough to lead to sizeable consequences in output?

It turns out that the key lies once again with imperfect competition. Consider the effects of a decrease in aggregate demand on the behaviour of monopolist firms, illustrated in Figure 15.5. Taking the behaviour of all other firms as given, this will make any given firm want to set a lower price. If there were no costs of adjustment, the firm would go from point A in Figure 15.5 to point C. If the firm doesn't adjust at all, it would go to point B. It follows that its gain from adjusting would be the shaded triangle. If the menu cost is greater than that, the firm would choose not to adjust.

But what is the social cost of not adjusting? It is the difference in consumer surplus corresponding to a decrease in quantity from point C to point B. This is given by the area between the demand curve D' and the marginal cost curve, between B and C. This is *much* bigger than the shaded triangle! In other words, the social loss is much bigger than the firm's loss from not adjusting, and it follows that small menu costs can have large social effects.⁶

Another type of rigidity emphasised by New Keynesians are real rigidities (as distinct from the nominal kind). These correspond to a low sensitivity of the desired price to aggregate output. If the

Figure 15.5 Welfare effects of imperfect competition

desired price doesn't change much with a change in output, the incentive to adjust prices will be lower. (Think about the slope of the marginal cost curve in Figure 15.5). If there are no costs of adjustment (i.e. nominal rigidities), that doesn't matter, of course; but the real rigidities amplify the effect of the nominal ones. These real rigidities could come from many sources, such as the labour market. If labour supply is relatively inelastic (think about low levels of labour mobility, for instance), we would have greater real rigidities. (This actually sets the stage for us to consider our next topic in the study of cyclical fluctuations: labour markets and unemployment.)

In sum, a combination of imperfect competition, nominal rigidities (menu costs), and real rigidities implies that aggregate demand policy is both desirable and feasible. We will now turn to a very brief discussion of how this view of the world has been embedded into full-fledged dynamic stochastic general equilibrium (DSGE) models such as those introduced by the RBC tradition to give birth to the modern New Keynesian view of fluctuations.⁷

15.4 | New Keynesian DSGE models

We express the modern New Keynesian DSGE (NK DSGE) model in its canonical (microfounded) version, combining the New Keynesian IS curve, the New Keynesian Phillips curve, and a policy rule. We show the continuous-time and discrete-time versions of the model.

New Keynesian DSGE models embody the methodological consensus underpinning modern macroeconomics. It has become impossible to work in any self-respecting Central Bank, for instance, without coming across a New Keynesian DSGE model. But modern, state-of-the-art DSGE models are very complicated. If you thought that RBC models were already intricate, consider the celebrated NK DSGE model by Smets and Wouters (2003), originally developed as an empirical model of the Euro area. It contains, in addition to productivity shocks, shocks to adjustment costs, the equity premium,

wage markup, goods markup, labor supply, preferences, and the list goes on and on. Another difficulty is that there is little consensus as to which specific model is best to fit real-world fluctuations. So what we do here is consider a few of the key ingredients in NK DSGE models, and explain how they combine into what is often called the canonical New Keynesian model.

15.4.1 | The canonical New Keynesian model

We first develop the model in continuous time, which is simpler and allows for the use of phase diagrams, so that we can readily put the model to work and develop some intuition about its operation and dynamics. Later, we turn to discrete time, and write down the version of the model that is most commonly used in practical and policy applications.

The demand side of the canonical New Keynesian model is very simple. We start from our model of consumer optimisation, which by now we have seen many times. You will recall the Euler equation of the representative consumer.

$$\dot{C}_t = \sigma (r_t - \rho) C_t, \quad (15.37)$$

where C_t is consumption, $\sigma > 0$ is the elasticity of intertemporal substitution in consumption, and ρ is the rate of time discounting. In a closed economy with no investment, all output Y_t is consumed. Therefore,

$$C_t = Y_t, \quad (15.38)$$

and

$$\dot{Y}_t = \sigma (i_t - \pi_t - \rho) Y_t, \quad (15.39)$$

where we have used the definition $r_t \equiv i_t - \pi_t$, and i_t is the nominal interest rate, taken to be exogenous and constant for the time being. If we define the output gap as,

$$X_t \equiv \frac{Y_t}{\bar{Y}_t}, \quad (15.40)$$

where \bar{Y}_t is the natural or long run level of output, then the output gap evolves according to

$$\frac{\dot{X}_t}{X_t} = \frac{\dot{Y}_t}{Y_t} - g, \quad (15.41)$$

where g is the percentage growth rate of the natural level of output, assumed constant for now. Finally, letting small-case letters denote logarithms, using the Euler equation (15.39), we have

$$\dot{x}_t = \sigma (i_t - \pi_t - r^n), \quad (15.42)$$

where $r^n \equiv \rho + \sigma^{-1}g$ is the natural or Wicksellian interest rate, which depends on both preferences and productivity growth. It is the interest rate that would prevail in the absence of distortions, and corresponds to a situation in which output is equal to potential.

This last equation, which we can think of as a dynamic New Keynesian IS equation (or NKIS) summarises the demand side of the model. The NKIS equation says that output is rising when the real interest rate is *above* its long-run (or natural) level. Contrast this with the conventional IS equation, which says that the *level* of output (as opposed to the *rate of change* of output in the equation above) is above its long-run level when the real interest is *below* its long-run (or natural) level.

The NKIS differs from traditional IS in other important ways. First, it is derived from micro-founded, optimising household behaviour. Second, the relationship between interest rates and output emerges from the behaviour of consumption, rather than investment, as was the case in the old IS. Intuitively, high interest rates are linked to low output now because people decide that it is better to postpone consumption, thereby reducing aggregate demand.

Turn now to the supply side of the model. We need a description of how prices are set in order to capture the presence of nominal rigidities. There are many different models for that, which are typically classified as time-dependent or state-dependent. State-dependent models are those in which adjustment is triggered by the state of the economy. Typically, firms decide to adjust (and pay the menu cost) if their current prices are too far from their optimal desired level. Time-dependent models, in contrast, are such that firms get to adjust prices with the passage of time, say, because there are long-term contracts. This seems slightly less compelling as a way of understanding the underpinnings of price adjustment, but it has the major advantage of being easier to handle. We will thus focus on time-dependent models, which are more widely used.

There are several time-dependent models, but the most popular is the so-called Calvo model. Calvo (1983) assumes that the economy is populated by a continuum of monopolistically-competitive firms. Each of them is a point in the $[0, 1]$ interval, thus making their ‘total’ equal to one. The key innovation comes from the price-setting technology: each firm sets its output price in terms of domestic currency and can change it only when it receives a price-change signal. The probability of receiving such a signal s periods from now is assumed to be independent of the last time the firm got the signal, and given by

$$\alpha e^{-\alpha s}, \quad \alpha > 0. \quad (15.43)$$

If the price-change signal is stochastically independent across firms, we can appeal to the ‘law of large numbers’ to conclude that a share α of firms will receive the price-change signal per unit of time. By the same principle, of the total number of firms that set their price at time $s < t$, a share

$$e^{-\alpha(t-s)} \quad (15.44)$$

will not have received the signal at time t . Therefore,

$$\alpha e^{-\alpha(t-s)} \quad (15.45)$$

is the share of firms that set their prices at time s and have not yet received a price-change signal at time $t > s$.

Next, let v_t be the (log of the) price set by an individual firm (when it gets the signal), and define the (log of the) price level p_t as the arithmetic average of all the prices v_t still outstanding at time t , weighted by the share of firms with the same v_t :

$$p_t = \alpha \int_{-\infty}^t v_s e^{-\alpha(t-s)} ds. \quad (15.46)$$

It follows that the price level is sticky, because it is a combination of pre-existing prices (which, because they are pre-existing, cannot jump suddenly).

How is v_t set? Yun (1996) was the first to solve the full problem of monopolistically-competitive firms that must set prices optimally, understanding that it and all competitors will face stochastic price-setting signals. Getting to that solution is involved, and requires quite a bit of algebra.⁸

Here we just provide a reduced form, and postulate that the optimal price v_t set by an individual firm depends on the contemporaneous price level p_t , the expected future paths of the (log of) expected relative prices, and of the (log of) the output gap:

$$v_t = p_t + \alpha \int_t^{\infty} [(v_s - p_s) + \eta x_s] e^{-(\alpha+\rho)(s-t)} ds, \quad (15.47)$$

where, recall, ρ is the consumer's discount rate and $\eta > 0$ is a sensitivity parameter.⁹ So the relative price the firm chooses today depends on a discounted, probability-weighted average of all future relative prices $(v_s - p_s)$ and all output gaps x_s . This is intuitive. For instance, if the output gap is expected to be positive in the future, then it makes sense for the firm to set a higher (relative) price for its good to take advantage of buoyant demand.

Note from this expression that along any path in which the future x_s and v_s are continuous functions of time (which we now assume), v_t is also, and necessarily, a continuous function of time. We can therefore use Leibniz's rule to differentiate the expressions for p_t and v_t with respect to time, obtaining¹⁰

$$\dot{p}_t = \pi_t = \alpha (v_t - p_t), \quad (15.48)$$

and

$$\dot{v}_t - \dot{p}_t = -\alpha \eta x_t + \rho (v_t - p_t). \quad (15.49)$$

Combining the two we have

$$\dot{v}_t - \dot{p}_t = -\alpha \eta x_t + \frac{\rho}{\alpha} \pi_t. \quad (15.50)$$

Differentiating the expression for the inflation rate π_t , again with respect to time, yields

$$\dot{\pi}_t = \alpha (\dot{v}_t - \dot{p}_t). \quad (15.51)$$

Finally, combining the last two expressions we arrive at

$$\dot{\pi}_t = \rho \pi_t - \kappa x_t, \quad (15.52)$$

where $\kappa \equiv \alpha^2 \eta > 0$. This is the canonical New Keynesian Phillips curve. In the traditional Phillips curve, the rate of inflation was an increasing function of the output gap. By contrast, in the Calvo-Yun NKPC the *change* in the rate of inflation is a decreasing function of the output gap! Notice, also that while p_t is a sticky variable, its rate of change π_t is not; it is intuitive that π_t should be able to jump in response to expected changes in relevant variables.

Solving this equation forward we obtain

$$\pi_t = \int_t^{\infty} \kappa x_s e^{-\rho(s-t)} ds. \quad (15.53)$$

So the inflation rate today is the present discounted value of all the future expected output gaps. The more "overheated" the economy is expected to be in the future, the higher inflation is today.

To complete the supply side of the model we need to specify why the output gap should be anything other than zero – that is, why firms can and are willing to supply more output than their long-term profit maximizing level. The standard story, adopted, for instance, by Yun (1996), has two components. Output is produced using labour and firms can hire more (elastically supplied) labour in the short-run to enlarge production when desirable. When demand rises (recall the previous section of this

chapter), monopolistically-competitive firms facing fixed prices will find it advantageous to supply more output, up to a point.

This NKPC curve and the dynamic NKIS curve, taken together, fully describe this model economy. They are a pair of linear differential equations in two variables, π_t and x_t , with i_t as an exogenous policy variable. In this model there is no conflict between keeping inflation low and stabilising output. If $i = r^n$, then $\pi_t = x_t = 0$ is an equilibrium. Blanchard and Galí (2007) term this the divine coincidence.

The steady state is

$$\bar{\pi} = i - r^n \text{ (from } \dot{x}_t = 0 \text{)} \quad (15.54)$$

$$\rho \bar{\pi} = \kappa \bar{x} \text{ (from } \dot{\pi}_t = 0 \text{)} \quad (15.55)$$

where overbars denote the steady state. If, in addition, we assume $i = r^n$, then $\bar{\pi} = \bar{x} = 0$. In matrix form, the dynamic system is

$$\begin{bmatrix} \dot{\pi}_t \\ \dot{x}_t \end{bmatrix} = \Omega \begin{bmatrix} \pi_t \\ x_t \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma (i - r^n) \end{bmatrix} \quad (15.56)$$

where

$$\Omega = \begin{bmatrix} \rho & -\kappa \\ -\sigma & 0 \end{bmatrix}. \quad (15.57)$$

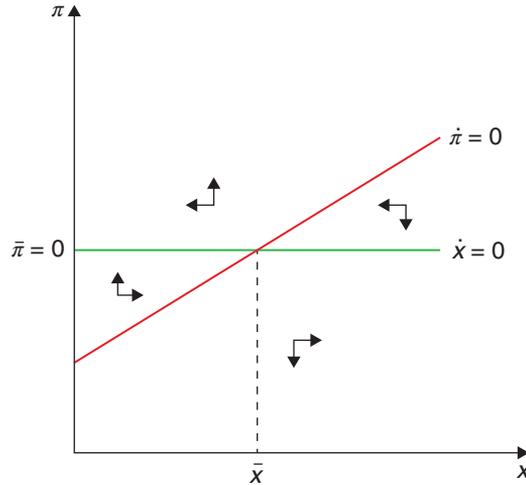
It is straightforward to see that $\text{Det}(\Omega) = -\sigma\kappa < 0$, and $\text{Tr}(\Omega) = \rho > 0$. It follows that one of the eigenvalues of Ω is positive (or has positive real parts) and the other is negative. This means the system exhibits saddle path stability, in other words that for each π_t there is a value of x_t from which the system will converge asymptotically to the steady state. But remember that here both x and π are jump variables! This means that we have a continuum of perfect-foresight convergent equilibria, because we can initially choose both π_t and x_t .

The graphical representation of this result is as follows. When drawn in $[\pi_t, x_t]$ space, the Phillips curve is positively-sloped, while the IS schedule is horizontal, as you can see in the phase diagram in Figure 15.6. If $x_0 > \bar{x}$, there exists a $\pi_0 > \bar{\pi}$ such that both variables converge to the steady state in a south-westerly trajectory. The converse happens if $x_0 < \bar{x}$. Along a converging path, inflation and output do move together, as in the standard Phillips curve analysis. To see that, focus for instance on the south-west quadrant of the diagram. There, both output and inflation are below their long run levels, so that a depressed economy produces unusually low inflation. As output rises toward its long-run resting point, so does inflation.

But the important point is that there exists an infinity of such converging paths, one for each (arbitrary) initial condition! An exogenous path for the nominal interest, whichever path that may be, is not enough to pin down the rate of inflation (and the output gap) uniquely. What is the intuition for this indeterminacy or nonuniqueness? To see why self-fulfilling recessions may occur, suppose agents believe that output that is low today will gradually rise towards steady state. According to the NKPC, New Keynesian Phillips curve, a path of low output implies a path of low inflation. But with the nominal interest rate exogenously fixed, low expected inflation increases the real rate of interest and lowers consumption and output. The initial belief is thus self-fulfilling.

15.4.2 | A Taylor rule in the canonical New Keynesian model

In Chapter 19 we further discuss interest rate policy and interest rate rules. Here we simply introduce the best-known and most-widely used rule: the Taylor rule, named after Stanford economist John

Figure 15.6 Indeterminacy in the NK model

Taylor, who first proposed it as a description of the behaviour of monetary policy in the U.S. In Taylor (1993), the rule takes the form

$$i_t = r_t^n + \phi_\pi \pi_t + \phi_x x_t, \quad (15.58)$$

where ϕ_π and ϕ_x are two coefficients chosen by the monetary authority. The choice of r_t^n requires it be equal to the normal or natural real rate of interest in the steady state. In what follows we will often assume $\phi_\pi > 1$, so that when π_t rises above the (implicit) target of 0, the nominal interest rises more than proportionately, and the real interest goes up in an effort to reduce inflation. Similarly, $\phi_x > 0$, so that when the output gap is positive, i_t rises from its normal level. Using the Taylor rule in the NKIS equation (15.42) yields

$$\dot{x}_t = \sigma \left[(r_t^n - r^n) + (\phi_\pi - 1) \pi_t + \phi_x x_t \right], \quad (15.59)$$

so that the rate of increase of the output gap is increasing in its own level and also increasing in inflation (because $\phi_\pi - 1 > 0$). The resulting dynamic system can be written as

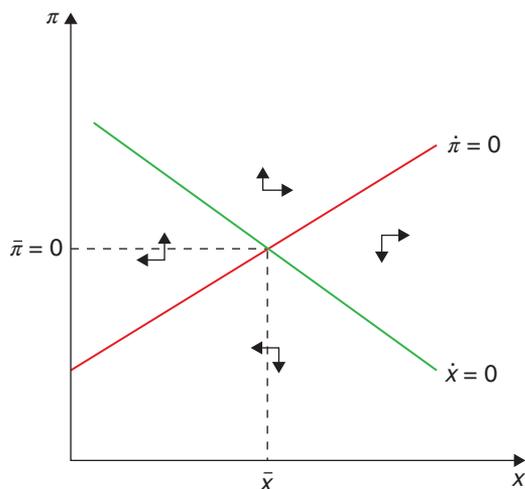
$$\begin{bmatrix} \dot{\pi}_t \\ \dot{x}_t \end{bmatrix} = \Omega \begin{bmatrix} \pi_t \\ x_t \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma (r_t^n - r^n) \end{bmatrix} \quad (15.60)$$

where

$$\Omega = \begin{bmatrix} \rho & -\kappa \\ \sigma (\phi_\pi - 1) & \sigma \phi_x \end{bmatrix}. \quad (15.61)$$

Now $\text{Det}(\Omega) = \rho \sigma \phi_x + \sigma (\phi_\pi - 1) \kappa > 0$, and $\text{Tr}(\Omega) = \rho + \sigma \phi_x > 0$. It follows that $\phi_\pi > 1$ is sufficient to ensure that both eigenvalues of Ω are positive (or have positive real parts). Because both π_t and x_t are jump variables, the steady state is now unique. After any permanent unanticipated shock, the system just jumps to the steady state and remains there!

As you can see in the phase diagram in Figure 15.7, the $\dot{x}_t = 0$ schedule (the NKIS) now slopes down. All four sets of arrows point away from the steady state point – which is exactly what you need to guarantee uniqueness of equilibrium in the case of a system of two jump variables!

Figure 15.7 Active interest rule in the NK model

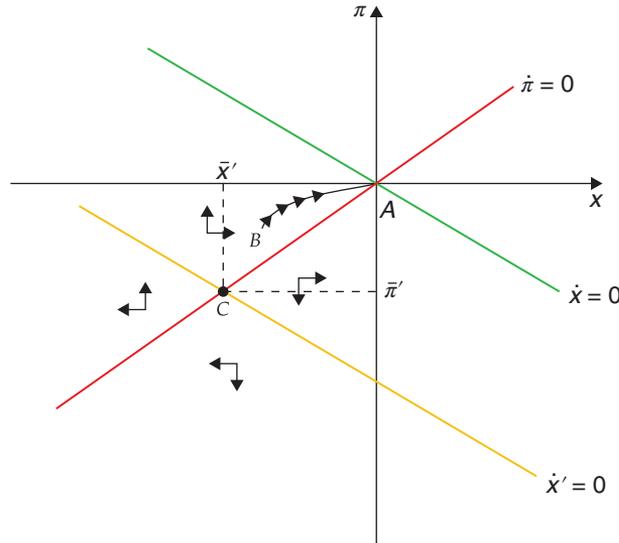
Go back to the expression $\text{Det}(\Omega) = \rho\phi_x + \sigma(\phi_\pi - 1)\kappa$, which reveals that if $\phi_\pi < 1$ and ϕ_x is not too large, then $\text{Det}(\Omega) < 0$. Since, in addition, $\text{Tr}(\Omega) > 0$, we would have a case of one positive and one negative eigenvalue, so that, again, multiplicity of equilibria (in fact, infinity of equilibria) would occur.

So there is an important policy lesson in all of this. In the canonical New Keynesian model, interest rate policy has to be sufficiently activist (aggressively anti-inflation, one might say), in order to guarantee uniqueness of equilibrium – in particular, to ensure that the rate of inflation and the output gap are pinned down. In the literature, policy rules where $\phi_\pi > 1$ are usually called active policy rules, and those where $\phi_\pi < 1$ are referred to as passive policy rules. Active policy rules are said to satisfy the ‘Taylor principle’: sufficiently reactive interest rate policy pins down the equilibrium.

In this very simple model, which boils down to a system of linear differential equations, the uniqueness result is simple to derive and easy to understand. In more complex versions of the Keynesian model, –for instance, in non-linear models which need to be linearised around the steady state – or in circumstances in which the zero lower bound on the nominal interest rate binds, dynamics can be considerably more complicated, and the condition $\phi_\pi > 1$ in the Taylor rule need not be sufficient to guarantee uniqueness of equilibrium. For a more detailed treatment of these issues, see Benhabib et al. (2001a), Benhabib et al. (2001b), Benhabib et al. (2002), Woodford (2011), and Galí (2015).

Before ending this section, let us put this model to work by analysing a shock. Let’s imagine a monetary tightening implemented through a transitory exogenous increase in the interest rate (think of the interest moving to $r_t^n + z$, with z the policy shifter), or, alternatively, imagine that at time 0, the natural rate of interest suddenly goes down from r^n to \underline{r}^n , where $0 < \underline{r}^n < r^n$, because the trend rate of growth of output, g , has temporarily dropped. After $T > 0$, either z goes back to zero, or the natural rate of interest goes back to r^n and remains there forever. How will inflation and output behave on impact, and in the time interval between times 0 and T ? The phase diagram in Figure 15.8 below shows the answer to these questions in a simple and intuitive manner.

Notice that either of these changes imply a leftward shift of the \dot{x} equation. So, during the transition the dynamics are driven by the original $\dot{\pi}$ and new \dot{x} equations which intersect at point C.

Figure 15.8 A reduction in the natural rate

The system must return to A exactly at time T . On impact, inflation and the output jump to a point, such as B, which is pinned down by the requirement that between 0 and T dynamics be those of the system with steady state at point C. That is, during the temporary shock the economy must travel to the north-east, with inflation rising and the output gap narrowing, in anticipation of the positive reversion of the shock at T . Between 0 and T the negative shock, intuitively enough, causes output and inflation to be below their initial (and final, after T) steady-state levels. If initially $i^n = r^n$, so $\bar{\pi} = 0$ and $\bar{x} = 0$, as drawn below, then during the duration of the shock the economy experiences deflation and a recession (a negative output gap).¹¹

What happens to the nominal interest rate? Between 0 and T , both inflation and the output gap are below their target levels of zero. So, the monetary authority relaxes the policy stance in response to both the lower inflation and the negative output gap. But that relaxation is not enough to keep the economy from going into recession and deflation. We return to this issue in Chapter 22.

15.4.3 | Back to discrete time

The canonical New Keynesian model has a natural counterpart in discrete time, which is more broadly used for practical applications. In discrete time the Phillips curve becomes (see Galí (2015) for the detailed derivation)

$$\pi_t = \beta E_t \pi_{t+1} + \kappa x_t, \quad (15.62)$$

where $0 < \beta = \frac{1}{1+\rho} < 1$ is the discount factor, E_t is the expectations operator (with expectations computed as of time t), and the output gap is again in logs. To derive the IS curve, again start from the Euler equation, which in logs can be written as

$$y_t = E_t y_{t+1} - \sigma \log(1 + r_t) + \sigma \log(1 + \rho), \quad (15.63)$$

where we have already used $c_t = y_t$. If we recall the fact that for small r_t and ρ , $\log(1 + r_t) \approx r_t$, and $\log(1 + \rho) \approx \rho$, equation (15.63) becomes

$$y_t = E_t y_{t+1} - \sigma (r_t - \rho). \quad (15.64)$$

Finally, subtracting \bar{y} from both sides yields

$$x_t = E_t x_{t+1} - \sigma (i_t - E_t \pi_{t+1} - \rho), \quad (15.65)$$

where we have used the definition $r_t = i_t - E_t \pi_{t+1}$ and the fact that $x_t = y_t - \bar{y}$. If the natural rate of output is not constant, so that

$$\bar{y}_{t+1} = \bar{y}_t + \Delta, \quad (15.66)$$

(15.65) becomes

$$x_t = E_t x_{t+1} + \Delta - \sigma (i_t - E_t \pi_{t+1} - \rho) \quad (15.67)$$

or

$$x_t = E_t x_{t+1} - \sigma (i_t - E_t \pi_{t+1} - r^n), \quad (15.68)$$

where

$$r^n = \rho + \frac{\Delta}{\sigma}, \quad (15.69)$$

and again the variable r^n is the natural, or Wicksellian, interest rate, which can move around as a result of preference shocks (changes in ρ) or productivity growth (Δ). To close the model, we can again appeal to an interest rule of the form

$$i_t = i^n + \phi_\pi E_t \pi_{t+1} + \phi_x x_t. \quad (15.70)$$

As before, policy makers in charge of interest rate setting respond to deviations in expected inflation from the target (here equal to zero), and to deviations of output from the full employment or natural rate of output. Taylor argued that this rule (specifically, with $\phi_\pi = 1.5$, $\phi_x = 0.5$), and an inflation target of 2% is a good description of how monetary policy actually works in many countries – and, in particular, of how the Fed has behaved in modern times (since the mid-1980s).

There is an active research program in trying to compute optimal Taylor rules and also to estimate them from real-life data. In practice, no central bank has formally committed exactly to such a rule, but the analysis of monetary policy has converged onto some variant of a Taylor rule – and on interest rate rules more broadly – as the best way to describe how central banks operate.

Substituting the interest rate rule into the NKIS equation (15.70) (in the simple case of constant \bar{y}) yields

$$x_t = E_t x_{t+1} - \sigma [(\phi_\pi - 1) E_t \pi_{t+1} + \phi_x x_t + (i^n - r^n)]. \quad (15.71)$$

This equation plus the NKPC constitute a system of two difference equations in two unknowns. As in the case of continuous time, it can be shown that an interest rule that keeps i_t constant does not guarantee uniqueness of equilibrium. But, it again turns out that if $\phi_\pi > 1$, a Taylor-type rule does ensure that both eigenvalues of the characteristic matrix of the 2×2 system are larger than one. Since both π_t and x_t are jumpy variables, that guarantees a unique outcome; the system simply jumps to the steady state and stays there.

To analyse formally the dynamic properties of this system, rewrite the NKPC (15.62) as

$$E_t \pi_{t+1} = \beta^{-1} \pi_t - \beta^{-1} \kappa x_t. \quad (15.72)$$

Next, use this (15.71) to yield

$$E_t x_{t+1} = \sigma (\phi_\pi - 1) \beta^{-1} \pi_t + x_t + \sigma [- (\phi_\pi - 1) \beta^{-1} \kappa + \phi_x] x_t + \sigma (i^n - r^n). \quad (15.73)$$

(15.72) and (15.73) together constitute the canonical New Keynesian model in discrete time. In matrix form, the dynamic system is

$$\begin{bmatrix} E_t \pi_{t+1} \\ E_t x_{t+1} \end{bmatrix} = \Omega \begin{bmatrix} \pi_t \\ x_t \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma (i^n - r^n) \end{bmatrix} \quad (15.74)$$

where

$$\Omega = \begin{bmatrix} \beta^{-1} & -\beta^{-1} \kappa \\ \sigma \beta^{-1} (\phi_\pi - 1) & 1 + \sigma [- (\phi_\pi - 1) \beta^{-1} \kappa + \phi_x] \end{bmatrix}. \quad (15.75)$$

Now

$$\text{Det}(\Omega) = \beta^{-1} (1 + \sigma \phi_x) = \lambda_1 \lambda_2 > 1 \quad (15.76)$$

and

$$\text{Tr}(\Omega) = \beta^{-1} + 1 + \sigma [\phi_x - (\phi_\pi - 1) \beta^{-1} \kappa] = \lambda_1 + \lambda_2, \quad (15.77)$$

where λ_1 and λ_2 are the eigenvalues of Ω . For both λ_1 and λ_2 to be larger than one, a necessary and sufficient condition is that

$$\text{Det}(\Omega) + 1 > \text{Tr}(\Omega) \quad (15.78)$$

which, using the expressions for the determinant and the trace, is equivalent to

$$(\phi_\pi - 1) x + (1 - \beta) \phi_x > 0. \quad (15.79)$$

This condition clearly obtains if $\phi_\pi > 1$. So, the policy implication is the same in both the continuous time and the discrete time of the model: an activist policy rule is required, in which the interest rate over-reacts to changes in the (expected) rate of inflation, in order to ensure uniqueness of equilibrium.

For a classic application of this model to monetary policy issues, see Clarida et al. (1999). In later chapters of this book we use the model to study a number of issues, some of which require the S to the DSGE acronym: shocks! As we saw in our discussion of RBC models, the business cycle properties we obtain from the model will depend on the properties we assume for those shocks.

The kind of DSGE model that is used in practice will add many bells and whistles to the canonical version, in the description of the behaviour of firms, households, and policy-makers. In doing so, it will open the door to a lot of different shocks as well. It will then try to either calibrate the model parameters, or estimate them using Bayesian techniques, and use the model to evaluate policy interventions. You will find in Appendix C a basic illustration of this kind of model, so you can do it for yourself!

These models will nevertheless keep the essential features of a firm grounding on household and firm optimisation, which is a way to address the Lucas Critique, and also of the consequent importance of expectations. We discuss the issues in greater detail in Chapter 17 and thereafter.

15.5 | What have we learned?

We have gone over the basics of the Keynesian view of the business cycle, from its old IS-LM version to the modern canonical New Keynesian DSGE model. We saw the key role of imperfect price adjustment, leading to an upward-sloping aggregate supply curve, under which aggregate demand shocks have real consequences. We showed how imperfect competition and nominal (and real) rigidities are crucial for that. We saw how the Euler equation of consumption gives rise to the modern New Keynesian IS curve, while the Calvo model of price setting gives rise to the New Keynesian Phillips curve. Finally, we saw how we need to specify a policy rule (such as the Taylor rule) to close the model.

There is no consensus among macroeconomists as to whether the Keynesian or classical (RBC) view is correct. This is not surprising since they essentially involve very different world views in terms of the functioning of markets. Are market failures (at least relatively) pervasive, or can we safely leave them aside in our analysis? This is hardly the type of question that can be easily settled by the type of evidence we deal with in the social sciences.

Having said that, it's important to stress the methodological convergence that has been achieved in macroeconomics, and that has hopefully been conveyed by our discussion in the last two chapters. Nowadays, essentially all of macro deals with microfounded models with rational agents, the difference being in the assumptions about the shocks and rigidities that are present (or absent) and driving the fluctuations. By providing a unified framework that allows policy makers to cater the model to what they believe are the constraints they face, means that the controversy about the fundamental discrepancies can be dealt, in a more flexible way within a unified framework. Imagine the issue of price rigidity, which is summarised by Calvo's α coefficient of price adjustment. If you believe in no price rigidities, α has a specific value, if you think there are rigidities you just change the value. And nobody is going to fight for the value of α , are they? Worst case scenario, you just run it with both parameters and look at the output. No wonder then that the DSGE models have become a workhorse, for example, in Central Banking.

15.6 | What next?

Any number of macro textbooks cover the basics of the Keynesian model, in its IS-LM version. The textbook by Romer (2018) covers the topics at the graduate level, and is a great introduction to the fundamentals behind the New Keynesian view. For the canonical, modern New Keynesian approach, the book by Galí (2015) is the key reference.

Notes

¹ This is what's behind Keynes oft-quoted (and misquoted) statement (from Ch. 16 of the *General Theory*) that “‘To dig holes in the ground,’ paid for out of savings, will increase, not only employment, but the real national dividend of useful goods and services.” (Note, however that he immediately goes on to say that ‘it is not reasonable, however, that a sensible community should be content to remain dependent on such fortuitous and often wasteful mitigation when once we understand the influences upon which effective demand depends.’) Similarly, in Ch. 10 he states: “If the Treasury were to fill old bottles with banknotes, bury them at suitable depths in disused coalmines which are then filled up to the surface with town rubbish, and leave it to private enterprise on well-tried

principles of laissez-faire to dig the notes up again (the right to do so being obtained, of course, by tendering for leases of the note-bearing territory), there need be no more unemployment and, with the help of the repercussions, the real income of the community, and its capital wealth also, would probably become a good deal greater than it actually is. It would, indeed, be more sensible to build houses and the like; but if there are political and practical difficulties in the way of this, the above would be better than nothing.”

² As we will see, this is not exactly a Keynesian model; it was actually the opening shot in the rational expectations revolution. The New Keynesian approach, however, is the direct descendant of that revolution, by incorporating the rational expectations assumption and championing the role of aggregate demand policy under those conditions.

³ See Hicks (1937).

⁴ This is a simplifying assumption of certainty equivalence behaviour.

⁵ Note that this uses the law of iterated expectations, which states that $E(E(p)) = E(p)$: you cannot be systematically wrong in your guess.

⁶ The mathematical intuition is as follows: because the firm is optimising in point A, the derivative of its income with respect to price is set at zero, and any gain from changing prices, from the firm’s perspective, will be of second order. But point A does not correspond to a social optimum, because of imperfect competition, and that means that the effects of a change in prices on social welfare will be of first order.

⁷ Somewhat confusingly, people often refer to the modern New Keynesian view of fluctuations and to DSGE models as synonyms. However, it is pretty obvious that RBC models are dynamic, stochastic, and general-equilibrium too! We prefer to keep the concepts separate, so we will always refer to New Keynesian DSGE models.

⁸ See Benhabib et al. (2001b), appendix B, for a full derivation in continuous time.

⁹ Implicit in this equation is the assumption that firms discount future profits at the household rate of discount.

¹⁰ Leibniz’s rule? Why, of course, you recall it from calculus: that’s how you differentiate an integral. If you need a refresher, here it is: take a function $g(x) = \int_{a(x)}^{b(x)} f(x, s) ds$, the derivative of g with respect to x is: $\frac{dg}{dx} = f(x, b(x)) \frac{db}{dx} - f(x, a(x)) \frac{da}{dx} + \int_{a(x)}^{b(x)} \frac{df(x, s)}{dx} ds$. Intuitively, there are three components of the marginal impact of changing x on g : those of increasing the upper and lower limits of the integral (which are given by f evaluated at those limits), and that of changing the function f at every point between those limits (which is given by $\int_{a(x)}^{b(x)} \frac{df(x, s)}{dx} ds$). All the other stuff is what you get from your run-of-the-mill chain rule.

¹¹ Whatever i^n was initially, in drawing the figure below we assume the intercept does not change in response to the shock — that is, it does not fall as the natural interest rate drops temporarily.

References

- Benhabib, J., Schmitt-Grohé, S., & Uribe, M. (2001a). Monetary policy and multiple equilibria. *American Economic Review*, 91(1), 167–186.
- Benhabib, J., Schmitt-Grohé, S., & Uribe, M. (2001b). The perils of Taylor rules. *Journal of Economic Theory*, 96(1-2), 40–69.
- Benhabib, J., Schmitt-Grohé, S., & Uribe, M. (2002). Avoiding liquidity traps. *Journal of Political Economy*, 110(3), 535–563.

- Blanchard, O. & Galí, J. (2007). Real wage rigidities and the New Keynesian model. *Journal of Money, Credit and Banking*, 39, 35–65.
- Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3), 383–398.
- Clarida, R., Gali, J., & Gertler, M. (1999). The science of monetary policy: A New Keynesian perspective. *Journal of Economic Literature*, 37(4), 1661–1707.
- Galí, J. (2015). *Monetary policy, inflation, and the business cycle: An introduction to the New Keynesian framework and its applications*. Princeton University Press.
- Hicks, J. R. (1937). Mr. Keynes and the “classics”; A suggested interpretation. *Econometrica*, 147–159.
- Lucas, R. E. (1973). Some international evidence on output-inflation tradeoffs. *The American Economic Review*, 63(3), 326–334.
- Romer, D. (2018). *Advanced macroeconomics*. McGraw Hill.
- Smets, F. & Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the Euro area. *Journal of the European Economic Association*, 1(5), 1123–1175.
- Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy* (Vol. 39, pp. 195–214). Elsevier.
- Woodford, M. (2011). *Interest and prices: Foundations of a theory of monetary policy*. Princeton University Press.
- Yun, T. (1996). Nominal price rigidity, money supply endogeneity, and business cycles. *Journal of Monetary Economics*, 37(2), 345–370.

Unemployment

16.1 | Theories of unemployment

One of the most prominent features of the business cycle is the fluctuation in the number of people who are not working. More generally, at any given point in time there are always people who say they would like to work, but cannot find a job. As a result, no book in macroeconomics would be complete without a discussion on what determines unemployment.

The question of why there is unemployment is quite tricky, starting with the issue of whether it exists at all. Every market has frictions, people that move from one job to the next, maybe because they want a change, their business closed, or they were fired. This in-between unemployment is called *frictional unemployment* and is somewhat inevitable, in the same way that at each moment in time there is a number of properties idle in the real estate market.¹

Another difficulty arises when workers have very high reservation wages. This may come about because people may have other income or a safety net on which to rely.² When this happens, they report to the household survey that they want to work, and probably they want to, but the wage at which they are willing to take a job (their reservation wage) is off equilibrium.³ How should we classify these cases? How involuntary is this unemployment?

Now, when people would like to work at the going wages and can't find a job, and more so when this situation persists, we say there is *involuntary unemployment*. But involuntary unemployment poses a number of questions: why wouldn't wages adjust to the point where workers actually find a job? Unemployed individuals should bid down the wages they offer until supply equals demand, as in any other market. But that does not seem to happen.

Theories of unemployment suggest an explanation of why the labour market may fail to clear. The original Keynesian story for this was quite simple; wages were rigid because people have money illusion, i.e. they confuse nominal with real variables. In that view, there is a resistance to have the nominal wage adjusted downwards, preventing an economy hit by a shock that would require the nominal (and presumably the real) wage to decrease to get there. There may be some truth to the money illusion story⁴, but economists came to demand models with rational individuals. The money illusion story also has the unattractive feature that wages would be countercyclical, something that is typically rejected in the data.⁵

How to cite this book chapter:

Campante, F., Sturzenegger, F. and Velasco, A. 2021. *Advanced Macroeconomics: An Easy Guide*.

Ch. 16. 'Unemployment', pp. 243–258. London: LSE Press.

DOI: <https://doi.org/10.31389/lsepress.ame.p> License: CC-BY-NC 4.0.

The theories of unemployment we will see tell some more sophisticated stories. They can be classified as:

1. *Search/Matching models*: unemployment exists as a by-product of the process through which jobs that require different skills are matched with the workers who possess them.
2. *Efficiency-wage theories*: firms want to pay a wage above the market-clearing level, because that increases workers' productivity.
3. *Contracting/Insider-Outsider models*: firms cannot reduce the wage to market-clearing levels because of contractual constraints.

We will see that each of these stories (and the sub-stories within each class) will lead to different types of unemployment, for example frictional vs structural. They are not mutually exclusive, and the best way to think about unemployment is having those stories coexist in time and space. To what extent they are present is something that varies with the specific context. Yet, to tackle any of these models, we first need to develop the basic model of job search, to which we now turn.

A word of caution on unemployment data

Unemployment data is typically collected in *Permanent Household Surveys*. In these surveys, data collectors ask workers if they work (those that answer they do are called employed), if they don't work and have been looking for a job in the last x unit of time they are called unemployed. The sum of employed and unemployed workers comprise the labour force. Those that are not in the labour force are out of the labour force (this includes, children, the retired, students, and people who are just not interested in working). The key parameter here is the x mentioned above. It is not the same to ask about whether you looked for a job in the last minute (unemployment would probably be zero, considering that you are probably answering the survey!), in the last week or in the last year. The longer the span, the higher the labour force and the higher the unemployment rates. Methodological changes can lead to very significant changes in the unemployment figures.

16.2 | A model of job search

The theory of search solves the problem faced by an unemployed worker that is faced with random job offers. The solution takes the form of a reservation wage w_R . Only if the wage offer is larger than this reservation wage will the searcher take the job.

The specifics of the labour market have motivated the modelling of the process of job search. Obviously, how the market works depends on how workers *look* for a job. The theory of search tackles this question directly, though later on found innumerable applications in micro and macroeconomics.

Let's start with the basic setup. Imagine a worker that is looking for a job, and every period (we start in discrete time), is made an offer w taken from a distribution $F(w)$. The worker can accept or reject the offer. If he accepts the offer, he keeps the job forever (we'll give away with this assumption later). If he rejects the offer he gets paid an unemployment compensation b and gets a chance to try a

new offer the following period. What would be the optimal decision? Utility will be described by the present discounted value of income, which the worker wants to maximise

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t x_t, \quad (16.1)$$

where $x = w$ if employed at wage w , and $x = b$ if unemployed and $\beta = \frac{1}{1+\rho}$. This problem is best represented by a value function that represents the value of the maximisation problem given your current state. For example, the value of accepting an offer with wage w is

$$W(w) = w + \beta W(w). \quad (16.2)$$

It is easy to see why. By accepting the wage w , he secures that income this period, but, as the job lasts forever, next period he still keeps the same value, so the second term is that same value discounted one period. On the other hand, if he does not accept an offer, he will receive an income of b and then next period will get to draw a new offer. The value of that will be the maximum of the value of not accepting and the value of accepting the offer. Let's call U the value of not accepting (U obviously is motivated by the word unemployment):

$$U = b + \beta \int_0^{\infty} \max\{U, W(w)\} dF(w). \quad (16.3)$$

Since,

$$W(w) = w/(1 - \beta), \quad (16.4)$$

is increasing in w , there is some w_R for which

$$W(w_R) = U. \quad (16.5)$$

The searcher then rejects the proposition if $w < w_R$, and accepts it if $w \geq w_R$. Replacing (16.4) in (16.5) gives

$$U = \frac{w_R}{(1 - \beta)}. \quad (16.6)$$

But then combining (16.3) and (16.4) with (16.6) we have

$$\frac{w_R}{1 - \beta} = b + \frac{\beta}{1 - \beta} \int_0^{\infty} \max\{w_R, w\} dF(w). \quad (16.7)$$

Subtracting $\frac{\beta w_R}{1 - \beta}$ from both sides we get

$$w_R = b + \frac{\beta}{1 - \beta} \int_{w_R}^{\infty} (w - w_R) dF(w). \quad (16.8)$$

Equation (16.8) is very intuitive. The reservation wage needs to be higher than b , and how much depends on the possibility of eventually obtaining a good match; the better the prospects, the more demanding the searcher will be before accepting a match. On the other hand, a high discount factor, which means that waiting is painful will decrease the reservation wage.

An analogous specification can be built in continuous time. Here the value functions need to be understood as indicating the instantaneous payoff of each state. As time is valuable, these payoffs have to match the interest return of the value function. The analogous of the discrete time version are:

$$rW(w) = w, \quad (16.9)$$

$$rU = b + \alpha \int_0^{\infty} \max\{0, W(w) - U\} dF(w). \quad (16.10)$$

Notice how natural the interpretation for (16.9) is. The value of accepting a wage is the present discounted value of that wage $\frac{w}{r}$. The value of not accepting an offer is the instantaneous payment b plus α that needs to be interpreted as the probability with which a new opportunity comes along, times its expected value.⁶ We can also use our standard asset pricing intuition, which we first encountered when analysing the Hamiltonian. The asset here is the state “looking for a job” that pays a dividend of b per unit of time. The capital gain (or loss) is the possibility of finding a job, which happens with probability α and yields a gain of $\int_0^{\infty} \max\{0, W(w) - U\} dF(w)$.

As still $W(w_R) = U$,

$$W(w) - U = \left(\frac{w - w_R}{r} \right), \quad (16.11)$$

which can be replaced in (16.10) to give an expression for the reservation wage

$$w_R = b + \frac{\alpha}{r} \int_{w_R}^{\infty} (w - w_R) dF(w). \quad (16.12)$$

16.2.1 | Introducing labour turnover

The model can be easily modified to introduce labour turnover. If the worker can lose his job, we need to introduce in the equation for the value of accepting an offer the possibility that the worker may be laid off and go back to the pool of the unemployed. We will assume this happens with probability λ :

$$rW(w) = w + \lambda[U - W(w)]. \quad (16.13)$$

The equation for the value of being unemployed remains (16.10), and still $W(w_R) = U$. Because $rW(w_R) = w_R$ we know that $rU = w_R$. (16.13) implies that $W(w) = \frac{w + \lambda U}{(r + \lambda)}$, which replacing in (16.10) gives

$$rU = b + \frac{\alpha}{r + \lambda} \int_{w_R}^{\infty} [w - w_R] dF(w), \quad (16.14)$$

or

$$rW(w_R) = w_R = b + \frac{\alpha}{r + \lambda} \int_{w_R}^{\infty} [w - w_R] dF(w). \quad (16.15)$$

The reservation wage falls the higher the turnover; as the job is not expected to last forever, the searcher becomes less picky.

This basic framework constitutes the basic model of functioning of the labour market. Its implications will be used in the remainder of the chapter.

16.3 | Diamond-Mortensen-Pissarides model

The Diamond-Mortensen-Pissarides model describes a two-way search problem: that of workers and that of firms. A matching function $M(U, V)$ that relates the number of matchings to the unemployment and vacancies rates allows us to build a model of frictional unemployment.

We will put our job search value functions to work right away in one very influential way of analysing unemployment: thinking the labour market as a matching problem in which sellers (job-seeking workers) and buyers (employee-seeking firms) have to search for each other in order to find a match. If jobs and workers are heterogeneous, the process of finding the right match will be costly and take time, and unemployment will be the result of that protracted process.⁷

Let us consider a simple version of the search model of unemployment. The economy consists of workers and jobs. The number of employed workers is E and that of unemployed workers is U ($E + U = \bar{L}$); the number of vacant jobs is V and that of filled jobs is F . (We will assume that one worker can fill one and only one job, so that $F = E$, but it is still useful to keep the notation separate.) Job opportunities can be created or eliminated freely, but there is a fixed cost C (per unit of time) of maintaining a job. An employed worker produces A units of output per unit of time ($A > C$), and earns a wage w , which is determined in equilibrium. We leave aside the costs of job search, so the worker's utility is w if employed or zero if unemployed; the firm's profit from a filled job is $A - w - C$, and $-C$ from a vacant job.

The key assumption is that the matching between vacant jobs and unemployed workers is not instantaneous. We capture the flow of new jobs being created with a matching function

$$M = M(U, V) = KU^\beta V^\gamma, \quad (16.16)$$

with $\beta, \gamma \in [0, 1]$. This can be interpreted as follows: the more unemployed workers looking for jobs, and the more vacant jobs available, the easier it will be to find a match. As such, it subsumes the searching decisions of firms and workers without considering them explicitly. Note that we can parameterise the extent of the thick market externalities: if $\beta + \gamma > 1$, doubling the number of unemployed workers and vacant jobs more than doubles the rate of matching; if $\beta + \gamma < 1$ the search process faces decreasing returns (crowding).

We also assume an exogenous rate of job destruction, which we again denote as b . This means that the number of employed workers evolves according to

$$\dot{E} = M(U, V) - bE. \quad (16.17)$$

We denote a as the rate at which unemployed workers find new jobs and α as the rate at which vacant jobs are filled. It follows from these definitions that we will have

$$a = \frac{M(U, V)}{U}, \quad (16.18)$$

$$\alpha = \frac{M(U, V)}{V}. \quad (16.19)$$

The above describes the aggregate dynamics of the labor market, but we still need to specify the value for the firm and for the worker associated with each of the possible states. Here is where we will use

the intuitions of the previous section. Using the same asset pricing intuition that we used in (16.9) and (16.10), but now applied to both worker and firm, we can write

$$rV_E = w - b(V_E - V_U), \quad (16.20)$$

$$rV_U = a(V_E - V_U), \quad (16.21)$$

$$rV_V = -C + \alpha(V_F - V_V), \quad (16.22)$$

$$rV_F = A - w - C - b(V_F - V_V), \quad (16.23)$$

where r stands for the interest rate (which we assume to be equal to the individual discount rate). These equations are intuitive, so let's just review (16.20). The instantaneous value of being employed is the instantaneous wage. With probability b the worker can become unemployed in which case loses the utility $(V_E - V_U)$. The reasoning behind the other equations is similar, so we can just move on.

We assume that workers and firms have equal bargaining power when setting the wage, so that they end up with the same equilibrium rents:⁸

$$V_E - V_U = V_F - V_V. \quad (16.24)$$

16.3.1 | Nash bargaining

Let us start by computing the rents that will accrue to employed workers and employing firms, as a function of the wage, using (16.20)-(16.23):

$$r(V_E - V_U) = w - b(V_E - V_U) - a(V_E - V_U) \Rightarrow V_E - V_U = \frac{w}{a + b + r}, \quad (16.25)$$

$$r(V_F - V_V) = A - w - C - b(V_F - V_V) + C - \alpha(V_F - V_V) \Rightarrow V_F - V_V = \frac{A - w}{\alpha + b + r}. \quad (16.26)$$

The assumption of equal bargaining power (16.24) implies that the equilibrium wage must satisfy

$$\frac{w}{a + b + r} = \frac{A - w}{\alpha + b + r} \Rightarrow w = \frac{(a + b + r)A}{a + \alpha + 2b + 2r}. \quad (16.27)$$

The intuition is simple: a and α capture how easy it is for a worker to find a job, and for a firm to find a worker; their relative size determines which party gets the bigger share of output.

The equilibrium will be pinned down by a free-entry condition: firms will create job opportunities whenever they generate positive value. In equilibrium, the value of a vacant job will be driven down to zero. But how much is a vacant job worth to a firm? Using (16.22), (16.26), and (16.27) yields

$$\begin{aligned} rV_V &= -C + \alpha \frac{A - w}{\alpha + b + r} = -C + \alpha \frac{A - \frac{(a+b+r)A}{a+\alpha+2b+2r}}{\alpha + b + r} \\ &\Rightarrow rV_V = -C + \frac{\alpha}{a + \alpha + 2b + 2r} A. \end{aligned} \quad (16.28)$$

Now recall (16.18) and (16.19). We can turn these into functions of E , by focusing the analysis on steady states where E is constant. For this to be the case (16.17) implies that $M(U, V) = bE$, the numbers of jobs filled has to equal the number of jobs lost. It follows that

$$a = \frac{bE}{U} = \frac{bE}{\bar{L} - E}. \quad (16.29)$$

To find out what α is, we still need to express V in terms of E , which we do by using the matching function (16.16):

$$KU^\beta V^\gamma = bE \Rightarrow V = \left(\frac{bE}{KU^\beta} \right)^{\frac{1}{\gamma}} = \left(\frac{bE}{K(\bar{L} - E)^\beta} \right)^{\frac{1}{\gamma}}. \quad (16.30)$$

As a result, we obtain

$$\alpha = \frac{bE}{V} = \frac{bE}{\left(\frac{bE}{K(\bar{L} - E)^\beta} \right)^{\frac{1}{\gamma}}} = K^{\frac{1}{\gamma}} (bE)^{\frac{\gamma-1}{\gamma}} (\bar{L} - E)^{\frac{\beta}{\gamma}}. \quad (16.31)$$

Conditions (16.29) and (16.31) can be interpreted as follows: a is an increasing function of E because the more people are employed, the smaller will be the number of people competing for the new job vacancies and the easier it for an unemployed worker to find a job. Similarly, α is decreasing in E because the same logic will make it harder for a firm to fill a vacancy.

The final solution of the model imposes the free-entry condition, using (16.28), to implicitly obtain equilibrium employment:

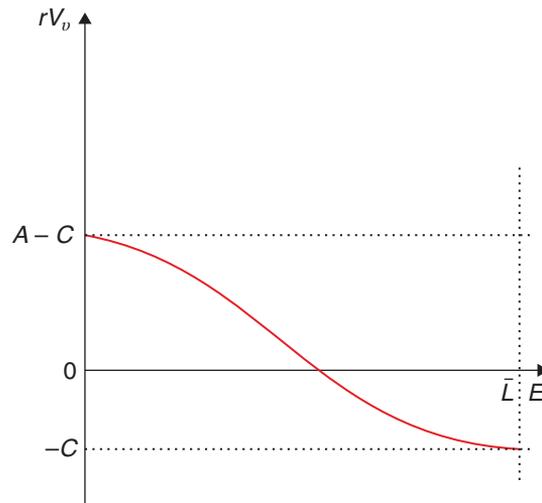
$$rV_V(E) = -C + \frac{\alpha(E)}{a(E) + \alpha(E) + 2b + 2r} A = 0. \quad (16.32)$$

What does the function $V_V(E)$ look like? It is negatively sloped, because

$$V'_V(E) = \frac{A}{r} \frac{\alpha'(E) [a(E) + 2b + 2r] - a'(E)\alpha(E)}{(a(E) + \alpha(E) + 2b + 2r)^2} < 0. \quad (16.33)$$

Intuitively, more employment makes it harder and more expensive to fill vacant jobs, reducing their value to the firm. When E is zero, filling a job is very easy, and the firm gets all the surplus $A - C$; when E is equal to \bar{L} (full employment), it is essentially impossible, and the value of the vacancy is $-C$. This can be plotted as in Figure 16.1.

Figure 16.1 Equilibrium employment in the search model



16.3.2 | Unemployment over the cycle

What happens when there is a negative shock in demand? We can illustrate this through analysing the effect of a drop in A , which a brief inspection of (16.28) shows corresponds to a leftward shift of the V_V curve in Figure 16.1 resulting in an increase in unemployment. Equation (16.27) in turn shows that the equilibrium wage will also fall as a result – though less than one for one, because of the effect of lower employment on the ease and cost of filling new vacancies. This means that the search models generate the kind of cyclical unemployment that characterises recessions.

We can also note that such a decrease in productivity will affect the equilibrium number of vacancies: (16.30) shows that there are fewer vacancies when employment is low. This seems to be consistent with the fact that you don't see many help wanted signs during recessions. (The negative link between unemployment and vacancies is often called Beveridge curve.) This happens in the model because a steady state with low employment is one in which the matching rate is low, and this means that firms will be discouraged from opening new vacancies that are likely to remain unfilled for a long time. This is precisely due to what is called thick market externalities: if there aren't many vacancies, people are unlikely to be looking for jobs, which discourages vacancies from being opened. People disregard the effect that their own job search or vacancy has on the thickness of the market, which benefits every other participant.

In any event, the unemployment described in these search models is what we call frictional unemployment – the by-product of the fact that it takes time to match heterogeneous workers and heterogeneous jobs. It is hard to think that long-term unemployment of the sort that often happens in real life will be due to this mechanism. Thus, we need other stories to account for a more stable unemployment rate. To these we now turn.

16.4 | Efficiency wages

The efficiency wage story builds on the idea that effort increases with wages. The firm may find it optimal to charge an above equilibrium wage to induce workers to exert effort. The chosen wage may lead to aggregate unemployment without firms having an incentive to lower it. Efficiency wages provide then a model of steady state unemployment.

The idea behind efficiency wages is that the productivity of labour depends on effort, and that effort depends on wages. Because of these links, firms prefer to pay a wage that is higher than the market equilibrium wage. But at this wage there is unemployment. The most basic version of this story – one that applies to very poor countries – is that a higher wage makes workers healthier as they can eat better. But there are many other ways to make the argument. For example, it is mentioned that Henry Ford paid his workers twice the running wage to get the best and reduce turnover, and, as we will see in the next section, sometimes firms pay a higher wage to elicit effort because they have difficulty in monitoring workers' effort.⁹

To see this, let us consider a general model in which the firm's profits are

$$\pi = Y - wL,$$

where

$$Y = F(eL),$$

with $F' > 0$ and $F'' < 0$. We denote by e the effort or effectiveness of the worker. The crucial assumption is that this effectiveness is increasing in the real wage:

$$e = e(w),$$

with $e' > 0$. With all these assumption we can rewrite the firm problem as

$$\text{Max}_{L,w} F(e(w)L) - wL,$$

which has first-order conditions

$$\frac{\partial \pi}{\partial L} = F'e - w = 0, \quad (16.34)$$

and

$$\frac{\partial \pi}{\partial w} = F'Le'(w) - L = 0. \quad (16.35)$$

Combining (16.34) and (16.35) we have

$$\frac{we'(w)}{e(w)} = 1.$$

The wage that satisfies this condition is called the *efficiency wage*. This condition means that the elasticity of effort with respect to wage is equal to one: a 1% increase in the wage translates into an equal increase in effective labour.

Why does this create unemployment? Notice that (16.34) and (16.35) is a system that defines both the wage and employment. If the optimal solution is w^* and L^* , total labour demand is NL^* where N indicates the number of firms. If the supply of labour exceeds this number, there is unemployment because firms will not want to reduce their wages to clear the market.¹⁰ We can also extend this model to include the idea that effort depends on the wage the firm pays relative to what other firms pay, or existing labour market conditions. Summers and Heston (1988) do this and the insights are essentially the same.

The model provides an intuitive explanation for a permanent disequilibrium in labour markets. What explains the relation between wages and effort? To dig a bit deeper we need a framework that can generate this relationship. Our next model does exactly that.

16.4.1 | Wages and effort: The Shapiro-Stiglitz model

The Shapiro-Stiglitz model builds a justification for efficiency wages on the difficulties for monitoring worker's effort. Labour markets with little monitoring problems will have market clearing wages and no unemployment. Markets with monitoring difficulties will induce worker's rents and steady state unemployment.

When you're at work, your boss obviously cannot perfectly monitor your effort, right? This means you have a moral hazard problem: the firm would like to pay you based on your effort, but it can only observe your production. It turns out that the solution to this moral hazard problem leads to a form of efficiency wages.

Following Shapiro and Stiglitz (1984), consider a model in continuous time with identical workers who have an instantaneous discount rate of ρ and a utility at any given instant that is given by

$$u(t) = \begin{cases} w(t) - e(t) & , \text{ if employed} \\ 0 & , \text{ otherwise} \end{cases} \quad (16.36)$$

where again w is wage and e is effort. For simplicity, we assume that effort can take two values, $e \in \{0, \bar{e}\}$. At any given point in time, the worker can be in one of three states: E means that she is employed and exerting effort ($e = \bar{e}$), S means that she is employed and shirking ($e = 0$), and U denotes unemployment. We assume that there is an exogenous instantaneous probability that the worker will lose her job at any instant, which is given by b . In addition, there is an instantaneous probability q that a shirking worker will be caught by the firm, capturing the monitoring technology. Finally, the rate at which unemployed workers find jobs is given by a . This is taken to be exogenous by individual agents, but will be determined in equilibrium for the economy as a whole. Firms in this model will simply maximise profits, as given by

$$\pi(t) = F(\bar{e}E(t)) - w(t) [E(t) + S(t)], \quad (16.37)$$

where $F(\cdot)$ is as before, and $E(t)$ and $S(t)$ denote the number of employees who are exerting effort and shirking, respectively.

In order to analyse the choices of workers, we need to compare their utility in each of the states, E , S , and U . Let us denote V_i the intertemporal utility of being in state i ; it follows that

$$\rho V_E = (w - \bar{e}) + b(V_U - V_E). \quad (16.38)$$

How do we know that? Again, we use our standard asset pricing intuition that we found in the first section of this chapter. The asset here is being employed and exerting effort, which pays a dividend of $w - \bar{e}$ per unit of time. The capital gain (or loss) is the possibility of losing the job, which happens with probability b and yields a gain of $V_U - V_E$. The rate of return that an agent requires to hold a unit of this asset is given by ρ . Using the intuition that that the total required return be equal to dividends plus capital gain, we reach (16.38). A similar reasoning gives us

$$\rho V_S = w + (b + q)(V_U - V_S), \quad (16.39)$$

because the probability of losing your job when shirking is $b + q$. Finally, unemployment pays zero dividends (no unemployment insurance), which yields¹¹

$$\rho V_U = a(V_E - V_U). \quad (16.40)$$

Solving the model

If the firm wants workers to exert effort, it must set wages such that $V_E \geq V_S$. The cheapest way to do that is to satisfy this with equality, which implies

$$\begin{aligned} V_E = V_S &\Rightarrow (w - \bar{e}) + b(V_U - V_E) = w + (b + q)(V_U - V_S), \\ &\Rightarrow (w - \bar{e}) + b(V_U - V_E) = w + (b + q)(V_U - V_E), \\ &\qquad\qquad\qquad \Rightarrow \bar{e} = q(V_E - V_U), \\ &\qquad\qquad\qquad \Rightarrow V_E - V_U = \frac{\bar{e}}{q} > 0. \end{aligned} \quad (16.41)$$

Note that wages are set such that workers are strictly better off being employed than unemployed. In other words, firms will leave rents to workers, so they are afraid of losing their jobs and exert effort as a result. What exactly is that wage? We can use (16.38), (16.40) and (16.41) to get

$$\begin{aligned}\rho V_E &= (w - \bar{e}) - b \frac{\bar{e}}{q}, \\ \rho V_U &= a \frac{\bar{e}}{q}.\end{aligned}\tag{16.42}$$

Subtracting these two equations yields

$$\rho \frac{\bar{e}}{q} = (w - \bar{e}) - b \frac{\bar{e}}{q} - a \frac{\bar{e}}{q} \Rightarrow w = \bar{e} + (\rho + a + b) \frac{\bar{e}}{q}.\tag{16.43}$$

Again this is very intuitive. The wage has to compensate effort, but then adds an extra which depends on the monitoring technology. For example if the monitoring technology is weak, the wage premia needs to be higher.

We know that a will be determined in equilibrium. What is the equilibrium condition? If we are in steady state where the rate of unemployment is constant, it must be that the number of workers losing their jobs is exactly equal to the number of workers gaining new jobs. If there are N firms, each one employing L workers (remember that all workers exert effort in equilibrium), and the total labour supply in the economy is \bar{L} , there are $a(\bar{L} - NL)$ unemployed workers finding jobs, and bNL employed workers losing theirs. The equilibrium condition can thus be written as:

$$a(\bar{L} - NL) = bNL \Rightarrow a = \frac{bNL}{\bar{L} - NL} \Rightarrow a + b = \frac{b\bar{L}}{\bar{L} - NL}.\tag{16.44}$$

Substituting this into (16.43) yields

$$w = \bar{e} + \left(\rho + \frac{b\bar{L}}{\bar{L} - NL} \right) \frac{\bar{e}}{q}.\tag{16.45}$$

This is the *no-shirking condition* (NSC), which the wage needs to satisfy in order to get workers to exert effort. Note that $\frac{\bar{L}-NL}{\bar{L}}$ is the unemployment rate in this economy, so that (16.45) implies that the efficiency wage will be decreasing in the unemployment rate; the greater the unemployment rate is, the more workers will have to fear, and the less their required compensation will be.¹² At full employment, an unemployed worker would instantly find a new job just like her old one, so she has nothing to fear from the threat of being fired. The premium is also decreasing in the quality of the monitoring technology, q , which also reduces the need for overcompensation.

We still need to figure out what L will be in equilibrium. Labor demand by firms will come from the maximisation of (16.37), which entails

$$\bar{e}F'(\bar{e}L) = w.\tag{16.46}$$

A graphic representation of the equilibrium is given in Figure 16.2.

In the absence of the moral hazard problem (or with perfect monitoring), the equilibrium occurs where the labour demand curve crosses the labour supply, which is horizontal at \bar{e} up until the full employment point \bar{L} , at which it becomes vertical. (The figure assumes that $\bar{e}F'(\bar{e}\bar{L}/N) > \bar{e}$.) This frictionless equilibrium point is denoted E^W , and it entails full employment. However, because of the moral hazard problem, the firms will optimally choose to leave rents to workers, which in turn means that some workers will be left unemployed because the wage rate will not go down to market-clearing levels.

Figure 16.2 Shapiro-Stiglitz model

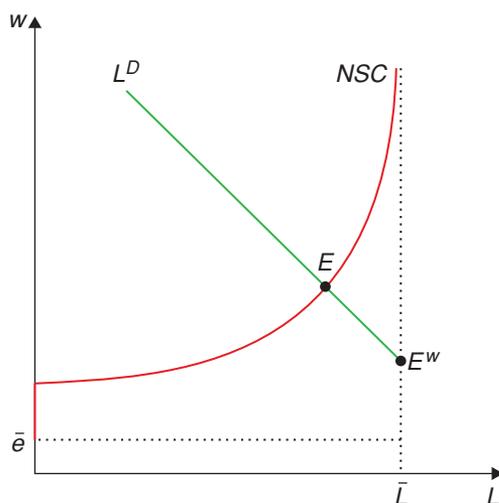
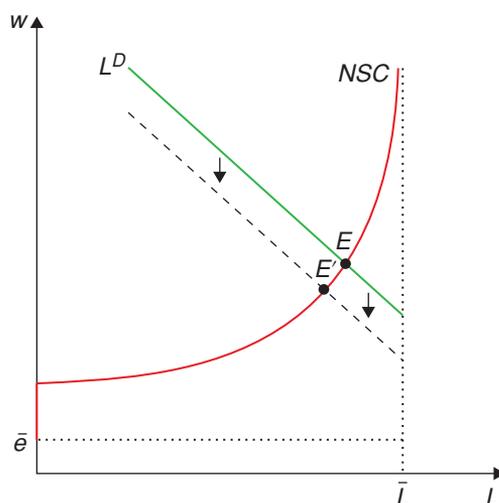


Figure 16.3 A decrease in labor demand



What happens when there is a negative shock to demand? This can be seen in Figure 16.3. We see that unemployment goes up, and the real wage goes down (as higher unemployment makes it cheaper to induce workers' effort).

Note that this unemployment is inefficient, as the marginal product of labour exceeds the marginal cost of effort. The first-best allocation is for everyone to be employed and to exert effort, but this cannot be simply implemented because of the informational failure.

This model, or rather a simple extension of it, is also consistent with a widespread feature of many labour markets, particularly (but not exclusively) in developing countries: the presence of dual labour markets. Suppose you have a sector where effort can be monitored more easily, say, because output is

less subject to random noise, and another sector (say, the public sector) where monitoring is harder. Then those working in the latter sector would earn rents, and be very reluctant to leave their jobs.

This model has some theoretical difficulties (e.g. Why don't people resort to more complicated contracts, as opposed to a wage contract? What if the problem is one of adverse selection, e.g. unobservable abilities, as opposed to moral hazard?) and empirical ones (e.g. calibration suggests that the magnitude of employment effects would be quite small). But it is still one of the better-known stories behind efficiency wages.

16.5 | Insider-outsider models of unemployment

The insider-outsider story builds an institutional theory of unemployment: unionisation transforms the labour market into a bilateral wage negotiation that may lead to higher than equilibrium wages. The unemployed, however, cannot bid the wage down because they are excluded from the bargaining game.

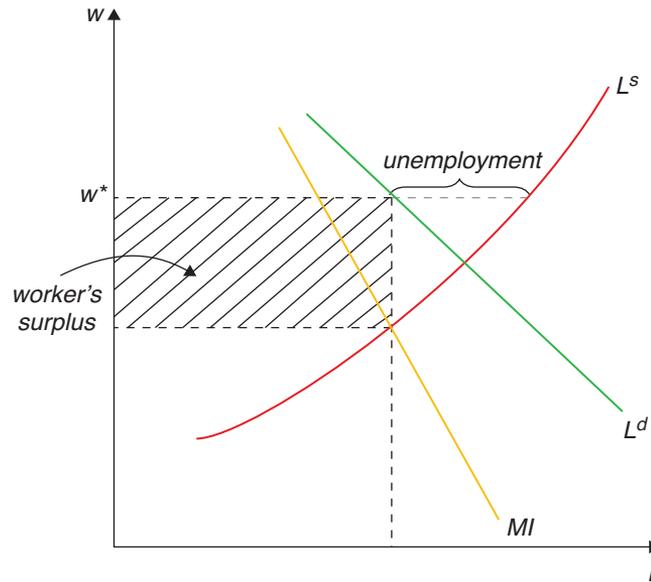
The insider-outsider model also speaks of a dual labour market, but for different reasons. A standard model of a dual market occurs when governments impose a minimum wage above the equilibrium rate leaving some workers unemployed. Alternatively, in the formal market, unionised workers choose the wage jointly with the firm in a bargaining process. The key assumption is that the firm cannot hire outsiders before it has all insiders (e.g. union members) working, and insiders have little incentive to keep wages low so that outsiders can get a job. As a result the equilibrium wage is higher than the market-clearing wage.

In these dual labour market stories, we may want to ask what is going on in the labour market for outsiders. That, in fact, is a free market, so there should be full employment there. In fact, for most developing countries unemployment is not a big issue, though a privileged number of formal workers do have higher wages. In other words, for the insider-outsider story to generate high economy-wide unemployment, you need the economy to have a very small informal sector. Examples could be European countries or Southern African countries.

At any rate, to see the insider-outsider story in action as a model of unemployment, consider an economy where all workers are unionised so that aggregate labour demand faces a unique supplier of labour: the centralised union. In Figure 16.4 we show that, as with any monopolist, the price is driven above its equilibrium level, and at the optimal wage there is excess supply of labour (unemployment). Notice that if the demand for labour increases the solution is an increase in the wage and in employment, so the model delivers a procyclical wage.

The role of labour market regulations on the functioning of the labour market is a literature with strong predicament in Europe, where unionisation and labour regulation were more prevalent than, say, in the U.S. In fact, Europe showed historically a higher unemployment rate than the U.S., a phenomenon called Eurosclerosis.

The literature has produced a series of interesting contributions surrounding labour market rigidities. One debate, for example has do to with the role of firing costs on equilibrium unemployment. Increasing firing costs increases or decreases unemployment? It increases unemployment, some would claim because it makes hiring more costly. Others would claim it reduces unemployment because it makes firing more costly. Bentolila and Bertola (1990) calibrated these effects for European labour markets and found that firing costs actually decrease firing and reduce the equilibrium unemployment rate. The debate goes on.

Figure 16.4 The distortion created by the insider

16.5.1 | Unemployment and rural-urban migration

Inspired by the slums of Nairobi, which swelled even further as the nation developed, Harris and Todaro (1970) developed the concept that unemployment was a necessary buffer whenever there were dual labour markets. It is a specific version of the insider-outsider interpretation. According to Harris and Todaro, there is a subsistence wage (back in the countryside) (w_s) that coexists with the possibility of a job in the formal sector (w_f). For the market to be in equilibrium, expected wages had to be equalised, i.e.

$$pw_f = w_s, \quad (16.47)$$

where p is the probability of finding a job in the formal sector. How is p determined? Assuming random draws from a distribution we can assume that

$$p = \frac{E}{E + U}, \quad (16.48)$$

where E stands for the total amount of people employed, and U for the total unemployed. Solving for U , using (16.47) in (16.48) we obtain that

$$U = E \left(\frac{w_f - w_s}{w_s} \right),$$

i.e. the unemployment rate is a function of the wage differential.

16.6 | What next?

The search theory of unemployment is well-covered in Rogerson et al. (2005) *Search-Theoretic Models of the Labor Market: A Survey*. Two textbooks you need to browse if you want to work on the topics discussed in this chapter are Pissarides (2000) *Equilibrium Unemployment Theory* and, of course, the graduate level textbook by Cahuc et al. (2014) *Labor Economics*. This will be more than enough to get you going.

Notes

- ¹ Lilien (1982) provides evidence that regions in the U.S. with larger sectorial volatility have higher unemployment rates, providing some evidence in favour of the frictional unemployment hypothesis.
- ² The running joke in the UK in the 70's claimed that if you were married and had three children you could not afford to be employed. In fact, there is ample evidence that the end of unemployment benefits strongly change the probability of someone finding a job.
- ³ A relevant case of this would be South Africa, which has unemployment rates in the upper 20's. In South Africa, large portions of the population live in communities far away from city centers, a costly legacy of the Apartheid regime, making commuting costs expensive in money and time. At the same time, large transfers to lower income segments combine with transportation costs to generate large reservation wages. People look for a job but find it difficult to find one that covers these fixed costs, leading to high and persistent unemployment.
- ⁴ A nice set of empirical experiments showing that nominal illusion is quite pervasive can be found in Shafir et al. (1997).
- ⁵ Another reason why countercyclical real wages are problematic can be seen with a bit of introspection: if that were the case, you should be very happy during a recession, provided that you keep your job – but that doesn't seem to be the case!
- ⁶ Technically, we would call this a *hazard rate* and not probability, as it is not limited to $[0,1]$. We abuse our language to aid in the intuition.
- ⁷ This is but one example of a general kind of problem of two-sided markets, which can be used to study all sorts of different issues, from regulation to poverty traps. The unemployment version was worth a Nobel prize in 2010 for Peter Diamond, Dale Mortensen, and Chris Pissarides, and the analysis of two-sided markets is one of the main contributions of 2014 Nobel laureate Jean Tirole.
- ⁸ This result is not arbitrarily imposed, but an application of the axiomatic approach to bargaining of Nash Jr (1950).
- ⁹ Akerlof and Yellen (1986) provide a comprehensive review of the literature on efficiency wages.
- ¹⁰ The mathematical intuition for why firms have a low incentive to adjust wages is similar to the argument for the effects of small menu costs under imperfect competition, which we saw in the last handout: because firms are in an interior optimum in the efficiency-wage case, the first-order gains from adjusting wages are zero.
- ¹¹ This assumes that, if employed, the wages will be enough that the worker will want to exert effort, which will be the case in equilibrium.
- ¹² This is not unlike the Marxian concept of the reserve army of labor!

References

- Akerlof, G. A. & Yellen, J. L. (1986). *Efficiency wage models of the labor market*. Cambridge University Press.
- Bentolila, S. & Bertola, G. (1990). Firing costs and labour demand: How bad is Euroclerosis? *The Review of Economic Studies*, 57(3), 381–402.
- Cahuc, P., Carcillo, S., & Zylberberg, A. (2014). *Labor economics*. MIT Press.
- Harris, J. R. & Todaro, M. P. (1970). Migration, unemployment and development: A two-sector analysis. *The American Economic Review*, 60(1), 126–142.
- Lilien, D. M. (1982). Sectoral shifts and cyclical unemployment. *Journal of Political Economy*, 90(4), 777–793.
- Nash Jr, J. F. (1950). The bargaining problem. *Econometrica*, 155–162.
- Pissarides, C. A. (2000). *Equilibrium unemployment theory*. MIT Press.
- Rogerson, R., Shimer, R., & Wright, R. (2005). Search-theoretic models of the labor market: A survey. *Journal of Economic Literature*, 43(4), 959–988.
- Shafir, E., Diamond, P., & Tversky, A. (1997). Money illusion. *The Quarterly Journal of Economics*, 112(2), 341–374.
- Shapiro, C. & Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *The American Economic Review*, 74(3), 433–444.
- Summers, R. & Heston, A. (1988). A new set of international comparisons of real product and price levels estimates for 130 countries, 1950–1985. *Review of Income and Wealth*, 34(1), 1–25.