2

Trends and Cycles, Shocks and Stability

Alas! I knew how it would end: I've mixed the cycle with the trend,
And fear that, growing daily skinnier, I have at length become non-linear.
I scarcely now, a pallid ghost, can tell ex-ante *from* ex-post:
My thoughts are sadly inelastic, my acts incurably stochastic.

D. H. Robertson (1955)

Measures of the level of real activity have two primary characteristics: Over time, they trend upward, but on occasion they swing noticeably, declining before eventually recovering and then continuing more or less along an upward march. These fluctuations are called business cycles, and they have occupied the attention of economists at least since the early nineteenth century.

In the United States the "dating" of business cycle peaks and troughs – and hence of recessions and expansions – is handled by the private National Bureau of Economic Research (NBER), where a committee within the NBER study a number of monthly indicators of real activity in order to identify turning points in the series. Because the NBER wait until such turning points are clearly discernable, the announcement of these dates typically comes well after a turning point is reached. The indicators currently used by the NBER to date business cycles are real personal consumption, total industrial production, real personal income less (government) transfers, real wholesale and retail trade sales, and two measures of employment (of late, income and nonfarm payroll employment have tended to receive the most weight). The rationale for using a number of indicators reflects the NBER's definition of a recession, which is a significant downturn in real activity that is spread throughout the entire economy and that therefore manifests itself across a

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

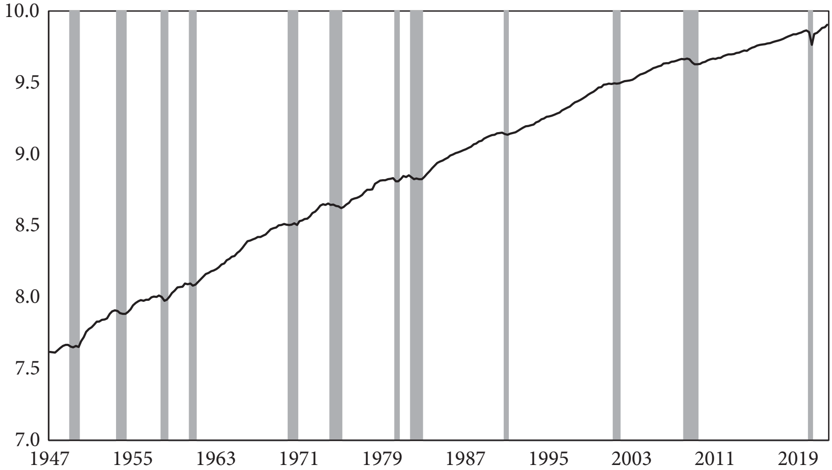

Figure 2.1 Natural log of real GDP. (Shaded bars denote recessions as defined by the NBER.)

broad set of aggregate measures of employment, output, and spending.[1] For reference, Figure 2.1 plots the log of real GDP together with the periods the NBER have identified as recessions (the gray regions).

By contrast, economists often seek a more statistically informed definition of the business cycle. One important reason is that the NBER dates do not identify the "recovery" phase of the business cycle, in which activity is retracing the losses incurred during the recession and has not yet settled back to a more normal rate of advance. A second reason is that the trend itself is of independent interest, so we would like to be able to extract and study that component of real activity as well. And a third reason is that most time-series modelling requires series to be stationary, which requires some way of dealing with the nonstationary ("trend") component.[2]

Like most topics in economics, there is essentially no consensus about the nature of business cycles, why they happen, or even how best to measure them. In this chapter and Chapter 3, we will take a look at some of the more plausible answers to these questions.

[1] Hence, the conventional definition of a recession as two back-to-back quarters of declining GDP is not used by the NBER. The NBER also use quarterly values of their cyclical indicators (along with real GDP and real gross domestic income) to identify the *quarter* of a turning point; on rare occasions, the month of the turning point won't fall in the quarter that the NBER identify as a turning point.

[2] See Watson (1986). Note that a trend in this context is just the nonstationary portion – for example, even though it can't rise or fall without limit, the unemployment rate can still be nonstationary and have a "trend" if its long-run mean is different over different periods.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

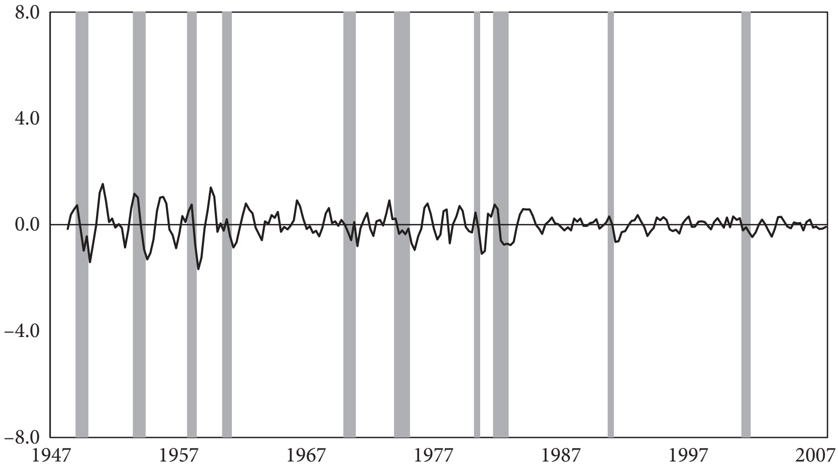

Figure 2.2 Cyclical component from a Beveridge–Nelson decomposition.

2.1 Measuring Trend and Cycle

For the sake of argument, let's assume that it makes sense to decompose a measure of real activity – say the log of real GDP, y_t – into a trend component τ_t and a cyclical component c_t, so that $y_t = \tau_t + c_t$. If we assume that the trend component follows a random walk with drift μ and that in the absence of any additional shocks cyclical fluctuations would eventually die away and real activity would return to its (stochastic) trend, then a reasonable estimate of the trend is given by

$$\tau_t = \lim_{n\to\infty} E_t y_{t+n} - n\mu \equiv \tau_t^{BN}. \tag{2.1}$$

This is the Beveridge–Nelson (1981) trend. Implicitly, it requires y_t to be I(1) (in particular, we don't want the drift term μ to change). Other than that, though, a decomposition like this is in principle consistent with various alternative views regarding the sources of business cycles. To estimate the Beveridge–Nelson trend, all we need is a forecasting model for y_t; after that, we can define the cycle as $y_t - \tau_t^{BN}$. One straightforward way to model y_t is as an ARIMA(p,1,q) process; if we do so, the cycle that we obtain looks like the solid line in Figure 2.2.[3]

For some economists, results like those in Figure 2.2 were rather baffling. Taken at face value, these results imply that the "cycle" is close to nonexistent,

[3] The specific model used here is an ARIMA(2,1,2) fit over the period 1947:Q4 to 2007:Q4; this choice of endpoint means that the estimates are not influenced by the 2007–2009 recession.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

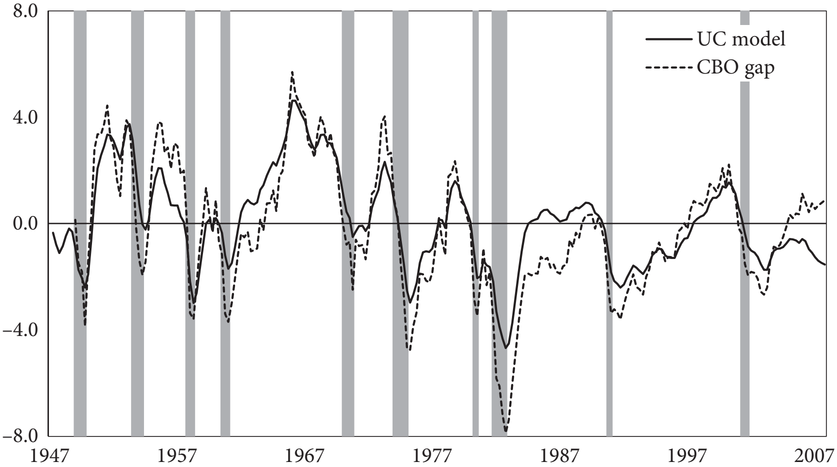

Figure 2.3 Cyclical component from an unobserved components model and CBO's output gap estimate.

and that movements in the "trend" account for most of the observed variation in output; in fact, many of the ordinary fluctuations of the cyclical component are as large as what occur during NBER-designated recessions. Moreover, alternative statistical detrending procedures that modelled τ_t and c_t as unobserved components (using a Kalman filter to back them out) found estimates of the cyclical component that were similar to the solid line in Figure 2.3 and so were both larger and more persistent than the Beveridge–Nelson cycle; such estimates also looked more like the kinds of output gap measures produced by the Congressional Budget Office (the dashed line in Figure 2.3) and other policy institutions.[4]

One reconciliation of these results was provided by Morley, Nelson, and Zivot (2003). These authors noted that a reasonably general trend–cycle decomposition could be written as

$$\begin{aligned} y_t &= \tau_t + c_t \\ \tau_t &= \mu + \tau_{t-1} + \eta_t \\ \phi_p(L)c_t &= \theta_q(L)\varepsilon_t, \end{aligned} \tag{2.2}$$

with the cycle explicitly modelled as an ARMA(p, q), and where the time-t innovations to the trend and cycle were allowed to be correlated. It turns out

[4] The Congressional Budget Office, or CBO, currently maintain an estimate of potential GDP that is intended to measure the level of production that would take place under full-employment conditions; the deviation of actual output from potential is referred to as an output gap. (We will look at this approach to measuring trend output in Chapter 4.)

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

that what we assume about this correlation is key. Unobserved components models of the sort used to generate Figure 2.3 typically assumed that the trend and cycle innovations had *no* correlation, which reflected the notion that the trend component was determined by the supply side of the economy while business cycles were largely the result of shocks to aggregate demand. But this assumption was not actually tested.

To make things concrete, let's assume that we fit an unobserved components model in which the cycle follows an AR(2) process (so $p = 2$ and $q = 0$) and where the correlation between η_t and ε_t (call it $\rho_{\eta\varepsilon}$) is restricted to be zero. (This is the same model that was used to generate the solid line in Figure 2.3.) In general, when $p = 2$ and $q = 0$ in model 2.2, it is possible to show that y_t will follow an ARIMA(2,1,2) process. However, imposing $\rho_{\eta\varepsilon} = 0$ restricts the coefficients in this process in ways that turn out to be rejected by the data. Specifically, if we allow $\rho_{\eta\varepsilon}$ to be an estimated parameter, we find that it is *negative* (with a value on the order of –0.9), and that the resulting unobserved components model generates a measure of the cycle that looks like Figure 2.2.[5] Equivalently, if we fit y_t using an *unrestricted* ARIMA(2,1,2) process and then use it to compute the Beveridge–Nelson decomposition for y_t – as was done to produce the solid line in Figure 2.2 – we obtain an estimate of the cycle that is identical to what we get from the unobserved components model with $\rho_{\eta\varepsilon}$ freely estimated.

Why does any of this matter? Well, if the trend and cycle innovations for a series like real GDP are strongly negatively correlated, it provides support for the hypothesis that real shocks are the dominant driving force behind economic fluctuations. As Morley, Nelson, and Zivot (2003) point out, a positive real shock – say a technological innovation – will shift up the long-run path of output in the period that it hits, leaving actual output to catch up from below. Hence, if real shocks dominate, positive shocks to the trend will be associated with negative shocks to the cyclical portion of the series, and vice versa. By contrast, shocks that have a transitory effect on output – a shift to an expansionary monetary policy stance, say – will only affect the cyclical component. So if these sorts of shocks were a major driver of output fluctuations, we would find that the correlation between the trend and cycle components is close to zero.

Before we reach such a strong conclusion, though, it is useful to consider some extensions to the original Beveridge–Nelson approach. In order to apply this particular trend–cycle decomposition, we need to have a

[5] It is possible to show that we will be able to estimate $\rho_{\eta\varepsilon}$ if the ARMA process for c_t has $p \geq q + 2$, which is satisfied here (see Morley, Nelson, and Zivot, 2003, pp. 236–237).

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

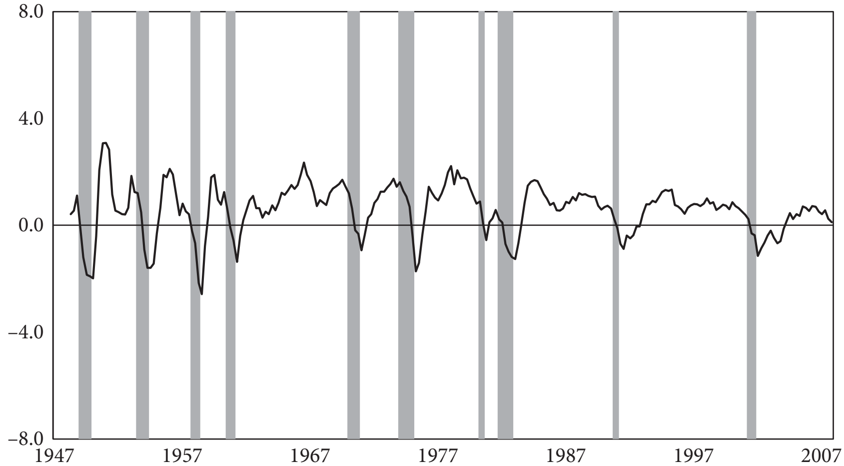

Figure 2.4 Cyclical component from a VAR-based Beveridge–Nelson decomposition.

good forecasting model for output growth (recall Equation 2.1). Moreover, as Evans and Reichlin (1994) point out, the Beveridge–Nelson cycle is essentially the forecastable "momentum" of the series, which means that the share of the observed fluctuations in output that we attribute to the cycle is closely related to how well we can forecast output growth.

To illustrate this point, note that the Beveridge–Nelson definition of the trend naturally carries over to a VAR context (instead of using an ARIMA model to predict a single series, we can use a VAR to predict a group of series). To keep things parsimonious, then, let's consider a two-variable VAR in real GDP and employment (we use employment because it is another measure of real activity that varies over the business cycle in a way that might be informative about output). The resulting "cycle" for real GDP is plotted in Figure 2.4; it's clearly evident that this measure of the cycle is smoother and more persistent than the cycle from the univariate model in Figure 2.2 (it also lines up better with the NBER dates), though its amplitude is not especially large.[6]

A second extension to the original Beveridge–Nelson procedure focuses on the trend component itself. Again, the reason that we end up with a cyclical component like the one in Figure 2.2 is that the decomposition attributes

[6] Employment is nonfarm payroll employment from the BLS establishment survey. Both output and employment enter the VAR as log-differences; the VAR has four lags and is estimated from 1948:Q2 to 2007:Q4.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

such a large portion of the observed movement in output to changes in the trend. Put differently, because the estimated cyclical component is so small (and continues to be so even after 2007 – not shown), the univariate decomposition implies a trend estimate that follows actual GDP extremely closely, including over most of the 2007–2009 recession and subsequent sluggish recovery.[7] But Kamber, Morley, and Wong (2018) argue that we might view it as *a priori* implausible that the Beveridge–Nelson decomposition implies such large variability in the trend. To give a flavor of their argument, assume that we use an AR(p) model to forecast output growth Δy_t,

$$\Delta y_t = \mu + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \varepsilon_t, \tag{2.3}$$

and note that the time-t change in the estimated Beveridge–Nelson trend from a unit ε_t shock will be equal to $1/(1-\phi_1-\phi_2-\cdots-\phi_p)$, as this is the eventual increase in the *level* of y relative to the level that would obtain absent a shock. If we denote this term as ψ, then the ratio of the variance of a trend shock to the variance of the forecast error ε will equal ψ^2. Hence, if the sum of the AR coefficients is on the order of 0.3 (which isn't too unreasonable in US data), then this ratio – which can be thought of as a signal-to-noise ratio – will be around 2. What that implies, then, is that trend shocks are considerably more variable than the one-quarter-ahead forecast error for (log) real GDP.

Beyond the fact that it generates an output gap that looks silly, why might we view such a high signal-to-noise ratio with suspicion? After all, lag selection procedures generally suggest that an AR(1) or AR(2) model fits output growth reasonably well, and "silly" is ultimately in the eye of the beholder. What Kamber, Morley, and Wong argue, however, is that certain types of processes for output growth – such as an MA process with a near-unit root – will not be well captured by a finite-order AR process, while the parameters of richer ARMA-type models can be poorly tied down in finite samples. They therefore develop a "Beveridge–Nelson filter" that allows one to impose a low signal-to-noise ratio as a (dogmatic) prior, and also to control

[7] Extending the estimation period through 2019:Q4 for the unobserved components model (2.2) with uncorrelated trend and cycle causes it to break down completely: If the post-2007 data are included in the sample, the model estimates that output is 5 percent above trend, on average, for most of the sample period before plunging 5 percent below trend in the wake of the 2007–2009 recession. Here, the model's constant average trend growth rate implies that the trend rises too slowly over much of the sample and too quickly toward the end of the sample.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

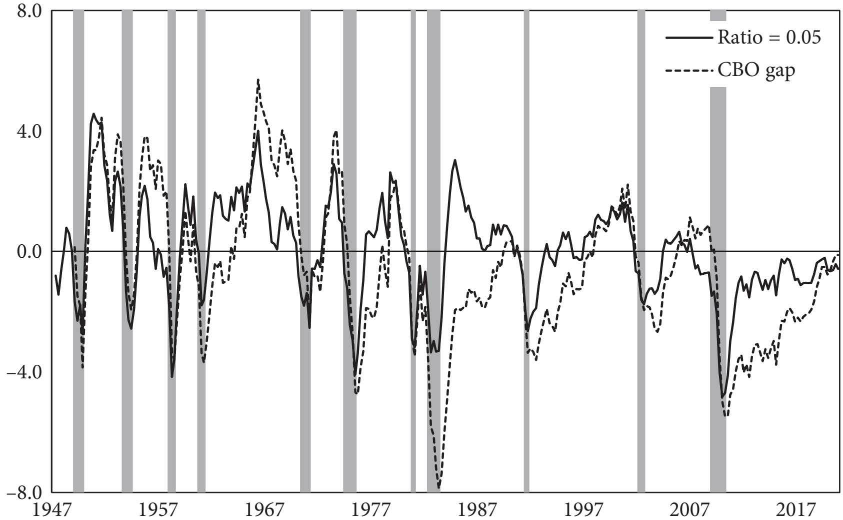

Figure 2.5 Cyclical component from Kamber–Morley–Wong Beveridge–Nelson filter (with signal-to-noise ratio fixed at 0.05) together with CBO's output gap estimate.

for possible trend breaks. Figure 2.5 plots the cyclical component obtained using their baseline filter specification with the signal-to-noise ratio fixed at 0.05 (a relatively small value).[8] The cycle is reasonably well correlated with the CBO output gap (also plotted in the figure through 2019:Q4), though some important differences are apparent (namely, the estimated depth of the 1981–1982 recession and the speeds of recovery from the 1990–1991 and 2007–2009 recessions).

We have now reached the point where we've been able to coax a respectable business cycle out of the Beveridge–Nelson decomposition, though to do so we have needed to venture uncomfortably close to the realm of vulgar curve fitting.[9] Why is this accomplishment worth celebrating? Well, from either a time series or an economic perspective, this definition of the trend (or long-run mean) is extremely intuitive and compelling – so much so that it's not easy to come up with other sensible definitions. In economic terms, the concept seems very close to an equilibrium notion – in other words, the value that a variable would return to once any adjustment mechanisms

[8] Kamber, Morley, and Wong (2018) propose an automatic selection procedure, which in this case yields a signal-to-noise ratio of 0.24; using this value instead yields results that are very similar to those shown in Figure 2.5.

[9] That job is better left to the Hodrick–Prescott filter, which we discuss in Section 2.2.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

or other frictions have fully played through (and taking into account any persistent effects on the equilibrium itself). Empirically, the method dovetails well with the standard techniques and approaches of time-series analysis, while using a stochastic trend to model the persistent movement in an economic aggregate seems much more defensible than any sort of deterministic alternative. On the other hand, in order to obtain a time-series gap measure that roughly resembles the CBO's output gap, we have needed to drastically reduce the variability of the trend component. That fact raises two questions: First, are statistical detrending procedures like the ones just considered actually useful; and second, why should we view the CBO measure as an appropriate benchmark?

Taking the second question first, the magnitude and persistence of the CBO gap lines up well with another important cyclical indicator, the unemployment rate, and that fact alone is a strong argument in its favor.[10] Regarding the broader usefulness of statistical detrending procedures, the answer is "it depends." Many time series approaches to trend–cycle decomposition, especially the univariate ones, do seem to allow the trend to respond too much to persistent movements in the actual series (we'll see some other examples soon).[11] Ultimately, though, which approach you prefer will probably depend strongly on your priors. Those who believe that the productive capacity of the economy is an important determinant of actual output and is itself subject to large quarter-to-quarter fluctuations (say because of productivity, efficiency, or utilization shocks) will be comfortable with the idea that the trend is highly variable and the cycle relatively small. By contrast, those who think that full-employment output evolves more slowly (say because it is determined by things like growth in the labor force and other additions to existing productive capacity) and that business cycles largely reflect demand-driven departures from full employment will be less likely to accept a procedure that places all of the action in the trend.

If we do continue to confine ourselves to purely statistical methods for measuring the business cycle, there is one other approach to isolating the trend and cyclical components of a time series that is worth thinking about. The fact that we refer to the idea of a business *cycle* calls to mind something that has a more or less regular periodicity. While we wouldn't want to

[10] An exception is the 2007–2009 recession, where the peak-to-trough drop in the output gap implied by the CBO measure seems somewhat less pronounced than the corresponding rise in the unemployment rate when compared with what happened in previous recessions.

[11] That said, some sort of statistical detrending procedure is often used one way or another in more-structural approaches to measuring the economy's supply side.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

take this idea literally – earlier writers on business cycles referred to them as "recurrent but not periodic" – from 1945 to 2019 the NBER's dating implies that postwar US business cycles have ranged from six quarters to twelve years in duration, with most lasting less than ten years.[12] As a way of measuring the business cycle, therefore, we can apply a particular type of statistical filter (known as a bandpass filter) that takes a series and extracts the portion of its fluctuations that is attributable to cycles with periodicities that fall in this range.[13] The reason we might want to do something like this is that macroeconomic aggregates typically manifest relatively small spectral peaks at business-cycle frequencies; using a suitable filter can throw these cyclical movements into sharper relief. (In an important sense, this is why we seek to detrend a series in the first place.)

For this particular application therefore, we want a filter that "passes through" cycles with periods longer than (say) two years and shorter than (say) twelve years. The problem of how to approximate such a filter as a weighted moving average with a finite number of terms has been worked out by several authors – including Baxter and King (1999) and Christiano and Fitzgerald (2003) – under different approximation criteria. Figure 2.6 plots the results obtained by applying the Baxter–King bandpass filter to log real GDP (times 100) and the unemployment rate, using a window width of 81 quarters – so ten years on either side – and cutoffs equal to two and twelve years.[14] A clear contemporaneous correlation between the series is apparent (the correlation coefficient is –0.91), reflecting the tendency of these series to move closely together (but in opposite directions) over the business cycle.[15]

[12] The quote comes from Burns and Mitchell (1946). In their definition of a business cycle, Burns and Mitchell identified the duration of a business cycle as "more than one year to ten or twelve years," but prior to the Great Depression, Mitchell (1927) had used "three to about six or seven years." For the postwar United States (and at the time of this writing), if we define cycles on a trough-to-trough basis, their duration ranges from 2.3 to 10.8 years; on a peak-to-peak basis, the range is 1.5 to 12.2 years.

[13] We know from time-series analysis that we can express a variable as the weighted sum of periodic (cosine and sine) functions with different frequencies (or, equivalently, periods) – very intuitively, the idea is similar to a Taylor expansion, which uses a weighted sum of polynomials to approximate a function.

[14] A filter like this is a two-sided moving average. Using a wider window increases the filter's ability to pass through only the desired frequencies but means that more data are lost from the beginning and end of the sample. One way to deal with this problem is to "pad" the sample with additional observations in the form of forecasts and backcasts from a time-series model; in the figures, the filter is applied to the series starting in 1959:Q1 and ending in 2007:Q4, so only actual data are used.

[15] This phenomenon is related to an empirical regularity called Okun's law, which we will return to in Chapter 5.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

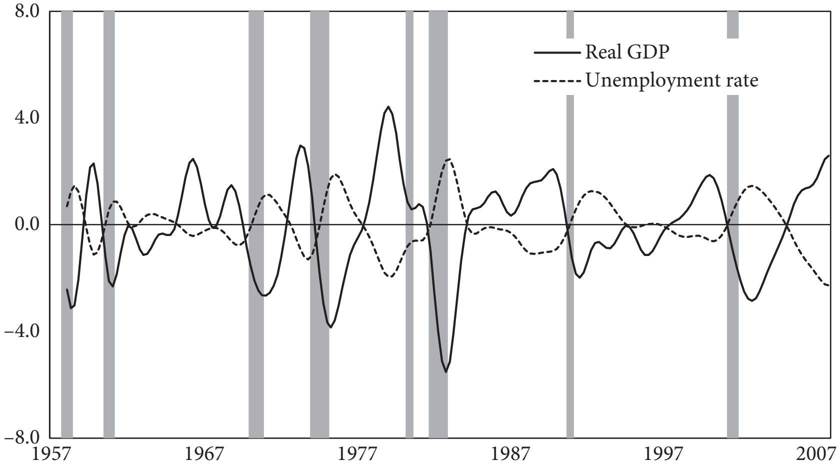

Figure 2.6 Filtered log real GDP (times 100) and unemployment rate obtained from the Baxter–King bandpass filter with cutoffs equal to two and twelve years (window width is 81 quarters).

One problem with an approach like this one is that it is probably too dogmatic regarding the periodicity of "the" business cycle. If we invoke the Burns and Mitchell definition (or use more recent NBER dates) to set the filter cutoffs, we are already buying into that definition of a cyclical contraction or expansion and forcing our filtered series to conform to it. As a check of this definition as it pertains to the comovement of real GDP and the unemployment rate, we can use the cross-spectrum to estimate pairwise coherences for these variables.[16] Figure 2.7 plots the coherence between log-differenced real GDP and the change in the unemployment rate over the 1959 to 2019 period; the x-axis of the chart is given as fractions of π, so the leftmost hill between 0.07 and 0.23 implies a large pairwise coherence at periods between 2.1 and 7.2 years, with a steep dropoff at around 9 years.[17] That's somewhat shorter than the average business cycle duration implied by the NBER dates (and used in the bandpass filter) – and the range is narrower as well – though given the imprecision surrounding estimates of the cross-spectrum, especially at lower frequencies, the correspondence isn't too bad.

[16] The coherence between two variables tells us how much of the variation in a series at a particular frequency can be explained by another series.

[17] The period (in quarters) is given by $2/f$, where f is the fraction of π given on the x-axis. (To compute the periods cited in the text, I used values of f expressed to more decimal places than are reported here.)

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

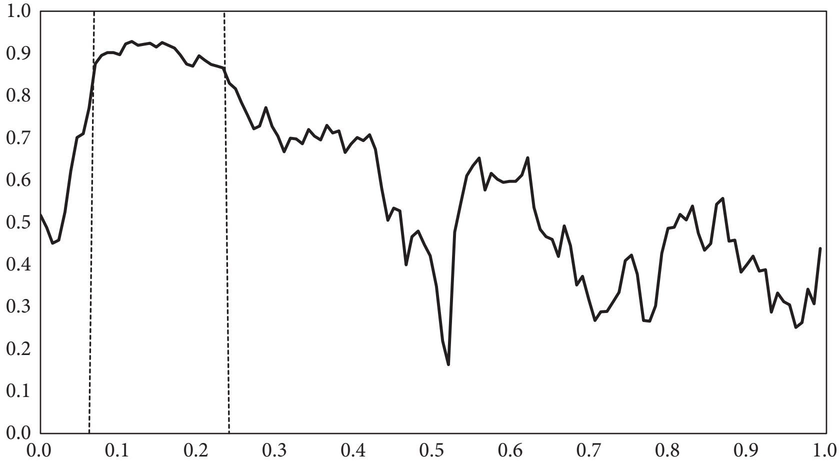

Figure 2.7 Estimated coherence between real GDP growth and the change in the unemployment rate. Values along the x-axis are fractions of π; vertical lines are at 0.07 and 0.23.

Another, related issue that arises with statistical filters can be seen from Figure 2.8, which plots the actual unemployment rate against its "low-pass" or "trend" component (that is, the component that reflects cycles with periods greater than twelve years). There is a distinct dip in the trend series over the first part of the 1960s as the reduction in the actual unemployment rate is interpreted by the filter as being too persistent to be a cyclical phenomenon (an interpretation that is difficult to square with the history of this period). Likewise, at the end of the sample the low-pass component is pulled up by the large and persistent rise in the unemployment rate that resulted from the 2007–2009 recession.

This last result highlights a deeper question, which is whether it is entirely sensible to treat cyclical and trend movements as largely separate phenomena. Even if we think that cyclical fluctuations are mostly demand-driven (as opposed to being the result of supply-side shocks), it is still the case that the demand and supply sides are likely to be linked over periods longer than a business cycle but shorter than the "long run" – for instance, as an investment boom translates into an increase in the economy's productive capacity, or as a prolonged slump causes workers to be persistently "scarred" as their skills atrophy and their connection to the labor force weakens. Conversely, we can imagine that certain types of supply-side disturbances (an oil price shock, say) could contribute to a recession by reducing aggregate demand. With only a dozen postwar business cycles at our disposal, each of which

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

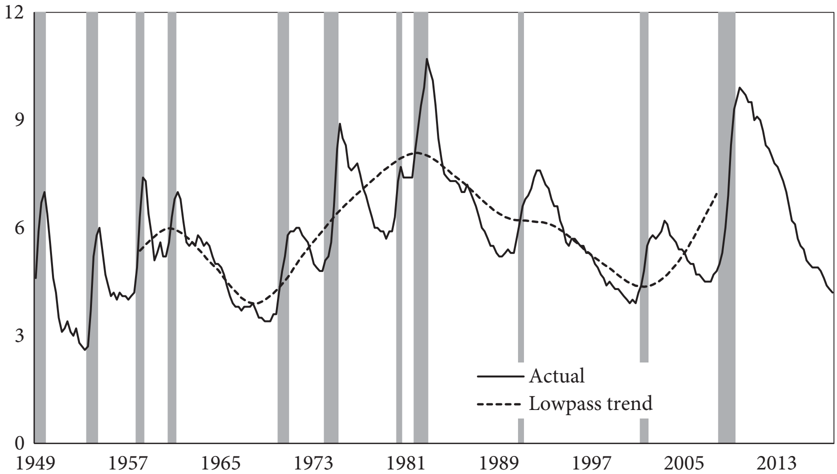

Figure 2.8 Unemployment rate and estimated trend from lowpass filter (cutoff is twelve years).

had its own specific features and causes, it is probably too much to expect that we can pin down the characteristics of "the" business cycle well enough to calibrate the parameters of a bandpass filter, or to use any purely statistical procedure to cleanly separate trends and cycles.[18]

2.2 Digression: "Why You Should Never Use the Hodrick–Prescott Filter"

An alternative method for detrending a time series that was popularized by Hodrick and Prescott (1997) involves fitting a smooth trend component through the series.[19] Specifically, for a series y_t Hodrick and Prescott suggested setting the trend component τ_t to minimize the following criterion,

$$\min \left\{ \sum_{t=1}^{T} \left(y_t - \tau_t\right)^2 + \lambda \sum_{t=1}^{T} \left(\Delta \tau_t - \Delta \tau_{t-1}\right)^2 \right\}. \tag{2.4}$$

18 Related issues have been discussed in the literature. For example, Murray (2003) shows that in a model like (2.2), a bandpass filter does not fully exclude the trend component from the measured cycle and so can overstate or understate the importance of cyclical dynamics at the assumed business-cycle frequencies. Murray argues that this problem arises whenever a nonstationary trend is present; however, the appendix to Trimbur and McElroy (2022) shows that the actual issue is that a stochastic trend contains frequency components that extend into the intermediate, "cyclical" portions of the spectrum and that are therefore passed through by a bandpass filter.

19 Hodrick and Prescott had initially proposed their method in a 1981 working paper.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

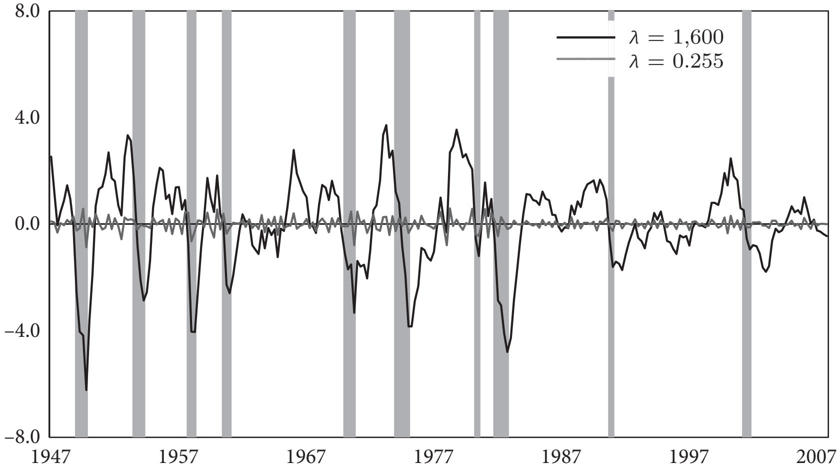

Figure 2.9 Cyclical component from Hodrick–Prescott filter with λ =1,600 (black line) and freely estimated λ =0.255 (gray line).

If we assume that y_t is in logarithms, then Equation (2.4) expresses a trade-off between keeping the trend as close as possible to the actual series while also keeping the growth rate of the trend smooth. The parameter λ governs this trade-off: When λ is zero, then the trend will be equal to y_t itself so as to minimize the first part of the criterion; as λ becomes very large, then the goal is to have the growth rate of the trend be nearly constant (in the limit where $\lambda \to \infty$, $\Delta\tau_t$ becomes a constant and the trend is a straight line). In most applications (and following Hodrick and Prescott's suggestion), λ is set equal to 1,600 in quarterly data; Figure 2.9 shows the resulting cyclical component when the filter is applied to 100 times the log of real GDP from 1947:Q1 to 2007:Q4.

It turns out that an exact closed-form solution to the minimization problem (2.4) can be computed for a specified λ; the solution is a matrix equation that implies that τ_t will be a linear function of all of the values of y. However, as Hodrick and Prescott pointed out, an alternative way to compute τ_t involves fitting a Kalman smoother to the following state-space representation,

$$\begin{aligned} y_t &= \tau_t + c_t \\ \tau_t &= 2\tau_{t-1} - \tau_{t-2} + \nu_t, \end{aligned} \tag{2.5}$$

where c_t denotes the cyclical component, c_t and ν_t are mutually uncorrelated white noise, and λ equals to the ratio of their variances, $\lambda = \sigma_c^2/\sigma_\nu^2$.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

As Hamilton (2018) demonstrates, either method will yield the same value of τ_t so long as the same value of λ is used and a diffuse prior is used for the initial state in the Kalman smoother.[20]

From this description, we can intuitively discern two potential problems with the Hodrick–Prescott filter.[21] First, the fact that a high degree of smoothing (a large value of λ) will cause the trend to be nearly linear hints that we could easily run into trouble, as linearly detrending a random walk is a recipe for generating spurious dynamics. (As Hamilton also demonstrates, the Hodrick–Prescott filter does in fact induce spurious dynamics even for the smaller values of λ that are typically used, with additional distortions introduced at the end of the sample.) More importantly, we might be suspicious that an estimation procedure that assumes the cyclical component is white noise (recall the state-space representation given by the system 2.5) would yield cycles that look like those in Figure 2.9, which are relatively smooth and persistent. The reason is that we are imposing a value of λ that is wildly at odds with what we obtain from freely estimating it; for example, fitting the system (2.5) by maximum likelihood yields $\sigma_c^2 = 0.125$ and $\sigma_\nu^2 = 0.491$. That implies a value of λ equal to 0.255 (not 1,600), a trend component that is essentially identical to the series itself, and the nonexistent cyclical component shown in Figure 2.9.

Hamilton (2018) also proposes a different way to isolate the cyclical component of a nonstationary time series that does not share the various problems associated with the Hodrick–Prescott filter (the method bears a passing relationship to the Beveridge–Nelson decomposition). Whatever one thinks about Hamilton's suggested alternative, though, one thing is clear: Anyone who shows you empirical results that use the Hodrick–Prescott filter is basically just wasting your time.

2.3 Shocks and Their Propagation

If we are willing to accept that business cycles represent large and persistent swings in real activity relative to a reasonably smooth trend, a natural next question is how to think about them – and especially recessions – in economic terms. In particular, we would like to be able to explain why it is that most business cycles tend to exhibit three "phases" or states, with sharp contractions in real activity (recessions) followed first by a period of rapid

[20] Note that we can normalize σ_ν^2 to be unity and set $\sigma_c^2 = \lambda$.

[21] These are discussed formally and in broader terms by Hamilton (2018); unsurprisingly, Hodrick (2020) strongly disagrees with Hamilton's conclusions.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

recovery and then by a return to a period of more moderate growth (the expansion state that the US economy typically finds itself in).

Since the 1930s, the dominant paradigm for thinking about macroeconomic fluctuations is one where erratic shocks or "impulses" feed through a stable propagation mechanism with a determinate (single) equilibrium.[22] The idea is that we can model the macroeconomy as a system of dynamic equations; this system (the *propagation mechanism*) summarizes how macro aggregates are related to each other, and is ultimately related to the assumed behavioral responses, technological constraints, and policy reaction functions that link together groups of agents in the economy. Shocks or impulses that arise from outside the system give rise to dynamic responses as the system is first moved away from equilibrium and then returns to it. (Of course, a continuous stream of these shocks will ensure that the system is never actually in equilibrium – or at least won't be in equilibrium for very long.)

In order to explain business cycles in these terms, though, we essentially have to argue that recessions reflect large negative shocks while recoveries represent the dynamic path that the economy follows as the effects of a recessionary shock play out. This view certainly has its adherents; Temin (1989), for example, even goes so far as to explain the Great Depression in these terms.[23] And more generally, this is how many real business cycle (RBC) and new-Keynesian models "explain" downturns. In the former, shocks to productivity or other real-side variables drive fluctuations in activity; the propagation mechanism reflects market-clearing responses to these real shocks.[24] New-Keynesian models assume a slightly different propagation mechanism (market clearing under the assumption of nominal rigidities), but estimated versions of these models typically rely on shocks to household discount

[22] This conception was developed by Frisch (1933) in an extremely influential paper. Frisch interpreted business cycles as resulting from repeated random shocks hitting a propagation mechanism that itself generated damped cycles, and claimed that a simple calibrated model could give rise to cycles with periods roughly equal to those observed at the time. However, as Zambelli (2007) points out, Frisch's solution is incomplete and there actually aren't any cycles in his model (the system simply returns monotonically toward its equilibrium state). Zambelli speculates that the development of macroeconomic theory might have taken a somewhat different course had this fact been known by Frisch's contemporaries.

[23] Temin explicitly eschews the explanation that the Great Depression reflected some sort of inherent instability in the economy. Instead, he argues that a very large shock – the First World War – changed elements of the world economic order in such a way that maintaining the prewar policy regime (the gold standard) required overly deflationary fiscal and monetary policies that were maintained for too long.

[24] One reason why RBC modellers favor the univariate Beveridge–Nelson characterization of the business cycle is that it leaves very little cycle to actually be explained through this mechanism.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

factors, investment-specific technological change, or wedges between the central bank's policy rate and the interest rate that enters spending decisions in order to generate large cyclical swings. (None of these shocks is observed; rather, they are inferred from the model's inability to fit the data.)

One reason to be skeptical of the "big negative shock" interpretation of recessions is that it seems to require that these shocks would all be one-sided; that is, we do not tend to see large *positive* exogenous shocks that sharply push up real activity – like a "negative recession" – when the economy is in its typical expansionary state. (Of course, it isn't impossible that the distribution of shocks to the economy has a long left tail, but then we would want to be able to explain or model the source of this skewness.) Another reason to be skeptical of this way of looking at recessions is that it assumes the propagation mechanism itself doesn't change as a result of the shock. The amplitude of most recessions and speed with which they occur suggest that they might involve something more than just the economy's typical continuous response to a shock; instead, recessions seem closer to a regime change, in which something happens to induce a discontinuous or nonlinear shift into a separate recession state. In other words, the propagation dynamics that result in a recession seem to be fundamentally different to those that prevail under ordinary circumstances.[25]

We can get a hint that something like this might be going on by considering a statistical model that explicitly permits such a regime change to take place. In 1989, Hamilton proposed describing the dynamics of a measure of real activity (say real GDP growth, Δy_t) with a Markov switching model:

$$\begin{aligned} \phi_p(L)\Delta y_t &= \mu(s_t) + \varepsilon_t \\ \mu(s_t) &= \mu_0 + \mu_1 s_t, \end{aligned} \tag{2.6}$$

where s_t is an indicator variable that equals zero or one. The idea is that the economy transitions between two regimes; in either regime, Δy_t follows an AR process, but the unconditional mean of the process is allowed to be different across regimes. The transition *between* regimes is then assumed to be governed by a Markov process with fixed transition probabilities,

$$\Pr[s_t = 1 \mid s_t = 1] = p; \quad \Pr[s_t = 0 \mid s_t = 0] = q. \tag{2.7}$$

[25] Note that a trend–cycle decomposition that is based on a model like Equation (2.2) implicitly assumes that cycles are symmetric since a single ARMA process is used to model the cyclical component.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

Hamilton originally wanted to use this approach to model a trend break in real GDP growth (specifically, the post-1970s productivity slowdown). However, when he fit the model to US real GDP, he obtained estimated probabilities for being in one of the states that looked very similar to the NBER recession bars shown in Figure 2.1. Hamilton realized that the model was capturing a switch into a *recession state*, with a reduction in mean GDP growth that goes away after the economy returns to a nonrecession state. In other words, recessions appeared to involve a discrete change in the process for output growth, which in this simple model *is* the propagation mechanism for shocks.[26]

We can get a flavor of Hamilton's results with a slightly different model that – like the NBER – uses several measures of real activity in order to pin down cyclical turning points.[27] Roughly, the idea is to model the behavior of the NBER's main cyclical indicators (nonfarm payrolls, real manufacturing and trade sales, industrial production, and real personal income less transfers, all of which are available monthly) as being driven by a common factor whose mean growth rate changes when the economy enters a recession state. Figure 2.10 plots the estimated probability that the economy was in a recession at a specified date; these probabilities line up tolerably well with the NBER's dating. The model's transition probabilities imply that we should expect the economy to be in a recession 10 percent of the time, with the average duration of a recession equal to 7.3 months and the average duration of an expansion equal to 67 months. For reference, the NBER's dating implies that the US economy has spent 13 percent of the time in a recession over the 1959–2019 period, with a mean recession duration of just under a year and a mean expansion duration of seventy-eight months.[28]

[26] An old piece of folk wisdom among macroeconomic forecasters holds that the best way to forecast real activity over the next couple of years is to determine whether a recession is in the offing. If a recession does seem likely, then you should project a downturn roughly in line with average postwar experience; otherwise, assume growth will continue at its current average pace. Hamilton's switching model shows why this approach tended to work reasonably well in practice (so long as you were good at calling a recession).

[27] The model is described in chapter 10 of Kim and Nelson (1999) and is fit over the period January 1959 to December 2019.

[28] The model's estimated transition probabilities are $p = 0.985$ and $q = 0.863$, where state zero is defined to be the recession state (these are the posterior medians). With a Markov process, the average fraction of time that the economy spends in a given state equals $x/(2 - p - q)$, with $x = 1 - p$ for a recession and $x = 1 - q$ for an expansion; once a given state is entered, its expected duration in months is given by $1/(1 - y)$, with $y = q$ for a recession and $y = p$ for an expansion.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

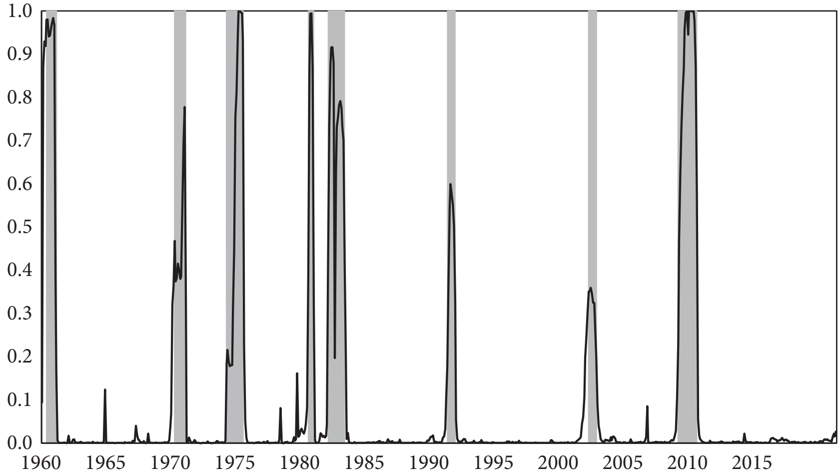

Figure 2.10 Recession probabilities implied by Markov switching model.

Taken literally, the assumption that contractions and expansions evolve according to a Markov process implies that duration dependence is not a feature of the business cycle: No matter how long an expansion has lasted, the probability of entering a recession is always the same (and vice versa). There have been a host of studies that attempt to explicitly test whether duration dependence characterizes recessions or expansions; in general, the results are inconclusive. That probably shouldn't be too surprising: Once again, we only have a dozen data points (postwar recessions) that we can look at, and the proximate causes of these recessions include such disparate events as financial crises, oil shocks, pandemics, and deliberately contractionary monetary policy. It seems unlikely then that duration dependence would be a *structural* feature of the business cycle, and while it is certainly possible that imbalances could build up over long expansions that would make a recession more likely, or that policy responses might become more aggressive in a prolonged slump, these kinds of mechanisms are better studied on their own rather than as "average" properties of cyclical fluctuations.

Empirical models like these capture one sort of asymmetry in the business cycle by allowing real activity to have different underlying dynamics in and out of recessions: In Hamilton's model and its variants, a recession reflects an intercept shift, so a recovery represents the restoration of a normal rate of trend growth. However, we might instead prefer a description of the business cycle that permits the expansion phase of the business cycle to have two distinct pieces: a relatively rapid recovery of activity that occurs once a

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

recession ends; and a slower, trend rate of advance that takes hold after the recovery is complete (and that represents the economy's usual expansion state).[29] Following section 5.6 of Kim and Nelson (1999), we can describe such a model as follows:

$$\begin{aligned} y_t &= \tau_t + x_t \\ \phi_p(L)x_t &= \gamma s_t + \varepsilon_t \\ \tau_t &= \mu_{t-1} + \tau_{t-1} + \eta_t \\ \mu_t &= \mu_{t-1} + \nu_t, \end{aligned} \tag{2.8}$$

where y_t is the logarithm of real GDP; ε_t, η_t, and ν_t are uncorrelated innovations; and s_t is a one–zero indicator that follows a Markov process.

This model is similar to the unobserved components model (Equation 2.2) that was used earlier to implement a trend–cycle decomposition, but with two notable differences. First, the rate of trend growth μ_t is allowed to vary (it is modelled as a driftless random walk rather than as a constant). Second, the cyclical component is modelled as an AR process together with a shift term that follows a Markov process. In addition, if we assume that $\sigma_\varepsilon = 0$, then the level of real activity that prevails during the expansion phase will be more like a "normal" or "ceiling" level, rather than a statistical trend for which fluctuations on either side of the trend net out to zero over long enough periods. (We will return to this notion when we discuss the productive potential of an economy in Chapter 4.) Estimates of the cyclical component x_t from this model are shown in Figure 2.11. The model reveals some evidence of asymmetry; the results also indicate that output remained below trend throughout the entire 1970s, which could reflect the model's difficulty in disentangling the post-1965 slowdown in trend output growth from the relatively large and frequent recessions seen over this period. The relatively slow recoveries that followed the 1990–1991 and 2001 recessions are also captured by the model; similarly, output appears to take somewhat longer to recover from the 2007–2009 recession compared with the equally deep Volcker recessions of 1980–1982, even though the model's estimate of trend output (not shown) slows noticeably in the wake of the financial crisis.

The various Markov switching models that we have been looking at represent simple statistical characterizations, not deep structural models of the

[29] This three-phase characterization of the business cycle is known as the "plucking" model of recessions (after Milton Friedman) or the "Joe Palooka" model (after Alan Blinder); as originally described, it tried to capture the idea that deeper recessions tended to be followed by correspondingly stronger recoveries.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

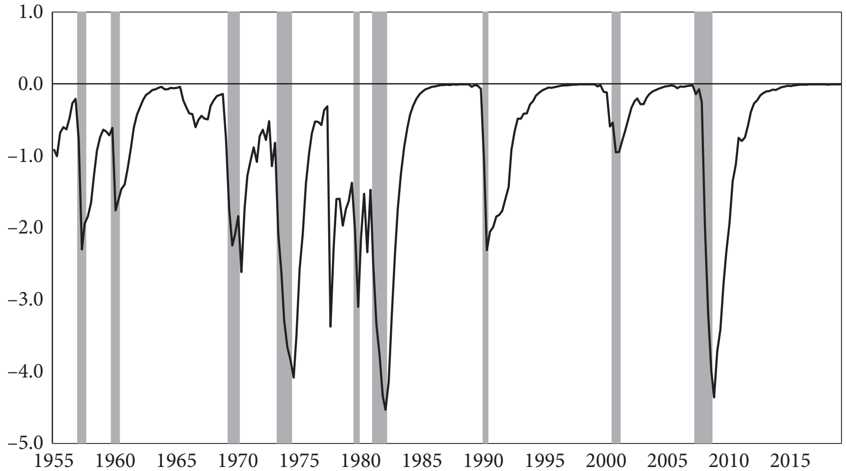

Figure 2.11 Deviations of log real GDP from trend implied by a "plucking" model of recessions.

business cycle. For example, including the pandemic period causes a model like the one used for Figure 2.10 to completely break down: Because the pandemic recession was so deep and rapid compared to previous recessions, the model doesn't identify any prior periods as belonging to a recession state. (Hamilton's original model broke down much earlier – around the late 1980s – likely reflecting both a change in the amplitude of business cycles and slower trend growth.) It is also very likely that the dynamics of a recovery are themselves determined by the *nature* of the shock that causes a recession – for example, the financial crisis almost certainly affected the economy in a way that made the recovery from the 2007–2009 recession more protracted.[30] All that said, switching models like these do strongly suggest that a recession is something more than just the simple propagation of an especially large shock, and instead involves a discontinuous shift in household and firm behavior following a shock. If so, then the idea that business cycles can be modelled as market-clearing equilibrium phenomena seems even harder to entertain.

[30] That observation hints that Lucas's (1977) hypothesis that "business cycles are all alike" might be wrong in an important sense, and – in the case of the slow recovery that followed the financial crisis – further suggests that the response of fiscal and monetary policy might be an important determinant of how a recovery plays out.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

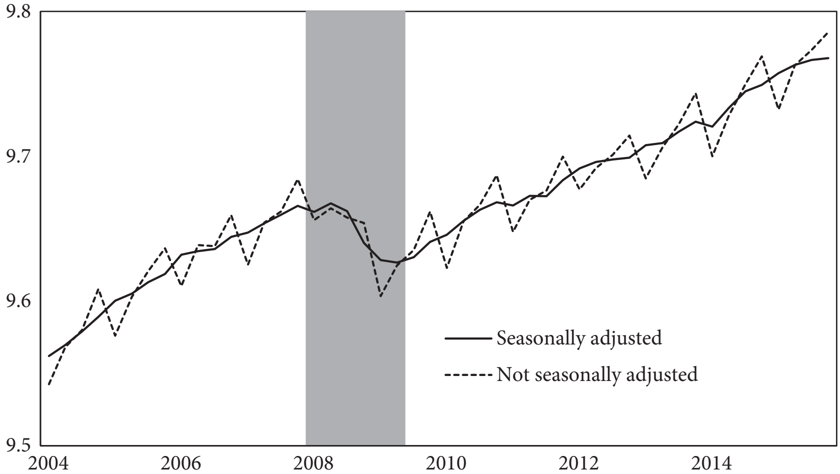

Figure 2.12 Seasonally adjusted and unadjusted log real GDP.

2.4 What Seasonal Cycles Suggest about Business Cycles

It seems that certain events (or combinations of events) are able to tip the economy into a recession state, while others aren't. What other evidence can we bring to bear regarding what's needed to induce an economic downturn?

One suggestive clue is given by Figure 2.12, which plots the logarithm of seasonally adjusted and not seasonally adjusted real GDP. Every year, the United States enters into a deep downturn in the first quarter, followed by a gradual recovery over the remainder of the year. These swings largely reflect consumption patterns associated with the winter holidays, and they are extremely large: On average, the peak-to-trough decline implied by the seasonal cycle (about 4 percent for the period shown) is as large as the *entire* decline in output that occurred during the 2007–2009 recession.[31]

The magnitude of seasonal cycles implies that the realization of a large shock is not by itself sufficient to push the economy into a recession state. The reason, of course, that seasonal swings don't result in actual recessions is that they are largely predictable; they represent anticipated shocks that, while very large, don't cause households and firms to significantly revise

[31] As a technical aside, not seasonally adjusted GDP doesn't actually exist: Published GDP is produced using seasonally adjusted source data, while not seasonally adjusted GDP is put together separately by replacing these source data with their seasonally unadjusted counterparts. However, the effects of seasonality are similarly pronounced in other seasonally unadjusted measures of real activity.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

their assessment of the current and prospective state of the economy, and that can also be planned for well in advance.[32] In this sense, then, the economy appears to be conditionally stable in the face of certain types of disturbances; namely, those that are not too far out of line with past experience and current anticipations. But whether it is stable more broadly – that is, whether there are natural corrective forces that lead the economy to recover in the wake of *any* shock – remains an open question.[33]

2.5 Recessions and Stability

As an empirical fact, the US economy has recovered from every postwar recession (though some recoveries have been noticeably slower than others). It's far from clear, however, why this would be. In periods where demand is depressed because output is low, and output is low because demand is, it seems as though the economy could easily get "stuck" in a low-activity state. And in theoretical terms, the fact that microtheorists have been unable to come up with a mechanism that would tend to push a model economy toward its (Walrasian) equilibrium also – and correctly – suggests that macroeconomists would have trouble coming up with a convincing theoretical description of what causes the economy to recover following a recession, let alone what would bring it back to a state of full employment.

In early Keynesian analysis, there were various "traps" the economy could find itself in, and no tendency for the economy to right itself absent active policy intervention.[34] This reflected Keynes's vision of the nature of the business cycle, in which a demand multiplier magnified the effects of swings in business investment and investment was assumed to be largely determined by the sentiment of businesspeople (and so largely exogenous). One of the earliest attempts to demonstrate that there *would* be self-correcting tendencies came from Pigou (1943), and was the centerpiece of Patinkin's (1965) analysis; the mechanism these authors had in mind was one where the disinflationary effects of a slump would cause real money balances (or real wealth more generally) to increase, thereby stimulating consumption.[35]

[32] This is not a new observation: Burns and Mitchell (1946, chapter 3) used it as a justification for seasonally adjusting economic aggregates before trying to analyze the business cycle.

[33] Seasonal cycles have been used to try to gain other insights into the nature of the business cycle, though not completely convincingly; see Miron and Beaulieu (1996) and the book-length treatment by Miron (1996) for two somewhat dated contributions.

[34] Keynes did allow that depreciation of the capital stock and liquidation of inventories would, under normal conditions, help to stimulate a recovery (Keynes 1936, pp. 317–318).

[35] If these references seem old, it's because they are – stability is as neglected a topic in macro as it is in micro.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

From an empirical and theoretical standpoint, the "Pigou–Patinkin" effect is unlikely to be an effective way to end a recession (or stabilize the economy). Empirically, wealth effects on consumption are too weak; and in the case of real balances, most of the money in the economy is generated by bank lending (and so is likely to move with real activity). Given those two facts, the decline in prices needed to appreciably boost consumption would be so large as to cause widespread bankruptcies among producers if labor costs were unchanged (thus reducing wealth), and would also put many households in distress if wages declined too and households had any sort of previously contracted debt obligations.[36]

Moreover, from the standpoint of general equilibrium theory Patinkin's "solution" to the stability problem turns out to be deficient. As Grandmont (1983) has shown, the real balance effect might be too weak *even in theory* to stabilize the economy. What is then needed is for agents' expectations regarding future prices and interest rates to respond in the "right" way (for some agents, not at all) to current prices – a condition that is highly implausible. Grandmont concludes that in a monetary economy, "… full price flexibility may not lead to market clearing after all … [because] there may not exist a set of prices and interest rates that would equate Walrasian demands and supplies."

We can get a flavor of this argument with the following simple example.[37] Consider an exchange economy with two periods and with individual endowments e_t, consumption c_t, prices p_t, and initial money holdings $\overline{m}$. A consumer therefore faces the following intertemporal budget constraint:

$$p_1 c_1 + p_2 c_2 \leq p_1 e_1 + p_2 e_2 + \overline{m}, \tag{2.9}$$

which is given as the line going through the points $\alpha\,\beta\,\gamma$ in Figure 2.13; the consumer's preferences are given by a utility function $u(c_1, c_2)$. Now assume that consumers expect future prices to be higher than current prices such that the price ratio p_2/p_1 exceeds the marginal rate of substitution at the endowment point α, as shown in panel A. If all consumers have these expectations, there will be pervasive excess demand in the market – everyone will want to consume more than their initial endowment e_1. Likewise, if we assume that consumers expect future prices to decline, a situation like the one in panel B can arise, where there is pervasive excess supply. Let's focus on the excess supply case. If p_1 declines in response to the excess supply,

[36] See Kalecki (1944).

[37] This discussion is taken from section 1.4 of Grandmont (1983).

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

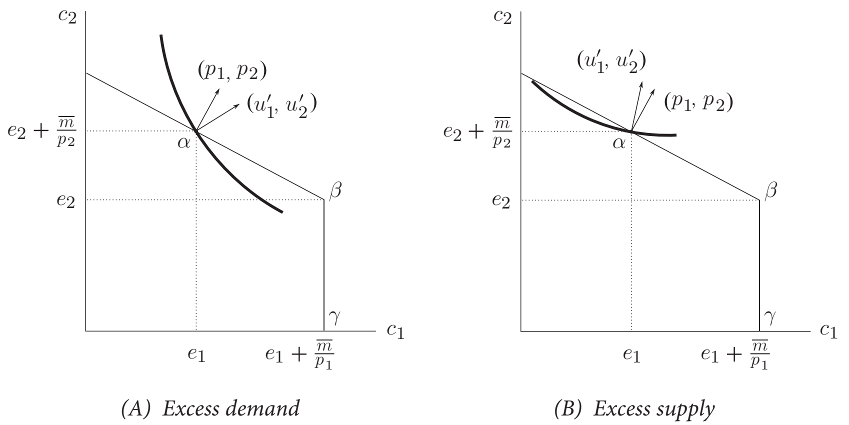

Figure 2.13 An example where the real-balance effect is too weak and price flexibility fails to deliver market clearing.

it will shift out line $\beta\,\gamma$; what happens to $\alpha\,\beta$ depends on how consumers' expectations of p_2 change in response. For example, if the decline in the price level is assumed to be permanent, then the *relative* price p_2/p_1 is unchanged and so $\alpha\,\beta$ shifts out in a parallel fashion. In this case, the real balance effect can be too weak to equilibrate the market.[38]

The preceding example makes it clear that price flexibility does not guarantee market clearing or a return to full employment. The following example and discussion, which is taken from Bénassy (1982, chapter 10), demonstrates how the presence of intertemporal decisionmaking can cause other sorts of problems. Consider the simplest possible setup: one firm, one consumer, and two periods. The consumer's utility function is given by

$$U(c_1, c_2, m_2) = (\alpha - \delta)\ln c_1 + (\alpha + \delta)\ln c_2 + \beta m_2 \qquad (2.10)$$

(putting the consumer's terminal money holding in their utility function is a trick to ensure the existence of a well-defined price level, and assuming $\delta > 0$ implies that the consumer wants to save in period 1). Production q is related to labor l by $q = F(l)$, and labor in both periods is in fixed supply and equal to $\bar{l}$. The firm can store any unsold output in inventory, and remits all profits to the consumer. Finally, there is a fixed money stock $\overline{m}$.

[38] This is immediately apparent for homothetic preferences; more generally, if the marginal rate of substitution along the vertical line at e_1 is bounded below by a positive value ε, then if $p_2/p_1 < \varepsilon$, no Walrasian equilibrium can exist for any value of the current price system p_1.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

Full employment in this economy implies that the firm produces $\overline{q} = F(\bar{l})$ in both periods; the Walrasian price and real wage are also the same in both periods and given by

$$p^* = \frac{\alpha \overline{m}}{\beta \overline{q}} \quad \text{and} \quad w^*/p^* = F'(\bar{l}). \tag{2.11}$$

However, even if the prevailing price and real wage are both equal to these values, it is very unlikely that a full-employment equilibrium will result. The reason is that demand in the two periods is

$$c_1 = \left(1 - \frac{\delta}{\alpha}\right)\overline{q} \quad \text{and} \quad c_2 = \left(1 + \frac{\delta}{\alpha}\right)\overline{q}, \tag{2.12}$$

which implies that the firm will need to produce $\overline{q}$ in period 1 and store $(\delta/\alpha)\overline{q}$ in inventories in (correct) anticipation that higher demand will materialize in period 2. Put differently, the firm will need to respond to a level of demand in period 1 that is less than its output by purposely accumulating inventories to meet higher demand next period, which is the same as saying that the firm must forecast that demand will exceed the full-employment level of output in the second period despite its having fallen short in the first period. Since the firm receives no signal from the future to tell it that demand will be higher next period, it is more likely to just cut production next period under the assumption that demand in period 2 will be similar to demand in period 1. And all this occurs despite prices and wages being at their "correct" intertemporal (Walrasian) values.

There is actually a strong case to be made that sticky wages or prices are largely irrelevant to business cycles. (As we will see in Chapter 6, though, sticky prices probably *are* needed to explain certain features of the inflation process.) In particular, there are other reasons that don't involve price maladjustment that can explain why workers and firms might be off their Walrasian supply and demand curves and why markets might fail to clear.[39] Similarly, the lack of any proof of stability even for the idealized Walrasian system suggests that an economy might not immediately return to full employment even when the shocks that pushed it into a recession have

[39] Examples include pervasive uncertainty; imperfect information and incomplete markets; the presence of nonallocative prices or institutions that favor quantity adjustments over price adjustments (see Bewley, 1999 for some evidence from the labor market); and the price system's broader inability to coordinate the actions of workers, producers, and consumers over nontrivial lengths of time.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

abated (at least, no one has ever come up with a convincing description of how this would happen, or why it's bound to occur in a market economy).

We might at least expect that the eventual adjustment of wages and prices in a new-Keynesian model would move economic activity back to its normal level. That also turns out to be not quite right. In many versions of these models, the monetary authority determines the economy's inflation rate and output gap through its choice of policy targets; the policy feedback rule then serves as a regulating mechanism to bring the economy toward these targets (assuming the nominal policy rate does not have to fall below zero to do so). In empirical implementations (such as Smets and Wouters, 2007), this process is helped along because recessions are "explained" in terms of fancifully labelled residuals that are themselves assumed to be stationary; hence, over history much of what brings the model out of a recession simply involves having these "structural" shocks die out or reverse themselves.

The slow recovery from the 2007–2009 recession, which contrasted with the rapid recoveries seen after previous deep recessions, renewed interest in microfounded optimizing models involving multiple equilibria. In these models, the economy can find itself stuck in a (locally stable) low-activity equilibrium that can only be exited through a policy intervention (such as a fiscal stimulus that pushes the economy back to a high-activity equilibrium) or an exogenous shift in expectations. Although models of this sort are intriguing, the specific mechanisms that are used tend to be either too contrived to take seriously or generate counterfactual predictions. For example, early work in this vein by Diamond (1982) assumes that production and consumption can be modelled as a search equilibrium, which doesn't seem like the best way to capture the insight that these activities invariably involve separate parties and so can be difficult to coordinate. Later work by Farmer (2010) uses a model with labor-market search to derive "Keynesian" conclusions about slumps; however, the model also predicts countercyclical real wages (p. 41) and – if the monetary authority respects the Taylor principle – an *upward*-sloping relation between wage growth and unemployment (p. 162).[40]

[40] Farmer does provide an interesting extension in the form of a "beliefs function" whereby self-fulfilling expectations influence aggregate demand through their effect on the value of capital and hence household wealth. (The model needs a wealth-based channel because it relies on a permanent-income framework.) While the notion that the public's beliefs can influence the level of activity probably has some truth to it, Farmer's conclusion that the Federal Reserve should target stock prices is harder to take seriously.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press

So what does cause a recession to end, and is an exit from a recession inevitable? The short answer is that nobody really knows. At best, we might plausibly argue that longer-run expectations about income and demand might help to anchor spending and production and contribute to a revival of "animal spirits" among consumers and producers; in addition, policy stimulus (especially when earlier policy actions were themselves a contributor to the recession) likely plays some role in fostering a recovery as well. But anyone who had read the *General Theory* could have probably told you that much.

https://doi.org/10.1017/9781009465779.003 Published online by Cambridge University Press