

Chapter 18

Nominal frictions and business cycles

Alisdair McKay and Morten Ravn

18.1 Introduction

The Great Depression was a key event in the development of macroeconomics as a field. To many observers, this episode revealed that market economies can perform very poorly as the mass unemployment of the depression was a clear sign that something was very wrong. It was then natural to ask what led to this failure and what could be done to improve the situation.

The Keynesian school of thought emerged as a result of the depression and provides a perspective on business cycle fluctuations of all magnitudes not just crises like the depression. One core idea of Keynesian economics is that the productive factors of the economy may not be used at the optimal level. A second core idea in Keynesian economics is that the level of nominal demand influences the level of production. A central piece of most economic theory is that, in the *long run*, nominal values are unimportant and simply a choice of units. However, going back at least to Hume's 1752 essay *Of Interest*, it has also been recognized that changes in nominal demand such as a change in the money supply may have real effects, at least temporarily, because the process of adjusting prices is neither perfect nor immediate.

This chapter serves several purposes. We will review the evidence that (a) prices adjust only infrequently and do not immediately react to changes in market conditions and (b) changes in nominal variables (here nominal interest rates) affect real outcomes such as real output. We also present the modern incarnation of the Keynesian tradition in the form of the workhorse New Keynesian model. Using this model, we will explain why imperfect price adjustment leads to a role for nominal demand in determining real outcomes, why the market equilibrium can deviate from the efficient level of production, and what types of policies can lead to better outcomes.

18.2 Empirical evidence on price rigidity

Here we will review the empirical evidence on price rigidity. We will concentrate on nominal rigidities in prices because this is also the focus of most of our theoretical analysis, but many of the same issues that apply to the prices of goods and services also apply to wages.

The research we review here examines data on the prices of individual goods as opposed to economy-wide price indices. We discuss aggregate evidence later in the chapter.

The literature on the adjustment of individual prices developed significantly since the early 2000's. In a line of important papers focused on the U.S., [Bils and Klenow \(2004\)](#), [Klenow and Kryvtsov \(2008\)](#) and [Nakamura and Steinsson \(2008\)](#) exploited access to the survey-based micro data underlying the construction of the Consumer Price Index (CPI) to document a series of facts about goods-level price adjustments. Much of this work has now been extended to other countries.¹

The data underlying the U.S. CPI is collected by the Bureau of Labor Statistics (BLS). The BLS surveys are carried out monthly or bi-monthly in 75 urban areas collecting price quotes from about 23,000 retail and service establishments for about 85,000 individual items. These individual items are then classified into about 300 goods categories known as entry level items (ELIs).

One feature of these data that has received particular attention is the frequency of price changes. Suppose a price has a probability $1 - \theta$ of changing each month. The expected time between price adjustments is then

$$\sum_{t=1}^{\infty} (1 - \theta)\theta^t t = \frac{\theta}{1 - \theta}.$$

Thus, observing the length of time a price of a good remains at a certain level is informative about the frequency of price changes for this good. In the data, such price change frequencies depend on the category of goods. For instance, magazine prices tend to change very infrequently while the prices of energy and food items tend to be adjusted much more frequently. The mean and median price duration of all items in the CPI are 6.2 months and 3.4 months, respectively (see [Klenow and Malin, 2010](#)).² The very considerable difference between the mean and the medians indicates significant heterogeneity across categories.

One of the features of these data is that prices have memory—sellers often return the price of an item to a level they have set in the past. One key source of this memory is sales, i.e. temporary discounts. The key problem presented by sales is whether such episodes should be excluded or not when estimating the frequency of price changes. If firms introduce sales as a direct response to a temporary change in market conditions, it would seem appropriate to include sales. In contrast, if the timing and size of the temporary price discount are not affected by economic conditions, price changes associated with sales may not reflect true flexibility in prices and should therefore be excluded. The literature therefore typically reports estimates with or without controlling for sales.

Many individual products change over time and are replaced by new versions. Such product replacements imply that for a subset of goods, while their prices have been observed in the past, prices cannot be observed this month (and future months) because they are no longer for sale. In constructing the CPI, the BLS will substitute a new comparable product

¹In this respect, an important effort has been made by the European Central Bank in its Inflation Persistence Network which has been summarized in [Dhyne, Alvarez, Le Bihan, Veronese, Dias, Hoffmann, Jonker, Lunnemann, Rumler, and Vilmunen \(2006\)](#).

²The statistics reported here are computed from the price change frequencies within ELIs weighted using CPI weights. These results relate to survey prices in the three largest cities in the U.S.

for the unavailable product. These product substitutions are often associated with price adjustments. It is an open question whether and how one should control for such product substitutions when computing measures of price rigidity. To the extent that product turnover presents an opportunity to adjust prices in the face of shocks, it would seem natural to include such price adjustments in measures of price flexibility. At the same time, by its very nature, when there is product substitution, the precise nature of the good changes and so the price change may be unrelated to market conditions.

When temporary sales prices are excluded, the estimates of the mean and median price durations rise to 8.0 months and 6.9 months, respectively (see [Klenow and Malin, 2010](#)). Hence, it is clear that whether sales are included or not makes a big difference especially if one concentrates on the median price duration estimates. These estimates treat price quotes as referring to the same item even in the face of product substitutions. Eliminating these as well, the mean and median price durations increase further to 10.1 and 8.3 months, respectively.

In summary, judging from the evidence across goods, there is considerable evidence of infrequent price changes with the range of empirically plausible estimates of price durations for the U.S. ranging from 6 months to 11 months depending on whether one focuses on the mean or the median and depending on the price measure. Estimates for European economies are typically in the upper range of this interval (see [Dhyne et al., 2006](#)).

What is not clear from this evidence is whether prices change infrequently because there are infrequent shocks to “market conditions” or because there are frictions that prevent prices from adjusting. [Eichenbaum, Jaimovich, and Rebelo \(2011\)](#) use data from a large retailer to estimate how the prices of products respond to the replacement cost of the goods and find that the price-cost margin varies within a fairly narrow range implying prices respond quickly to changes in market conditions.

Movements in individual prices are large relative to the changes in the aggregate price level. For example, the median absolute price change is 11.5 percent (see [Klenow and Kryvtsov, 2008](#)). The obvious interpretation of this fact is that most price changes reflect conditions in a specific market as opposed to macroeconomic conditions. It is possible that prices can respond strongly and quickly to sector-specific shocks while adjusting slowly and imperfectly to aggregate shocks.³ [Boivin, Giannoni, and Mihov \(2009\)](#) find evidence in support of this idea. They study disaggregated data on consumer and producer prices and examine how consumer prices respond to common and sector-specific shocks. They find that most price changes are driven by sector-specific developments rather than aggregate conditions and disaggregated prices are sticky in response to macroeconomic conditions.

18.3 The New Keynesian model

Early Keynesian theories were based on simple relationships between aggregate variables that did not have a close connection to the microeconomic decisions made by individuals.

³This may occur due to information frictions, e.g. sellers are unaware of all the macroeconomic developments that may impact the optimal price. Price adjustments may also be imperfect if firms want to keep their prices near those of other firms who are not simultaneously updating their prices. This force is known as “real rigidities” and it amplifies the effect of nominal rigidities.

Research in the 1980s and 1990s developed the **New Keynesian** model that is based on the same principles of optimization and equilibrium that underlie modern macroeconomics and incorporate frictions that interfere with the immediate and perfect adjustment of prices.

Nominal rigidities can be modeled in several ways. One natural way to model them is to assume prices are fixed for a certain period of time. For example, many employment contracts are reviewed and adjusted at an annual frequency (see [Grigsby et al., 2021](#)). Apartment rents are another type of transaction with long-term contracts. The same may be the case for wholesale prices where prices of goods are subject to longer-term contracting between firms. Taylor's (1980) overlapping-contracts model adopts such a view of nominal rigidities and assumes that each price is fixed for some number n periods and therefore each period $1/n$ of the prices in the economy is updated. This assumption requires the modeler to keep track of n different prices. A simpler approach is to assume that a fraction $1 - \theta \in [0, 1]$ of the prices in the economy is updated each period, but unlike Taylor's model, the prices that are updated each period are randomly selected. This assumption, which was introduced by [Calvo \(1983\)](#), leads to a much more tractable model because it is no longer necessary to keep track of the previously set prices as we will explain below.

Environment. A representative household has preferences for consumption and leisure as represented by

$$U_0 = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{L_t^{1+\psi}}{1+\psi} \right], \quad (18.1)$$

where C_t denotes consumption and L_t is labor supply and $\beta \in [0, 1]$. The household chooses how much to work, how much to save, and how much to consume. The vehicle for savings is a one-period nominal bond that pays interest i_t in period $t + 1$ and is in zero net supply.

The household consumes a final good, which is produced by a representative competitive firm out of a continuum of intermediate inputs indexed by $j \in [0, 1]$. The final goods producer operates a constant elasticity of substitution production function. Letting Y_t be the quantity of final goods produced, the production function is given as:

$$Y_t = \left(\int_0^1 y_{j,t}^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon}{\varepsilon-1}}, \quad (18.2)$$

where $y_{j,t}$ is the quantity of the j th intermediate input used and $\varepsilon > 1$ is the elasticity of substitution between varieties. The intermediate inputs are produced according to the production function

$$y_{j,t} = A_t \ell_{j,t}, \quad (18.3)$$

where $\ell_{j,t}$ is the amount of labor used in producing good j and A_t is an aggregate productivity that follows a first-order Markov process. The real marginal cost of producing one unit of intermediate goods is common across firms and equal to w_t/A_t , where w_t is the real wage denominated in units of final goods.

To model nominal rigidities we necessarily have to depart from perfect competition. After all, if all firms are price takers, no firm can be said to set its price. Moreover, it is assumed that firms stand ready to satisfy demand at their posted prices, which a competitive firm may

not be willing or able to do. We therefore adopt monopolistic competition as a convenient framework in which to model price setting. We have seen the production function (18.2) before in Chapter 6. Using the argument that we introduce there, the final goods producer's cost-minimization problem yields a price index

$$P_t = \left(\int_0^1 P_{j,t}^{1-\varepsilon} dj \right)^{1/(1-\varepsilon)},$$

where P_t is the cost of producing one unit of the final good and $P_{j,t}$ is the price of a unit of the j th intermediate good.⁴ As the final goods producer is competitive, it sells its output at marginal cost, and P_t is therefore also the price of a unit of final goods.

It also follows from the final goods producer's cost-minimization problem that the demand for intermediate good j is given by the following iso-elastic function of the good's relative price:

$$y_{j,t} = \left(\frac{P_{j,t}}{P_t} \right)^{-\varepsilon} Y_t. \quad (18.4)$$

Note that the price elasticity of the demand facing an intermediate goods producer is ε .

Both $P_{j,t}$ and P_t are nominal prices meaning they are set in terms of units of money. Money does not feature in this version of the New Keynesian model except as the unit of account for prices and bonds. Here we are making use of the cashless-limit argument introduced in Chapter 17, in which case households hold zero cash balances yet money can serve as the numeraire. However, extending the model with money demand derived from a money-in-the-utility-function assumption would produce exactly the same equilibrium as long as the marginal rate of substitution between consumption and leisure is independent of real cash balances.⁵

The key assumption that distinguishes New Keynesian models is that they feature nominal rigidities—that is, one or more prices in the economy is sticky. These nominal rigidities can apply to goods prices, to wages, or both; but here we will assume that the prices of intermediate inputs are sticky with the price-setting friction modeled as in Calvo (1983). Specifically, an intermediate goods producer has an i.i.d. probability $1 - \theta \in [0, 1]$ of being able to update its price each period. In the monopolistic competition model, the intermediate goods producers who are allowed to adjust prices in a given period then set prices taking as given the actions of other firms, input prices, and all aggregate variables. Firms serve whatever demand they face given the price they are quoting. Thus, in the short run, output is “demand determined” in the New Keynesian model.

The government is the final participant in this economy. The government sets the nominal interest rate and its choices are often represented through an interest rate rule that specifies the interest rate as an explicit function of macroeconomic variables such as the rate of inflation. We defer making detailed assumptions about the monetary policy rule.

⁴Formally, P_t is the derivative of the cost function for the final goods producers with respect to the quantity produced.

⁵Adding money demand would simply determine money supply.

Decision problems. The household's decision problem is to maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{L_t^{1+\psi}}{1+\psi} \right],$$

subject to the budget constraint

$$B_{t+1} + P_t C_t = P_t w_t L_t + (1 + i_{t-1}) B_t + P_t \int m_{jt} dj \quad \forall t,$$

where B_{t+1} is the nominal bonds held from date t to date $t + 1$ and m_{jt} is the date- t profit earned by firm j . In addition to the budget constraint, the household must also satisfy a no Ponzi game condition of the kind introduced in Chapter 4. It is convenient to rewrite the budget constraint in real terms by dividing both sides by P_t

$$b_{t+1} + C_t = w_t L_t + \frac{1 + i_{t-1}}{1 + \pi_t} b_t + \int m_{jt} dj,$$

where $b_t \equiv B_t/P_{t-1}$ is the real value of savings in period $t - 1$ and $1 + \pi_{t+1} \equiv P_{t+1}/P_t$ is the gross rate of inflation between periods t and $t + 1$.⁶

This decision problem results in two optimality conditions. The Euler equation

$$C_t^{-\sigma} = \beta \mathbb{E}_t \left[\frac{1 + i_t}{1 + \pi_{t+1}} C_{t+1}^{-\sigma} \right] \quad (18.5)$$

and the intratemporal labor supply condition

$$C_t^{-\sigma} w_t = L_t^\psi. \quad (18.6)$$

The Euler equation in (18.5) implies that the intertemporal path of consumption is determined by the (gross) real interest rate, $(1 + i_t)/(1 + \pi_{t+1})$. Hence, household's intertemporal choices are no different than in standard flexible price models encountered earlier in this book. Likewise, the labor supply condition in eq. (18.6) is also standard and equalizes households' marginal rate of substitution between consumption and work with the real wage. Nominal rigidities affect these conditions only through equilibrium changes in real wages and real interest rates.

An intermediate goods firm chooses how to price its good when it has the opportunity to reset its price. As usual, it is only relative prices that matter so we will define $p_{jt} \equiv P_{jt}/P_t$. Suppose a firm has a relative price p_{jt} at date t . If at date $t + 1$ the firm does not update its nominal price, its new relative price will be

$$p_{jt+1} = \frac{P_{jt}}{P_{t+1}} = \frac{P_{jt}}{P_t} \frac{P_t}{P_{t+1}} = \frac{p_{jt}}{1 + \pi_{t+1}}.$$

If the firm instead does update its price at date $t + 1$, it faces the same decision problem as any other firm that is updating its price that period because there are no firm-specific

⁶Note that in the standard notation used here, i_t applies to savings from t to $t + 1$ while π_t refers to the inflation between $t - 1$ and t . Both variables are determined at date t .

state variables other than the price they set in the past, which is no longer relevant to the firm updating its price. Let P_{t+1}^R be the nominal price set by firms who update their prices in $t + 1$, which is known as the **reset price**. Let $p_{t+1}^R = P_{t+1}^R/P_{t+1}$ be the corresponding relative price.

In period t , the real profits of firm j are

$$m_{jt} = p_{jt}y_{jt} - w_t\ell_{jt} = (p_{jt} - w_t/A_t)y_{jt} = (p_{jt} - w_t/A_t)p_{jt}^{-\varepsilon}Y_t,$$

where we have used eq. (18.3) to substitute for ℓ_{jt} and eq. (18.4) to substitute for y_{jt} .

The value function of a firm with relative price p is

$$V(p, \mathcal{S}) = u'(C(\mathcal{S})) \left(p - \frac{w(\mathcal{S})}{A(\mathcal{S})} \right) p^{-\varepsilon} Y(\mathcal{S}) + \beta \mathbb{E} \left[\theta V \left(\frac{p}{1 + \pi(\mathcal{S}')}, \mathcal{S}' \right) + (1 - \theta) V(p^R(\mathcal{S}'), \mathcal{S}') \right],$$

where \mathcal{S} is the aggregate state, which evolves independently of an individual firm's decisions because of the monopolistic competition assumption. On the right-hand side of this Bellman equation, the first term is the period payoff, which is the profits earned during the period valued at the marginal utility of consumption of the representative household (the owner of the firm). The second term is the continuation value, which has two components. The first component represents the value if the firm does not update its price next period while the second component is the continuation value if it does update.

A firm that updates its price maximizes $V(p, \mathcal{S})$ therefore finds the price that solves:

$$p^R(\mathcal{S}) = \arg \max_p V(p, \mathcal{S}).$$

As shown in the appendix, the first-order necessary condition for the choice of the optimal reset price can be expressed as:

$$0 = \mathbb{E}_t \sum_{\tau=t}^{\infty} (\beta\theta)^{\tau-t} u'(C_\tau) Y_\tau \left[p_{t,\tau}^{R-\varepsilon} - \varepsilon p_{t,\tau}^{R-\varepsilon-1} \left(p_{t,\tau}^R - \frac{w_\tau}{A_\tau} \right) \right],$$

where $p_{t,\tau}^R \equiv P_t^R/P_\tau = p_t^R / \prod_{s=t+1}^{\tau} (1 + \pi_s)$, is the relative price at date τ of a firm that last updated its price at date t . This condition is interesting. First, the term in square brackets times Y_τ is the marginal change in profit from a change in $p_{t,\tau}^R$. In a flexible price model, this term would equal zero period by period. But, because prices are sticky, firms aim at hitting this condition “on average” which means weighting profits in future periods by the probability $\theta^{\tau-t}$ that price will remain in effect that far in the future (and discounting future marginal profit contributions by the owners' marginal rate of intertemporal substitution). Secondly, firms are more forward looking when prices are stickier. This result is intuitive. In a flexible price setting, firms simply solve static optimization problems that result in equalizing marginal revenue and marginal costs period by period. With sticky prices, they now need to consider how their current choice of price may affect future profits. For example, firms will adjust prices today in response to information about marginal costs in the future.

Rearranging, we obtain

$$p_t^R \mathbb{E}_t \sum_{\tau=t}^{\infty} (\beta\theta)^{\tau-t} u'(C_\tau) Y_\tau \left(\frac{P_\tau}{P_t} \right)^\varepsilon = \frac{\varepsilon}{\varepsilon - 1} \mathbb{E}_t \sum_{\tau=t}^{\infty} (\beta\theta)^{\tau-t} u'(C_\tau) Y_\tau \left(\frac{P_\tau}{P_t} \right)^{1+\varepsilon} \frac{w_\tau}{A_\tau}.$$

We can then express the condition for the optimal reset price as:

$$p_t^R = \frac{p_t^N}{p_t^D} \quad (18.7)$$

where:

$$p_t^N = \frac{\varepsilon}{\varepsilon - 1} \frac{w_t}{A_t} u'(C_t) Y_t + \theta \beta \mathbb{E}_t (1 + \pi_{t+1})^{1+\varepsilon} p_{t+1}^N$$

and

$$p_t^D = u'(C_t) Y_t + \theta \beta \mathbb{E}_t (1 + \pi_{t+1})^\varepsilon p_{t+1}^D.$$

Market clearing. There are three markets at each date: the labor market, the goods market, and the bond market. Labor market clearing requires that the total labor supplied by households equals the total labor used in production

$$L_t = \int \ell_{j,t} dj.$$

As consumption is the only use for final goods, goods market clearing requires that production equals total consumption

$$Y_t = C_t. \quad (18.8)$$

Finally, bond market clearing requires that the net demand for bonds is zero.

Aggregation. In the Calvo model, firms are given the opportunity to re-optimize their prices with probability $1 - \theta$ and it is assumed that the process for the arrival rate of this opportunity is memoryless (ie. it follows a Poisson process). Hence, every period, firms are split randomly into a group of firms that cannot adjust their prices and another group that can. Let \mathcal{J}_t be the set of firms that update their prices in period t . It follows then that the price level can be expressed as:

$$P_t^{1-\varepsilon} = \int_j P_{j,t}^{1-\varepsilon} dj = \underbrace{\int_{j \notin \mathcal{J}_t} P_{j,t}^{1-\varepsilon} dj}_{\text{non-adjusters}} + \underbrace{\int_{j \in \mathcal{J}_t} P_{j,t}^{1-\varepsilon} dj}_{\text{adjusters}}$$

where each of these two terms are given as:

$$\int_{j \notin \mathcal{J}_t} P_{j,t}^{1-\varepsilon} dj = \theta P_{t-1}^{1-\varepsilon}$$

and

$$\int_{j \in \mathcal{J}_t} P_{j,t}^{1-\varepsilon} dj = (1 - \theta) (P_t^R)^{1-\varepsilon}.$$

Here we have used the assumption that firms are chosen randomly to adjust their prices to derive the expression for $\int_{j \notin \mathcal{J}_t} P_{j,t}^{1-\varepsilon} dj$. The expression for $\int_{j \in \mathcal{J}_t} P_{j,t}^{1-\varepsilon} dj$ exploits the fact that all firms that optimize their prices at the same date, choose the same price. It follows from this that:

$$P_t = \left[\theta P_{t-1}^{1-\varepsilon} + (1 - \theta) (P_t^R)^{1-\varepsilon} \right]^{1/(1-\varepsilon)}. \quad (18.9)$$

This expression makes it clear that the price level is sticky since it displays inertia. The higher is the probability that the firm cannot adjust prices in any given period, θ , the larger is the backward looking component of the price level. Yet, despite this, inflation is purely forward looking in this model. To see this, we can also exploit the above result to derive a relationship between inflation and the relative reset price:

$$p_t^R = \frac{\left[\frac{P_t^{1-\varepsilon} - \theta P_{t-1}^{1-\varepsilon}}{1-\theta} \right]^{1/(1-\varepsilon)}}{P_t} = \left[\frac{1 - \theta(1 + \pi_t)^{\varepsilon-1}}{1 - \theta} \right]^{1/(1-\varepsilon)}$$

which we can rearrange to give:

$$1 + \pi_t = \left[\frac{1 - (1 - \theta) (p_t^R)^{1-\varepsilon}}{\theta} \right]^{1/(\varepsilon-1)}. \quad (18.10)$$

Since firms are purely forward looking when setting p_t^R , inflation is purely forward looking, too. As a result, we do not need to keep any record of past prices (or inflation) when determining the current (or future) inflation rates.

Since only a fraction of firms can adjust prices, there can be dispersion of prices across firms. This price dispersion across intermediate input producers is a source of inefficiency as the final goods producers will over-utilize those inputs that have low relative prices and under-utilize those with high relative prices. To see this, integrate eq. (18.3) across j

$$\int y_{j,t} dj = A_t \int \ell_{j,t} dj = A_t L_t,$$

where the second equality uses the labor market clearing condition. Now substitute (18.4) for $y_{j,t}$ to arrive at

$$Y_t = \frac{A_t}{D_t} L_t, \quad (18.11)$$

where

$$D_t \equiv \int \left(\frac{P_{j,t}}{P_t} \right)^{-\varepsilon} dj \geq 1$$

is a measure of price dispersion. In particular, if $P_{j,t}$ is common across all firms, as would happen in this model if intermediate goods prices were flexible, then there is no price dispersion $P_{j,t} = P_t$ for all j and $D_t = 1$. In this case it follows that $Y_t = A_t L_t$. When there is price dispersion, however, D_t will be greater than one, which reduces the effective productivity of the economy. This loss of efficiency is due to the concave production function for final goods and the misallocation of labor across intermediate varieties. By the logic that the updated prices are randomly selected we have

$$D_t = \int_{j \in \mathcal{J}_t} p_t^{R-\varepsilon} dj + \int_{j \notin \mathcal{J}_t} \left(\frac{P_{j,t-1} P_{t-1}}{P_{t-1} P_t} \right)^{-\varepsilon} dj = (1 - \theta) p_t^{R-\varepsilon} + \theta(1 + \pi_t)^\varepsilon D_{t-1}. \quad (18.12)$$

State variables. There are two state variables of the aggregate economy: the current level of TFP, A_t , and the degree of price dispersion D_{t-1} . These states evolve according to the exogenous law of motion for A_t and eq. (18.12), respectively.

Equilibrium. An equilibrium of this economy involves stochastic processes for $C_t, L_t, Y_t, w_t, \pi_t, i_t, A_t, D_t$ that satisfy (18.5), (18.6), (18.8), (18.10), (18.11), (18.12), the monetary policy rule, and the exogenous law of motion for A_t ; in addition, an equilibrium requires a policy rule p^R and a value function V that solve the price-setting problem.

Flexible-price equilibrium and output gap. In order to complete the specification of the environment we need a monetary policy rule. An important benchmark for policy in this context is the counterfactual outcome with fully flexible prices. We introduce that now as common monetary policy rules make reference to this benchmark.

The flexible-price economy is a special case where $\theta = 0$ so all prices are updated every period. When an intermediate goods firm sets its price in period t , it knows it will for certain be able to adjust its price again in $t + 1$ and the only consideration is maximizing profits in period t . This static profit maximization problem results in setting a constant markup $\mu \equiv \varepsilon/(\varepsilon - 1)$ over marginal cost as we saw in Chapter 6.⁷ So $p_t^R = \mu w_t/A_t$ and the markup is positive as long as $\varepsilon > 1$ which we have imposed earlier. As all firms face the same profit maximization problem when prices are flexible, they all choose to set the same price and the definition of the price index reduces to $P_t^n = P_t^{R,n}$ or $p_t^{R,n} = 1$ where we use the superscript “ n ” to indicate variables in the flexible price, or “natural,” equilibrium. It then follows that $w_t^n/A_t = \mu^{-1}$. Moreover, as all firms set the same price, there is no efficiency loss from price dispersion. We can then solve the following system of equations to obtain the equilibrium:

$$\begin{aligned} w_t^n &= A_t/\mu \\ Y_t^n &= C_t^n \\ Y_t^n &= A_t L_t^n \\ C_t^{n-\sigma} w_t^n &= (L_t^n)^\psi, \end{aligned}$$

where the second, third, and fourth equations are the aggregate resource constraint, the aggregate production function, and the intratemporal labor supply condition. The solution to this system is:

$$Y_t^n = \mu^{-1/(\sigma+\psi)} A_t^{(1+\psi)/(\sigma+\psi)}. \quad (18.13)$$

Let us also define the output gap as

$$x_t \equiv \log Y_t - \log Y_t^n \quad (18.14)$$

i.e. the log of level of output in the sticky-price equilibrium relative to the natural level.

⁷You can verify this by solving $\max_p \{(p - w_t/A_t) p^{-\varepsilon} Y_t\}$.

Interest rate rules. An interest rate rule describes how monetary policy is set. These can take several forms, but the simplest version is to impose a relationship between nominal interest rates and other variables in the model. For example,

$$i_t = \bar{i} + \phi_\pi(\pi_t - \pi^*) + \phi_x x_t,$$

where π^* is the inflation target and $\bar{i} = \beta^{-1}(1 + \pi^*) - 1$ is the nominal interest rate that would satisfy the Euler equation (18.5) in a steady state with constant inflation π^* . The coefficients ϕ_π and ϕ_x determine how strongly nominal interest rates react to deviations of inflation from its target and to the output gap. Interest rate rules are often called **Taylor rules** in light of Taylor’s (1993) finding that such a rule with $\phi_\pi = 1.5$ and $\phi_x = 0.5$ provided a good description of the monetary policy choices of the Federal Reserve between 1987 and 1992.⁸

Monetary policy shocks. Here we are building the baseline New Keynesian model. More elaborate versions of this model are often used in central banks to analyze potential monetary policy strategies. Researchers and policymakers are therefore often interested in questions of the form “what would happen if we raise interest rates taking the economic conditions as given?” To answer this question, we want to calculate the consequences of a shift in interest rates that is not the result of a shift in the economic fundamentals that appear in the interest rate rule. To engineer such a shift, we can add an exogenous shock to the interest rate rule.

The three-equation model. The model we have presented above is often called the **three-equation model** because a first-order approximation of the model around a zero-inflation steady state is summarized by the following three equations. First, we have a log-linearized version of the consumption Euler equation

$$x_t = \mathbb{E}_t x_{t+1} - \frac{1}{\sigma}(\hat{i}_t - \mathbb{E}_t \pi_{t+1} - r_t^n), \quad (18.15)$$

where x_t is the output gap and

$$r_t^n = -\log \beta + \frac{\sigma(1 + \psi)}{\sigma + \psi}(\hat{A}_t - \mathbb{E}_t \hat{A}_{t+1})$$

is the real natural rate of interest.⁹ This equation represents the demand-side of the economy and is sometimes called the **IS curve** in a reference to older IS-LM models. Notice that

⁸The coefficients in Taylor’s specification correspond to annual rates of inflation and annualized interest rates. Interest rates and inflation rates are measured in percentage points per unit of time while the output gap is simply measured in percentage points. Therefore, if you change the length of a unit of time, you need to adjust the coefficients accordingly. For example, in a quarterly model, the equivalent rule would be $\phi_\pi = 1.5$ and $\phi_x = 0.125$.

⁹To derive this equation, define $\hat{i}_t = \log(1 + i_t)$. Then take logs of both sides (18.5) noting that in a first-order perturbation solution we assume the shocks are arbitrarily small so we can take logs inside the expectation operator. Finally, we need the definition of the output gap from eqs. (18.13) and (18.14), which when linearized give $x_t = \hat{Y}_t - \frac{1+\psi}{\sigma+\psi}\hat{A}_t$. Using this to substitute for \hat{Y}_t yields the equation above. Note that r_t^n is equal to the real interest rate that solves the linearized Euler equation when consumption takes its natural level.

output at date t is forward-looking (it depends on $\mathbb{E}_t \hat{Y}_{t+1}$) and is decreasing in the real interest rate where $1/\sigma$ is the elasticity of intertemporal substitution.

Next, as we show in the appendix, we use the first-order condition of the firm's price-setting problem and other equilibrium conditions of the model to arrive at a New Keynesian Phillips curve

$$\pi_t = \kappa x_t + \beta \mathbb{E}_t [\pi_{t+1}], \quad (18.16)$$

where the composite parameter $\kappa \equiv (1 - \theta)(1 - \beta\theta)(\sigma + \psi)/\theta$ is usually referred to as the slope of the New Keynesian Phillips curve. Here we see that inflation is forward-looking and increasing in the current output gap. The slope of the Phillips curve, κ , is larger if prices are more flexible (lower θ) or if marginal cost is more sensitive to the level of production.

Lastly, we have an interest rate rule. For our simulations here we will assume

$$\hat{i}_t = \phi_\pi(\pi_t - \pi^*) + \phi_x x_t + \omega_t, \quad (18.17)$$

where ω_t is an AR(1) monetary policy shock. In addition to these three core equations, we also have exogenous processes for TFP and the monetary policy shock. The three-equation representation of the model, (18.15)-(18.17), is often used as a small-scale model of the economy to explore qualitative features of business cycles and monetary policy.

Determinacy. A key consideration for any monetary policy rule is whether it yields a unique equilibrium. Suppose that, for some reason, inflation expectations are elevated despite expectations of a zero output gap. Could that lead to high inflation today? It all depends on the behavior of the nominal interest rate. If the nominal interest rate is unresponsive, high inflation expectations translate to low real interest rates, which through the IS curve lead to a positive output gap. The positive output gap then puts upward pressure on current inflation through the Phillips curve. By this logic, some small expectation of higher inflation in the far future could justify expectations of high inflation (and positive output gaps) all the way back to the present. In the model, nothing pins down the expectations of inflation in the infinitely far future so we indeed have multiple equilibria when nominal interest rates are unresponsive. In order to prevent this outcome, the interest rate rule must respond sufficiently strongly so that the real interest rate *rises* when inflation rises thereby stabilizing the economy. The three-equation model has a unique equilibrium if and only if

$$(1 - \beta)\phi_x + \kappa(\phi_\pi - 1) > 0 \quad (18.18)$$

(see the appendix for a derivation). Note that $\phi_\pi > 1$ is sufficient for this condition to hold and necessary if $\phi_x = 0$. This condition is known as the **Taylor principle**.

The role of nominal demand in determining output. As we mentioned in the introduction to this chapter, when prices adjust imperfectly, changes in nominal variables can have real effects. Traditionally, this point is often discussed in terms of the money supply—if money is injected, a frictionless model would imply an immediate adjustment of the price level with no effect on any real variables. This outcome is often called the classical dichotomy between real and nominal sides of the economy. With nominal rigidities, this logic breaks

down as the increase in nominal demand from the increase in the money supply is not immediately undone by an increase in the price level. Here we will make a similar argument in terms of nominal interest rates. Suppose the government announces a lower path for nominal interest rates going forward (e.g. an expansionary monetary policy shock). As we saw in eq. (18.5), the path of aggregate consumption (and therefore aggregate output) is fully determined by the path of real interest rates. Therefore, if the classical dichotomy were to hold, it would have to be the case that path of expected inflation would immediately fall so as to leave real interest rates unchanged. But if we solve eq. (18.16) forward, we see inflation at each date reflects expectations of current and future output gaps, which would be zero under the classical dichotomy. In fact, the model predicts that with lower real interest rates there will be a positive output gap and an increase in inflation that can reinforce the decline in nominal interest rates absent a countervailing response of the monetary policy rule.

18.4 Monetary policy strategies

The New Keynesian model has an active role for government. After all, a key equation in the model is the monetary policy rule. We now describe what the role for policy is in the New Keynesian model and how this is reflected in some real-world monetary policy strategies.

18.4.1 Policy objectives

The equilibrium of the New Keynesian model can be inefficient for two reasons (i) the total level of production can deviate from the efficient level and (ii) labor can be misallocated across intermediate inputs. To see this in more detail, consider the planner's problem

$$\max_{C_t, Y_t, L_t} \frac{C_t^{1-\sigma}}{1-\sigma} - \frac{L_t^{1+\psi}}{1+\psi}$$

subject to

$$C_t = Y_t = A_t L_t.$$

This is a static problem because the planner only has to choose how to allocate labor each period and in the aggregate there is no way to move resources across time. The constraint on the planner uses the fact that when L_t units of labor are used to produce each intermediate good, the quantity of final goods that can be produced is $A_t L_t$. The first-order condition of this problem leads to

$$Y_t^* = A_t^{(1+\psi)/(\sigma+\psi)}, \tag{18.19}$$

where a star denotes the first-best allocation.

To focus on the possibility of an inefficient level of aggregate production, suppose there is no price dispersion so $Y_t = A_t L_t$. Then using (18.19), (18.6), and $C_t = Y_t$ we obtain

$$Y_t = \left(\frac{w_t}{A_t} \right)^{1/(\sigma+\psi)} Y_t^*.$$

In an efficient economy, the real wage would be equal to the marginal product of labor. The equation above shows that if the wage deviates from this efficient level, output will deviate from the efficient level. When the wage is too low, households will supply too little labor and too little will be produced. Comparing this to the natural level of output (18.13), we see that in the flexible price economy, output is too low because $w_t/A_t = 1/\mu < 1$. So monopoly power is one source of inefficiency. Notice that w_t/A_t is real marginal cost and if firms set prices equal to nominal marginal cost we would have $P_t = (P_t w_t)/A$ or $w_t/A_t = 1$.

Nominal rigidities affect the markups that firms charge and therefore the degree of inefficiency in the economy. To see this, suppose there is a decrease in nominal marginal cost. As some firms do not update their prices, their markups rise. In the aggregate, the increase in average markups is reflected in a lower w_t/A_t ratio and the economy is further from the efficient level of production. Over time, firms will update their prices and return their markups to the desired level μ and w_t/A_t will return to $1/\mu$.

Now consider the inefficiency due to price dispersion. Using the aggregate production function $Y_t = A_t L_t / D_t$, (18.19), (18.6), and $C_t = Y_t$ we obtain

$$Y_t = D_t^{-\psi/(\sigma+\psi)} \left(\frac{w_t}{A_t} \right)^{1/(\sigma+\psi)} Y_t^*.$$

This equation shows that there is an additional inefficiency, when $D_t \neq 1$. From the concavity of the final goods production function, the efficient use of labor is to produce equal amounts of all the intermediate inputs. The economy will deviate from this outcome if some intermediate goods producers have lower prices than others (see eq. 18.11). The reason one producer may have a lower price than another is, for example, that one set its price more recently than the other and, in the intervening time, market conditions changed leading to a change in the optimal reset price. As the reset price fluctuates, so too will the inflation rate (see eq. 18.10). One goal for monetary policy then is to stabilize the inflation rate because this minimizes the efficiency loss from relative price dispersion.¹⁰

Define the *welfare-relevant output gap* as $x_t^w \equiv \log Y_t - \log Y_t^*$. Using eqs. (18.13) and (18.19) we see that

$$x_t^w = \log Y_t - \log Y_t^n + \log Y_t^n - \log Y_t^* = x_t - \frac{1}{\sigma + \psi} \log \mu.$$

The welfare-relevant output gap differs from the output gap because the natural level of output is distorted by monopoly power.

In the literature that followed Phillips' original contribution, Phillips (1958), it was perceived that there was a stable relationship between (wage) inflation and resource utilization (as measured by unemployment) which therefore presented a trade-off to policymakers. This notion was challenged by Friedman (1968) and Phelps (1967) who posited that the long-run Phillips curve is vertical (at the natural rate of unemployment) as real wages, which should be independent of inflation in the long-run, determine employment. If monetary policy attempts to persistently raise the level of output above the natural level, the results will be

¹⁰The Calvo model of nominal rigidities can generate large efficiency losses from price dispersion because a firm can be stuck with a very old price that is far out of alignment with the current price level. Models of menu costs in which firms can choose to change their prices subject to a cost tend to generate less price dispersion because firms with very misaligned prices will choose to change their prices.

high inflation and no actual increase in output. For these reasons, a common view is that monetary policy should focus on stabilizing the economy around the flexible-price level of activity even though this may not be the first best. One way of formalizing this view is to give the policymakers another policy tool that can address long-run inefficiencies while leaving monetary policy responsible for responding to aggregate shocks. In the New Keynesian model, a simple extension of the model is to suppose there is lump-sum tax on households that finances a constant production subsidy for intermediate goods producers. In particular, suppose that intermediate goods producers are given a subsidy $\tau^l = 1/\epsilon \in (0, 1)$ so that their effective cost of labor is $(1 - \tau^l)w_t/A_t$. The production subsidy induces them to produce more and, in the steady-state, the subsidy undoes the distortion from monopoly power.

In summary, in the rest of this chapter we will take the goals for policy to be to (i) bring the aggregate level of production to the natural level and to (ii) minimize the efficiency loss from price dispersion by stabilizing inflation. Locally around a zero-inflation steady state, the welfare of the representative household in the basic New Keynesian model can be expressed as¹¹

$$U \approx -\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\pi_t^2 + \frac{\kappa}{\epsilon} x_t^2 \right]. \quad (18.20)$$

Welfare decreases with inflation or deflation because it results in price dispersion. Welfare also decreases with a positive or negative output gap because the level of production differs from the optimal level.

18.4.2 The divine coincidence

In the baseline New Keynesian model, it is possible to have zero inflation and zero output gaps at all times. The New Keynesian Phillips curve, (18.16), is the key to this argument. From that equation we can see that if there is no output gap at any date, then there will be no inflation today or in any future period. This aspect of the model—that there is no trade off between output gap and inflation stabilization—is known as the **divine coincidence**.

The standard interpretation of the divine coincidence is that the model is somewhat too simple and abstracts from the features of the economy that lead to a meaningful choice between inflation stabilization and output stabilization. What might those features be? To break the divine coincidence we need a *time-varying* wedge between the flexible-price level of output and the efficient level of output. Productivity shocks themselves do not create such a wedge because they affect both the flexible-price level of output as well as the efficient level of output in equal proportions. We say the wedge needs to be time-varying because a constant distortion requires a permanent change in the level of activity and is therefore not an issue for which monetary policy is well suited.

An extension of the model, which is often discussed due to its simplicity, involves shocks to the elasticity of substitution between varieties of intermediate goods. The resulting time-varying market power affects Y_t^n but does not change Y_t^* leading to a time-varying gap between the socially efficient level of production and the flexible-price level. In the linearized

¹¹Here we use a second order approximation around a steady state in which the monopoly distortion has been corrected through the labor subsidy $\tau^l = 1/\epsilon$. See [Woodford \(2003b\)](#).

version of the model, these shocks appear as a “cost-push” shock—an exogenous term that is appended to the Phillips curve

$$\pi_t = \kappa x_t + \beta \mathbb{E}_t [\pi_{t+1}] + \eta_t \quad (18.21)$$

(see [Steinsson, 2003](#), for a derivation). With this added term, it is no longer possible to stabilize inflation and output perfectly because a policy of setting $x_t = 0$ at all dates no longer leads to $\pi_t = 0$. The divine coincidence also breaks if the flexible-price economy does not respond to shocks in the efficient way; for example because there are frictions in the determination of wages (see [Blanchard and Galí, 2007](#)).

18.4.3 Inflation targets and price level targets

Around the world, many central banks follow a version of a monetary policy strategy called **flexible inflation targeting**. This strategy can be summarized as minimizing deviations of inflation from a target level while also minimizing deviations of output from its natural level. These goals can be formalized in the objective function

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t [(\pi_t - \pi^*)^2 + \lambda x_t^2], \quad (18.22)$$

where π^* is the inflation target and λ is a parameter that determines the relative weight placed on the two goals. The policymaker then seeks to minimize this loss function, which closely resembles eq. (18.20).

When inflation is unpredictable, it is risky to agree to a long-term nominal contract because the real values stipulated by the contract are unpredictable. This rationale would argue that it is more valuable for monetary policy to minimize the uncertainty over the price level than to target an inflation rate. Under price-level targeting, the central bank seeks to return the price level to a specific path. For example, consider the interest rate rule

$$i_t = \bar{i} + \phi_P (P_t - P_t^*) + \phi_x x_t$$

where P_t^* is the price-level target, which could be the price level in some base year scaled by a constant annual inflation rate. This rule dictates that the central bank raises interest rates whenever the price level is above target, which would, all else equal, put downward pressure on inflation and move the economy back toward the target. In addition to providing more certainty about the price level, price-level targeting is appealing because it makes clear that future policy will undo unwanted changes in the price level—something that is also useful under inflation targeting as we will see next.

18.4.4 Expectations, commitment, and time consistency

In the New Keynesian model, the private sector is forward looking: current inflation depends on expectations of future inflation and current demand depends on expectations of future real interest rates. Therefore, expectations of what monetary policymakers will do in the future affect the macroeconomic outcomes today. If the private sector expects interest rates to be high in the future *ceteris paribus*, output and inflation will be lower today.

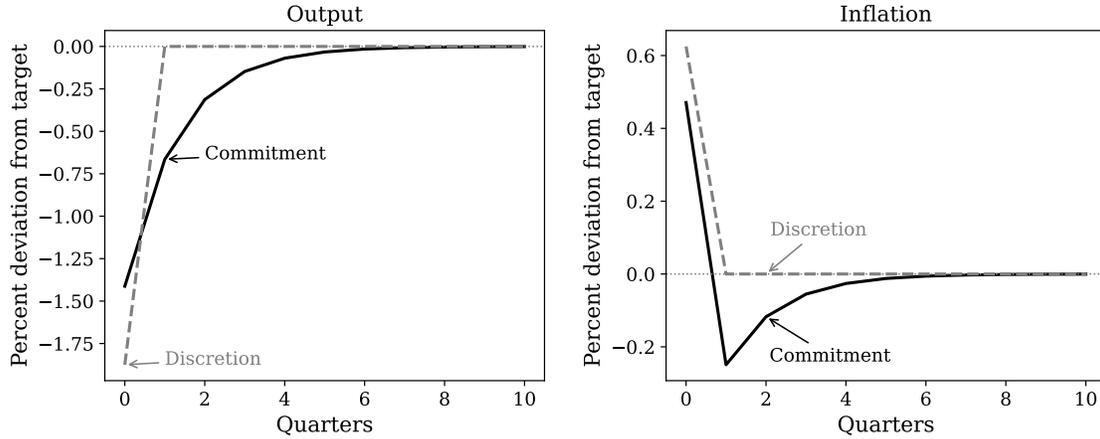


Figure 18.1: Response of output and inflation to a transitory cost push shock under commitment and discretionary policies.

Notes: These paths are simulated using the parameters $\beta = 0.99$, $\kappa = 0.2$, $\varepsilon = 3$.

Suppose a central bank pursues the inflation targeting objective (18.22) and the economy is hit by a cost-push shock at date 0. That is, we consider the Phillips curve (18.21) with $\eta_t > 0$ for just one period and zero thereafter. Now consider two possible monetary policy responses. In the first, the central bank raises interest rates at date 0 and then returns them to their long-run level. We will call this the “discretionary” policy for reasons that will become clear. In the second case, the central bank raises interest rates by less, but only gradually normalizes them. We will call this the “commitment” policy. Figure 18.1 plots the two cases. Under the discretionary policy, inflation rises at date 0 and then returns to zero. The linearized version of the model with the discretionary policy has no endogenous state variables so once the cost-push shock dissipates, the economy immediately returns to steady state.¹² The Phillips curve at date 0 is

$$\pi_0 = \kappa x_0 + \eta_0,$$

where we have used $\pi_1 = 0$ as the economy returns to steady state in the next period. In this case, the policymaker can only use x_0 to lean against the cost-push shock. As we see in the figure, output is reduced so as to dampen the inflationary effects of the shock.

Under the commitment policy, inflation is positive at date 0 and then negative for several future periods. The Phillips curve at date 0 is

$$\pi_0 = \kappa x_0 + \beta \pi_1 + \eta_0,$$

where now π_1 appears. Under this policy, the central bank reduces π_1 in order to reduce π_0 with less impact on x_0 . As we see in the figure, there is less inflation at date 0 but also a smaller decline in output at date 0. In order to reduce π_1 , the central bank keeps interest rates persistently higher than normal so that output and inflation are persistently lower.

¹²In the non-linear model, the degree of price dispersion is an endogenous state variable, but a first-order approximation of (18.12) around a zero-inflation steady state yields $D_t \approx 1$.

While the outcomes from the commitment policy are better than those from the discretionary policy at date 0, they are worse at future dates. Under the discretionary policy, there are no deviations in output or inflation at any date $t \geq 1$ while, under the commitment policy, output and inflation are too low relative to the targets. This brings us to a time-consistency problem. At date 0, the central bank would like to announce the commitment policy, but it would like to switch to the discretionary policy at date 1. If the private sector anticipates this switch in policy, then the benefits of the commitment policy at date 0 are unattainable because the central bank cannot convince the private sector that π_1 will be negative. The central bank can only achieve the better outcomes if it is able to commit at date 0 to have tight monetary policy at future dates even though it will not want to do that when those future dates arrive.

Finally, under the discretionary policy, the price level jumps at date 0 and then remains constant thereafter. Under the commitment policy, the price level rises, but then falls subsequently as inflation is negative for several periods. In fact, if you accumulate the inflation rates in the figure, you find that the long-run price level is unaffected. This result demonstrates that long-run price stability is useful for a central bank even if it is pursuing an inflation targeting framework because it serves to stabilize inflation expectations and therefore helps stabilize inflation rates.

18.5 Aggregate evidence of nominal rigidity

We now review the aggregate evidence on price rigidity and the response of real variables to nominal shocks in the form of changes in the nominal interest rate targeted by monetary policy. As we will see, the baseline model is qualitatively consistent with this evidence—in contrast to a flexible-price model—but also has shortcomings.

18.5.1 The macroeconomic effects of monetary policy shocks

One of the most studied issues in macroeconomics is the dynamic effects of monetary policy shocks, which lead to exogenous changes in nominal interest rates. This topic attracts so much interest because it speaks to the relevance of the nominal rigidities that underlie the Keynesian perspective. In most models with flexible prices, nominal variables including nominal interest rates have no bearing on real outcomes (e.g. see eq. (18.13)). Thus, the impact of changes in nominal interest rates on real variables provides information about the importance of nominal rigidities. Moreover, monetary policy plays a central role in modern macroeconomic policy and understanding the consequences of a change in interest rates is a crucial ingredient to real-world policy decisions.

The main challenge with assessing the effects of a change in interest rates is that monetary policy changes *in response* to developments in the economy. If we simply look at the correlations between variables, we will tend to find that higher nominal interest rates are associated with higher levels of inflation, but this may simply reflect endogeneity of monetary policy decisions (because central banks tend to increase interest rates when inflation rises). We are instead interested in the *causal* effect of interest rates on the economy. That is, we ask what would happen if interest rates increased for reasons unrelated to state of the

economy? To answer this question, we need to empirically identify monetary policy shocks as opposed to systematic movements in interest rates in response to economic conditions.

Researchers typically identify monetary policy shocks using information that allows them to estimate the systematic or predictable component of interest rates. Removing this systematic component from the actual change in interest rates yields a residual that is interpreted as a monetary policy shock. One strand of literature, which was pioneered by [Kuttner \(2001\)](#) and [Gürkaynak, Sack, and Swanson \(2005\)](#), is premised on (a) the fact that monetary policy decisions are announced at particular times known to researchers and (b) in a narrow time window around the announcement, changes in interest rates will be dominated by monetary policy as opposed to other economic news. Using financial data, we can make a forecast of the monetary policy decision just minutes before it is announced. This forecast reflects market expectations of how monetary policy will be conducted given current economic conditions—i.e. it is the market’s view of the systematic policy response. If the announced decision differs from the forecast, it is due to a deviation of monetary policy from its usual practice (as judged by financial markets).

Another strand of literature, starting with [Romer and Romer \(2004\)](#), uses central bank forecasts of inflation and other variables to estimate the endogenous component of policy. The motivation for this approach is that central banks often set policy based on their assessment of the economic outlook as reflected in economic forecasts. For example, if the forecast for inflation is elevated or unemployment is expected to be low, policymakers will raise interest rates. By regressing interest rate changes on the central bank forecasts, this approach estimates the systematic component of policy and the residuals from this regression can be interpreted as movements in interest rates that are not due to changes in the economic outlook.

Figure 18.2 plots impulse responses for output, inflation, and nominal interest rates following a monetary policy shock identified using the [Romer and Romer](#) method.¹³ The left panel of the figure shows that contractionary monetary policy shocks lead to persistently higher nominal interest rates. The center panel of the figure shows that aggregate output declines. The right panel shows that inflation declines so real interest rates rise more than nominal interest rates.¹⁴ So we find that a nominal shock affects real variables.

The basic New Keynesian model presented above is qualitatively consistent with the patterns revealed by the empirical estimates: higher nominal interest rates generate increases in real interest rates and reductions in aggregate demand and therefore a decline in output. With lower output, there is lower resource utilization leading to lower marginal costs and a decline in inflation. The timing of the responses, however, is quite different. Figure 18.2 shows model-simulated responses of interest rates, output, and inflation to a monetary policy shock (for now, just focus on the basic model and we will return to the sticky-wage model below). We chose the magnitude of the simulated shock and the monetary policy rule to roughly match the path of nominal interest rates that we estimated. In addition, we chose

¹³We use the implementation of this method by [Wieland and Yang \(2016\)](#) and regress the outcome of interest on the estimated monetary policy shocks and the lags of macroeconomic variables.

¹⁴The increase in inflation at very short horizons is known as the “price puzzle” and is a fairly common feature of empirical estimates of the effects of monetary policy shocks. It likely reflects reverse causation from inflation to interest rates that is not removed by the identification strategy. In some specifications, adding further control variables eliminates the price puzzle (see [Sims, 1992](#)).

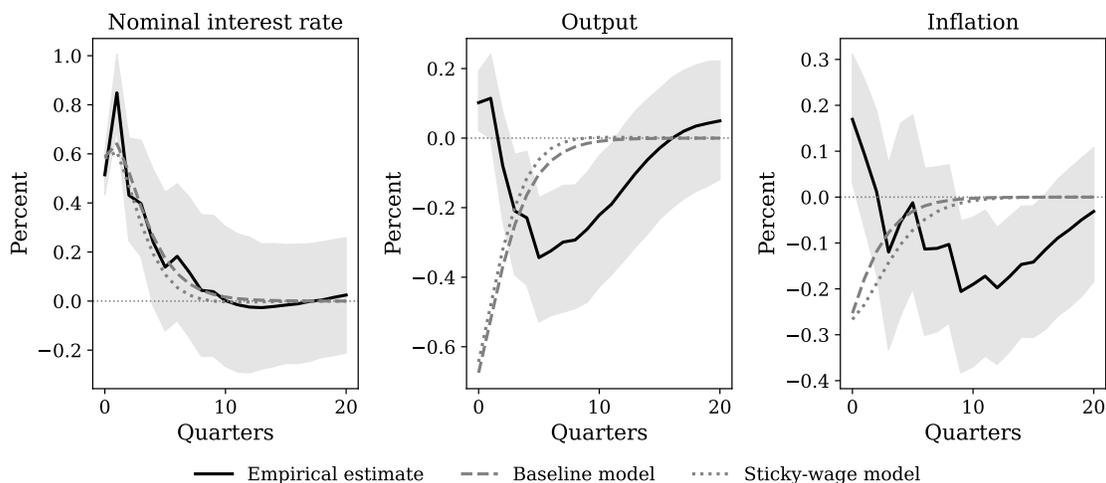


Figure 18.2: Empirical and simulated responses to a monetary policy shock.

Notes: The shaded areas around the empirical estimates are 90% confidence bands.

the other parameters of the model to roughly match the magnitudes of the responses of output and inflation.¹⁵ As shown in the figure, the model-generated responses of output and inflation immediately fall in response to the monetary shock and then gradually return to steady state whereas the empirical responses feature delayed responses of output and inflation.

18.5.2 Price rigidity in the aggregate

At the start of this chapter we discussed evidence on the frequency of price changes in micro data. But does this micro-level rigidity translate to rigidity in the aggregate?

One way to answer this question is to rely on a general equilibrium model as discussed above and ask which values of structural parameters allow one best to “match” the data? One might, for example, study the estimated impulse response functions in Figure 18.2 and estimate the slope of the Phillips curve by comparing the magnitude of the inflation response to the magnitude of the output response. One implementation of this idea was carried out by [Christiano, Eichenbaum, and Evans \(2005\)](#). These authors construct a rich New Keynesian model and fit its structural parameters to match the impulse response functions following a monetary policy shock. Their model includes many extensions of the basic New Keynesian framework some of which we describe in the next section. Overall, they find that the data indicate a value of θ that is consistent with an average contract length of 2.5 quarters which is in the range of values suggested by the literature on the frequency of price changes discussed earlier. Yet, at the same time, they also find that rigidities in wage setting are crucial to explain the dynamics of inflation following a monetary policy shock.

¹⁵Here we use eqs. (18.15) and (18.16) and a policy rule $\hat{i}_t = \gamma \hat{i}_{t-1} + (1 - \gamma)(\phi_\pi \pi_t + \phi_x x_t) + \omega_t$ and an AR(1) specification for ω_t to mimic the estimated path of interest rates. We use parameter values $\gamma = 0.65$, $\phi_\pi = 1.5$, $\phi_x = 0.125$, $\sigma = 5$, $\psi = 1$, $\theta = 1 - \frac{1}{8}$, $\beta = 0.99$, and an autocorrelation of 0.5 for ω_t .

An alternative approach is to estimate the New Keynesian Phillips curve directly in isolation from other parts of the model. Here, the question is how strongly changes in the output gap (or marginal costs of production) translate to changes in prices. Consider the Phillips curve

$$\pi_t = \kappa x_t + \beta \mathbb{E}_t [\pi_{t+1}] + \eta_t$$

where we have included the shock η_t as we would in general not expect the Phillips curve to hold exactly as an empirical relationship. Our interest is in estimating $\kappa = (1 - \theta)(1 - \theta\beta)(\sigma + \psi)/\theta$ as this speaks to the strength of nominal rigidities.

The direct estimation of the Phillips curve requires one to address several challenges. Any measure of the output gap is bound to be associated with measurement error which induces an attenuation bias in our estimate of κ if we were simply to apply ordinary least squares. However, if we have an instrumental variable that is correlated with the true output gap and uncorrelated with the measurement error we can sidestep this source of bias. Inflation expectations are not directly observable either. To measure inflation expectations, the typical practice is to use statistical techniques to construct a forecast of π_{t+1} on the basis of information that is available at date t . Finally, the presence of the cost-push shock complicates matters. Suppose there is an inflationary cost push shock, $\eta_t > 0$, and that the central bank responds to this by restraining demand and inducing a negative output gap. In this scenario, there is reverse causation from inflation to the output gap.

The literature has adopted a variety of methods to overcome these challenges and it remains an active area of research. Galí and Gertler (1999) is representative of the approach that has been followed by many papers in the literature. This approach rewrites the Phillips curve as

$$\pi_t = \kappa x_t + \beta \pi_{t+1} + \underbrace{\eta_t - \beta (\pi_{t+1} - \mathbb{E}_t [\pi_{t+1}])}_{\equiv \zeta_t},$$

where we have replaced expected inflation with the realization of π_{t+1} and included the expectational error $\pi_{t+1} - \mathbb{E}_t[\pi_{t+1}]$ in the error term, which we now denote ζ_t . If we assume $\mathbb{E}_{t-1}[\eta_t] = 0$, any lagged variable known at $t - 1$ that predicts x_t or π_{t+1} can be a valid instrument.¹⁶ This (strong) assumption allows us to use realized future inflation in place of expected inflation thereby obviating the need for measurement of inflation expectations and also prevents reverse causation as the lagged instruments are not correlated with η_t .

To measure the output gap, Galí and Gertler make use of the micro-foundations of the New Keynesian model, which say it is real marginal cost that is relevant in price setting. They start by positing a Cobb-Douglas aggregate production function:

$$Y_t = A_t K_t^{\alpha_K} L_t^{\alpha_L}$$

¹⁶Let z_{t-1} be the instrument. The orthogonality condition requires that this instrument be uncorrelated with ζ_t , which we can verify as follows

$$\text{cov}(z_{t-1}, \zeta_t) = \mathbb{E}[z_{t-1}\zeta_t] = \mathbb{E}[z_{t-1}\{\eta_t - \beta(\pi_{t+1} - \mathbb{E}_t[\pi_{t+1}])\}] = \mathbb{E}[z_{t-1}\eta_t] + \beta\mathbb{E}[z_{t-1}(\pi_{t+1} - \mathbb{E}_t[\pi_{t+1}])] = 0,$$

where $\mathbb{E}[z_{t-1}\eta_t] = \mathbb{E}[z_{t-1}\mathbb{E}_{t-1}[\eta_t]]$ by the law of iterated expectations and $\mathbb{E}_{t-1}[\eta_t] = 0$ by assumption. Similarly $\mathbb{E}[z_{t-1}(\pi_{t+1} - \mathbb{E}_t[\pi_{t+1}])] = \mathbb{E}[z_{t-1}(E_t[\pi_{t+1}] - \mathbb{E}_t[\pi_{t+1}])] = 0$ by the law of iterated expectations.

where K_t is the input of capital and α_K and α_L are the output elasticities with respect to the two factor inputs. Assuming that the labor input is flexible while the capital stock is predetermined in period t , cost minimization (together with firms being price takers in the input markets) implies that real marginal costs are given as:

$$mc_t = \frac{w_t}{\partial Y_t / \partial L_t} = \frac{s_t^L}{\alpha_L},$$

where $s_t^L = (w_t L_t) / Y_t$ is the labor share of income and w_t is the real wage. Thus, up to a first-order approximation, the log of real marginal costs are given as the log of the labor share.

Galí and Gertler estimate the Phillips curve using the generalized method of moments applied to quarterly U.S. data from 1960 to 1997. Their estimates imply an estimate of the frequency of price adjustment, $1 - \theta$, in the range of 0.085-0.171 per quarter, which indicates much longer average price contract length than the estimates from disaggregated prices discussed earlier.

The New Keynesian Phillips curve implies inflation is purely forward-looking. Before the development of the New Keynesian model, typical specifications of the Phillips curve instead included lagged inflation rather than inflation expectations. This so-called “accelerationist” Phillips curve implies a high degree of inflation persistence. Forward- and backward-looking Phillips curves have very different implications for how inflation can be controlled. In the forward-looking case, a credible central bank can immediately reduce inflation by committing to a low long-run inflation rate and zero output gaps. In the latter case, inflation can only be reduced by imposing negative output gaps on the economy. To explore the issue of inflation persistence, Galí and Gertler also considered a “hybrid” New Keynesian Phillips curve in which they assume that a certain fraction of price setters are purely backward looking and simply set prices by updating the past average reset price (of forward-looking firms) with the past inflation rate. In this extension, they estimate the fraction of firms that are backward looking as opposed to forward looking. Their estimates place much more weight on forward-looking behavior than on backward-looking behavior.

The work of Galí and Gertler (1999) has been influential because it establishes a link between inflation and marginal costs as consistent with the price setting condition embedded in the New Keynesian model. However, there remains considerable uncertainty surrounding the Phillips curve parameters as changes in the data series, the sample period, or the econometric specification can result in considerable changes in the parameter estimates, see (see Mavroeidis, Plagborg-Møller, and Stock, 2014). One issue of note is that the correlation between the labor share and inflation appears to have weakened over time. In Figure 18.3 we show the dynamic correlations between the labor share and leads and lags of the inflation rate for U.S. quarterly data, 1960:1-2019:4. We compute cross correlations for both the whole sample and for an early sample (ending in 1997) and a late sample (starting in 1998). Consistent with the results of Galí and Gertler (1999),¹⁷ there is a significant positive relationship between the labor share and inflation in the early sample period. It is this correlation that the estimates of the Phillips curve pick up. In the last part of the sample, however, the sign of the contemporaneous correlation is negative.

¹⁷Galí and Gertler (1999) measure inflation on the basis of the GDP deflator while we use the PCE deflator.

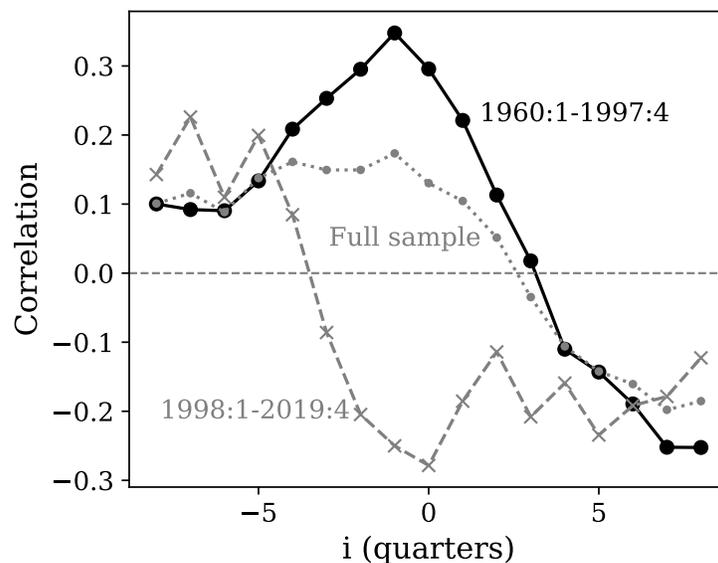


Figure 18.3: Correlation between labor share at t and inflation at $t + i$

Notes: The figure shows the dynamic correlation function between the labor share and lags and leads of the inflation rate. The data have been HP-filtered.

One possible explanation for the falling correlation of inflation and measures of marginal cost is that a smaller share of the variation in inflation now comes from changes in marginal costs (i.e. movements along the Phillips curve) as opposed to other factors (i.e. shifts of the Phillips curve). In this case, estimating κ by concentrating on the economy's response to structural shocks that move the economy along the Phillips curve, for example as pursued by [Christiano et al. \(2005\)](#), may be a preferable approach.¹⁸

18.6 Sticky wages and other extensions

The analysis in this chapter has focused on a simple version of the New Keynesian model with sticky prices. This model is useful for illustrating qualitative implications of nominal rigidities but is not a compelling quantitative explanation of the business cycle as we have seen in [Figure 18.2](#). We now describe some of the main extensions of the New Keynesian framework that allow for a richer description of aggregate dynamics. We will pay particular attention to frictions in wage setting as these are a key source of nominal rigidity in addition to price stickiness.

18.6.1 Sticky Wages

The New Keynesian model that we have examined so far includes nominal rigidities in goods prices. This focus on the goods market contrasts with the traditional Keynesian literature

¹⁸[Christiano et al.](#) specify an entire general equilibrium model. An alternative approach is to estimate the Phillips curve in isolation from the rest of the model using the information contained in impulse response functions following identified shocks. See [Barnichon and Mesters \(2020\)](#) and [Galí and Gambetti \(2020\)](#).

that, motivated by the high levels of unemployment observed during the Great Depression, focused on sticky nominal wages.¹⁹ The lack of instantaneous adjustment of nominal wages is consistent with the empirical fact that the distribution of wage changes observed in micro data tend to have a spike at zero, see e.g. [Kahn \(1997\)](#). Sticky wages result in periods when labor is under-utilized thereby providing an explanation for labor market “slack” in the sense that there are workers, although willing to work and actively searching for a job, who appear unable to find a job at the going wage. Sticky wages also hold some theoretical appeal as the basic sticky-price New Keynesian makes counterfactual predictions for the cyclicity of firm profits.²⁰

The Taylor model of overlapping contracts has considerable empirical and intuitive appeal for many parts of the labor market as many wages are adjusted at an annual frequency (see [Grigsby et al., 2021](#)). However, as we have argued earlier, the Calvo model offers much analytical convenience. As mentioned above, there may be asymmetries in nominal wage flexibility in terms of the wage change distribution being asymmetric with few downwards movements in wages (of continuing employee-employer relationships). Here, to keep the analysis simple, we will maintain the symmetric modeling of wage stickiness. Wage stickiness may also be an insurance mechanism that can arise in the absence of nominal rigidities. Firms may, for example, be prepared to offer workers insurance against variations in wages because they are better able and willing to absorb profit fluctuations than workers are at handling income changes. Such considerations will typically lead to sticky real wages rather than sticky nominal wages and therefore falls somewhat outside the aims of this chapter.

In the sticky-price model we introduced monopolistic competition in the goods market in order to reconcile the assumption of output being demand determined (in the short run) with firms’ willingness to supply goods at a possible predetermined price. A similar assumption is required when modeling sticky wages. The standard avenue taken in the literature is that households rent their labor supply to a continuum of labor unions that differentiate labor and rent it to firms setting a nominal wage that is above the cost that they have acquired it for. Households may therefore sometimes be unable to work the number of hours they would have chosen to in the absence of labor unions and nominal rigidities. However, households are assumed to be the ultimate owners of labor unions and are therefore compensated for this by profits received from these institutions.

In this setup, if, for the sake of discussion, we entertain the idea that wages are sticky and adjusted as in the Calvo model while prices are flexible, the model would be almost identical. In the sticky-price model, the marginal cost of production is the wage (adjusted for productivity) and the wage equals the marginal rate of substitution between consumption and leisure because the household is on its labor supply curve (eq. 18.6). In this setup there is a markup between the marginal cost and the price and due to nominal rigidities this markup can vary in response to shocks. We will now sketch out a sticky-wage model in which prices are equal to marginal costs (production is competitive) but there are frictions in the labor market that introduce a wedge between wages and the marginal rate of substitution between consumption and leisure. This approach to modeling sticky-wages has its roots in [Blanchard](#)

¹⁹A classic reference is [Keynes \(1936\)](#).

²⁰The sticky-price model predicts that firm profits decline following a positive demand shock. The resulting negative wealth effect on labor supply is actually central to the increase in labor supply with balanced-growth preferences, see [Broer, Hansen, Krusell, and Öberg \(2020\)](#).

and Kiyotaki (1987) and was developed in more detail by Erceg, Henderson, and Levin (2000).

The union sets its wage subject to Calvo-style nominal rigidities and stands ready to supply any amount of labor that is demanded at that wage. A competitive final goods producer combines these varieties of labor to produce a final good selling its output at marginal cost. Let W_t be the nominal wage-index that reflects the cost of a bundle of labor that allows the final goods producer to produce one more unit of goods. Since goods are sold at marginal cost, we have $P_t = W_t$ in all periods. It then follows that price inflation will be identical to wage inflation. This model yields the exact same Phillips curve as the sticky-price model. Where the model will differ is its implications for real wages, which are now constant as $w_t = W_t/P_t = 1$. In Section 18.4.1, we argued that the distance between the equilibrium level of output and the efficient level depends on the ratio w_t/A_t . Such a condition no longer holds. Instead, the relevant issue is how the marginal rate of substitution compares to productivity. In the sticky-wage model, households need not be on their labor supply curves and there can be times when they would like to work more, but wages are high and labor demand is therefore “too low,” in the sense of being below its efficient (flexible wage) level.

A model with *both* sticky wages and sticky prices is, however, somewhat different from the models with only one nominal rigidity. We describe this model in detail in Appendix 18.A.3. As we show there, inflation is now explained by three equations that replace the Phillips curve

$$\pi_t = \beta \mathbb{E}_t [\pi_{t+1}] + \xi^P (\hat{w}_t - \hat{A}_t) \quad (18.23)$$

$$\pi_t^w = \beta \mathbb{E}_t [\pi_{t+1}^w] - \xi^w (\hat{w}_t - \hat{A}_t) + \kappa^w x_t \quad (18.24)$$

$$\hat{w}_t = \hat{w}_{t-1} + \pi_t^w - \pi_t. \quad (18.25)$$

Eq. (18.23) is the price Phillips curve. It is similar to the standard New Keynesian Phillips curve but now depends on the real wage w_t rather than the output gap. Goods price setters choose their prices taking account of current and future marginal costs, which in this model is just the real wage relative to productivity.²¹ Eq. (18.24) is the wage Phillips curve where π_t^w is wage inflation (i.e. the growth rate of nominal wages). Wage setters will increase nominal wages if the marginal rate of substitution is high relative to real wages. Therefore eq. (18.24) is increasing in the output gap (the marginal rate of substitution rises as households work and consume more) and is decreasing in the real wage.²² This equation is forward looking for the same reason that the price Phillips curve is. Wage setters know their wage could remain fixed for a number of periods so they look ahead to future market conditions. Finally, eq. (18.25) is the log of the identity $w_t = W_t/P_t = (W_{t-1}/P_{t-1})(W_t/W_{t-1})(P_{t-1}/P_t)$. This equation relates the change in the real wage between any two periods to the difference between nominal wage inflation and nominal price inflation.

Figure 18.2 shows results of simulating the model with both sticky wages and sticky prices. There are two things worth noting here. First, we parameterize this model with

²¹If the model included decreasing returns in production or factors of production other than labor, there could also be an output gap term in eq. (18.23) in addition to the wage term.

²²Eq. (18.24) is also increasing in productivity because there is a wealth effect on labor supply that raises the marginal rate of substitution between leisure and consumption.

double the frequency of price and wage adjustments as in the basic model. In the sticky-wage model, prices and wages update once per year on average while in our calibration of the sticky-price model they updated every two years on average. When nominal rigidities layer on top of each other, the pass through from resource utilization to inflation becomes more gradual. Even when they are able to update their prices, intermediate goods firms only raise their prices to the extent their marginal costs rise and the change in their marginal costs is muted by the wage rigidities. Second, note that inflation is more persistent with the two rigidities. Wage inflation initially falls by more than price inflation leading real wages to fall. Thereafter, the low real wages exert a downward force on price inflation and the real wage only returns to its steady state value gradually.

18.6.2 Other extensions of the basic New Keynesian model

We showed above that the basic New Keynesian model is qualitatively consistent with empirical estimates of the impact of monetary policy shocks. However, as also made clear, quantitatively, the model does not manage to match the data. The same is the case for other structural shocks often studied in macroeconomics such as total factor productivity shocks, shocks to investment efficiency, fiscal shocks, uncertainty shocks, etc. Clearly this basic model fails to capture some important features of business cycle fluctuations. For that reason, much work in the area has considered so-called “medium-scale” models that extend the above framework with the hope of improving the model’s quantitative performance. Here we will discuss a few of these extensions and the underlying reason for their introduction into this line of work.

Consumption dynamics: A main feature of many empirical estimates of the macroeconomic impact of aggregate shocks is gradual adjustment over time. We see this above in Figure 18.2 in the hump-shaped response of output to the identified monetary policy shock, but such dynamics are standard findings in the literature also in response to other shocks. There are many ways in which macroeconomists have attempted to model such dynamics.

Under the permanent-income hypothesis, consumption is determined by permanent income and the path of interest rates. It is impossible then to explain a gradual change in consumption without a very particular path for interest rates. To generate a gradual consumption response, some authors replace the preferences in eq. (18.1) with a specification such as

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{(C_t - \gamma C_{t-1})^{1-\sigma}}{1-\sigma} - \frac{L_t^{1+\psi}}{1+\psi} \right], \quad (18.26)$$

where $\gamma \in [0, 1)$ and C_{t-1} refers to last period’s consumption. These preferences are interpreted as reflecting consumption habits. To see why, note that the term involving consumption can be rewritten as

$$C_t - \gamma C_{t-1} = (1 - \gamma)C_t + \gamma(C_t - C_{t-1})$$

i.e. as a weighted average of the level and the change in consumption. Thus, in habit formation models, households are concerned about smoothing both the level and the growth rate of consumption. When γ is large, households effectively worry about smooth growth

rates of consumption and in this case the level of consumption will tend to adjust partially to shocks over time.²³

An alternative way of generating richer dynamics for aggregate consumption is to abandon the complete markets assumptions underlying the permanent income hypothesis. When households face uninsurable idiosyncratic risks and borrowing constraints, their consumption choices will tend to reflect current labor market conditions. For example, borrowing-constrained agents will spend strongly out of current income so aggregate consumption will respond more strongly to aggregate disposable income. Similarly, an increase in the risk households face, say an increase in the risk of unemployment, will lead them to cut back on consumption for precautionary reasons. These issues are explored in the literature on heterogeneous-agent New Keynesian models.

Capital, investment, and adjustment costs: The model discussed so far has only one factor input: labor. Quantitative models typically also include capital accumulation both because capital is important as a savings vehicle and also because investment demand is an important part of aggregate demand. Typically, one assumes a Cobb-Douglas technology:

$$y_{j,t} = A_t k_{j,t}^\alpha \ell_{j,t}^{1-\alpha}. \quad (18.27)$$

Let's assume that capital is owned by households and rented out to firms at the (real) rental rate r_t^k . Now assume that capital accumulates over time according to a standard neoclassical specification:

$$K_{t+1} = (1 - \delta)K_t + I_t,$$

Unfortunately, this model implies that investment demand becomes extremely sensitive to variations in monetary policy. To see this, note that firms will continuously adjust their capital demand so as to equate the marginal product of capital with the cost of capital $r_t^k + \delta$. In steady state, investment satisfies $I_t = \delta K_t$ and with a small δ , the flow of investment is small relative to the capital stock. Small percentage changes in the desired capital stock then translate to large percentage changes in investment. The model as written would then imply that small changes in interest rates lead to large changes in investment and equilibrium output.

To limit such high investment volatility, it is common in medium-scale models to include sources of adjustment costs. A common specification is:

$$K_{t+1} = (1 - \delta)K_t + \xi \left(\frac{I_t}{K_t} \right) K_t$$

where the function $\xi(I_t/K_t)$ is assumed to be increasing but concave and captures adjustment costs.²⁴ Notice here that, because of concavity, it is costly to vary the investment rate. This specification is therefore able to generate gradual adjustments in the flow of investment and hump-shaped aggregate dynamics.

²³This specification is usually referred to as an internal habit model because the household understands that increasing C_t will affect its utility next period. Another approach treats the habit term as referring to past *aggregate* consumption (taken as given). That approach is called an “external habit” or “catching up with Joneses.”

²⁴Typically one assumes $\xi > 0, \xi' > 0, \xi'' < 0$, and that $\xi(\delta) = \delta$ so that Tobin's Q equals 1 in the deterministic steady state.

Empirically, the slope of the Phillips curve is low relative to what we would expect based on the rigidity of prices as measured in micro-data as we discussed in Section 18.5.2. This tension is exacerbated in the New Keynesian model with capital accumulation. In the short-run, the capital stock is predetermined and the only way to produce more in the aggregate is to use more labor. If we use the production function (18.27), the elasticity of output with respect to labor is $1 - \alpha$, which means we need to use $1/(1 - \alpha) > 1$ units of labor in order to produce one more unit of output. Thus, relative to the specification in (18.3), marginal costs are more sensitive to the quantity produced and the Phillips curve becomes steeper. One way to address this concern is to allow for variable capacity utilization whereby the effective capital services are not predetermined giving the economy an additional opportunity to adjust production beyond changes in labor effort.