How do mark-ups vary with demand?

By Clare Macallan and Miles Parker of the Bank's Structural Economic Analysis Division.

The Monetary Policy Committee's (MPC's) objective is to deliver price stability. In order to achieve that goal, it is necessary to understand how inflation reacts to economic events. In the long run, inflation is determined by monetary policy. But over a shorter time horizon, one important determinant of changes in inflation is the gap between the prices charged by businesses and the costs that they face: that 'mark-up' will influence how changes in demand relative to supply feed through into consumer price inflation. The evidence presented in this article suggests that mark-ups vary positively with excess demand. That will increase the sensitivity of inflation to changes in excess demand. But it could also increase the efficacy of monetary policy, since the level of excess demand is in part determined by the level of Bank Rate set by the MPC.

Introduction: why do we care about the mark-up?

In the United Kingdom, the Monetary Policy Committee's (MPC's) monetary policy objective is to deliver price stability, which is defined by the Government's 2% inflation target for the consumer prices index (CPI). That does not mean that the MPC seeks to keep inflation at precisely 2% every month: any attempt to do so would require large changes in interest rates and so would inject unnecessary uncertainty into the economy. Instead, the Committee aims to set Bank Rate such that when inflation is moved away from its 2% target by a disturbance (or 'shock'), it returns to target within a reasonable time period.

As this discussion makes clear, knowing how inflation is affected by shocks is critical to the Committee achieving their goal of price stability. And in order to understand the dynamics of inflation after a disturbance, it is important to understand how prices move in reaction to economic events. Businesses that have some degree of market power typically set price as a mark-up over marginal cost, with the aim of maximising profits. So crucial to understanding how prices move following a disturbance in the economy are the dynamics of the mark-up, and in particular how it responds to movements in aggregate demand and supply. For example, a key judgement in the May 2008 *Inflation Report* was the extent to which consumer-facing companies pass higher energy and import costs on into higher prices, or absorb them by pushing down on labour costs or accepting lower profits at a time of slowing demand and rising spare capacity.

Economic theory is ambiguous about how the mark-up reacts to movements in demand and supply; that is an empirical question. This article brings together some recent findings on the behaviour of the mark-up, focusing on how it varies with movements in demand relative to supply capacity, referred to in this article as 'excess demand'. The first section defines the mark-up and explains how it is related to the profit margin. The next section explains why theory cannot say how mark-ups respond to changes in excess demand. The third section analyses what the empirical evidence says about that response, drawing on data for both the economy as a whole and for individual industries. The final section draws together the results from this analysis and discusses the implications for monetary policy.

The mark-up and the profit margin: similar but different

The terms 'mark-up' and 'profit margin' are often used in the same context but they are not the same. The 'mark-up' is the gap between the price that a business charges and its marginal cost (the additional cost of producing an extra unit of output). That gap is likely to be lower for businesses which face greater competition: in more competitive markets businesses are more likely to undercut the prices charged by their rivals and that limits their ability to set price above marginal cost. That is not to say that changes in competitiveness are the only factor that drive changes in the mark-up; rather, if there is a disturbance to the economy then the degree of competition will determine the level at which the mark-up will eventually settle.

The 'profit margin' is usually understood to be the ratio of profits (or 'gross operating surplus') to gross output. The annex to this article discusses the relationship between that profit margin and the mark-up. Businesses often talk about raising and lowering their profit margins, and rarely make mention of their mark-up. But it is the mark-up, not the profit margin, which is the variable that each business has in mind when setting its price: given its expectations about costs, a business can, and does, alter its mark-up but it cannot directly control its profit margin. That will instead be determined by demand for the business's output, given its choice of mark-up. So it is the response of the mark-up which determines how changes in excess demand affect businesses' price decisions and not the profit margin. That means that understanding movements in the mark-up is important for understanding the dynamics of inflation and so maintaining price stability.

But measuring the mark-up is tricky: it cannot be directly observed so instead must be estimated; and any estimate will depend on how businesses are assumed to combine inputs (such as capital and labour) to produce output. In contrast, it is straightforward to calculate profit margins. And, as shown in the annex, if the profit margin increases when excess demand increases then the mark-up, in general, will rise too. That means that profit margins can serve as a useful cross-check when policymakers are trying to assess movements in mark-ups: using the reaction of profit margins to changes in excess demand to deduce the likely response of the mark-up limits the criticism that movements in the estimated mark-up reflect mismeasurement.

The response of mark-ups to changes in excess demand: what does theory say?

Economic theory is unable to say unambiguously how mark-ups and profit margins respond to changes in excess demand. That is for (at least) two reasons. First, for any given change in excess demand, the mark-up will be affected by how quickly wages and other costs adjust relative to the speed at which prices change. For an intuitive example, consider the case of an increase in excess demand. If prices adjust slowly but wages are flexible, then the mark-up will fall: in order to supply the extra output demanded businesses must employ more labour; that necessitates higher wages, raising marginal costs; and since prices respond more slowly than wages, costs rise faster than prices, reducing the mark-up. Theory cannot predict whether wages or prices will respond quickest to a change in excess demand and so cannot say how the mark-up will respond to that change.

Second, the level of mark-up that businesses would ideally like to set at the current time — their 'desired' mark-up — may depend on what they think will happen to excess demand going forward. That is because, for a given current mark-up, changes in the level of future excess demand will affect a business's total level of profits over time. There are various theories about how businesses will set their current mark-up given their expectations about excess demand. Some of these, such as the 'customer-market' model of Phelps and Winter (1970) predict that businesses' desired mark-ups will rise when current excess demand is high relative to future excess demand. In that model, a business's current choice of mark-up affects its level of profits, both today and in all future periods, via market share: the higher the mark-up and thus the price that the business charges, the more its customers will switch to rivals offering lower prices and so the lower its future profits will be. But by setting a higher mark-up today, the business will reap greater profits now. So there is a trade-off: a higher mark-up today increases current profits but depresses future profits. Should the business raise its mark-up? That depends on whether it expects excess demand to increase or decrease over the coming periods. If excess demand is expected to fall going forwards, then a given loss of market share will result in fewer lost sales and so the fall in future profits will be smaller than the rise in current profits; if excess demand is expected to rise, then the opposite will be true. That means that businesses will want to raise their mark-up when current excess demand is high relative to future expected excess demand, so mark-ups will be positively correlated with excess demand.

In contrast, other models predict that businesses will reduce their mark-ups when they expect excess demand to fall. For example, in the 'implicit collusion' model of Rotemberg and Woodford (1992), businesses (implicitly) collude over prices rather than set prices independently. That collusion is supported by the threat of lowering prices: if a business undercuts its competitors, then its profits will rise temporarily as it gains sales but fall in the longer run as the other businesses retaliate and cut prices too. The expectation that future demand will be lower than current demand increases each business's incentive not to collude: the temporary rise in profits from increased sales will be greater than the fall in future profits. So in order to keep the collusive agreement going, businesses must lower the current collusive price. That means that businesses' collusive, or desired, mark-up will fall when current demand is high relative to future demand, which implies that mark-ups are negatively correlated with excess demand.

So economic theory is ambiguous about how mark-ups respond to changes in excess demand. That means that it is necessary to examine the empirical evidence on the behaviour of mark-ups. The next section of the article draws on data for the United Kingdom, looking first at the mark-up across the economy as a whole and then studying data for individual industries.

Empirical evidence on the response of mark-ups to changes in excess demand

The economy-wide mark-up

Monetary policy makers focus on how key variables — such as the mark-up — behave across the economy as a whole. So this section examines how the 'aggregate' mark-up changes in response to changes in excess demand.

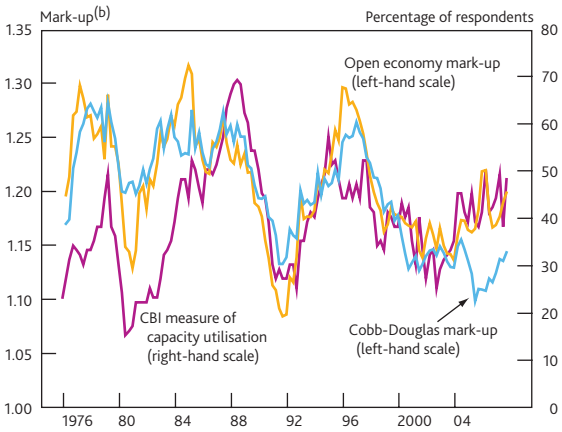
As noted in the annex, one method of estimating the mark-up is to make an assumption about how businesses combine inputs to produce output — their 'production function'. A simple starting case is the Cobb-Douglas function; given that assumption, the aggregate mark-up will be inversely proportional to the aggregate labour share, the ratio of total compensation paid to workers to the value of output produced. But that production function is likely to be too simple a representation of how businesses produce output. So this article also considers estimates of the mark-up implied by two more realistic production functions. The 'labour hoarding' version captures a feature of employment data: businesses are sometimes slow to reduce the number of workers they employ when demand begins to fall. The 'open economy' production function reflects the fact that businesses in the United Kingdom use some imported goods, such as energy and raw materials, to produce output.

Charts 1 and **2** plot the estimated mark-up for the private sector based on these three production functions, over the period 1976 Q2–2007 Q4. Also plotted in the charts is an indicator of excess demand from the *CBI Quarterly Industrial Trends* (*QIT*) survey: the percentage of businesses in the manufacturing sector reporting that their output is above capacity. This variable is often referred to as 'capacity utilisation'.(1) There are, of course, other indicators of excess demand, some of which are used later in the article. But the general picture presented in **Charts 1** and **2** is similar if any of these alternatives are used in place of capacity utilisation.

Chart 1 suggests that the measures of the mark-up derived from the Cobb-Douglas and open economy production functions fall when excess demand falls and rise when it strengthens. But interpreting the evidence in **Chart 2** for the mark-up derived from the production function adjusted for labour hoarding is less straightforward: that measure appears to trough a little before capacity utilisation does, such that it starts to rise while excess demand is still falling, especially in the first half of the sample period.

Table A examines the relationships between the measures of the mark-up and capacity utilisation more formally. It shows that the estimates of the mark-up based on the Cobb-Douglas and open economy production functions are positively correlated with the contemporaneous survey estimate of

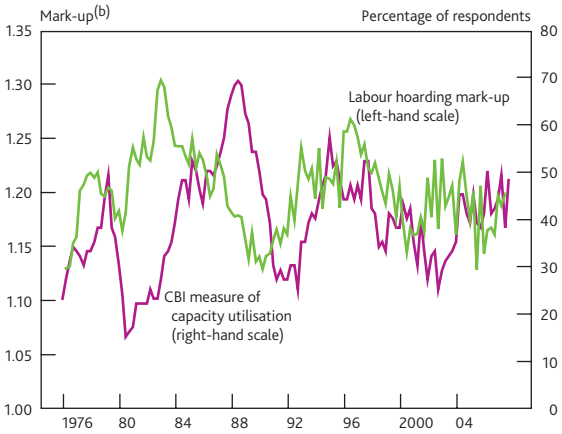
Chart 1 Estimates of the mark-up based on the Cobb-Douglas and open economy production functions and capacity utilisation(a)



Sources: Bank of England, CBI, ONS and authors' calculations.

(a) For full details of how these estimates of the mark-up are constructed, see Macallan, Millard and Parker (2008).
(b) Ratio of price to marginal cost.

Chart 2 Estimate of the mark-up based on the labour hoarding production function and capacity utilisation(a)



Sources: Bank of England, CBI, ONS and authors' calculations.

(a) See footnote (a) to **Chart 1**.
(b) See footnote (b) to **Chart 1**.

Table A Correlations of the estimated aggregate mark-ups based on different production functions with capacity utilisation(a)

	1976 Q2–2007 Q3	Since 1992
Cobb-Douglas	0.28**	0.43**
Labour hoarding	-0.09	0.30*
Open economy	0.44**	0.70**

Note: * denotes significant at the 5% level; and ** denotes also significant at the 1% level.

Sources: Bank of England, CBI, ONS and authors' calculations.

(a) See footnote (a) to **Chart 1**.

(1) This measure is the percentage of businesses responding 'no' to *Question 4* in the survey. It would be preferable to have an estimate of the level of capacity utilisation in the economy as a whole. But such a measure is not available for much of the sample period, since surveys of the services industries have a shorter back run than the *CBI QIT* survey. However, for the period when service sector data are available, there is quite a close correlation between capacity utilisation in the two sectors.

capacity utilisation. Against this, the measure that allows for labour hoarding appears to move in the opposite direction to changes in excess demand, although the correlation is not significant at the 10% level.

Given that there is some evidence to suggest that the mark-up varies positively with changes in excess demand, it is interesting to investigate whether these correlations have changed over time. In particular, did the relationship between the mark-up and demand alter after 1992? In that year, the United Kingdom adopted inflation targeting. Furthermore, between 1992 and the end of the sample period there were no recessions in the United Kingdom and the volatility of output was greatly reduced. So businesses were operating in a different economic environment and that might have caused them to change the way in which they set prices relative to costs.

The second column of **Table A** provides some insight into this issue. It shows that, for the period from 1992 to 2007 Q3, all three measures of the mark-up were positively correlated with capacity utilisation, a result which contrasts with those for the whole sample. So the evidence suggests that mark-ups are now more likely to respond positively to changes in excess demand than they were prior to 1992. That finding is broadly consistent with existing empirical evidence for the United Kingdom. For example, Haskel *et al* (1995) found a positive correlation between mark-ups in the manufacturing sector and excess demand.(1)

Industry-level mark-ups and profit margins

So far, the article has investigated how the aggregate mark-up varies with changes in excess demand. But it is also instructive to examine whether there are any differences in the way mark-ups respond to excess demand in different industries; if such differences exist then the behaviour of the mark-up across the economy as a whole will depend on how shocks to the economy affect individual industries and sectors. For example, a slowdown led by the financial and consumer sectors could have a very different effect on the aggregate mark-up than one led by a downturn in the construction sector.

Small (1997) has investigated mark-ups and profit margins at the industry level. Using a slightly different method to that outlined in the annex, he calculated mark-ups for six broad industries within the services sector and ten, more narrowly defined, industries within manufacturing from 1968 to 1991. These estimated mark-ups were positively correlated with indicators of excess demand such as capacity utilisation, and to a similar extent in all 16 industries.

Small's (1997) analysis can be extended using data from the Bank of England industry data set (BEID).(2) This data set divides the market sector of the economy into 31 industries, including both manufacturing and service sector industries. The data are annual and cover the period from 1970 to 2003. Using the BEID, the relationship between mark-ups and excess demand can be studied in finer detail than Small's (1997) study; and the analysis can be extended to investigate how mark-ups have changed since 1992, a period not covered by Small's (1997) sample.

Table B shows how mark-ups vary on average across industries in response to changes in three different indicators of aggregate excess demand. A range of indicators is used because excess demand cannot be directly observed and so must be estimated. One common method for doing that is to take data on actual output, or labour input, and use statistical techniques to remove 'trend' or 'normal' variation. The three indicators in **Table B** reflect that method: the deviation of log private sector output from its trend, as estimated by a statistical method called the Hodrick-Prescott filter; and the deviation of log private sector hours from its trend, which is either assumed to be linear or measured by a Hodrick-Prescott filter.(3)

Table B Estimates of the response of mark-ups on average across industries to changes in excess demand

Excess demand indicator	Average market sector industry response of mark-ups to changes in excess demand indicator[a]
Hodrick-Prescott filtered GDP	0.013**
Linearly detrended hours	0.038**
Hodrick-Prescott filtered hours	0.048**

Note: ** denotes significant at the 1% level.

Sources: Bank of England and authors' calculations.

(a) Absolute change in mark-ups on average across the market sector industries in response to a 1 percentage point change in the indicator of excess demand.

What can these indicators say about the relationship between mark-ups and excess demand? A positive number in **Table B** suggests that mark-ups, on average across industries, vary positively with excess demand. So the BEID data support Small's (1997) findings: mark-ups in the 31 market sector industries on average respond in the same direction as changes in excess demand.

Unfortunately, the BEID does not contain sufficient years of data to examine how mark-ups respond to excess demand in each industry separately. But it is possible to examine how profit margins react at this level of detail. That is useful because, as noted in the annex, under fairly general conditions changes in excess demand will cause the mark-up and the

(1) These results are somewhat at odds with evidence for the United States, the country studied in most of the existing literature. For example, Rotemberg and Woodford (1999) find that estimates of the mark-up calculated from the more realistic production functions move in the opposite direction to excess demand.
(2) For a detailed description of the BEID, see Oulton and Srinivasan (2005).
(3) The indicator of excess demand used in the charts above, capacity utilisation as measured by the *CBI QIT* survey, cannot be used here because the method used to estimate the results in **Table B** requires more years of data than are available.

profit margin in any particular industry to move in the same direction. The BEID data suggest that profit margins are, in general, positively correlated with various indicators of excess demand, although there are only a few industries in which the correlation is particularly high (**Table C**). That appears to be because, at the industry level, profit margins have sometimes been subject to sudden changes caused by sector-specific events and these movements tend to dominate any other variation in profit margins.

Table C Market sector industries in which the correlation between gross profit margins and capacity utilisation is significant at the 10% level[(a)]

Industry	Correlation coefficient
Non-metallic mineral products	0.68**
Basic metals and metal goods	0.65**
Paper, printing and publishing	0.60**
Chemicals and pharmaceuticals	0.60**
Agriculture	0.55**
Food, drink and tobacco	0.45*
Vehicles	0.41*
Retailing	0.39*
Business services	0.39*
Hotels and catering	0.34
Mechanical engineering	0.34
Water supply	0.32
Communications	0.32

Note: * denotes also significant at the 5% level; and ** denotes also significant at the 1% level.

Sources: Bank of England, CBI and authors' calculations.

(a) Capacity utilisation is the CBI measure used in **Charts 1** and **2**, averaged over the year.

The result that profit margins respond in the same direction as excess demand is also supported by business-level evidence. Using annual company accounts data for the period from 1972 to 1992, Small (1997) finds that profit margins calculated at this level of disaggregation vary positively with a variety of excess demand indicators.

Implications for monetary policy

The evidence presented above suggests that the mark-up in the United Kingdom is positively correlated with excess demand, both at the aggregate and at the industry level, consistent with the 'customer-market' model outlined above. That means that, other things being equal, mark-ups will tend to add to inflationary pressures when demand is growing more strongly than supply, and put downward pressure on inflation as the amount of spare capacity increases. The MPC seeks to achieve its inflation target in part by changing Bank Rate to influence the level of demand in the economy. So the relationship between mark-ups and excess demand is an important part of the transmission mechanism of monetary policy.

The results also suggest that the effect of an increase in input costs on consumer prices will depend on the state of the economy: if costs rise at a time when demand is relatively weak then businesses will seek to maintain sales by dampening the impact of those cost increases on their prices — that is, by reducing their mark-ups; and that reduction in mark-ups will delay the full pass-through of the rise in costs into inflation. That is consistent with the central projection in the May 2008 *Inflation Report* which assumes that businesses' profit margins and mark-ups will contract as the amount of spare capacity in the UK economy increases. But that contraction only partially offsets the impact of the sharp rises in energy and import costs on inflation.

Conclusion

This article has examined how mark-ups and profit margins vary with excess demand. Understanding those movements is crucial to understanding how inflation responds to changes in economic disturbances and so helps the Monetary Policy Committee to achieve their goal of price stability. Since theory does not make clear predictions about how the mark-up varies with excess demand, answering this question becomes an empirical matter.

The evidence presented in this article, consistent with previous work, suggests that the mark-up in the United Kingdom is positively correlated with excess demand, both at the aggregate and at the industry level. Furthermore, the correlation between the aggregate mark-up and capacity utilisation appears to have strengthened since 1992, which could suggest that the mark-up has become more responsive to changes in excess demand over the past fifteen years. These results support the assumption in the May 2008 *Inflation Report* that businesses' profit margins contract somewhat as the amount of spare capacity increases and as businesses' non-wage costs rise.

That mark-ups vary positively with excess demand also implies that either prices respond to economic events faster than do wages and/or that businesses' desired mark-ups also vary positively with excess demand. Trying to distinguish between these alternatives could be a useful avenue for future work. Indeed, the Bank has recently carried out a survey of the price-setting behaviour of around 700 companies partly in order to obtain an answer to this question. The results from this survey, including a comparison to a similar survey conducted in the mid-1990s, will be published in due course.

Annex
Understanding the aggregate mark-up and the aggregate profit margin

The aggregate mark-up

The term 'mark-up' is typically used to refer to the gap between the price that a business charges, P, and its marginal cost, MC (the additional cost when output is increased by one unit). The mark-up cannot be calculated from this definition, since marginal costs are unobservable. But it can be shown that, for a profit-maximising business, the mark-up will be given by:

$$\text{Mark-up, } m = \frac{P}{MC} = \frac{\varepsilon_{yh}}{s} \qquad \textbf{(A1)}$$

That is, the mark-up, m, is equal to the elasticity of output with respect to labour input, ε_{yh} (which reflects how much extra output can be produced when the business hires an additional unit of labour), divided by the labour share, s (the amount of nominal GDP that accrues to workers in the form of compensation). This relationship will hold in the economy as a whole, as well as for an individual business.

An economy-wide labour share can be easily estimated using data in the National Accounts. But calculating the elasticity of output is trickier: the formula used in that calculation will change depending on what one assumes about the 'representative' business's production function — that is, how inputs are combined to produce output. One of the simplest assumptions to make is that businesses have a 'Cobb-Douglas' production function:

$$y = Ah^{\alpha}k^{(1-\alpha)} \qquad \textbf{(A2)}$$

Here, h is the business's labour input in hours and k its capital input; A represents its level of technology; and α is a fixed parameter which reflects the proportions in which the business combines labour and capital to produce its output, y. One property of the Cobb-Douglas production function is that the elasticity of output with respect to labour input (ε_{yh}) is a constant (α). And that means that the mark-up will be proportional to the inverse of the labour share.

In practice, the Cobb-Douglas production function is likely to be too simple a representation of how businesses produce output. For example, it assumes that businesses always employ just enough labour in order to produce their desired level of output. But businesses may hoard labour — that is, they do not fire workers immediately when demand for their output falls, probably because the fall in demand could prove temporary and it is costly to hire and fire workers. Furthermore, the United Kingdom is an open economy and so businesses will usually import goods, such as energy and raw materials, in order to produce output. The production function can be adapted so that it allows for these alternative assumptions.

Using alternative production functions does not mean that equation **(A1)** cannot be used to estimate the mark-up. But with different assumptions, the mark-up will no longer be proportional to the inverse of the labour share.

The aggregate profit margin

The term 'profit margin' is used to denote the ratio of profits (or 'gross operating surplus') to some other variable. One common choice for that other variable is the capital stock; that measure of the profit margin is also termed the 'rate of return on capital'. But since it is difficult to obtain reliable capital stock data, this article focuses instead on the ratio of total profits to gross output, a measure sometimes termed the 'gross profit share'. So if the business above, which sells its output, y, at price P, pays a wage of W to each unit of labour input and uses another composite intermediate input, x, which costs q per unit, then its gross profit margin will be:

$$\frac{Py - Wh - qx}{Py} \qquad \textbf{(A3)}$$

The relationship between the mark-up and the profit margin

How does the mark-up relate to the profit margin? It can be shown that:

$$\text{Gross profit margin} = \frac{Py - Wh - qx}{Py} = 1 + r\frac{k}{y} - \frac{1}{m} \qquad \textbf{(A4)}$$

Equation **(A4)** highlights a key relationship between the mark-up and the profit margin: the response of the profit margin to a change in demand will depend on how the opportunity cost of capital (r), the capital to output ratio ($\frac{k}{y}$) and the mark-up react to that change in demand. Data show that capital remains relatively fixed over the cycle, so the capital to output ratio is unlikely to increase when demand rises; the data also suggest that there is little correlation between the cost of capital and excess demand. That means that the only way that the gross profit margin can rise when demand rises is if its third determinant, the mark-up, also rises with demand.

References

Haskel, J, Martin, C and Small, I (1995), 'Price, marginal cost and the business cycle', *Oxford Bulletin of Economics and Statistics*, Vol. 57, pages 25–41.

Macallan, C, Millard, S and Parker, M (2008), 'The cyclicality of mark-ups and profit margins for the United Kingdom: some new evidence', forthcoming *Bank of England Working Paper*.

Oulton, N and Srinivasan, S (2005), 'Productivity growth in UK industries, 1970–2000: structural change and the role of ICT', *Bank of England Working Paper no. 259*.

Phelps, E S and Winter, S G (1970), 'Optimal price policy under atomistic competition', in Phelps, E S *et al* (eds), *Microeconomic foundations of employment and inflation theory*, New York, W W Norton and Co.

Rotemberg, J and Woodford, M (1992), 'Oligopolistic pricing and the effects of aggregate demand on economic activity', *Journal of Political Economy*, Vol. 100, No. 6, pages 1,153–207.

Rotemberg, J and Woodford, M (1999), 'The cyclical behaviour of prices and costs', in Taylor, J B and Woodford, M (eds), *Handbook of macroeconomics*, Vol. 1B, Amsterdam, Elsevier.

Small, I (1997), 'The cyclicality of mark-ups and profit margins: some evidence for manufacturing and services', *Bank of England Working Paper no. 72*.