Working Paper no. 300

Elasticities, markups and technical progress: evidence from a state-space approach

Colin Ellis

July 2006

Bank of England

Elasticities, markups and technical progress: evidence from a state-space approach

Colin Ellis

Working Paper no. 300

Monetary Analysis Division, Bank of England, Threadneedle Street, London, EC2R 8AH.
Email: colin.ellis@bankofengland.co.uk.

The views expressed are those of the author and do not necessarily reflect those of the Bank of England or Monetary Policy Committee members. I am very grateful to Richard Barwell, Charles Bean, Charlotta Groth, Richard Harrison, Neal Hatch, Richard Pierse, Simon Price, Fergal Shortall, and two anonymous referees for their help, comments and suggestions. I am also in debt to Rhys Foulkes, Charlotte James, Sylaja Srinivasan and Richard Windram for their help with collecting data. Needless to say, that does not implicate anyone else in any mistakes I have made. This paper was finalised on 19 April 2006.

The Bank of England's working paper series is externally refereed.

Information onthe Bank's working paper series can be found at
www.bankofengland.co.uk/publications/workingpapers/index.htm.

Publications Group, Bank of England, Threadneedle Street, London, EC2R 8AH; telephone +44 (0)20 7601 4030, fax +44 (0)20 7601 3298, email mapublications@bankofengland.co.uk.

ISSN 1749-9135 (on-line)

Contents

Abstract

Conventional techniques for estimating the elasticity of substitution between capital and labour in the production process typically focus on factor-demand equations. An implicit assumption in this approach is normally that the markup is stationary. But that may not be true. This paper considers a new approach that models the markup as an unobserved variable. Using the factor-demand equations for capital and labour, technical progress can also be estimated as a stochastic process, rather than just imposing a time trend. The resulting estimates of the whole-economy markup for the UK economy suggest that it has fallen over the past 30 years, and this result appears to withstand a variety of robustness checks. The estimated elasticity is somewhat lower than most previous estimates. This implies that conventional techniques may be misleading.

Key words: Markups, factor demands, technical progress.

JEL classification: C32, D40, E23, E30.

Summary

UK monetary policy is concerned with keeping inflation – the rate of increase in prices – on target at 2% a year. So it is important for policymakers to consider how firms set prices. Typically, economists work with models that assume companies set their output prices as a markup over marginal cost – that is, the cost of producing an extra unit of output.

In most economic models, that markup is assumed to be fixed, at least on average over a long period of time. But in practice, it is possible that the markup could have changed over time, for example if competition between companies becomes more intense. At the same time, standard economic models often impose an assumption about production technology: in particular, how easy companies find it to swap between machines (capital) and workers (labour) when they produce their output. This is called the elasticity of substitution in production. In fact, any assumption about the markup will affect the estimated elasticity of substitution in a model, and vice versa.

This paper proposes a new approach, where the markup and elasticity are jointly estimated. In particular, the markup is allowed to (potentially) vary over the past 30 years. The model is estimated using so-called 'state-space' techniques, which allow the unobserved markup to be modelled using UK data on prices, wages and other macroeconomic variables. The estimation results are very different from what standard approaches find – in particular, the state-space approach suggests that the aggregate markup in the UK economy has fallen by around a quarter since the early 1970s, and that firms find it harder to swap between capital and labour than is often assumed. In addition, the model also lets technical progress in the economy – a gauge of the efficiency with which firms use capital and labour to make output – be estimated in a more realistic manner than in most models. This turns out to be crucial – the usual approach in other work, of simply including a time trend in the model, is shown to give misleading results.

The key results from using the state-space model are robust to a number of consistency checks, such as the degree of tightness in the labour market, looking at the private sector rather than the economy as a whole, and measuring how useful machines are in production, rather than what they are worth. Given that the model focuses on long-run effects, data from the 19th century are used to check that running the model from 1970 is not misleading. Finally, the model is applied to US data, again retrieving plausible results.

This new approach of treating the markup as unobserved and estimating it at the same time as production technology yields several insights. First, the markup in the United Kingdom has fallen over the past 30 years or so. This implies that the unit labour cost of production – essentially the pay workers receive for each unit of output they produce – has not always been a good guide to the marginal cost of production, despite it being widely used to proxy marginal cost in previous work. Second, firms find it harder to swap between capital and labour in production than most other estimates suggest. Finally, using a time trend to proxy technical progress can be very misleading.

1 Introduction

Markups are an important concern for monetary policy. In standard models, firms set prices as a markup over marginal cost. So if markups vary over time, either with the cycle or due to structural changes, policymakers need to be aware of them. Otherwise they could misjudge inflationary pressures in the economy: for example if costs remain weak, but the markup is rising when policymakers believe it is fixed, then inflation will be higher than expected. Changes in competition, which can affect the markup, can have important implications for policy, as shown by Khan and Moessner (2005). Markups tend to be investigated in the context of the firm's factor-demand equation for labour: for the United Kingdom examples include Price (1992) and Ellis and Price (2003). Changes in the markup are generally proxied by including some cyclical variable (eg Smith (2000)) or a role for competitor's (import) prices (eg Agoloskoufis *et al* (1990)).

Yet these approaches often ignore important concerns when considering markups. A standard assumption is that production technology is Cobb-Douglas: and apart from recent evidence that refutes this (see Ellis and Price (2004)), *any* imposed assumption about technology is crucial for the implied estimate of the markup. Calibrating the elasticity of substitution at different values will affect the resulting markup estimate. In addition, the fact that the markup is also present in the firm's capital-demand equation is generally ignored.

In this paper I address the problem of the estimated production technology and implied markup approximations being dependent on each other. This is done by treating the markup as fundamentally unobserved, and using state-space techniques to generate estimates. While such techniques have been used to estimate NAIRUs (eg Greenslade *et al* (2003)) and potential output (Kuttner (1994)), their application to markups is an innovation. I show that relying solely on the labour-demand equation can lead to misleading results, and hence that the capital-demand equation should also be considered when investigating technology and markups.

The remainder of the paper is set out as follows. In the next section I consider what the first-order conditions for a profit-maximising firm imply in the long run, and note some common techniques – and problems – in estimating them, before proposing a state-space method that overcomes these. I then apply the model to a variety of data. Section 3 presents results for the UK economy as a whole over the past 30 years or so. Section 4 uses simulation techniques to investigate results from different models, and Section 5 discusses some possible concerns with the technique, and offers solutions. Section 6 examines the private sector, and Section 7 takes a longer-run approach, applying the technique to data that stretch back to the 19th century in order to shed light on the shorter-sample results. Section 8 then briefly sets out the results from applying it to US data. Finally, Section 9 concludes.

2 Theory

2.1 Factor-demand relationships

In the neoclassical model, the key long-run relationships where markups play a role are the first-order conditions (FOCs) for capital and labour for a profit-maximising firm. These long-run relationships do not incorporate structural dynamics, although under certain conditions factor-demand equations can be derived that are driven by the static FOCs.(1)

Simple economic theory states that the price an imperfectly competitive firm optimally charges will increase as the demand curve becomes more inelastic. In other words, anything that makes demand less elastic will increase the profit-maximising markup.

Using a constant returns to scale (CRS), constant elasticity of substitution (CES) production function with labour augmenting technical progress (a),(2) output (Y) is defined as:

$$Y^S = [\alpha K^{-\theta} + (1-\alpha)(Ne^a)^{-\theta}]^{-1/\theta} \qquad \textbf{(1)}$$

where K denotes capital, N denotes labour, α is the distribution parameter in the production function(3) and $1/(1+\theta)= \sigma$ is the elasticity of substitution between capital and labour in production.

Using a constant elasticity demand curve

$$Y^D = P^{-\varepsilon} \qquad \textbf{(2)}$$

where P denotes (output) price and ε is the elasticity of demand, we can derive the first-order conditions for a profit-maximising firm with respect to the two factor inputs, capital (K) and labour (N). The first-order condition for labour yields

$$\frac{\partial}{\partial N}\{Y^{1-\frac{1}{\varepsilon}}\} = W \qquad \textbf{(3)}$$

where W denotes the cost of employing labour. (**3**) solves as

$$W = P(1-\frac{1}{\varepsilon})(1-\alpha)e^{-\theta a}(\frac{Y}{N})^{1+\theta} \qquad \textbf{(4)}$$

Taking logs of (**4**) and re-arranging yields the standard factor pricing equation for labour:

---

(1) See Ellis and Price (2004). Later on the role of dynamic factors will be addressed.
(2) Technical progress is normally assumed to be labour-augmenting in CES production functions, rather than capital-augmenting. Simple justifications for this are based on single good models (see Barro and Sala-I-Martin (1995) and Solow (1999)).
(3) This governs the allocation of income between capital and labour.

$$w-p=-\mu+\ln(1-\alpha)+\frac{\sigma-1}{\sigma}a+\frac{1}{\sigma}(y-n) \qquad (5)$$

where lower case denotes natural logarithms and μ is the (log of the) markup, defined as

$$\mu=\ln(\frac{\varepsilon}{\varepsilon-1}) \qquad (6)$$

The markup is a function of the elasticity of demand (ε). So any change in the markup reflects a change in the elasticity of demand, or equivalently a change in competition. (**5**) can be re-written as the factor-demand equation for labour:

$$y=n+\sigma(w-p)+(1-\sigma)a-\sigma\ln(1-\alpha)+\sigma\mu \qquad (7)$$

A special case of (**7**) arises with Cobb-Douglas technology, where the elasticity of substitution is unity (σ = 1). In this instance output per worker (hereafter labour productivity) equates to real wages (*w-p*) plus the markup.

$$y-n=w-p+\mu+\beta \qquad (7a)$$

Subject to the capital accumulation identity,

$$K_{t+1}=(1-\delta)K_t+I_t \qquad (8)$$

the factor-demand equation for capital can also be derived:

$$y=k+\sigma r-\sigma\ln(\alpha)+\sigma\mu \qquad (9)$$

where r denotes the (log of the) real user cost of capital (RCC).(4) Both the labour (**7**) and capital (**9**) factor-demand equations should hold in long-run equilibrium – the main difference between the two equations is the lack of technical progress in the latter. This follows from the assumption that technical progress (TP) is labour-augmenting.(5)

2.2 Model assumptions and identifying changes in the markup

Based on these long-run conditions, what would we expect to see if the markup changed? The simplest case to consider is when technology is Cobb-Douglas and the relative price of capital is fixed. In that instance, the factor-demand equations are:(6)

$$\begin{aligned} y-n&=w-p+\mu \\ y-k&=r+\mu \end{aligned} \qquad (10)$$

(4) The real user cost is described in more detail in the data appendix of Ellis and Price (2004).
(5) In a Cobb-Douglas framework, it does not matter if technical progress is capital or labour-augmenting (or both) as the term drops out. But in a more general CES framework, it does.
(6) The distribution parameter terms (α) have been dropped for simplicity.

Note that all of the variables in (**10**) may (at least partly) be endogenous. But that does not prevent us identifying and exploiting the long-run relationships.

From (**10**), changes in the markup will also be evident in other variables – for example, an increase in μ would be visible as a fall in the labour share. Similarly, the cost of capital and/or the capital-output ratio would also change.

However, when relative prices are changing over time the picture is more complicated. In particular, suppose that the relative price of capital is falling, and hence that the capital-output ratio is rising over time. A fall in the markup could also be related to a rise in the capital-output ratio. So it may not be simple to distinguish between these two events.

Of course, in that instance we could still refer back to the labour share. But suppose also that technology was not Cobb-Douglas. Now, changes in real wages may not be matched one for one by changes in labour productivity – and hence the labour share may not be constant over time – without any changes in the markup.[7]

In just two short steps we have moved from a world in which it is easy to identify changes in the markup from the labour share, or the capital-output ratio, to one where it is much harder. The simple world of Cobb-Douglas and fixed relative prices is understandably appealing – but it may not be the real world (see Ellis and Price (2004) and Ellis and Groth (2003)). And the simple implications of a change in the labour share for the markup in a Cobb-Douglas world do <u>not</u> carry across to a more general model.

2.3 A more general model

How are markups identified in the more general CES model? Factor pricing equations state that firm's optimal prices are the markup over marginal cost. But the markup over average cost (ie the profit share) could be a good guide, certainly in long-run equilibrium when all inputs to production are variable and marginal cost should equal average cost.[8] In fact, stationary markups are generally imposed in steady-state growth models: and in a CES framework, that implies that labour productivity and technical progress grow at the same rate. This is illustrated by (**11**), which is a simple transformation of (**7**).

$$\mu = (p-w) + (y-n) + \left(\frac{1-\sigma}{\sigma}\right)(y-n-a) + \ln(1-\alpha) \qquad \textbf{(11)}$$

Assuming the distribution parameter is fixed for simplicity, a simple interpretation of (**11**) is:

$$\textit{Markup} = \textit{inverse labour share} + \left(\frac{1-\sigma}{\sigma}\right)*(\textit{labour productivity} - \textit{technical progress}) \qquad \textbf{(12)}$$

---

[7] This is discussed in more detail in the next section. Note that with CES technology and a falling relative price of capital, the economy is not on a balanced growth path. For forecasting purposes, Harrison *et al* (2005) resolve this by assuming that the relative price stabilises at some point in the future. Bakhshi and Larsen (2001) impose Cobb-Douglas technology, thereby avoiding this problem.
[8] Average markups, or margins, are sometimes used in empirical analysis (eg Leith and Malley (2003)).

By definition, in steady state the markup will not vary over time. Similarly, the labour share is normally assumed to be constant in steady state. So we may interpret technical progress – hereafter TP – as 'trend' labour productivity in steady state.

The expression offers a simple intuition for change in the markup. When labour productivity is above trend, the markup is rising (conditional on the labour share). When it is below trend, the markup is falling. And the harder it is to swap between capital and labour in production (the lower σ is), the bigger any change in the markup will be.[9]

But the markup itself, μ, is an unobserved variable. It is often assumed to be constant in estimation; this essentially imposes steady-state behaviour. Alternatively, the markup is modelled by including other variables such as capacity utilisation or import prices: see Bank of England (2000) or the discussion in Ellis and Price (2003). In the former case, this approach is based on the belief that changes in markups may be related to the business cycle. In the latter, the inclusion of import prices is often justified on grounds of increased competition, particularly from abroad in the case of a small open economy. But essentially these approaches 'add-on' these other variables to a constant, and therefore the resulting implied estimate of the markup is restrictive – it corresponds to the (transformed) behaviour of the variables that are used to model the markup. In general, there is no reason to believe that the behaviour of the markup must correspond exactly to these augmenting series. A less restrictive approach would be to allow the markup to vary over time in a non-dependent framework. The next section discusses how this could be achieved.

2.4 A method for estimating markups and technology

In estimation, (**7**) can be written as

$$y_t - n_t = \beta_1(p_t - w_t) + \beta_2 t + \beta_1\mu_t + \beta_0 \qquad \textbf{(13)}$$

where the markup is assumed to be time-varying and technical progress is proxied using a time trend (t). We need to model the time-series properties of the markup in some way, and crucial to this is whether the markup is stationary or not. If it is, then we can model the markup using a time-series process, such as an autoregressive (AR) or moving average (MA) model (or an ARMA model). This is a simple – and normally reasonably good – way of modelling time-series data: see Blanchard and Fischer (1989). But if the markup is non-stationary (I(1)), then we would need to model the *first difference* of the markup as a stationary time-series process, and then integrate this up to obtain a time series for the markup itself.

The idea of the markup being I(1) is unconventional. In steady-state models, it is fixed. But any change in the degree of competition would affect the markup. So it is plausible that the markup could be non-stationary in-sample if competition has changed, regardless of its imposed

[9] Note however that this 'intuition' says nothing explicitly about causality – it is merely coincident observation.

steady-state behaviour.(10) And, if the degree of competition has increased steadily over time, we would expect to observe a long-running decline in the markup.

Despite this, modelling the markup as non-stationary may seem counterintuitive. While in short samples the markup may be I(1), if we take non-stationarity seriously then estimates could be explosive. It is worth remembering that any estimate of the markup will be tied down by the observed series, such as the price to labour cost ratio and labour productivity. But in an unbounded non-stationary model, the behaviour of the estimated markup will serve to act as a robustness check on the results. If the estimate were explosive – such as a 1,000% fall, which is possible under a random-walk assumption – then that would indicate the model is likely to be misspecified.

For example, a simple model for the markup would be a random walk. That model can be represented in state-space form:

$$\begin{aligned} (y-n)_t &= \beta_0 + \beta_1 (p-w)_t + \beta_2 t + \beta_1 \mu_t + \eta_t \\ \mu_t &= \mu_{t-1} + \lambda_t \end{aligned} \qquad \textbf{(14)}$$

where η and λ are independent normal disturbances:

$$\begin{aligned} E(\eta) &= 0 \\ E(\lambda) &= 0 \\ E(\eta^2) &= \sigma_\eta^2 \\ E(\lambda^2) &= \sigma_\lambda^2 \\ E(\eta, \lambda) &= 0 \end{aligned} \qquad \textbf{(15)}$$

Recall that equation (**13**) is a long-run relationship. There are likely to be cyclical movements around that relationship, which could play an important role in small samples.(11) So some gauge of those cyclical movements might also be useful, although this may require imposing structure on the model. In estimation, the 'smoothness' of the markup estimate can be influenced by imposing the noise to signal ratio on the relative magnitude of the two disturbance terms:

$$\kappa = \frac{\sigma_\eta}{\sigma_\lambda} \qquad \textbf{(16)}$$

Investigating the time series of the markup under different values of κ is another way of testing the robustness of any results: essentially the state-space estimates should just be smoother when the model is estimated with a higher noise to signal ratio.

(10) From Section 2.3, note also that even with a non-stationary markup, the implied behaviour of the labour share would not necessarily be non-stationary as well, depending on production technology and technical progress.
(11) If the sample includes relatively few complete cycles, the estimation results could be biased.

The model described above is deliberately simple. But in practice, the markup estimate could be affected by a number of factors that are not allowed for in the model. For example, the use of a time trend to model technical progress, or the precise definition of capital, could result in the markup 'estimate' reflecting factors other than the markup itself. Later on in the paper, a number of refinements will be explored to address these issues.

I use the Kalman Filter to build the likelihood function, which is then maximised in estimation. This allows the markup and the elasticity of substitution to be estimated jointly. The Kalman Filter has been used in previous studies to estimate other unobserved economic variables of interest, such as the NAIRU (see Greenslade *et al* (2003)) and potential output (see Kuttner (1994)). Importantly, to start with I impose no priors on the markup apart from whether it is I(1) or I(0): I simply use the observed components of the structural factor-demand equations that we know the markup is part of. So any 'true' movement in the markup that is normally captured by proxy variables should also be visible in any state-variable estimates. Crucially, this approach returns a markup and elasticity that have been jointly estimated, rather than one being dependent on an assumption about the other. The next section reports results for the UK economy, based on 30 years of quarterly data.

3 Results for the UK economy: the past 30 years

3.1 Data

Having set up the model, I estimate it for the UK economy as a whole. The GDP deflator is used as the measure of whole-economy prices. Data for the price to labour cost ratio and labour productivity are shown in Charts 1 and 2 below (note that the charts have log scales). The labour productivity measure is based on the Labour Force Survey (LFS) measure of employment, which starts in 1971 (see Lindsay and Doyle (2003)). The sample finishes at the end of 2003.

Chart 1
Ratio of price to labour cost (*p*-*w*)

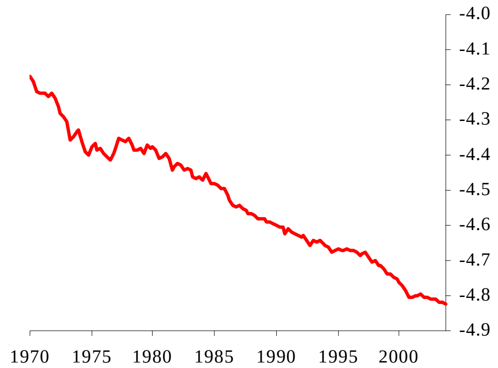
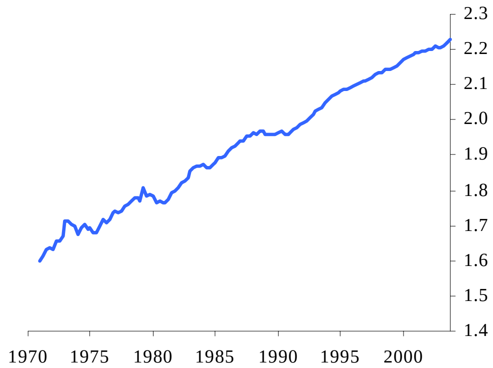


Chart 2
Labour productivity (*y*-*n*)



In addition, I will also estimate the capital-demand equation. That requires data on the user cost of capital and the capital-output ratio: these series are shown in Charts 3 and 4 (which also have log scales).(12)

Chart 3
The capital-output ratio (*k*-*y*)

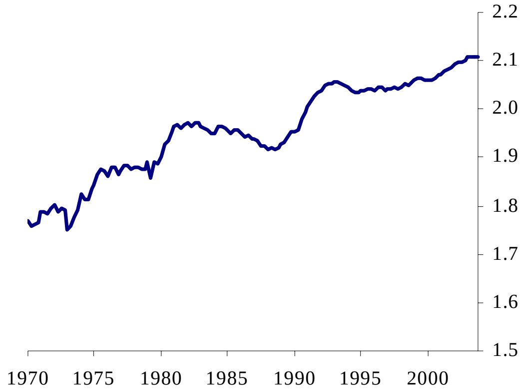
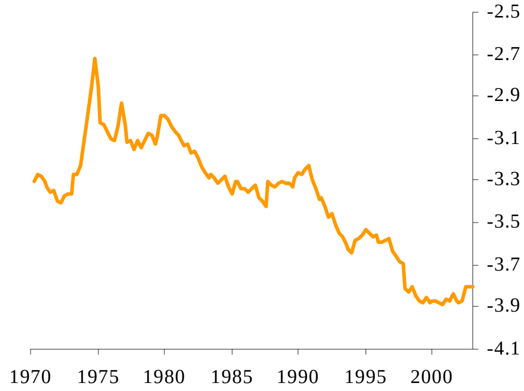


Chart 4
Real user cost of capital (*r*)



3.2 Stationarity of markups and irreducible cointegration

A key question is whether the markup should be modelled as a stationary process or a non-stationary one. Given the possibility of a change in the markup – ie non-stationary behaviour in-sample – there is no obvious answer. One thing that may yield insight is the notion of irreducible cointegration.

Irreducible cointegration was developed by Davidson (1994, 1998). A cointegrating vector is irreducible if none of the cointegrating variables can be omitted without the loss of the cointegration property. The converse of this is that if a new variable is added to a vector that is not cointegrating, and the enlarged vector becomes cointegrating as a result, that new variable is part of the irreducible vector.

How does this help? If the price to labour cost ratio and labour productivity are not cointegrated, that would suggest that the markup term is part of the irreducible vector, and hence that it should be modelled as I(1). The Kalman Filter will ensure that the resulting residual from the model is stationary, ie the markup estimate will cointegrate with the price to labour cost ratio and labour productivity.

The number of lags for the cointegration test was determined by the serial correlation criterion. Conditioning on a filtered measure of capacity utilisation – UK GDP minus a Hodrick-Prescott trend (Chart 5) – four lags were sufficient for an unrestricted VAR of the price to labour cost ratio and productivity. But tests did not reveal evidence of cointegration at the 10% significance level.(13) And indeed, even using a different number of lags there was no significant evidence of

(12) I use a measure of the capital stock excluding dwellings, based on Oulton and Srinivasan (2003). More detail on the construction of the user cost series can be found in the data appendix of Ellis and Price (2004). That paper also investigates the role of investment in adjustment to equilibrium, which is not examined here.
(13) I included a time trend to proxy for technical progress, but the result held even when this was dropped.

cointegration (Table A). That suggests modelling the markup as I(1) in-sample, rather than as a stationary process.[14]

Chart 5
A filtered measure of capacity utilisation

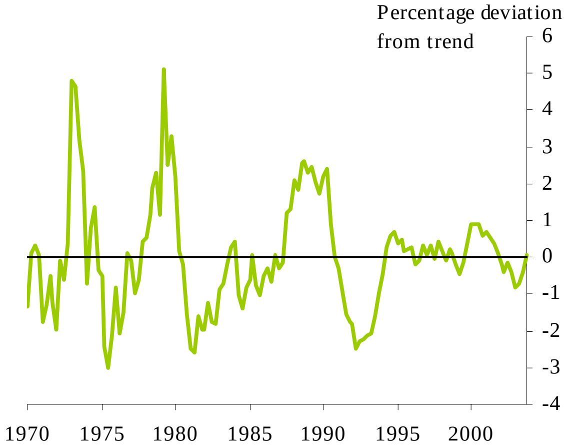


Table A
Cointegration test results

Number of lags included	Null hypothesis: no cointegration	
	Trace statistic	Max-eigen statistic
1	12.4	11.3
2	10.2	9.9
3	12.9	12.4
4[(a)]	13.5	13.2
5	15.3	14.8
6	13.6	11.4

* indicates rejection of null hypothesis at the 10% significance level (null is never rejected in table).
(a) Number of lags where VAR is free of serial correlation.

3.3 Results from standard approaches and other work

Before proceeding with the state-space estimation, a quick re-examination of two standard approaches may be useful at this point to compare and contrast with the state-space results discussed later. Both of these are based solely on the labour-demand equation (**7**), to match how markups are typically estimated.

The first of these is the usual cointegrating framework of estimating a vector error correction mechanism (VECM). Using four lags and the Hodrick-Prescott filtered measure of capacity utilisation, the VECM approach suggested an elasticity of substitution of around 1.2.[15]

A second approach for estimating the elasticity of substitution is dynamic ordinary least squares (DOLS). This is a simple method that yields consistent estimates of the long-run cointegrating parameters, although it is uninformative about the dynamics and adjustment. Using this approach yielded an estimate of the elasticity of substitution of around 1.1.

Other research suggests that the elasticity of substitution in the United Kingdom may be somewhat lower. For example, Barrell and Pain (1997) report an estimate of 0.48 for the private sector, and Ellis and Price (2004) an estimate of 0.44, while NIGEM incorporates a slightly higher estimate of 0.66 (see NIESR (2002)). Thomas (1997) suggests an elasticity of around 0.1-0.2, lower than all of these. Finally, Harrison *et al* (2005) use a calibrated elasticity of substitution of 0.32 in the recently developed Bank of England Quarterly Model (BEQM).

---

[14] Technical progress could also be stochastic, which could potentially influence this result. This is addressed later.
[15] Note however that there was no evidence of cointegration at standard significance levels (see Section 3.2).

3.4 State-space results using a single factor-demand relationship

Having estimated the elasticity using some standard approaches, I now proceed with estimation of the state-space model. The results presented below assume the simple random walk process for the markup in (**14**). Experiments with richer time-series processes found that the extra coefficients were generally insignificant.(16) A random walk appears to offer enough richness for the data. I experimented with different noise to signal ratios, but the impact on the parameter estimates was small. Similarly, different starting values did not change the maximum likelihood estimates.(17) The results presented throughout this paper appear to be the global maxima that maximise the likelihood function, rather than local maxima.

The estimated elasticity of substitution from the labour-demand model was around 0.10, much lower than the estimates from conventional techniques and somewhat lower than the majority of evidence from other research discussed earlier. In addition, the estimate of the markup from the state-space model was broadly flat over the sample (Chart 6). The smoothness of the measures is in marked contrast to the Cobb-Douglas measure of the markup, the inverse of the labour share, but that reflects the imposed noise to signal ratio. The key observation is that the estimated model suggests that markups have not changed over the past 30 years. This result would be consistent with the standard VECM and DOLS approaches discussed earlier, which impose a constant markup by assumption. But the state-space estimate of the elasticity is very different from that found using those standard approaches.

Chart 6
State-space estimate of the markup from labour-demand equation

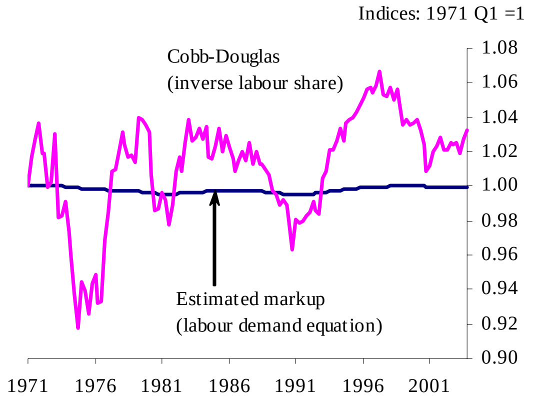
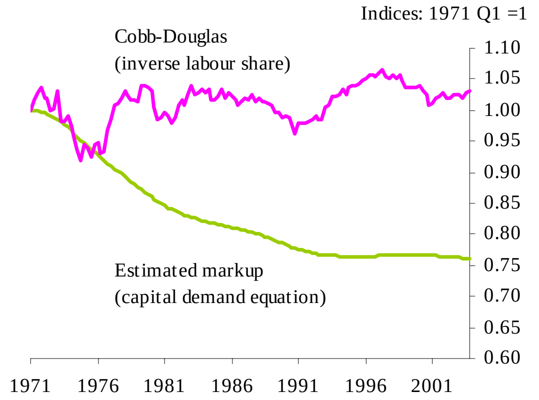


Chart 7
State-space estimate of the markup from capital-demand equation



Of course, if these results are correct then they should be replicable using the capital-demand equation (**9**). But when this was tried, the results were somewhat different. In particular, while the estimated elasticity was only a little higher at 0.23, the markup estimate fell over the sample, rather than being broadly flat (Chart 7).

(16) An example of one of these models is reported in Appendix 2.
(17) Starting values for all parameters were zero, and the Berndt, Hall, Hall, and Hausman algorithm was used to maximise the likelihood function. The reported results use a noise to signal ratio of eight: results from the key models in this note are available in Appendix 2, and other results are available on request. All results were checked to ensure they represented global maxima in the likelihood function by grid-searching across starting values.

Table B
Results from approaches with single demand equations

Model specification	Estimated elasticity of substitution	Percentage change in estimated markup
Labour demand equation	0.10	-0.1%
Capital demand equation	0.23	-23.8%

The different results from the two single-equation approaches (Table B) raises concerns about the use of a time trend to proxy technical progress in the labour-demand equation: this is discussed later. It also suggests that modelling the individual FOCs independently may be ignoring the cross-equation restriction that is needed to correctly identify the model. The next section explores this avenue.

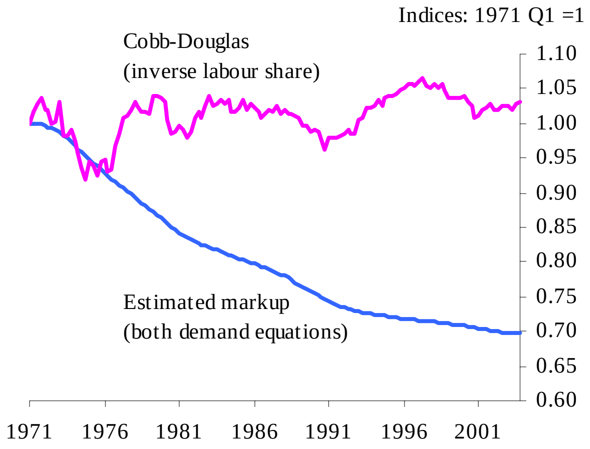
3.5 State-space results using two factor-demand relationships

When the labour and capital-demand equations are estimated separately, the resulting estimates of the markup are very different. Using both of the factor-demand equations together (**(7)** and **(9)**), we can exploit the cross-equation restriction on the markup. The state-space model can now be written as:[18]

$$
\begin{aligned}
y_t - n_t &= \beta_1\mu_t - \beta_1(p_t - w_t) + (1-\beta_1)\beta_2 t \\
y_t - k_t &= \beta_1\mu_t + \beta_1 r_t + \beta_0 \\
\mu_t &= \mu_{t-1} + \lambda_t
\end{aligned}
\qquad \textbf{(17)}
$$

As before, restrictions were placed on the disturbance terms. There are now three of these – one in the state equation, and one in each of the signal equations. Once again, I experimented with different restrictions, but the key results were broadly unaffected.

Chart 8
Markup estimate from dual factor-demand equations



[18] The disturbance terms on the signal equations have been dropped for simplicity.

This time the elasticity of substitution estimate was 0.21, close to the estimate from the capital equation. The resulting markup estimate was also closer to that from the capital-demand equation (Chart 8): over the whole sample it falls by around 30%. However, the cross-equation restriction on the model – that the elasticity parameter was the same in both the labour and capital equations – was rejected.

These results are interesting. But there is still room for improvement. In particular, technical progress is unlikely to be a time trend. In fact, Bean (1994) notes that using a time trend can result in identification problems for factor-demand equations. The next section examines what happens when technical progress is modelled as a stochastic process, rather than a deterministic trend.

3.6 The treatment of technical progress

The estimated elasticity of substitution from the dual factor-demand model is in the ball-park of the single-equation estimates, although somewhat lower than in the majority of other research. But the treatment of technical progress is important. The use of a time trend to proxy technical progress is not unusual, but neither is it ideal. It is more likely that technical progress is stochastic, and should be modelled as such. Estimating this at the same time as the unobserved markup requires a more complex model with extra restrictions.

One temptation would be to combine the two factor-demand relationships. While both (**7**) and (**9**) contain the markup, a linear combination of them does not. Or in other words, by subtracting (**9**) from (**7**) we have:

$$k - n = \sigma(w - p) - \sigma r + (1-\sigma)a + \sigma[\ln(\alpha) - \ln(1-\alpha)] \qquad \textbf{(18)}$$

This could now be estimated as a model with a single state variable (technical progress).

However, this approach throws away information: in particular, the model cannot identify exactly how the state variable affects each factor-demand equation.[19] In order to maintain its structure, we must estimate a version of the model with two state variables (the markup and TP) and two FOCs (**19**). Technical progress was modelled as a random walk with drift, with α as the drift term, based on the prior that it should tend to rise over time. However, the drift coefficient was freely estimated.

$$\begin{aligned}
y_t - n_t &= \beta_1\mu_t - \beta_1(p_t - w_t) + (1-\beta_1)\eta_t \\
y_t - k_t &= \beta_1\mu_t + \beta_1 r_t + \beta_0 \\
\mu_t &= \mu_{t-1} + \lambda_t \\
\eta_t &= \eta_{t-1} + \alpha + \phi_t
\end{aligned} \qquad \textbf{(19)}$$

Now there are separate disturbance terms, λ and Φ, for each of the state-space processes: as before, they were assumed to have independent normal distributions.

[19] The state variable in (**18**) might then be interpreted as a linear combination of two stochastic trends, one from each factor-demand equation.

When this model was estimated the elasticity of substitution was 0.22, broadly unchanged from the previous result but still significantly different from Cobb-Douglas. This time, the cross-equation restriction on the model was accepted: so the treatment of technical progress affects this test. Furthermore, the drift term for TP was significant, and the estimated series itself does exhibit significant variation (Chart 9): this demonstrates that the time-trend assumption from the previous section is inadequate. In fact, the TP series appears to be procyclical. Over the sample as a whole, technical progress grows by a little over 2% a year. That is very close to labour productivity growth over the sample:[20] so in a steady-state out-of-sample forecast (recall (**12**)), the model would return a constant markup.

Chart 9
Technical progress estimate

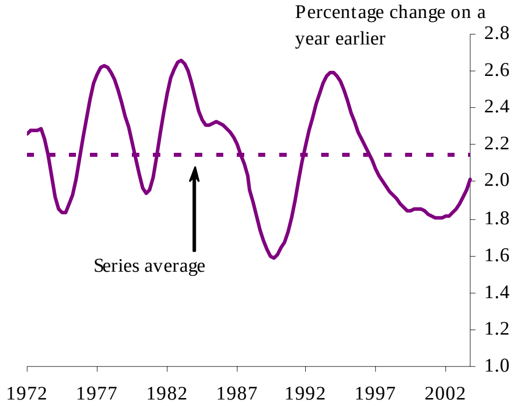
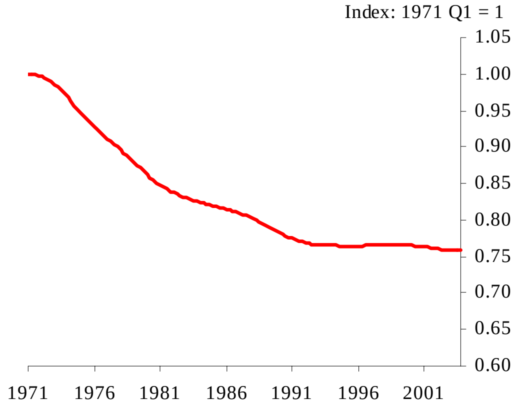


Chart 10
Markup estimate



However, the resulting in-sample estimate of the markup is not constant. In fact, it falls by 24% over the sample as a whole (Chart 10), although most of the fall occurs prior to the 1990s. Such a fall in the markup is notable, but the profile is not explosive. This suggests that, in practice, the theoretical concerns about the potential explosiveness of the state-space estimates (discussed in Section 2.4) were unwarranted. In practice, this estimate of the markup is very similar to that found from the capital-demand equation. On closer inspection this is unsurprising – the markup is the only state variable in that expression, so discrepancies between the capital-output ratio and the user cost are likely to be attributed to it.

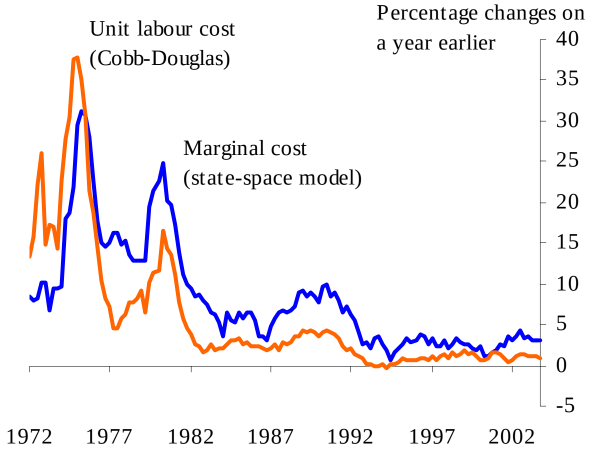
It is worth noting that the previous estimates of the elasticity discussed earlier normally assume constant markups. Given the falling markup returned from this model, a lower elasticity ensures that the model is consistent with the data – so in some sense the difference between these results and previous elasticity estimates is to be expected. This demonstrates the fact that the usual ‘fixed markup’ assumption will affect the elasticity estimate.

Should the markup fall? By the accounting identity (**12**), it should if labour productivity is growing slower than technical progress. That is precisely what the state-space model finds.[21] Of course, while the estimated markup falls, the price series is unaffected. So that implies that marginal cost grows faster in the state-space model than under the assumption of Cobb-Douglas

(20) Indeed, the restriction that the estimated drift term was equal to its steady-state calibration was easily accepted.
(21) Note that the model does not establish causality – it merely exhibits consistent co-movement across variables.

technology (Chart 11). Or in other words, unit labour costs are not a particularly good guide to marginal cost. That also implies that the labour share may not be a good guide to the markup, despite it being used to proxy the markup in a wide variety of empirical work (for example in Gagnon and Khan (2005)).[22]

Chart 11
Cost measures



So the model returns a relatively low elasticity of substitution, a stochastic estimate of technical progress, and a falling markup. These results are in contrast to the behaviour that is often imposed or assumed upon these variables: namely that technology is Cobb-Douglas, technical progress follows a time trend, and the markup is fixed.

4 Testing the estimation methods: simulation results

So far, using a state-space model to estimate the markup has yielded interesting results. But the results from different specifications have sometimes been markedly different, as shown in Table C. In particular, the estimated markup from the labour-demand model (**14**) was flat, while the estimate from the capital-demand (**9**) and dual-demand (**19**) models fell by around a quarter.

Table C
Results from state-space approaches

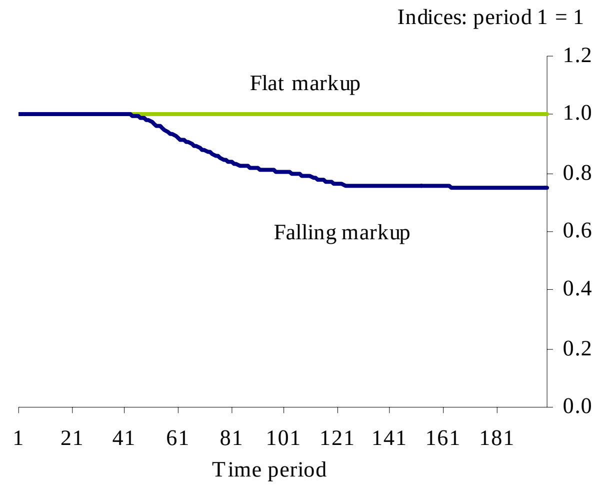
Model specification	Estimated elasticity of substitution	Percentage change in estimated markup
Labour demand equation	0.10	-0.1%
Capital demand equation	0.23	-23.8%
Both demand equations *(stochastic TP)*	0.22	-24.2%

(22) I am very grateful to the two anonymous referees for emphasising this point.

How should these different results be interpreted? If our belief is that the markup is flat, we may want to believe the labour-demand results, despite the low elasticity estimate. But in fact, that model could be the most misleading. Bean (1994) notes that using a time trend can result in identification problems in labour-demand equations – in particular, the time trend can act to 'mop up' any variation in the model. This could also bias down the estimate of the elasticity.

So how can we identify which form of the model is the best – or which gives the right answer? One avenue is to use Monte Carlo simulations. I generated two sets of data, based on an underlying data generating process (DGP) with two factor-demand equations, a fixed elasticity of substitution (0.45) and stochastic technical progress. But the underlying markup variable in the model was different in the two data sets: in one it was constant, but in the other it fell by around a quarter (Chart 12).[23] These markup variables were assumed to be unobservable, as they are in the real world. The observable data variables – $p\text{-}w$, $y\text{-}n$, $y\text{-}k$ and r – were generated based on UK data calibrations.

Chart 12
Assumed profiles for the simulated markup



Using the generated data, I then estimated the three different models: a labour-demand model (**14**), a capital-demand model (**9**) and a dual factor-demand model (**19**).[24] The estimation results from the simulations would reveal which model returns the most accurate results.

Table D shows the estimation results when the underlying markup was flat, based on 10,000 simulations. Note that the markup was identically flat (as in Chart 12) in <u>each</u> of the simulations – rather, the random component was restricted to the observable variables.

All three estimated models returned markup estimates that – on average – were broadly flat. Yet the distribution of the estimated markup change from the labour-demand model is significantly skewed, despite the median being close to zero. The labour model also returns a median elasticity estimate that is far too low. So these simulation results appear to validate the

[23] The profile of the falling markup was determined arbitrarily.
[24] The dual factor-demand model imposed the steady-state drift on technical progress, based on the generated labour productivity series. The signal-noise ratios in all the models were set by calibrating the noise variance against UK data (the calibration was identical in all of the models).

identification concerns noted in Bean (1994). In contrast, the capital and dual-demand models perform well.

Table D
Simulation results with a flat markup

	Elasticity of substitution (median estimate)	5th percentile	95th percentile	Change in markup (median estimate)	5th percentile	95th percentile
True model (underlying DGP)	0.450	n.a.	n.a.	0.0	n.a.	n.a.
State-space models:						
just labour demand	0.014	0.001	0.557	-0.6	-25.2	1.0
just capital demand	0.450	0.418	0.478	0.0	-12.3	9.6
both factor demands	0.450	0.432	0.468	0.0	-5.9	5.4

(a) Approximated by log differences.

What about when the true markup falls over time? Table E presents results from another 10,000 simulations. Once again, on average the labour-demand model returns a broadly flat markup – but now this is simply wrong. Similarly, the average elasticity is once again too low and the distribution of estimates massively skewed.

The capital-demand and dual-demand models perform much better than the labour-demand model when the markup falls, both in terms of the estimated elasticity and the estimated fall in the markup. The simulations suggest that if anything the two models appear to understate the extent of the fall in the latter, and overestimate the elasticity slightly. But the true coefficients are well within normal confidence intervals and are much closer to the underlying data than those from the labour-demand model. Both the capital and dual-demand models also provided a good steer when the markup was flat (Table D). The dual-demand model is more efficient than the capital-demand model for estimating both the size of the change in markup and the elasticity of substitution: the range between the 5th and 95th percentiles is smaller.

Table E
Simulation results with a falling markup

	Elasticity of substitution (median estimate)	5th percentile	95th percentile	Change in markup[a] (median estimate)	5th percentile	95th percentile
True model (underlying DGP)	0.450	n.a.	n.a.	-28.8	n.a.	n.a.
State-space models:						
just labour demand	0.013	0.001	0.624	-1.1	-21.7	1.8
just capital demand	0.464	0.411	0.507	-23.0	-46.3	-3.6
both factor demands	0.467	0.433	0.495	-21.6	-36.1	-9.7

(a) Approximated by log differences.

These simulation results show that just relying on a labour-demand model can yield very misleading results. And this is in a model structure where the markup is allowed to vary over time, let alone models where the markup is assumed to be fixed (like the conventional techniques discussed in Section 3.3). The dual-demand model performs the best in terms of efficiency. So in proceeding, I will focus on that model. However, while the simulation results suggest it is the best of the three models, there could still be shortcomings in my current approach. The next section examines some refinements to the model that address these shortcomings.

5 Refinements to the model

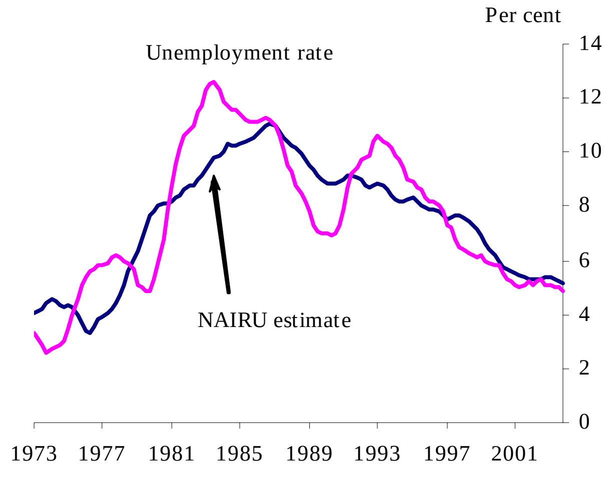
So far, this paper has focused on the long-run relationships that arise from a firm's profit-maximising behaviour. There are a number of ways in which the model could be refined, or checked for consistency. A simple example is varying the distribution parameter (α) in the production function, rather than holding it constant in estimation.[25] But the impact on the results was minimal – because it appears (with opposite signs) in both the capital and labour-demand equations.

However, there are other refinements to the model that can be addressed: these include labour market disequilibrium, capital services, and capital adjustment costs.

5.1 Labour market disequilibrium

In the benchmark model (**19**), we focus solely on the long-run relationships. But this may be unsatisfactory, particularly in the case of the labour market. When firms are driven off their factor-demand equations, any adjustment back to equilibrium via the labour market can take a very long time (see Friedman (1968)). So not allowing for disequilibrium in the labour market – any pressure on wages and prices not captured by the factor-demand equations – may bias the results.

Chart 13
An estimate of the NAIRU[(a)]



(a) Based on Greenslade *et al* (2003).

The obvious solution is to include a simple gauge of labour market disequilibrium from previous work. In particular, the deviation of unemployment from an updated estimate of the NAIRU in

---

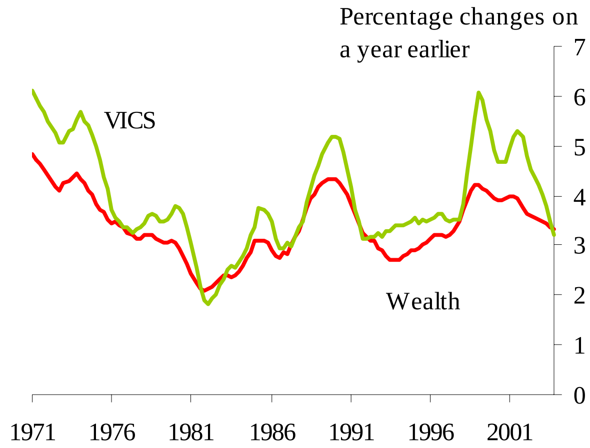
[(25)] NIGEM calibrates α at 0.14 for the United Kingdom (see NIESR (2002)), while Harrison *et al* (2005) choose a value of 0.31.

Greenslade *et al* (2003, see Chart 13) was included in the model, to allow for these effects.[26] But the results from the model were broadly unaffected.

5.2 Capital services

A second refinement to the model could be to use a measure of the flow of productive services from capital, rather than a wealth measure of the asset stock. The former is conceptually superior (see Oulton (2001)), and volume indices of capital services (VICS) have been constructed by Bank of England staff in previous work. A measure of capital services is shown in Chart 14, alongside the wealth measure used previously – the VICS measure has tended to grow at a faster rate, and been more (pro)cyclical than the wealth measure. VICS weights asset types by rental prices, rather than asset prices, and more weight is given to assets such as computers, and less to assets like buildings.[27]

Chart 14
Measures of capital



To implement the VICS measure, we also need to use a different proxy for the 'depreciation' rate, based on the shares of assets in VICS at base period prices, as opposed to shares in wealth. In terms of capital accumulation, the relevant investment series in a VICS world is also one which weights assets by rental prices. But accounting for these, it is relatively simple to use the services measure of capital, rather than the wealth one.

When VICS data was employed instead of the wealth measure, the estimated fall in the whole-economy markup was somewhat larger, at 37%. And the estimated elasticity was higher at 0.29. These differences in results between the wealth and services capital models is not surprising, given the different profiles for capital (Chart 14).

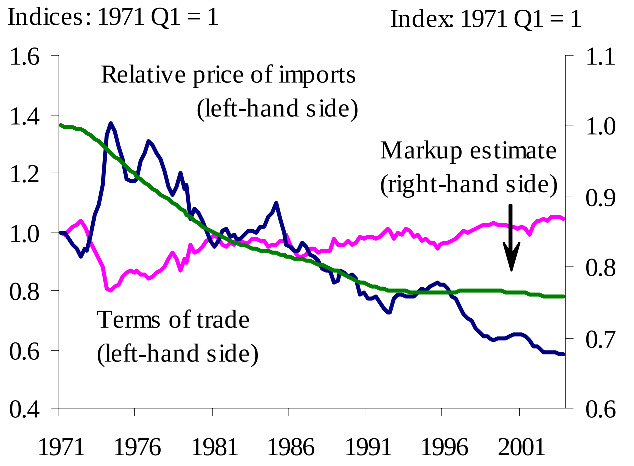
[26] Chart 13 shows a NAIRU estimate based on the consumption deflator. Using estimates based on RPIX and the AEI had no significant impact: see Greenslade *et al* (2003) for more details of these different estimates.
[27] For more information see Oulton and Srinivasan (2003). Note that it is the level of VICS, not the change, that appears in the production function, in an analogous manner to the wealth measure.

5.3 Open-economy concerns and the markup

The estimated fall in the markup appears to be robust to different model specifications. But is there anything we can relate this fall in the markup to? Given that the United Kingdom is a small open economy, competitive pressures from abroad could influence UK markups. So we may observe a correlation between some gauge of those competitive pressures and our estimated markup.

As mentioned earlier, some measure of import prices is sometimes included in factor-demand equations to attempt to capture these competitive pressures. Chart 15 shows the markup estimate alongside the UK terms of trade and the price of UK imports, relative to the whole-economy GDP deflator.

Chart 15
The estimated markup and foreign price variables



At first glance, there does appear to be a relationship between the relative price of imports and the markup. But there are important differences. In particular, the markup is broadly flat since 1990, despite the sharp appreciation in the sterling effective exchange rate in 1996 and 1997, which is evident in the relative price of imports.

Also, it may be more appropriate to consider the terms of trade, rather than the relative price of imports. The terms of trade reflects the price UK exporters charge in foreign markets, compared to the price foreign exporters charge for their products when sold in the United Kingdom. The terms of trade have risen gently over the past 30 years, implying that UK exporters have been able to raise their prices, relative to foreign exporters to the United Kingdom. That is consistent with greater trade specialisation, or other factors such as an expected improvement in productivity: see Dury *et al* (2003). But it is more difficult to tell a story about foreign competition squeezing markups when UK producers' prices are faring better in export markets – which may be a more competitive environment than the domestic economy – than the prices of foreign exporters to the United Kingdom.

This highlights a danger with simply including a foreign price term in factor-demand equations to 'capture' changes in the markup. In practice, it is of course likely that foreign prices do affect domestic ones. But the precise mechanism by which foreign prices affect UK prices – and in particular whether this is via the markup – can be hard to identify (see Ellis and Price (2003)). One strength of the state-space approach is that it is less restrictive than simply including such defined variables in a regression.

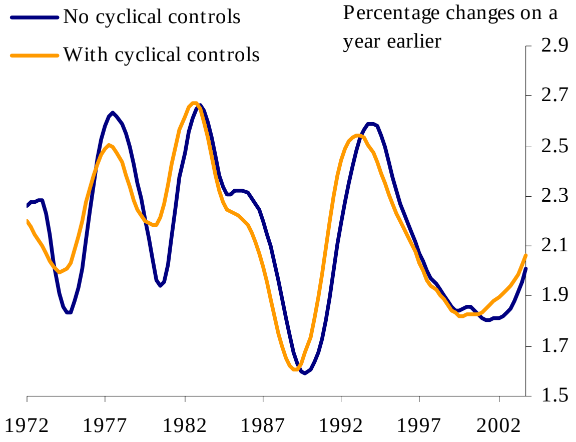
5.4 Cyclical variation and capital adjustment costs

As previously noted, the baseline model ignores any process of adjustment to the long run, or any stationary process that may end up in the signal disturbances. One sign of this could be the very pro-cyclical growth in the estimate of technical progress (recall Chart 9). To account for this, I tried including a measure of capacity utilisation (and its lags) in the model. If the model results are to be trusted, it should not have a big impact on the key variables.

There is no long-run survey measure of whole-economy capacity utilisation, but we can use the previous approximation of the difference between output data and a Hodrick-Prescott (HP) filtered version of the same series (Chart 5). Although this will suffer from end-point problems, it may offer a good approximation to other capacity utilisation measures over the sample as a whole, such as those derived from a production function.

Including the utilisation measure can also address another concern. While labour market disequilibria do not appear to have a big impact on the main results, perhaps of more concern is the presence of capital adjustment costs, which may affect firms' investment decisions. But, in order to incorporate these explicitly, estimation becomes very complex.[28] If investment and other capital adjustment effects – ie the dynamic adjustment to long-run equilibrium – vary with the cycle, then including the cyclical utilisation measure may control for some of that adjustment.

Chart 16
Estimates of technical progress



[28] Depending on the form of adjustment cost, the first-order condition can be very complex and non-linear. Such adjustment costs are typically modelled in a GMM framework; see Groth (2005).

When the filtered capacity utilisation term was included in the model, the estimate of TP was broadly unchanged (Chart 16). The same was also true for the estimated elasticity of substitution and markup series. So cyclical factors do not appear to be affecting the key results.

An alternative method of investigating the impact of adjustment costs would be to estimate the model over a far longer period of time, where such dynamic factors are likely to be less important. This will be discussed later on. In the next section, I examine the private sector.

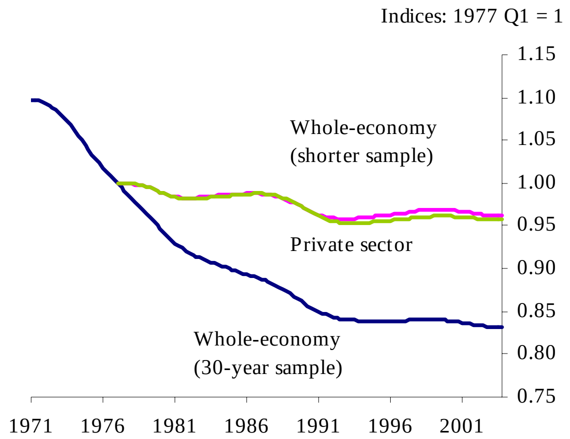
6 The private sector

So far I have pursued my investigation using whole-economy data. But it may be more appropriate to look at private sector data from an inflation-targeting perspective. That is because the vast majority of goods and services in the consumer prices index (CPI) are produced by the private sector. Analysis of the private sector plays an important role in the Bank's forecasting process – for example, BEQM uses a private sector production function, rather than a whole-economy one (see Harrison *et al* (2005)).

6.1 Estimating the private sector model

Unfortunately the private sector output data (as used in Harrison *et al* (2005)) are only available from the late 1970s. But, using this shorter sample, the model can be estimated in the same way as before. The resulting estimate of the elasticity was a little lower at 0.20. But the estimated markup was now much flatter over the sample, in marked contrast to the whole-economy results. The private sector estimate falls by around 5% since 1977, compared to around 17% using whole-economy data.[29]

Chart 17
Markup estimates



However, part of this discrepancy appears to reflect short-sample bias: when the whole-economy model was estimated over the shorter sample (corresponding to private sector data), the markup

[29] This result was robust to using whole-economy starting values in estimation.

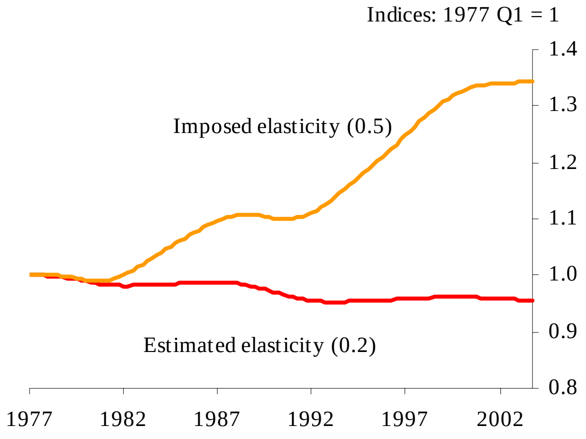
also fell by the same amount, in contrast to the earlier results (Chart 17). In part, this is because a substantial part of the fall in the markup happens in the early 1970s, before the private sector sample starts.

When the private sector model was re-estimated using a VICS measure for the private sector, the estimated fall in the markup was around 7%, a little larger than the estimated fall when the wealth measure of capital was employed. The estimated elasticity was unchanged at 0.20. At first glance this suggests that the distinction between wealth and services measures of capital does not appear to be a concern for the private sector model. But once again this may be affected by short-sample problems: similar results were retrieved when the whole-economy model was estimated over the shorter sample, in contrast to the longer-sample results.

6.2 Imposing the elasticity

While the fall in the markup over the shorter sample is much smaller, the estimated elasticity of substitution broadly corresponds to previous results in this paper. An alternative to estimating the elasticity would be to impose a calibrated value; indeed, one attractive feature of the model is that such judgements are easy to implement. Chart 18 shows the implied markup from imposing an elasticity of 0.5, broadly in line with results from other research (see Section 3.3): the markup now rises sharply over time. And even when steady-state behaviour on the markup is imposed by calibrating the drift term in technical progress, the rise is still evident.[30] So, overall, the model finds that an elasticity of 0.5 – in the region of many previous estimates – is not consistent with a constant markup.

Chart 18
Estimated private sector markups



It is possible that the model results are affected by the potential errors discussed previously, such as labour market disequilibrium or capital adjustment costs. But when the previous measure of labour market disequilibrium (Chart 13) was included in the model, or when public sector employment was taken as given and a private sector unemployment gap was calculated from the

[30] In practice, the estimated drift term in the unrestricted model was very close to its calibrated steady-state value.

aggregate, the effect on the results was small. Similarly, controlling for cyclical factors did not have a big impact.

However, the concerns about capturing the impact of adjustment costs in a short sample (see Section 5.4) are even more applicable to the shorter private-sector sample. One way of investigating this would be to estimate the model over a far longer time period, where such dynamic factors are likely to be less important. Returning to whole-economy data, the next section explores this line of enquiry.

7 Results for the UK economy: the past 150 years

As we are focusing solely on the long-run relationships that arise from a firm's profit-maximising behaviour, a natural avenue of enquiry would be to use a long run of data. For instance, there may be concerns about small-sample bias affecting the previous results – a 30-year sample may not be enough for the long-run relationships to dominate dynamic effects. To overcome this problem, I re-estimated the model on annual data running back to the 19th century.

7.1 The long-run data

The long-run data set was supplied by Charles Bean, and is based mainly on Mitchell (1988). In general there is a bigger health warning over historical data, given the improvements in sampling and methodology over the past 100 years or so. Based on inspection, the historical data do not appear to be implausible. One form of cross-check was enabled by the overlap between the historical data set and the latest vintages of data published by the Office for National Statistics (ONS). For most variables, such as GDP, investment and wages, the difference in growth rates between the historical and current data were very small. In that instance the growth rates of the historical data were spliced backwards onto the current ONS data. One exception was the relative price of investment, which is part of the user cost of capital. So the historical relative price series was adjusted for the discrepancy between it and the National Accounts data (see Appendix 1).

Chart 19
Real wages

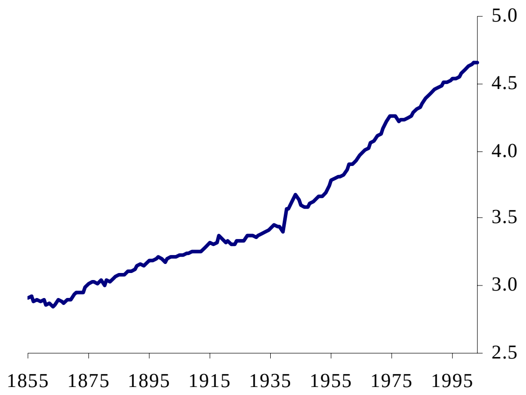


Chart 20
Output per worker

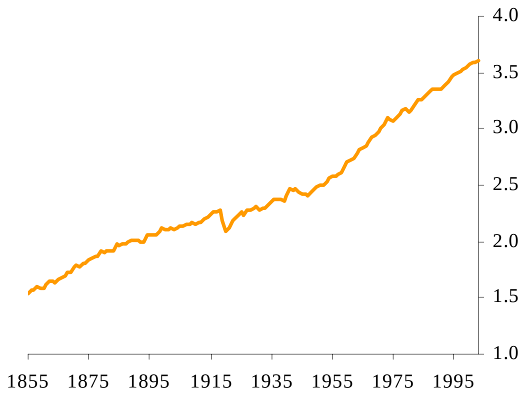


Chart 21
Capital-output ratio

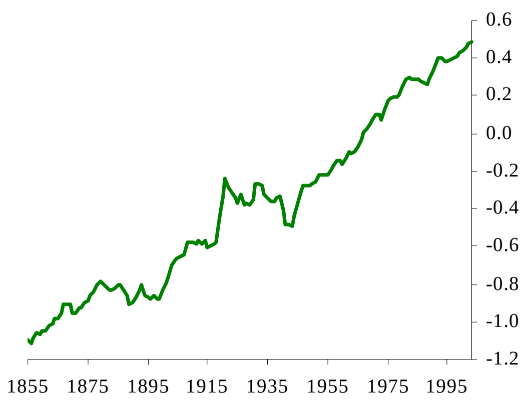
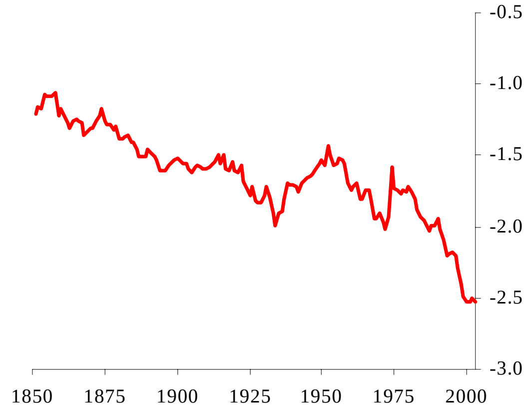


Chart 22
Real user cost of capital

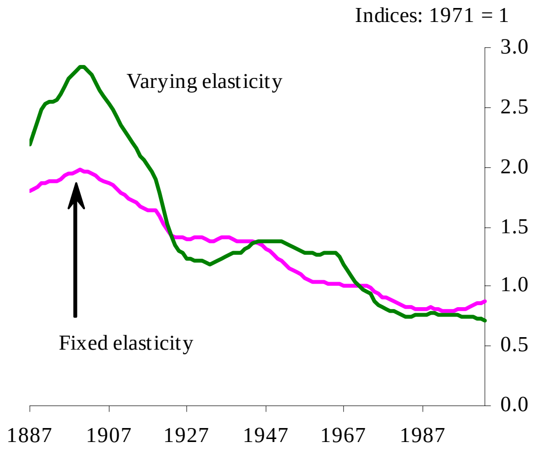


The key variables from the long-run data set are shown in Charts 19 to 22 (all charts have log scales). Unsurprisingly real wages and output per worker rise over time. So does the capital-output ratio, which would suggest that – unless markups have behaved very strangely – the user cost of capital has fallen over time. Chart 22 shows the user cost of capital, which decreases over time due to the falling relative price of investment.

7.2 Estimating the model

Using these long-run data, I estimated the dual factor-demand model with stochastic technical progress (**19**). The retrieved estimate of the elasticity was 0.76, much higher than the short-sample result. But imposing a single elasticity for the 150-year sample as a whole may not be appropriate – in such a long sample, it is far more likely that the elasticity of substitution between capital and labour may have changed; in shorter samples (eg around 30 years) a constant elasticity is likely to be less problematic.[(31)] To allow for this, I also estimated a model where the elasticity could evolve over time.[(32)]

Chart 23
Markup estimates



[(31)] I am grateful to Rich Barwell and Charlotta Groth for this observation.
[(32)] This was done by using rolling estimations to give a path for the elasticity.

When I pursued this approach, at the end of the sample the estimated elasticity was around 0.3, not too far from the short-sample results. The profile of the resulting markup estimate was unsurprisingly more volatile, but not markedly different (Chart 23). And, since 1971, the estimated markup falls by around 28% using the varying-elasticity approach, broadly consistent with the 25% fall found over the short sample. Interestingly, the variation in the estimated elasticity of substitution from 1971 onwards was fairly limited, suggesting that the fixed-elasticity approach may be appropriate for the short-sample estimation.

There are naturally considerable uncertainties around these long-run results, emanating both from concerns about data consistency and the lack of other long-run evidence to examine. I am not suggesting that the results should be taken as definitive. But they do suggest that their short-sample counterparts are not wildly misleading.

8 Applying the technique to US data

So far, the technique for estimating markups and technology appears to have yielded well-determined results. But how else can we test the technique? One line of enquiry is to examine other countries. In particular, I applied the model to US data for the non-farm business sector.[33]

Chart 24
Estimated markup in the United States

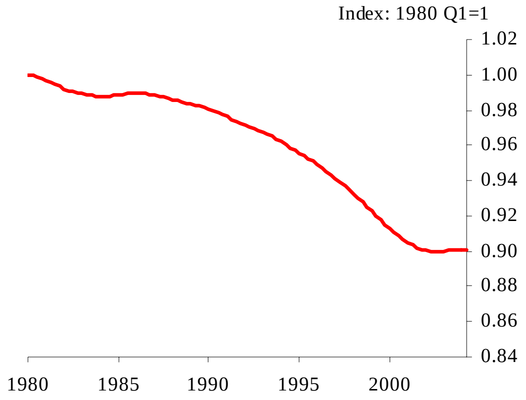
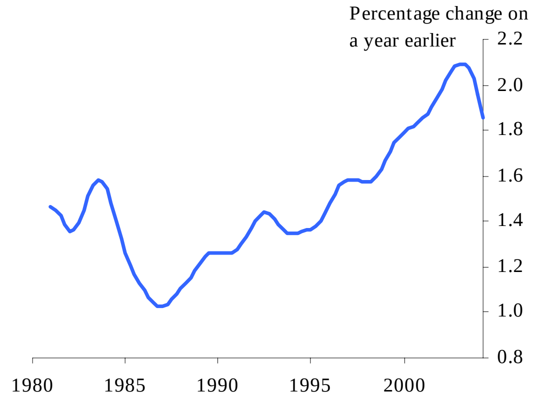


Chart 25
Estimated TP in the United States



The estimated elasticity of substitution for the United States was 0.27, in the ballpark of the UK results. It is a little lower than other US estimates – for example, Chirinko *et al* (2004) estimate an elasticity of 0.4, while Chirinko and Mallick (2005) report an estimate of 0.31 – but not substantially different. Once again the estimated markup was non-stationary, as shown in Chart 24: it falls by around 10% since 1980. It is hard to know how plausible this is, although the timing fits broadly with the launch of the North American Free Trade Association (NAFTA)

[33] The US data were taken from Datastream and NIGEM. Concerns about short-sample bias do apply to the results in this section – but the timing of the markup falls are more centred in the sample than in the UK results. To impose more structure on the results, the drift term in technical progress was calibrated at its steady-state value. The US model was also modified to replace capital with investment in the second factor-demand relationship, as discussed in Ellis and Price (2004).

in 1994.[34] But one check on the results is the estimate of technical progress from the model, shown in Chart 25. Growth in technical progress dips in the mid-1980s, but picks up during the late 1990s – consistent with the documented acceleration in US productivity (see eg Basu *et al* (2003)). This lends support to the model and the estimation technique.

9 Conclusions

Typically, the markup is assumed to be constant when estimating the elasticity of substitution between capital and labour in production. This paper has challenged this presumption and instead modelled the markup explicitly using state-space techniques, by incorporating information from both the labour and capital factor-demand equations. This approach also allows technical progress to be modelled as a stochastic process, rather than as a deterministic trend.

Table F
A summary of results

Model specification	Estimated elasticity of substitution	Percentage change in markup (and sample)
UK whole-economy	0.22	-24% (1971-2003)
Shorter sample	*0.22*	*-4% (1977-2003)*
UK private sector	0.20	-4% (1977-2003)
Using VICS	*0.20*	*-7% (1977-2003)*
UK whole-economy	n.a.	-67% (1887-2003)
US non-farm business sector	0.27	-10% (1980-2003)

This paper has applied the technique to a variety of data. The key results from the main models are summarised in Table F. In all cases, the assumption of Cobb-Douglas technology is rejected, and the estimated elasticity of substitution is generally lower than many – but not all – estimates from previous work. The resulting markup estimates all fall over time, but to differing extents. Simulation results suggest that my preferred model, with both factor-demand equations, is the best at efficiently identifying both the elasticity of substitution and the size of any change in the markup. I have also demonstrated that specifying technical progress as a deterministic process, rather than a stochastic one, can yield misleading results when we focus solely on the labour-demand equation.

Two main results from the paper – a low elasticity and a falling, not fixed, markup – are common to all the estimated models. This illustrates the dangers with conventional estimation techniques, which impose behaviour on at least one of these variables. Modelling the markup as fixed, or imposing Cobb-Douglas technology (or some other inappropriate elasticity), will consequently yield misleading estimation results. Importantly, the results presented here also imply that unit labour costs may not be a good measure of marginal cost.

---

[34] Because of the imposed noise to signal ratio, the model smoothes any changes in the markup.

Appendix 1: The long-run data set

As noted in the main text, the backrun of the long-run data is based on Mitchell (1988). That contains GDP, investment, earnings and price data (both whole-economy and investment deflators) back to 1830. It also contains capital data[35] back to 1850 and employment data back to 1855. All of these series are available on an annual basis until around 1990.

ONS data on these series[36] are available from the 1960s. That means there is an overlap of around 25 years between the two data sets, where the reported series can be compared. For most variables, the difference between the historical and current data (growth rates) was small. So splicing the two series together was uncontroversial. The key exception was the relative price of investment.

The relative price of investment has been falling since around 1980 in the National Accounts (see Ellis and Groth (2003)). But the same is not true in the historical data set. Chart A1 shows the two relative price measures.

Chart A1
Measures of the relative price of investment

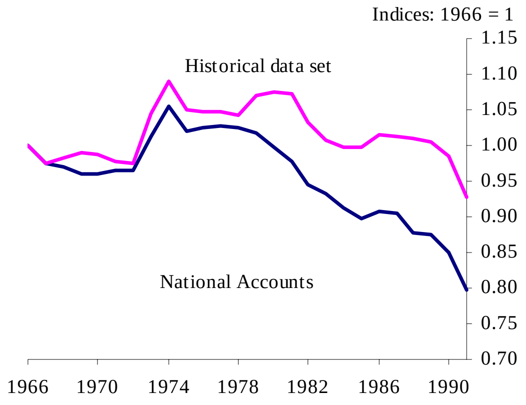
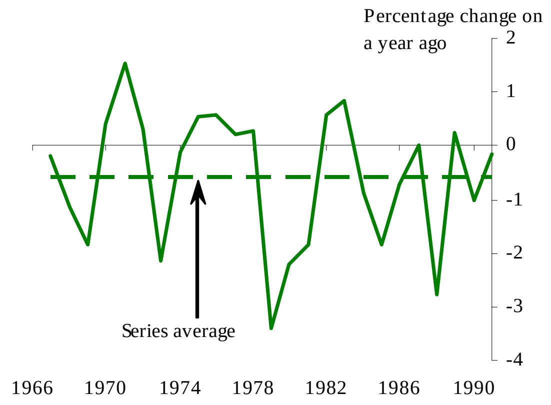


Chart A2
The ratio of the relative price measures



But while the two measures diverge, the ratio of the two appears to be I(1). Chart A2 shows the change in the ratio of the two measures, which appears to be stationary and passes standard unit root tests. On average, the historical measure of the relative price grew 0.6% a year more quickly than the National Accounts measure. This observation presents a possible solution: we can adjust the growth rate of the historical series and splice it on to the National Accounts data.

The relative price of investment is one component of the real user cost of capital.[37] Other components are harder to come by. In the end, I used a simplified measure:

$$RCC = \frac{P_K}{P_Y}(c + \delta)$$

---

[35] This needed interpolating during the Second World War.
[36] The capital stock data is based on Oulton and Srinivasan (2003).
[37] See the data appendix of Ellis and Price (2004) for more detail on the user cost.

where P_K/P_Y denotes the relative price of capital, c the real cost of finance and δ the depreciation rate. An estimate of δ can be backed out from the investment and capital data in the historical data set.[(38)]

For the cost of finance I used a weighted average cost of capital (WACC) measure, as in Ellis and Price (2004).[(39)] The WACC comprises of a cost of equity finance and a cost of debt finance. The former can be calculated back to 1850, using a simple dividend discount model.[(40)] For the latter I backcast the series using macro variables, as in Ellis and Price (2004). A heroic assumption to calculate the WACC was that the weights of the two cost of finance measures were unchanged. Arbitrarily varying these did have some impact on the user cost series, but the impact on the estimated model was relatively small. In any event, it is unclear exactly how these weights would have evolved over time.

---

(38) I experimented with different variants, for example based on different splicing methods, but the impact on the user cost was very small. I also experimented with adding the measures of the expected change in the relative price term to the user cost, but again this only had a small impact.
(39) The WACC is a textbook finance concept – see eg Brealey and Myers (2000). I was discouraged from using a simple *ex-post* real interest rate, as it was very different from the WACC over the available sample.
(40) I used a global series for the dividend yield from 1923. But due to data limitations, it was spliced onto the dividend yield for the Bank of England prior to this.

Appendix 2: Key estimation results

This appendix reports the key state-space results described in the main body of the note. First, Table A1 shows the results from the dual factor-demand model with stochastic technical progress, described in Section 3.6.[41]

Table A1
Whole-economy model results

Sample: 1971 Q1 to 2003 Q4

	Coefficient	Standard error	Z-statistic	P-value
Elasticity of substitution	**0.22**	**0.01**	**18.9**	**0.0**
Constant term in labour equation	-8.2	61015.3	0.0	1.0
Drift term in technical progress (log form)	-5.3	0.3	-18.1	0.0
Model statistics				
Log likelihood	452.7	Akaike info criterion		-6.8
Diffuse priors	2	Schwarz criterion		-6.7

Table A2 presents results when the same model is estimated, but including extra AR terms when in the time-series process for the markup. As discussed in the main text, the extra terms are all insignificant. But the (insignificant) extra AR terms also impact on the estimated elasticity, suggesting the model is misspecified.

Table A2
Results for a whole-economy model with a richer markup process

Sample: 1971 Q1 to 2003 Q4

	Coefficient	Standard error	Z-statistic	P-value
Elasticity of substitution	**-0.08**	**0.02**	**-4.4**	**0.0**
Constant term in labour equation	3.9	13030802.3	0.0	1.0
Drift term in technical progress (log form)	-4.9	0.1	-74.2	0.0
AR terms in markup variable				
AR(2)	0.0	0.0	0.9	0.4
AR(3)	0.0	0.0	0.1	0.9
AR(4)	0.0	0.0	0.0	1.0

[41] Note that all constant terms reflect random starting values for the estimated state variables.

Table A3 presents estimation results for the private sector model, as detailed in Section 6.1.

Table A3
Private sector model results

Sample: 1977 Q1 to 2003 Q4	Coefficient	Standard error	Z-statistic	P-value
Elasticity of substitution	**0.20**	**0.01**	**14.3**	**0.0**
Constant term in labour equation	-12.7	41661.2	0.0	1.0
Drift term in technical progress (log form)	-5.3	0.2	-33.1	0.0
Model statistics				
Log likelihood	441.7	Akaike info criterion		-8.1
Diffuse priors	2	Schwarz criterion		-8.0

Finally, Table A4 shows the results from estimating the same model over a long run of data, as described in Section 7.2 (the fixed-elasticity version).

Table A4
Results from estimating over long-run data

Sample: 1886 to 2003	Coefficient	Standard error	Z-statistic	P-value
Elasticity of substitution	**0.76**	**0.04**	**19.1**	**0.0**
Constant term in labour equation	0.7	17346.8	0.0	1.0
Drift term in technical progress (log form)	-3.3	0.1	-40.9	0.0
Model statistics				
Log likelihood	264.6	Akaike info criterion		-3.5
Diffuse priors	2	Schwarz criterion		-3.4

References

Agoloskoufis, G, Martin, C and Pittis, N (1990), 'Pricing and product market structure in open economies: an empirical test', *CEPR Discussion Paper,* No. 486.

Bakhshi, H and Larsen, J (2001), 'Investment-specific technological progress in the United Kingdom', *Bank of England Working Paper no. 129*.

Bank of England (2000), *Economic models at the Bank of England: September 2000 update*, Bank of England.

Barrell, R and Pain, N (1997), 'Foreign direct investment, technological change, and economic growth within Europe', *Economic Journal*, Vol. 107, pages 1,770-86.

Barro, R and Sala-I-Martin, X (1995), *Economic growth*, McGraw Hill.

Basu, S, Fernald, J, Oulton, N and Srinivasan, S (2003), 'The case of the missing productivity growth: or, does information technology explain why productivity accelerated in the US but not the UK?', *NBER Working Paper no. 10010*.

Bean, C (1994), 'European unemployment: a survey', *Journal of Economic Literature*, Vol. 32, pages 573-619.

Blanchard, O and Fischer, S (1989), *Lectures on macroeconomics*, MIT Press.

Brealey, R and Myers, S (2000), *Principles of corporate finance*, Irwin McGraw-Hill.

Chirinko, R, Fazzari, S and Meyer, A (2004), 'That elusive elasticity: a long-panel approach to estimating the capital-labor substitution elasticity', *CESIFO Working Paper no. 1240*.

Chirinko, R and Mallick, D (2005), 'The substitution elasticity, growth theory, and the low-pass filter panel model', *mimeo*.

Davidson, J (1994), 'Identifying cointegrating regressions by the rank condition', *Oxford Bulletin of Economics and Statistics*, Vol. 56, pages 103-08.

Davidson, J (1998), 'Structural relations, cointegration and identification: some simple results and their application', *Journal of Econometrics*, Vol. 87, pages 87-113.

Dury, K, Piscitelli, L, Sebastia-Barriel, M and Yates, T (2003), 'What caused the rise in the UK terms of trade?', *Bank of England Quarterly Bulletin,* Summer, pages 164-76.

Ellis, C and Groth, C (2003), 'Long-run equilibrium ratios of business investment to output in the United Kingdom', *Bank of England Quarterly Bulletin*, Summer, pages 177-87.

Ellis, C and Price, S (2003), 'The impact of price competitiveness on UK producer price behaviour', *Bank of England Working Paper no. 178*.

Ellis, C and Price, S (2004), 'UK business investment and the user cost of capital', *The Manchester School*, Supplement, Vol. 72, pages 72-93.

Friedman, M (1968), 'The role of monetary policy', *American Economic Review*, Vol. 58, No. 1, pages 1-17.

Gagnon, E and Khan, H (2005), 'New Phillips curve under alternative production technologies for Canada, the United States, and the euro area', *European Economic Review*, Vol. 49, pages 1,571-602.

Greenslade, J, Pierse, G and Saleheen, J (2003), 'A Kalman filter approach to estimating the UK NAIRU', *Bank of England Working Paper no. 179*.

Groth, C (2005), 'Estimating UK capital adjustment costs', *Bank of England Working Paper no. 258*.

Harrison, R, Nikolov, K, Quinn, M, Ramsay, G, Scott, A and Thomas, R (2005), *The Bank of England Quarterly Model*, Bank of England.

Khan, H and Moessner, R (2005), 'Competitiveness, inflation, and monetary policy', *Bank of England Working Paper no. 246*.

Kuttner, K (1994), 'Estimating potential output as a latent variable', *Journal of Business and Economic Statistics*, Vol. 12(3), pages 361-67.

Leith, C and Malley, J (2003), 'A sectoral analysis of price-setting behaviour in US manufacturing', *University of Glasgow Discussion Paper in Economics*, No. 03-07.

Lindsay, C and Doyle, P (2003), 'Experimental consistent time series of historical Labour Force Survey data', *Labour Market Trends*, ONS, September.

Mitchell, B (1988), *British historical statistics*, Cambridge University Press.

NIESR (2002), 'World model manuals', NIESR, London, *mimeo*.

Oulton, N (2001), 'Measuring capital services in the United Kingdom', *Bank of England Quarterly Bulletin*, Autumn, pages 295-309.

Oulton, N and Srinivasan, S (2003), 'Capital stocks, capital services, and depreciation: an integrated framework', *Bank of England Working Paper no. 192*.

Price, S (1992), 'Costs, prices and profitability in UK manufacturing', *Applied Economics*, Vol. 23, pages 839-50.

Smith, P (2000), 'Output price determination and the business cycle', *Economic Modelling*, Vol. 17, pages 49-69.

Solow, R (1999), 'Neoclassical growth theory', in Taylor, J and Woodford, M (eds), *Handbook of macroeconomics: Volume 1A*, North-Holland.

Thomas, R (1997), 'The demand for M4: a sectoral analysis. Part 2 – the corporate sector', *Bank of England Working Paper no. 62*.