Microeconometric Models of Investment and Employment

Stephen Bond
Nuffield College, Oxford and Institute for Fiscal Studies

John Van Reenen
Centre for Economic Performance, London School of Economics and CEPR

December 2003

JEL classification: D92, G31, J23, D21, C50, O33

Keywords: Investment, employment, panel data

Address for correspondence:

Stephen Bond, IFS, 7 Ridgmount Street, London WC1E 7AE, UK

steve.bond@nuffield.ox.ac.uk

Abstract

We survey recent microeconometric research on investment and employment that has used panel data on individual firms or plants. We focus on model specification and econometric estimation issues, but we also review some of the main empirical findings. We discuss advantages and limitations of microeconomic data in this context.

We briefly review the neoclassical theory of the demand for capital and labour, on which most of the econometric models of investment and employment that we consider are based. We pay particular attention to dynamic factor demand models, based on the assumption that there are costs of adjustment, which have played a prominent role especially in the microeconometric literature on investment. With adjustment costs, current choices depend on expectations of future conditions. We discuss the challenges that this raises for econometric model specification, and some of the solutions that have been adopted. We also discuss estimation issues that arise for dynamic factor demand equations in the context of micro panel data for firms or plants.

We then discuss a number of topics that have been the focus of recent microeconometric research on investment and employment. In particular, we review the literatures on investment and financing constraints, relative price effects on investment and employment, investment and uncertainty, investment in research and development (R&D), elasticities of substitution and complementarity between technology, capital and skilled and unskilled labour, and recent work on models with non-convex adjustment costs.

Contents

1. Introduction

This chapter surveys the application of econometric methods to study investment and employment decisions, using microeconomic data at the level of individual firms or plants. We discuss a range of models and issues that have been at the centre of microeconometric research on company factor demand behaviour over the last decade. We do not attempt to review the extensive econometric research on investment and employment that uses more aggregated data at the sectoral or macroeconomic levels. Chirinko (1993a) and Caballero (1999) provide surveys of recent work on aggregate investment, whilst Hamermesh (1993) provides a comprehensive survey of work on aggregate employment.

Microeconomic data offers several important advantages for the study of investment and employment behaviour. First, it allows us to eliminate the impact of aggregation over firms or plants when estimating a particular model. Second, there may be cross-sectional variation in explanatory variables that helps to identify parameters of interest. Perhaps more importantly, the availability of micro data allows the researcher to investigate heterogeneity in behaviour between different types of firms or plants that would simply not be possible with more aggregated data.

On the other hand, microeconomic data sources are often unrepresentative in their coverage (typically with a bias towards larger units), and severely limited in the type of information that they provide. These limitations have shaped the research questions that micro data sets have been widely used to address. A basic reason for estimating models of investment or labour demand is to quantify how the employment of capital or labour inputs responds to changes in factor prices - for example, the response of investment to interest rates or to an investment tax credit, or the response of employment to minimum wage legislation. These questions require elasticities of substitution to be estimated - serious investigation requires a flexible representation of substitution possibilities to be specified, and estimation requires detailed information on a range of factor inputs or cost shares, and the corresponding factor prices. Unfortunately many micro data sources, particularly those obtained from company accounts, provide only crude or partial measures of factor inputs or cost shares, and little or no information on the factor prices faced by individual firms or plants.

Largely for this reason, much of the microeconometric work has focused on estimating a single equation for company investment or company employment, rather than the more ambitious systems of interrelated factor demand or share equations that are often estimated using aggregate or industry level data.[1] Also, where this literature has investigated elasticities of substitution, considerable ingenuity has often been required, either to obtain measurable variation across firms in factor prices, or to specify models which can address some questions of interest without price information at all. Examples of the former include variation in the cost of capital due to a different mix between tax-favoured and unfavoured types of investment, or variation in the effective impact of taxes due to asymmetries between the tax treatment of profits and losses; examples of the latter include work on technology-skill complementarity, which we discuss in section 7.5 below.

Notwithstanding these limitations, there are a range of interesting questions that can be and have been addressed using microeconomic data. We are interested not only in how much investment or employment will respond to a change in factor prices, but also in how quickly. Once we recognise that complete adjustment does not occur immediately, the question of how current investment and employment decisions depend on expectations of future prices and demand conditions becomes important, and controlling for these unobserved expectations presents a particular challenge for econometric modelling. Indeed for some policy questions, such as the evaluation of temporary tax incentives, the characterisation of adjustment

[1] See Berndt (1991) for an excellent introduction to the interrelated factor demand literature.

dynamics is of crucial importance. The recognition that aggregation can distort the underlying dynamic relationship has motivated the use of micro data to study adjustment processes for both capital and labour.[2] Moreover, as we discuss in section 3, models have been developed that allow adjustment parameters to be estimated without necessarily restricting or identifying the long-run elasticities of substitution.

Another major use of micro data has been to test some of the simplifying assumptions that are made in specifying traditional factor demand models. Leading examples include the question of whether firms can finance their investment spending in 'perfect' capital markets, or whether they may face important financing constraints; and the question of whether firms determine employment along a labour demand schedule, or whether employment levels are subject to bargaining with workers. Micro data allows heterogeneity across firms - for example, between small and large firms, or between unionised and non-unionised plants - to be exploited in testing these specifications, and the importance of differences among sub-samples of firms or plants to be investigated. Again these questions can be addressed without fully specifying the nature of the production technology.

The interest in testing hypotheses about the nature of the adjustment process or the environment in which investment and employment decisions are taken has led to a focus on structural models of investment and employment dynamics, in which the optimal evolution of the firm's stock of capital or labour inputs is derived from some underlying theoretical model of the firm, and this is used to obtain an econometric model whose parameters reflect the firm's technology. As we discuss further in section 3.6, reduced form dynamic models of investment or employment generally compound structural adjustment parameters with the process by which current expectations of future demand or prices are formed, which makes

[2]See Nickell (1978, 1986) for a discussion of aggregation biases in the context of dynamic investment and labour demand equations, and Blundell and Stoker (this volume) for a survey of aggregation issues more generally.

it difficult to draw firm conclusions about the nature of the adjustment process or the role of, say, financial variables in an investment equation. Nevertheless, this *relative* disadvantage of reduced form models compared to actual (as distinct from ideal) structural models should not be overstated. As we will emphasise, the most commonly used structural dynamic models of investment and employment are based on extreme simplifying assumptions, and are frequently rejected when subjected to mild empirical testing. Moreover the recent empirical work on models with non-convex adjustment costs, whilst largely descriptive, has cast doubt on many of the structural models developed in the 1980s.

Two further limitations of commonly used microeconomic data sets should be noted at the outset. First, data on publicly traded companies or manufacturing plants already reflect a considerable degree of aggregation over investment decisions in many different types of equipment and structures, and employment decisions over many different types of workers. For large units, there may also be aggregation over inputs used in different lines of business, and for annual data there is the further question of aggregation over time. Although these micro data may be the most appropriate level of aggregation for investigating some questions - such as the relationship between investment and share prices, debt structures, or other aspects of corporate finance and corporate governance - we should recognise that they may still be too aggregated for identifying other parameters of interest. Whilst these concerns are also present in the study of household level data,[3] they are probably more severe in the case of data on large firms or large plants. Conversely, though, we should also recognise that if the object of interest is how aggregate investment or employment responds to some change in wages or prices, then the adjustment that we observe within existing firms or plants may be only one component of the aggregate response. In particular, adjustment which takes

[3] See, for example, Bourguignon and Chiappori (1992) and the discussion in Blundell et al. (this volume)

the form of the entry or exit of more or less capital intensive firms or plants is likely to be missed with commonly used data sets, and for the same reason there may be no simple relationship between aggregate adjustment dynamics and those observed at the micro level.[4] Although there is much we can learn from the study of investment and employment adjustment in microeconomic data sets, we should be cautious in extending these findings to address macroeconomic questions.

In this chapter we concentrate principally on the issues that have been the focus of recent econometric research using firm or plant level data. We present the state of the art as we perceive it and we do not hesitate to point out what we consider to be important weaknesses and omissions in this literature. Today's gaps are tomorrow's opportunities for important progress to be made in research. Whilst we recognise that important advances have been made in the specification and estimation of microeconometric investment and employment models in recent years, this is an area where the development of new data resources is presenting new challenges to traditional approaches, as well as exciting opportunities for richer structural models to be developed.

The chapter is organised as follows. Section 2 sets out the basic neoclassical theory of factor demand, on which most of the econometric models we consider are based. Section 2.1 briefly reviews the static factor demand literature, whilst section 2.2 introduces dynamic factor demand models based on the assumption that changing the level of factor inputs involves adjustment costs. Section 3 illustrates how this approach has been used to derive dynamic econometric investment equations. Sections 3.1 - 3.4 discuss alternative structural models based on strictly convex costs of adjustment, including the popular Q model of investment and the Euler equation approach; section 3.5 discusses more recent work on models with non-convex adjustment costs; and section 3.6 discusses the use of reduced form dynamic models in this context. Section 4 discusses some econometric issues that

[4] See Caballero (1992) and Campbell and Fisher (2000) for further discussion.

arise in the specification and estimation of these dynamic factor demand equations, particularly those of stochastic specification, and estimation using micro panel data for individual firms or plants. Section 5 discusses the sources of such data, and its limitations. Section 6 discusses some topics in the recent empirical literature on investment. Section 6.1 presents some basic empirical findings; section 6.2 discusses the literature on testing for financing constraints; section 6.3 discusses some recent research on taxes and investment; section 6.4 discusses some recent work on uncertainty and investment; and section 6.5 discusses microeconometric models of research and development (R&D) investment. Section 7 discusses some topics in the recent empirical literature on employment. Section 7.1 considers wage elasticities; section 7.2 discusses models of employment determination with union bargaining; section 7.3 discusses models of employment dynamics; section 7.4 discusses whether adjustment costs are important for net changes in the level of employment or for gross flows of hiring and firing; and section 7.5 discusses research on skill-biased technical change. Section 8 presents our summary of the main themes, omissions and opportunities for future research in this area.

2. Theoretical framework

We begin this chapter with a brief exposition of the neoclassical theory of factor demand. The model we consider is simplified in many respects. The firm's objective is to maximise the value of the equity owned by its shareholders, so that a host of interesting corporate control issues are assumed away. These shareholders are assumed to be risk neutral, so that the effects of risk on the firm's required rate of return are not considered. The firm issues no debt and pays no taxes, so that corporate financial policy is not considered. The firm operates in competitive markets and in a world characterised by symmetric information, so that strategic behaviour is not considered, and the firm is able to issue as much new equity as it chooses at an exogenously given required rate of return, determined by the riskless interest rate. Hence internal finance from retained profits and external finance from new share issues are perfect substitutes, and there is separability between the firm's real and financial decisions, as in the Modigliani-Miller (1958, 1961) theorems. It is not our intention to suggest that these omitted considerations are unimportant or uninteresting. We will touch on some of these issues in later sections, but to do full justice to them would take us well beyond the scope of this chapter.

We distinguish between three types of factors of production: capital assets, which are owned by the firm and provide productive services for several time periods; labour inputs, which are hired by the firm each period; and current inputs, which are purchased by the firm but which are fully consumed in contributing to the current period's production. Capital assets, which may include both tangible assets like equipment and structures and intangible assets like knowledge and reputation, are durable, whilst labour and current inputs are not. However a more important distinction is whether the level of these inputs can be costlessly and immediately adjusted in response to new information. We first examine the static case which abstracts from any adjustment costs or delays.

2.1. Static factor demand

It is useful to briefly review static models of the firm's demand for capital and labour, in order to introduce some important concepts and to clarify how the structural dynamic models we consider later generalise this static framework. These static models also form the basis for most reduced form dynamic factor demand equations, which we discuss further in section 3.6.

The basic factor demand model we consider can be characterised by the following optimisation problem for the firm

$$V_t(K_{t-1}) = \left\{ \max_{I_t, L_t, M_t} \Pi_t\left(K_t, L_t, M_t, I_t\right) + \beta_{t+1} E_t\left[V_{t+1}\left(K_t\right)\right] \right\} \qquad (2.1)$$

where V_t is the maximised value of the firm in period t, $\Pi_t(.)$ is the firm's net revenue function in period t, $K_t = \left(K_t^1, ..., K_t^N\right)$ is a vector of N types of capital inputs, $L_t = \left(L_t^1, ..., L_t^R\right)$ is a vector of R types of labour inputs, $M_t = \left(M_t^1, ..., M_t^S\right)$ is a vector of S types of current inputs, $I_t = \left(I_t^1, ..., I_t^N\right)$ is a vector of gross investments in each type of capital, $\beta_{t+1} = (1+\rho_{t+1})^{-1}$ is the firm's discount factor, where ρ_{t+1} is the risk-free rate of interest between period t and period $t+1$, and $E_t[.]$ denotes the expected value conditional on information available in period t, where the expectation is taken over the distribution of future prices and interest rates.

The equation of motion for the capital inputs is

$$K_t^i = \left(1 - \delta^i\right) K_{t-1}^i + I_t^i \text{ for } i = 1, ..., N \tag{2.2}$$

where δ^i is the rate of depreciation for capital of type i, assumed to be exogenous and fixed. Note that gross investment may be positive or negative, so that disinvestment is also assumed to be costless.

In the absence of any adjustment costs, the net revenue function may take the form

$$\Pi_t(K_t, L_t, M_t, I_t) = p_t F(K_t, L_t, M_t) - p_t^K I_t - w_t L_t - p_t^M M_t \tag{2.3}$$

where $F(K_t, L_t, M_t)$ is the production function, p_t is the price of the firm's output, $p_t^K = \left(p_t^{K,1}, ..., p_t^{K,N}\right)$ is a vector of prices for each type of capital goods, $w_t = \left(w_t^1, ..., w_t^R\right)$ is a vector of wage rates for each type of labour and $p_t^M = \left(p_t^{M,1}, ..., p_t^{M,S}\right)$ is a vector of prices for each type of current inputs. Note that capital inputs are assumed to be purchased and owned by the firm, whilst labour inputs are assumed to be hired.[5]

The solution to the optimisation problem (2.1) subject to the constraints (2.2)

[5] The model can of course accommodate capital inputs that are leased; these would be treated in a similar way as labour inputs are treated here.

can be characterised by the first-order conditions

$$-\left(\frac{\partial \Pi_t}{\partial I_t^i}\right) = \lambda_t^i \text{ for } i = 1, ..., N \tag{2.4}$$

$$\lambda_t^i = \left(\frac{\partial \Pi_t}{\partial K_t^i}\right) + \left(1 - \delta^i\right) \beta_{t+1} E_t \left[\lambda_{t+1}^i\right] \text{ for } i = 1, ..., N \tag{2.5}$$

$$\left(\frac{\partial \Pi_t}{\partial L_t^i}\right) = 0 \text{ for } i = 1, ..., R \tag{2.6}$$

$$\left(\frac{\partial \Pi_t}{\partial M_t^i}\right) = 0 \text{ for } i = 1, ..., S \tag{2.7}$$

where $\lambda_t^i = \frac{1}{1-\delta^i}\left(\frac{\partial V_t}{\partial K_{t-1}^i}\right)$ is the shadow value of inheriting one additional unit of capital of type i in period t. Equation (2.4) shows that the cost of acquiring additional units of each type of capital in period t will be equated to their shadow values. Equation (2.5) describes the evolution of these shadow values along the optimal path for the capital stocks, whilst equations (2.6) and (2.7) are standard first-order conditions for the non-durable factors of production, equating the price of these inputs with their marginal revenue products (see equation (2.3)).

For a price-taking firm, we have $-\left(\frac{\partial \Pi_t}{\partial I_t^i}\right) = p_t^{K,i}$ and $\left(\frac{\partial \Pi_t}{\partial K_t^i}\right) = p_t\left(\frac{\partial F}{\partial K_t^i}\right)$. Substituting these expressions into (2.4) and (2.5) respectively, combining these equations to eliminate λ_t^i and $E_t\left[\lambda_{t+1}^i\right]$ from (2.5) and rearranging yields

$$\left(\frac{\partial F}{\partial K_t^i}\right) = \frac{p_t^{K,i}}{p_t}\left(1 - \left(\frac{1-\delta^i}{1+\rho_{t+1}}\right) E_t\left[\frac{p_{t+1}^{K,i}}{p_t^{K,i}}\right]\right) = \frac{r_t^i}{p_t} \text{ for } i = 1, ..., N. \tag{2.8}$$

This shows that if the level of capital inputs can be freely adjusted, the marginal product of capital of type i will be equated in each period with the real user cost of capital $(\frac{r_t^i}{p_t})$ for capital of type i (Jorgenson (1963)). The user cost depends on the relative price of capital goods of type i, the firm's required rate of return, the depreciation rate for capital of type i, and the expected rate of change in the price of capital goods of type i. This is also the equilibrium price at which capital goods of type i could be rented for use in period t in a competitive rental market, so the user cost is also known as the rental price of capital.[6]

[6] The static model in which the firm purchases durable capital inputs is formally equivalent to

2.1.1. Functional forms

To derive useful factor demand equations that can be estimated, we then need to parameterise the static production function $F(K_t, L_t, M_t)$. First, we consider the popular Constant Elasticity of Substitution (CES) functional form (Arrow et al. (1961)). To illustrate this we use a two factor production structure in which there is a single capital good (K_t), a single labour input (L_t), and no other current inputs.[7] Assuming constant returns to scale, this production function has the form

$$Y_t = F(K_t, L_t) = (a_K K_t^{\rho} + a_L L_t^{\rho})^{\frac{1}{\rho}} \tag{2.9}$$

where $\rho = \left(\frac{\sigma-1}{\sigma}\right)$ and σ is the elasticity of substitution between capital and labour. To ensure that the firm's value maximisation problem has a solution in the absence of adjustment costs, we also assume that there is some degree of monopolistic competition and the firm faces a downward sloping demand curve for its output (Y_t) of the isoelastic form

$$p_t = BY_t^{-\frac{1}{\eta^D}} \tag{2.10}$$

where B is a demand shift parameter and $\eta^D > 1$ is the price elasticity of product demand. Under these conditions the demands for capital and labour have the convenient forms

$$K_t = a_K^{\sigma} Y_t \left(\frac{r_t}{p_t(1-\frac{1}{\eta^D})}\right)^{-\sigma} \tag{2.11}$$

$$L_t = a_L^{\sigma} Y_t \left(\frac{w_t}{p_t(1-\frac{1}{\eta^D})}\right)^{-\sigma} \tag{2.12}$$

a static model in which capital inputs are leased at the rental price (r_t^i). A voluminous literature has considered how various tax structures impact on the user cost of capital. See, for example, Hall and Jorgenson (1967), King (1974), King and Fullerton (1984) and Jorgenson and Landau (1993).

[7] Or more realistically, we treat $F(K_t, L_t)$ as a production function for value-added rather than gross output.

giving the log-linear equations

$$\ln K_t = \sigma \ln a_K(1 - \frac{1}{\eta^D}) + \ln Y_t - \sigma \ln \left(\frac{r}{p}\right)_t \tag{2.13}$$

$$\ln L_t = \sigma \ln a_L(1 - \frac{1}{\eta^D}) + \ln Y_t - \sigma \ln \left(\frac{w}{p}\right)_t \tag{2.14}$$

which can be used as a basis for estimating the elasticity of substitution, or the responsiveness of factor intensities to changes in relative prices. Interpreted as expressions for the desired levels of factor inputs in the long run, these static factor demand equations form the basis for many reduced form models of investment and employment, as we discuss further in section 3.6 below.

Next, we consider the case of more than two factors of production. It is simple to extend the basic CES production function to this case, but this imposes the unappealing restriction that the elasticity of substitution between all pairs of inputs is the same. To allow different patterns of substitution (or complementarity) between different factors requires the use of a more flexible functional form. In this context it is convenient to consider the dual of the firm's profit maximisation problem, in which the firm is assumed to minimise its costs taking the level of output as given. To illustrate this dual approach we will assume that the cost function can be written as a translog (Christensen et al. (1971, 1973)), which is a second-order approximation to an arbitrary functional form.[8]

For n variable factors of production $X_t = (X_{1t}, ..., X_{nt})$ - for example, the capital, labour and current inputs considered above - and their associated vector of factor prices $W_t = (W_{1t}, ..., W_{nt})$ - for example, the user costs, wage rates and input prices considered above - the translog cost function has the form

$$\ln C_t = \ln \alpha_0 + \sum_{i=1}^{n} \alpha_i \ln W_{it} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln W_{it} \ln W_{jt} \tag{2.15}$$

[8] This has been a popular choice in the applied microeconometric literature. Alternative functional forms include the two-level CES and the Generalised Leontief (see Hamermesh (1993) or Berndt (1991) for a more extended discussion).

$$+\alpha_Y \ln Y_t + \frac{1}{2}\gamma_{YY}(\ln Y_t)^2 + \sum_{i=1}^{n} \gamma_{iY} \ln W_{it} \ln Y_t$$
$$+\phi_\tau t + \frac{1}{2}\phi_{\tau\tau} t^2 + \phi_{\tau Y} t \ln Y_t + \sum_{i=1}^{n} \phi_{\tau W_i} t \ln W_{it}$$

where the coefficients on time (t) represent technical change. The ϕ_τ and $\phi_{\tau\tau}$ coefficients reflect factor-neutral technical change, whilst the $\phi_{\tau W_i}$ coefficients reflect technical change that is biased towards factor i.[9] The cost minimising choices of input demands (X_{it}) are then conveniently expressed as log-linear cost share equations. From Shephard's (1953) Lemma we have

$$\frac{\partial \ln C_t}{\partial \ln W_{it}} = \frac{W_{it}}{C_t}\frac{\partial C_t}{\partial W_{it}} = \frac{W_{it} X_{it}}{C_t} = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln W_{jt} + \gamma_{iY} \ln Y_t + \phi_{\tau W_i} t \quad (2.16)$$

where $\frac{W_{it} X_{it}}{C_t} = S_{it}$ is the share of factor i in total costs.

There are a series of economic restrictions that can be imposed on the system of equations in (2.16). In order to correspond to a well behaved production function, a cost function should be homogeneous of degree one in the vector of factor prices. That is, for a given level of output, total cost must increase proportionally when all prices increase proportionally. This implies the following relationships among the parameters

$$\sum_{i=1}^{n} \alpha_i = 1; \quad \sum_{i=1}^{n} \gamma_{iY} = 0 \quad (2.17)$$
$$\sum_{j=1}^{n} \gamma_{ij} = \sum_{i=1}^{n} \gamma_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{n} \gamma_{ij} = 0.$$

The returns to scale (μ_t) can be computed as the inverse of the elasticity of costs with respect to output. Specifically

$$\mu_t = \left(\frac{\partial \ln C_t}{\partial \ln Y_t}\right)^{-1}. \quad (2.18)$$

[9] See Chambers (1988) for an extensive discussion, and section 7.5 below for a detailed treatment of skill-biased technical change.

Therefore, constant returns to scale implies that (for all i)

$$\gamma_{iY} = 0 \text{ and } \gamma_{YY} = 0 \tag{2.19}$$

and hence the output term drops out of the share equations (2.16) under this restriction.

Uzawa (1962) has shown that the Allen partial elasticities of substitution between two inputs i and j (Allen (1938)) can be computed from the cost formula

$$\sigma_{ijt} = \frac{C_t(\partial^2 C_t/\partial W_{it}\partial W_{jt})}{(\partial C_t/\partial W_{it})(\partial C_t/\partial W_{jt})}. \tag{2.20}$$

For the translog cost function we have

$$\sigma_{ijt} = \frac{\gamma_{ij} + S_{it}S_{jt}}{S_{it}S_{jt}}, \qquad i \neq j \tag{2.21}$$

and

$$\sigma_{iit} = \frac{\gamma_{ii} + S_{it}^2 - S_{it}}{S_{it}^2} \tag{2.22}$$

The own and cross-price factor demand elasticities are given by $\epsilon_{ijt} = S_{jt}\sigma_{ijt}$ and are also easily calculated from the estimated parameters. We discuss examples of the implementation of this structure for the three-factor case with capital, skilled labour and unskilled labour in section 7.5 below. Notice that the basic CES production function corresponds to the restriction that $\sigma_{iit} = \sigma_{ijt} = \sigma$ for all factors i and j and time periods t. The Cobb-Douglas production function (Cobb and Douglas (1928)) imposes the further restriction that $\sigma = 1$.

2.2. Dynamic factor demand

We now introduce adjustment costs. The basic motivation for introducing costs of adjustment is to account for the observation that adjustment in the level of factor inputs takes time to complete, or more specifically for the empirical failure of models which assume adjustment to be costless and immediate. The assumption that an activity is costly is a natural way for an economist to rationalise why more of it does not take place, but it is not the only possibility. Alternative models of investment have, for example, introduced exogenous delays (e.g. delivery lags) to explain investment dynamics, or considered limited substitution possibilities between installed capital and variable factors of production.[10]

We will continue to assume that current inputs are variable factors, in the sense that the level of these inputs can be varied immediately and without paying any adjustment costs. We will assume that capital inputs are quasi-fixed factors, in the sense that variations in their level impose costs of adjustment on to the firm, which will tend to delay and may possibly prevent their adjustment in response to new information. Thus at any point in time the level of capital inputs may differ from those which satisfy the static first-order conditions (like those illustrated for a price-taking firm in equation (2.8)); and if there are adjustment costs associated with, for example, replacement investment, then the steady-state level of capital inputs may also differ from the solution to the static problem with costless adjustment. We will also allow labour inputs to be subject to adjustment costs. Define $H_t = \left(H_t^1, ..., H_t^R\right)$ as a vector of gross hiring in each type of labour, and the equation of motion for labour as

$$L_t^i = \left(1 - \gamma^i\right) L_{t-1}^i + H_t^i \text{ for } i = 1, ..., R \tag{2.23}$$

where γ^i is the quit rate for labour of type i, also assumed to be exogenous and fixed.

The value maximisation problem analogous to (2.1) is now

$$V_t(K_{t-1}, L_{t-1}) = \left\{ \max_{I_t, H_t, M_t} \Pi_t\left(K_t, L_t, M_t, I_t, H_t\right) + \beta_{t+1} E_t\left[V_{t+1}\left(K_t, L_t\right)\right] \right\} \tag{2.24}$$

where the dependence of net revenue on gross investment and gross hiring now reflects the presence of adjustment costs. The net revenue function is now specified

[10] See Jorgenson (1971) and Nickell (1978) for comprehensive accounts of these approaches in the investment context.

as

$$\Pi_t\left(K_t, L_t, M_t, I_t, H_t\right) = p_t\left[F\left(K_t, L_t, M_t\right) - G\left(I_t, H_t, K_t, L_t\right)\right] - p_t^K I_t - w_t L_t - p_t^M M_t \tag{2.25}$$

where $G\left(I_t, H_t, K_t, L_t\right)$ is the adjustment cost function, with adjustment costs assumed to take the form of foregone production and initially assumed to be strictly convex in gross investment and gross hiring.

Given this specification, the solution to the firm's value maximisation problem continues to be characterised by first-order conditions (2.4), (2.5) and (2.7). The new first-order conditions for hiring and labour are

$$-\left(\frac{\partial \Pi_t}{\partial H_t^i}\right) = \mu_t^i \text{ for } i = 1, ..., R \tag{2.26}$$

$$\mu_t^i = \left(\frac{\partial \Pi_t}{\partial L_t^i}\right) + \left(1 - \gamma^i\right) \beta_{t+1} E_t\left[\mu_{t+1}^i\right] \text{ for } i = 1, ..., R \tag{2.27}$$

where $\mu_t^i = \frac{1}{1-\gamma^i}\left(\frac{\partial V_t}{\partial L_{t-1}^i}\right)$ is the shadow value of inheriting one additional unit of labour of type i in period t.

We have assumed that these costs of adjustment are strictly convex and differentiable, which will tend to smooth the adjustment of quasi-fixed factors to new information, since a series of small adjustments is assumed to be cheaper than a single large change in the level of these inputs. This limitation on the flexibility of factor inputs also rationalises forward-looking factor demand behaviour: since a change in the capital stock, for example, will be costly to reverse subsequently, the response of investment to a given change in the cost of capital will be different depending on whether that change is expected to be temporary or permanent. Hence these models predict that expectations of future demand and prices will be important determinants of current investment and employment decisions, which presents a particular challenge for econometric modelling. In the next section we discuss several approaches to this problem that have been used in the investment literature, beginning with those that can be obtained as special cases of the general factor demand model presented here. Specifications of dynamic labour demand models are discussed more briefly in section 7.

Whilst most structural models of investment and employment dynamics that have been widely used in modelling factor demands at the firm level in the last twenty years have been based on the assumption of strictly convex adjustment costs, it should be noted that this specification was introduced into the literature principally as a matter of analytical convenience. More recent models have considered the implications of non-convex costs of adjustment, either by assuming (partial) irreversibility of current investment decisions, or by introducing a fixed cost component in the specification of adjustment costs. These models, which predict large but infrequent adjustments, will be considered further in section 3.5 below.

3. Dynamic investment models

Most of the structural models of investment used in empirical analysis of firm level data can be obtained as special cases of the general factor demand model outlined in the previous section. Most of these models assume a single (homogeneous) capital input, and treat this as the only quasi-fixed factor used by the firm. The most popular of these models has been the Q model, which requires assumptions under which the unobserved shadow value of capital is simply related to the observed market-to-book or average q ratio. Specialising the assumption of strictly convex adjustment costs to a symmetric, quadratic functional form then yields a convenient, linear equation based on (2.4), relating investment to observed average q. Although still widely used, dissatisfaction with the empirical performance of the Q model has led to interest in less restrictive implementations of the basic adjustment costs model, such as the approach proposed by Abel and Blanchard (1986) and the Euler equation approach, introduced into the investment literature by Abel (1980). We consider each of these models in turn.

3.1. The Q model

To illustrate the Q model we assume that the firm's only quasi-fixed input is a single homogeneous capital good. Most realistically, this can be thought of as a fixed coefficients aggregate of the different capital goods used by the firm.[11] Provided there are no adjustment costs associated with labour and current inputs, the model can straightforwardly allow for many types of these inputs. We consider the case of multiple quasi-fixed factors in section 3.4 below.

Based on the net revenue equation (2.25) and the assumption of perfectly competitive markets, we then obtain

$$\left(\frac{\partial \Pi_t}{\partial I_t}\right) = -p_t \left(\frac{\partial G}{\partial I_t}\right) - p_t^K \tag{3.1}$$

which substituted into (2.4) yields

$$\left(\frac{\partial G}{\partial I_t}\right) = \left(\frac{\lambda_t}{p_t^K} - 1\right) \frac{p_t^K}{p_t} = (q_t - 1) \frac{p_t^K}{p_t}. \tag{3.2}$$

Moreover, solving equation (2.5) forward by repeated substitution yields

$$\lambda_t = E_t \left[\sum_{s=0}^{\infty} (1-\delta)^s \beta_{t+s} \left(\frac{\partial \Pi_{t+s}}{\partial K_{t+s}}\right) \right] \tag{3.3}$$

where β_{t+s} is the discount factor that discounts period t+s revenues back to period t.

To interpret these expressions, notice that the shadow value of an additional unit of capital (λ_t) is a forward-looking measure of current and expected future values of the marginal revenue product of capital, where the discounting reflects the diminution of each current unit of capital over time through depreciation, as well as the standard compensation for delay. In the static factor demand model,

[11] For example, the firm's technology may be such that it must always combine 2 units of equipment with 1 unit of structures, but it can substitute between this capital aggregate and other labour and current inputs.

the optimal capital stock was characterised by $\lambda_t = p_t^K$, or by $q_t = \frac{\lambda_t}{p_t^K} = 1$, where this ratio of shadow value to purchase cost is known as *marginal* q. With strictly convex costs of adjustment, marginal adjustment costs $\left(\frac{\partial G}{\partial I_t}\right)$ are an increasing function of current gross investment, so equation (3.2) shows that investment is an increasing function of the deviation between the actual value of marginal q and this desired value in the absence of adjustment costs. Moreover we have the striking result that all influences of expected future profitability on current investment are summarised in marginal q, through the shadow value of capital.

To obtain an empirical investment model we require an explicit form for marginal adjustment costs, and a way of measuring marginal q. Primarily for convenience, most implementations of the Q model have assumed that the cost of adjusting the capital stock is symmetric and quadratic about some 'normal' rate of investment, which may or may not be related to the rate of depreciation. More fundamentally, the basic Q model requires the adjustment cost function $G(I_t, K_t)$ to be homogeneous of degree one in (I_t, K_t), consistent with constant returns to scale. One popular functional form that has these properties, suggested by Summers (1981), is

$$G(I_t, K_t) = \frac{b}{2}\left[\left(\frac{I}{K}\right)_t - a\right]^2 K_t \tag{3.4}$$

where the parameter b reflects the importance of adjustment costs. Using this specification in (3.2) gives the linear model

$$\left(\frac{I}{K}\right)_t = a + \frac{1}{b}\left[(q_t - 1)\frac{p_t^K}{p_t}\right]. \tag{3.5}$$

The distinctive feature of the Q model is the equality between marginal q and average q established by Hayashi (1982). The basic requirement is that the net revenue function $\Pi_t(K_t, L_t, M_t, I_t)$ is homogeneous of degree one, sufficient conditions for which are that both the production function and the adjustment cost function display constant returns to scale, and the firm is a price taker in all markets.[12] In this case we can combine equations (2.4) and (2.5) to obtain

$$\lambda_t (K_t - I_t) = \left(\frac{\partial \Pi_t}{\partial I_t}\right) I_t + \left(\frac{\partial \Pi_t}{\partial K_t}\right) K_t + \beta_{t+1} E_t \left[(1-\delta)\lambda_{t+1} K_t\right] \tag{3.6}$$

or

$$(1-\delta)\lambda_t K_{t-1} = \Pi_t (K_t, L_t, M_t, I_t) + \beta_{t+1} E_t \left[(1-\delta)\lambda_{t+1} K_t\right] \tag{3.7}$$

since $\left(\frac{\partial \Pi_t}{\partial L_t}\right) = \left(\frac{\partial \Pi_t}{\partial M_t}\right) = 0$ for the variable factors of production. Solving forward by repeated substitution gives

$$(1-\delta)\lambda_t K_{t-1} = E_t \left[\sum_{s=0}^{\infty} \beta_{t+s} \Pi_{t+s} (K_{t+s}, L_{t+s}, M_{t+s}, I_{t+s})\right] = V_t \tag{3.8}$$

where V_t is again the maximised value of the firm. Thus we have

$$\lambda_t = \frac{V_t}{(1-\delta) K_{t-1}} \text{ or } q_t = \frac{V_t}{(1-\delta) p_t^K K_{t-1}} \tag{3.9}$$

so that marginal q is equal to the ratio of the maximised value of the firm in period t to the replacement cost value in period t of the capital stock that the firm inherits from the previous period. This ratio, known as average q or Tobin's q (Brainard and Tobin (1968), Tobin (1969)), can in principle be measured.[13] The usual implementation further requires that share prices are not affected by bubbles or fads, so that the 'fundamental' value of the firm given in (3.8) can be measured by its stock market valuation.[14]

Substituting average q for marginal q in (3.5) then gives the basic Q investment equation as

$$\begin{aligned} \left(\frac{I}{K}\right)_t &= a + \frac{1}{b}\left[\left(\frac{V_t}{(1-\delta) p_t^K K_{t-1}} - 1\right)\frac{p_t^K}{p_t}\right] \\ & a + \frac{1}{b} Q_t. \end{aligned} \tag{3.10}$$

[12] The presence of strictly convex adjustment costs in this model ensures that the value-maximisation problem has a solution, even with perfect competition and constant returns.

[13] Abel and Eberly (1994) show more generally that average q is proportional to marginal q if the net revenue function is homogeneous of degree k. The Hayashi (1982) equality result is the special case with k = 1.

[14] The model extends straightforwardly to incorporate exogenous debt policies, in which case the numerator of average q becomes the maximised value of the firm's capital assets, and the firm's equity market valuation has to be adjusted by an estimate of the firm's outstanding debt. The model can also be extended for various forms of taxation. See, for example, Summers (1981) and Hayashi (1982).

Notice that if share prices do correctly reflect fundamentals then the structure of the Q model implies that all relevant expectations of future profitability are summarised by the firm's stock market valuation, and the prediction that Q_t defined in (3.10) should be a sufficient statistic for investment.

One further point to notice is that the Q model identifies the parameters (a, b) of the adjustment cost function (3.4). These parameters are identified without requiring any functional form for the gross production function to be specified, given the assumptions of perfect competition and constant returns to scale. This may be an advantage or a disadvantage, depending on the context. If the objective is to quantify the importance of adjustment costs, or to test this specification, then this robustness to different functional forms for the production function may be an advantage. On the other hand if the objective is to estimate the response of investment to some change in tax rates or other component of the user cost of capital, then it is not sufficient to know the parameters of the adjustment cost function. Simulating the effects of a tax change on investment would require additional information about the elasticity of substitution between capital and other factors of production that is not identified by estimation of the Q model alone.[15]

We have highlighted the restrictive structure required to equate marginal q and average q, and the particular structure of adjustment costs needed to obtain a linear relationship between investment and marginal q. The advantages of the Q model compared to reduced form models are that the influence of expectations on current investment decisions is explicitly modelled, and that the parameters identified by estimating equation (3.10) are technological parameters of the adjustment cost function, which should be invariant to structural breaks in the underlying processes generating prices and interest rates. These are impor-

[15] See Summers (1981) and Salinger and Summers (1983) for examples of tax simulation in the context of the Q model.

tant advantages in the literature which tests the null hypothesis of perfect capital markets against an alternative in which financing constraints are important, as we discuss in section 6.2 below.

Nevertheless there is no shortage of reasons why the Q model may be seriously mis-specified. Adjustment costs may not be well described by the symmetric, quadratic functional form that is commonly imposed - thus the relationship between investment rates and Q_t may be non-linear and asymmetric even *within* the convex adjustment costs framework. Perfect competition and constant returns to scale may not be adequate assumptions, in which case average q ceases to be a sufficient statistic for the influence of expectations although marginal q may still be.[16] Stock market valuations may differ from fundamental values without necessarily violating weaker forms of the efficient markets hypothesis (Summers (1986)) - for example, share prices may be affected by rational bubbles (Blanchard and Watson (1982)) or liquidity traders (Campbell and Kyle (1993)). This would introduce severe measurement error problems if the average q ratio is constructed using share price data, and could undermine identification of the adjustment cost parameters.[17] Note also that for share prices to measure marginal q appropriately, the stock market must have the same expectations as the firm, in particular about the future path of the firm's capital stock. More mundane measurement error issues may also be important - the capital stock and debt measures that can be constructed from company accounts data are likely to be subject to substantial errors, and accurately measuring the (tax-adjusted) prices of capital goods relevant for individual firms is generally not possible with public data sets. It may be inappropriate to treat a single capital aggregate as the only quasi-fixed factor of production, and it may be necessary to distinguish between installed capital and new capital when specifying the substitution possibilities between capital and

[16] See Hayashi (1982).

[17] See Erickson and Whited (2000) and Bond and Cummins (2001) for further discussion.

other factors of production. Firms may operate in imperfect capital markets, and the objectives of their managers may not always coincide with shareholder value maximisation.

Thus it may not be completely surprising that the empirical performance of the Q model has generally been disappointing. Estimates of equation (3.10) have generally yielded extremely low values for the coefficient $(\frac{1}{b})$, suggesting extraordinarily high marginal costs of adjustment and implausibly slow adjustment of the actual capital stock. The prediction that Q_t is a sufficient statistic for investment has generally been rejected in empirical tests, and the explanatory power of the Q_t variable is often found to be very weak when other variables such as sales and cash flow are added to the econometric model. Most attempts to control for measurement error in Q_t have been unsuccessful in reversing these basic results, although Cummins et al. (1994) find that the Q model provides a more satisfactory description of investment behaviour in periods when the variation in measured Q_t is dominated by tax changes. Recently Erickson and Whited (2000) and Bond and Cummins (2001) have also reported more favourable results when accounting for forms of measurement error suggested by the possibility of bubbles or fads in stock market valuations. These findings are consistent with the interpretation that variation in Q_t as normally measured is dominated by uninformative 'noise' in share prices, although as we have emphasised there are many other reasons why the Q model may be mis-specified.

3.2. The Abel and Blanchard model

If the concern is that the conditions required to equate marginal q and average q may not hold, or if measures of average q based on stock market valuations are suspect, then an alternative approach is to attempt to measure marginal q itself, and estimate equation (3.5) directly.

Abel and Blanchard (1986) suggested constructing an estimate of the shadow value of capital using an auxiliary econometric model, based on equation (3.3). This procedure requires a specification for the marginal revenue product of capital in terms of observable variables, and a forecasting model for these variables. Notice that this forecasting model does not need to yield accurate predictions, but rather needs to mimic the expectations of future marginal revenue products on which the firm's investment decisions are based. How well this can be done using the information available to the econometrician is not entirely clear. Given a set of forecasts of the future marginal revenue products of capital, these are discounted back to the current period to yield an estimate of λ_t. This is then used to construct an estimate of marginal q, which can be used in place of the average q ratio to estimate the investment equation.

This procedure avoids the use of share price data, and can in principle be used to relax the assumptions of perfect competition and constant returns to scale if a suitable form for the marginal revenue product of capital is specified. These assumptions are replaced by a specification for the marginal revenue product, and the need to specify an auxiliary forecasting model. The linear specification of the investment model still relies heavily on the assumption of symmetric, quadratic costs of adjustment.

3.3. The Euler equation

The Euler equation approach introduced by Abel (1980) can also relax the linear homogeneity of the net revenue function and avoid the use of share price data. Perhaps more importantly, this approach avoids the need to parameterise the expectations-formation process. This is achieved by using the first-order condition for investment (2.4) to eliminate the shadow value of capital from the Euler equation (2.5), and then estimating the Euler equation itself rather than a model based on (2.4).

For the case of a single capital input this gives the expression

$$-\left(\frac{\partial \Pi_t}{\partial I_t}\right) = -\left(1-\delta\right)\beta_{t+1}E_t\left[\left(\frac{\partial \Pi_{t+1}}{\partial I_{t+1}}\right)\right] + \left(\frac{\partial \Pi_t}{\partial K_t}\right). \tag{3.11}$$

Using the net revenue function (2.25) and assuming perfectly competitive markets then gives

$$\left(\frac{\partial G}{\partial I_t}\right) = E_t\left[\psi_{t+1}\left(\frac{\partial G}{\partial I_{t+1}}\right)\right] + \left[\left(\frac{\partial F}{\partial K_t}\right) - \left(\frac{\partial G}{\partial K_t}\right) - \left(\frac{r}{p}\right)_t\right] \tag{3.12}$$

where $\psi_{t+1} = \left(\frac{1-\delta}{1+\rho_{t+1}}\right)\frac{p_{t+1}}{p_t}$ is a real discount factor and $\left(\frac{r}{p}\right)_t$ is the user cost of capital, as defined in equation (2.8).

Comparing equations (3.12) and (3.2) shows that the two terms on the right hand side of (3.12) contain essentially the same information as marginal q. In particular, given the current difference between the marginal product of capital and the user cost, all relevant information about expected future profitability is here summarised by the one-step ahead forecast of discounted marginal adjustment costs.

Using the adjustment cost function (3.4) yields

$$\left(\frac{I}{K}\right)_t = a\left(1 - E_t\left[\psi_{t+1}\right]\right) + E_t\left[\psi_{t+1}\left(\frac{I}{K}\right)_{t+1}\right] + \frac{1}{b}\left[\left(\frac{\partial F}{\partial K_t}\right) - \left(\frac{\partial G}{\partial K_t}\right) - \left(\frac{r}{p}\right)_t\right]. \tag{3.13}$$

To implement this model, the one-step ahead expected values can be replaced by the realised values of these variables in period t+1, which introduces forecast errors that will be orthogonal to information available in period t under the assumption of rational expectations. Assuming constant returns to scale, the marginal product of capital can be substituted without assuming a parametric form for the production function, as in Bond and Meghir (1994). Alternatively the model can be implemented by assuming some form for the production function, as in Abel (1980). Whilst this is more restrictive, it shows that in principle substitution parameters can be identified from the Euler equation, which is not the case for the Q model. The Euler equation model can also be extended to allow for imperfectly competitive product markets and/or for decreasing returns to scale.

3.4. Multiple quasi-fixed factors

All the models we have considered thus far in this section have treated capital as a single quasi-fixed input, and assumed that all other inputs can be adjusted costlessly. The more general model outlined in section 2.2 can be used to show how these models are affected by the presence of more than one quasi-fixed factor. To illustrate this, we consider the case in which the firm can substitute between two types of capital (e.g. equipment and structures), both of which are subject to adjustment costs. The implications of treating labour as a quasi-fixed factor of production are essentially similar.

Combining equations (2.4) and (2.5) as we did to obtain equation (3.6), assuming $\Pi_t\left(K_t^1, K_t^2, L_t, M_t, I_t^1, I_t^2\right)$ is again homogeneous of degree one, and summing across the two types of capital yields

$$\sum_{i=1}^{2}\left(1-\delta^i\right)\lambda_t^i K_{t-1}^i = \Pi_t + \beta_{t+1}E_t\left[\sum_{i=1}^{2}\left(1-\delta^i\right)\lambda_{t+1}^i K_t^i\right] = V_t. \tag{3.14}$$

Thus marginal q for the first type of capital can be expressed as

$$q_t^1 = \frac{\lambda_t^1}{p_t^{K,1}} = \frac{V_t}{\left(1-\delta^1\right)p_t^{K,1}K_{t-1}^1} + \frac{1}{p_t^{K,1}}\left(\frac{\partial \Pi_t}{\partial I_t^2}\right)\left(\frac{1-\delta^2}{1-\delta^1}\right)\left(\frac{K_{t-1}^2}{K_{t-1}^1}\right) \tag{3.15}$$

and similarly for q_t^2. Assuming for simplicity that the adjustment cost function is additively separable in the two types of investment, such as

$$G\left(I_t^1, I_t^2, K_t^1, K_t^2\right) = \frac{b_1}{2}\left[\left(\frac{I_t^1}{K_t^1}\right) - a_1\right]^2 K_t^1 + \frac{b_2}{2}\left[\left(\frac{I_t^2}{K_t^2}\right) - a_2\right]^2 K_t^2 \tag{3.16}$$

we obtain a model for investment in the first type of capital as

$$\begin{aligned}\left(\frac{I_t^1}{K_t^1}\right) \quad = \quad & a_1 + \frac{1}{b_1}\left[\left(\frac{V_t}{\left(1-\delta^1\right)p_t^{K,1}K_{t-1}^1} - 1\right)\frac{p_t^{K,1}}{p_t}\right] \\ & -\frac{b_2}{b_1}\left(\frac{1-\delta^2}{1-\delta^1}\right)\left(\frac{I_t^2}{K_t^2}\right)\left(\frac{K_{t-1}^2}{K_{t-1}^1}\right) + \frac{b_2 a_2}{b_1}\left(\frac{1-\delta^2}{1-\delta^1}\right)\left(\frac{K_{t-1}^2}{K_{t-1}^1}\right) \\ & -\frac{1}{b_1}\left(\frac{1-\delta^2}{1-\delta^1}\right)\left(\frac{p_t^{K,2}}{p_t}\right)\left(\frac{K_{t-1}^2}{K_{t-1}^1}\right)\end{aligned} \tag{3.17}$$

and similarly for investment in the second type of capital.

This shows that the basic Q model is mis-specified when there is more than one quasi-fixed factor. Whilst this system of equations could in principle be estimated, the literature has looked instead for restrictions on the form of adjustment costs that allow a single equation for total investment to be estimated. In particular, Hayashi and Inoue (1991) obtain conditions under which the structure of the basic Q model is preserved, with the aggregate capital stock measure being constructed as a Divisia index of the individual capital stocks, rather than as their sum.[18]

Both the Abel and Blanchard approach and the Euler equation approach extend more straightforwardly to the case of additively separable adjustment costs. An expression analogous to (3.3) holds for each type of capital. This can be used to obtain an estimate of marginal q for each type of capital, provided one is willing to specify the marginal revenue product for each type of capital. This will again yield a system of equations for each type of investment. For the Euler equation case, expressions analogous to (3.11) and (3.12) hold for each type of capital.

These approaches can also accommodate the case of interrelated adjustment costs, provided one is willing to specify the form of the adjustment cost function. See, for example, Shapiro (1986) for a system of Euler equations with interrelated adjustment costs.

3.5. Non-convex adjustment costs

Several considerations have motivated researchers to consider models with non-convex costs of adjustment. As noted earlier, the assumption of strictly convex adjustment costs was introduced primarily for analytical convenience. Descriptive evidence on the time series behaviour of investment in very disaggregated data sets has questioned the empirical validity of this simplifying assumption. Analytical techniques that characterise optimal investment behaviour in the presence of

[18] See also Galeotti and Schiantarelli (1991) and Chirinko (1993b).

non-convex adjustment costs have become more familiar to economists. And the empirical performance of the structural dynamic models based on strictly convex adjustment costs has been sufficiently problematic to encourage investigation of alternatives.

Census of production data provides evidence at the level of manufacturing establishments (plants, or related groups of plants), which is much more disaggregated than accounting data on companies. Doms and Dunne (1998) use data from the Longitudinal Research Database (LRD), covering 12,000 US manufacturing plants over the period 1972-89, to illustrate several aspects of the 'lumpy' adjustment of capital. For example, more than half of all plants experience a year in which the capital stock increases by over 35%; the two largest investment 'spikes' are often observed in consecutive years; and the incidence of these investment spikes is highly correlated with the time series of aggregate investment for these establishments. Anti Nilsen and Schiantarelli (1998) report similar findings using plant level data for Norway. Their results also show that around 30% of Norwegian plants have zero investment in an average year, although this proportion falls to only 6% if they focus on main production plants, or aggregate over multiple plants belonging to the same company. Although aggregation over plants to the company level tends to smooth some of the discreteness observed for individual plants, Abel and Eberly (1996) find that the distribution of investment rates is positively skewed in a large sample of publicly traded US companies from the Compustat database, and the two highest investment rates are observed in consecutive years for about half the firms in their sample.

This evidence of infrequent and lumpy adjustment can be explained by the presence of adjustment cost components that are not strictly convex in the level of investment.[19] Complete irreversibility occurs when gross investment is restricted

[19] The same evidence could also be explained by indivisibilities, ruling out small purchases of capital. See Nickell (1978) for further discussion.

to be non-negative. As Abel and Eberly (1996) report that some disinvestment (sales of capital) occurs in about 50% of their observations, this assumption appears too extreme to characterise the investment behaviour of most companies.[20] Partial irreversibility occurs when used capital goods can only be sold for less than their true replacement cost value, for example as a result of adverse selection in second-hand capital goods markets (cf. Akerlof (1970)). This wedge between the purchase price and the sale price of capital goods introduces a piecewise linear cost of adjustment, that is kinked at the point of zero gross investment. This kink is sufficient to explain why zero investment may occur even though the firm is not at its desired capital stock, but does not explain why observed adjustment should be lumpy. The latter can be explained by assuming a fixed cost component to adjustment costs, that is independent of the level of investment undertaken, although the size of the fixed cost may depend on whether the adjustment is upwards (positive investment) or downwards (disinvestment).

Theoretical models of investment have long since considered behaviour in the presence of irreversibility and linear adjustment costs (see, for example, Arrow (1968) and Nickell (1978)). However it is only more recently, and notably through the work of Dixit and Pindyck (1994) and Bertola and Caballero (1994), that these models have begun to have a major impact on the empirical literature.

Abel and Eberly (1994, 1996) have extended the Q model of investment to obtain a structural model of investment dynamics in the presence of non-convex adjustment costs. As in the traditional Q model, they assume constant returns to scale and perfect competition in all markets. Their adjustment cost function has the form

$$G\left(I_t, K_t\right) = a^+ K_t + b^+ I_t + c^+ \left(\frac{I}{K}\right)_t^2 K_t \text{ if } I_t > 0 \qquad (3.18)$$

[20] This is not to suggest that models with complete irreversibility are not useful in characterising some investment decisions, for example the decision to develop an offshore oil field. See, for example, Pesaran (1990) and Favero and Pesaran (1994).

$$= a^{-}K_t + b^{-}I_t + c^{-}\left(\frac{I}{K}\right)_t^2 K_t \text{ if } I_t < 0$$

where aK_t denotes a fixed cost of adjustment that is paid if investment is non-zero, bI_t denotes a linear adjustment cost, $c\left(\frac{I}{K}\right)_t^2 K_t$ denotes a strictly convex adjustment cost, and each of the parameters is allowed to take different values depending on the sign of gross investment. In this case the investment rate can be characterised as

$$\begin{aligned} \left(\frac{I}{K}\right)_t &= \frac{1}{c^{+}}\left(\lambda_t - p_t^K - b^{+}\right) \text{ if } \lambda_t > \overline{\lambda}_t\left(a^{+}, p_t^K + b^{+}\right) \qquad (3.19) \\ &= 0 \text{ if } \underline{\lambda}_t\left(a^{-}, p_t^K - b^{-}\right) \leqslant \lambda_t \leqslant \overline{\lambda}_t\left(a^{+}, p_t^K + b^{+}\right) \\ &= \frac{1}{c^{-}}\left(\lambda_t - p_t^K - b^{-}\right) \text{ if } \lambda_t < \underline{\lambda}_t\left(a^{-}, p_t^K - b^{-}\right) \end{aligned}$$

where λ_t is the shadow value of capital, $\overline{\lambda}_t$ is an upper threshold value below which the firm does not find it worthwhile to undertake positive gross investment, and $\underline{\lambda}_t$ is a lower threshold above which the firm does not find it worthwhile to undertake disinvestment. Moreover since this adjustment cost function is homogeneous of degree one in (I_t, K_t), the model can be implemented by equating marginal q with average q in the standard way. This model suggests a monotonic but non-linear relationship between investment and average q, with a region of inactivity where gross investment is zero between the two threshold values.

Caballero and Leahy (1996) have criticised the Abel and Eberly formulation. They show that the assumptions of perfect competition and constant returns to scale take on a more crucial role in the presence of fixed costs of adjustment. Not only is linear homogeneity of the net revenue function required to equate marginal and average q, but in the presence of fixed adjustment costs this restriction is also needed to obtain a monotonic relationship between investment and *marginal* q. If the net revenue function is concave, they show that marginal q becomes a non-monotonic function of investment and the inverse function, investment as a function of marginal q, does not exist. Thus the model described by (3.19) would be fundamentally mis-specified, however one measures the shadow value of capital.

Caballero and Leahy (1996) clearly regard the combination of perfect competition and constant returns to scale as 'unlikely' to be appropriate assumptions, particularly in the context of large firms. This is an interesting departure from Lucas (1967), who viewed this combination of assumptions as a strength of the underlying q theory. Without linear homogeneity of the net revenue function, the firm has an optimal size, whereas with linear homogeneity of the net revenue function the firm has no optimal size. This reconciles the model with evidence suggesting that changes in firm size are difficult to predict.[21] Recall that these assumptions have been made in the vast majority of structural investment models derived in the absence of fixed costs. Given this, testing a model which maintains these assumptions whilst introducing fixed costs does not seem to be an unreasonable project *a priori*. Whilst linear homogeneity may not hold exactly, it may provide a sufficiently good approximation to be useful, and the fact that this assumption greatly simplifies the characterisation of optimal investment in the presence of fixed adjustment costs would be no small advantage if this is the case. In our view, conclusive evidence that linear homogeneity should be abandoned in the investment literature has not yet been presented.

Abel and Eberly (1996) and Barnett and Sakellaris (1998) have estimated non-linear models relating the investment rate to average q using data on US publicly traded companies. Both studies find significant non-linearities. Barnett and Sakellaris (1998) find a concave relationship, with a flatter slope at higher values of average q. Abel and Eberly (1996) find an S-shaped sigmoidal relationship, with a steeper slope at values of average q around unity and a flatter response at both higher and lower values.[22]

At first sight this seems inconsistent with the predictions of the model with

[21] The proposition that firm size follows a random walk is known as Gibrat's Law, and a large literature in empirical industrial organisation has sought to test this hypothesis (see Sutton (1998)). Whilst the results are somewhat inconclusive, the fact that a large literature exists suggests that the proposition is not without interest, and is not rejected out of hand.

[22] See also Eberly (1997).

non-convex adjustment costs, which implies no response of investment to average q in the intermediate range of inaction, and a response of investment to average q only when average q moves outside this range. Abel and Eberly (1996) suggest that this apparently paradoxical finding can be explained by aggregation over different types of capital goods. The basic idea is that with non-convex adjustment costs, adjustment occurs at both the extensive margin (the decision to invest or not to invest) and the intensive margin (how much to invest or disinvest given that an adjustment is worthwhile). At intermediate values, a rise in average q will trigger adjustment at the extensive margin for some types of capital, with spending on those types of investment increasing from zero to some level that is large enough to justify paying the corresponding fixed cost. At higher values, a further rise in average q produces adjustment only at the intensive margin for all types of capital, so that the response of total investment spending to the change in average q may be flatter.

This argument is ingenious, but it should be noted that there are censoring and measurement error problems that can also account for the non-linear shape found by Abel and Eberly (1996), even if the true relationship is linear. Gross investment in their model is allowed to be negative, and should be so if average q is low enough to justify disinvestment. However the investment data used to estimate this relationship measures only positive capital expenditures, without subtracting any measure of disinvestment. Thus the measure of gross investment used in the empirical work is effectively censored below at zero. This censoring, which is not accounted for in their estimation, can potentially explain a flattening in the response of measured investment to average q at low values of q: once the dependent variable is close to zero, it cannot go much lower, whatever happens to average q. Measurement error in average q can also explain the flattening observed at the upper end of the distribution, particularly if, as seems likely, the measurement error has a higher variance at higher values of average q - for example, if high average q ratios are more likely to be affected by bubbles or fads in share prices. More generally, measurement error in the explanatory variable can change the shape of the underlying relationship, as well as introducing an attenuation bias (see, for example, Chesher (1991)). As noted earlier, measurement error is likely to be a serious problem with the average q ratio. Bond and Cummins (2001) suggest that noise in share prices can account for findings of non-linearity in this context, when average q is measured using stock market valuations.

The censoring problem noted here may also be more fundamental than it appears. The measure of disinvestment that we would like to subtract from capital expenditures is the replacement cost value of the capital goods that are sold or disposed of. The proceeds from sales of capital goods will underestimate this if there are adverse selection problems in second-hand capital goods markets, and the book value of disposals will underestimate their replacement cost value by an unknown amount depending on the age of the goods that are sold and the inflation rates that have been experienced. Thus whilst it may be important to allow for non-linearities in the relationship between *measured* investment and its determinants, it is not clear that these non-linearities can identify the shape of the underlying adjustment cost function, at least over the range where disinvestment is important.

Caballero et al. (1995) adopt a somewhat more *ad hoc* approach to test for the importance of fixed adjustment costs.[23] Their empirical analysis is motivated by the following observations. In a model with (possibly asymmetric) fixed costs of adjustment, but no linear or strictly convex components, capital adjustment if it occurs will take the capital stock to the same target level, K_t^*, whether this adjustment occurs from above or from below. Conditional on the fixed cost levels, positive investment will occur when the ratio $Z_t = \frac{K_t^*}{K_t}$ exceeds some critical threshold, and disinvestment will occur when Z_t goes below some threshold. However if

[23] See also Cooper et al. (1999) and Anti Nilsen and Schiantarelli (2003).

the actual fixed cost levels are stochastic and not observed by the econometrician, all we can say is that the probability of observing investment or disinvestment should be increasing in the absolute value of $z_t = \ln Z_t$.

Caballero et al. (1995) implement this test using LRD data on US manufacturing plants. This requires constructing a measure of the target capital stock (K_t^*), and hence the 'gap' (z_t). Although in general this target will not be equal to the optimum capital stock that the firm would choose in the absence of adjustment costs, Caballero et al. (1995) base their measure of the target on the predicted values from a static demand for capital equation of the form (2.13). They find that the probability of observing investment is clearly increasing with positive values of z_t, consistent with the presence of the fixed costs. They also find an asymmetry, with the probability of observing disinvestment remaining low even at large negative values of z_t, although as they note this asymmetry could reflect mis-measurement of disinvestment in the data. This 'gap' methodology has also been criticised by Cooper and Willis (2001), who suggest that the findings may be sensitive to mis-specification of the target level to which the actual capital stock is assumed to adjust.

Finally in this section we should mention interesting recent work by Cooper and Haltiwanger (2000). They also use US plant level data to investigate capital stock adjustment, but rather than directly estimating an investment equation they use indirect inference to estimate parameters of a general adjustment cost function. In essence, this asks what forms of adjustment costs are required to match the non-linear relationship between investment and profitability found in their data set. Perhaps not surprisingly, they find that a general specification of adjustment costs is required to fit this relationship, combining both convex and non-convex components with irreversibility. This finding, like the descriptive evidence on plant-level investment discussed earlier, casts doubt on the structural econometric investment models that have maintained much simpler adjustment cost specifications. The challenge remains to develop structural investment equations that are consistent with richer forms of adjustment costs, without requiring the net revenue function to be homogeneous of degree one.

3.6. Reduced form models

The actual adjustment process is likely to be extremely complex, particularly when we consider that the data on investment at the firm level are an aggregate over many types of capital goods, and possibly over multiple plants. The structural models of investment dynamics that have been proposed to date have not been conspicuously successful in characterising this adjustment process, possibly because they have neglected these aggregation issues, and/or the implications of some of the non-convexities discussed in the previous section. Nevertheless it is clear that capital cannot be adjusted costlessly and immediately, so it is not appropriate to resort to static models.[24] An intermediate possibility is to rely on dynamic econometric specifications that are not explicitly derived as optimal adjustment behaviour for some particular structure of adjustment costs. A favourable interpretation of such reduced form models is that they represent an empirical approximation to some complex underlying process that has generated the data. A less favourable interpretation is that they compound parameters of the adjustment process with parameters of the expectations-formation process, and are subject to the Lucas (1976) critique. In any case it is useful to be aware of the form of these models, and some of their limitations.

One approach which has been widely used in the investment literature is based on first-differencing a static factor demand model to obtain an investment equation, for example

$$\left(\frac{I_t}{K_{t-1}}\right) - \delta \approx \Delta k_t = \Delta k_t^* \qquad (3.20)$$

[24] Unless perhaps one is willing to treat all variables of interest as co-integrated non-stationary variables and one has sufficiently long time series to appeal to co-integration results (cf. Engle and Granger (1987)).

where k_t^* is the logarithm of the optimal capital stock, which may for example be given by equation (2.13). If k_t^* is log-linear in the level of output, this leads to versions of the popular accelerator model, in which investment is related to output growth. Recognising that the actual capital stock does not adjust fully and immediately to changes in the desired level, so-called flexible accelerator models introduce distributed lags in Δk_{t-s}^* and possibly Δk_{t-s}. This gives a dynamic specification of the form

$$a(L)\Delta k_t = b(L)\Delta k_t^* \tag{3.21}$$

where $a(L)$ and $b(L)$ are polynomials in the lag operator (i.e. $L^s x_t = x_{t-s}$). Flexible accelerator models of this type have been estimated using firm data by, for example, Eisner (1977) and Mairesse and Dormont (1985).

An alternative is to specify a simple partial adjustment model for the level of the capital stock, such as

$$\left(\frac{I_t}{K_{t-1}}\right) - \delta \approx \Delta k_t = \theta\left(k_t^* - k_{t-1}\right) \tag{3.22}$$

in which some constant fraction θ of the gap between the actual and desired levels of the capital stock is closed in each period. This is obviously very restrictive, but a more flexible dynamic adjustment model which nests both partial adjustment and accelerator models as special cases is an error correction model, such as

$$\alpha(L)\Delta k_t = \beta(L)\Delta k_t^* + \theta\left(k_{t-s}^* - k_{t-s-1}\right) \tag{3.23}$$

where $\alpha(L)$ and $\beta(L)$ are again polynomials in the lag operator, the form of which can be chosen empirically.[25] Error correction models were introduced into the investment literature by Bean (1981), and have been considered in the context of firm data by Bond et al. (1999) and Bond et al. (2003).

The connection between error correction models and co-integration techniques have popularised these adjustment models in the time series literature, but the

[25] See Hendry (1995, Chapter 7) for further discussion of alternative dynamic models.

error correction model was introduced into econometrics long before the literature on co-integration developed.[26] In fact, the error correction model is nothing more than a particular parameterisation of an autoregressive distributed lag (ADL) model. For example, the ADL(1,1) model

$$k_t = \alpha_1 k_{t-1} + \beta_0 k_t^* + \beta_1 k_{t-1}^* \tag{3.24}$$

can always be re-parameterised as

$$\Delta k_t = -\beta_1 \Delta k_t^* + (1-\alpha_1)(k_t^* - k_{t-1}) \tag{3.25}$$

under the long-run proportionality restriction $\left(\frac{\beta_0+\beta_1}{1-\alpha_1}\right) = 1$.[27]

Implementation of the error correction model will require a specification for the target level of the capital stock. Subject again to the observation that the target capital stock in the presence of adjustment costs is not necessarily equal to the desired capital stock in the absence of adjustment costs, this can be based on a static factor demand specification such as (2.13). For example, combining (2.13) and (3.25) gives

$$\begin{aligned} \Delta \ln K_t &= (1-\alpha_1)\,\sigma \ln a_K (1 - \frac{1}{\eta^D}) - \beta_1 \Delta \ln Y_t + \beta_1 \sigma \Delta \ln \left(\frac{r}{p}\right)_t \\ &\quad + (1-\alpha_1)(\ln Y_t - \ln K_{t-1}) - (1-\alpha_1)\,\sigma \ln \left(\frac{r}{p}\right)_t . \end{aligned} \tag{3.26}$$

Notice that these models can be used to estimate the long-run elasticity of the capital stock with respect to the user cost of capital, whilst allowing for the fact that this adjustment does not occur immediately. In principle these models can be extended to incorporate non-linear and asymmetric dynamics, for example by allowing the parameters α_1 and β_1 to take different values depending on the sign of $(k_t^* - k_{t-1})$ and whether the absolute value of this gap is large or small.

[26] See, for example, Sargan (1964) and Davidson et al. (1978).

[27] This restriction can be tested by including an additional term in either k_t^* or k_{t-1} on the right hand side of (3.25).

Since these reduced form models are empirical generalisations of static factor demand specifications, and the adjustment cost models considered previously are theoretical generalisations of static factor demand theory, it is no surprise that the two approaches are related. Nickell (1978, Chapter 11) shows that for a symmetric, quadratic adjustment cost function, the level of investment can be obtained approximately as

$$I_t - \delta K_{t-1} = \Delta K_t \approx \phi\left(\widehat{K}_t - K_{t-1}\right) \tag{3.27}$$

where

$$\widehat{K}_t = (1-\gamma)\, E_t\left[\sum_{s=0}^{\infty} \gamma^s K^*_{t+s}\right] \tag{3.28}$$

and K^*_t is the capital stock level that the firm would have chosen in the absence of adjustment costs. Thus the investment decision is approximately described by a partial adjustment mechanism, in which the 'target' level of the capital stock is itself a function of both current and expected future levels of the static optimum.[28] Nickell (1985) characterises restrictive conditions for the K^*_t process under which this generates either a partial adjustment or an error correction model relating investment to $(K^*_t - K_{t-1})$. Whilst this relationship is of interest, it does not provide a very appealing motivation for the use of these reduced form empirical models. If this structure for the adjustment costs is taken seriously, we have seen how this structure can be identified and tested more directly. The attraction of the reduced form models is that they may provide an empirical approximation to a much more complex adjustment process, whose structure we have not yet been able to characterise satisfactorily as the outcome of a richer dynamic optimisation problem.

The potential disadvantages of the reduced form approach can now be illustrated as follows. Suppose for simplicity that the structural adjustment process

[28] Recall that this dependence of current investment decisions on expected future profitability was also evident from the first-order conditions (see equation (3.3)).

was characterised by

$$I_t = \alpha K_{t-1} + \beta_0 Y_t + \beta_1 E_t \left[Y_{t+1} \right] \tag{3.29}$$

which is a special case of (3.27) in which we have omitted expected values beyond period t+1, and assumed that the static optimum capital stock is proportional to output. Suppose also that expectations of future output are formed according to

$$E_t \left[Y_{t+1} \right] = \pi_0 Y_t + \pi_1 Y_{t-1} + \pi_2 X_t \tag{3.30}$$

where X_t is a vector containing any additional variables that happen to be useful in forecasting future output. Then substituting (3.30) into (3.29) we obtain the reduced form investment equation

$$I_t = \alpha K_{t-1} + (\beta_0 + \beta_1 \pi_0) Y_t + \beta_1 \pi_1 Y_{t-1} + \beta_1 \pi_2 X_t. \tag{3.31}$$

This shows how the reduced form models compound the parameters of the structural adjustment process $(\alpha, \beta_0, \beta_1)$ with the parameters of the expectations-formation process (π_0, π_1, π_2).[29] This has two potentially important consequences. First, if the parameters of the expectations-formation process are not stable, then the parameters of the reduced form investment equation will also be unstable, even though the parameters of the structural adjustment process may have been constant. For example, if there is a structural break in the process generating output, perhaps as a result of entry into the market or some change in macroeconomic policy, this will induce parameter instability in the reduced form investment model. This simply illustrates that the reduced form models are subject to the Lucas (1976) critique. In principle this is not the case for the structural investment models we considered in sections 3.1-3.3, whose parameters are 'deep' parameters describing the adjustment cost technology (see (3.10) or (3.13)), and which are

[29] If the structural adjustment process were really as simple as equation (3.29), we could of course substitute the realised value of future output for the one-step ahead expectation. However this will not generally be possible when the structural model relates current investment to expectations of output or profitability in the distant future, as suggested by (3.27) or (3.3).

expected to be invariant to changes in the processes generating, for example, output and the user cost. In practice, this claim relies on these structural models being correctly specified, and even then we may find that the parameters of the adjustment cost function are not constant over time.

The second consequence is that X_t variables will appear to be significant in reduced form equations like (3.31), even though they play no role in the structural model for investment, and their only role in the reduced form equation is to help to forecast future values of the fundamental determinants of investment. This is clearly problematic if we want to draw any inferences about the nature of the underlying structural model. For example, finding that financial variables have significant coefficients in a reduced form investment equation does not identify whether these financial variables are important structural determinants of investment spending - perhaps as a result of financing constraints - or whether they simply help to forecast future output or profitability. We discuss some possible solutions to this identification problem further in section 6.2 below. Here we note that whilst the problem is particularly transparent in the context of reduced form models, a similar issue will affect structural models that are not correctly specified, and which therefore do not fully control for all influences of expected future output or profitability on the level of current investment.

4. Econometric issues

4.1. Stochastic specification and identification

One important issue in the implementation of econometric investment and employment models concerns the sources of stochastic error terms. In this section we discuss in particular the stochastic specifications that have been considered in microeconometric applications of the Q and Euler equation models of investment, and their implications for the consistent estimation of these models.

The Q model derived in equation (3.10) is a deterministic relationship between the investment rate and the Q variable. The intercept in this equation is a parameter of the adjustment cost function, interpreted as a 'normal' rate of investment at which costs of adjustment are zero. The standard way of introducing stochastic variation into the Q model is to treat this parameter as stochastic, and to interpret the error term in the Q investment equation as reflecting shocks to the adjustment cost function. Simply replacing the constant a with $a_{it} = a + e_{it}$ for firm i in time period t gives the econometric model

$$\left(\frac{I}{K}\right)_{it} = a + \frac{1}{b}Q_{it} + e_{it} \tag{4.1}$$

where e_{it} is an additive shock to the 'normal' rate of investment, or equivalently an additive shock to marginal adjustment costs.

This approach is convenient, if somewhat *ad hoc*. Nevertheless it has several implications which should be taken seriously when estimating the Q model. Perhaps most importantly, it implies that Q_{it} should be an endogenous variable in the econometric model (4.1). Current shocks to adjustment costs will affect the current period's net revenue (Π_{it}), and therefore the current value of the firm (V_{it}).[30] This endogeneity of Q_{it} will therefore need to be taken into account in order to obtain consistent estimates of the adjustment cost parameters.

A second implication is that technological shocks need not be statistical innovations. There may be permanent differences across firms in their 'normal' investment rates, and there may be common trends in the nature of adjustment costs that affect all firms in the same way, perhaps as a result of business cycle fluctuations. Thus for estimation of the Q model using company panel data it is not inconsistent with the underlying theory to include firm-specific and time-specific error components. Simply letting $e_{it} = \eta_i + \zeta_t + \varepsilon_{it}$ gives the error components

[30] See equation (3.8). Alternatively, the current adjustment cost shock affects the current marginal revenue product of capital, and therefore the current shadow value of capital (λ_{it}). See equation (3.3).

specification

$$\left(\frac{I}{K}\right)_{it} = a + \frac{1}{b}Q_{it} + \eta_i + \zeta_t + \varepsilon_{it}. \tag{4.2}$$

Moreover there is no compelling reason for the idiosyncratic, time-varying component of adjustment cost shocks (ε_{it}) to be serially uncorrelated. For example, if these shocks follow an $AR(1)$ process, $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + \nu_{it}$, with ν_{it} serially uncorrelated, then we obtain the dynamic specification

$$\left(\frac{I}{K}\right)_{it} = a(1-\rho) + \rho\left(\frac{I}{K}\right)_{i,t-1} + \frac{1}{b}Q_{it} - \frac{\rho}{b}Q_{i,t-1} + \eta_i(1-\rho) + \zeta_t - \rho\zeta_{t-1} + \nu_{it}, \tag{4.3}$$

i.e. a dynamic model relating the current investment rate to both current and lagged Q and the lagged dependent variable, subject to a non-linear common factor restriction. More generally, the underlying theory does not rule out a dynamic relationship between investment rates and the Q variable, provided these dynamics are consistent with some serial correlation process in the adjustment cost shocks.

A second potential source of stochastic variation in the Q model is measurement error. In view of both the assumptions required to measure marginal q, using either the average q ratio or an auxiliary econometric forecasting model, and the limitations of most publicly available datasets,[31] the likelihood of significant measurement error in the Q variable does indeed seem overwhelming. Denoting the true value of the explanatory variable on the right-hand side of (3.5) or (3.10) by Q^*_{it}, and the measured value by $Q_{it} = Q^*_{it} + m_{it}$, where m_{it} is an additive measurement error, then gives an econometric model of the same form as (4.1), with $e_{it} = -\frac{m_{it}}{b}$. Again a principal implication is that measured Q_{it} will be correlated with the error term, and this endogeneity should be allowed for in estimation.[32]

[31] See section 5 below.

[32] Though it is perhaps worth noting that whilst the theory predicts a positive correlation between current Q_{it} and adjustment cost shocks in the absence of measurement error, the correlation between measured Q_{it} and $-\frac{m_{it}}{b}$ will be negative. Thus the direction of the potential simultaneity bias becomes ambiguous if both adjustment cost shocks and measurement errors are present.

Moreover the measurement error may also have firm-specific, time-specific and serially correlated components. Unfortunately the residual measurement error component may still not have properties that are convenient for estimation, and alternative techniques such as those considered in Erickson and Whited (2000), or alternative measures such as those considered in Bond and Cummins (2001), may still be required to identify the parameters of the underlying model.

In contrast to the Q model, the Euler equation approach considered in section 3.3 has an intrinsically stochastic specification when the one-step ahead expected values are replaced by their realised values. Assuming the real discount factor ψ_{t+1} is a constant parameter ψ and denoting $\left(\frac{I}{K}\right)_{i,t+1} = E_{it}\left[\left(\frac{I}{K}\right)_{i,t+1}\right] + \epsilon_{i,t+1}$, where $\epsilon_{i,t+1}$ is the error made by firm i when forecasting its period t+1 investment rate using information available in period t, the Euler equation in (3.13) becomes

$$\left(\frac{I}{K}\right)_{it} = a(1-\psi) + \psi\left(\frac{I}{K}\right)_{i,t+1} + \frac{1}{b}\left[\left(\frac{\partial F}{\partial K_{it}}\right) - \left(\frac{\partial G}{\partial K_{it}}\right) - \left(\frac{r}{p}\right)_{it}\right] - \psi\epsilon_{i,t+1}. \tag{4.4}$$

The forecast error $\epsilon_{i,t+1}$ will certainly be correlated with $\left(\frac{I}{K}\right)_{i,t+1}$, but under weak rational expectations should be orthogonal to information available in period t. One implication is that this forecast error cannot be serially correlated, and cannot contain a permanent firm-specific component.[33] On the other hand, since firms are likely to be subject to common shocks reflecting, for example, business cycle surprises, these forecast errors are likely to be correlated across firms in each period. This suggests that identification of the Euler equation may be problematic without long time series of data for individual firms, since independence of the error terms across individuals is typically required for consistent estimation in panels when the number of time periods is fixed. The inclusion of time dummies will only permit consistent estimation in the special case where the effects of aggregate shocks are perfectly correlated across individual firms. Of course, the

[33] Although firm-specific measurement errors or omitted variables *may* rationalise firm-specific effects in empirical applications of Euler equation models of factor demand.

assumption that technological shocks and measurement errors are independent across firms (conditional on time dummies) may also be problematic in the context of the Q model and other factor demand specifications, but this assumption is particularly difficult to reconcile with the underlying economic structure when the stochastic disturbances contain a forecast error.[34]

It should also be noted that unobserved heterogeneity in the real discount factor ψ could undermine the identification of the Euler equation. Measured heterogeneity in ψ, reflecting for example differences in the mix of capital assets used by different firms or differences in their required rates of return, can be allowed for by including suitable interaction terms in the estimated model. Unobserved heterogeneity in ψ is problematic because, as is clear from (4.4), this interacts with the endogenous variable $\left(\frac{I}{K}\right)_{i,t+1}$. Unrestricted unobserved heterogeneity (ψ_{it}) would therefore leave the Euler equation unidentified. Restricting the unobserved heterogeneity to be permanent and firm-specific ($\psi_{it} = \psi_i$) would allow consistent estimation if long time series data were available, but even this restriction would leave an identification problem in panels with a small number of time periods. As has been emphasised by Pesaran and Smith (1995), heterogeneous slope coefficients will invalidate the instruments typically used to estimate dynamic models from short panels.[35] Again we should acknowledge that a similar issue may arise in the context of the Q model if the adjustment cost parameter (b) is itself heterogeneous across firms.

The potential sources of stochastic variation in reduced form investment equa-

[34] Interested readers are referred to Altug and Miller (1990, 1991) and Blundell et al. (1996) for further discussion of this issue.

[35] Although it is less clear from (4.4), a similar issue can arise if there is unobserved heterogeneity in the adjustment cost parameter a. The source of this potential problem is the $\left[\left(\frac{\partial F}{\partial K_{it}}\right) - \left(\frac{\partial G}{\partial K_{it}}\right)\right]$ term. If, as in Bond and Meghir (1994), linear homogeneity of the production and adjustment cost functions is used to measure this marginal product, the effect is to introduce a linear term in $a\left(\frac{I}{K}\right)_{it}$. Indeed it should be noted that this form of the Euler equation is inconsistent with the presence of the kind of adjustment cost shocks that are typically used to motivate stochastics in the Q model.

tions are less transparent. If these models are regarded as empirical approximations to some complex adjustment process, then the residual can be viewed as an approximation error, but this gives little guidance as to its statistical properties. One approach in this context is to regard certain properties of the error term (e.g. lack of serial correlation) and certain properties of the model parameters (e.g. stability) as features of a desirable approximation to the data generation process, and to treat these properties as key objectives of an empirical specification search. Since this raises issues that are common to reduced form models in a wide range of econometric applications, we refer interested readers to Hendry (1995) for further discussion.

4.2. Estimation

Many of the issues that arise in estimating dynamic factor demand equations using company panel data can be illustrated by considering a model of the form

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \beta x_{it} + u_{it} \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T \qquad (4.5) \\ u_{it} &= \eta_i + v_{it} \end{aligned}$$

where i indexes firms and t indexes time, and η_i is an unobserved firm-specific effect.[36] For example, if y_{it} is the investment rate, x_{it} is Q_{it} as defined in (3.10) and $\alpha = 0$ then we have the basic Q model, whilst if $\alpha \neq 0$ and x_{it} is a vector containing both Q_{it} and $Q_{i,t-1}$ we have the Q model with an $AR(1)$ component to the adjustment cost shock. The Euler equation model (4.4) has this form if we normalise on $\left(\frac{I}{K}\right)_{i,t+1}$ rather than $\left(\frac{I}{K}\right)_{it}$,[37] and the reduced form models considered in section 3.6 are typically just more general dynamic equations of this type. For simplicity, this section will focus on the case where x_{it} is a scalar.

Since the available company panels typically contain a large number of firms observed for a relatively small number of time periods, we will concentrate on the

[36] See Arellano and Honore (2001) for a more detailed discussion of the methods we review here.

[37] Similar estimation issues will arise whichever normalisation is adopted.

estimation issues that arise when N is large and T is small, assuming that the error term v_{it} is distributed independently across firms.[38] OLS will give biased parameter estimates since $y_{i,t-1}$ is necessarily correlated with η_i (and x_{it} may be), and the Within estimator will give biased parameter estimates for small T since $\widetilde{y}_{i,t-1}$ is necessarily correlated with $\widetilde{u}_{it}$ (and $\widetilde{x}_{it}$ may be).[39,40]

If x_{it} is strictly exogenous with respect to u_{is}, in the sense that $E\left[x_{it}u_{is}\right]=0$ for all s,t (or some strictly exogenous instrument z_{it} is available), the parameters (α,β) can be estimated consistently by using the vector $(x_{i1},...,x_{iT})$ - or $(z_{i1},...,z_{iT})$ - as instrumental variables for each of the levels equations in (4.5). If x_{it} is correlated with η_i but strictly exogenous with respect to v_{is} (or some instrument z_{it} with these properties is available), the parameters (α,β) can be estimated consistently by taking first-differences of (4.5) and then using the vector $(x_{i1},...,x_{iT})$ - or $(z_{i1},...,z_{iT})$ - as instrumental variables for each of the resulting first-differenced equations. More commonly, in the absence of strictly exogenous instruments, identification of (α,β) relies on assuming some limited serial correlation in the v_{it} disturbances.

For example, consider the case in which x_{it} is correlated with both η_i and v_{it}. Assuming that $E[y_{i1}v_{it}]=E[x_{i1}v_{it}]=0$ for $t=2,...,T$ and that $E[v_{is}v_{it}]=0$ for $s\neq t$ yields the moment conditions

$$E\left[w_{i,t-s}\Delta v_{it}\right]=0 \text{ for } s\geqslant 2 \text{ and } t=3,...,T \tag{4.6}$$

where $w_{it}=(y_{it},x_{it})$. This allows the use of lagged values of endogenous vari-

[38] Unless otherwise indicated, asymptotic properties will hold as N goes to infinity for fixed T.

[39] The Within estimator is obtained as OLS after subtracting firm-means of each variable, so that $\widetilde{y}_{it}=y_{it}-\frac{1}{T-1}\sum_{s=2}^{T}y_{is}$ and $\widetilde{y}_{i,t-1}=y_{i,t-1}-\frac{1}{T-1}\sum_{s=1}^{T-1}y_{is}$. This is equivalent to the Least Squares Dummy Variables (LSDV) estimator, obtained by including a dummy variable for each firm.

[40] In the special case where $\beta=0$, we can say that the OLS estimate of α will be biased upwards (Hsiao (1986)) and the Within estimate of α will be biased downwards (Nickell (1981)). Informally these estimates provide guidance about the likely range of values for α, so that candidate consistent estimators which lie well outside this range can often be regarded with suspicion.

ables dated $t-2$ and earlier as instrumental variables for the equations in first-differences.[41] For $T > 3$ the parameters are over-identified, and alternative Generalised Method of Moments (GMM) estimators are defined by different ways of weighting the moment conditions. If instead of assuming no serial correlation we allow v_{it} to be $MA(q)$, the implication is that only lagged endogenous variables dated $t-2-q$ and earlier can be used as instruments.

This first-differenced GMM estimator will provide consistent estimates of the parameters (α, β) as the number of firms becomes large. In some contexts these estimators have good finite sample properties, but this will not be the case when the lagged values of the series are only weakly correlated with subsequent first-differences. When the instruments available are weak the GMM estimator can exhibit large finite sample biases, as well as imprecision.[42] In the context of equation (4.5) this is particularly likely to be a problem when the individual series for y_{it} and x_{it} are highly persistent, and when the time series dimension of the panel is very short.[43]

Alternative estimators are available that have better small sample properties in these cases, although they have been less commonly used in applied work to date. Alonso-Borrego and Arellano (1999) propose a symmetrically-normalised GMM estimator that is median-unbiased even in the case of weak instruments. Ahn and Schmidt (1995) note that estimators based on the linear moment conditions (4.6) are not efficient under the standard 'error components' assumption that $E(\eta_i v_{it}) = 0$, and that additional non-linear moment conditions are available in this case. The resulting non-linear GMM estimator is asymptotically efficient relative to the linear first-differenced GMM estimator, and can be expected to

[41] See Holtz-Eakin et al. (1988) and Arellano and Bond (1991). $x_{i,t-1}$ could be used as an additional instrument if x_{it} is predetermined with respect to v_{it}. Additional instruments are available for the levels equations if x_{it} (or Δx_{it}) is uncorrelated with η_i.

[42] For general results, see Nelson and Startz (1990a, 1990b) and Staiger and Stock (1997). For the case of dynamic panel data models, see Blundell and Bond (1998).

[43] In the special case with $\beta = 0$, Blundell and Bond (1998) show that the first-differenced GMM estimator of α has a serious downward bias in these cases.

have better small sample properties.[44]

Estimators with better properties can also be obtained in some contexts if one is able to impose more restrictive assumptions than those considered above. Arellano and Bover (1995) note that Δx_{it} may be uncorrelated with the unobserved firm-specific effects, even when the level of x_{it} is correlated with η_i. Combined with the assumption of limited serial correlation, this allows suitably lagged values of the first-differences $\Delta x_{i,t-s}$ to be used as additional instruments for the equations in levels. These additional moment conditions are over-identifying restrictions that can be tested, and where they are valid they can significantly improve on both the asymptotic and small sample properties of the first-differenced GMM estimator, particularly when the parameters (α, β) are only weakly identified from the first-differenced equations. Blundell and Bond (1998, 2000) note that if the initial conditions y_{i1} also satisfy the stationarity restriction $E[\Delta y_{i2}\eta_i] = 0$ then suitably lagged values of $\Delta y_{i,t-s}$ as well as $\Delta x_{i,t-s}$ are available as instruments for the levels equations.[45] The resulting linear estimator, which combines equations in levels with equations in first-differences and which they label 'system GMM', is shown to provide dramatic gains both in asymptotic efficiency and in small sample properties, compared to both the linear first-differenced GMM estimator and to the non-linear GMM estimator of Ahn and Schmidt (1995).

Another potentially important issue in the context of panel data on companies or plants is that of non-random entry and exit. Entry into a panel of establishments may reflect the decision by a firm to enter a particular market, whilst entry into a panel of firms may reflect other economic choices such as the decision to ob-

[44] Again for the special case with $\beta = 0$, Ahn and Schmidt (1995) show that the gain in asymptotic efficiency is largest in the cases where α is only weakly identified from the linear moment conditions (4.6) - i.e. when α is high and T is small.

[45] In the special case with $\beta = 0$, this initial condition restriction requires that the series $(y_{i1}, ..., y_{iT})$ have a constant first moment for each firm. In the multivariate context, Blundell and Bond (2000) show that a sufficient condition is for the series (y_{it}, x_{it}) to both have stationary means. However this is not necessary. For example, both y_{it} and x_{it} may have non-stationary means provided that the Δx_{it} series is always uncorrelated with η_i and the conditional model (4.5) has generated the y_{it} series for a sufficiently long pre-sample period.

tain a stock market listing. Exit from a panel of plants may reflect plant closure or acquisition, whilst exit from a panel of companies may reflect mergers, takeovers and bankruptcies. Many of these economic events are potentially correlated with shocks that affect investment and employment decisions.

Non-random entry into the sample does not present a serious estimation problem, however, since the entry decision is a function of variables dated at the time of entry, which can be regarded as fixed over the subsequent sample period. The variables which determine entry may be correlated with the firm-specific effects in factor demand equations, but this fixed correlation merely shifts the firm-specific effects and can be controlled for using the estimation methods described above. The determinants of entry may also be correlated with subsequent shocks to the factor demand equation if the latter are serially correlated, but this correlation can be controlled for by dropping a limited number of initial periods.

Non-random exit from the sample could cause a potentially more serious attrition bias, that would only be corrected by controlling for firm-specific effects under very restrictive assumptions. This would remain an issue whether estimation is based on the 'unbalanced' panel, including those firms which exit, or the 'balanced' panel containing only the subset of firms which survive through to the end of the sample period. The development of tests and controls for attrition bias in panels of firms and plants is an important area for future research, although it should be recognised that this may not be a purely statistical issue. For example, in principal-agent models of company behaviour with incomplete contracts, where managers retain some flexibility not to pursue shareholder value-maximisation, it may be the case that investment decisions are directly influenced by the risk of being taken over or going bankrupt, since managers who dislike these events may choose actions to reduce their risk. Thus the nature of the exit processes may affect the specification of factor demand models, and not only the estimation methods required for consistent estimation.

5. Data

The advantages of microeconomic data sets are that data is often available for a large number of individual firms or plants, and aggregation problems are generally reduced compared to industry level or aggregate data sets. Moreover it is possible to move beyond a representative firm framework, and to test models that imply heterogeneous behaviour across firms, as for example may occur in the presence of financing constraints. The disadvantages are that the available measures of factor inputs and outputs are often crude, and key variables like factor prices are generally not measured at the firm or plant level. This is a major reason why much of the microeconometric literature on investment and employment has focused on issues such as the nature of adjustment dynamics and the presence of financing constraints, which can be investigated without microeconomic variation in factor prices.

Until very recently, most microeconometric factor demand studies relied on publicly available company data sets, generally obtained from company accounts. Examples include Compustat data for US firms and Datastream data for UK firms. Company accounts are not produced for the benefit of econometric research, and the measures that are directly available are often inappropriate for testing economic models. Measures based on recorded cash flows are likely to be most reliable, but even these can present problems for econometric research. For example, whilst data on sales may be accurate, there may be too little information on changes in inventories to infer the value of production. Similarly there may be insufficient information on the cost of current inputs to infer value-added. However flow measures that are based on changes in accounting valuations of assets and liabilities present more severe measurement issues, as does the use of the book values of these stocks themselves.

These book values of assets and liabilities are generally based on historic cost valuations, which may deviate substantially from current economic values in the case of long-lived capital assets and long-term debts. The historic cost valuation of the firm's capital stock is based on the prices at which assets were originally purchased, and so neglects both general price inflation and relative price changes over the intervening period. A related concern affects the depreciation deductions reported in company accounts. Even in commercial accounts, these may be based on cautious assumptions about the length of useful asset lives, and in some cases the only information available is based on depreciation rules required for tax purposes. For these reasons economic researchers have often preferred to construct their own capital stock estimates, based on cumulating the observed investment flows in a perpetual inventory formula that can allow for inflation and alternative estimates of economic depreciation rates. A similar problem affects the valuation of inventories, although here it is important to know the valuation method that has been used to construct the company accounts, which varies across countries and may vary across firms within the same country. The historic cost valuation of debt is based on the amounts borrowed and not yet repaid. The market value of the firm's outstanding debt may be different if interest rates have changed over the intervening period. Again economic researchers have sometimes preferred to construct alternative estimates of the market value of debt, although this is problematic without knowledge of the maturity structure. Many of these valuation problems come together in accounting measures of profit, with some of the principal concerns being the deduction of nominal rather than real interest payments, the inclusion of inflationary gains on holdings of inventories, and the use of historic cost depreciation charges.[46] Not to mention the possibility that the timing of some charges against profits may be manipulated to manage the release of news about the company to the financial markets.

Of particular concern in the factor demand context is the poor quality of

[46] See Edwards et al. (1987) for a comprehensive discussion of accounting measures of profit.

accounting information on the use of various inputs. For example, company accounts may report expenditures on direct purchases of fixed capital, but may give little information on the breakdown of these expenditures by type of asset, or may provide little information on the value of fixed capital obtained through the acquisition of other firms, or may provide little information on the value of fixed capital sold or scrapped. Thus the available data on investment may be subject to measurement errors whose importance may differ across sectors and over the business cycle. As we discussed in section 3.5, the absence of good data on disposals of capital may be particularly important when testing for asymmetries in upward and downward adjustments. The available data on employment in company accounts is often limited to a snapshot of the number of employees: it is comparatively rare to have information on hours worked, skill composition or flow measures of hiring and separations, all of which would be desirable when investigating the structure of labour demand. In company accounts data it is also unusual to have data on the usage of factors other than capital and labour.

A distinct source of micro data is provided by the large establishment level data sets that are compiled by government statistical agencies in order to estimate the aggregate levels of production, investment and employment in different sectors of the economy. Typically these are populations of all establishments above a certain size threshold, and stratified random samples of smaller establishments. An establishment may comprise one or more plants, which may account for all or part of a firm's activities. In the last ten years, these data sources have increasingly been made available for empirical research. Examples include the Longitudinal Research Database (LRD) in the US; the Annual Respondents Database (ARD) in the UK; the Norwegian data used by Anti Nilsen and Schiantarelli (1998); and the French data underlying the Abowd, Kramarz and Margolis (1999) study.[47]

There are several advantages of these establishment data sets over company

[47] See Abowd and Kramarz (1999) for more details.

accounts data. First, they provide data at a more disaggregated level, with the unit of observation generally being a plant or plants in the same geographical location. The LRD also provides quarterly observations. Secondly, these data sources provide far more coverage of the activities of smaller firms than do most company databases, which are often limited to publicly traded firms. Thirdly, the establishment data usually disaggregate information on factor inputs to a greater degree, allowing some consideration of multiple types of capital and labour. On the other hand, some variables may only be measured at the level of the firm, such as stock market valuations and tax payments. Coverage may also be limited to manufacturing or production industries, excluding the service sector which accounts for a large and growing share of investment and employment in developed economies.

As we have stressed in this chapter, a major problem facing microeconometric research on factor demand is the absence of comprehensive micro data on factor prices. This is partly a conceptual problem and not just a data limitation: if each factor of production were bought and sold in a single competitive market, then all firms would face the same price for each factor, and there would be no cross-section variation in factor prices that could be used to identify factor demand parameters. In practice there may be regional and sectoral differences in factor prices, although this is less useful in the context of large firms which may operate in several locations and industries. Thus considerable ingenuity is typically required to identify compelling sources of exogenous price variation. Examples from the investment literature include variation in the effects of taxes across different firms, either because they use different mixes of capital subject to different tax treatments, as in Auerbach and Hassett (1992), Cummins et al. (1994) and Chirinko et al. (1999); or because they are affected differently by non-linearities in the tax system, as in Devereux (1989) and Devereux et al. (1994). Examples from the labour demand literature include regional variation in minimum wage legislation, and variation in the extent of unionisation.

Our final remarks on sources of microeconomic data for factor demand studies concern different levels of aggregation at which data is available. Large companies may own many subsidiary firms, so that even with company accounts data there is an important distinction between the *unconsolidated* accounts reported by individual subsidiaries, and the *consolidated* accounts reported by the group as a whole. Individual firms may also operate multiple plants, either in different locations or performing different activities. Data at different levels of aggregation may be most suitable for addressing different questions. For example, if the objective is to uncover the structure of the adjustment cost technology, then it seems appropriate to use the most disaggregated data available, to avoid the tendency for discrete adjustments to be smoothed by aggregation across plants or types of capital or labour.[48] As noted previously, commonly used micro data on companies or large plants may still be too aggregated for this purpose. However if the objective is to investigate the impact of financing constraints, it may well be appropriate to consider consolidated data for the company as a whole: even in the presence of a financing constraint, the spending by an individual subsidiary firm or plant may not be constrained by its own cash flow, since the company can reallocate financial resources between different parts of the group. More generally, the relevant locus for at least some aspects of corporate decision making may be at the level of the firm rather than the plant, and it is worth noting that value-maximising behaviour at the level of the company need not imply value maximisation for individual subsidiaries or plants. There are also likely to be important advantages from combining both establishment and firm level data, for example to investigate the impact of aggregation on investment and employment dynamics. Finally, as we noted in the introduction, important aspects of aggregate adjustment may take place through the entry and exit of individual firms

[48] See, for example, Hamermesh (1989) and Anti Nilsen and Schiantarelli (2003).

or plants, so evidence from microeconomic data will not necessarily provide the correct answers to macroeconomic questions.

6. Topics in investment

6.1. Some basic findings

Many studies have used company panel data to evaluate the Q model of investment. This model appeared particularly well suited to company data sets, since stock market valuations are readily and accurately measured for companies with publicly traded shares, and in contrast to the user cost of capital, there is rich cross-section variation in the average q ratio that should help to identify the model. There was also initial optimism that some of the apparent failings of the Q model that had been reported in earlier studies using aggregate data may be due to aggregation problems that arise when investment rates and average q, both specified in ratios, are constructed using aggregate data.

In fact, as we noted in section 3.1, most of the empirical problems found with the aggregate data have been reproduced in microeconometric studies. These include very low coefficients on the Q variable, suggesting incredibly high marginal costs of adjustment, and violation of the prediction that Q should be a sufficient statistic for investment. In most micro studies, additional variables such as cash flow or sales have been found to be informative after controlling for Q, and in some cases the Q variable becomes insignificant when these other variables are added to the empirical model. Similar findings have been reported independently using data for a wide variety of countries and time periods, and this has also been the case in the relatively small number of studies that have recognised the potential importance of endogeneity and (transient) measurement error in average q. For example, both Hayashi and Inoue (1991) and Blundell et al. (1992) have used first-differenced GMM estimators of the type described in section 4.2 to estimate versions of the Q model, for panels of Japanese and UK listed manufacturing companies respectively. Hayashi and Inoue (1991) consider a version of the Q model that allows for multiple capital goods as well as the basic specification, whilst Blundell et al. (1992) estimate a version of the Q model that allows for an $AR(1)$ component in the error term. Both papers report very low coefficients on their Q variables, and find either cash flow or sales terms to be highly significant additional regressors, even allowing these variables to be endogenous and correlated with firm-specific effects.

One potentially important exception to this general pattern of results is provided by Cummins et al. (1994), who focus on periods around major tax reforms in the US and report much higher coefficients on their Q variable in these years. This is consistent with their interpretation that major tax reforms provide quasi-experiments that help to identify the effects of economic 'fundamentals' on investment, so that during these periods fluctuations in measured Q are dominated by informative changes in tax parameters rather than uninformative measurement errors. Notice that if this interpretation is correct, the conventional findings discussed in the previous paragraph can only be explained if there are substantial and highly persistent measurement errors in average q that are not easily controlled for by the use of lagged instruments.

Two recent papers have developed alternative approaches to estimating the Q model which take seriously this possibility of persistent measurement error in average q. Erickson and Whited (2000) consider a GMM estimator based on higher order moment conditions. Their approach can allow for persistent measurement error in the stock market valuation as a measure of the firm's fundamental value, for example as a result of asset price bubbles, provided that the difference between the two values is independent of the fundamental value. They find that additional cash flow variables are no longer significant when this form of measurement error is allowed for in their sample. Bond and Cummins (2001) consider identification of the Q model in the presence of share price bubbles. They show that the parameters may not be identified, using conventional measures of average q, in the more problematic case where the bubble component is itself correlated with the fundamental value of the firm. To deal with this case, they consider using a direct estimate of the firm's fundamental value, based on forecasts of future profits published by securities' analysts.[49] When using this alternative measure, they find a much higher coefficient on their average q variable than is usually obtained, and neither cash flow nor sales variables are found to be significant.

Gilchrist and Himmelberg (1995) estimate a version of the Abel and Blanchard (1986) model, based on equation (3.3), for a panel of listed US manufacturing companies. Unfortunately it is not clear whether the model they estimate has a structural interpretation, since in their specification of $\left(\frac{\partial \Pi_t}{\partial K_t}\right)$they assume that net revenue Π_t is homogeneous of degree one in K_t, although their specification of adjustment costs depends on current investment and is not homogeneous of degree one in K_t. Certainly in the standard formulation of the Q model, net revenue is homogeneous of degree one in the pair (I_t, K_t), and not homogeneous of degree one in K_t alone.[50] Nonetheless their results are interesting in that they also find a much larger coefficient on their constructed measure of marginal q than on a standard measure of average q, and conditional on this measure of marginal q they also find smaller and weaker coefficients on a cash flow variable for at least some sub-samples of their data. These findings are also suggestive of a potentially severe measurement error problem for conventional measures of average q constructed using stock market valuations.

Results based on the Euler equation approach have been mixed. Unrestricted

[49] See also Cummins, Hassett and Oliner (1999) and Bond and Cummins (2000).

[50] Gilchrist and Himmelberg (1995) use the same form for adjustment costs as that in equation (3.4) - see their equation (7). Of course the formulation for $\left(\frac{\partial \Pi_t}{\partial K_t}\right)$ based on the assumption that Π_t is homogeneous of degree one in K_t may provide a good *approximation* if adjustment costs are sufficiently small. Abel and Blanchard (1986) also used this approximation in one of their applications to time series data, but they were careful not to claim that they were identifying a structural adjustment cost function. See their footnote 5.

estimates of investment dynamics have generally been difficult to reconcile with the Euler equation implied by the symmetric, quadratic adjustment costs model, in the sense that estimated coefficients on leads or lags of the investment rate have not implied plausible values for the discount factor (cf. equation (3.13)). Tests of the over-identifying restrictions implied by the Euler equation have often been rejected, at least for large sub-samples of the data. On the other hand some papers have reported more reasonable estimates of the structural parameters and non-rejection of the over-identifying restrictions for particular sub-samples, and suggested that these results are consistent with a financing constraints interpretation. See Gilchrist (1991), Whited (1992), Bond and Meghir (1994) and Hubbard et al. (1995) for examples of this approach.

One conclusion from this literature seems to be that the standard implementations of structural models based on the assumption of symmetric, quadratic costs of adjustment do not provide an adequate characterisation of the observed investment data, at least for a large part of the company data sets that have been used. Recent research in the Q framework suggests that measurement error in stock market valuations, as measures of the expectations of future profits relevant for investment decisions, should be taken seriously - reflecting the intrinsic difficulty of controlling for *firms'* expectations of future conditions. However it would be too early to conclude that this is the only source of mis-specification. As we emphasised in section 3.1, there are many candidate explanations for empirical rejections of these models. These results have motivated the huge empirical literature on financing constraints that we consider in the next section, as well as the more recent empirical work on non-convex adjustment costs that we discussed in section 3.5.

6.2. Financing constraints

A major topic of interest in recent microeconometric research on company investment has been to test for the possibility that investment spending is subject to significant financing constraints. In each of the basic investment models outlined above, capital markets were assumed to be perfect in the sense that the firm can raise as much investment finance as it desires at some required rate of return (ρ_{t+1}) that is given exogenously to the firm. In this case the firm's real investment decision is separable from its financial decisions, and investment depends only on the price (i.e. the required rate of return) at which finance is available. Quantitative indicators of the availability of internal finance, such as current profits or cash flow, should affect investment only to the extent that they convey new information about its likely future profitability; and if the maintained structure of the Q model were correct, these financial variables should not appear as significant explanatory variables in an investment model after controlling for a measure of (marginal) q.

This separability between real and financial decisions no longer holds if the firm faces 'imperfect' capital markets, in which internal and external sources of investment finance are not perfect substitutes. We define a firm's investment to be *financially constrained* if a windfall increase in the supply of internal funds (i.e. a change which conveys no new information about the profitability of current investment) results in a higher level of investment spending. Clearly firms are not constrained in this sense in the Q model of equation (3.2), where given current prices and interest rates, investment depends only on the current and expected future marginal revenue products of capital, as summarised in marginal q through the shadow value of an additional unit of capital. However firms' investment may be financially constrained in 'hierarchy of finance' or 'pecking order' models of corporate finance, in which external sources of finance (for example, from new share issues or borrowing) are assumed to be more expensive than internal sources of finance (for example, from retained earnings).[51]

6.2.1. A simple hierarchy of finance model

To illustrate the implications of this assumption rigorously, but as simply as possible, we maintain all the assumptions used to obtain the Q model, except that we introduce an additional cost associated with using external finance. We continue to assume that the firm issues no debt, pays no taxes and is characterised by symmetric information, but we introduce an explicit transaction cost (f_t) per unit of new shares issued. Similar results can be obtained in models where the cost premium reflects asymmetric information or differential taxes, and can be extended to models with debt provided that lending to the firm becomes a risky proposition for lenders beyond some level of debt (i.e. there is some risk of default, all debt is not fully collateralized, and there are 'deadweight' costs associated with defaulting on unsecured debt).[52]

Recognising the distinction between dividends paid (D_t) and the value of new shares issued (N_t), the value of the firm's equity is given by the expected present value of net distributions to shareholders, as

$$V_t = E_t \left[\sum_{s=0}^{\infty} \beta_{t+s} \left(D_{t+s} - N_{t+s} \right) \right] \tag{6.1}$$

whilst the sources and uses of funds identity links dividends and new share issues to the net revenue (Π_t) generated in period t, according to

$$D_t = \Pi_t + (1 - f_t) N_t. \tag{6.2}$$

Introducing non-negativity constraints on dividends and new share issues, with as-

[51] See, for example, Myers (1984). Notice that this assumption does not require rationing to be present in any of the external capital markets, although it can incorporate rationing (cf. Stiglitz and Weiss (1981)) as a special case in which the cost of external funds becomes infinitely high.

[52] See Hayashi (1985a) and Bond and Meghir (1994) for extensions to models with debt.

sociated shadow values (ν_t^D) and (ν_t^N), the firm's optimisation problem becomes[53]

$$V_t(K_{t-1}) = \left\{ \max_{I_t, L_t, M_t, N_t} \begin{pmatrix} \Pi_t(K_t, L_t, M_t, I_t) - f_t N_t + \nu_t^N N_t \\ +\nu_t^D \left[\Pi_t(K_t, L_t, M_t, I_t) + (1-f_t) N_t\right] \\ +\beta_{t+1} E_t \left[V_{t+1}(K_t)\right] \end{pmatrix} \right\}. \qquad (6.3)$$

The first-order condition for optimal investment becomes

$$-(1+\nu_t^D)\left(\frac{\partial \Pi_t}{\partial I_t}\right) = \lambda_t \qquad (6.4)$$

and the Euler equation for λ_t becomes

$$\lambda_t = (1+\nu_t^D)\left(\frac{\partial \Pi_t}{\partial K_t}\right) + (1-\delta)\beta_{t+1} E_t\left[\lambda_{t+1}\right]. \qquad (6.5)$$

In addition we now have a first-order condition for optimal new share issues, which for $N_t > 0$ gives

$$(1+\nu_t^D) = \frac{1}{1-f_t} \qquad (6.6)$$

and for $D_t > 0$ gives $\nu_t^N = f_t$.

Assuming perfect competition and the same expression for net revenue as in (2.25), the first-order condition for investment becomes

$$\left(\frac{\partial G}{\partial I_t}\right) = \left(\frac{q_t}{1+\nu_t^D} - 1\right)\frac{p_t^K}{p_t} \qquad (6.7)$$

where $q_t = \frac{\lambda_t}{p_t^K}$ is marginal q, as before.

This model has three distinct financial regimes. Retained earnings are the cheapest source of finance, so if the firm has sufficient earnings to finance its desired investment, it will issue no new shares. In this case the non-negativity constraint on dividends is not binding, and the shadow value of an additional unit of internal finance (ν_t^D) is zero. In this regime the basic Q model given by equation (3.2) describes the firm's investment.

If the firm does not have sufficient earnings to finance its desired investment, the non-negativity constraint on dividends is binding, and the shadow value of

[53] Notice that if $f_t = 0$, these non-negativity constraints are redundant. The problem (6.3) reduces to that considered in sections 2.2 and 3.1, and the firm's financial policy is indeterminate. This is a manifestation of the Modigliani-Miller (1958, 1961) irrelevance theorems.

internal funds is strictly positive. In this case the firm has to decide whether or not to finance additional investment by using the more expensive external source of finance. If the investment projects that would be foregone by not issuing shares are sufficiently profitable compared to the higher cost of external funds, the firm will choose to issue shares, and using (6.6) its investment in this regime will be described by

$$\left(\frac{\partial G}{\partial I_t}\right) = ((1 - f_t)q_t - 1)\,\frac{p_t^K}{p_t}. \tag{6.8}$$

However if the investment projects foregone by not issuing new shares are not sufficiently profitable to warrant paying the higher cost of external funds, the firm will be in a financially constrained position, in which both dividends and new share issues are zero. From the sources and uses of funds condition (6.2) and the net revenue function (2.25), the level of investment expenditure is constrained to the level of cash flow (i.e. $p_t\,(F_t - G_t) - w_t L_t - p_t^M M_t$). Thus in this constrained regime, windfall changes in cash flow have a direct effect on the level of investment, holding marginal q constant. Allowing the firm to borrow will tend to weaken this sensitivity of investment to windfall fluctuations in cash flow, but will only eliminate it in the special case where debt acts as a perfect substitute for finance from retained earnings.[54]

These results are illustrated in Figure 1 <Figure 1>, which is adapted from Hayashi (1985a). Investment rates (I/K) are shown on the horizontal axis, and marginal adjustment costs are assumed to be linear in the investment rate, as in (3.4). Values of $\left(\frac{q_t}{1+\nu_t^D} - 1\right)\frac{p_t^K}{p_t}$ are shown on the vertical axis, and the diagram is drawn for a given level of marginal q. The firm can finance investment rates up to (C/K) from internal funds. If the firm with marginal q illustrated has marginal adjustment costs given by the schedule G_I^1, it will choose the investment rate $(I/K)_1$, pay strictly positive dividends and issue no new shares. If the firm has much lower marginal adjustment costs given by the schedule G_I^3, it will choose

[54]See Hayashi (1985a) or Bond and Meghir (1994).

the investment rate $(I/K)_3$, financed partly (and, at the margin, completely) by issuing new shares, and pay zero dividends. However if the firm faces the intermediate marginal adjustment cost schedule G_I^2, it neither issues shares nor pays dividends, and investment is constrained at the rate (C/K). In this position, a windfall increase in cash flow that allows the firm to finance investment rates up to (C'/K) from internal funds does indeed cause an increase in the firm's investment rate, holding marginal q constant.[55] Whilst we have illustrated these financial regimes by considering different levels of adjustment costs for a given level of marginal q, the same conclusions can be reached by considering different levels of marginal q for a given schedule of marginal adjustment costs.

This model indicates that the simple relationship between investment and marginal q described by equation (3.2) or (3.5) no longer holds in the presence of financing constraints. Given the maintained structure of the Q model, a simple test of the null hypothesis that there are no financing constraints can be obtained by including additional financial variables, such as cash flow, on the right hand side of (3.10). Under the null the estimated coefficients on additional financial variables should be insignificantly different from zero,[56] whilst under the alternative financial variables will be informative about investment if some firms in the sample are in a financially constrained position. Fazzari et al. (1988) and many subsequent papers have exploited this property of the Q model to develop 'excess sensitivity' tests for the importance of financing constraints on firms' investment spending.[57]

[55] Given our timing convention, with current investment immediately productive, this experiment can be thought of as resulting from some exogenous grant or 'helicopter drop' of money to the firm; changes in cash flow resulting from changes to current prices would generally also affect the profitability of current investment. In richer specifications, windfall changes in cash flow could arise from unrelated lines of business (Lamont (1997)) or from certain tax changes (Calomiris and Hubbard (1995)).

[56] Although even under the null, current financial variables are likely to be endogenous (e.g. current profits or cash flow would be affected by adjustment cost shocks), and this endogeneity should be allowed for when implementing such tests.

[57] Chirinko (1997) correctly notes that the hierarchy of finance model typically does not imply

It is worth stressing that it is the simple relationship between investment and marginal q that breaks down in the presence of financing constraints, not necessarily the equality between average and marginal q. In the simple model we have described here, we can show that the equality between average and marginal q is maintained, despite the presence of financing constraints, although this need not be the case in richer versions of the hierarchy of finance model. Combining the first-order conditions (6.4) and (6.5) and using the linear homogeneity of the net revenue function, as we did to derive equation (3.7), gives the expression

$$\begin{aligned}(1-\delta)\,\lambda_t K_{t-1} &= (1+\nu_t^D)\Pi_t\left(K_t, L_t, M_t, I_t\right) + \beta_{t+1}E_t\left[(1-\delta)\,\lambda_{t+1}K_t\right] \quad (6.9)\\ &= E_t\left[\sum_{s=0}^{\infty}\beta_{t+s}(1+\nu_{t+s}^D)\Pi_{t+s}\left(K_{t+s}, L_{t+s}, M_{t+s}, I_{t+s}\right)\right]\end{aligned}$$

which depends on current and expected future shadow values of internal funds. However it is straightforward to show that the relation

$$(1+\nu_t^D)\Pi_t = D_t - N_t$$

holds in each of the three financial regimes, so using (6.1) we obtain the same expression for marginal q as in equation (3.9), regardless of which regime the firm is currently in or expects to experience in the future. Thus in this model the basic average Q model (3.10) continues to characterise the investment of those firms that are currently paying positive dividends and issuing no new shares, and in principle the parameters of the adjustment cost function could be identified from this sub-sample. However the literature has focused on testing the null hypothesis that there are no significant financing constraints, for which this equality between average and marginal q under the alternative is not essential.

---

a linear relationship between investment rates, Q and financial variables, but this is not necessary to motivate these excess sensitivity tests. It is sufficient to show that the simple linear relationship between investment rates and Q is mis-specified under the financing constraints alternative.

6.2.2. Excess sensitivity tests

These tests are justified formally because under the maintained structure of the Q model, the Q variable summarises all the information about expected future profitability that is relevant for the current investment decision. This is an important advantage of structural investment models like the Q model or the Euler equation compared to reduced form investment equations in the context of testing for the presence of financing constraints. As we illustrated in section 3.6, significant effects from financial variables in a reduced form investment equation may simply reflect their role in forecasting future demand or profitability.

Nevertheless this distinction should not be exaggerated, since it relies heavily on all the assumptions that were used to derive the Q model, and becomes blurred once we recognise that these assumptions may be invalid. Thus average Q, for example, would not be a sufficient statistic for investment if the firm operates in imperfectly competitive markets or subject to decreasing returns to scale;[58] standard measures of average Q would not be a sufficient statistic if share prices are subject to bubbles or fads;[59] and the simple linear relation between investment rates and Q would be mis-specified if adjustment costs are not symmetric and quadratic. In all these cases, financial variables may contain additional information that helps to explain investment after controlling for a linear Q term, even under the null hypothesis of no financing constraints.

Fazzari et al. (1988) were certainly aware that adding financial variables to the basic Q model is a joint test of all the maintained assumptions of the model, and not simply the assumption of no financing constraints. For this reason, they proposed a test that exploits cross-sectional differences between firms in the relationship between investment and financial variables. The basic idea is that even if

[58] See Hayashi (1982), Schiantarelli and Georgoutsos (1990) and Cooper and Ejarque (2001), for example.

[59] See Blanchard, Rhee and Summers (1993), Galeotti and Schiantarelli (1994) and Bond and Cummins (2001), for example.

the Q model is mis-specified, it may be mis-specified for all firms in a similar way, so that *differences* in the estimated coefficients on additional financial variables in an investment-Q equation may be an indication of *differences* in the impact of financing constraints.

6.2.3. Sample-splitting tests

Formally this 'sample-splitting' test can be justified most easily in the following context. Suppose we identify one group of firms for whom the cost premium for external finance is likely to be negligible, and another group of firms for whom the cost premium for external finance may be high. Then if all the assumptions of the Q model were correct apart from the possibility that some firms face financing constraints, we should expect that additional financial variables are insignificant for the first sub-sample; but such financial variables may be significant for the second sub-sample if some of these firms do indeed face binding financing constraints.

This sample-splitting test is analogous to that found in the literature which tests for the effect of liquidity constraints on household consumption by investigating heterogeneity in the relationship between consumption and current income across high wealth and low wealth households.[60] Several problems with the test have been noted in the literature. The most favourable outcome would be if we found no evidence of mis-specification for the sub-sample of firms that were considered on *a priori* grounds not to face a cost premium for external finance, and evidence of excess sensitivity to cash flow[61] for the sub-sample of firms that were considered on *a priori* grounds to be potentially subject to financing constraints. Even in this case, it is possible that the result reflects some other source of mis-specification that is only relevant for the second group of firms. For example, Fazzari et al. (1988) use the dividend payout ratio as a sample-splitting criterion,

[60] See, for example, Hayashi (1985b) and Zeldes (1989).

[61] i.e. significant coefficients on cash flow variables conditional on Q, and controlling for the endogeneity of current cash flow.

arguing that firms facing a high cost premium for external finance will tend to choose a low dividend payout ratio. However it has been noted that their sub-sample of firms with a low dividend payout ratio also tend to be younger and smaller than average, and it may be that share prices are subject to greater pricing errors for this sub-sample. If that were the case, it could explain why average Q is less informative, and additional financial variables more informative, for that sub-sample, even if they are not subject to significant financing constraints.[62]

A second potential problem is that the sample-splitting criterion used may not be exogenous for the investment equations estimated. If the allocation of firms to a particular sub-sample is correlated with shocks to the investment equation, then estimation of the investment model on the endogenously selected sub-sample will be subject to a sample selection bias of the type discussed in Heckman (1979). This suggests that selection criteria based on current financial characteristics or current size may give potentially misleading results unless care is taken to control for the endogeneity of the selection.[63]

An important concern in practice has been the difficulty of finding any sub-samples for which there is no evidence of mis-specification of the basic structural model. The typical finding in studies based on the Q model has been that cash flow variables have significantly larger coefficients in the sub-samples that are considered more likely to be financially constrained, but that such terms also have coefficients that are significantly different from zero for the sub-samples that are considered less likely to be financially constrained. One interpretation is that the Q model is mis-specified for both sub-samples, but in different ways. Suppose that the Q model is mis-specified even for firms that do not face financing con-

[62] This point was noted by James Poterba in his Brookings Panel discussion of Fazzari et al. (1988). This particular problem could be avoided by using either an Abel-Blanchard model or Euler equation in place of the average Q model, but the general point remains that the effects of *other* sources of mis-specification of the basic structural model tested may not be common across different sub-samples of firms.

[63] This point was noted by Alan Blinder in his Brookings Panel discussion of Fazzari et al. (1988).

straints, perhaps because they have market power or non-convex costs of adjustment. This accounts for the significant effect of cash flow even for the sub-sample that is maintained not to be financially constrained. Provided the effects of this mis-specification are similar for the two sub-samples, however, the presence of financing constraints for one group of firms could plausibly explain a significantly higher coefficient for this sub-sample. The difficulty with this interpretation lies in establishing whether the effects of general model mis-specification are indeed similar for the two sub-samples. One possibility would be to investigate directly whether the current (or lagged) financial variables included in the investment equation are more informative predictors of future demand or profitability for one of the two sub-samples; if not, this would cast doubt on one of the leading alternative explanations for their differential importance in the investment equation.[64]

A different interpretation of the same finding is that both sub-samples are subject to financing constraints, but to differing degrees. Thus all firms may face a cost premium for external finance, but some firms may face a much higher cost premium than others. Other things being equal, a bigger transactions cost on new share issues would increase the probability that a firm finds itself in the financially constrained regime in the model we outlined above, and this would tend to increase the sensitivity of investment to fluctuations in cash flow. This interpretation seems reasonable in the simple model of financing constraints we have considered, and many papers have presented evidence of differential cash flow sensitivities as being consistent with the presumption of a higher cost premium for one sub-sample of firms. However there has been some recent controversy about this interpretation, with Kaplan and Zingales (1997) claiming that a higher cost premium for external finance may actually be associated with lower sensitivity of

[64] Gilchrist and Himmelberg (1995) investigated the forecasting role of cash flow in their study. Notice that this approach can be applied in the context of reduced form investment equations, as well as mis-specified structural models. See Bond et al. (1999) for an application using error correction models.

investment to cash flow.[65] To understand this Kaplan-Zingales critique, and its limitations, it is necessary to briefly consider more realistic models of financing constraints in which the firm has access to external finance by issuing debt as well as new shares.

6.2.4. The Kaplan and Zingales critique

A standard model with debt finance, and the one analysed by Kaplan and Zingales (1997), is illustrated in Figure 2 <Figure 2>.[66] In the absence of financing constraints, this is simply a representation of the first-order condition (2.8) *for a static model of investment*: the downward-sloping line represents the marginal product of capital in the current period,[67] which is equated with the user cost of capital. Given the firm's current cash flow, suppose that borrowing provides a perfect substitute for retained earnings up to the level of investment indicated by X, but becomes increasingly expensive at higher levels of borrowing - perhaps because there is an increasing risk of default and there are deadweight costs associated with bankruptcy. Firms wishing to invest more than X will again find themselves in a financially constrained regime, in which investment is sensitive to windfall fluctuations in cash flow. In this case, a windfall increase in cash flow would increase the level of investment that can be financed without resorting to more expensive debt, say from X to X'. This lowers the marginal cost of external finance for all levels of investment above X, and results in the firm optimally choosing a higher level of investment. Moreover it appears that firms facing a higher risk premium in the cost of borrowing will display greater sensitivity of investment to cash flow than firms facing a lower risk premium in the cost of borrowing: in Figure 2, a given windfall increase in cash flow has a greater impact

[65] See also Fazzari, Hubbard and Petersen (2000) and Kaplan and Zingales (2000).

[66] See, for example, Hubbard (1998). For rigorous treatments of debt finance in the presence of default risk, see Hayashi (1985a) and Bond and Meghir (1994).

[67] Given the capital stock inherited from the previous period, there is a one-to-one association between the current period's investment level and capital stock.

on the level of investment for a firm facing the cost schedule A than for a firm facing the cost schedule B.

Kaplan and Zingales (1997) have pointed out that this last conclusion depends on the presumed linearity of the marginal product of capital schedule, as illustrated in Figure 2, and need not hold under alternative assumptions about the production function. To illustrate this possibility, suppose that the marginal product of capital has the piecewise linear form shown in Figure 3 <Figure 3>. In this case, the sensitivity of investment to a given windfall increase in cash flow is greater for a firm facing the cost schedule B than for a firm facing the cost schedule A. More generally, the degree of the financing constraint faced by the firm, as measured by the slope of its cost of external funds schedule, cannot be inferred simply from the sensitivity of investment to cash flow.

This conclusion is certainly correct in the static model analysed by Kaplan and Zingales (1997), but several limitations of the result should be noted. First, the result is obtained under the alternative hypothesis that firms are indeed subject to important financing constraints, and does not undermine the basic excess sensitivity test of the null hypothesis of no financing constraints. As we discuss below, recent papers that have not relied on the average Q model and have controlled for the endogeneity of current cash flow have been more successful in finding some sub-samples of firms for which there is *no evidence* of excess sensitivity to cash flow, whilst the same methods yield significant evidence of excess sensitivity for other sub-samples of the data. Secondly, the Kaplan-Zingales result does not itself provide any alternative explanation for the common empirical finding that there are stronger effects of cash flow on investment for some types of firms than for others; nor does their result rule out the *possibility* that these differential cash flow sensitivities could indeed reflect differences in the severity of financing constraints. Thirdly, their result is obtained in a static model with no adjustment costs, and depends on the curvature of the marginal product of capital schedule. It is not clear that their result extends to a model with strictly convex adjustment costs, in which the first-order condition characterising optimal investment relates marginal adjustment costs to marginal q, as in equation (3.2), rather than equating the marginal product of capital to the user cost. In particular, it is not clear that there is a similar result in the model with symmetric, quadratic adjustment costs, in which marginal adjustment costs are a linear function of the investment rate. This limitation is potentially important, since much of the empirical work in this area, including that reported by Kaplan and Zingales (1997) themselves, is based on the assumption of symmetric, quadratic adjustment costs, and not on a static investment model. Finally, their result is obtained in a model where the cost of external funds is increasing at the margin. This may be a reasonable assumption if debt is the only source of external finance. However in models where the firm can issue new equity as well as debt, and in which the cost of new equity finance is above the cost of internal finance but is not increasing at the margin, the probability of the firm finding itself in the financially constrained regime again depends on the cost premium for using new equity finance - as was the case in the simple model we discussed in section 6.2.1 above.[68] In these models, a higher sensitivity of investment to cash flow may reflect a higher cost premium for using new equity finance simply because the firm is more likely to be in the financially constrained regime, whether or not investment also responds more to a given change in cash flow within that regime.

In summary, the Kaplan and Zingales (1997) critique is limited to the claim that differing cash flow sensitivities reveal different degrees of financing constraints under the alternative hypothesis that these types of firms are both subject to significant financing constraints; and whilst they present one model in which a greater sensitivity of investment to cash flow could be associated with a lower

[68] See Hayashi (1985a), Fazzari et al. (1988) and Bond and Meghir (1994) for examples of models of this type.

cost premium for external finance, it is not clear whether this result generalises to models with adjustment costs or a richer specification of the firm's financial policy. Nevertheless, even in the most favourable scenario - for example, in studies which find no excess sensitivity to cash flow for one sub-sample and significant excess sensitivity for another - this sample-splitting test would not establish that heterogeneity in the cost premium for external finance is the correct explanation for this difference. As we stressed in the previous section, it remains quite possible that the different relationship between cash flow and investment among sub-samples of firms can be explained by other sources of mis-specification in the basic structural models that have been used as the basis for these tests.

6.2.5. Empirical results

Comprehensive surveys of recent empirical work investigating the impact of financing constraints on company investment can be found in Hubbard (1998) and Schiantarelli (1996). As we have discussed, much of this literature has followed Fazzari et al. (1988) in considering excess sensitivity tests based on adding financial variables to the average Q model, and has tested for differences in the sensitivity of investment to financial variables between sub-samples of firms that are considered more likely or less likely to be affected by significant financing constraints *a priori*. Sample-splitting criteria that have been considered include dividend payout ratios (Fazzari et al. (1988)); firm size, age or growth (Devereux and Schiantarelli (1990)); the firm's credit rating (Whited (1992)); the dispersion in the firm's share ownership (Schaller (1993)); whether the firm is affiliated to a larger corporate grouping (Hoshi et al. (1991)); and whether the firm has a relationship with a particular bank (Elston (1993)). We do not attempt to review this extensive literature here, but rather focus on a selection of recent papers that have extended this basic methodology.

Gilchrist and Himmelberg (1995) present results based on a version of the

Abel-Blanchard (1986) model, as well as results for an average Q model. They also consider a range of possible sample-splitting criteria, including firm size, dividends and credit ratings. Their results are for a sample of publicly traded US manufacturing companies from Standard & Poor's Compustat database, over the period 1979-89. They also use a first-differenced GMM estimator of the kind described in section 4.2 above, to allow for the presence of unobserved firm-specific effects and the endogeneity of both current Q and cash flow variables. Interestingly, when they use their measure of marginal q, they find no significant effects from cash flow for their sub-samples of large firms, and firms with either a bond rating or a commercial paper rating; whilst for their sub-samples of small firms, and firms with no bond or commercial paper ratings, they find significant effects from cash flow conditional on their measure of marginal q (and also conditional on average q). Moreover, Gilchrist and Himmelberg (1995) confirm that these differences in the coefficients on cash flow in their investment equations cannot be accounted for simply by current cash flow being a better predictor of future profitability in some sub-samples than in others; in contrast, they show that the relationship between current cash flow and future cash flow is very similar across the different samples. Although we noted above that their model may not have a structural interpretation, their paper does provide some very interesting evidence on the relationship between investment and cash flow for different groups of US firms.

Bond et al. (1999) present results based on a reduced form error correction model, for samples of British and German manufacturing companies. They use a system GMM estimator (Blundell and Bond (1998)), which again allows for both firm-specific effects and the endogeneity of their current sales and cash flow variables. They also find no significant effects when cash flow terms are added to their basic error correction specification for investment by German companies, although they do find significant cash flow effects in the same specification for investment by British companies.[69] Again they confirm that this difference does not reflect current cash flow being a better predictor of future cash flow or future sales for the British sample. Bond et al. (1999) also find that, within their sample of UK companies, there is significantly less sensitivity of investment to cash flow for the sub-sample of firms that perform R&D. They suggest that this finding is consistent with a 'deep pockets' explanation for which firms participate in R&D, so that the R&D-performing firms are a self-selected sub-sample where financing constraints associated with the cost of external finance may be less significant.

Bond and Meghir (1994) present a more direct test of the empirical implications of the hierarchy of finance model. They note that this model predicts that the same firm may be financially constrained in some periods and not in others; and that the firm's current dividend and new share issuing behaviour should signal which financial regime the firm is currently in. Thus in Figure 1, for example, the firms in the financially-constrained regime should be paying zero dividends and issuing no new shares; whilst the firms in the unconstrained regimes should be either paying positive dividends or issuing new shares.[70] Using an Euler equation specification for publicly traded UK manufacturing companies, and a GMM estimator that allows for the endogeneity of current financial choices, Bond and Meghir (1994) find that excess sensitivity to financial variables is concentrated in periods when firms pay unusually low dividends (relative to their average payout ratios), and issue no new shares.[71]

Whilst most of the results in this literature appear to be consistent with the

[69] Other cross-country comparisons of microeconometric investment equations include Hall et al. (1999) and Bond et al. (2003).

[70] More generally, whilst most of the sample-splits that have been used in the literature are interpreted as indicating whether or not a particular sub-sample of firms faces a significant cost premium for external finance, those based on current (or average) dividend payout behaviour can be interpreted as indicating whether a particular sub-sample of firms is currently (or predominantly) in a financially-constrained position. The latter tests would continue to have power even if all firms faced potentially significant financing constraints.

[71] Other microeconometric studies based on the Euler equation approach include Gilchrist (1991), Whited (1992) and Hubbard et al. (1995).

possibility of significant financing constraints, at least for some types of firms in some periods, we have emphasised that these tests could also be detecting other sources of mis-specification in the underlying investment models.[72] It should also be noted that there are alternative models in corporate finance, such as Jensen's (1986) 'free cash flow' theory, that could potentially account for the excess sensitivity of investment to cash flow and other financial variables.[73] An important limitation of this literature is that it has not yet developed a convincing structural specification for company investment under the alternative hypothesis that some form of financing constraints is important. This is no easy task, since it would require both a model to allocate firms to different financial regimes, and a rigorous characterisation of optimal financial behaviour in the presence of bankruptcy costs and possibly asymmetric information. Nevertheless these developments will be important if we are to obtain more compelling evidence in favour of the financing constraints hypothesis, and to obtain useful models for policy simulation analysis in the presence of financing constraints.

6.3. Taxes and the user cost of capital

Compared to the voluminous literature on financing constraints and investment, there has been a dearth of microeconometric studies that focus on estimating the sensitivity of investment to changes in taxes, interest rates or other components of the user cost of capital. This does not reflect any lack of interest in the topic - the relationship between investment and interest rates is crucial to understanding the transmission mechanism of monetary policy, and the effects of taxes and subsidies

[72] A different possibility, stressed by Gomes (2001), is that cash flow variables may contribute little additional explanatory power in an investment equation, even in the presence of financing constraints. Although accounting for the insignificance of cash flow variables has not been the primary concern of this empirical literature to date, the possibility that these tests have very low power does seem to be present in the simulation model used by Gomes (2001).

[73] In the free cash flow approach, managers have non-value maximising objectives (which may include over-investment) and are subject to less effective monitoring when spending internal funds than external funds.

on investment are crucial for the evaluation of tax policies and potential tax reforms. The frequency of tax changes that are intended to influence investment behaviour suggests that policy makers believe in the effectiveness of this policy, but this view has not received overwhelming support from decades of econometric research based on aggregate time series data.[74]

The limited contribution of microeconometric research to date rather reflects the difficulty of measuring relevant cross-section variation in the user cost of capital. The risk-free nominal interest rate is common to all firms in the same country. Firms producing different products may experience variation in the own-price real interest rate, but measuring this variation requires time series on the prices charged by individual firms which are not widely available. There is more potential for measuring differences in risk-premia across firms, at least to the extent that the relevant risk-premia are well characterised by standard asset pricing models. This is one area where the increasing availability of high-frequency data on stock returns for individual firms may provide a promising direction for future research. Similarly it may be possible to exploit differences across firms in depreciation rates, although the accuracy of measured differences in accounting data may be questionable.

The focus in existing micro studies has generally been on measuring differences in the effects of taxes on the user cost of capital for different firms. One source of this variation is the asymmetry in most corporate tax systems between the treatment of profits and losses. The effective value of tax allowances is reduced for firms in a loss-making or 'tax-exhausted' position by the delay between incurring expenses and being able to claim tax deductions. Unfortunately it is not straightforward to identify which firms are currently in a tax-exhausted position from most publicly available data sources, and it is still harder to forecast when these

[74] See Chirinko (1993a), Hassett and Hubbard (1996) and Caballero (1999) for recent reviews of this evidence.

firms will resume paying taxes. Devereux (1989) and Devereux et al. (1994) have investigated the impact of this type of tax variation on the investment behaviour of UK firms.

Variation also arises because firms use a different mix of capital inputs, some of which receive a more favourable tax treatment than others. Thus firms which tend to use tax-favoured types of capital face a relatively low user cost of capital. Ideally we would want to use investment data disaggregated by type of asset, but even if this is not available it may be possible to exploit the resulting variation across firms or industries in the user cost of capital. Auerbach and Hassett (1992), Cummins et al. (1994) and Chirinko et al. (1999) have exploited differences in the composition of investment across US industries to measure variation in tax-related components of both tax-adjusted Q variables and the user cost of capital.[75] Cummins et al. (1994) estimate significant effects from the cross-section variation that occurs in this measure of the user cost in periods of major tax reforms, with an implied long-run elasticity of the capital stock with respect to the user cost between -0.5 and -1.0 (Hassett and Hubbard (1996)). Chirinko et al. (1999) report statistically significant but smaller estimates, around -0.25, although the estimated returns to scale implied by their reduced form models suggest some doubt about these findings.

Caballero et al. (1995) also exploit this source of tax variation to estimate long-run elasticities of the capital stock with respect to the user cost. Their econometric approach is quite different, in that they rely on co-integration methods to estimate a long-run relationship between the capital-output ratio and the user cost of capital, both in logarithms, as in equation (2.13) above. This is estimated using time series data for individual US plants in the Longitudinal Research Database, imposing the restriction that the elasticity is equal for all plants within each two-digit industry. Caballero et al. (1995) report estimates ranging from -0.01 to

[75] See also Cummins et al. (1996).

-2.0 across different sectors, with the average being about -1.0. This estimate is similar to that found using co-integration methods on aggregate manufacturing data by Bertola and Caballero (1994) and Caballero (1994).

Whilst it is certainly the case that these micro studies have found evidence consistent with taxes and the user cost of capital having an economically significant effect on capital intensities in the long run, it is perhaps a little early to agree with Hassett and Hubbard (1996) that there is a new 'consensus' on the size and robustness of this effect.

6.4. Uncertainty

The relationship between investment and uncertainty has attracted considerable interest in recent theoretical research, and has been investigated in some recent microeconometric studies. Renewed interest in this topic has followed from the development of the literature on 'real options', which stresses that the option of waiting to invest until more information has been revealed can be valuable.[76] If so, then extinguishing this option by investing today should be viewed as a cost. Moreover the value of this foregone option will be greater at higher levels of uncertainty, so that current investment may be deterred by a higher level of uncertainty.

This intuitive prediction contrasts with earlier results on the relationship between uncertainty and investment in the context of the Q model. Abel (1983) showed that a higher level of uncertainty would be associated with higher investment in the Q model, although this effect would of course be fully reflected in the behaviour of q, so that in this case there should be no additional effect of uncertainty variables after conditioning on q. Caballero (1991) reconciles these theoretical results by showing that much depends on the nature of the net revenue function and the adjustment costs facing the firm. By maintaining strictly

[76] See Dixit and Pindyck (1994) for an excellent introduction to this literature.

convex adjustment costs and a linear homogeneous net revenue function, the Q model rules out any value associated with the option of delaying investment. In contrast, real options become valuable when there are both non-convex costs of adjustment, such as (partial) irreversibility, and a concave net revenue function. Even then, the effects of a higher level of uncertainty on the average level of the capital stock in the long run are found to be ambiguous (Abel and Eberly (1999); Caballero (1999)), essentially because both investment and disinvestment actions may be deterred by these real option effects. As stressed by Bloom el al. (2001), however, a less ambiguous prediction is that a higher level of uncertainty will be associated with slower adjustment of the capital stock, and in particular with a smaller impact effect of demand shocks on current investment.

Empirical work has been limited by the difficulty of finding convincing empirical couterparts to the concepts of 'uncertainty' used in this theoretical literature. Leahy and Whited (1996) considered measures of uncertainty based on the volatility in stock market returns for publicly traded US firms. They found that investment rates were negatively related to these uncertainty measures in simple specifications, but that this effect became insignificant when they conditioned on a standard measure of average q. One concern with this kind of measure is that stock market returns may be subject to 'excess volatility', if share prices are indeed subject to bubbles, fads or other influences over and above firms' fundamental valuations. Nevertheless, Bloom et al. (2001) found a smaller impact effect of sales growth on current investment for publicly traded UK firms facing higher volatility in stock returns, although this is in the context of a reduced form, error correction model of investment with no explicit controls for the effect of expected future levels of profitability. A concern in this context is that lower uncertainty (stock market volatility) may be associated with greater optimism about the firm's future prospects. Guiso and Parigi (1999) address both these concerns using data from a specially conducted survey of Italian firms, which asked firms both about expected levels and perceived dispersion in future demand. They also found a weaker effect of expected demand on current investment for firms that perceived greater uncertainty about this future demand, although their results are based on a single cross-section of firms, and the time horizon of these expectational variables is quite short.

Whilst these papers present findings that appear to be consistent with the predicted effects of real options, the development of more convincing measures of uncertainty and more rigorous testing strategies will be required before we can be confident that these effects are statistically or economically significant. In view of the considerable theoretical and policy interest in this topic, this would seem to be a promising area for further research.

6.5. R&D investment

In modern economies a large proportion of firms' investment is in intangible assets. One of the most important of these are research and development (R&D) expenditures. Economists have long regarded technical change as the most important driver of economic growth, so much of the early microeconometric work on firm level R&D naturally focused on analysing the *effect* of R&D on measures of firm performance (such as productivity, firm market value or patenting activity).[77] R&D is generally cumulated into a 'knowledge stock' and then treated as one type of capital input amongst others. More recently there has been renewed interest in understanding the determinants of firms' R&D decisions, as growth theory re-focused on the endogenous decisions of firms to invest in R&D (e.g. Romer (1986); Aghion and Howitt (1992)). There are relatively few papers which look at the demand for R&D using the dynamic structural models discussed in section 3,[78] although clearly the results can be extended if we are prepared to

[77] See, for example, the collection of papers in Griliches (1984).

[78] Hall (1992, 1993), Harhoff (1998), Klette and Moen (1999) report some results for the Euler equation for R&D. Himmelberg and Peterson (1994) have a non-structural version of the Q

treat R&D symmetrically with physical investment. For example, equation (3.17) would have cross terms in the tangible and R&D capital stocks in addition to Q in the R&D equation.

A key issue here is whether many of the problems inherent with modelling fixed capital are just exacerbated in the case of R&D capital or whether they are qualitatively different. For example, R&D is typically a highly risky and uncertain investment, has large adjustment costs, enjoys many government subsidies (e.g. tax privileges and direct grants), and is subject to strategic gaming (e.g. patent races). But all these issues also arise with fixed capital. Perhaps altogether they add up to a difference in kind, but this is unclear *a priori*.

The main difference between fixed capital and R&D capital is probably in regard to externalities. R&D creates knowledge which is difficult to fully appropriate by the firm making the investment. The 'knowledge spillovers' to other firms create the fundamental public goods problem which gives a rationale for governments to subsidise R&D. Much of the literature on R&D and productivity is motivated by the idea that the social returns to R&D exceed the private returns and there is a large body of work on the empirical search for spillovers (see Griliches (1998) for a summary).

Another problem with comparing R&D to fixed capital is measurement. First, in building up a replacement cost capital stock measure there is typically a benchmark year when either (i) there was a survey of replacement costs (e.g. fire insurance values) or, more typically, (ii) historic cost stock data can be used. There is no equivalent for R&D so researchers are forced to make assumptions about the pre-sample growth of R&D along with an assumed 'knowledge depreciation' rate. A second problem is that disclosure of R&D in company accounts is far more limited than investment, especially outside the US. Typically, researchers have to deal with the fact that in many countries R&D is a voluntarily disclosed

model.

item in company accounts, and therefore subject to serious selectivity biases. Finally, what counts as R&D is less clear than physical investment as about 90% of reported outlays are current costs and 50% are wages and salaries.

There are large empirical literatures on the effects of firm size, product market structure and labour market institutions on R&D. Since these have been surveyed elsewhere[79] we focus on three issues here - VAR approaches, financial constraints and taxes.

6.5.1. VAR approaches

In light of these modeling and data difficulties, several authors take a vector autoregression (VAR) approach to examining R&D. Lach and Schankerman (1989) focus on unraveling the pattern of Granger causality between R&D and fixed investment by projecting the log of both current R&D and current investment against lagged investment, sales, R&D and other variables. They find that investment does not 'Granger-cause' R&D, but R&D does Granger-cause investment. Although they find corroborating evidence at the industry level in Lach and Robb (1996), others find very different results. Using British data, Nickell and Nicolitsas (1996) find that industry R&D (rather than firm R&D) predicts investment. Toivanen and Stoneman (1998) find the exact opposite result (investment predicts R&D and not *vice versa*). The atheoretical structure of the VAR is problematic here and the interpretation of the correlation pattern (even if it were robust) is difficult. In this context the paper by Pakes (1985) is more satisfactory as his application of dynamic factor demand theory does place more restrictions on the data (a three equation system for R&D, patents and market value). Even though the restrictions are not rejected by the data, it has proved harder to push the theory much further in this direction (e.g. Griliches et al. (1991)) as the framework is fundamentally driven by unobserved stochastic shocks which are only poorly

[79] See Cohen and Levin (1989) and Menezes-Filho et al. (1998)

tracked by the observables in the system.

6.5.2. Financing constraints and R&D

A key area of interest for R&D models is the role of financing constraints. It has long been recognised that the asymmetric information problems that lie at the heart of credit constraints may be more important for R&D investments. Uncertainty, lack of collateral and the danger of losing one's ideas to competitors make it likely that firms will rely on internal sources of finance for R&D more than for other types of investment. On the other hand, the larger adjustment costs for R&D make it unlikely that transitory cash flow shocks will have a very large impact on firms' R&D decisions. Indeed, to the extent that firms only participate in performing any R&D if they can be reasonably sure of not encountering financial constraints, we may see less sensitivity to cash flow by R&D performing firms than non-R&D performing firms.

Hall (1992) analyses a panel of large US manufacturing firms and finds that R&D is significantly correlated with cash flow using a variety of model specifications (reduced form and Euler equations). Himmelberg and Peterson (1994) find evidence that R&D is sensitive to cash flow for small US firms in the high tech sector. The evidence outside the US is less clear. Hall et al. (1999) use a bi-variate VAR approach to examine 'causality' patterns in samples of US, Japanese and French firms in high tech sectors. They find that the cash flow correlation is far stronger in the US than in the other countries. Mulkay et al. (2001) also find that the R&D-cash flow correlation is stronger in the US than in France. Bond et al. (1999) find no effect of cash flow on R&D in their samples of British and German firms. They do find, however, evidence that in Britain cash flow seems to matter for the decision to participate in R&D, whereas it has no effect in the German sample. So the upshot of these studies is that the influence of cash flow on R&D appears stronger in the Anglo-American countries than in Continental

Europe or Japan, subject to the concerns we discussed in section 6.2 above.

6.5.3. Tax-price of R&D

There is a wide variation over time and across countries in the user cost of R&D capital. This is driven by the special treatment of R&D for tax purposes - many countries have tax credits for R&D, super-deductions and accelerated depreciation schedules (see Bloom et al. (1998) for a survey). Since the tax rules for claiming these benefits often differ depending on corporation tax liabilities, size of firm, current and past R&D spending, region and industry, these tax rules imply that there is a cross sectional distribution of user costs facing firms in a given year. Hall (1993) and Hines (1994) use US firm panel data to investigate the impact of changes in the user cost on R&D. Dagenais et el. (1997) implement a similar methodology using Canadian company data. These authors uncover significant effects of the tax price on R&D, with a price elasticity of around unity in the long run.[80]

A motivation for these studies is that changes in tax policy may cause some exogenous variation in the price of R&D. Unfortunately, a problem with these studies is that the user cost cannot be taken as truly exogenous as the tax position of individual firms will depend on current shocks which could also influence their R&D decisions. Thus, one still has to use some kind of instrumental variable procedure of the kind discussed in section 4.2.

7. Topics in employment

The demand for labour is a particular case of the general model outlined in section 2 above. Labour demand is particularly interesting from a policy perspective - the social and political consequences of a 20% fall in the relative demand for less

[80] Hall (1993) finds larger long-run elasticicies than Dagenais et al. (1997). Bloom et al. (2002) also find long-run elasticities of around unity using macro data across eight countries.

skilled workers will have greater interest than a 20% fall in the relateve demand for less sophisticated capital equipment.

A popular approach here is a version of the 'reduced form' models discussed in section 3.6. For example, analogously to (3.24)

$$l_t = \alpha_1^L l_{t-1} + \beta_0^L l_t^* + \beta_1^L l_{t-1}^*. \tag{7.1}$$

Using (2.14) we have

$$l_t = \alpha_0^L + \alpha_1^L l_{t-1} + \beta_0^L y_t + \beta_1^L y_{t-1} - \beta_0^L \sigma(w-p)_t - \beta_1^L \sigma(w-p)_{t-1} \tag{7.2}$$

where $\alpha_0^L = (1-\alpha_1^L)\sigma \ln a_L(1-\frac{1}{\eta^D})$. This can also be re-written (assuming constant returns) in error correction form

$$\begin{aligned} \Delta l_t &= \alpha_0^L + \beta_0^L \Delta y_t - \beta_0^L \sigma \Delta(w-p)_t \\ &\quad + \left(1-\alpha_1^L\right)\left[(y-l)_{t-1} - \sigma(w-p)_{t-1}\right]. \end{aligned} \tag{7.3}$$

Again, equations of this form can be justified explicitly in a dynamic optimising framework under quadratic adjustment costs (Nickell (1985, 1986); Bresson et al. (1992)). Often, researchers (especially in the UK) have assumed a Cobb Douglas production function and substituted out output for capital. This has the advantage that it is more reasonable to treat capital as predetermined in labour demand equations than output which, in general, must be treated as endogenous. Versions of the Euler equation analogous to (3.11) have also been estimated, although these are less common than in the investment literature.[81]

There are several existing surveys on labour demand. We examine some topics of particular interest arising since the publication of Nickell (1986), Hamermesh (1993) and Hamermesh and Pfann (1996), which give a summary of the literature as it stood at the beginning of the 1990s.[82] The issues of exogenous factor

[81] See, for example, Machin et al. (1993), Meghir et al. (1996) and Alonso-Borrego (1998).

[82] We do not attempt to survey the literature which focuses on the adequacy of general equilibrium models of the demand for skills. Interested readers are referred to Heckman and Sedlacek (1985) for an example which examines the selection bias inherent in aggregate studies of wages and labour demand.

price variation, union bargaining, adjustment dynamics, gross and net flows, and heterogeneity by skill are discussed in turn.

7.1. Variation in wages

One important difference between the firm level study of investment and employment is that firm level data often has some information on wages. Typically, these are quite crude measures such as the average wage (wage bill divided by number of workers) but this is a major advantage over capital where variation in the cost of capital between firms has to be constructed by the econometrician as it is absent from firm accounts. Of course, some of the variation in the average firm compensation cost will be variation in the quality mix of workers in the firm (e.g. by skill, gender or ethnic groups) which is conflated with genuine changes in the price of labour facing the firm. Increasingly though, the availability of matched worker-firm panels is enabling researchers to improve their measurement of firm level wage rates.

What are the exogenous sources of variation in the price of labour facing firms? In many contexts changes in the institutional structure surrounding wage determination offer scope for instrumenting firm level wages. Union power, minimum wage changes and regional variation (due to partial labour immobility) offer a much wider range of possibilities than with investment. Unfortunately, when some of these institutionally determined variations in the price of labour are used to examine labour demand, the results have been mixed. The survey in Card and Krueger (1995), for example, illustrates that it is very difficult to ascertain any clear evidence of significant wage elasticities in minimum wage studies.

7.2. Union bargaining

A large sub-literature has developed in estimating employment equations to examine different models of union behaviour. The traditional model of unionisation keeps to the neoclassical framework where the firm chooses employment unilaterally. The union, however, has some influence over how the wage is set. The monopoly union model allows the union complete power to set the wage whereas the more general 'right to manage' model allows for genuine bargaining over the wage rate (see Pencavel (1991) or Booth (1995) for an extensive discussion of these models). These models have the convenient property that the basic structure of the static labour demand model still holds, but the wage will have some firm level variation due to differences in union power between firms.

It is well known that these models are not Pareto efficient and a set-up allowing the firm and union to bargain over both wages and employment can lead to utility gains for both sides (Leontief (1946)). A second class of 'efficient bargaining' models allows explicitly for such contracts. Ashenfelter and Brown (1986) pointed out that in this case employment will, in general, depend both on the bargained wage and the 'outside' wage (the income received in the event of a breakdown in bargaining). Thus the presence of the outside wage in an employment equation is potentially a test of 'efficient bargaining'. Further generalisations of these models are possible to allow for differential degrees of bargaining power in the wage decision and the employment decision (Manning (1987)) - this essentially means also including an extra term in union power in the employment equation (Alogoskoufis and Manning (1991)).

There are various criticisms of these approaches. For one the presence of the outside wage in the employment equation could be due to many reasons other than efficient bargaining, such as efficiency wages (Nickell and Wadhwani (1991) attempt to test between the general bargaining model and efficiency wages). Another criticism is that the testing procedure breaks down once we allow for forward-looking behaviour in the presence of adjustment costs, as the alternative wage can enter the dynamic employment decision rule (Machin et al. (1993)). Most serious, however, is that it is very difficult to measure what is actually the true alternative wage facing union bargainers. It is quite likely that the average regional or industry wage is an extremely poor measure of this (see MaCurdy and Pencavel (1986)). This applies even more to measures of union power. The rather inconclusive and fragile results in this literature are likely to stem from this basic problem.[83]

A small, closely related literature seeks to test the adequacy of the neoclassical model of employment determination by analysing behaviour across different ownership structures. Probably the best example of this is Craig and Pencavel (1994) who examine the differences between co-operatives and conventional firms within a single industry - the Pacific Northwest Plywood industry. They found that the standard static model based on profit maximisation was a reasonable description of employment behaviour for the conventional firms, but was quite inadequate for the co-operatives (for example, there was no effect of wages on employment in the latter group).

7.3. Dynamics of adjustment

Asymmetries in adjustment costs have been more of an issue in the labour context because of various regulations aimed at increasing firing costs (e.g. Pfann and Palm (1993); Pfann and Verspagen (1989); Burgess and Dolado (1989)). Most European countries place a large number of restrictions on the ability of firms to shed labour. Pfann and Verspagen (1989) argued for keeping the assumption of convex adjustment costs but allowing for asymmetries. They suggested an adjustment cost function of the form:

$$G(\Delta L) = \frac{1}{2}b(\Delta L)^2 - c\Delta L + e^{c\Delta L} - 1 \tag{7.4}$$

[83] See Card (1986), Abowd (1989), Christofides and Oswald (1991) and Boal and Pencavel (1994) for good examples of attempts to examine different models of union behaviour. Another basic problem is that very different models may apply in different industries and bargaining contexts.

where b and c are parameters in the adjustment cost function. If costs are symmetric then $c = 0$. If the marginal cost of a positive adjustment is greater than the marginal cost of a negative adjustment then $c > 0$. Substitution of this adjustment cost function into the net revenue function (2.25) leads to a non-linear Euler equation. This can either be estimated directly (Schiantarelli and Sembenelli (1993); Jaramillo et al. (1993)) or with approximation methods (e.g. Pfann and Palm (1993)). Studies using this approach have found evidence for asymmetries (see Hamermesh and Pfann (1996)).

In an influential paper, Hamermesh (1989) examined monthly data on 7 manufacturing plants. Although the aggregate series appeared smooth, individual plant level employment adjustment was extremely lumpy. Davis and Haltiwanger (1992) produced the first large scale analysis of plant level employment changes in the US. They emphasised the fact that during times of net employment expansion there were very large numbers of firms who were cutting employment, and *vice versa* in recessions. As with Hamermesh (1989), much of the plant level employment changes occurred over short periods.

One branch of the literature has gone on to focus on non-convex adjustment costs as a reason for this heterogeneity (see section 3.5 above). Caballero et al. (1997) take a similar approach to analysing plant level employment changes as they do for investment (the dataset they use is essentially the same, the Longitudinal Research Database). A crucial issue is how empirically to measure the 'gap' (or labour shortage) term; that is, the difference between the actual level of employment and the target/desired level of employment. They assume that hours per worker can be used to infer a reliable measure of this gap, or rather the difference between hours intensity in the current quarter and the plant's mean hours per worker over the sample period.[84] For a given target level of employ-

[84] An alternative method is to write down explicitly the firm's dynamic optimisation problem and use this to calculate the gap. This requires many assumptions about the parameters to calibrate the optimisation problem. See Caballero and Engel (1993, 1999) for examples of this

ment, a large change in employment will lower the size of the gap and therefore the deviation of hours per worker from its long-run mean. They therefore use the coefficient from a regression of the change in employment on the change in hours to build a measure of the gap and then characterise the degree to which a plant actually adjusts towards the optimal level of employment as a function of the size of this gap. They find that the 'adjustment hazard' is not constant as the partial adjustment model would predict but rather increases with the gap. That is, the probability (and proportionate size) of the adjustment increases with the scale of the shortage. They also find two modes of the distribution of employment changes, one at (practically) zero adjustment and another at full adjustment. They conclude that this is consistent with (S,s) types of adjustment behaviour.

One problem with this approach is that the OLS regression of employment changes on hours changes, which is critical in defining the gap term, is subject to endogeneity. Productivity shocks will increase the target level of employment and lead to simultaneous increases in jobs and hours, leading to a strong upwards bias in the relevant parameter. Caballero et al. (1997) attempt to deal with this by conditioning their regression sample only on observations where there have been very large changes in both employment and hours. They argue that in these periods, the changes in employment targets will be swamped by the effects of large changes in hours and employment. Cooper and Willis (2001) suggest that this approach may be misleading by analysing some simulated data generated from a purely convex adjustment cost model in a stochastic environment. In such an environment, the only periods of large employment and hours changes are exactly those in which there are large changes in the employment target levels. Cooper and Willis (2001) also suggest that (mis)measuring the gap using the Caballero et al. (1997) methodology on their simulated data produces the kind of non-linearities in the aggregate data that Caballero et al. interpreted as evidence

approach.

of non-convexities. However their simulation results are based on samples that cover more time periods than those used in empirical work, and the degree to which mis-specification of the gap can result in the appearance of quantitatively significant non-linearities remains controversial.

Cooper and Willis (2001) argue for a more explicit structural approach to address the issues of non-convex adjustment costs. An example of this is contained in Rota (1998) who uses data on Italian firms. These firms are on average smaller than the plants studied in Caballero et al. (1997). There is a large mode at (absolutely) zero adjustment and Rota (1998) uses this to define three regimes (adjust up, adjust down and don't adjust). The adjustment regimes are characterised by an Euler equation analogous to that described in Section 3.3 above and the selection rule into regimes is determined by an ordered probit. Estimation of the Euler equations requires using the estimated parameters from the ordered probit to correct for the endogenous selection into adjustment regimes. Apart from the usual problems discussed above, her study raises several additional issues. One worry is how the regime selection rule is separately identified from the employment rule (this problem also arises in Hamermesh (1989)). A second (and related) issue is the fact that the model is silent on the structural form of the selection rule which determines when an adjustment takes place. An attempt to explicitly implement a dynamic discrete model of employment adjustment is Aguirregabiria and Alonso-Borrego (1999),[85] who consider a model with linear (but not fixed) adjustment costs. The adjustment costs are amplified by the firing costs due to Spanish labour regulations and the authors examine the impact of a reduction in these for workers on temporary contracts. They estimate the productivity shock through a first stage production function and then use this explicitly in the dynamic optimisation problem for the employment decision in a second stage. They found important effects of the reform in improving employment and job mobil-

[85] See also Aguirregabiria (1997).

ity. Although this approach is an advance it does hinge critically on the correct specification and estimation of a production function to identify the technology shock.[86]

7.4. Gross vs. net flows

Hamermesh and Pfann (1996) emphasise that most studies of adjustment costs consider net rather than gross employment flows. A firm might have no net change in employment but hire 1000 workers and sack 1000 workers. It is likely that this firm will bear more adjustment costs than one which had no gross changes in employment at all.[87]

Information on turnover is rare in most micro datasets, but a few studies have started to examine the issue in more detail. Hamermesh et al. (1996) analysed a large sample of Dutch plants and decomposed aggregate gross flows. They found that firm level gross flows accounted for a substantial proportion of the total. Abowd et al. (1999) examined French data on the entry and exit of workers from firms and attempted to use information on the size of costs associated with the movements of workers. They found that there were very high fixed costs associated with firing workers and most adjustment was through varying the hiring rate.

7.5. Heterogeneous labour and skill-biased technical change

Most firm level datasets only have information on total employment. Aggregation over different types of labour can cause many problems, for example the appearance of spurious dynamics (Nickell (1986)). If there is access to data disaggregated by skill type, for example, a whole range of important questions are opened up. Most obviously there is the question of how the wage elasticity of labour demand differs between different groups of workers. Bresson et al. (1992), for example,

[86] For discussions of some of the many problems with estimating production functions with firm level data, see Griliches and Mairesse (1998) and Blundell and Bond (2000).

[87] Note that this is a different meaning of gross job flows than that found in the job creation and destruction literature discussed above.

estimated employment equations for 3 different types of labour in 586 French manufacturing firms - they found that the wage elasticities were greatest for the least skilled workers.

We have discussed the issue of multiple quasi-fixed factors in section 3.4 above and the analysis extends in a straightforward manner to multiple types of labour. Many authors have been interested in how adjustment costs might differ for different groups of workers: for example, are they higher for skilled than for unskilled workers? The issue of dynamic complementarity and substitutability has been considered (Nadiri and Rosen (1969)).

The debate over the (static) elasticities of complementarity and substitution between heterogeneous labour, capital and technology has been a topic of long standing interest to economists. Discussions have been enlivened in recent years by the rapid increase in the college-high school premium in the United States, Great Britain and many other countries. Many researchers have argued that this is primarily due to skill-biased technological change. We do not intend to review all the arguments here (see Autor and Katz (1999) for a survey) but we will focus on a more narrow set of questions. First we consider some of the meanings of technology-skill complementarity. Second, we critically consider methodologies for testing its size and existence. Finally we review some of the empirical results, particularly those based on micro data.

We examine the neoclassical analysis of skill-biased technical change initially in the context of the static factor demand model described in section 2.1. A closely related issue is capital-skill complementarity.[88] At issue here is the Hicks-Allen partial elasticity of substitution discussed above. One could clearly take the same approach to technology, viewing it as a form of (partially appropriable) 'knowledge capital' (Griliches (1998)) and modelling it as simply another form of

[88] Griliches (1969) is the pioneering paper here finding evidence in favour. Weiss (1977), by contrast, did not find consistent evidence across sectors using more disaggregated data by skill type.

capital input. The alternative approach is more traditional, treating technology as a free good available to all firms in the economy. Note that the key difference is that in the first approach technology is essentially a choice variable for the firm: one factor among others. It may have special features (such as non-rivalry as emphasised in the endogenous growth literature) but can be considered as a choice variable for the firm. In the second approach technology is 'manna from heaven'. It is exogenous and does not change with economic conditions (or at least is treated as such).[89] Both these notions are discussed in the following simple model.

Consider a production function for value-added (Y) with three factors (skilled labour (S), unskilled labour (U) and capital (K)). Using the results from section 2.1.1 the share equations derived from the translog cost function (equation (2.15)) are

$$S_S = \alpha_S + \gamma_{SS} \ln W_S + \gamma_{SU} \ln W_U + \gamma_{SK} \ln W_K + \gamma_{SY} \ln Y + \phi_{\tau W_S} t$$

$$S_U = \alpha_U + \gamma_{US} \ln W_S + \gamma_{UU} \ln W_U + \gamma_{UK} \ln W_K + \gamma_{UY} \ln Y + \phi_{\tau W_U} t \qquad (7.5)$$

$$S_K = \alpha_K + \gamma_{KS} \ln W_S + \gamma_{KU} \ln W_U + \gamma_{KK} \ln W_K + \gamma_{KY} \ln Y + \phi_{\tau W_K} t.$$

The restrictions that can be placed on the parameters of the share equations are as follows. Symmetry will mean $\gamma_{ij} = \gamma_{ji}$. Homogeneity means we also have $\sum_{i=S,U,K} \gamma_{ij} = 0$ for all j factors, $\sum_{i=S,U,K} \gamma_{iY} = 0$, $\sum_{i=S,U,K} \phi_{\tau W_i} = 0$ and $\sum_{i=S,U,K} \alpha_i = 0$. Coupled with the fact that the shares add up to unity, one equation becomes redundant and we need only estimate the system

$$S_S = \alpha_S + \gamma_{SS} \ln(W_S/W_U) + \gamma_{SK} \ln(W_K/W_U) + \gamma_{SY} \ln Y + \phi_{\tau W_S} t \qquad (7.6)$$

$$S_K = \alpha_K + \gamma_{KS} \ln(W_S/W_U) + \gamma_{KK} \ln(W_K/W_U) + \gamma_{KY} \ln Y + \phi_{\tau W_K} t.$$

[89]The exogenous technology approach is where the term 'skill-biased technical change' originated from. The endogenous technology approach is closer to the idea of complementarity proper.

Constant returns to scale implies the restriction that $\gamma_{SY} = \gamma_{KY} = 0$. Symmetry ($\gamma_{SK} = \gamma_{KS}$) implies a further cross equation restriction. Given estimates of the parameters we can calculate all the elasticities of substitution from the formulae in section 2.1.1.

A positive coefficient γ_{SK} implies substitutability, but complementarity between two factors requires not only that the coefficient γ_{SK} be negative, but also more negative (in absolute value) than the product of the factor shares.[90] The bias of technical change depends on the values of the $\phi_{\tau W}$ parameters. The price elasticities are $S_j\sigma_{ij}$ and are also easily calculated from the estimated parameters (and the predicted shares).

There are various ways to bring dynamic considerations into this equation. A simple way is to treat capital as quasi-fixed.[91] Thus, instead of the long-run cost function many researchers would consider a short-run variable cost function (e.g. Brown and Christensen (1981)). This recognises that the quasi-fixed factors are not at their long-run optimal values (without being explicit over the adjustment dynamics). In comparison to equation (2.15) we replace the cost of capital with the quantity of capital

$$\begin{aligned}
\ln VC &= \ln\alpha_0 + \sum_{i=1}^{n}\alpha_i \ln W_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_{ij}\ln W_i \ln W_j \qquad (7.7)\\
&\quad +\alpha_Y \ln Y + \frac{1}{2}\gamma_{YY}(\ln Y)^2 + \sum_{i=1}^{n}\gamma_{iY}\ln W_i \ln Y + \gamma_{KY}\ln K \ln Y\\
&\quad +\gamma_K \ln K + \frac{1}{2}\gamma_{KK}(\ln K)^2 + \sum_{i=1}^{n}\gamma_{iK}\ln W_i \ln K\\
&\quad +\phi_\tau t + \frac{1}{2}\phi_{\tau\tau}t^2 + \phi_{\tau Y} t \ln Y + \sum_{i=1}^{n}\phi_{\tau W_i} t \ln W_i + \phi_{\tau K} t \ln K
\end{aligned}$$

where $i = S, U$ (n=2). Using the same logic as before, using Shephard's (1953)

[90] See equation (2.21). Even if all factors were substitutes, one might still be interested in whether the elasticity of substitution with capital was greater for unskilled workers than for skilled workers.

[91] This also has the practical advantage that, as we discussed in section 6.3, measuring exogenous variation in the user cost of capital is extremely difficult.

Lemma for the variable factors and imposing homogeneity and symmetry, we end up with a single (variable cost) share equation

$$S_S = \alpha_S + \gamma_{SS} \ln(W_S/W_U) + \gamma_{SK} \ln K + \gamma_{SY} \ln Y + \phi_{\tau W_S} t \tag{7.8}$$

where S_S is now the share of skilled workers in variable costs (the wage bill). If we want to impose constant returns here $\gamma_{SK} = -\gamma_{SY}$. Assuming that this is true, we can write the wage bill share equation as

$$S_S = \alpha_S + \phi_{\tau W_S} t + \gamma_{SS} \ln(W_S/W_U) + \gamma_{SK} \ln(K/Y). \tag{7.9}$$

As before, the Allen partial elasticity of substitution between skilled and unskilled labour is greater (less) than unity as $\gamma_{SS} > (<)\ 0$. The coefficient on the capital intensity variable should be positive to be consistent with capital-skill complementarity. Explicit calculation of the size of the elasticity of substitution/complementarity requires additional information, such as direct estimates of the cost function parameters. If the sign of $\phi_{\tau W_S}$ (essentially the time trend) is positive, this is consistent with skill-biased technical change.

There are two major problems with using this method as a way for examining skill-biased technical change. First, the time trend could be picking up many other aspects of the economic environment rather than just technical change. This is the standard problem of treating technology as a residual. A natural response to this is to find more direct proxies for technology. Clearly all the usual problems arise in that there is no perfect measure of technical change, but some observable measure (R&D, patents, etc.) seems preferable to assuming the residual trend is all technology. Once we do use explicit measures of technical change, however, we run into the second problem that firms have influence over technical progress. R&D for example, as discussed in section 6.5, is also a choice variable. One could set the problem up as a model where we consider two capital stocks, knowledge capital (G) and physical capital (K), in the variable cost function.[92] This would

[92] The distinction between skill-biased technical change and capital-skill complementarity can

imply adding an extra term $\ln(G/Y)$ to (7.9), giving[93]

$$S_S = \alpha_S + \phi_{\tau W_S} t + \gamma_{SS} \ln(W_S/W_U) + \gamma_{SK} \ln(K/Y) + \gamma_{SG} \ln(G/Y). \qquad (7.10)$$

We have not discussed the method of including technology variables in earnings equations as a way of examining skill biases (e.g. Krueger (1993)). The omission is deliberate (see Chennells and Van Reenen (1997) for a longer discussion). The theoretical basis of such an equation is unclear. In the neoclassical model technology shocks can affect the demand for labour, but the wage is exogenous to the firm as it is determined in the aggregate labour market (this is implicit in the structure discussed above). Under this view the significance of technology indicators in individual earnings equations is likely to capture unobserved ability of individuals which is correlated with both wages and the use of new technologies. There is evidence that this is indeed the case (e.g. DiNardo and Pischke (1997)).

Although it uses industry data, a key paper in this area is Berman et al. (1994), who estimate versions of equations (7.9) and (7.10) on 4 digit US manufacturing data in long differences.[94] They use R&D expenditures and computer investment as their measures of technical change. These proxies for technology are found to have a positive and significant association with the growth in the wage bill share of non-production workers, the computer variable accounting for about a third of the increase in the share. Autor et al. (1998) extend this study over a longer time period (from the 1940s to early 1990s) and to non-manufacturing industries. They corroborated the importance of technical change (especially computer use)

become murky. In the standard approach technology is exogenous and capital is chosen by the firm. Capital-skill complementarity is a conjecture about the shape of isoquants holding output constant. Technical change, however, causes a move to a new isoquant. Embodied technical change makes the distinction even less clear.

[93] Calculation of the elasticities of substitution/complementarity with two quasi-fixed factors is complex. The sign of the cross-elasticity will depend on both γ_{SK} and γ_{SG} (see Brown and Christensen (1981)).

[94] Although (like much of the subsequent literature) they replace wages with time dummies due to the problem that industry wage changes reflect a mix of genuine change in the price of labour and changes in the skill mix.

in accounting for the increase in skilled workers as a proportion of the wage bill. Machin and Van Reenen (1998) compare the US to 6 other countries (Denmark, France, Germany, Japan, Sweden and the UK). They find results which broadly support the importance of skill bias using their measure of R&D intensity. Other papers with country-specific analyses have also tended to find evidence of skill-biased technical change (e.g. Hansson (1997) for Sweden and Gera et al. (1999) for Canada), but Goux and Maurin (1995) are more sceptical about its importance in France.

Aggregation may be a serious problem for these industry studies. The Longitudinal Research Database (LRD), a manufacturing panel dataset for the population of larger plants, has been a prime resource in the USA. Doms et al. (1997) and Dunne et al. (1997) both find evidence of skill bias, but Doms et al. (1997) stress that they cannot find evidence for significant effects in the time series dimension of their data. This is a worrying result, because it does suggest that some other unmeasured factor may be driving both skills and technology. On the other hand, measurement error issues and the fact that they use counts of production technologies (rather than computer usage) might account for their results. Indeed, when they use measures of computer capital instead of the count-based measure they find evidence of significant skill bias even in the time series dimension. Adams (1999) focuses on firms mainly operating in the chemical industry. He finds that firm R&D in the same product field as that produced by the plant is associated with skill bias.

Duguet and Greenan (1997) use an innovations survey to estimate cost share equations for a panel of French manufacturing firms, 1986-1991, in long differences. They find evidence for skill bias and argue that it comes primarily from the introduction of new products, although their results here are mixed. One problem with subjective innovations surveys is the comparability of the notion of innovation across different firms. An interesting extension, given the increasing availability of this type of innovation survey, would be to use the longitudinal aspect of the panel when the innovation questions are asked to the same firms in future. Machin (1996) uses the British Workplace Industrial Relations Survey (WIRS) panel 1984-1990, which contains information on the introduction of computers and also finds evidence for skill bias. Haskel and Heden (1999) use data on about 10,000 British manufacturing establishments from the Annual Respondents Database (ARD) panel to estimate (7.10). They also find that changes in the wage bill share of non-production workers are correlated significantly with the intensity of investment in computer technology. Aguirregabiria and Alonso-Borrego (1998) use firm level Spanish panel data and attempt to control for some of the non-convexities discussed above. They find effects from the first introduction of 'technological capital', but they find no effects from subsequent increases in the stock of this capital or from R&D.

Taken as a whole we draw three conclusions from this body of empirical work. First, there does appear to be considerable support for the notion of skill-biased technical change across a range of studies, and these are usually (but not always) robust to controlling for fixed effects. Secondly, there have been few attempts to find instrumental variables to deal with the potential endogeneity of technology. Candidate instruments could include government-induced schemes to alter the incentives to accumulate technological capital (such as R&D tax credits, government grants, etc.). Thirdly, there are surprisingly few studies which try to analyse the mechanisms by which technological change translates into higher demand for skills[95]. One mechanism is through organisational changes such as delayering, decentralisation and giving greater autonomy to workers. These organisational factors have been found to be important in the case study evidence and in the literature on the productivity paradox (investigating why computers have not

[95]For a recent exception see Autor, Levy and Murnane (2003) who argue that IT substitutes for both manual and non-manual 'routine' tasks.

raised measured productivity by as much as might have been expected). Some preliminary work suggests that this organisational restructuring could be the link between technology and labour demand (Bresnahan et al. (2002); Caroli and Van Reenen (2001)).

8. Conclusions

We can summarise the main themes from this chapter as follows. Structural microeconometric models of investment and employment are useful for testing hypotheses about the environment in which firms make decisions about their factor inputs: is investment spending subject to financing constraints? are employment levels subject to union bargaining? Given that complete adjustment of capital stock and employment levels does not occur immediately, these structural models need to be dynamic in nature. This presents a major challenge for econometric modelling, since current decisions depend on unobserved expectations of future demand conditions and factor prices.

Structural dynamic models developed in the 1980s generally assumed that firms face strictly convex costs of adjustment. These models rationalise slow adjustment, and allow structural econometric specifications to be derived that control for the role of unobserved expectations. Examples include the Q model of investment, and the Euler equation models that have been used in both the investment and employment literatures. However these models predict a smooth, gradual pattern of adjustment. Recent work, particularly that which uses plant level data, has suggested that a pattern of infrequent, large adjustments may be more relevant for both capital and labour. Moreover the structural dynamic models based on strictly convex adjustment costs have generally been rejected in microeconometric tests. Initial optimism that the rejection of these models with more aggregated data may be attributed to aggregation biases does not appear to have been well founded.

Whilst there is now a broad consensus that these traditional structural models appear to be inadequate, there is less agreement on which sources of potential mis-specification are most important. Does this reflect the importance of capital market imperfections, or non-convex components of adjustment costs, or something else? How important is the measurement error introduced by different approaches to controlling for firms' expectations of future conditions, particularly when stock market valuations are used? How important are the simplifying assumptions typically made concerning market structure? Whilst a lot of research in the last decade has implemented rigorous tests of relatively simple structural specifications against specific alternatives, surprisingly little progress has been made in developing richer structural models that incorporate these features. This balance will need to be redressed if we are to provide more convincing evidence that particular features of the firm's technology and environment are important for understanding investment and employment behaviour.

As in other areas of microeconometric research, such as household consumption and labour supply behaviour, it is important to recognise that all aggregation problems are not circumvented by the use of data on microeconomic units. Annual investment spending by a large, publicly traded company is clearly aggregated both over time and over different types of capital goods, and may be aggregated over multiple plants or subsidiary firms. Total employment is also aggregated over heterogeneous types of workers, and a constant level of employment may disguise significant inflows and outflows. Identification of structural models of investment and employment dynamics may require more serious attention to be paid to these aggregation issues than has generally been the case in previous research.

Another striking feature of this literature is the limited attention that has been paid to directly estimating long-run price elasticities of demand for capital and labour inputs. This is largely explained by the paucity of microeconomic data on factor prices. New data sources may allow significant progress to be made in this area. In the employment context, the development of matched panels covering both individual workers and individual firms should provide more accurate information on wage rates paid by individual firms than has been available hitherto. On the investment side, high frequency data on stock returns may allow cross-firm variation in the risk premium component of the user cost of capital to be exploited.

Another major area of interest that merits further research is the impact of technological and organisational change on the demand for capital and labour. On the employment side, there has been considerable research into the impact of skill-biased technical change, but surprisingly little micro research has addressed the relationship between technological opportunities and investment. Given the enormous policy interest in the effects of technical progress, additional work is required on the nature of these effects and the transmission mechanisms through which technical change affects investment and employment.

Finally we note that whilst the vast majority of microeconometric research has used data for firms and plants in the manufacturing sector, manufacturing industry now accounts for a comparatively small and declining share of aggregate investment, employment and GDP in most developed economies. Similarly the globalisation of business activities has meant that multinational corporations now account for a significant and growing share of total domestic investment and employment in many countries. Multinational firms have opportunities to substitute between domestic and foreign factors of production, which may make their investment and employment behaviour qualitatively different from that of purely domestic firms. Greater emphasis on the behaviour of multinational companies and firms in service sectors is likely to be required if these microeconometric studies are to provide useful insights into the broader behaviour of investment and employment.

Acknowledgements:

We thank Manuel Arellano, Nick Bloom, Jeffrey Campbell, Jason Cummins, Jan Eberly, Eduardo Engel, Dan Hamermesh, Lars Hansen, Jim Heckman, Steve Nickell, John Pencavel and Frank Wolak for helpful comments and discussions. We thank the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at the Institute for Fiscal Studies for financial support. All errors are our own.

References

[1] Abel, A.B. (1980), "Empirical investment equations: an integrative framework", in: K. Brunner and A. Meltzer, eds., On the State of Macroeconomics, Carnegie-Rochester Conference Series 12:39-93.

[2] Abel, A.B. (1983), "Optimal investment under uncertainty", American Economic Review 73:228-233.

[3] Abel, A.B. and O.J. Blanchard (1986), "The present value of profits and cyclical movements in investment", Econometrica 54:249-273.

[4] Abel, A.B. and J.C. Eberly (1994), "A unified model of investment under uncertainty", American Economic Review 84:1369-1384.

[5] Abel, A.B. and J.C. Eberly (1996), "Investment and q with fixed costs: an empirical analysis", mimeo, The Wharton School, University of Pennsylvania.

[6] Abel, A.B. and J.C. Eberly (1999), "The effects of irreversibility and uncertainty on capital accumulation", Journal of Monetary Economics 44:339-377.

[7] Abowd, J.M. (1989), "The effect of wage bargains on the stock market value of the firm", American Economic Review 79(4):774-809.

[8] Abowd, J.M., P. Corbel, and F. Kramarz (1999), "The entry and exit of workers and the growth of employment: an analysis of French establishments", Review of Economics and Statistics 81(2):170-187.

[9] Abowd, J.M. and F. Kramarz (1999), "The analysis of labor markets using matched employer-employee data", in: O. Ashenfelter and D. Card, eds., Handbook of Labor Economics, vol. 3B (North-Holland, Amsterdam).

[10] Abowd, J.M., F. Kramarz and D.N. Margolis (1999), "High wage workers and high wage firms", Econometrica, 67(2):251-333.

[11] Adams, J. (1999) "The structure of firm R&D and the factor intensity of production", Review of Economics and Statistics 81:499-510.

[12] Aghion, P. and P. Howitt (1992), "A model of growth through creative destruction", Econometrica 60:323-351.

[13] Aguirregabiria, V. (1997), "Estimation of dynamic programming models with censored dependent variables", Investigaciones Economicas XXI(2):167-208.

[14] Aguirregabiria, V. and C. Alonso-Borrego (1998), "Employment, occupational stucture, technological capital and reorganisation of production" Universidad Carlos III de Madrid Working Paper.

[15] Aguirregabiria, V. and C. Alonso-Borrego (1999), "Labor contracts and flexibility: evidence from a labor market reform in Spain", University of Chicago, mimeo.

[16] Ahn, S.C. and P. Schmidt (1995), "Efficient estimation of models for dynamic panel data", Journal of Econometrics 68:5-28.

[17] Akerlof, G. (1970), "The market for lemons: quality uncertainty and the market mechanism", Quarterly Journal of Economics 84:488-500.

[18] Allen, R. (1938), Mathematical Analysis for Economists (Macmillan, London).

[19] Alogoskoufis, G. and A. Manning (1991), "Tests of alternative wage-employment bargaining models with an application to the UK aggregate labour market", European Economic Review 35:23-37.

[20] Alonso-Borrego, C. (1998), “Demand for labour inputs and adjustment costs: evidence from Spanish manufacturing firms”, Labour Economics 5:475-497.

[21] Alonso-Borrego, C. and M. Arellano (1999), “Symmetrically normalised instrumental-variable estimation using panel data”, Journal of Business and Economic Statistics 17:36-49.

[22] Altug, S. and R.A. Miller (1990), “Household choices in equilibrium”, Econometrica: 58(3):543-570.

[23] Altug, S. and R.A. Miller (1991), “Human capital, aggregate shocks and panel data estimation”, Institute for Empirical Macroeconomics, Discussion Paper no. 47, Federal Reserve Bank of Minneapolis.

[24] Anti Nilsen, O. and F. Schiantarelli (2003), “Zeroes and lumps in investment: empirical evidence on irreversibilities and non-convexities”, Review of Economics and Statistics 85(4):1021-1037 .

[25] Arellano, M. and S.R. Bond (1991), “Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations”, Review of Economic Studies 58:277-297.

[26] Arellano, M. and O. Bover (1995), “Another look at the instrumental-variable estimation of error component models”, Journal of Econometrics 68:29-52.

[27] Arellano, M. and B. Honore (2001), “Panel data models: some recent developments”, in: J.J. Heckman and E.E. Leamer, eds., Handbook of Econometrics, vol. 5 (North-Holland, Amsterdam).

[28] Arrow, K.J. (1968), "Optimal capital policy with irreversible investment", in: J.N. Wolfe, ed., Value, Capital and Growth: Papers in Honour of Sir John Hicks (Edinburgh University Press).

[29] Arrow, K.J., H.B. Chenery, B. Minhas and R.M. Solow (1961), "Capital-labor substitution and economic efficiency", Review of Economics and Statistics 43:225-254.

[30] Ashenfelter, O. and J. Brown (1986), "Testing the efficiency of employment contracts", Journal of Political Economy 94(3):S40-S87.

[31] Auerbach, A.J. and K.A. Hassett (1992), "Tax policy and business fixed investment in the United States", Journal of Public Economics 47:141-170.

[32] Autor D. and L. Katz (1999), "Changes in wage inequality", in: O. Ashenfelter and D. Card, eds., Handbook of Labor Economics, vol. 3A (North-Holland, Amsterdam).

[33] Autor, D., R. Levy and D. Murnane (2003), "The skill content of recent technological change: an empirical exploration", Quarterly Journal of Economics 118(4), forthcoming.

[34] Autor, D., L. Katz and A. Krueger (1998), "Computing inequality: have computers changed the labor market?", Quarterly Journal of Economics 113:1169-1214.

[35] Barnett, S.A. and P. Sakellaris (1998), "Nonlinear response of firm investment to Q: testing a model of convex and non-convex adjustment costs", Journal of Monetary Economics 42:261-288.

[36] Bean, C.R. (1981), "An econometric model of manufacturing investment in the UK", Economic Journal 91:106-121.

[37] Berman, E., J. Bound and Z. Griliches (1994), "Changes in the demand for skilled labor within US manufacturing: evidence from the annual survey of manufacturers", Quarterly Journal of Economics 109:367-98.

[38] Berndt, E. (1991), The Practice of Econometrics (Addison-Wesley, MA).

[39] Bertola, G. and R.J. Caballero (1994), "Irreversibility and aggregate investment", Review of Economic Studies 61:223-246.

[40] Blanchard, O.J., C. Rhee and L.H. Summers (1993), "The stock market, profit and investment", Quarterly Journal of Economics CVIII:115-134.

[41] Blanchard, O.J. and M. Watson (1982), "Bubbles, rational expectations and financial markets", in: P. Wachtel, ed., Crises in the Economic and Financial Structure (Lexington, MA).

[42] Bloom, N., S.R. Bond and J. Van Reenen (2001), "The dynamics of investment under uncertainty", Working Paper no. W01/05, The Institute for Fiscal Studies, London.

[43] Bloom, N., L. Chennells, R. Griffith and J. Van Reenen (1998), "The effects of tax treatment on the changing cost of R&D: evidence from eight countries", in: H. Lawton-Smith, ed., The Economics of Regulation of High Technology Industries (Oxford University Press, Oxford).

[44] Bloom, N., R. Griffith and J. Van Reenen (2002), "Do R&D tax credits work: evidence from an international panel of countries 1979-97", Journal of Public Economics, 85:1-31.

[45] Blundell, R.W. and S.R. Bond (1998), "Initial conditions and moment restrictions in dynamic panel data models", Journal of Econometrics 87:115-143.

[46] Blundell, R.W. and S.R. Bond (2000) "GMM estimation with persistent panel data: an application to production functions", Econometric Reviews 19:321-340.

[47] Blundell, R.W., S.R. Bond, M.P. Devereux and F. Schiantarelli (1992), "Investment and Tobin's Q: evidence from company panel data", Journal of Econometrics 51:233-257.

[48] Blundell, R.W., S.R. Bond and C. Meghir (1996), "Econometric models of company investment", in: L. Matyas and P. Sevestre, eds., The Econometrics of Panel Data (Kluwer Academic Publishers, Dordrecht).

[49] Blundell, R.W., T.E. MaCurdy and C. Meghir (this volume), "Estimation and specification in labor supply models", in: J.J. Heckman and E.E. Leamer, eds., Handbook of Econometrics, vol. 6 (North-Holland, Amsterdam).

[50] Blundell, R.W. and T. Stoker (this volume), "Models of aggregate economic relationships that account for heterogeneity", in: J.J. Heckman and E.E. Leamer, eds., Handbook of Econometrics, vol. 6 (North-Holland, Amsterdam).

[51] Boal, W.M. and J. Pencavel (1994), "The effect of labor unions on employment, wages, and days of operation: coal mining in West Virginia", Quarterly Journal of Economics 109:267-298.

[52] Bond, S.R. and J.G. Cummins (2000), "The stock market and investment in the new economy: some tangible facts and intangible fictions", Brookings Papers on Economic Activity 2000(1):61-124.

[53] Bond, S.R. and J.G. Cummins (2001), "Noisy share prices and the Q model of investment", Working Paper no. W01/22, The Institute for Fiscal Studies, London.

[54] Bond, S.R., J. Elston, J. Mairesse and B. Mulkay (2003), "Financial factors and investment in Belgium, France, Germany and the UK: a comparison using company panel data", Review of Economics and Statistics 85(1):153-165.

[55] Bond, S.R., D. Harhoff and J. Van Reenen (1999), "Investment, R&D and financing constraints in Britain and Germany", Working Paper no. W99/5, The Institute for Fiscal Studies, London.

[56] Bond, S.R. and C. Meghir (1994), "Dynamic investment models and the firm's financial policy", Review of Economic Studies 61:197-222.

[57] Booth, A. (1995), The Economics of the Trade Union (Cambridge University Press, Cambridge).

[58] Bourguignon, F. and P.A. Chiappori (1992), "Collective models of household behaviour: an introduction", European Economic Review 36: 355-365.

[59] Brainard, W. and J. Tobin (1968), "Pitfalls in financial model building", American Economic Review 58(2): 99-122.

[60] Bresnahan, T.F., E. Brynjolfsson and L.M. Hitt (2002), "Information technology, workplace organisation and the demand for skilled labor: firm-level evidence", Quarterly Journal of Economics 117:339-376.

[61] Bresson, G., F. Kramarz and P. Sevestre (1992), "Heterogeneous labour and the dynamics of aggregate labour demand: some estimations using panel data", Empirical Economics 17:153-168.

[62] Brown, R. and L. Christensen (1981), "Estimating elasticities of substitution in a model of partial static equilibrium: an application to US agriculture 1947-74", in: C. Field and E. Berndt, eds., Modelling and Measuring Natural Resource Substitution (MIT Press, Cambridge).

[63] Burgess, S. and J. Dolado (1989), "Intertemporal rules with variable speeds of adjustment: an application to UK manufacturing employment", Economic Journal 99:347-365.

[64] Caballero, R.J. (1991), "On the sign of the investment-uncertainty relationship", American Economic Review 81:279-288.

[65] Caballero, R.J. (1992), "A fallacy of composition", American Economic Review 82:1279-1292.

[66] Caballero, R.J. (1994), "Small sample bias and adjustment costs", Review of Economics and Statistics 76(1):52-58.

[67] Caballero, R.J. (1999), "Aggregate investment", in: J.B. Taylor and M. Woodford, eds., Handbook of Macroeconomics, vol. 1B (North-Holland, Amsterdam).

[68] Caballero, R.J. and E. Engel (1993), "Microeconomic adjustment hazards and aggregate dynamics", Quarterly Journal of Economics 108:313-358.

[69] Caballero, R.J. and E. Engel (1999), "Explaining investment dynamics in U.S. manufacturing: a generalized (S,s) spproach", Econometrica 67:783-826.

[70] Caballero, R.J., E. Engel and J. Haltiwanger (1995), "Plant-level adjustment and aggregate investment dynamics", Brookings Papers on Economic Activity 1995(2):1-54.

[71] Caballero, R.J., E. Engel and J. Haltiwanger (1997), "Aggregate employment dynamics: building from microeconomic evidence", American Economic Review 87:115-137.

[72] Caballero, R.J. and J.V. Leahy (1996), "Fixed costs: the demise of marginal q", National Bureau of Economic Research Working Paper no. 5508.

[73] Calomiris, C.W. and R.G. Hubbard (1995), "Internal finance and firm-level investment: evidence from the undistributed profits tax of 1936-37", Journal of Business 68:443-482.

[74] Campbell, J.R. and J.D.M. Fisher (2000), "Aggregate employment fluctuations with microeconomic asymmetries", American Economic Review 90:1323-1325.

[75] Campbell, J.Y. and A.S. Kyle (1993), "Smart money, noise trading and stock-price behaviour", Review of Economic Studies 60:1-34.

[76] Card, D. (1986), "Efficienct contracts with costly adjustment", American Economic Review 76:1045-1071.

[77] Card, D. and A. Krueger (1995), Myth and Measurement (Princeton University Press, Princeton).

[78] Caroli, E. and J. Van Reenen (2001), "Skill-biased organisational change? Evidence from a panel of British and French establishments", Quarterly Journal of Economics 116:1449-1492.

[79] Chambers, R. (1988), Applied Production Analysis (Cambridge University Press, Cambridge).

[80] Chennells, L. and J. Van Reenen (1997), "Technical change and earnings in British establishments", Economica 64:587-604.

[81] Chesher, A. (1991), "The effect of a measurement error", Biometrika 78:451-462.

[82] Chirinko, R.S. (1993a), "Business fixed investment spending: modelling strategies, empirical results and policy implications", Journal of Economic Literature 31:1875-1911.

[83] Chirinko, R.S. (1993b), "Multiple capital goods, Q and investment spending", Journal of Economic Dynamics and Control 17:907-928.

[84] Chirinko, R.S. (1997), "Finance constraints, liquidity, and investment spending: theoretical restrictions and international evidence", Journal of the Japanese and International Economies, 11:185-207.

[85] Chirinko, R.S., S.M. Fazzari and A.P. Meyer (1999), "How responsive is business capital formation to its user cost? An exploration with micro data", Journal of Public Economics 74: 53-80.

[86] Christensen, L.R., D.W. Jorgenson and L.J. Lau (1971), "Conjugate duality and the transcendental logarithmic production function", Econometrica 39:255-256.

[87] Christensen, L.R., D.W. Jorgenson and L.J. Lau (1973), "Transcendental logarithmic production frontiers", Review of Economics and Statistics 55:28-45.

[88] Christofides, L. and A. Oswald (1991), "Efficient and inefficient employment outcomes: a study based on Canadian contract data", Research in Labor Economics 12:173-190.

[89] Cobb, C. and P.H. Douglas (1928), "A theory of production", American Economic Review 18:139-165.

[90] Cohen, W. and R. Levin (1989), "Empirical studies of innovation and market structure", in: R. Schmalensee and R. Willig, eds., The Handbook of Industrial Organisation, vol.1 (North-Holland, Amsterdam).

[91] Cooper, R.W. and J. Ejarque (2001), "Exhuming Q: market power vs. capital market imperfections", National Bureau of Economic Research Working Paper no. 8182.

[92] Cooper, R.W. and J.C. Haltiwanger (2000), "On the nature of capital adjustment costs", National Bureau of Economic Research Working Paper no. 7925.

[93] Cooper, R.W., J.C. Haltiwanger and L. Power (1999), "Machine replacement and the business cycle: lumps and bumps", American Economic Review 84:921-946.

[94] Cooper, R.W. and J.L. Willis (2001), "The economics of labor adjustment: mind the gap", National Bureau of Economic Research Working Paper no. 8527, forthcoming in American Economic Review.

[95] Craig, B. and J. Pencavel (1994), "The empirical performance of orthodox models of the firm: conventional firms and worker co-operatives", Journal of Political Economy 102:718-744.

[96] Cummins, J.G, K.A. Hassett and R.G. Hubbard (1994), "A reconsideration of investment behaviour using tax reforms as natural experiments", Brookings Papers on Economic Activity 1994(2):1-59.

[97] Cummins, J.G, K.A. Hassett and R.G. Hubbard (1996), "Tax reforms and investment: a cross-country comparison", Journal of Public Economics 62:237-273.

[98] Cummins, J.G., K.A. Hassett and S.D. Oliner (1999), "Investment behavior, observable expectations, and internal funds ", New York University, mimeo.

[99] Dagenais, M., P. Mohnen and P. Therrien (1997), "Do Canadian firms respond to fiscal incentives to research and development?", GREQAM Working Paper 97b05.

[100] Davidson, J.E.H., D.F. Hendry, F. Srba and S. Yeo (1978), "Econometric modelling of the aggregate time-series relationships between consumers' expenditure and income in the United Kingdom", Economic Journal 88:661-692.

[101] Davis, S. and J. Haltiwanger (1992), "Gross job creation, gross job destruction and employment reallocation", Quarterly Journal of Economics 107:819-863.

[102] Devereux, M.P. (1989), "Tax asymmetries, the cost of capital and investment: some evidence from UK panel data", Economic Journal 99(supplement):103-112.

[103] Devereux, M.P., M.J. Keen and F. Schiantarelli (1994), "Corporation tax asymmetries and investment: evidence from UK panel data", Journal of Public Economics 53:395-418.

[104] Devereux, M.P. and F. Schiantarelli (1990), "Investment, financial factors and cash flow: evidence from UK panel data", in: R.G. Hubbard, ed., Asymmetric Information, Corporate Finance and Investment (University of Chicago Press).

[105] DiNardo, J. and J.S. Pischke (1997), "The returns to computer use revisited: have pencils changed the wage structure too?", Quarterly Journal of Economics 112:291-303.

[106] Dixit, A.K. and R.S. Pindyck (1994), Investment under Uncertainty (Princeton University Press).

[107] Doms, M. and T. Dunne (1998), "Capital adjustment patterns in manufacturing plants", Review of Economic Dynamics 1(2):409-429.

[108] Doms, M., T. Dunne and K. Troske (1997), "Workers, wages and technology", Quarterly Journal of Economics 112:253-289.

[109] Duguet, E. and N. Greenan (1997), "Skill biased technological change: an econometric study at the firm level", CREST Working Paper.

[110] Dunne, T., J. Haltiwanger and K. Troske (1997), "Technology and jobs: secular changes and cyclical dynamics", Carnegie-Rochester Conference Series on Public Policy 46:107-178.

[111] Eberly, J.C. (1997), "International evidence on investment and fundamentals", European Economic Review 41:1055-1078.

[112] Edwards, J.S.S., J.A. Kay and C.P. Mayer (1987), The Economic Analysis of Accounting Profitability (Oxford University Press).

[113] Eisner, R. (1977) "Cross section and time series estimates of investment functions", Annales de l'Insee 30/31:99-129.

[114] Elston, J. (1993), "Firm ownership structure and investment: theory and evidence from German manufacturing", WZB Discussion Paper no. FS IV 93-28, Berlin.

[115] Engle, R.F. and C.W.J. Granger (1987), "Cointegration and error correction: representation, estimation, and testing", Econometrica 55:251-276.

[116] Erickson, T. and T.M. Whited (2000), "Measurement error and the relationship between investment and q ", Journal of Political Economy 108:1027-1057.

[117] Favero, C.A. and M.H. Pesaran (1994), "Oil investment in the North Sea", Economic Modelling 11:308-329.

[118] Fazzari, S.M., R.G. Hubbard and B.C. Petersen (1988), "Financing constraints and corporate investment", Brookings Papers on Economic Activity 1988(1):141-195.

[119] Fazzari, S.M., R.G. Hubbard and B.C. Petersen (2000), "Investment-cash flow sensitivities are useful: a comment on Kaplan and Zingales", Quarterly Journal of Economics 115(2):695-705.

[120] Galeotti, M. and F. Schiantarelli (1991), "Generalised Q models for investment", Review of Economics and Statistics 73:383-392.

[121] Galeotti, M. and F. Schiantarelli (1994), "Stock market volatility and investment: do only fundamentals matter?", Economica 61:147-165.

[122] Gera, S., W. Gu and Z. Lin (1999), "Technology and the demand for skills: an industry-level analysis", Statistics Canada Working Paper no. 28.

[123] Gilchrist, S. (1991), "An empirical analysis of corporate investment and financing hierarchies using firm-level panel data", mimeo, Board of Gevernors of the Federal Reserve System.

[124] Gilchrist, S. and C.P. Himmelberg (1995), "Evidence on the role of cash flow for investment", Journal of Monetary Economics 36:541-572.

[125] Gomes, J.F. (2001), "Financing investment", American Economic Review 91:1263-1285.

[126] Goux, N. and E. Maurin (1995), "The decline in demand for unskilled labour: an empirical analysis method and its application to France", CREST mimeo.

[127] Griliches, Z. (1969), "Capital-skill complementarity", Review of Economics and Statistics 51:465-468.

[128] Griliches, Z.(1984), R&D, Patents and Productivity (Chicago University Press, Chicago).

[129] Griliches, Z. (1996), "Productivity, R&D and basic research at the firm level in the 1970s", American Economic Review 76(1):141-54.

[130] Griliches, Z. (1998) R&D and Productivity: The Econometric Evidence (Chicago University Press, Chicago).

[131] Griliches, Z., B.H. Hall and A. Pakes (1991), "R&D, patents and market value revisited: is there a technological opportunity factor?", Economics of Innovation and New Technology 1:183-201.

[132] Griliches, Z. and J. Mairesse (1998), "Production functions: the search for identification", in S. Strøm, ed., Econometrics and Economic Theory in the 20th Century (Cambridge University Press, Cambridge).

[133] Guiso, L. and G. Parigi (1999), "Investment and demand uncertainty", Quarterly Journal of Economics 114(1):185-227.

[134] Hall, B.H. (1992), "Investment and research and development at the firm level: does the source of financing matter?", Working Paper no. 92-194, Department of Economics, University of California, Berkeley.

[135] Hall, B.H. (1993), "R&D tax policy during the 1980s: success or failure", Tax Policy and the Economy 7:1-36.

[136] Hall, B.H., J. Mairesse, L. Branstetter and B. Crepon (1999), "Does cash flow cause investment and R&D: an exploration using panel data for French, Japanese and United States firms in the scientific sector", in D. Audretsch and A.R. Thurik, eds., Innovation, Industry Evolution and Employment (Cambridge University Press, Cambridge).

[137] Hall, R.E and D.W. Jorgenson (1967), "Tax policy and investment behaviour", American Economic Review 57:391-414.

[138] Hamermesh, D. (1989), "Labor demand and the structure of adjustment costs", American Economic Review 79:674-689.

[139] Hamermesh, D. (1993), Labor Demand (Princeton Academic Press).

[140] Hamermesh, D, W. Hassink and J. Van Ours (1996), "Job turnover and labor turnover: a taxonomy of employment dynamics", Annales d'Économie et de Statistique 41/42:21-40.

[141] Hamermesh, D. and G. Pfann (1996), "Adjustment costs in factor demand", Journal of Economic Literature 34:1264-1292.

[142] Hansson, P. (1997), "Trade, technology and changes in employment of skilled labor in Swedish manufacturing", in: J. Fagerberg, P. Hansson, L. Lundberg and A. Melchior, eds., Technology and International Trade (Cheltenham: Edward Elgar).

[143] Harhoff, D. (1998), "Are there financing constraints for innovation and investment in German manufacturing firms?", Annales d'Économie et de Statistique 49/50:421-456.

[144] Haskel, J.E. and Heden, Y. (1999), "Computers and the demand for skilled labour: industry and establishment panel evidence for the UK", Economic Journal 109:C68-C79.

[145] Hassett, K.A. and R.G. Hubbard (1996), "Tax policy and investment", National Bureau of Economic Research Working Paper no. 5683.

[146] Hayashi, F. (1982), "Tobin's average q and marginal q: a neoclassical interpretation", Econometrica 50:213-224.

[147] Hayashi, F. (1985a), "Corporate finance side of the Q theory of investment", Journal of Public Economics 27:261-280.

[148] Hayashi, F. (1985b), "Tests of liquidity constraints: a critical survey", in: T. Bewley, ed., Advances in Econometrics, Vol. II, (Cambridge University Press).

[149] Hayashi, F. and T. Inoue (1991), "The relation between firm growth and q with multiple capital goods: theory and evidence from Japanese panel data", Econometrica 59(3):731-754.

[150] Heckman, J.J. (1979), "Sample selection bias as a specification error", Econometrica 47:153-161.

[151] Heckman, J.J. and G. Sedlacek (1985), "Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market", Journal of Political Economy 93:1077-1125.

[152] Hendry, D.F. (1995), Dynamic Econometrics (Oxford University Press).

[153] Himmelberg, C. and B. Peterson (1994), "R&D and internal finance: a panel data study of small firms in high tech industries", Review of Economics and Statistics 76(1):38-51.

[154] Holtz-Eakin, D., W. Newey and H.S. Rosen (1988), "Estimating vector autoregressions with panel data", Econometrica 56:1371-1395.

[155] Hoshi, T., A.K. Kashyap and D. Scharfstein (1991), "Corporate structure, liquidity and investment: evidence from Japanese industrial groups", Quarterly Journal of Economics CVI:33-60.

[156] Hines, J. (1994), "No place like home: tax incentives and the location of R&D by American multinationals", Tax Policy and the Economy 8:65-104.

[157] Hsiao, C. (1986) Analysis of Panel Data (Cambridge University Press).

[158] Hubbard, R.G. (1998), "Capital-market imperfections and investment", Journal of Economic Literature 36:193-225.

[159] Hubbard, R.G., A.K. Kashyap and T.M. Whited (1995), "Internal finance and firm investment", Journal of Money, Credit and Banking 27:683-701.

[160] Jaramillo, F., F. Schiantarelli and A. Sembenelli (1993), "Are adjustment costs for labour asymmetric? An econometric test on panel data for Italy", Review of Economics and Statistics 75:640-648.

[161] Jensen, M. (1986), "Agency costs of free cash flow, corporate finance, and takeovers", American Economic Review 76:323-329.

[162] Jorgenson, D.W. (1963), "Capital theory and investment behaviour", American Economic Review 53:247-259.

[163] Jorgenson, D.W. (1971), "Econometric studies of investment behaviour", Journal of Economic Literature 9:1111-1147.

[164] Jorgenson, D.W. and R. Landau (1993), Tax Reform and the Cost of Capital (Brookings Institution, Washington DC).

[165] Kaplan, S.N. and L. Zingales (1997), "Do investment-cash flow sensitivies provide useful measures of financing constraints?", Quarterly Journal of Economics 112(1):169-216.

[166] Kaplan, S.N. and L. Zingales (2000), "Investment-cash flow sensitivities are not valid measures of financing constraints", Quarterly Journal of Economics 115(2):707-712.

[167] King, M.A. (1974), "Taxation and the cost of capital", Review of Economic Studies 41:21-35.

[168] King, M.A. and D. Fullerton (1984), The Taxation of Income from Capital (University of Chicago Press).

[169] Klette, T.J. and J. Moen (1999), "From growth theory to technology policy: coordination problems in theory and practice", Nordic Journal of Political Economy 25:53-74.

[170] Krueger, A. (1993), "How have computers changed the wage structure: evidence from micro data, 1984-89", Quarterly Journal of Economics 108:33-60.

[171] Lach, S. and R. Rob (1996), "R&D, investment and industry dynamics", Journal of Economics and Management Strategy 5(2):217-249.

[172] Lach, S. and M. Schankerman (1989), "The dynamics of R&D investment in the scientific sector", Journal of Political Economy 97:880-904.

[173] Lamont, O.A. (1997), "Cash flow and investment: evidence from internal capital markets", Journal of Finance 52:83-109.

[174] Leahy, J.V. and T.M. Whited (1996), "The effect of uncertainty on investment: some stylised facts", Journal of Money, Credit and Banking 28:64-83.

[175] Leontief, W. (1946), "The pure theory of the guaranteed annual wage contract", Journal of Political Economy 54:76-79.

[176] Lucas, R.E. (1967), "Adjustment costs and the theory of supply", Journal of Political Economy 75:321-334.

[177] Lucas, R.E. (1976), "Econometric policy evaluation: a critique", in: K. Brunner and A. Meltzer, eds., The Phillips Curve and Labour Markets, Carnegie-Rochester Conference Series 1:19-46.

[178] Machin, S. (1996), "Changes in the relative demand for skills in the UK labour market", in: A. Booth and D. Snower, eds., Acquiring Skills (Cambridge University Press, Cambridge).

[179] Machin, S. and J. Van Reenen (1998), "Technology and changes in the skill structure: evidence from seven OECD countries", Quarterly Journal of Economics 113:1215-1244.

[180] Machin. S., A. Manning and C. Meghir (1993), "Dynamic models of employment based on firm level panel data", in: J. Van Ours, G. Pfann and G. Ridder, eds., Labor Demand and Equilibrium Wage Formation (Elsevier Science Publishers, Amsterdam).

[181] MaCurdy, T. and J. Pencavel (1986), "Testing between competing models of wage and employment determination", Journal of Political Economy 94:S1-S39.

[182] Mairesse, J. and B. Dormont (1985), "Labor and investment demand at the firm level: a comparison of French, German and US manufacturing, 1970-79", European Economic Review 28:201-231.

[183] Manning, A. (1987), "An integration of trade-union models in a sequential bargaining framework", Economic Journal 97:121-139.

[184] Meghir, C., A. Ryan and J. Van Reenen (1996), "Job creation, technological innovation and adjustment costs", Annales d'Économie et de Statistique, 41/42, 255-274.

[185] Menezes-Filho, N., D. Ulph and J. Van Reenen (1998), "R&D investment and union bargaining: evidence from British companies and establishments", Industrial and Labor Relations Review 52(1):45-63 .

[186] Miller, M.H. and F. Modigliani (1961), "Dividend policy, growth and the valuation of shares", Journal of Business 44:411-433.

[187] Modigliani, F. and M.H. Miller (1958), "The cost of capital, corporation finance, and the theory of investment", Americal Economic Review 48:261-297.

[188] Mulkay, B., B.H. Hall and J. Mairesse (2001), "Investment and R&D in France and the United States", in H. Heinz and R. Strauch, eds., Investing Today for the World of Tomorrow (Springer Verlag).

[189] Myers, S.C. (1984), "The capital structure puzzle", Journal of Finance 39:575-592.

[190] Nadiri, M. and S. Rosen (1969), "Interrelated factor demand functions", American Economic Review 59:547-571.

[191] Nelson, C.R. and R. Startz (1990a), "Some further results on the exact small sample properties of the instrumental variable estimator", Econometrica 58:967-976.

[192] Nelson, C.R. and R. Startz (1990b), "The distribution of the instrumental variable estimator and its t-ratio when the instrument is a poor one", Journal of Business and Economic Statistics 63:5125-5140.

[193] Nickell, S.J. (1978), The Investment Decisions of Firms (Cambridge University Press, Cambridge).

[194] Nickell, S.J. (1981), "Biases in dynamic models with fixed effects", Econometrica 49:1417-1426.

[195] Nickell, S.J. (1985), "Error correction, partial adjustment and all that: an expository note", Oxford Bulletin of Economics and Statistics 47(2):119-129.

[196] Nickell, S.J. (1986), "Dynamic models of labor demand", in: O. Ashenfelter and R. Layard, eds., Handbook of Labor Economics, vol.1 (North-Holland, Amsterdam).

[197] Nickell, S.J. and D. Nicolitsas (1996), "Does innovation encourage investment in fixed capital?", Centre for Economic Performance Discussion Paper No. 309, London School of Economics.

[198] Nickell, S.J.and S. Wadhwani (1991), "Employment determination in British industry: investigations using micro-data", Review of Economic Studies 58:955-970.

[199] Pakes, A. (1985), "On patents, R&D and the stock market rate of return", Journal of Political Economy 93:390-409.

[200] Pencavel, J. (1991), Labor Markets Under Trade Unionism (Oxford University Press, Oxford).

[201] Pesaran, M.H. (1990), "An econometric analysis of exploration and extraction of oil in the UK continental shelf", Economic Journal 100:367-390.

[202] Pesaran, M.H. and R. Smith (1995), "Estimating long-run relationships from dynamic heterogeneous panels", Journal of Econometrics 68:79-113.

[203] Pfann, G. and F. Palm (1993), "Asymmetric adjustment costs in non-linear labour demand models of the Netherlands and UK manufacturing sectors", Review of Economic Studies 60:397-412.

[204] Pfann, G. and B. Verspagen (1989), "The structure of adjustment costs for labour in the Dutch manufacturing sector", Economics Letters 29:365-371.

[205] Romer, P.M. (1986), "Increasing returns and long run growth", Journal of Political Economy 94(5):1002-1037.

[206] Rota, P. (1998), "Dynamic labour demand with fixed costs of adjustment", mimeo, University College London.

[207] Ruiz-Arranz, M. (2002), "Wage inequality in the U.S.: Capital-skill complementarity vs. skill-biased technological change", Harvard University mimeo.

[208] Salinger, M.A. and L.H. Summers (1983), "Tax reform and corporate investment: a microeconometric simulation study", in: M. Feldstein, ed., Behavioural Simulation Methods in Tax Policy Analysis (University of Chicago Press).

[209] Sargan, J.D. (1964), "Wages and prices in the UK: a study in econometric methodology", in: P.E. Hart, G. Mills and J.K. Whitaker, eds., Econometric Analysis for National Economic Planning (Butterworths, London).

[210] Schaller, H. (1993), "Asymmetric information, liquidity constraints and Canadian investment", Canadian Journal of Economics 26:552-574.

[211] Schiantarelli, F. (1996), "Financial constraints and investment: a critical survey of the international evidence", Oxford Review of Economic Policy 12(2):70-89.

[212] Schiantarelli, F. and D. Georgoutsos (1990), "Imperfect competition, Tobin's q and investment: evidence from aggregate UK data", European Economic Review 34:1061-1078.

[213] Schiantarelli, F. and A. Sembenelli (1993), "Estimation of Euler equations for employment when adjustment costs are asymmetric: evidence from panel

data for UK companies", in: J. Van Ours, G. Pfann and G. Ridder, eds., Labor Demand and Equilibrium Wage Formation (Elsevier Science Publishers, Amsterdam).

[214] Shapiro, M.D. (1986), "The dynamic demand for capital and labor", Quarterly Journal of Economics CI:513-542.

[215] Shephard, R. (1953) Cost and Production Functions (Princeton University Press).

[216] Staiger, D. and J.H. Stock (1997), "Instrumental variables regression with weak instruments", Econometrica 65:557-586.

[217] Stiglitz, J.E. and A. Weiss (1981), "Credit rationing in markets with imperfect information", American Economic Review 71:393-410.

[218] Summers, L.H. (1981), "Taxation and corporate investment - a Q-theory approach", Brookings Papers on Economic Activity 1981(1):67-140.

[219] Summers, L.H. (1986), "Does the stock market rationally reflect fundamental values?", Journal of Finance 41:591-600.

[220] Sutton, J. (1998), Technology and Market Structure (MIT Press, Cambridge).

[221] Tobin, J. (1969), "A general equilibrium approach to monetary theory", Journal of Money, Credit and Banking 1:15-29.

[222] Toivanen, O. and P. Stoneman (1998), "Dynamics of R&D and investment: UK evidence", Economic Letters 58:119-126.

[223] Uzawa, H. (1962), "Production functions with constant elasticities of substitution", Review of Economic Studies 29:291-299.

[224] Weiss, R. (1977) "Elasticities of substitution across capital and occupations in US manufacturing", Journal of the American Statistical Association 72:764-771.

[225] Whited, T.M. (1992), "Debt, liquidity constraints and corporate investment: evidence from panel data", Journal of Finance 47:1425-1460.

[226] Zeldes, S.P. (1989), "Consumption and liquidity constraints: an empirical investigation", Journal of Political Economy 97:305-346.

Figure 1. The Q model with financial regimes

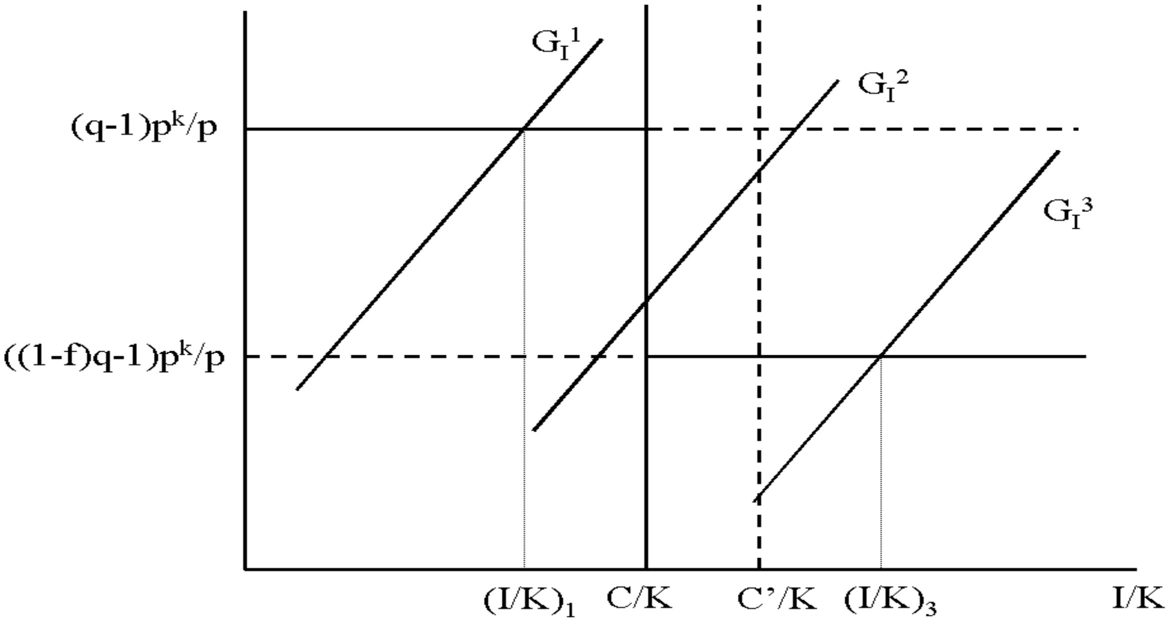

Figure 8.1:

Figure 2. Static model, costly debt, linear MPK

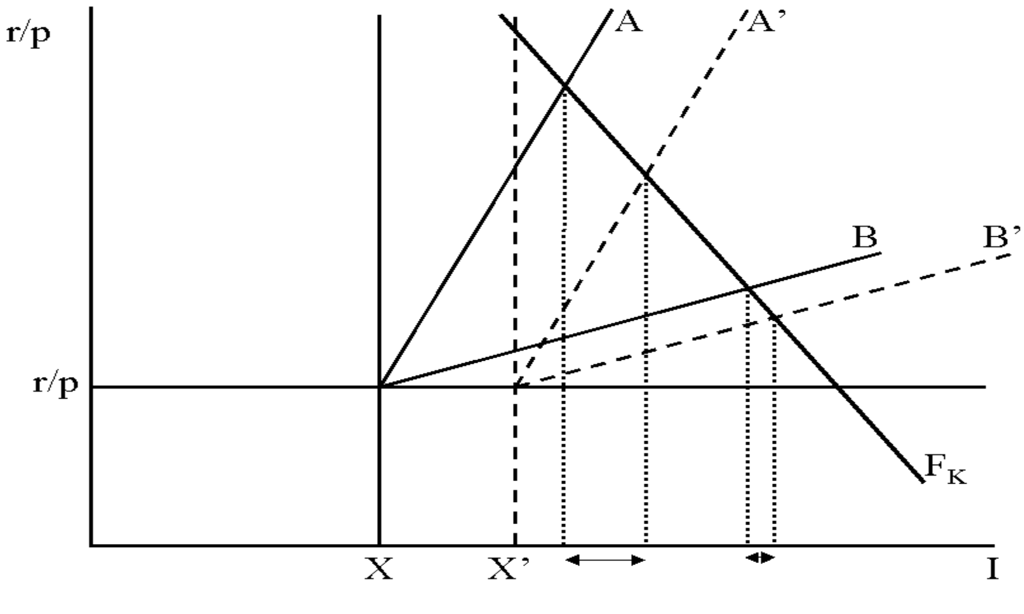

Figure 8.2:

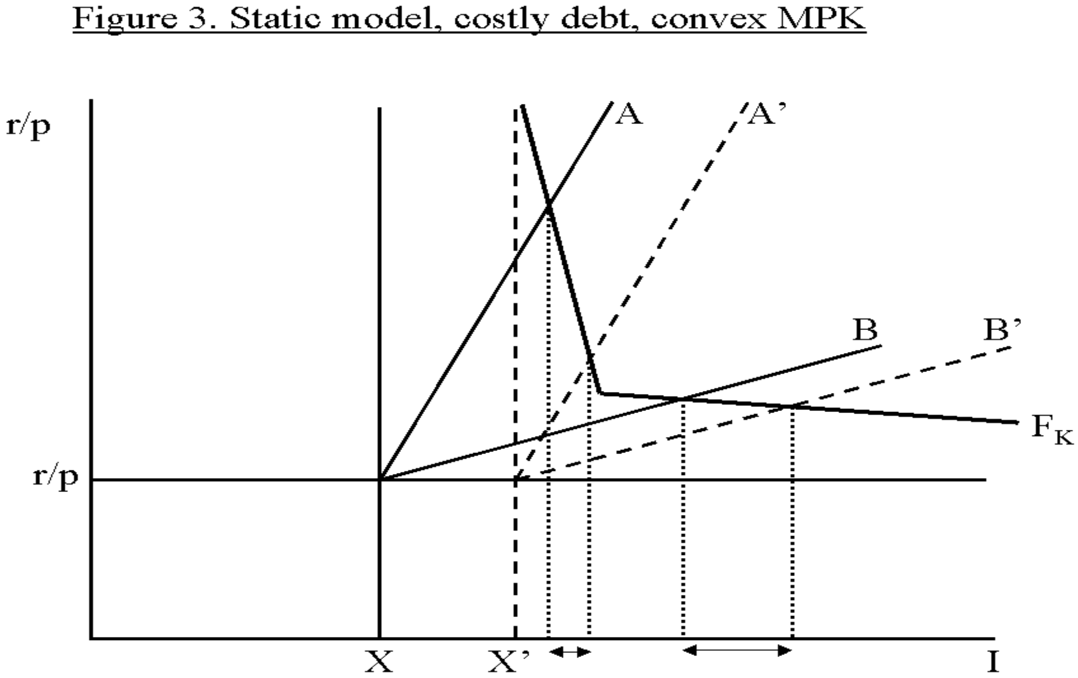
Figure 3. Static model, costly debt, convex MPK
r/p
A
A'
B
B'
F_K
r/p
X
X'
I

Figure 8.3: