The Impact of Financing Constraints on Investment

by

J.E.S. Brown

London School of Economics and Political Science

Thesis submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy at the University of London

November 1998

Abstract

This thesis is an empirical and theoretical analysis of the impact of financing constraints on firm-level investment behaviour. Its primary objectives are to model this impact, and to test the restrictions these models place on the data. Chapter 1 contains a discussion of these themes, and provides an overview of the thesis. Chapter 2 addresses the empirical question of whether innovative firms are financially constrained. To answer this question, several structural investment equations are tested, and the sensitivity of physical investment expenditures to internal finance is compared across innovative and non-innovative firms. The investment expenditures of innovative firms are found to be more sensitive to cash flow than those of non-innovative firms. These results support the hypothesis that innovative firms are financially constrained. The third chapter builds a theoretical model to explain a widely reported fact in the inventory literature, which is that the variance of production exceeds the variance of sales. This fact contradicts a prediction of the standard Linear-Quadratic model of inventory investment, and for this reason is often referred to as the "excess variance of production" puzzle. In this chapter, a model of inventory investment is built. It is shown that when financing constraints are imposed on the model, it can explain the excess variance of production puzzle. In the absence of these constraints, the model does not deliver this result. The fourth chapter returns to the theme of identifying financially constrained firms. A weakness of existing tests of financing constraints is that they are not both direct and structural. This chapter addresses that criticism by constructing a model of investment from which is derived a simple and direct, structural test of the null hypothesis that a group of firms is financially constrained. The test is implemented on a panel of U.S. manufacturing firms. The results support the findings of existing tests.

Acknowledgements

When I began this thesis at the LSE, I had the extremely good fortune of being offered a position at the Financial Markets Group by David Webb. That position gave me financial support, and access to resources I needed for my research. More importantly, through the position I met many colleagues who stimulated me intellectually, and from whom I learned a great deal, not only about economics but also about how to undertake research. I am deeply indebted to David for having provided me with this position, and for the guidance and support he has given me throughout my time at the FMG. I would also like to thank my supervisor, John Hardman Moore, for his guidance and encouragement throughout my PhD. Of those who have helped me with my research, I would particularly like to express my thanks to Steve Bond, who very generously has given me much helpful advice on the empirical sections of my thesis. Others who I would like to thank are Anders Damgaard, Martin Evans, Urs Haegler, Nobu Kiyotaki, Bob Nobay, Danny Quah, Sven Rady, and Ian Tonks. In particular, I would like to thank my office-mates Richard Payne and Alex Stremme. Not only have they both been extremely generous in offering me help and encouragement with my research, but it has truly been a pleasure to have shared an office with them.

Finally, I would like to thank my parents, my sister Lindsey, and my fiancée Robynne. Without their love and support, I would not have been able to complete this thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis is comprised of three essays which model and test the impact of financing constraints on firm level investment behaviour.

Chapter 2 is an application of the testing methodology currently predominant in the empirical literature on financing constraints. It addresses the question of whether firms which are innovative, i.e. those that devote resources towards the development of new products or production processes, are more likely to be financially constrained than non-innovative firms. Although on theoretical grounds innovative firms are natural candidates to face large cost premia on external finance, most empirical studies have shown that R&D expenditures are surprisingly insensitive to fluctuations in internal finance. However, this insensitivity may be due to technological reasons. Consequently, Chapter 2 examines the behaviour of the physical capital investment of innovative and non-innovative firms. The empirical results can be interpreted as evidence that innovative firms do face binding financing constraints.

Chapter 3 is a purely theoretical analysis, devoted to the explanation of an empirical puzzle in the inventory literature. A robust and widely reported fact is that the variance of a firm's production is greater than the variance of its sales. This contradicts a prediction of the standard model of inventories, the Linear-Quadratic model. Chapter 3 builds a model of inventory investment, and shows that when financing constraints are imposed, the model can explain this puzzle.

The testing methodology employed in Chapter 2 has several weaknesses. These weaknesses, discussed in the next section of this introduction, are addressed in Chapter 4. In that chapter, a structural model of investment is constructed from which a direct test for capital market imperfections can be derived. The test is based on theoretical results regarding the economic exogeneity of the firm's cash flow, and takes the form of a simple Granger causality test. As this test is both a direct test of financing constraints, and is derived explicitly from a structural model, it is not subject to the same criticisms made of existing tests. In the latter part of the chapter the test is performed on a panel of US manufacturing firms. The results are identical to those reported in the majority of papers in the empirical literature on financing constraints. The null hypothesis that financing constraints are binding fails to be rejected for firms classified *a priori* as financially constrained. For other firms, the hypothesis is rejected.

The remainder of this chapter provides a brief survey of the literature in two areas. In the first section, the empirical literature on financing constraints is reviewed, and the strengths and weaknesses of various tests discussed. In the second section, the the empirical and theoretical literature on inventory investment is reviewed.

1.1 Testing for Capital Market Imperfections: Methodological Issues

In the 1970's and 1980's, theoretical developments in the field of asymmetric information were applied to the study of capital markets. A series of papers demonstrated that the presence of asymmetric information in these markets would force firms to pay a premium for external sources of finance.[1] In extreme circumstances, access to external finance may even be rationed or prohibited altogether. Participants in these markets were thought to face a "hierarchy of finance" in which internal finance was cheaper than external finance, and perhaps where the costs of different forms

[1]For a review of the theoretical literature on asymmetric information and capital markets, see Freixas and Rochet (1997).

of external finance differed as well. The key empirical prediction of these models is that fluctuations in internal finance will affect investment expenditures through a "supply of finance" channel, in addition to the standard "demand" channel in which the fluctuations proxy for changes in the set of investment opportunities facing the firm.

The intuition behind this prediction is illustrated in Figure 1.1. In that figure, the supply of finance schedule is perfectly elastic up to the point where the firm exhausts its internal sources of finance at X. The cost of internal finance, r, is the appropriate risk-adjusted interest rate at which, in a perfect capital market, finance would be provided. At X the supply schedule becomes upward-sloping to reflect the cost premium on external finance.[2] The demand schedule is drawn such that the financing constraint is binding. That is, the firm actually spends I^c, whereas it would have spent I^* had it not been constrained.

The equilibrium in this market is a function of two exogenous disturbances: a demand shock μ, and a supply shock η. For generality, it is assumed that the former shock shifts the supply curve through its affect on current period profits. The latter shock may be interpreted as a cash windfall, i.e. a shock to the firm's current profits which contains no information regarding the firm's investment opportunities. The shocks are assumed to be orthogonal.

If capital markets are perfect, then only shifts in the demand schedule cause changes in investment expenditures.[3] However, as can be seen from Figure 1, if capital markets are imperfect then shifts in the supply schedule will also affect investment. In these circumstances, the demand shock affects investment through two channels: the "demand" channel represented by shifts in the demand curve, and the "supply" channel, shown by shifts in the supply curve. Also in these circumstances, the pure supply shock, η, will affect investment spending.

[2]For the purposes of this discussion, the shape of the supply curve is unimportant. What matter is that it is not perfectly elastic.

[3]To simplify this discussion, r is assumed to be constant over time.

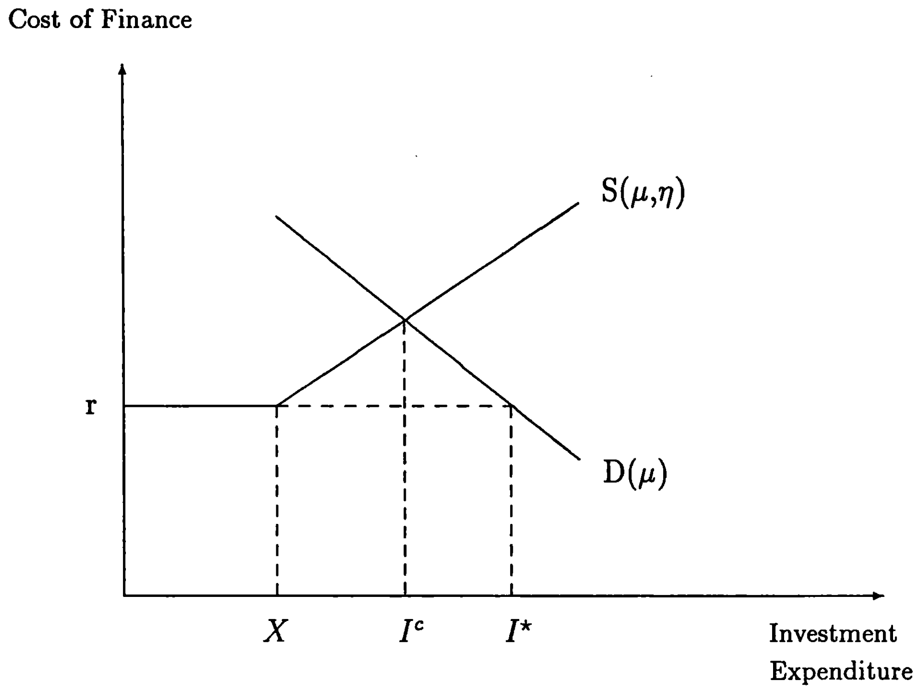

Figure 1.1: The Identification Problem

Viewed from this perspective, it can be seen that in order to provide conclusive evidence on capital market imperfections, shifts in the supply of finance schedule first must be identified. The main criticism of the existing empirical papers on financing constraints is directed at the weaknesses in the identification techniques they have employed. Indeed, many of the papers in this literature do not directly address the identification problem. For the purposes of critically reviewing this literature, papers are classified into the following four groups: (i), indirect reduced-form tests, (ii), indirect structural tests, (iii), direct reduced-form tests, and (iv), direct structural tests.[4]

The first type of test was used by Fazzari, Hubbard, and Petersen (1988), the paper which spawned the modern empirical literature on financing constraints.[5] In this

[4]For more comprehensive reviews of the empirical literature on financing constraints, see Hubbard (1997) and Schiantarelli (1996). For a review of the empirical investment literature, see Chirinko (1993).

[5]There are some earlier empirical studies on the importance of financial determinants of invest-

test, reduced-form investment regressions are run. Variables such as Tobin's q, Jorgenson's user cost of capital, or first differences in sales are used to control for shifts in investment demand, while variables such as cash flow are used to capture the effect of shifts in the supply of finance schedule. As no formal attempt is made to identify movements in the two schedules, any economic interpretation of the coefficients in these regressions will be ambiguous. This is the main objection to this technique. In these regressions, it is unclear the extent to which variables such as cash flow are proxying for shifts in the demand schedule.

However, inference in this technique is not based upon the coefficients themselves, but rather upon a comparison of their magnitudes across different classes of firms. The justification for this inference is similar to that used in classical experiment design. That is, if a sample firms can be stratified according to a specific criterion, then differences in the regression results across strata can be attributed to the differences identified by the stratification criterion. The criteria that have been used are variables which, it is argued, will separate firms according to whether or not the assumptions underlying the hypothesis of perfect capital markets are violated. Variables which have been used include firm size, firm age, existence of a bond or commercial paper rating, and the values taken by certain ratios such as dividends to earnings.

The difficulty with comparing reduced-form coefficients across classes is that unlike in classical experiment design, it is impossible to ensure that those determinants which are not controlled for by the design of the experiment have identical distributions for all classes of firms. Certain structural items may differ systematically between groups. For example, the stochastic process governing demand for the "group-average" firm's output may differ across groups. As reduced-form coefficients are functions of underlying structural parameters, a version of the Lucas critique is applicable here. These parameters must either be controlled for, as in classical experiments, or identified in the regressions themselves. If neither is done, then it

---

ment expenditures, most notably Meyer and Kuh (1957).

remains unclear the extent to which the difference in reduced-form coefficients can be explained by differences in the values of structural parameters, as opposed to differences in premia on external finance.

In response to these criticisms, subsequent research focused on estimating structural investment equations, all of which were based on the neoclassical Adjustment Costs model of investment under the assumption of perfect capital markets (Eisner and Strotz (1963), Lucas (1967), Gould (1968)). Two structural equations can be derived from this model: the q equation, and an Euler equation.[6] If all of the assumptions underlying the derivation of these equations are valid, then the restrictions they place on the data will not be rejected. These include restrictions on which variables should enter the regression significantly, the signs of their coefficients, restrictions on the relative magnitudes of the coefficients (within a group), and over-identification restrictions. The second group of papers test these restrictions.

Like the first test, this test does not attempt to directly identify shifts in the supply curve. The goal of this test is to isolate the reasons why structural restrictions may be rejected by the data for a certain sample of firms. To do this many of these papers also stratify their samples.[7] However, unlike the first test, these regressions are structural and therefore not subject to the Lucas critique.

In an attempt to identify the causes of rejection, these papers introduced modifications to the basic Adjustment Costs model. Some papers focused on the q equation, still using market value data to measure q. For example, Blundell, Bond, Devereux, and Schiantarelli (1992), addressed the issues of econometric endogeneity and measurement error by directly examining the conditions under which lags (and leads) of q were valid instruments. Hayashi and Inoue (1991) modified the model to allow for heterogeneity in capital stock. Their estimation technique explicitly permitted the coefficient on q to vary over time.[8] Finally, Schiantarelli and Georgoutsos (1990)

[6]See the appendix of Chapter 2 for a derivation of these equations.

[7]Nevertheless, not all papers in this group use stratified samples. Notable exceptions are Blundell, Bond, Devereux, and Schiantarelli (1992) and Bond and Meghir (1994).

[8]This coefficient is a function of the market rate of interest, and therefore will vary. Most

relaxed the assumption of perfect competition, by assuming that the firm operates in a monopolistically competitive product market. All of these papers report that certain restrictions implied by their respective structural models were rejected, at least for a sub-sample of firms.

Other papers focused on tests of the Adjustment Costs model which do not rely on market data to formulate q. One group of papers test the Euler equation of the adjustment costs model. Bond and Meghir (1994) and Whited (1992) and Hubbard, Kashyap, and Whited (1995) both found that the data rejected certain restrictions implied by their versions of the model. Abel and Blanchard (1986) and Gilchrist and Himmelberg (1995), tested the q equation without using market data to construct q.[9] Both these papers found that the Adjustment Costs model was nevertheless rejected.

Although these papers test structural equations, they do not directly test the hypothesis that capital markets are imperfect. Despite the modifications of the Adjustment Costs model made by many of the papers in the literature, all of their null hypotheses are still joint hypotheses.[10] Thus, when the model is rejected in the data for a particular group of firms, it is impossible to identify which assumption has been violated. Consequently, one is forced to used reduced-form evidence, such as that provided by the first test, in support of the argument that the perfect capital markets assumption has been violated. In contrast, if one of the assumptions in the joint hypothesis was that capital markets were imperfect, then a failure to reject the null would be unambiguous evidence of financing constraints.

The pitfalls of using the indirect structural test of the second technique are illustrated by the model in Chapter 4, and that in Caballero and Leahy (1996). Both

---

studies include time dummies to partially account for interest rate changes.

[9]Under certain conditions, marginal q can be expressed as the expected present discounted value of the cash flow-capital ratio. Using standard techniques for predicting geometric distributed lags, (see Sargent (1987), p.303) marginal q can be estimated from company accounts data.

[10]Some other assumptions included in these hypotheses in addition to that of perfect capital markets are the linear homogeneity of certain functions, no fixed costs of adjustment, and the stationarity of the stochastic processes governing structural disturbances.

models show that when some of the assumptions of the Adjustment Costs model are changed, but the perfect capital markets assumption retained, then average q is no longer the only sufficient statistic for investment. Other variables, such as the cash flow-capital ratio, are sufficient statistics as well. Moreover, under more general assumptions on the stochastic processes of the exogenous shocks, neither q nor cash flow are sufficient statistics. In other words, both variables would enter a "q equation" regression significantly, despite there being perfect capital markets. In this case, it is the failure of assumptions other than that of perfect capital markets which has lead to the rejection of the standard Adjustment Costs model.

Despite their weaknesses, Chapter 2 is an application of the first two tests to a sample of UK firms. The sample has been divided into innovative and non-innovative firms. Both the Euler equation and the q equation of the Adjustment Costs model are tested on each sub-sample. Each of these equations is rejected in each sub-sample. Thus, using the second test one cannot distinguish between the two groups of firms. However, the investment expenditures of innovative firms are significantly more sensitive to fluctuations in cash flow than those of non-innovative firms. Using the arguments of the first test, this is interpreted as evidence that innovative firms are financially constrained.

The third group of papers addresses the weakness of the indirect structural test by attempting to identify η, the shocks which shift only the supply curve. If capital markets are imperfect then η, or variables which are only a function of η, will enter significantly into an investment regression. Thus, these papers directly test the hypothesis of imperfect capital markets by testing the significance of these variables in reduced-form investment equations.

The difference between the first test and this test is that the papers using the latter attempt to identify observable variables which are functions only of η. For example, Blanchard, Lopez-de Silanes, and Schleifer (1994) use data on the receipt of windfalls from court judgements. Lamont (1997) uses data from diversified oil companies. He argues that shocks to the cash flow of oil-related activities can be interpreted as supply of finance shocks for the investment expenditures of non-oil activities. Hubbard, Kashyap, and Whited (1995) use tax payments as an instrument for cash flow. Froot and Stein (1991) use exchange-rate shocks to the internal finance of multinational firms. Finally, Fazzari and Petersen (1993) use the co-movement between investment in working capital and physical capital to identify η. All of these papers found that investment was positively correlated with their chosen proxies for η.

However, these papers share the generic weakness of reduced-form tests, which is that they do not identify η in a structural economic model. Consequently, they are subject to the same reservations as the original Fazzari, Hubbard, and Petersen (1988), namely that there is no theoretical argument supporting the assertion that the chosen data items are independent of the μ shock. In the absence of an economic model, one still cannot be certain as to what interpretation to give to the coefficients of the variables chosen to proxy for η. Moreover, the stricter is the criteria for selecting such variables, the smaller the sample becomes on which the final test is run. This makes it more difficult to argue that capital market imperfections are widespread enough to be of importance.

The final type of test is both direct and structural. As such, it is not subject to the same criticisms made of the first three types of test. Since the test is structural, cross-group comparisons can be made. Since it is direct, failure to reject its null hypothesis can be interpreted unambiguously as evidence of capital market imperfections.

Some papers in the financing constraints literature have imposed financing constraints on the Adjustment Costs model in an attempt to derive structural variables which are affected by the underlying shocks only through the "supply channel". The most obvious example of such a variable is the Lagrange Multiplier on the financing constraint. The Lagrange Multiplier enters both the first-order condition and the Euler equation only when the financing constraint is binding, or is expected to bind. Thus, if the Multiplier were an observable variable, it could be used as the basis of a direct structural test. But that is the problem with this approach. The multiplier is not observable. In order for the test to be structural, the Multiplier must be explicitly expressed in terms of observable variables.[11] Tests which use proxies for the Lagrange Multiplier, such as Whited (1992), Hubbard, Kashyap, and Whited (1995), and Gilchrist and Himmelberg (1998) are not structural.

Chapter 4 of this thesis is the only paper which currently falls into this category of test. In that chapter, a slightly different approach to the identification problem is taken. Rather than identifying η, and testing the sensitivity of investment expenditures to this shock, this chapter specifies conditions in the context of an investment model under which it is possible to identify η. It is shown that if financing constraints always bind, these conditions are not satisfied. If, on the other hand, financing constraints never bind, the conditions are satisfied, and η can be identified using a restriction first employed in Blanchard and Quah (1989). These theoretical results place a testable restriction on the VAR representation for certain observable variables generated by the investment model. This permits an extremely simple test of the null hypothesis that a group of firms is financially constrained.

As was mentioned previously, the vast majority of papers in this literature find evidence of capital market perfections. Currently, there are only two papers which find evidence to the contrary: Kaplan and Zingales (1997) and Cummins, Hassett, and Oliner (1997). Both these papers make an important contribution to the literature by providing results which demonstrate the inherent weaknesses of the reduced-form tests discussed above. However, it is equally true that both papers suffer from precisely these same weaknesses. As a result, they do not bring us closer to understanding whether or not capital market imperfections exist. The debate over the existence of potential imperfections in capital markets will continue until evidence based on direct structural tests is discovered. Chapter 4 is a first step in this direction.

---

[11]For example, the equivalence of marginal q, an unobservable variable, with average q, an observable variable, is explicitly derived under certain restrictions of the Adjustment Costs model under perfect capital markets.

1.2 Financing Constraints and Inventory Investment

The theoretical literature on inventory investment is dominated by models which are based on real as opposed to financial factors. However, there are several stylised facts concerning inventories which suggest that capital market frictions may play an important role in our understanding of the determinants of inventory investment. The first of these is the role inventories play in accounting for business cycles. In developed countries, inventory disinvestment typically accounts for about 50% of the declines in aggregate output during post-war recessions.[12] This is true despite the fact that the stock of inventories accounts for only about 1% of the level of aggregate output in a given year.

A growing body of theoretical research in macroeconomics emphasizes the potential for capital market imperfections to explain both the magnitude and persistence of fluctuations in aggregate output.[13] However, these theories typically abstract away from inventory investment. Given the importance of inventories in accounting for fluctuations in aggregate output, it is a natural step to analyse the impact of financing constraints on firm level inventory investment.

This direction for research becomes more tightly focused when one considers another stylised fact about inventories, which is that sales and inventory investment are positively correlated.[14] An implication of this fact is that the variance of production exceeds the variance of sales.[15] This can be seen from the accounting identity $y_t \equiv s_t + \Delta n_t$, where y_t is production, s_t is sales, and n_t is the stock of inventories. From this equation we get $var(y) = var(s) + var(\Delta n) + cov(s, \Delta n)$.

---

[12]For US data see Blanchard and Fischer (1989), and Blinder and Maccini (1991). For data from other developed countries, see Ramey and West (1998).

[13]See for example Kiyotaki and Moore (1995), Greenwald and Stiglitz (1993), Gertler (1992), Calomiris and Hubbard (1990), Bernanke and Gertler (1989).

[14]This fact is manifested in aggregate data by the procyclicality of inventory investment, a fact which has lead some economists to speculate on whether inventory investment is a source of destabilization.

[15]This fact is very widely reported. See, for example, Blinder and Maccini (1991) or Ramey and West (1998).

The standard model of inventories, the Linear-Quadratic Model (Holt et al. (1960)), predicts exactly the opposite. In that model firms hold inventories in order to smooth production. This implies that the variance of production is lower than the variance of sales. Moreover, the production smoothing prediction rests on very weak assumptions, the main one being a convex cost function. For these reasons this stylised fact is often referred to as the "excess variance of production" puzzle.

Chapter 3 shows that this puzzle can be explained by the effect of capital market imperfections on inventory investment. The intuition behind this explanation is very similar to the intuition of the "financial accelerator" underlying many macro models of business cycles. Following below average sales realisation, financially constrained firms do not have sufficient funds to replace all the units which were sold, and consequently "under-invest" in inventories. Conversely, following a high sales realisation firms are in a financial position to "over-invest". It is this behaviour which leads inventory investment to be positively correlated with sales, and hence explains the excess variance of production puzzle.

Accelerator motives are very common in the inventory literature.[16] They are usually derived from the interaction of a costly adjustment technology (of the stock of inventories), and inventory target levels based on sales expectations. However, structural estimates of the costs of adjusting the stock of inventories are implausibly high.[17] The model of Chapter 3 can be viewed as an additional source of accelerator effects. A possible explanation of the high estimates of adjustment costs is that they reflect the interaction of both technological and financial frictions to adjustment.

Thus, one motivation for Chapter 3 is that it provides a link between a strand the macro literature on business cycles, and an explanation of the observed behaviour of one the most important components of fluctuations in aggregate output. However, there is also considerable empirical motivation for the theory of inventory investment presented in that chapter. The empirical literature on inventories provides

[16] A very early contribution is Metzler (1941).

[17] See Blinder and Maccini (1991) and Ramey and West (1998).

quite limited support for existing explanations of the excess variance of production puzzle.[18] For example, one way to explain the puzzle is to introduce cost shocks into the Linear-Quadratic model (e.g. Eichenbaum (1989)). However, there is very little evidence which shows that observable cost shocks, such as shocks to real wages or interest rates, have a significant effect on inventory investment (Ramey and West (1998)). An alternative explanation is that the cost function is non-convex (e.g. Ramey (1991)). Yet many studies find that estimated cost functions are convex (Blanchard (1983), West (1986)).

If it can be shown empirically that capital market imperfections are an important determinant of inventory investment, this would provide strong support for the theoretical argument that capital market frictions are essential to the understanding of aggregate fluctuations. There exists a body of reduced-form evidence to this effect. For example, Gertler and Gilchrist (1994) find evidence that small manufacturing firms draw down their inventory stocks heavily following a monetary contraction, whereas large firms appear to borrow in order to smooth the impact of a downturn on their inventory behaviour. Similarly, using US panel data, Carpenter, Fazzari, and Petersen (1994) find that the inventory investment of small firms is more sensitive to cash flow than is the inventory investment of large firms. Of course, these studies are subject to the reservations discussed in the previous section of this chapter.

The excess variance of production puzzle is a robust and widely reported fact. It is highly likely that it is not caused by one of the competing theories exclusively. Therefore, a key goal of research in this area is to identify restrictions on the data which can discriminate between theories. The empirical prediction of the model of Chapter 3 which distinguishes it from other models of inventory investment is the following. A shock to the supply of the firm's internal finance which is orthogonal to shocks affecting the firm's desired level of inventories will, nevertheless, affect inventory investment. Thus, to test this model, one must first solve precisely the same identification problem as that discussed in the previous section.

[18]Existing models can be grouped into three categories: (i) modifications of the Linear-Quadratic model, (ii) stockout-avoidance models, and (iii) (S,s) models.

Chapter 2

R&D Intensity and Finance: Are Innovative Firms Financially Constrained?

2.1 Introduction

Arguably, the class of firms for which the assumption of perfect capital markets is least likely to be satisfied is the class of innovative firms - firms which devote resources towards the production of innovative products or processes. For this reason economists have long thought that a firm's supply of internal finance will be one of the main determinants of its R&D expenditures. However, there is little empirical support for this belief. R&D expenditures are far less sensitive to fluctuations in internal finance than other expenditures, such as investment in physical capital. In their survey of the R&D literature, Kamien and Schwartz (1982) conclude that "the empirical evidence that either liquidity or profitability are conducive to innovative effort or output appears slim".

These findings are consistent with the assumption of perfect capital markets, together with the hypothesis that different expenditures are driven by different factors. However, they are also consistent with the alternative hypothesis that financially constrained firms adjust different expenditures disproportionately in response to a common factor, namely changes in supply of internal finance. Innovative firms may choose to smooth R&D expenditures by using other expenditures as a buffer against fluctuations in internal finance, either because the costs of adjusting R&D are higher or because the elasticity of the marginal benefit to R&D is higher than that of other factors of production.

Simply by comparing the statistical properties of the R&D and investment processes, it is not possible to distinguish between these hypotheses. However, an implication of the perfect capital markets assumption is that each of the firm's expenditures should be equally insensitive to fluctuations in the firm's supply of internal finance. Therefore, a possible test is to compare individually the sensitivities to internal finance of different expenditures of innovative firms with those of non-innovative firms.

For the purpose of this study, an innovative firm is defined to be a firm with a research agenda to develop global innovations. While the assumptions underlying the theory on capital market imperfections may apply equally to firms which develop local innovations, much of this innovative activity may not be carried out within a formal R&D framework, and therefore would not be identified by the stratification criterion used in this chapter. The criterion uses the ratio of R&D to total investment as a measure of R&D intensity. Firms which reported an R&D intensity greater than their respective industry average in each year from 1990 to 1993 are classified as innovative.

The sensitivity of physical capital investment expenditures to fluctuations in internal finance is then compared across the two classes of firms.[1] For robustness, three different investment equations are estimated. Two of these are derived from the Adjustment Costs model of investment, Tobin's q equation and the Euler equation

[1]In this sample the number of observations on R&D across time is insufficient to enable a similar analysis of the sensitivity of R&D to internal finance. A new accounting standard, SSAP 13, requiring large firms to disclose their R&D expenditures in their company accounts, came into effect in the UK only in 1989.

corresponding to the model, and the third is derived from an Accelerator model.

The results are consistent with the alternative hypothesis that innovative firms are financially constrained. Both the results from the q equation and from the Accelerator model show that the sensitivity of investment expenditures to internal finance is much higher for innovative firms than for non-innovative firms. The Euler equation results show that the Adjustment Costs model is rejected for both sub-samples. However, there is evidence that the model does fit the data of non-innovative firms better than that of innovative firms. In particular, the investment expenditures of innovative firms are much more sensitive to measures of free cash flow (cash flow less tax and interest payments), than is the investment of non-innovative firms.

The test in this paper is not a direct test of whether or not financially constrained firms underinvest in R&D. However, viewed in conjunction with other empirical work, the results provide indirect evidence that this is true. The presence of financing constraints may reduce the returns to R&D by hindering the ability of innovative firms to establish a large market share for a successful innovation before competitors introduce rival products. The responses to a survey of innovative firms in the US conducted by Levin, Klevorick, Nelson, and Winter (1987) show sales and service efforts to be one of the most important means of appropriating the returns to R&D.

The rest of the chapter is organised as follows. Theoretical issues regarding capital market imperfections and the determinants of R&D expenditures and physical capital investment are addressed in section 2.2. The empirical models and estimators are discussed in section 2.3. A description of the sample selection criteria and the results are given in section 2.4. The final section concludes.

2.2 Theoretical Issues

2.2.1 Capital Market Imperfections

One of the main assumptions upon which rest many of the results in the literature on capital market imperfections is the assumption of asymmetric information. When firms possess more information about the quality of an investment project than do potential investors, or when the firm can control variables which are not observable to the investor but which affect the return to the project, capital markets will be inefficient. Arguably, such conditions are most likely to be satisfied by firms which devote resources to innovating. The production of an innovation is more difficult to predict from observable inputs than is the production of most other types of output. Thus, there is greater scope for inputs which are not observable to all parties, such as a researcher's skill level or the choice of research agenda, to affect the returns to an investment in the development of an innovation. Moreover, given that many innovations are produced in technologically advanced industries, there are potentially large differences in the information sets of the different parties to a financial contract. This will limit the extent to which monitoring can reduce possible agency problems.

Under the assumption that managers have an informational advantage over investors regarding the quality of the potential investment projects the firms may undertake, Myers and Majluf (1984) show that equity markets will be inefficient. Given its informational disadvantage, the market requires all firms to issue equity at a discount. The discount can imply such a heavy dilution of the existing shareholders stake in the existing assets of the firm that it is not in their interest to undertake a positive NPV project. Stiglitz and Weiss (1981) show that asymmetric information leads to similar outcomes in debt markets. Again the key assumption in this model is that the market is at an informational disadvantage vis-à-vis the firm regarding the quality of the investment project for which debt finance is being sought (specifically, projects differ according to the variance of their returns). Creditors react to excess demand by rationing some borrowers rather than by raising interest rates. Raising interest rates increases the riskiness of the average investment project in the pool of credit applicants because applicants with "safe" projects drop out. Again in equilibrium positive NPV projects will be forgone.

An inherent part of an R&D project is the accumulation of knowledge. Knowledge is a public good, and the existence of patent systems is typically justified as a mechanism whereby firms which invest in knowledge capital can protect their investment (in legal parlance, the firm's intellectual property). However, patents work only imperfectly. In a survey of R&D investing firms in the US, Levin, Klevorick, Nelson, and Winter (1987) report that managers believed non-patent methods of protecting knowledge capital to be more important than patents. Those methods include the lead time a firm has over its rivals (i.e. differences in their knowledge capital), and the speed with which they accumulate knowledge. According to their study, innovative firms clearly possess intellectual property which is unprotected by patents, and which has an important impact on the value of its investment projects. It is equally true that such property cannot be appropriated by another party; it is the inalienable property of the firm.[2]

Hart and Moore (1994) have shown that, even in a model of debt with full information, positive NPV projects may still be forgone. The results of this model rest upon two assumptions: first, that the entrepreneur possess an asset which a creditor is unable to appropriate, and second, that this "inalienable" asset affect the value of assets that can be appropriated (i.e the firm's collateralisable assets). The threat that the entrepreneur may withdraw the inalienable asset from the production process can limit the debt capacity of the firm below the cost of the investment project.[3] Therefore, whether or not such an investment project is undertaken depends upon

---

[2]Strictly speaking, this type of intellectual property is the asset of the research staff within the firm. For simplicity, I assume that the researchers are also owner-managers of the firm.

[3]The strength of the threat to withdraw the inalienable asset depends upon the outside options available to the two parties. Levin, Klevorick, Nelson, and Winter (1987) report that one of the main channels of information spillover is through the hiring of rival firms R&D staff. This would indicate that the firm's bargaining position vis-à-vis creditors is strong. Other things being equal, such a firm's debt capacity would be low.

the amount of internal finance available to the entrepreneur.

Even if innovative firms could mitigate the effect of capital market imperfections by, for example, revealing some of their knowledge capital to parties outside the firm, doing so may not be optimal. Levin, Klevorick, Nelson, and Winter (1987) report that secrecy is also an important way firms protect their intellectual property, particularly for process innovations. Indeed, the importance of lead time over rivals suggests that revealing information may substantially reduce the value of the innovation. Bhattacharya and Ritter (1983) and Horstmann, MacDonald, and Slivinski (1985) present theoretical models in which it is not optimal for a firm to reveal all of its information, either through a third party such as a financial intermediary, or through patenting its innovations. Thus, innovative firms may prefer to use internal finance in order to protect their knowledge capital.

These theoretical arguments imply that internal funds will be an important source of finance for innovative firms. However, to what extent will firms be able to separately finance different investment projects? Firms which conduct R&D typically produce the product innovations and implement the process innovations which are the corresponding outputs of the firm's R&D input. Hence, the firm's innovations will affect the returns to its physical capital, and the returns to investment in new physical capital will depend upon the firm's future innovations. It is therefore unlikely that firms will be able to separately finance R&D projects and physical capital investment projects. If capital markets are imperfect for R&D projects, those imperfections will impact upon the firm's physical investment projects due to the interdependence of the returns to the two types of investment.

2.2.2 Determinants of R&D and Investment

The above theoretical arguments suggest that fluctuations in internal finance should be highly correlated with at least some of the expenditures of innovative firms. The arguments do not imply that all expenditures will be sensitive to internal finance. Moreover, they do not place any testable restrictions on differences in the statistical properties of different expenditures. Therefore, one cannot test the hypothesis of imperfect capital markets by comparing the statistical properties of different expenditures such as R&D and physical capital investment.

There is considerable empirical evidence showing that actual expenditures on R&D and physical capital respond to different, as well as to common, factors. For example, Lach and Schankerman (1989) show that while both R&D and investment respond to a shock which is permanent (in the sense that its impact is highly persistent), investment is influenced strongly by an idiosyncratic shock as well.[4]

Such evidence is consistent with both hypotheses. In general, under the hypothesis that innovative firms are not financially constrained, desired expenditures on R&D and physical capital will respond to different, as well as common, factors. For example, physical capital may be subject to productivity shocks which do not affect the expected marginal value of knowledge capital. However, it is equally plausible that firms adjust only their physical capital expenditures in response to a shock to its internal finance. Under the alternative hypothesis that innovative firms are financially constrained, the responses of actual expenditures on R&D and physical capital to shocks to the firm's internal finance will be proportionate only in the restricted case where the functions governing the firm's production and its costs of adjusting inputs are homothetic. There does not seem to be much evidence that this restriction is valid. If innovative firms' production functions were homothetic, the ratio of R&D to other inputs would be a constant function of firm size. However, as Pakes and Schankerman (1984) note, a stylised fact in the empirical literature is the "observation that the coefficient of [cross-sectional] variation of research intensity [R&D to sales ratio] is an order of magnitude larger than those of traditional inputs."

Moreover, there are reasons why financially constrained firms would prefer to adjust

[4]Similar evidence has been reported by other authors. Pakes (1985) found that changes in R&D were associated with large changes in the market value of the firm, indicating the market expected a persistant change in profits. Himmelberg and Petersen (1994) found evidence that R&D expenditures were sensitive to permanent (i.e highly persistent) changes in cash flow, not transitory changes. They interpreted this as evidence of higher costs of adjustment for R&D, although it is also consistent with the above null hypothesis.

physical capital investment. Given the importance firms place on secrecy and lead time over competitors, the implicit costs of laying off R&D staff may be much higher than the costs of adjusting physical investment. Researchers who are laid off may transmit valuable knowledge about firm's research programs to its competitors. Levin, Klevorick, Nelson, and Winter (1987) report that hiring a competitor's R&D personnel is viewed by many firms as one of the most effective channels of information spillover. In addition to possessing knowledge valued by competitors, it is likely that researchers acquire a great deal of firm specific knowledge which would imply substantial training costs associated with hiring new staff. A widely reported fact which is consistent with these observations is that within-firm variance of R&D is much lower than that of investment (see, for example, Lach and Schankerman (1989), Himmelberg and Petersen (1994), Hall (1992)).

In conclusion, the observed stochastic properties of investment and R&D are consistent with both hypotheses: (i) capital markets are perfect, but some of the determinants of desired R&D and investment expenditures differ (as do the properties of the stochastic processes governing the respective determinants), and (ii) capital markets are imperfect, but actual R&D and investment expenditures respond differently to the common determinant of internal finance shocks. Therefore, an appropriate test to distinguish between them is to examine the sensitivity of the expenditures to changes in the firm's supply of internal finance. In view of the lower within-firm variability of R&D, performing this test on investment will be more powerful.

2.3 Empirical Specification

2.3.1 Model Specification

To distinguish between the alternative and null hypotheses, this chapter examines how different empirical models of investment fit the data of different sub-samples of firms. In common with much of the empirical investment literature, two of the empirical equations used in this chapter are derived from the Adjustment Costs model of investment (expositions of this model are given in Eisner and Strotz (1963), Lucas (1967), and Gould (1968)): the q equation, and the Euler equation corresponding to the Adjustment Costs model.

The q equation is the first-order condition of the Adjustment Costs model under three assumptions: (i) linear homogeneity of the profit and cost of adjustment functions, (ii) the firm is a price-taker in both input and output markets, and (iii), perfect capital markets.[5] Under the third assumption the market value of the firm equals its fundamental value (the expected present discounted value of future cash flows), while under the first two assumptions observable average q and unobservable marginal q are equal (see Lucas and Prescott (1971), Hayashi (1982)). The following quadratic form for the cost of adjustment function is assumed, $G(I, K; \varepsilon) = \frac{\varphi}{2}(\frac{I}{K} - \alpha - \epsilon)^2 K$, where I is investment, K the existing stock of capital, α a firm specific constant, and ϵ a random variable. With this functional form, the first order condition can be written as,

$$\left(\frac{I}{K}\right)_{it} = \alpha_i + \beta Q_{it} + \epsilon_t, \tag{2.1}$$

where β is the inverse of the marginal cost of adjustment, φ, and Q_{it} is the ratio of the market value of the firm to the replacement cost of its existing assets.

Under the null, the theory predicts that Q_{it} is a sufficient statistic for the firm's investment rate. In practice, however, the error involved in measuring Q may be very large. Factor analytic empirical studies have shown that the factor common to the stock price and investment accounts for only a small percentage to the total variation in the stock price. The vast majority of price variation is accounted for by an idiosyncratic factor (see Blanchard, Rhee, and Summers (1990) and Lach and Schankerman (1989)). An advantage of estimating the Euler equation is that it can be done without using observations on the firm's stock price.

[5]These functions are defined in the appendix, where both the first-order condition and the Euler equation are derived. Sufficient conditions for the linear homogeneity of the profit function are constant returns to scale and perfect competition in all input and output markets. An assumption in this particular derivation of the model is that newly installed capital becomes productive immediately.

To make the Euler equation estimable, either a functional form for the profit function can be assumed (as in Abel (1980)), or the linear homogeneity assumption made above can be maintained (as in Bond and Meghir (1994)). In fact, the linear homogeneity of the profit function can be relaxed to allow for imperfect competition in the output market.[6] With this linear homogeneity assumption, and using the same functional form for the costs of adjustment function, the Euler equation can be written as,

$$\begin{aligned}\left(\frac{I}{K}\right)_{it} &= \beta_1\left(\frac{I}{K}\right)_{it-1}+\beta_2\left(\frac{I}{K}\right)^2_{it-1}+\beta_3\left(\frac{CF}{K}\right)_{it-1} \\ &\quad +\beta_4\left(\frac{Y}{K}\right)_{it-1}+\alpha_i+d_t+v_t, \end{aligned} \tag{2.2}$$

where I represents investment, CF cash flow, and Y output. All variables are weighted by the firm's capital stock, K. The term α_i picks up firm specific effects, while d_t is a time dummy to control for fluctuations in the omitted price and user cost of capital terms (see appendix 2.A). One interpretation for the error term, v_t, is that it is an expectational error (i.e it represents the effect of new information acquired in period t). As such v_t should be i.i.d. However, v_t may also incorporate the effects of random variables such as the error term in the costs of adjustment function, and therefore may not be i.i.d.

The theoretical model places the following restrictions on the signs of the coefficients in the above equation: (i) the coefficient on the lagged investment rate, β_1, is positive and greater than one; (ii) the coefficient on the square of the lagged investment rate, β_2, is negative and greater that one in absolute value; (iii) the coefficient on cash flow, β_3, is negative, and (iv) the coefficient on output, β_4, is positive if there is imperfect competition (otherwise it is zero).

A disadvantage of testing the structural equations from the Adjustment Costs model is that both equations are not well specified under the alternative. That is, if the restrictions implied by the model are rejected by the data it is not possible to

[6]Both the production function and the costs of adjustment functions are still assumed to be linear homogeneous, and input markets perfectly competitive.

determine which assumption has been violated. It is possible that assumptions other than that of perfect capital markets have been violated. Therefore, as a check on the results of the regressions based on the Adjustment Costs model, a third and simpler model of investment is estimated: the Sales Accelerator model. A theoretical justification for this model can be derived under similar assumptions above, but without the assumption of adjustment costs (a good reference for this and other investment models is Nickell (1978)). Dropping that assumption eliminates the need to deal with expectations since the capital stock can be costlessly adjusted instantaneously. The factors driving investment are then the exogenous processes driving prices, the user cost of capital, and demand for the firm's output. Moreover, with no adjustment costs and under the null, changes in output are driven entirely by exogenous shocks to demand. The empirical model can be written as,

$$\left(\frac{I}{K}\right)_{it} = \alpha_i + d_t + \beta\left(\frac{\Delta Y}{Y}\right)_{it} + \mu_t \tag{2.3}$$

where again α_i picks up a firm specific effect, d_t picks up fluctuations in prices and the user cost of capital, and the growth rate of sales, $\Delta Y/Y$, captures exogenous demand shocks. Under the null hypothesis, cash flow contains the same information as the growth rate of sales and therefore should not enter significantly. However, under the alternative hypothesis cash flow will enter significantly if financing constraints are binding.

2.3.2 Estimators

An instrumental variables estimator is required for both of the empirical equations derived from the Adjustment Costs model. In the q equation, if ϵ_t is realised at the beginning of the period (i.e. is in the firm's information set in period t), then Q_{it} will be endogenous through the effect ϵ_t has on the actual amount of capital installed.[7] Valid instruments for Q_{it} will be variables with a lag or lead of j or greater, where

[7]The error term, ϵ_t, is interpreted as factors which are observable to the firm, but not to the econometrician. Alternatively, it can be assumed that ϵ_{t-1} is in the firm's information set and that ϵ is serially correlated.

j is the lowest lag for which $cov(\epsilon_t, \epsilon_{t-j}) = 0$. The theory places no restrictions on serial correlation in the ϵ process.

Similarly in the Euler equation, the standard within-groups estimator is biased due to the presence of the lagged dependent variable. Differencing the model to remove the fixed effect necessitates the use of instruments for the lagged dependent variable terms. As in the estimation of the q equation, lagged values can be used as instruments. The validity of the instruments will depend on the degree of serial correlation in the error term.

Unlike the first two models, an instrumental variables estimator is not theoretically required for the Sales Accelerator equation, since sales growth in that model is assumed to be strictly exogenous. However in this equation, as in the previous two models, under the alternative hypothesis the cash flow terms may be endogenous. Therefore, as in the other regressions the cash flow terms are instrumented with their lags, and so can be interpreted as the effect of *predictable* cash flow on investment. Instrumenting cash flow in this way should eliminate any "informational" effect it may have on investment.

All of the above models are differenced to remove the firm-specific effect, α_i, and an Anderson-Hsiao estimator is used (Anderson and Hsiao (1982))(except for the basic Sales Accelerator model, which is estimated by OLS). While this estimator is not efficient, it is consistent. Monte Carlo experiments done by Arellano and Bond (1991) show that in dynamic models the loss of efficiency is not great if the coefficient on the lagged dependent variable is not too close to one (a condition which is satisfied in all of the estimated dynamic models).

The Anderson-Hsiao estimator differs from the efficient estimator in two ways. First, the GMM estimator which imposes the moment condition for each available instrument *in each time period* is the efficient estimator in the class of instrumental variable estimators which use only linear combinations of instrumental variables (see Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bond (1991)).[8] In contrast, the

[8]For example, if there were five observations per firm, and instruments of lag two or greater

the moment conditions imposed by the Anderson-Hsiao estimator are the sum of the individual time period moment conditions corresponding to a particular lagged value of the instrument (i.e. there is one moment condition per instrument, as opposed to one moment condition for each year in which the instrument is available).

Second, the efficient GMM estimator uses an estimate of the optimal weighting matrix based on the residuals generated by an initial consistent estimator. The (arbitrarily chosen) weighting matrix for the Anderson-Hsiao estimator has two's on the principal diagonal, and one's on the main off-diagonals.[9] The standard errors of this estimator are White-corrected for heteroscedasticity.

The choice of the less efficient Anderson-Hsiao estimator was made for two reasons. First, the standard errors of the efficient GMM estimator produced by the estimation routine employed in this paper are biased downwards.[10] Second, in practice the large set of valid instruments typically available to the efficient estimator can raise difficulties. One of the problems encountered in this paper was the unreliability of test statistics, such as the Sargan statistic, when a large set of instruments were used.

One of the specification tests reported is the Sargan statistic (referred to as the J-statistic in Hansen (1982)). However, because the validity for the instruments depends crucially on serial correlation in the error term, two estimates of this serial correlation are also reported: a test for first-order serial correlation, the $m1$ statistic, and a test for second-order serial correlation, the $m2$ statistic (see Arellano and Bond (1991)). If the error term in the undifferenced model is white noise, then the error term in the differenced model should exhibit first-order, but not second-order, serial

---

were valid, then this estimator would impose six moment conditions: there is one valid instrument for the observation in period 3 (lag 2), two for the observation in period 4 (lags 2 and 3), and three for the observation in period 5 (lags 2, 3 and 4). If the instrument set were restricted to only the second lag, there would be three moment conditions.

[9] If the models are exactly identified, then the orthogonality conditions are set exactly to zero and the weighting matrix is the identity matrix.

[10] Estimation was performed using the DPD routine developed by Manuel Arellano and Stephen Bond (Arellano and Bond (1988)). The bias in the standard errors of the two-step GMM estimator is reported in Arellano and Bond (1991).

correlation.

2.4 Data and Results

2.4.1 Sample Design

The empirical methodology is the same as that which is commonly used in the literature on financing constraints and investment. A sample of firms is first identified, and then stratified according to a particular criterion, which in this analysis is whether or not the firm is innovative. Conclusions are then based on a comparison of the results from the group of "controls" (non-innovative firms) with those from the group of "experimentals" (innovative firms). This methodology requires sampling choices be made concerning the class of firms from which the initial sample is drawn, and the criteria used to stratify the sample.

For the purpose of this study, an innovative firm is defined to be a firm with a research agenda to develop *global* innovations.[11] While the assumptions underlying the theory on capital market imperfections may apply equally to firms which develop local innovations, much of this innovative activity may not be carried out within a formal R&D framework, and therefore would not be identified by a stratification criterion based on firm R&D expenditures. For this reason each firm, even those in the control group, was required to have at least one positive observation on R&D to be included in the sample. This is designed to reduce the number of firms which develop local innovations (and therefore may be financially constrained), but which either do not expense the costs under R&D in their accounts, or are not required to report their R&D expenditures.[12]

---

[11] Paul Stoneman offers the following distinction. " Global innovation would be the first occurrence in an economy (or even wider in the world economy) of a particular event...Local innovation would be the the first occurrence of the event in the unit of observation." Stoneman (1995), p.3.

[12] In January 1989 a new standard of accounting practice was introduced in the United Kingdom, the Statement of Standard Accounting Practice (SSAP 13), which requires certain firms to report their R&D expenditures. SSAP 13 did not alter the conditions under which R&D may be capitalized rather than expensed. Few firms in the sample reported capitalized R&D expenditures.

By focusing on the aims of the firm's research agenda, the definition of an innovative firm includes firms which may not actually develop a global innovation. The successful development of an innovation may serve to mitigate the effect of capital markets imperfections. For example, in an hidden information context successful innovation may reveal the firm's type. Consequently, constraints on the access to external capital may not differ between groups of firms identified as innovative and non-innovative according to a definition based on innovative output (patent counts or innovation counts).[13]

Accordingly, the stratification criterion appropriate for this study uses only information on a firm's inputs to the innovation process. The following measure of a firm's R&D intensity is used: the ratio of R&D expenditures to total investment (physical investment plus R&D expenditures). Innovative firms were defined to be those firms whose R&D intensity was above its corresponding industry mean for each year in which the firm reported a positive R&D expenditure. Implicit in this stratification criterion is the assumption that firms attempting to develop global innovations must spend more on R&D than other firms within the same industry. Firms with a low R&D intensity are assumed to have research programs aimed at replicating the existing technology of competitors. Using this criterion, 42 of 144 firms in the sample were identified as innovative. Table 2.1 shows the breakdown of observations for the entire sample, and for the two subsamples. Details of the sample selection procedures are given in the Data Appendix.

2.4.2 Results

Tables 2.2 and 2.3 show the results for the Q regression (equation (2.1)) for innovative and non-innovative firms respectively. Comparing the first column of the two tables we see that the point estimates of β are very similar for the two classes of firms. However, the test statistics for the class of non-innovative firms clearly reject

---

[13] Using data on actual innovations, Blundell, Griffith, and Van Reenen (1993) found that the probability of a firm producing an innovation in a period is greatly increased if the firm had innovated in previous periods.

the instrument set. The Sargan statistic rejects at the 1% level. The $m2$ statistic narrowly fails to reject at the 5% level. To deal with the potential serial correlation, it is assumed that the error term in equation (2.1) follows an AR(1) process, $\epsilon_t = \rho\epsilon_{t-1} + \nu_t$, and a Cochrane-Orcutt transformation is performed. Using equation (2.1) to solve for ϵ gives,

$$\left(\frac{I}{K}\right)_{it} = (1-\rho)\alpha_i + \beta Q_{it} + \rho\left(\frac{I}{K}\right)_{it-1} - \rho\beta Q_{it-1} + \nu_t. \tag{2.4}$$

This specification is used in column 2 of Table 2.3. Although the $m2$ statistic now gives no indication of second order serial correlation, the Sargan statistic still rejects the validity of the instrument set. In column 3 the second lag is dropped from the instrument set, and in column 4 the fourth lag is dropped. The Sargan statistic fails to reject either of these two new instrument sets. However, the $m1$ statistic fails to reject first-order serial correlation for the specification of column 3. Since the model is estimated in first-differences, this indicates misspecification. Therefore, the preferred specification for the class of non-innovative firms is that of column 4. Although the common factor restriction on the coefficients in equation (2.4) is not tested, it seems likely that it would not be rejected since the product of the coefficients on Q_t and $(I/K)_{t-1}$ is 0.0039.

To test the null hypothesis against the alternative, free cash flow was included in these regressions. A comparison of column 2 in Table 2.2 with column 5 in Table 2.3 shows that, while the null hypothesis can be rejected for both classes of firms, the effect of free cash flow on the investment of innovative firms is quantitatively larger (the coefficient on current free cash flow is 0.5712 for innovative firms as opposed to 0.1502 for non-innovative firms). Lagging the free cash flow term does not change this result, nor does dropping the second lag of the regressor from the instrument set for innovative firms (columns 5 to 7 in Table 2.2). A Wald test on the joint significance of both free cash flow terms is highly significant for innovative firms (19.627 for the specification of column 4, and 13.131 for the specification of column 7). For non-innovative firms the Wald test is 1.834, rejecting the joint significance of the cash flow terms at the 1% level.

These results are consistent with the hypothesis that innovative firms either have a preference for internal finance or face greater constraints on their access to external capital. The significance of cash flow for non-innovative firms indicates that the assumption of perfect capital markets may be inappropriate for this class of firms as well. However, it is possible that some innovative firms were included in the non-innovative group.[14]

The results for the Euler equation regression, given in Table 2.4, are consistent with the Q equation results.[15] Looking at the first and fourth columns, it can be seen that the Adjustment Costs model does not fit the data well for either class of firm. None of the estimated coefficients are within two standard errors of the range predicted by the theory. Moreover, the coefficient on the cash flow term has the wrong sign. The Euler equation results show that the previous results are not due to the poor empirical performance of Q.

As with the Q equation results, there is some evidence that the model fits the data for non-innovative firms somewhat better than that for innovative firms. In the second and fifth columns, the second lag is removed from the set of instruments. Although the specification tests do not reject the validity of instruments dated $t-2$, dropping these instruments changes the point estimates of the two lagged endogenous variable terms for the non-innovative class.[16] The point estimates are now within the range predicted by the theory. Though less precisely estimated, both coefficients are still significant at the 5% level. By comparison, in column 5 the coefficients estimated from the data on innovative firms remain well outside the range predicted by the

[14]The estimated coefficients on the cash flow terms of non-innovative firms is very close to those reported by Blundell, Bond, Devereux, and Schiantarelli (1992) in their study of the Q equation using UK manufacturing firms.

[15]In their paper on firm investment, Bond and Meghir (1994) show that the firm's debt policy can be incorporated into the Euler equation by including the term $(B/K)^2_{t-1}$. All the regressions reported in Table 2.4 were run with this term included, but its coefficient was never statistically significant.

[16]If the error term were an MA(1) process, values of the regressors in period $t-2$ would be invalid, but higher order lags would remain valid. Most of the bias arising from using the invalid $t-2$ instruments is likely to be manifested in the coefficients on the lagged endogenous terms. However, the Hausman test statistic also fails to reject the validity of these instruments, taking a value of 3.118 with 2 degrees of freedom.

theory.

It is not possible to deduce from this evidence which of the assumptions underlying the model is violated.[17] However, the most likely reason for the difference in the model's fit to the data of the different classes of firms the greater impact of financing constraints on innovative firms. In columns 3 and 6, the cash flow term is replaced by free cash flow (the difference being interest payments and taxes). This has no effect on the coefficient estimates for non-innovative firms, but increases the value of the estimate of the cash flow coefficient of innovative firms. This indicates that revenue lost to taxes and interest payments affects the investment expenditures of these firms. Second, while it is plausible that other factors which could explain this difference, such as differences in the costs of adjusting physical capital, may be associated with R&D intensity, it is more likely that much of the variation in other factors is accounted for by variation in industry averages. Since the stratification rule compared a firm's R&D intensity to its industry average, most industries are represented in both subsamples. To control for any residual inter-industry variation industry dummies were included in the estimation.

The results for the Accelerator Model, presented in Table 2.5, also support the hypothesis that innovative firms are financially constrained. Columns 1 and 4 show the results for the basic model, estimated using ordinary least squares. The sales growth term is positive and significant in both regressions. The results with free cash flow included in the specification are in columns 2 and 5.[18] As with the Q model results, the effect of free cash flow on the investment of innovative firms is quantitatively much larger (0.676 as opposed to 0.2996). Although this is not a forward looking model, by instrumenting free cash flow any information about the

[17]Note that because the (Y/K) controls for either imperfect competition or departures from a constant returns to scale production function, the only remaining assumptions are the assumed form of the cost of adjustment function, and perfect capital markets.

[18]Unlike the Q model, there is no theoretical argument that the sales growth term is endogenous. The free cash flow terms, however, may still be endogenous. Therefore, cash flow is instrumented with its lagged values in the subsequent regressions. The $m2$ statistic indicates that these instruments may not be valid for low intensity firms. A Cochrane-Orcutt transformation was performed. This eliminated the second order serial correlation. The point estimates of the coefficients on the free cash flow and sales growth terms were virtually identical.

future marginal productivity of capital has been purged from the contemporaneous free cash flow terms. The results in columns 3 and 6 show that lagged free cash flow is also significant, although it has a negative sign for innovative firms. This confirms that the results in the Q equation regressions are not due to the poor empirical performance of the measured Q variable.

2.4.3 Robustness Checks

Reservations concerning the validity of the inferences made thus far arise from two potential problems in the sample design. The estimator used may be inconsistent due to endogenous sampling (selection bias). Alternatively, the results may be spurious due to inadequate control of other exogenous firm characteristics which affect the firm's access to external capital.

There are two potential sources of selection bias: (i) survivorship in the panel, and (ii) the endogeneity of the sample stratification rule. The first source is due to the new accounting standard, SSAP 13. Prior to its introduction in 1989 no firm was obligated to report R&D expenditures. Consequently, firms which conducted R&D prior to 1990, but which did not survive until this time, are excluded from the sample. Moreover, SSAP 13 does not require all firms to report their R&D expenditures, though it applies to all the firms in this sample.[19]

Unfortunately, simple tests of selection bias suffer heavily from a lack of power.[20] However, it is not necessarily the case that non-randomness in the sample implies that conventional estimators are inconsistent. Inconsistency results only if the sample selection rule is dependent on the endogenous variables in the structural equa-

[19]The standard applies "in effect to companies which are public limited companies, or special category companies, or subsidiaries of such companies, or which exceed by a multiple of ten the criteria for defining a medium-sized company under the Companies Act 1985".

[20]Verbeek and Nijman (1992b) suggest "quasi-hausman" tests comparing, for example, the estimates from a fixed effect estimator for the unbalanced panel and the balanced sub-panel. The term "quasi-hausman", and the tests' lack of power, stem from the fact that unlike Hausman tests both estimators are inconsistent under the alternative.

tions.[21] Unlike samples derived from survey data such as the income experiment data used for early studies of attrition bias (e.g. Hausman and Wise (1979)), there are few a priori reasons to believe this is true for this sample. Moreover, potentially biased estimates due to survivorship or attrition affects all studies which use panel data collected from publically available company accounts. Restricting analysis to balanced panels, as many studies have done, does not affect the consistency of the estimator.

Another possible source of selection bias is the potential endogeneity of the rule used to split the sample into high and low intensity classes of firms. Since the measure of R&D intensity was defined to be the ratio of R&D to R&D plus investment, this measure may not be independent of the endogenous variables in the structural equations (i.e. $\Delta(I/K)_{it}$). However, the potential asymptotic bias for the fixed-effect estimator used in the analysis may not be very severe. Since the the stratification rule compared the R&D intensity in each year over the period 1990 to 1993 with the firm's corresponding industry average over the same period, most of the variation in R&D intensity is between-firm.[22] Firm specific fixed effects explain approximately 95% of the variance of R&D intensity.[23] The greater the proportion of R&D intensity explained by a firm-specific effect, the smaller will be the asymptotic bias of the fixed-effect estimator (see Verbeek and Nijman (1992b)). In the limit the fixed-effect estimator will be consistent.

Another reservation about the conclusions drawn is the degree to which they may be ascribed to factors other than innovativeness. One factor which has been commonly found to be associated with the presence of financing constraints is firm size. Table 2.6 show the average firms sizes for the two groups at in 1987 and 1993 (over these

[21]For a detailed discussion see Verbeek and Nijman (1992a).

[22]This will have removed much of the within-firm source of variation. However, because the period over which R&D intensity was averaged is so short (for some firms there was only 1 observation) not all of the within-firm source of variation will have been removed.

[23]This estimate is from a simple one way error components model. All data on R&D intensity was used, even that for firms excluded from the sample. Other empirical studies have found similar results. For example, using a slightly different measure of R&D intensity, the R&D to sales ratio, Pakes and Schankerman (1984) found that over 95% of the structural variance in R&D intensity was accounted for by the variance of a firm-specific structural parameter.

seven years the panel was at its maximum width). Looking at either the book value of the capital stock or total sales the average non-innovative firm is twice as large as the average innovative firm. Perhaps the difference in the results of the structural regressions is due to this size difference.

To check this the sample was stratified according to the initial observation of a firm's capital stock. The 48 firms in the lowest third of the distribution were classified as small. The results of the Q equation regressions for small and large firms are given in Table 2.7. In contrast to many of the results in the investment literature, the coefficient on free cash flow is insignificant for small firms, but is strongly significant for large firms. Since the division of small and large firms is roughly proportionate across the innovative and non-innovative classes of firms (thirteen small firms are in the group of innovative firms) it is unlikely that those results may be attributed to the discrepancy in average size.

2.5 Conclusions

The assumption of perfect capital markets is least likely to be satisfied for the class of firms which devote resources towards the development of innovative products or processes. An implication of this assumption is that each of the firm's expenditures should be insensitive to fluctuations in internal finance. To test this hypothesis existing empirical work either has examined the sensitivity of R&D expenditures to internal finance or has compared the properties of the statistical process of R&D with those of the processes of other expenditures such as physical capital investment. Neither of these procedures can distinguish between the two hypotheses: (i) that capital markets are perfect, and that different factors drive the firm's different expenditures, and (ii) that capital markets are imperfect, and that the different expenditures of the firm repond disproportionately to a common factor, namely shocks to the supply of internal finance.

To distinguish between these hypotheses, the sensitivity of physical investment expenditures to internal finance have been compared across innovative and non-innovative firms. For robustness, several investment equations were estimated. The results from the Q model and the Accelerator model show that the sensitivity of the investment of innovative firms to internal finance is much higher than that of non-innovative firms. The Euler equation results show that, although the Adjustment Costs model is rejected for both classes of firms, it fits the data of non-innovative firms better than that of innovative firms. There is evidence that the difference in the fit across the different classes of firm is due to the impact of financing constraints on innovative firms.

Table 2.1: Number of Observations in Samples

Full Sample

No. of obs	7	8	9	10	11	12	13	15	16	18	19	21	22
No. of companies	9	10	7	4	4	5	3	2	2	4	2	5	87

High Intensity Firms

No. of obs	7	8	9	10	11	12	15	18	19	21	22
No. of companies	4	3	2	1	2	1	1	2	2	1	21

Low Intensity Firms

No. of obs	7	8	9	10	11	12	13	15	16	18	21	22
No. of companies	5	7	5	3	2	4	3	1	2	2	4	66

Table 2.2: Q Equation - High Intensity Firms

	1	2	3	4	5	6	7
Q_t	0.0161	0.0066	0.0093	0.0086	0.0076	0.0038	0.0032
	(0.0064)	(0.0081)	(0.0044)	(0.0054)	(0.0107)	(0.0077)	(0.0079)
$(FCF/K)_t$	-	0.5712	-	0.1306	0.6174	-	-0.2545
	-	(0.1506)	-	(0.4791)	(0.1763)	-	(0.6389)
$(FCF/K)_{t-1}$	-	-	0.3707	0.2958	-	0.5684	0.7478
	-	-	(0.0832)	(0.2897)	-	(0.1385)	(0.437)
ml	-3.652	-3.53	-3.772	-3.296	-3.676	-3.815	-4.326
m2	0.614	0.902	0.187	0.321	0.933	-0.171	-0.356
Sargan	2.773	1.664	2.894	2.471	1.543	0.656	0.269
prob	(0.25)	(0.797)	(0.576)	(0.481)	(0.462)	(0.72)	(0.604)

1. The dependent variable is $(I/K)_t$. Sample period is 1972-1993. There are 686 observations.
2. Time and Industry dummies included in estimation. White-corrected standard errors are in brackets.
3. Lags 2,3, and 4 of the regressor are used in columns 1 to 4. Lags 3 and 4 in columns 5 to 7.
4. The $m1$ and $m2$ test statistics are normally distributed around zero.
5. The Wald test statistic for the joint significance of the two cash flow terms in column 4 is 19.627, and in column 7 is 13.131.

Table 2.3: Q Equation - Low Intensity Firms

	1	2	3	4	5	6	7
Q_t	0.0193	0.0221	0.129	0.0223	0.0177	0.0171	0.0136
	(0.0071)	(0.008)	(0.1446)	(0.0077)	(0.0065)	(0.0064)	(0.0107)
$(I/K)_{t-1}$	-	0.2011	-0.2331	0.1751	0.1674	0.1717	0.1985
	-	(0.036)	(0.2102)	(0.0414)	(0.0414)	(0.0387)	(0.0444)
Q_{t-1}	-	-0.0092	-0.0746	-0.005	-0.004	-0.0054	-0.0159
	-	(0.0042)	(0.0925)	(0.0042)	(0.0042)	(0.0041)	(0.0118)
$(FCF/K)_t$	-	-	-	-	0.1502	-	-1.2373
	-	-	-	-	(0.0778)	-	(1.1255)
$(FCF/K)_{t-1}$	-	-	-	-	-	0.1061	0.9045
	-	-	-	-	-	(0.0486)	(0.7542)
m1	-6.393	-7.575	-1.165	-7.459	-7.281	-7.236	-1.807
m2	-1.838	-0.589	-1.967	-0.777	-0.867	-0.812	0.9
Sargan	9.261	11.01	0.045	2.875	5.777	5.57	0.706
prob	(0.01)	(0.012)	(0.831)	(0.09)	(0.123)	(0.135)	(0.703)

1. The dependent variable is $(I/K)_t$. Sample period is 1972-1993. There are 1894 observations.
2. Time and Industry dummies included in estimation. White-corrected standard errors are in brackets.
3. Lags 2,3, and 4 of the regressors are used in columns 1 and 2, lags 3 and 4 in columns 3, and lags 2 and 3 in columns 4 to 7.
4. The $m1$ and $m2$ test statistics are normally distributed around zero.
5. The Wald test statistic for the joint significance of the two cash flow terms in column 7 is 1.834.

Table 2.4: Euler Equation

	Low Intensity			High Intensity		
$(I/K)_{t-1}$	0.5554	1.2064	0.5864	0.0818	0.3659	0.1055
	(0.134)	(0.5387)	(0.1386)	(0.2112)	(0.9165)	(0.2182)
$(I/K)^2_{t-1}$	-0.5192	-1.919	-0.5577	-0.0224	-0.0956	-0.0243
	(0.2127)	(0.9493)	(0.2205)	(0.2836)	(1.3172)	(0.2928)
$(CF/K)_{t-1}$	0.2198	0.3232	0.2195	0.2475	0.1839	0.3363
	(0.056)	(0.1084)	(0.0726)	(0.0714)	(0.2155)	(0.078)
$(Y/K)_{t-1}$	-0.0067	-0.0165	0.0001	0.0105	0.0015	0.0199
	(0.0055)	(0.0098)	(0.0048)	(0.0114)	(0.0151)	(0.0078)
m1	-7.443	-2.215	-7.604	-4.522	-1.726	-4.479
m2	-0.749	-1.875	-0.826	1.259	0.699	1.177
Sargan	5.441	0.708	7.284	2.816	2.976	3.47
prob	(0.71)	(0.95)	(0.506)	(0.945)	(0.562)	(0.902)

1. The dependent variable is $(I/K)_t$. Sample period is 1972-1993. There are 686 observations for high intensity firms and 1894 observations for low intensity firms.
2. Time and Industry dummies included in estimation. White-corrected standard errors are in brackets.
3. Lags 2,3, and 4 of the regressors are used in columns 1,3,4 and 6. Lags 3 and 4 in columns 2 and 5.
4. The $m1$ and $m2$ test statistics are normally distributed around zero.
5. The Hausman test statistic for the validity of the lag 2 instruments is 3.118 (2) for low intensity firms and 0.557 (2) for high intensity firms.

Table 2.5: Sales Accelerator Model

	Low Intensity			High Intensity		
$(\Delta Y/Y)_t$	0.0608	0.037	0.0541	0.1245	0.0442	0.1381
	(0.0264)	(0.0251)	(0.0258)	(0.0408)	(0.0448)	(0.0415)
$(FCF/K)_t$	-	0.2996	-	-	0.676	-
	-	(0.0871)	-	-	(0.154)	-
$(FCF/K)_{t-1}$	-	-	0.2116	-	-	-0.0706
	-	-	(0.0598)	-	-	(0.0224)
m1	-6.267	-6.209	-6.219	-3.549	-3.614	-3.529
m2	-1.859	-1.955	-1.793	0.634	0.88	0.663
Sargan	-	3.035	4.137	-	0.018	1.029
prob	-	(0.219)	(0.126)	-	(0.991)	(0.598)

1. The dependent variable is $(I/K)_t$. Sample period is 1972-1993. There are 686 observations for the high intensity firms and 1894 observations for the low intensity firms.
2. Time dummies included in estimation. White-corrected standard errors are in brackets.
3. Lags 2,3, and 4 of free cash flow are used in columns 2,3,5, and 6. OLS is used in columns 1 and 4.
4. The $m1$ and $m2$ test statistics are normally distributed around zero.

Table 2.6: Summary Statistics

	Innovative Firms		Non-Innovative Firms	
	Mean	S.d.	Mean	S.d.
Capital Stock 1987	99.2	22.8	284.9	93.6
Capital Stock 1993	238	73.9	566.7	189.5
Total Sales 1987	371.6	76.4	861.2	309.1
Total Sales 1993	689.7	158.7	1348.2	489.6
R&D Intensity	0.257	0.196	0.239	0.187

Table 2.7: Q Equation - Robustness Check

	Small Firms			Large Firms		
Q_t	0.0177	0.0156	0.015	0.0358	0.0294	0.03
	(0.0057)	(0.005)	(0.0046)	(0.009)	(0.0081)	(0.0075)
$(FCF/K)_t$	-	0.1102	-	-	0.3778	-
	-	(0.1054)	-	-	(0.1186)	-
$(FCF/K)_{t-1}$	-	-	0.076	-	-	0.2221
	-	-	(0.0727)	-	-	(0.0688)
m1	-4.441	-4.378	-4.481	-5.504	-5.387	-5.746
m2	-0.456	-0.456	-0.466	-1.034	-0.965	-1.292
Sargan	0.124	0.079	0.153	1.718	3.565	4.24
prob	0.725	0.961	0.926	0.19	0.168	0.12

1. The dependent variable is $(I/K)_t$. Sample period is 1972-1993. There are 686 observations.
2. Time and Industry dummies included in estimation. White-corrected standard errors are in brackets.
3. Lags 2,3, and 4 of the regressor are used in columns 1 to 4. Lags 3 and 4 in columns 5 to 7.
4. The $m1$ and $m2$ test statistics are normally distributed around zero.

2.A Adjustment Costs Model

The firm's problem is to maximise,

$$V_t = E_t \sum_{j=0}^{\infty} \beta^{-j} \Pi_{t+j}, \tag{2.5}$$

where β is a discount factor, assumed constant, and

$$\Pi_t = p_t F(K_t, L_t; \mu_t) - p_t G(I_t, K_t; \epsilon_t) - w_t L_t - p_t^I. \tag{2.6}$$

$F(\cdot, \cdot; \cdot)$ is the firm's production function, $G(\cdot, \cdot; \cdot)$ the cost of adjustment function, I investment, K the capital stock, L labour, p output price, p^I the price of new capital, w the wage, and μ and ϵ are random disturbances. The firm's problem can be formulated as the following dynamic programming problem,

$$V(K_{t-1}; \mu_t, \epsilon_t) = \max_{I,L} \{\Pi_t + \beta E_t[V(K_t; \mu_{t+1}, \epsilon_{t+1})]\}, \tag{2.7}$$

subject to (2.6), the law of motion for capital,

$$K_{t+1} = (1 - \delta)K_t + I_t, \tag{2.8}$$

and the stochastic processes governing the two disturbance terms. Assuming that capital invested in period t also becomes productive in period t, the first-order condition for investment is,

$$\left(\frac{\partial \Pi}{\partial I}\right)_t + \left(\frac{\partial \Pi}{\partial K}\right)_t + \beta E_t \left(\frac{\partial V}{\partial K}\right)_{t+1} = 0, \tag{2.9}$$

and the Euler equation is,

$$\left(\frac{\partial V}{\partial K}\right)_t = (1 - \delta) \left(\frac{\partial \Pi}{\partial K}\right)_t + \beta(1 - \delta) E_t \left(\frac{\partial V}{\partial K}\right)_{t+1}. \tag{2.10}$$

Using equation (2.10), the first-order condition can be re-written as,

$$-(1-\delta)\left(\frac{\partial\Pi}{\partial I}\right)_t = \left(\frac{\partial V}{\partial K}\right)_t . \tag{2.11}$$

This is the equation tested by the empirical q equation. To derive the latter from equation (2.11), assume that both the production function and the costs of adjustment function are linearly homogeneous in their arguments. This implies that the profit function $\Pi(I, K, L)$ is also homogeneous of degree one. Under this assumption it is straightforward to show that the value function is also homogeneous of degree one in K (see Stokey, Lucas, and Prescott (1989)). From Euler's theorem it follows that the marginal value of capital, $\partial V/\partial K$, equals the average value of capital, V/K. Assuming efficient capital markets allows market prices to be used to construct the average value of capital. Finally, assuming the functional form for $G(\cdot,\cdot;\cdot)$ gives equation (2.1).

To derive the Euler equation used in the text, substitute equation (2.11) into equation (2.10) to get,

$$(1-\delta)E_t\left(\frac{\partial\Pi}{\partial I}\right)_{t+1} = \left(\frac{\partial\Pi}{\partial I}\right)_t + \left(\frac{\partial\Pi}{\partial K}\right)_t . \tag{2.12}$$

Under the same linear homogeneity assumptions, the assumption of a perfectly competitive labour market and a monopolistic output market, and the assumed functional form for $G(\cdot,\cdot;\cdot)$ equation (2.2) in the text is derived.

2.B Data Appendix

The data used is from the published accounts of UK listed firms and was collected from Datastream. From the sampling frame of firms covered by Datastream, a primary sample of firms was identified with the following criterion: the firms chosen for the sample were required to have at least one positive recorded R&D expenditure between the years 1990–1993. Very few firms reported R&D expenditures prior to the introduction of the new accounting standard, SSAP 13, in 1989. The initial sample consisted of approximately 340 firms, the vast majority of which had fewer than four positive observations on R&D .

Two further selection criteria were applied to this primary sample. First, all firms with fewer than seven consecutive positive observations on all of the main variables were deleted. Data on firms which have been acquired, or which have gone out of business, are not readily available from Datastream. Consequently, all firms have seven or more observations leading up to 1993 (the final observation for each firm is 1993). Second, firms with large one period changes in their capital stock were also deleted. This was done to exclude firms with major mergers or acquisitions. Firms whose change in their book value of lay outside three times the interquartile range above and below the median were removed. The final sample is an unbalanced panel consisting of 144 firms with observations ranging between seven and twenty-two. Approximately half of the sample had observations available for the full sample period (1972–1993).

In splitting the sample, Datastream's finest industrial classification, level 6,was used.

The definitions of the data used in the construction of the variables is as follows (Datastream codes are in square brackets).

Tobin's Q (Q): The market value of ordinary shares [HMV] plus the book value of total loan capital [321] less deferred tax [301] divided by the book value of capital [339].

Investment (I) : Total new fixed assets [435].

Cash Flow (CF) : Provision for depreciation of fixed assets [136], plus operating profit before tax, interest and preference dividends [137].

Free Cash Flow (FCF) : Cash Flow less total interest charges [153] and total tax charge [172].

Output (Y) : Total sales [104].

R&D : Research and Development (expensed) [119].

Capital Stock (K) : Book value of net total fixed assets [339].

Market value of ordinary shares (V) : Historic market value of the firm [HMV].

Chapter 3

An Explanation of the Excess Variance of Production Puzzle

3.1 Introduction

A widely reported fact is that the variance of production exceeds the variance of sales.[1] This contradicts the standard linear-quadratic model of inventory investment (e.g. Holt, Modigliani, Muth, and Simon (1960)), which predicts that in order to minimize costs firms will smooth production over time using inventories as a buffer against demand shocks. For this reason the fact is often referred to as the 'excess variance of production puzzle'.

In this chapter,[2] we build a model of inventory investment and impose constraints on the firm's access to external sources of finance. It is found that the presence of these financing constraints can explain the excess variance of production puzzle in a model which otherwise would not deliver this result. In addition, the model with financing constraints predicts that inventory investment and sales covary positively. Moreover,

---

[1]This fact has been found using data of all levels of aggregation. For studies which report this finding in US data see Blinder and Maccini (1991), Blanchard (1983), and Blinder (1981). For UK data see Guariglia and Schiantarelli (1995). For evidence to the contrary see Fair (1989) and Krane and Braun (1991).

[2]This chapter is joint work with Urs Haegler.

even though the stochastic process governing the demand for the firm's output is specified to be independently and identically distributed (i.i.d.), the endogenous sales process exhibits positive serial correlation.[3] Both of these predictions are observed in data.

Evidence that financial factors influence inventory investment has been documented by a number of recent empirical studies. This paper provides a theoretical link between the evidence presented in these studies, and the fact that production varies more than sales. Gertler and Gilchrist (1994) find evidence that small manufacturing firms draw down their inventory stocks heavily following a monetary contraction, whereas large firms appear to borrow in order to smooth the impact of a downturn on their inventory behaviour. Using US panel data, Carpenter, Fazzari, and Petersen (1994) find that the inventory investment of small firms is more sensitive to cash flow than is the inventory investment of large firms. Kashyap, Stein, and Wilcox (1993) include the ratio of bank loans to commercial paper in several structural models of inventory investment, and find that it has a significant effect. In an examination of the 1982 US recession, Kashyap and Stein (1994) find that the ratio of liquid to total assets has significant explanatory power for the inventory investment of firms without bond ratings, but not for those firms with bond ratings. Taking different structural models to UK panel data, Guarigla (1996) and Guariglia and Schiantarelli (1995) find that financial factors have an important effect on the inventory behaviour of only those firms which may be in financial distress as indicated by a low coverage ratio (the ratio of cash flow to total interest expense). The latter study presents evidence that firms with high coverage ratios are more likely to smooth production.

From a macroeconomic perspective, inventories are a crucial component of fluctuations in aggregate output. For example, Blanchard and Fischer (1989) report that while the stock of inventories makes up only 1% of US GNP, declines in the stock account for 50% of the drop in output in recessions.[4] Clearly any model of fluc-

[3]In models where there is the possibility that the firm may stock out in a given period sales and demand are generally not identical. Such models are referred to as stockout-avoidance models (e.g. Abel (1985), Kahn (1987)).

[4]Similarly, for the UK Sensier (1996) reports figures of 3% and 30%, respectively.

tuations in aggregate output must be able to explain the behaviour of firm level inventory investment.

One of the shortcomings shared by many theoretical models of aggregate fluctuations is the weakness of their propagation mechanisms.[5] To strengthen this mechanism many recent papers have considered the effects of capital market imperfections.[6] In light of this development in the theoretical literature, it seems natural to ask the following question. Would imposing financing constraints on a partial equilibrium model of firm inventory investment explain the excess variance of production puzzle? This chapter provides an answer to that question.

Existing attempts to explain the excess variance of production can be put into three classes: (i) the linear-quadratic model modified with the introduction of either non-convex costs or cost shocks (e.g. Blinder (1986), Eichenbaum (1989), Ramey (1991), Hall (1996), Bresnahan and Ramey (1994)), (ii) the stockout-avoidance model with a demand process that exhibits serial autocorrelation (e.g. Kahn (1987)), and (iii) models of the (s,S) type (e.g. Blinder (1981), Caplin (1985)). Although much of the empirical research has been directed towards the first class of models, satisfactory evidence in support of these models is scarce. For example, with the exception of Ramey (1991), most authors have estimated marginal costs to be upward sloping (e.g. Blanchard (1983), Eichenbaum (1989), West (1986)) suggesting that increasing returns are not a source of nonconvex costs.[7] Similarly, little evidence has been found that shocks to observable costs have a significant effect on inventory investment (e.g. Blinder and Maccini (1991), Miron and Zeldes (1988)).[8]

[5]This is particularly true of Real-Business-Cycle models. See Cogley and Nason (1995).

[6]See, for example, Kiyotaki and Moore (1995), Greenwald and Stiglitz (1993), Gertler (1992), Calomiris and Hubbard (1990), Bernanke and Gertler (1989).

[7]Moreover, Bils and Kahn (1996) show that if marginal costs are decreasing, firm behaviour which minimizes quadratic costs produces the counterfactual prediction that the ratio of inventories to sales is procyclical. More promising models which incorporate non-convex costs while retaining increasing marginal costs have only been tested with data from the automobile industry (e.g. Hall (1996), Bresnahan and Ramey (1994)).

[8]However, Eichenbaum (1989) found no evidence against the version of the linear-quadratic model in which *unobservable* cost shocks are incorporated.

In contrast, much less testing has been performed on models in the other two classes.[9] However, the model developed here delivers testable predictions which distinguish it from all three categories above. In the second class of model, for instance, sales which do not contain information about expected future sales do not affect production. This is not true in our model, since such sales change the amount of internal funds available to finance production. Similarly, the (s,S) model predicts that the covariance between inventory investment and sales is zero, whereas the model here predicts that this covariance should be positive.

The model is presented in the next section of the chapter. To illustrate the intuition behind our results, a small example and its solution is described in the third section. Then, in section 3.4, we discuss properties of the value and the policy function in the general model. Section 3.5 extends the results of the example to the general model, and discusses further predictions of the model. The final section concludes.

3.2 The Model

Consider a firm which produces an (imperfectly) storable good at a constant unit cost (which we normalise to 1 without loss of generality). It attempts to sell its output each period at price p_t. We assume that $p_t = p > 1$ $\forall t$. The firm enters period $t \geq 1$ with a stock of goods G_t (those goods not sold in the previous period) and a stock of a liquid asset, M_t, which for convenience we will call money. The timeline of events in each period is shown in Figure 3.1.

At the beginning of each period, the stock of goods depreciates according to a depreciation technology represented by a nondecreasing and convex function δ : $\Re_+ \to \Re_+$, $G \mapsto \delta(G)$. We impose $\delta(0) = 0$ and $\delta(G) < G$, i.e. the firm never loses all unsold goods through depreciation. Thus, after depreciation in period t the firm is left with $G_t - \delta(G_t) > 0$ goods.[10]

---

[9]Kahn (1992) tests the stockout-avoidance model, and Mosser (1990) tests the (s,S) model.

[10]An equivalent assumption would be to require the firm to pay a pecuniary storage cost $\delta(G)$, financed by its money holdings. One then has to allow the firm to be able to 'reverse-engineer'

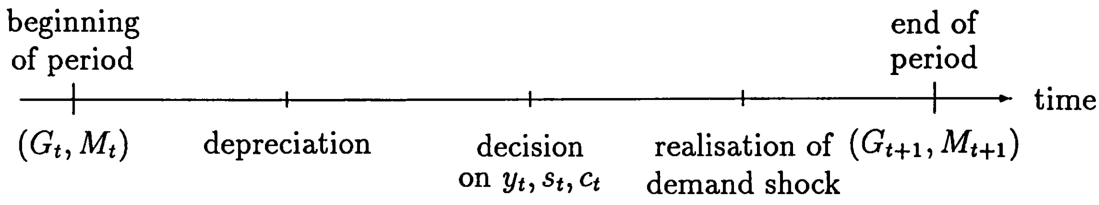

Figure 3.1: Timing of Events in Period t

The firm then makes its *gross* production decision, y_t, its savings decision, s_t, and its consumption decision, c_t, subject to the following financing constraints,

$$c_t \geq 0 \tag{3.1}$$

$$s_t \geq 0 \tag{3.2}$$

$$y_t \geq 0 \tag{3.3}$$

$$c_t + s_t + y_t \leq M_t \tag{3.4}$$

The non-negativity constraint on consumption can be interpreted as preventing the firm from raising equity capital from shareholders by issuing negative dividends.[11] The non-negativity constraint on savings is simply a borrowing constraint. The non-negativity constraint on gross production implies that the firm cannot 'reverse-engineer' and thereby consume (or save) out of its beginning-of-period stock of goods. The final inequality is the budget constraint.

Once production has taken place the demand realisation, z_{t+1}, occurs. As usual in stockout-avoidance models we impose the following non-negativity constraint on the stock of goods (i.e. the firm is not allowed to short sell its output). Sales, x_t, are given by

$$x_{t+1} = \min[z_{t+1}, n_t], \tag{3.5}$$

finished goods (generating one unit of money per unit) in order to pay the storage cost in those periods when the stock of money is insufficient to cover them.

[11] Alternatively, it could be interpreted as restricting the firm's access to trade credit.

where n_t is the total amount of goods the firm makes available for sale,

$$n_t = G_t - \delta(G_t) + y_t. \tag{3.6}$$

The demand realisation z_{t+1} is identically and independently distributed in each period t on a subset $\mathcal{Z} \subseteq \mathbb{R}_+$, according to the probability density function $\phi : \mathcal{Z} \to \mathbb{R}$. The associated cumulative distribution function is denoted by $\Phi : \mathcal{Z} \to [0,1]$. For convenience we define $\underline{z} = \inf \mathcal{Z} \geq 0$, and $\bar{z} = \sup \mathcal{Z}$.

The stock of money next period is the sum of savings in period t, s_t, and the revenue generated from sales. Thus, the law of motion for the money stock is[12]

$$M_{t+1} = s_t + px_{t+1}, \tag{3.7}$$

The law of motion for the stock of goods is

$$G_{t+1} = G_t - \delta(G_t) + y_t - x_{t+1}, \tag{3.8}$$

which can be rewritten as $G_{t+1} = n_t - x_{t+1}$.

The objective of the firm is to choose gross production and savings to maximise the present discounted value of consumption, i.e.

$$\max E_0 \sum_{t=0}^{\infty} \beta^t c_t \tag{3.9}$$

subject to (3.1), (3.2), (3.3), (3.4), (3.5), (3.7), (3.8), with the discount factor $\beta \in (0,1)$.

Since production is not directly observable in most firm data, it is defined empirically to be the change in inventories stocks plus sales.[13] Using the law of motion for the

---

[12]For simplicity we have implicitly assumed the gross interest rate on money to be equal to 1. The results are not affected by a gross interest rate R different from unity, as long as $\beta R < 1$.

[13]See, for example, the discussion in Blinder and Maccini (1991), p.77.

stock of goods, (3.8), observable production, q_t, is defined to be

$$q_t = G_{t+1} - G_t + x_{t+1} = y_t - \delta(G_t) \tag{3.10}$$

Thus, observable production is gross production, y_t, net of depreciation. Our primary interest is to compare the distribution of observable production q_t with that of sales x_t.

3.3 An Example

The main features of the general model can be illustrated with a simple example in which all variables are restricted to be integers. In any given period the number of goods demanded is either 0, 1, or 2, with probabilities ϕ_0, ϕ_1, and ϕ_2, respectively, such that $\phi_0 + \phi_1 + \phi_2 = 1$. The firm can store only one unit of inventory between periods. If the firm has unsold inventories in excess of this storage capacity, they depreciate completely (i.e. the depreciation function is parameterised as $\delta(G) = \max\{0, G-1\}$).

We first postulate the following policy, and then demonstrate its optimality for a set of given parameter values. The firm provides two units of goods for sale whenever this is feasible, and one otherwise. If after setting n to 2 there is money left over, first unit is saved, i.e.. $s = 1$. If there are still remaining funds they are spent on consumption.

Table 3.1 summarizes this information. In a stationary equilibriumthe firm can assume eight possible states, (G, M), which are listed in the first column. The second column is the (unique) sales realization, x, that has brought the firm into that state. Columns four to six list the policies we have postulated for each state. For each state an action is a triplet (n, s, c). For example, in state $(1, 3)$ the firm puts up two goods for sale, saves one unit of money, and uses one unit for consumption (the action is (2,1,1)). Column three lists the net production decisions, q, implied by these actions. Column seven shows the flow utility enjoyed by the firm in that

Table 3.1: The Postulated Policy and its Implications

(G, M)	x	q	n	s	c	u	x'	(G', M')
							0	(1,0)
(1,0)	0	0	1	0	0	0	1	(0,2)
							1	(0,2)
							0	(1,0)
(2,0)	0	-1	1	0	0	0	1	(0,2)
							1	(0,2)
							0	(2,0)
(2,1)	0	0	2	0	0	0	1	(1,2)
							2	(0,4)
							0	(2,0)
(0,2)	1	2	2	0	0	0	1	(1,2)
							2	(0,4)
							0	(2,1)
(1,2)	1	1	2	1	0	0	1	(1,3)
							2	(0,5)
							0	(2,1)
(1,3)	1	1	2	1	1	1	1	(1,3)
							2	(0,5)
							0	(2,1)
(0,4)	2	2	2	1	1	1	1	(1,3)
							2	(0,5)
							0	(2,1)
(0,5)	2	2	2	1	2	2	1	(1,3)
							2	(0,5)

state. Finally, the last two columns show the possible sales realizations, x', and next period's state, (G', M'), associated with each of the three realizations.

Our goal is to solve for the distributions of sales, x, and net production, q. To do so, we have to derive conditions under which the policy we have postulated is optimal. Then, we choose a parameter set $\Gamma = \{p, \beta, \phi_0, \phi_1, \phi_2\}$ for which these conditions are satisfied. Finally, given this parameter vector one can solve for the stationary joint distribution of the state variables, and derive distributions for x and q from there.

3.3.1 Optimality Conditions

To ensure that the financing constraint (3.1) will be binding, we first need to ensure that it is optimal to provide $n = 2$ in the first best case when there is no financing constraint. For this to be true, the chosen vector of parameter values must satisfy the following conditions:

$$2\phi_2 + \phi_1 - \phi_0 \geq \beta^{-1}, \tag{3.11}$$

$$\phi_2 - (\phi_1 + \phi_0) \leq \beta^{-1}. \tag{3.12}$$

The first inequality ensures that it is optimal to provide two units for sale at the margin. The second ensures that it is never optimal to provide three units.

Summing across states, there are 40 feasible actions available to the firm.[14] We want to derive restrictions on the parameter values which are sufficient for the optimality of the actions we have chosen for the firm. For example, in state $(1, 2)$, under what conditions will the proposed action, $(2, 1, 0)$, be preferred by the firm to the alternative action, $(2, 0, 1)$? Since the proposed policy involves five different actions, conditions such as this one must be derived for the remaining 35 feasible actions.

To express the firm's preferences over actions in terms of parameter values, a nec-

[14]Due to the non-negativity restriction on gross production, y, the action (1,0,0) is the only feasible action for states (1,0) and (2,0).

states	action preference	condition
(0,4),(1,3)	$(2,1,1) \succ (2,2,0)$	$1-\beta(1-\phi_0) > \beta\phi_0(v_3 - v_2)$
(1,2)	$(2,1,0) \succ (1,2,0)$	$1-\beta(1-\phi_0) < \beta\phi_0(v_2 - v_1)$
(0,2),(2,1)	$(2,0,0) \succ (1,1,0)$	$\phi_2 > \phi_0(v_2 - v_1)$
(0,2),(2,1)	$(2,0,0) \succ (1,0,1)$	$1 < \beta(1-\phi_0)(v_3 - v_2)$

Table 3.2: Independent Optimality Conditions

essary first step is to express the difference in the value of two states in terms of the parameters. Several states have equivalent values. For example, compare $v(1,0)$ with $v(2,0)$ in Table 3.1. In both states flow utility $u = 0$, and the expected discounted value of next period is $\beta(\phi_1 + \phi_2)v(0,2)$. Thus $v(1,0) = v(2,0)$, and we call the set $\{(1,0),(2,0)\}$ an equivalence class of states. For simplicity we refer to the value of this equivalence class as v_1. Using the same reasoning we get another four equivalence classes

$$\{(0,2),(2,1)\}, \quad \{(1,2)\}, \quad \{(1,3),(0,4)\}, \quad \{(0,5)\}.$$

We refer to the values of these equivalence classes as v_2, v_3, v_4, and v_5, respectively. Note that $v_5 - v_4 = v_4 - v_3 = 1$. In other words, the value function is linear over those states where consumption is positive.

In appendix 3.A, it is shown that the comparisons of different feasible actions imply only four independent conditions which must be satisfied by the parameter values in order for our proposed policy to be optimal. The four independent optimality conditions are summarised in Table 3.2.

Table 3.2 should be read as follows. For example, in states (0,4) and (1,3) the action (2,1,1) is preferred to the action (2,2,0) if the condition $[1-\beta(1-\phi_0)] > \beta\phi_0(v_3 - v_2)$ holds (i.e. consuming the marginal unit of money yields greater value than saving it). This condition is derived by simply comparing the values of each state under the alternative actions.

Thus, a chosen set of values for the parameters must satisfy the four conditions in Table 3.2 as well as conditions (3.11) and (3.12) in order for our proposed policy to be optimal. We assume that the set of parameters Γ takes on the values $\{p, \beta, \phi_0, \phi_1, \phi_2\} = \{2, 0.96, 0.15, 0.25, 0.6\}$. These values satisfy the above conditions.

3.3.2 Stationary Distributions

Define f_s, $s \in S := \{(1,0),(2,0),(2,1),(0,2),(1,2),(1,3),(0,4),(0,5)\}$, $\sum_{s \in S} f_s = 1$, as the probability with which the firm finds itself in state s at any point in time.

To obtain stationarity one has to ensure that, given the optimal policy, in each period the probability of leaving any state matches the probability of coming into it. In other words, the following system of equations has to be satisfied.

$$\begin{aligned}
\phi_0 f_{20} &= (1-\phi_0) f_{10}, \\
\phi_0(f_{02} + f_{21} &= f_{20}, \\
\phi_0(f_{12} + f_{13} + f_{04} + f_{05}) &= f_{21}, \\
(1-\phi_0)(f_{10} + f_{20}) &= f_{02}, \\
\phi_1(f_{21} + f_{02} &= f_{12}, \\
\phi_1(f_{12} + f_{04} + f_{05} &= (\phi_0 + \phi_2) f_{13}, \\
\phi_2(f_{21} + f_{02}) &= f_{04}, \\
\phi_2(f_{12} + f_{13} + f_{04}) &= (1-\phi_2) f_{05}.
\end{aligned}$$

Since these equations are not linearly independent, the system has to be 'pinned down' by the adding up constraint

$$\sum_{s \in S} f_s = 1.$$

f_{10}	f_{20}	f_{21}	f_{02}	f_{12}	f_{13}	f_{04}	f_{05}
0.0039	0.0219	0.1242	0.0219	0.0365	0.2070	0.0877	0.4969

Table 3.3: Stationary Distribution of the States

$y=-1$	$y=0$	$y=1$	$y=2$
0.0219	0.1281	0.2436	0.6064

$x=0$	$x=1$	$x=2$
0.1500	0.2655	0.5845

Table 3.4: Distributions of Net Production and Sales

The solution to the system is given by

$$
\begin{aligned}
f_{20} &= \left[\frac{2-\phi_0}{1-\phi_0} + \frac{1-\phi_0}{\phi_0}\left(1+\frac{1}{\phi_0}\right)\right]^{-1}, \\
f_{10} &= \frac{\phi_0}{1-\phi_0} f_{20}, \\
f_{21} &= \frac{1-\phi_0}{\phi_0} f_{20}, \\
f_{02} &= f_{20}, \\
f_{12} &= \frac{\phi_1}{\phi_0} f_{20}, \\
f_{13} &= \frac{\phi_1(1-\phi_0)}{\phi_0^2}, \\
f_{04} &= \frac{\phi_1}{\phi_0} f_{20}, \\
f_{05} &= \frac{\phi_2(1-\phi_0)}{\phi_0^2} f_{20}.
\end{aligned}
$$

The values chosen for the set of parameters Γ gives us the stationary joint distribution for the state variables shown in Table 3.3. From this distribution, the distributions for (net) production and sales can be derived. There are shown in Table 3.4.

The mean of production and sales is $\bar{q} = \bar{x} = 1.4345$, and the variances are $var(q) = 0.6334 > 0.5447 = var(x)$. Thus, in this example the firm exhibits production counter-smoothing.

3.3.3 Financing Constraints and the Excess Variance of Production

It is the presence of the financing constraints in this model which delivers the excess variance of production result. If capital markets were perfect, then the firm's optimal policy for production would be such that $var(q) = var(x)$. Each period the firm would simply replace what had be sold and what had depreciated. This implies that net production, q, is set equal to sales, x, in each period, which, in turn, implies that the number of goods put up for sale each period is constant.

When the firm is financially constrained such a policy is not possible. After particularly low sales realizations depreciation is so high that the firm does not have sufficient funds available to replace what had been sold and what had depreciated. In these circumstances net production, q, will be less than sales, x, and the amount of goods put up for sale next period drops below the unconstrained amount. We say that in this case the firm 'underproduces'. Now suppose that after a low sale the firm experiences a relatively high sale. In this case, it will have more than a sufficient amount of cash to replace all that had been sold and depreciated. It will now be in a position to rebuild its inventory back up to its unconstrained level, or at least close to it. Under these circumstances, the firm will 'overproduce'. What drives the variance of the firm's production above that of sales is the association of underproduction with low sales realizations, and overproduction with medium sales realizations.

Table 3.1 reveals that there is a discrepancy between x and q only for states $(2,0)$ and $(0,2)$. In $(2,0)$, where the sales realization $(x = 0)$ is associated with production $(y = -1)$, underproduction of one unit occurs. In this state the firm would like to make two units available for sale, but is prevented from doing so by the financing constraint. This, in turn, causes the firm to overproduce by one unit in state $(0,2)$.

Figure 3.2 helps to illustrate the intuition behind this result. In effect, the financing constraints cause a low sales realization to be mapped into an even lower production realization, and a medium sales realization to be mapped into a high production realization. If we compare this to the distribution of sales and net production when the nonnegativity constraint on consumption is absent, as depicted in Figure 3.3, we see this effect clearly.

3.3.4 The Nature of Financing Constraints

To what extent do the results in this paper depend on the type of financing constraints used? Here we offer only informal arguments that our results will not alter under different assumptions about the financing constraints. The basis for this is that regardless of their nature, financing constraints will be binding primarily after low sales realizations. Thus, underproduction will still be associated only with low sales realizations, which is the central feature of the excess variance of production result.

More specifically, consider two different ways to model the financing constraints. For one possibility, we may assume that there is a perfectly elastic supply of external finance, but that it is more costly than internal finance. For the other alternative, we may think of the firm being able to enter information-constrained insurance contracts.

The first case, where the firm faces a hierarchy of finance, is closer to our model. In fact, the financing constraints in our model could be interpreted as representing a finance hierarchy in which the premium on external funds is so high that it is never optimal for the firm to use external finance. Suppose instead that the premium were low enough to make the use of external finance attractive in some situations. In this case the firm can find itself in three qualitatively different regions. In the first region, internal funds will be so low that the benefit of the marginal good put up for sale is high enough to warrant the use of external finance. In the second region the firm is still constrained, but the value of the marginal good put up for sale is

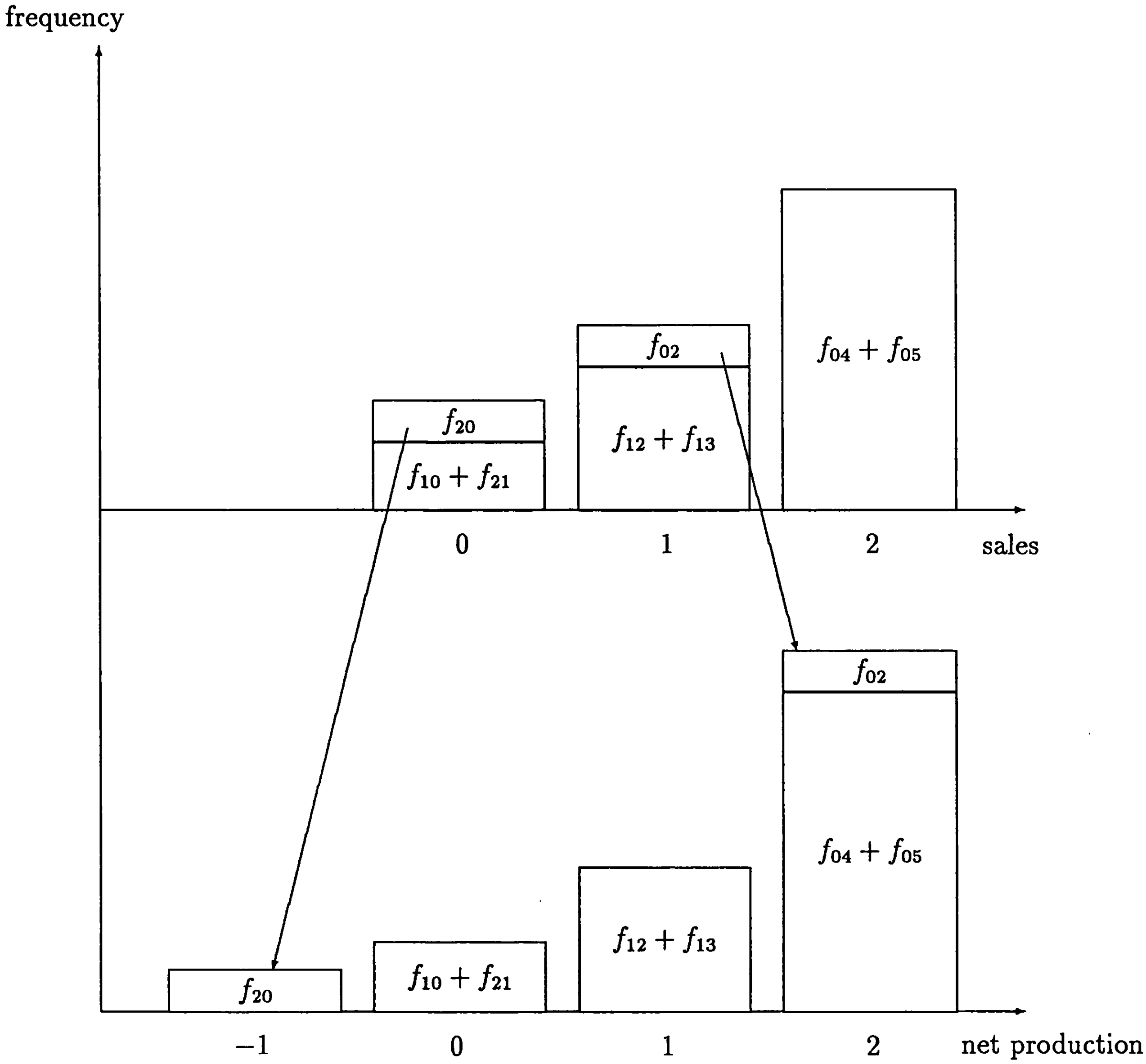

Figure 3.2: Distributions with Financing Constraint

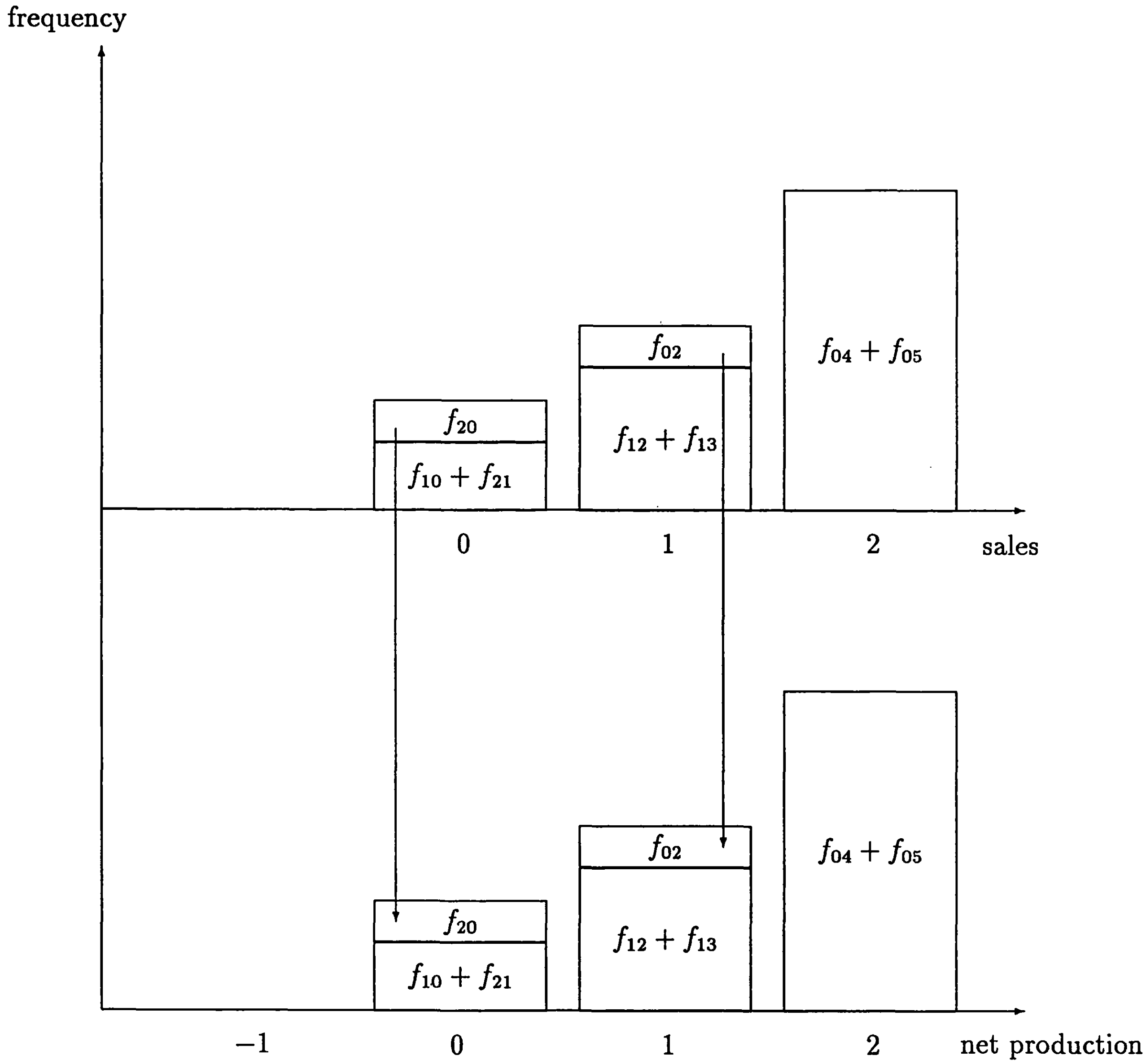

Figure 3.3: Distributions without Financing Constraint

not high enough to justify the use of external funds. Here the firm behaves exactly as it would in our model. In the third region the firm is unconstrained. Note that, because external finance is more costly, the optimal amount of goods put up for sale will be lower when the marginal unit is financed externally than when the marginal unit is financed with internal funds. Thus, if the firm begins with the unconstrained amount of goods for sale and has a low sales realization it will underproduce. In other words, underproduction will still be associated with low sales realizations.

For the second case, which is somewhat further from our model, our argument is only suggestive. Suppose that lenders can observe inventories at only two points in time: after sales *and* depreciation, and just before sales. In other words, lenders can observe the firm's gross production decision, y_t. However, lenders cannot observe sales and depreciation.[15] In this setup there is an incentive for the firm to report a bad sales realization when in fact there had been a good sales realization. Typically, the optimal contract in such a setup would place restrictions on the observable decision variable in order to ensure that the borrower truthfully reports good outcomes. This usually implies that in bad outcomes the level of the decision variable is lower than it would be if all variables were observable, in order to introduce a cost to reporting a good outcome as a bad outcome. We know that if all variables are observable, then the optimal policy of the firm is to set gross production such that it equals the sale plus depreciation. In the information-constrained framework, an optimal constract would force the firm to set gross production lower than this for low sales realizations. Thus, this informal argument suggests that the association of low sales with underproduction would remain, and that the excess-variance-of-production result is preserved.

[15]This argument is consistent with depreciation taking the form of a fixed storage capacity as in the example.

3.4 The General Model

We can specify either gross production, y_t, or, equivalently, the amount of goods made available for sale, n_t, as a choice variable since $n_t = G_t - \delta(G_t) + y_t$. Therefore, taking n_t and s_t as the choice variables, the Bellman equation for the problem outlined in section 3.2 is

$$\begin{aligned} v(G_t, M_t) &= \max_{n_t, s_t} \Big\{ M_t - s_t - (n_t - G_t + \delta(G_t)) + \beta \int_{\underline{z}}^{n_t} v(G_{t+1}, M_{t+1})\, \phi(dz_{t+1}) \\ &\quad + \beta \int_{n_t}^{\bar{z}} v(0, M_{t+1})\, \phi(dz_{t+1}) \Big\} \end{aligned} \tag{3.13}$$

subject to (3.1), (3.2), (3.3),(3.7), and (3.8).

3.4.1 The Unconstrained Problem

Consider removing the non-negativity constraint on consumption (3.1) from the problem above.[16] In this case, the following policy would be optimal. Optimal savings would be to set $s^*(G, M) = 0$ in each period since $\beta R < 1$. The optimal choice of n_t will always be an interior solution, so we can differentiate the Bellman equation with respect to n_t. Assuming that δ is differentiable (denoting with δ' its first derivative) and using the envelope conditions, $v_{M_t} = 1$ and $v_{G_t} = (1 - \delta'(G_t))$, we obtain

$$\int_{\underline{z}}^{n_t} (1 - \delta'(G_{t+1}))\, \phi(dz_{t+1}) + \int_{n_t}^{\infty} p\, \phi(dz_{t+1}) = \beta^{-1} \tag{3.14}$$

as the Euler equation. Since equation (3.14) involves neither M_t nor G_t the optimal policy for the firm is to set $n^*(G, M) = N$, a constant. This implies that in each

[16]We refer to this problem as the unconstrained problem, and to the non-negativity constraint on consumption as the financing constraint. Relaxing the latter allows the firm to perfectly insure its desired production expenditure against negative demand shocks through negative consumption in those periods where it does not have sufficient cash on hand. It is precisely the availability of this insurance we wish to remove with the constraint on consumption. If, instead, we had relaxed the non-negativity constraint on savings (3.2), the firm's optimal policy would have been simply to borrow as much as possible in the first period, due to the linear utility function and $\beta R < 1$.

period net production, q_t, is set equal to sales, x_t. Therefore, if this financing constraint, is relaxed $var(q) = var(x)$.

3.4.2 The Constrained Problem

When the non-negativity constraint on consumption is imposed we can no longer guarantee the differentiability of the value function since there is no way to guarantee that the optimal policies will be interior solutions. Thus we cannot simply analyse a version of equation (3.14). However, the following properties of the solution can be established, and they are sufficient to prove that the variance of production exceeds the variance of sales when all of the financing constraints are imposed.

Proposition 3.4.1 *The Bellman operator defined in (3.13) is a contraction mapping with a unique fixed point. The fixed point, v, satisfies the following properties:*

1. *v is increasing in M and G.*

2. *v is concave.*

3. *v is strictly concave in the region of constrained states.*

The proof of the proposition and the first two properties are fairly strightforward, and therefore are given in appendix 3.B. To show the strict concavity of v in the constrained region, we adopt the method of a concavity proof in Abel (1985).[17] Imagine there are three firms, firm A which puts up n_A for sale, firm B which put up n_B, and firm C, which puts up $\alpha n_A + (1-\alpha)n_B$. The third firm always gets α of firm A's demand z_t^A, and $(1-\alpha)$ of firm B's demand z_t^B. If neither firm A nor firm B stocks out, then the value of firm C will equal $\alpha v(n_A) + (1-\alpha)v(n_B)$. The same will be true if both A and B stock out. However, suppose firm A stocks out but firm B does not. In this circumstance, firm C will be able to meet the

[17] In that paper production occurs with a lag. Consequently, only the stock of inventories which the firm has at the beginning of the period are available for sale. Moreover, the stock of inventories is the only state variable.

excess demand of firm A, $\alpha(z_t^A - n_A)$, by selling it's unsold inventories from firm B's demand, $(1-\alpha)(n_B - z_t^B)$. Therefore, $v(\alpha n_A + (1-\alpha)n_B) > \alpha v(n_A) + (1-\alpha)v(n_B)$.

To adapt this proof to our model we have to incorporate optimal policies into the argument. Consider two states in which the firm is constrained, (G_0, M_0) and (G_1, M_1) where $G_1 - \delta(G_1) + M_1 > G_0 - \delta(G_0) + M_0$. A set of optimal policies is associated with each state. Choose one policy from each set. Let (n_0, s_0) be the chosen policy from the set associated with (G_0, M_0), and (n_1, s_1) be the chosen policy for (G_1, M_1). Since the firm is constrained in both states, it cannot be the case that the chosen optimal policies are identical. Consider a third state, $(G_\theta, M_\theta) = ((1-\theta)G_0 + \theta G_1, (1-\theta)M_0 + \theta M_1)$. Assume that the policy the firm follows in this state is $(n_\theta, s_\theta) = ((1-\theta)n_0 + \theta n_1, (1-\theta)s_0 + \theta s_1)$. Call v_θ the value attained by this policy in the state (G_θ, M_θ).

Suppose that in state (G_θ, M_θ) the firm receives $(1-\theta)$ of the demand shock z_0 for state (G_0, M_0), and θ of the demand shock z_1 for state (G_1, M_1). If the firm stocks out in neither state (G_0, M_0) nor (G_1, M_1), or if it stocks out in both of these states, then

$$\begin{aligned} G_\theta = (1-\theta)(n_0 - x_0) + \theta(n_1 - x_1) &= (1-\theta)G_0 + \theta G_1 \\ M_\theta = (1-\theta)(s_0 + px_0) + \theta(s_1 + px_1) &= (1-\theta)M_0 + \theta M_1 , \end{aligned}$$

which implies that in these circumstances

$$v_\theta(G_\theta, M_\theta) = (1-\theta)v(G_0, M_) + \theta v(G_1, M_1) .$$

As in Abel's argument, however, if the firm stocks out in only one state then

$$\begin{aligned} G_\theta &> (1-\theta)G_0 + \theta G_1 \\ M_\theta &> (1-\theta)M_0 + \theta M_1 , \end{aligned}$$

because the firm can use unsold goods from one state to meet the excess demand of the other state. For these outcomes,

$$v_\theta(G_\theta, M_\theta) > (1-\theta)v(G_0, M_0) + \theta v(G_1, M_1).$$

Since the policy (n_θ, s_θ) is not necessarily optimal it is true that $v(G_\theta, M_\theta) \geq v_\theta(G_\theta, M_\theta)$. Therefore, the value function is strictly concave in the region of the state space where the firm is constrained. □

In the example of section (3.3) we were able to compute a stationary distribution of states. Showing that an ergodic distribution exists in our general setting is a rather formidable task. We therefore simply assume existence of stationarity in the general model.

The following propositions and corollaries characterise the optimal policy functions, and are used to derive a number of statements about the distribution of sales and net production in the next section.

Proposition 3.4.2 *There exist upper bounds to the optimal policies, n and s, referred to as $\bar{n}, \bar{s}$. If $M - \bar{s} > \bar{n} - G + \delta(G)$ the remaining money is used for into consumption.*

Proof: Suppose the firm never consumes. Since $v(G, M)$ is then strictly concave in both its arguments, the marginal return to savings and to production will be decreasing in both state variables. The value of an infinitesimal unit of money would tend to β as $M \to \infty$, since the probability of encountering a sequence of sales realisations in which the non-negativity constraint on consumption will be binding tends to zero (effectively, the firm becomes unconstrained). Since $\beta < 1$, there will be a level of savings, $\bar{s}$, beyond which the firm will prefer consumption to further savings. Similarly, as $G \to \infty$ the marginal value of inventories tends to a value below one, since the probability of selling the marginal unit goes to zero. Thus, there also exists a level of goods the firm puts up for sale, $\bar{n}$, beyond which the firm prefers consumption to further production.[18] □

Corollary 3.4.3 *There exist optimal policy functions, $s^*(G,M)$ and $n^*(G,M)$, which map each element of the state space into the space of feasible actions.*

Proof: The one-period return function is concave in the state variables, and the feasibility set for the choice variables is convex. Over the region where consumption is zero v is strictly concave. It follows that the maximum in (3.13) in this region is attained by unique choices of s_t and n_t. From Proposition 3.4.2, it follows that when consumption is positive, $s_t^* = \bar{s}$ and $n_t^* = \bar{n}$. □

Corollary 3.4.4 *The mean of production, $\hat{q}$, equals the mean of sales, $\hat{x}$.*

Proof: Assume $\hat{q} > \hat{x}$ is true. Then the firm would accumulate inventories indefinitely, contradicting the existence of $\bar{n}$. Conversely, if $\hat{q} < \hat{x}$ were true, then $G \to 0$ as $t \to \infty$, which cannot be optimal since $v(G,M)$ is increasing in G. From an intuitive viewpoint, this corollary holds because each unit of output that the firm produces is either sold, or depreciates which is accounted for as negative production.
□

For the purposes of the next Proposition, define the total funds available to the firm in period t to be $G_t - \delta(G_t) + M_t$.

Proposition 3.4.5 *The optimal policy functions $n^*(G,M)$ and $s^*(G,M)$ are non-decreasing in the total funds available to the firm.*

Proof: We will first present the proof for $n^*(G,M)$. The statement holds when the non-negativity constraint on consumption is not binding, since in those states additional money holdings are simply used for consumption, without changing the amount $\bar{n}$ made available for sale. For states in which the non-negativity constraint

[18]The existence of $\bar{s}$ and $\bar{n}$ implies the existence of endogenous upper bounds to the state space, $\bar{M}$ and $\bar{G}$.

on consumption is binding, we prove the statement by contradiction. Note that in those states the return function is equal to zero, and we can therefore write

$$\begin{aligned} v(G, M) &= \beta \int_{\underline{z}}^{n^*(G,M)} v(n^*(G, M) - z, s^*(G, M) + pz)\phi(z)dz \\ &\quad + \beta[1 - \Phi(n^*(G, M))]v(0, s^*(G, M) + pn^*(G, M)). \end{aligned} \tag{3.15}$$

Assume that in some (G, M) the optimal policy is to choose values (n, s). Also suppose that, given some small $\varepsilon > 0$, the optimal policy in $(G, M + \varepsilon)$ is $(n - \mu, s + \varepsilon + \mu)$, for some small $\mu > 0$, which implies

$$\int_{\underline{z}}^{n-\mu} v(n - \mu - z, s + \varepsilon + \mu + pz)\phi(z)dz + [1 - \Phi(n - \mu)]v(0, s + \varepsilon - (p - 1)\mu + pn) >$$

$$\int_{\underline{z}}^{n} v(n - z, s + \varepsilon + pz)\phi(z)dz + [1 - \Phi(n)]v(0, s + \varepsilon + pn). \tag{3.16}$$

But then the firm could do better by shifting funds μ from production to savings in (G, M) as well, i.e.

$$\int_{\underline{z}}^{n-\mu} v(n - \mu - z, s + \mu + pz)\phi(z)dz + [1 - \Phi(n - \mu)]v(0, s + (p - 1)\mu + pn) >$$

$$\int_{\underline{z}}^{n} v(n - z, s + pz)\phi(z)dz + [1 - \Phi(n)]v(0, s + pn). \tag{3.17}$$

Hence, (n, s) cannot be an optimal choice in state (G, M), which contradicts the initial assumption. Inequality (3.17) is proven in appendix 3.C.

As for the function $s^*(G, M)$, assume that there is a range of states over which the function is decreasing. The only way s^* could be decreasing in the constrained region is if there was an overproportionate increase of n in response to a small increase of funds available (be it in the form of more money or of higher goods inventories). Due to strict concavity of the value function, however, the negative effect on the expected value of v next period due to a decrease in s would more than offset the positive effect of an increase in n. We conclude that such a policy cannot be optimal.

□

Corollary 3.4.6 *Each state (G, M) has a unique sales realisation which moves the firm from other states into this state. In other words, there exists a function, call it $x^*(G, M) : (G, M) \mapsto x$ which maps each element of the state space into the set of feasible sales realisations.*

Proof: Suppose that the sale x' moves the firm to the state (G, M) from the post-production pair (n', s'), and that a different sale x'' moves the firm to the same state $(\tilde{G}, \tilde{M})$ from a different post-production pair (n'', s''). Then

$$\begin{aligned} G = n' - x' = n'' - x'' &\quad\Rightarrow\quad x'' - x' = n'' - n' \\ M = s' + px' = s'' + px'' &\quad\Rightarrow\quad x'' - x' = \frac{s' - s''}{p} \end{aligned}$$

However, from Proposition 3.4.5 we know that if $n'' > n'$ then $s'' \geq s'$. Thus the above equations cannot be satisfied simultaneously when $x' \neq x''$. □

3.5 Variance Results

We are now in a position to extend the results of the example concerning the relative variances of production and sales to the general model. In addition, we demonstrate that the sales process will be positively serially correlated.

Proposition 3.5.1 *The variance of production exceeds the variance of sales. The covariance of production and sales also exceeds the variance of sales.*

Define $Q_{\hat{x}} := \{(G, M) \mid x^*(G, M) = \hat{x}\}$. Define the expected difference between production and sales, given $x^*(G, M) = \hat{x}$, to be

$$h(\hat{x}) = \int_{Q_{\hat{x}}} (q(e) - \hat{x}) \, f(e) \, de \,, \tag{3.18}$$

where $f(e)$ is the stationary density of the state $e = (G, M)$.

Using this definition of $h(\hat{x})$, and letting $X = [\underline{z}, \bar{n}]$, we can write $var(q)$ as

$$\begin{aligned} var(q) &= \int_X (h(\hat{x}) + \hat{x} - \mu)^2 \, d\hat{x} \\ &= \int_X (\hat{x}^2 + \mu^2 - 2\mu\hat{x}) \, d\hat{x} + \int_X h(\hat{x})^2 \, d\hat{x} - 2\mu \int_X h(\hat{x}) \, d\hat{x} + 2 \int_X h(\hat{x}) \, \hat{x} \, d\hat{x} \\ &= var(x) + \int_X h(\hat{x})^2 \, d\hat{x} + 2 \int_X h(\hat{x}) \, \hat{x} \, d\hat{x} \end{aligned}$$

We can write out a similar expression for $cov(x, q)$.

$$\begin{aligned} cov(x, q) &= \int_X (h(\hat{x}) + \hat{x})(\hat{x} - \mu^2) \, d\hat{x} \\ &= \int_X h(\hat{x}) \, \hat{x} \, d\hat{x} - \mu \int_X h(\hat{x}) \, d\hat{x} + \int_X \hat{x}^2 \, d\hat{x} - \mu \int_X \hat{x} \, d\hat{x} \\ &= \int_X h(\hat{x}) \, \hat{x} \, d\hat{x} + var(x) \end{aligned}$$

From these expressions it can be seen that sufficient condition for $var(q) > var(x)$, and a necessary and sufficient for $cov(x, q) > var(x)$ is

$$\int_X h(\hat{x}) \, \hat{x} \, d\hat{x} > 0 \,. \tag{3.19}$$

From Proposition 3.4.2, there is an upper bound to sales which lead the firm to produce less than it has sold. Define the function $\hat{a}(n)$ to be the lowest sale x for which the firm will set its net production $q = x$, given that it had provided n goods for sale. Suppose in period $t - 1$ the firm is unconstrained, and therefore puts up $\bar{n}$ goods for sale. Then $\hat{a}(\bar{n})$ is the lowest sale for which the firm will put up $\bar{n}$ for sale in period t. That is, $\hat{a}(\bar{n})$ will be the sale which satisfies the following equation,

$$px_t = \bar{n}_t - G_t + \delta(G_t) \,.$$

Substituting for G_t we can rewrite the above equation as

$$(p - 1)x_t = \delta(\bar{n}_t - x_t) \,. \tag{3.20}$$

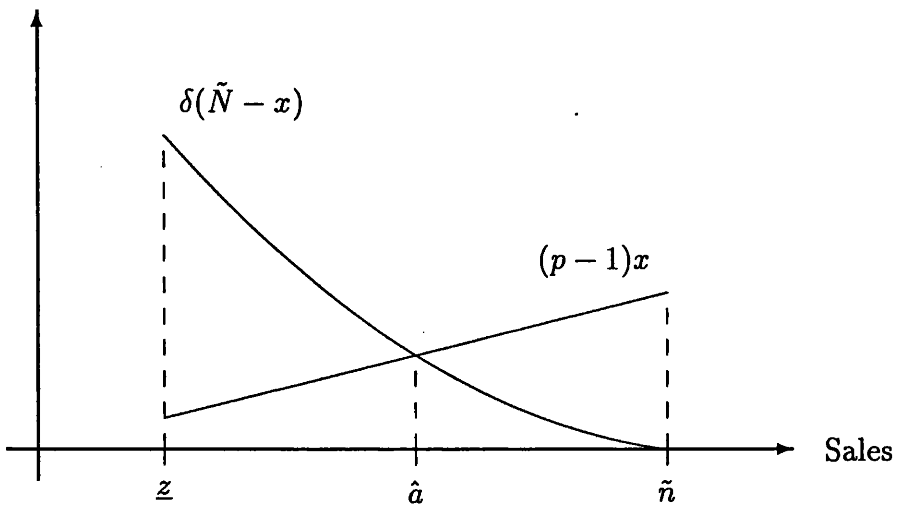

Figure 3.4: Existence of $\hat{a}$

Thus $\hat{a}(\bar{n}) = \delta(\bar{n} - \hat{a}(\bar{n}))/(p-1)$. The sale $\hat{a}(\bar{n})$ generates just the amount of revenue required to get the firm back to $\bar{n}$. To see that $\hat{a}(\bar{n}) \in (\underline{z}, \bar{n})$ look at Figure 3.4, which is the graph of equation (3.20). Since $p > 1$, $\hat{a}(\bar{n})$ is always strictly in this interval. Moreover, for $n < \bar{n}$, the function $\delta(n - \hat{a}(n))$ shifts towards the origin, thereby reducing $\hat{a}(n)$ for smaller n. Thus, if the firm provides less than $\bar{n}$ for sale, and sells more than $\hat{a}(\bar{n})$ it will either increase the amount it puts up for sale next period (i.e. over-produce), or increase its savings, or both (it may also start to consume if $(\bar{n}, \tilde{s})$ is an interior point in its new feasibility set). If the firm has provided $\bar{n}$ for sale, and sells more than $\hat{a}(\bar{n})$, the firm will again provide $\bar{n}$, and will add to either savings or consumption (or both).

Therefore, under-production will occur only in the interval $[\underline{z}, \hat{a})$, whereas over-production will occur in the interval $(\underline{z}, \bar{n})$. Thus, $h(x)$ can assume negative values only in the interval $[\underline{z}, \hat{a})$. Over the interval $[\hat{a}(\bar{n}), \bar{n})$, $h(x) > 0$. Moreover, since the firm never wants to provide more that $\bar{n}$ for sale, it will never overproduce after it sells $\bar{n}$. Therefore, $h(\bar{n}) = 0$. Finally, we know from corollary 3.4.4 that the means of production and sales are equal. Therefore, it is true that

$$\int_X h(\hat{x})\, d\hat{x} = 0\,. \tag{3.21}$$

Rewriting equation (3.21) gives us

$$\int_{\underline{z}}^{\hat{a}(\bar{n})} h(\hat{x})\, d\hat{x} + \int_{\hat{a}(\bar{n})}^{\bar{n}} h(\hat{x})\, d\hat{x} = 0\,.$$

From the properties of $h(x)$, it is true that

$$\int_{\underline{z}}^{\hat{a}(\bar{n})} h(\hat{x})\hat{x}\, d\hat{x} + \int_{\hat{a}(\bar{n})}^{\bar{n}} h(\hat{x})\hat{x}\, d\hat{x} > 0\,,$$

which is what we desire to prove. □

Corollary 3.5.2 *Changes in the stock of inventory put up for sale, Δn_t, covary positively with sales, x_t.*

Proof: Writing out the expression for $cov(\Delta n_t, x_t)$ we get

$$\begin{aligned} cov(\Delta n_t, x_t) &= cov(n_t, x_t) - cov(n_{t-1}, x_t) \\ &= cov(q_t + G_t, x_t) - cov(n_{t-1}, x_t) \\ &= cov(q_t, x_t) + cov(n_{t-1} - x_t, x_t) - cov(n_{t-1}, x_t) \\ &= cov(q_t, x_t) - var(x_t)\,, \end{aligned}$$

which is positive from the above Proposition.

Proposition 5.2: *Sales exhibit positive first-order autocorrelation.*

Proof: Let $x_t = \hat{x}$ and take $Q(\hat{x})$ to be defined as above. For each $q(G, M)$ such that $(G, M) \in Q(\hat{x})$, the set of feasible sales next period is the interval $[\underline{z}, n(G, M)]$. Define the probability that $x_{t+1} = x'$ conditional on the firm providing $n(G, M)$

goods for sale to be

$$\lambda_n(x') = \begin{cases} \phi(x') & \text{if } x' < n(G,M) \\ 1-\Phi(n) & \text{if } x' = n(G,M) \\ 0 & \text{if } x' > n(G,M) \end{cases}$$

Thus we can write the expected difference of next period's sale from the unconditional mean, μ, conditional on this period's sale $x_t = \hat{x}$ as

$$m(\hat{x}) = \int_{Q(\hat{x})} \int_{\underline{z}}^{n(e)} (x' - \mu)\,\lambda_n(x')\,f(e)\,de\,dx'\,,$$

where again $f(e)$ is the stationary density of the state $e = (G, M)$. The covariance of sales with its first lag is then

$$\begin{aligned} cov(x', x) &= \int_X m(\hat{x})(\hat{x} - \mu)\,d\hat{x} \\ &= \int_X (m(\hat{x})\,\hat{x} - m(\hat{x})\,\mu)\,d\hat{x} \\ &= \int_X m(\hat{x})\,\hat{x}\,d\hat{x}\,, \end{aligned} \tag{3.22}$$

where the third line follows from the fact when we integrate the conditional mean of next period's sales over all possible sales this period we get the unconditional mean. In other words,

$$\int_X m(\hat{x})\,d\hat{x} = 0\,. \tag{3.23}$$

From Proposition 3.4.5, it follows that an increase in the sales realisation leads to a target inventory, n, which is at least as high as that without the increase. Therefore, the conditional mean $m(x)$ is nondecreasing in x. Moreover, since the firm does not always put up $\bar{n}$ for sale, it must be true that $m(\bar{n}) > 0$. This implies that there is a sale, $\hat{w} \in [\underline{z}, \bar{n}]$ such that for all $x > \hat{w}$, the conditional expectation of next period's differential between sales and the unconditional mean of sales is positive, i.e. $m(\bar{n}) > 0$. Thus, positive weight is attached to higher values of x and the same negative weight (in absolute terms) to lower values of x. This implies that

$$\int_X m(\hat{x})\,\hat{x}\,d\hat{x} > 0. \tag{3.24}$$

□

3.6 Conclusions

This paper has presented a model of inventory investment in which the firm's access to external sources of finance is constrained. This model can explain the excess variance of production puzzle, and yields the prediction of positive co-variation between inventory investment and sales. Both of these facts are widely reported, and contradict the predictions of the standard Linear-Quadratic model of inventory investment. An equally important result derived from the model is that the endogenous sales process exhibits positive serial correlation, even though the underlying (and unobservable) demand process is exogenously specified to be serially uncorrelated.

The intuition behind these results was illustrated with an example. In that example, after a low sales realization, the firm is left simultaneously with a low level of internal funds and a high level of depreciation. Consequently, it does not have sufficient funds to be able to replace both the goods that were sold and those which depreciated. Therefore, the firm underproduces. After a medium sales realization, the firm has more revenue from sales and less depreciation occurs. In these circumstances, the firm could overproduce if it chose to do so. After a very high sales realization, the firm neither overproduces nor underproduces since it has more than enough funds to provide the first best amount of goods for sale. Thus, since underproduction is associated with low sales realizations and overproduction with medium sales realizations, the financing constraints effectively cause the distribution of production to be a mean-preserving spread of the sales distribution, thereby explaining the excess variance of production puzzle.

3.A Appendix A

In this section, we demonstrate that there are only four independent conditions which must be satisfied in order for the postulated policy to be optimal. First we compute the difference in value between the equivalence classes of states. The difference between the first two classes is given by

$$v_2 - v_1 = [(1-\phi_0)(v_3 - v_2) + \phi_2], \tag{3.25}$$

A demand of 1 or 2, which happens with probability $(1-\phi_0)$, brings a firm from equivalence class v_2 to class v_3, but a firm from v_1 only into v_2. Moreover, the former enjoys one unit of consumption if demand is 2 (probability ϕ_2).

For the difference between the second and the third class we have

$$v_3 - v_2 = \beta[\phi_0(v_2 - v_1) + (1-\phi_0)]. \tag{3.26}$$

A demand of 1 or 2, which happens with probability $(1-\phi_0)$, yields a firm from equivalence class v_3 one consumption unit more than a firm from v_2. Moreover, with zero demand (probability ϕ_2), the former ends up in v_2 whereas the latter is thrown back to v_1.

Thus, we obtain a system of two linear equations in the two unknowns $v_2 - v_1$ and $v_3 - v_2$, the solution to which is given by

$$v_2 - v_1 = \frac{\beta^2(1-\phi_0)^2 + \beta\phi_2}{1 - \beta^2\phi_0(1-\phi_0)}, \tag{3.27}$$

and

$$v_3 - v_2 = \frac{\beta^2\phi_0\phi_2 + \beta(1-\phi_0)}{1 - \beta^2\phi_0(1-\phi_0)}. \tag{3.28}$$

Let us assume parameter values $\{p, \beta, \phi_0, \phi_1, \phi_2\} = \{2, 0.96, 0.15, 0.25, 0.6\}$. Substituting them into (3.27) and (3.28) yields

$$v_2 - v_1 = 1.4072$$

and

$$v_3 - v_2 = 1.0266,$$

respectively. Stepping into the next higher equivalence class, v_4, simply means an additional unit of consumption, and therefore $v_4 - v_3 = 1 < 1.0266 < 1.4072$. The same is true for the difference between v_5 and v_4. Thus, the value increments are strictly decreasing over the first four equivalence classes and remain constant thereafter. This illustrate the strict concavity of the value function in the range of assets for which the constraint $c \geq 0$ is binding and its linearity beyond that region.

For the same reasons as in the unconstrained model, it is never optimal to choose $n > 2$ or $n = 0$. However, due to the the additional constraint it is no longer preferable to always set $s = 0$.

That it is still optimal to provide $n = 2$ whenever possible is shown in the following arguments. Consuming the resource unit freed up by providing only $n = 1$ instead is not optimal if and only if

$$1 < \beta[(1 - \phi_0)(v_3 - v_2) + \phi_2]. \tag{3.29}$$

Since $1 < 1.4137$, this is clearly the case with our parameter values.

Due to the decreasing increments in value established above the *change* in expected valuation from saving it (rather than investing it in production) is bounded by $\beta[\phi_0(v_2 - v_1) + \phi_1(v_3 - v_3) + \phi_2(v_3 - v_4)]$. The second term is, of course, zero, and stated here only for expositional reasons. If the demand shock is equal to 1 the firm will end up in state $(1,2)$ regardless of its decisions about saving and investment.

Since $v_3 - v_4 = -1$, this expression is negative if and only if

$$\phi_2 > \phi_0(v_2 - v_1). \tag{3.30}$$

For our parameter values this translates to $0.6 > 0.2111$. Thus, it is optimal to set $n^*(N, S) = 2 \ \forall \ (G, M)$ such that this is feasible. It can be shown that, as a consequence, 29 out of the 40 feasible actions can be excluded from the optimal set on those grounds.

Next consider states $(0,4)$ and $(1,3)$, where $n = 2$ is clearly feasible and will therefore be chosen. Instead of setting $(n, s, c) = (2,1,1)$ the firm's decision could choose $(n, s, c) = (2,2,0)$. This is dominated by the former choice if and only if

$$\beta[\phi_0(v_3 - v_2) + \phi_1 + \phi_2] < 1,$$

or

$$\beta\phi_0(v_3 - v_2) < 1 - \beta(1 - \phi_0). \tag{3.31}$$

which is equivalent to $0.9638 < 1$, given Γ.

For $(2,0,2)$ not to be better than $(2,1,1)$,

$$\beta[\phi_0(v_2 - v_1) + \phi_1 + \phi_2] > 1,$$

or

$$\beta\phi_0(v_2 - v_1) > 1 - \beta(1 - \phi_0). \tag{3.32}$$

must hold $(1.0186 > 1)$.

Inequality (3.31) also ensures that in state $(0,5)$ the choice $(2,1,2)$ is preferred to $(2,2,1)$; and since it implies that

$$\beta[\phi_0(1 + v_3 - v_2) + 2\phi_1 + 2\phi_2] < 2,$$

$(2,1,2)$ will neither be dominated by $(2,3,0)$. (The additional unit saved has expected marginal valuation of $\beta < 1$ as it is used for consumption with certainty next period.)

Inequality (3.32) is required for $(2,0,3)$ not to be preferred to $(2,1,2)$. Thus, we have disposed of another 5 actions.

In $(G, M) = (1, 2)$, again only choices s.t $n = 2$ need to be considered. The choice of $(n, s, c) = (2, 1, 0)$ over $(n, s, c) = (2, 0, 1)$ requires (3.32) to hold, which we know to be true. Thus, the latter can be eliminated.

That the postulated action $(2, 0, 0)$ is optimal in the states $(0, 2)$ and $(2, 1)$ is true because all other possibilities involve $n < 2$ (i.e.. 6 actions excluded).

In the remaining two states, $(2, 0)$ and $(1, 0)$, $(1, 0, 0)$ is the only feasible action and therefore trivially optimal.

In total we have excluded $29 + 5 + 1 = 35$ suboptimal actions, leaving us with 5 optimal ones across all states.

3.B Appendix B

Proof of Propostion 3.4.1

Due to the non-negativity constraints (3.2) and (3.3), the state space for this problem is bounded below by zero. We assume arbitrary upper bounds to the state space, $\bar{M}$ and $\bar{G}$.[19] With this assumption it is straightforward to show that the conditions for Lemma 9.5 and Theorem 9.6 in Stokey, Lucas, and Prescott (1989) (pp.261-64) are satisfied. Thus, the Bellman operator defined on the right-hand side of (3.13) is a contraction mapping with a unique fixed point.

Because of the law of motion for money (3.7), the one period return to the firm, $M - s - (n - G + \delta(G))$, and the feasibility constraints are all increasing in the state variable, M, v is also increasing in M by theorem 9.7 in Stokey, Lucas, and Prescott (1989) (p.264). Similarly, since the law of motion for goods (3.8), and the firm's one-period return are also increasing in G, v is increasing in G as well.

[19]Below we show that solution to the problem is characterised by upper bounds on n and s, which implies that M and G are bounded, too.

3.C Appendix C

In this section we prove that inequality 3.17 holds. First, note that for very small μ, the decrease of the first term of the left-hand side due to the change μ in the (upper) integral boundary is offset by the increase of the second term of the left-hand side due to the same change μ in the (lower) integral boundary.

Since v is strictly concave when the non-negativity constraint on consumption is binding, the stockout terms (without probabilities) of the two inequalities compare as follows:

$$v(0,s+\varepsilon+(p-1)\mu+pn)-v(0,s+\varepsilon+pn)\leq v(0,s+(p-1)\mu+pn)-v(0,s+pn). \quad (3.33)$$

Moreover, conditional on not stocking out, reducing n in favour of s is at least as valuable at lower savings as it is at higher savings (because of depreciation). Hence, $\forall\ z \leq n-\mu$,

$$v(n-\mu-z,s+\varepsilon+\mu+pz)-v(n-z,s+\varepsilon+pz)\leq v(n-\mu-z,s+\mu+pz)-v(n-z,s+pz). \quad (3.34)$$

This is true because, by approximating both sides of the inequality arbitrarily closely, we can rewrite it as

$$-[v(n-z,s+\varepsilon+pz)-v(n-\mu-z,s+\varepsilon+pz)]\mu+[v(n-z,s+\varepsilon+\mu+pz)-v(n-z,s+\varepsilon+pz)]\mu$$

$$\leq -[v(n-z,s+pz)-v(n-\mu-z,s+pz)]\mu+[v(n-z,s+\mu+pz)-v(n-z,s+pz)]\mu, \quad (3.35)$$

or

$$[v(n-z,s+\varepsilon+\mu+pz)-v(n-z,s+\varepsilon+pz)]-[v(n-z,s+\mu+pz)-v(n-z,s+pz)]$$

$$\leq [v(n-z,s+\varepsilon+pz)-v(n-\mu-z,s+\varepsilon+pz)]-[v(n-z,s+pz)-v(n-\mu-z,s+pz)]. \quad (3.36)$$

Due to strict concavity of v, both sides of the inequality measure the extent to which the value gain, due to the funds being larger by ε, is reduced by an increase in savings and production, respectively. More precisely, the left-hand side of the inequality represents the *decrease in the value gain* (implied by an ε-increase) that stems from adding even more funds, μ, to savings. The right-hand side is equal to the *decrease in the value gain* (implied by an ε-increase) due to an increase μ of the target inventory. The former decrease is larger in absolute terms since the gross unit return on additional savings μ is 1 $\forall$ z, whereas the gross unit return on an additional goods provision of μ is strictly smaller than 1 for at least some z that do not lead to a stockout. An analogous argument can be made when there is a jump from state (G, M) to a state $(G + \varepsilon, M)$, as this would be equivalent to leaving G unchanged and increasing M by some fraction of ϵ (which depends on the depreciation technology).

Chapter 4

A New Test for Capital Market Imperfections

4.1 Introduction

Since the publication of Fazzari, Hubbard, and Petersen (1988), a large empirical literature on investment and financing constraints has developed. The overwhelming majority of these papers report findings which support the conclusion that the investment expenditures of some firms are constrained by their inability to raise sufficient amounts of external finance. Yet despite the preponderance of papers reporting such findings, many economists remain skeptical of this conclusion.[1]

The source of much of this skepticism can be traced to an identification problem. The main prediction distinguishing financially constrained from unconstrained firms is the following: shocks to a constrained firm's supply of internal finance will affect its investment expenditures even if they do not alter its expectations of the future profitability of its investment opportunities. This is not true of unconstrained firms.[2]

[1]Indeed, Kaplan and Zingales (1997) go so far as to allege that "it is likely that a publication selection bias exists in this literature."

[2]This prediction may also be caused by an agency problem between the managers and the owners of a firm, referred to as the problem of free cash flow (Jensen (1986)). Both explanations of this empirical prediction, capital market imperfections and free cash flow, share the same identification

Therefore, to test whether or not a firm is financially constrained, supply shocks to internal finance first must be identified, then their correlation with investment examined. Sceptics argue that the identification problem has not yet been solved.

Existing tests can be roughly classified into two categories: reduced-form tests, and structural tests. A generic drawback of reduced-form tests is that one cannot give a specific economic interpretation to the regression coefficients. Papers in the financing constraints literature attempt to mitigate this problem by first splitting their samples into different classes of firms, then comparing the magnitudes of coefficients across classes. Differences in the magnitude on cash flow variables are interpreted as evidence of financing constraints. However, reduced-form coefficients are functions of the structural parameters of the underlying economic model. Differences in coefficient magnitudes may also be explained by differences in the values assumed by these parameters in different samples. Since the identification problem is not solved, it is unclear that any of the difference in coefficient magnitudes can be attributed to financing constraints. To ascribe the difference to financing constraints, it must be assumed that the unidentified structural parameters are identical across different samples of firms.

The second group of tests are based on structural investment models, and thus are robust to this critique.[3] However, these tests are indirect because they test investment models under the assumption of perfect capital markets. The basic assumptions of the investment model, the assumptions required to derive testable restrictions from the model, and the assumption of perfect capital markets all enter the null hypothesis of these tests. Rejection of the null hypothesis may therefore be due to a violation of any of these assumptions. Indeed, some of the assumptions made in order to derive testable restrictions are particularly restrictive, and therefore likely to be rejected. Moreover, it is not possible to identify which restriction has been rejected from other empirical results (such as coefficient magnitudes).

---

problem.

[3]Most existing work in the literature has used the neo-classical Adjustment Costs model (e.g. Lucas (1967), Gould (1968), Hayashi (1982)).

The financing constraints test developed in this chapter is both structural and direct. That is, one of the assumptions in the null hypothesis of this test is that capital markets are *imperfect*. Thus, since failure to reject the null implies that the data is consistent with all of the assumptions which make up the null hypothesis, this test is robust to the criticism made of existing structural tests.

The test is based on a modified version of the neo-classical Adjustment Costs model. One of the modifications is the introduction of two disturbance terms to the model, one of which is non-stationary. It is assumed that both disturbances are observable by the firm, and that their innovations are orthogonal. Realisations of the non-stationary disturbance permanently affect the levels of variables such as profits, sales, and physical capital. However, although these variables are non-stationary, their ratios are stationary. In other words, the logs of these variables are cointegrated.

A vector moving-average representation is derived for the first difference of the log of cash flow, and the log of the cash-flow capital ratio. When the firm is unconstrained, both of these variables are endogenous in the sense that each Granger causes the other. One way to understand this is that both cash flow and the capital stock adjust in order to maintain the long-run equilibrium relationship between the two variables. Moreover, since both disturbances have a transitory effect on the cash flow-capital ratio, but only the non-stationary disturbance has a permanent effect on the level of cash flow, the identification restriction of Blanchard and Quah (1989) can be used recover the structural disturbances from actual data.

In contrast, when financing constraints are binding, the first difference of the log of cash flow is exogenous. In other words, it is not Granger caused by the log of the cash flow-capital ratio. This is because, when financing constraints are binding, the firm spends all of its available profits on investment. Therefore, the only variable which is correcting the error in the long-run equilibrium between the two variables is the capital stock. Moreover, in this system both structural disturbances have a long-run effect on the level of cash flow. Thus, a Blanchard-Quah decomposition of cash flow does not exist, and the structural disturbances are unidentified.

To test whether the data of a group of firms is consistent with the hypothesis that those firms are financially constrained, one tests the null hypothesis that the first difference of the log of cash flow is not Granger caused by the log of the cash flow-capital ratio. Granger causality tests are very common and easily implemented. Therefore, an additional strength of this test is its simplicity.

The test was performed on an unbalanced panel of US manufacturing firms sampled at a quarterly frequency between 1980 and 1992. The sample was stratified according to three criteria: size, existence of a bond rating, and existence of a commercial paper rating. For the sample of small firms, and the sample of firms without a bond rating, the null hypothesis that the cash flow-capital ratio does not Granger cause the first difference of cash flow fails to be rejected. For both the sample of large firms, and the sample of firms with a bond rating, the null hypothesis is easily rejected. The null hypothesis is also rejected for both samples of firms with and without commercial paper ratings.

The rest of the chapter is organised as follows. The next section describes the investment model, states the main theoretical propositions, In the third section the financing constraints test is derived. In the following section the data is described, and the Granger causality tests run. The final section concludes. Proofs of all propositions are found in the Appendix.

4.2 A Modified Adjustment Costs Model of Investment

By far the most widely used model in the empirical literature on financing constraints is the Adjustment Costs model of Investment, which was first introduced by Eisner and Strotz (1963), Lucas (1967) and Gould (1968). What can be uncontroversially concluded from the results in this literature is that the model does not explain the data very well, particularly for firm classified *a priori* as financially constrained. The controversy surrounding this literature has to do with its attribution of the model's failure to the violation of the perfect capital markets assumption. This assumption is one of several in the joint null hypotheses tested in the literature. The following three assumptions, in addition to the perfect capital markets assumption, underlie what will be referred to as the standard formulation of the Adjustment Costs model:[4] (i) linearity of the firm's reduced-form profit function, (ii), stationarity of the stochastic processes governing the exogenous disturbances, and (iii), no fixed costs of adjusting the firm's capital stock. Violation of any of these three assumptions would lead to a rejection of the model.

In designing a new test for financing constraints it is desirable to use a model which can potentially explain existing results without appealing to capital market imperfections. In order to do so, the model used in this paper changes the first two assumptions. The popularity of the first assumption is derived from the result that under it, unobservable marginal q and observable average q are equivalent (Hayashi (1982), Lucas and Prescott (1971)). However, known conditions which ensure that the profit function is linear are quite restrictive. The firm must have a constant returns to scale production technology, and act as a price taker in both factor and output markets. In this chapter, the profit function is assumed to be strictly concave in the capital stock, which does not require both conditions to be satisfied simultaneously.

The assumption made in this model that one of the underlying disturbances is non-stationary can ultimately be justified on empirical grounds. Indeed, as will be shown later, for a substantial fraction of firms in the dataset used in the empirical analysis, the null hypothesis of a unit root in both the cash flow and capital stock series cannot be rejected. However, it is also an assumption which is particularly germane to the issue of financing constraints. It is likely that a high growth rate is an important characteristic of firms which are classified as constrained. Not all papers in the literature have reported the growth rates of the firms in their samples, but of those that have, it is typically the case that the average growth rate of constrained firms

[4]All of the papers in the financing constraints literature incorporate at least one of these assumptions into their null hypotheses.

is higher than that of unconstrained firms.[5]

The focus on growing firms is sensible, since a firm's demand for capital must be high enough to make potential financing constraints bind (or very nearly bind). However, there is no good reason to assume that this growth is deterministic. Indeed, there are reasons to prefer the choice of stochastic growth in a model of investment aimed at explaining the behaviour of firms which have been classified as constrained in the financing constraints literature. One would expect that external investors have greater difficulty forecasting the earnings of constrained firms as opposed to unconstrained firms. While this difficulty may in part be due to a shortage of data, a unit root in the firms' reported earnings series will add to the uncertainty of the forecasts. Moreover, as mentioned above, if some firms are growing stochastically, this will lead to a rejection of the some of the testable restrictions derived from the standard formulation of the Adjustment Costs model. For example, in the model in this paper, *both* average q and the cash flow-capital ratio are sufficient statistics for the investment rate of *unconstrained* firms. This provides an alternative explanation of why the cash-flow capital ratio has been found to be so important in explaining the investment behaviour of firms classified as constrained. Thus, from a methodological perspective, a reason to prefer the assumption of stochastic growth is that it can explain some of the findings in this literature without having to resort to the assumption of imperfect capital markets.

4.2.1 The Investment Model

This section outlines the model of investment without financing constraints, and derives some basic theoretical results. All proofs are in the Appendix.

The key components of the Adjustment Costs model are the functions describing

[5]This is certainly true of the original Fazzari, Hubbard, and Petersen (1988) article. Kaplan and Zingales (1997) investigate the importance of firm growth rates in their critique of that article. They conclude that "any splitting criterion that sorts firms into subsamples with differential outliers in growth rates...may be biased toward finding a difference in coefficients on cash flow. This bias may partially account for the large body of evidence finding a higher investment-cash flow sensitivity in fast growing companies, that tend to be classified as financially constrained."

the firm's profits (per time period) and the costs of adjusting the stock of physical capital, as well as the stochastic processes governing the exogenous shocks. The firm's profits in period t are represented by the function $\Pi(K_t, \theta_t) : \Re^{++} \times \Re^{++} \rightarrow \Re$, where K_t is the capital stock, and θ_t is an exogenous random variable. The profit function is assumed to be increasing in both variables, and, as mentioned above, concave in the capital stock, K_t. In addition, it is assumed to be concave in θ_t, and linear homogeneous in both K_t and θ_t. As only the capital stock and profits are observable, the linear homogeneity assumption is, in effect, a normalisation of θ.

θ is an exogenous non-stationary disturbance which follows the process given by

$$\frac{\theta_t}{\theta_{t-1}} = \tilde{\mu}_t. \tag{4.1}$$

The term $\mu = \log(\tilde{\mu})$ is an i.i.d. innovation distributed over support $[\underline{\mu}, \overline{\mu}]$ with continuous density function $\phi(\mu)$. It does not necessarily have mean zero.

The final modification to the standard Adjustment Costs model is the introduction of a transitory component to the firm's earnings. This is done by redefining the firm's profits to be $\tilde{\eta}_t \Pi(K_t, \theta_t)$, where $\tilde{\eta}_t = e^{\eta_t}$, and η is a white noise process distributed over support $[\underline{\eta}, \overline{\eta}]$ with continuous density function $\varphi(\eta)$. It is assumed that the firm can observe both shocks.

Adjustment costs are assumed to be the same as in the standard formulation of the model. Let $\Psi(I, K)$ represent the nominal costs of adjusting the capital stock where I is real investment. $\Psi(I, K)$ is homogeneous of degree one in I and K. For some $a \in (0, 1)$, $\lim_{I \rightarrow aK} \Psi_I(I, K) = \infty$. The price of capital is assumed to be an exogenously given constant which is normalised to 1. Therefore, the total amount spent on investment in period t is given by $I_t + \Psi(I_t, K_t)$.

The discount factor is given by β. It is assumed that capital does not depreciate. Therefore, the law of motion for the capital stock is given by,

$$K_{t+1} = K_t + I_t. \tag{4.2}$$

The firm's objective is to

$$\max_{\{I_t\}_{t=0}^{\infty}} E_0\left(\sum_{t=0}^{\infty}\beta^t(\tilde{\eta}_t\Pi(K_t,\theta_t)-(I_t+\Psi(I_t,K_t))\right), \tag{4.3}$$

subject to (4.2), and (4.1).

This problem can be solved using dynamic programming techniques. The Bellman equation corresponding to this problem is

$$\begin{aligned} V(K_t,\theta_t;\eta_t) = \; & \max_{I_t} \; \{\tilde{\eta}_t\Pi(K_t,\theta_t)-(I_t+\Psi(I_t,K_t)) \\ & + \; \beta\int\int V(K_{t+1},\theta_t\mu_{t+1};\eta_{t+1})\,\phi(\mu)\,\varphi(\eta)\,d\mu\,d\eta\} \end{aligned}$$

subject to (4.2).

The first proposition establishes the existence of the solution to this equation, and some of its properties.

Proposition 4.2.1 *There exists a unique function, $V(K,\theta;\eta)$, which solves the Bellman equation. This function has the following properties.*

1. *$V(K,\theta;\eta)$ is homogeneous of degree one in K and θ.*
2. *$V(\cdot,\cdot;\eta)$ is increasing.*
3. *$V(\cdot,\theta;\eta)$ is strictly concave.*
4. *The optimal investment policy correspondence, $I^*(K,\theta)$, is a continuous single-valued function which is homogeneous of degree 1 in (K,θ).*
5. *$V(K,\theta;\eta)$ is differentiable with respect to K.*

Any sequence of the capital stock $\{K_t\}_{t=0}^{\infty}$ generated by the policy function $I^*(K,\theta)$ and equation (4.1) obviously will be non-stationary. However, it can be shown that a stationary solution for the sequence $\{(K/\theta)_t\}_{t=0}^{\infty}$ exists. To do this it is convenient to make the following definitions. Let $k\equiv\frac{K}{\theta}$, $i\equiv\frac{I}{K}$, $v(k)\equiv V(k,1)$, $\psi(i)\equiv\Psi(i,1)$

and $\pi(k) \equiv \Pi(k, 1)$. Using these definitions and the linear homogeneity of the value function, we can rewrite the Bellman equation as

$$v(k_t; \eta_t) = \max_{i_t} \left\{ \tilde{\eta}_t \pi(k_t) - (i + \psi(i_t))k_t + \beta \int \int v(k_{t+1}; \eta_{t+1}) \, \tilde{\mu}_{t+1} \, \phi(\mu) \, \varphi(\eta) \, d\mu \, d\eta \right\}, \tag{4.4}$$

subject to

$$k_{t+1} = \frac{(1 + i_t)k_t}{\tilde{\mu}_{t+1}} \tag{4.5}$$

and (4.1).

It is relatively straightforward to show that a stationary distribution for k exists. This result is stated in Proposition (4.2.2).

Proposition 4.2.2 *There exists a unique invariant distribution for k over the finite interval* $[\underline{k}, \overline{k}]$.

The first order condition for this problem is

$$1 + \psi'(i_t) = \beta \int \int v' \left(\frac{(1 + i_t)k_t}{\tilde{\mu}_{t+1}}; \eta_{t+1} \right) \phi(\mu) \, \varphi(\eta) \, d\mu \, d\eta. \tag{4.6}$$

To understand how the model works, first note that for each value of θ, there is a unique value of K which maximize $\Pi(K, \theta)$. Consider an above average realisation of μ_{t+1}. Since μ is white noise, the expectation of $\tilde{\mu}_{t+1}$ does not affect i_t. However, from equation (4.5), a high realisation of $\tilde{\mu}_{t+1}$ lowers k_{t+1}. Note that the strict concavity of $V(K, \theta; \eta)$ with respect to K implies that $v'(k; \eta)$ is strictly decreasing. Therefore, an above average realisation of $\tilde{\mu}_{t+1}$ increases the marginal value of capital in period $t + 1$. This induces an immediate increase in i_{t+1} which, because of the one period lag before new capital becomes productive, lowers the marginal value of capital in period $t+2$ cateris paribus. If there were no costs of adjusting the capital stock, the marginal value of capital would be an i.i.d. process (and if newly installed capital immediately became productive it would be constant). Given that μ is assumed to be i.i.d., richer dynamics in the series of the marginal value of capital and other endogenous variables are due to adjustment costs. The assumptions made about adjustment costs imply that the capital stock gradually adjusts towards its new optimal level.

There are several features of this formulation of the Adjustment Costs model which are interesting to contrast with the standard formulation. First, note that from equation (4.6), the optimal investment function $i^*(k)$ is a strictly decreasing function of k. Therefore, any monotonic function of k will be a sufficient statistic for investment, I/K. Since $\Pi/K = \Pi(1, k^{-1})$, and $\Pi(K, \theta)$ is strictly increasing in its second argument, Π/K is strictly decreasing in k. A similar argument is true for $V(K, \theta)$. Hence, both Π/K and V/K are sufficient statistics for investment. In the standard formulation of the model only average q, V/K, is a sufficient statistic for I/K.

This is the same result as that derived in Caballero and Leahy (1996) under the assumption of fixed costs of adjustment. As discussed in that paper, the sufficiency result depends on the ability of both models to reduce the investment problem down to a problem with only one state variable, k, which, in turn, depends on the assumptions of linear homogeneity and of i.i.d. disturbances. If either of these assumptions are dropped, then there will be more than one state variable in the problem. In these circumstances, it is likely that no single observable variable will act as a sufficient statistic for investment. Rather, several regressors will be required to "summarise" all the information relevant to investment demand.

This result has serious implications for the interpretations given to some empirical results in the financing constraints literature. Typical results in the literature on financing constraints are that when both Π/K and V/K are included as regressors with I/K as the dependent variable, both are statistically significant. While this is a rejection of the standard formulation of the model, it does not reject this formulation of the model. Therefore, what has been interpreted as evidence of imperfect capital markets may instead be evidence that other assumptions in the standard formulation are invalid.

In fairness, the capital market imperfections interpretation rests not only on the rejection of the standard formulation, but also on a comparison of the coefficients on the Π/K term across different classes of firms. This coefficient is much larger for the groups of firms classified *a priori* as constrained than for other groups of firms. However, it is improper to base statistical inference about the validity of an economic model on comparisons of reduced-form coefficient magnitudes. Differences in the magnitudes of these reduced-form coefficients may simply be due to differences in the values assumed by the structural parameters across groups. Thus, the error in ascribing differences in these magnitudes to differences in cost premia on external finance may be understood as a version of the Lucas critique.

A second feature of this formulation which differs from that of the standard formulation is that average q, marginal q, and the average and marginal profitability of capital are endogenous.[6] In the standard formulation, they are all functions *only* of the exogenous disturbances. This is due to the assumption that the profit function is linear in capital. Therefore, in the standard formulation, shocks to marginal q induce a change in investment, but investment has no effect on marginal q. One of the implications of this assumption is that the exogenous disturbances affecting marginal q must be stationary in order for ratios of observable variables such as average q and Π/K to be stationary. In contrast, in this formulation both variables affect each other. Indeed, it is the effect of investment which causes marginal q to return to its unconditional mean. It is perhaps more intuitive to think of the process for the marginal value of capital as being driven by the reaction of investment to exogenous changes in the firm's set of investment opportunities, rather than as a strictly exogenous process.

Indeed, investment plays a crucial role in the stationarity of this problem. Note that since $V(K,\theta) = KV(1,k^{-1})$, proposition (4.2.2) implies that the ratio $V(K,\theta)/K$ is stationary random variable. The same is true for the ratios I/K and Π/K. By

[6]Endogeneity is used here in the economic sense associated with the idea of Granger causality (see Sargent (1987)). In both models, it is likely that these variables will be endogenous in the econometric sense. That is, they are likely to be correlated with the error term in an investment regression.

construction, none of the variables in the model are stationary in log levels. However, they are all stationary in log differences. Therefore, an implication of this model is that the log levels of I, V, K, Π are cointegrated, and that the cointegrating vector between each pair is $(1, -1)$. Viewing the model from this perspective, the structural shocks μ and η drive the ratios away from their unconditional means. However, there are two components of the error correction mechanism which maintain the cointegrated relationship between $V(K, \theta), K$, and $\Pi(K, \theta)$. The first component is investment itself. When there is an above average realisation of μ, so that both V/K and Π/K are above their equilibrium values, the firm invests, thereby raising the level of K, until both ratios return to their equilibrium. The second component of the error correction mechanism is the decay of the stationary shock η. Consider the impulse response of these ratios to a shock in η. Since η is white noise, the initial shock drives the ratio away from its equilibrium value for only the first period, after which the equilibrium is restored.

4.2.2 The Financially Constrained Model

The ultimate goal of this model is to form the basis of a test for capital market imperfections. Consequently, the discussion of a financially constrained firm is restricted to the simplest possible model. As will be shown below, the modelling assumptions made in this section are sufficient for the derivation of a testable restriction to distinguish constrained from unconstrained firms. Whether or not this restriction can also be derived from more general models of financially constrained firms must be left for future research.

The key feature of the investment behaviour of constrained firms which the test exploits is the exhaustion of the firm's (cheaper) internal funds. Thus, the financing constraint imposed is assumed to take an extremely simple form. In a given period a firm cannot spend more on investment than a given proportion, $\lambda \in (0, 1)$, of the profits it earns in that period. In other words, the constraint

$$\tilde{\eta}_t \Pi(K_t, \theta_t) - I_t - \Psi(I_t, K_t) \geq (1 - \lambda)(\tilde{\eta}_t \Pi(K_t, \theta_t)), \tag{4.7}$$

must be satisfied in every period. The firm neither has access to external funds, nor can it save. More general financing constraints would permit access to external finance at a premium. An interpretation of this constraint is that the cost premium on external finance is infinite. The term on the right-hand side of the above inequality can be interpreted as minimum dividend payment the firm must make to its owners. The size of this payment as a proportion of earnings, $1-\lambda$, is determined outside the model.

Thus, the problem facing a constrained firm is to maximise (4.3) subject to (4.2), (4.5), and (4.7).

Proposition 4.2.3 *There exists a unique function, $V(K,\theta,\eta)$, which solves the Bellman equation (4.4) subject to (4.2), (4.5) and (4.7). This function has the following properties.*

1. *$V(K,\theta,\eta)$ is homogeneous of degree one in K and θ.*
2. *$V(\cdot,\cdot,\eta)$ is increasing.*
3. *$V(K,\cdot,\eta)$ is strictly concave.*
4. *The optimal investment policy correspondence, $I^*(K,\theta,\eta)$, is a continuous single-valued function which is homogeneous of degree 1.*

A key difference between the unconstrained and constrained problems is that it cannot be proven that the value function is differentiable in the solution to latter. This greatly restricts one's ability to characterise the optimal investment function. In turn, knowledge of this function greatly simplifies both the proof of a stationary distribution for k in the constrained case, and the derivation of the financing constraints test. Therefore, the remainder of the discussion will focus on the case where the financing constraint (4.7) binds in every period. When this is the case, inequality (4.7) becomes an equality which implicitly defines the optimal investment function.

To establish the existence of a stationary distribution for k for the case where the financing constraint always binds, first it must be shown that there exists a specification of the model (i.e. a vector of parameter values) such that desired investment expenditures exceed profits in every period. The next lemma establishes that result. Proposition (4.2.5) then states the desired result for this case.

Lemma 4.2.4 *In the solution to the unconstrained problem, for each $\eta \in [\underline{\eta}, \overline{\eta}]$ there exists a value of k, $\tilde{k}(\eta)$, such that $\lambda\eta\pi(\tilde{k}(\eta))/\tilde{k}(\eta) = i(\tilde{k}(\eta)) + \psi(i(\tilde{k}(\eta)))$, and for $k < \tilde{k}(\eta)$, $\lambda\eta\pi(k)/k < i(k) + \psi(i(k))$. There exists a specification of the model in the constrained problem in which $\overline{k}_c < \tilde{k}(\overline{\eta})$, where $\overline{k}_c$ is the upper bound of the distribution of the state variable k in the constrained problem.*

According to lemma (4.2.4) over the interval $(0, \tilde{k}(\overline{\eta}))$ the firm would like to spend more than $\lambda\eta\pi(k)/k$ on installing new capital. Thus, for values of k in this region, the financing constraint (4.7) would be binding. The second part of the lemma states that we can find a $\lambda \in (0,1)$ such that the stationary distribution for the state variable in the constrained problem k_c, $[\underline{k}_c, \overline{k}_c] \in (0, \tilde{k}(\overline{\eta}))$. The financing constraint will always be binding for such a firm. The next proposition states that there are specifications of the support for μ such that in these circumstances a stationary distribution for k still exists.

Proposition 4.2.5 *There exists a unique invariant distribution for k in the constrained problem over the finite interval $[\underline{k_c}, \overline{k_c}]$.*

We now have all of the information about the solution to the firm's investment problem required for a test to distinguish between the two cases.

4.3 A Test for Financing Constraints

The methodology of the test is to demonstrate that when the financing constraint always binds, this places a testable restriction on the Wold representation of a given vector process of observable variables generated by the investment model. This is done by first deriving the "true" representation of the process when constraints never bind, and then when they always bind. The "true" representation refers to the MA representation with the error vector $(\mu_t,\ \eta_t)'$. It is then shown that the difference in these representations imply a testable difference in the corresponding Wold representations.

As a first step in deriving the "true" representations, consider the law of motion for the state variable k. Taking logs of equation (4.5) gives

$$\log(k_t) = \log(1 + i(k_{t-1})) + \log(k_{t-1}) - \log\mu_t. \tag{4.8}$$

Since the optimal policy function $i^\star(k)$ is monotonically decreasing in k, we can let $f(\log(k_{t-1})) = i^\star(\exp(\log(k_{t-1}))$. Taking a first order Taylor expansion around the unconditional mean of k, $\bar{k}$, we get,

$$i^\star(k_{t-1}) - i^\star(\bar{k}) = f'(\log(\bar{k}))(\log(k_{t-1}) - \log(\bar{k})). \tag{4.9}$$

Noting that $\log(1 + i^\star(k_{t-1})) \approx i^\star(k_{t-1})$ we can rewrite equation (4.8) as

$$\log(k_t) = \delta + \rho\log(k_{t-1}) - \mu_t, \tag{4.10}$$

where $\delta = i^\star(\bar{k}) + f'(\log(\bar{k}))$ and $\rho = (1 + f'(\log(\bar{k})))$, and where $\mu = \log(\tilde{\mu})$. Note that since $i^\star(k)$ is decreasing, $f'(\log(k)) < 0$, and therefore $\rho < 1$. The constant term δ can also incorporate the unconditional mean of μ if it is not equal to zero.

Define $CF_t = \tilde{\eta}_t\Pi(K_t, \theta_t)$. Using this definition, we can derive the following equations,

$$\begin{aligned}
\Delta\log(CF)_t &= \alpha\Delta\log(k)_t + \mu_t + \Delta\eta_t, && (4.11)\\
\log(CF/K)_t &= (\alpha - 1)\log(k)_t + \eta_t. && (4.12)
\end{aligned}$$

With these equations and equation (4.10), we can derive a vector moving-average representation for $(\Delta \log(CF)_t, \log(CF/K)_t)'$. For simplicity, assume that the functional form taken by the profit function is Cobb-Douglas, i.e. $\Pi(K_t, \theta_t) = K_t^{\alpha}\theta_t^{(1-\alpha)}$. With this assumption, $(\Delta \log(CF)_t, \log(CF/K)_t)'$ is a vector ARMA(1,1), and has the following moving average representation,

$$\begin{bmatrix} \Delta \log(CF_t) \\ \log(CF/K)_t \end{bmatrix} = \begin{bmatrix} \frac{1-\alpha+(\alpha-\rho)L}{1-\rho L} & \frac{(1-\rho)(1-L)}{1-\rho L} \\ \frac{1-\alpha}{1-\rho L} & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \eta_t \end{bmatrix} \tag{4.13}$$

where $\log(CF/K)_t$ now represents deviations from its unconditional mean. Since $\log(CF)_t$ and $\log(K)_t$ are cointegrated, $\Delta \log(CF)_t$ and $\Delta \log(K)_t$ will have a vector error-correction representation. Equation (4.13) is referred to as the companion form to the error-correction representation, and is easily derived from the latter (see Cochrane (1997)).

What does the moving-average representation (4.13) imply about the Wold representation for the process? Note that the polynomial in the upper right corner has a unit root. This implies that η has no long-run effect on the level of $\log(CF)$. Thus, the restriction used in Blanchard and Quah (1989) can be used to recover the "true" representation (4.13) from the Wold representation. In effect, this identification restriction performs a particular type of orthogonal permanent-transitory decomposition of the $\log(CF)$ series. The investment model implies that when the firm is unconstrained such a decomposition exists for $\log(CF)$. From Theorem 4.1 in Quah (1992), such a decomposition for $\log(CF)$ exists if and only if $\log(CF/K)_t$ Granger causes $\Delta \log(CF)_t$.[7] Thus, in a regression of $\Delta \log(CF)_t$ on its own lags and lags of $\log(CF/K)_t$, the latter group of terms should be jointly significant if the firm is unconstrained.

---

[7]The essence of the proof of this theorem is that there is a unique orthonormal matrix, Q, which transforms the Wold representation of $(\Delta \log(CF)_t, \log(CF/K)_t)'$ into its "true" representation (4.13). If $\log(CF/K)_t$ does not Granger cause $\Delta \log(CF)_t$, then the polynomial matrix associated with the Wold representation is lower triangular. This implies that Q is the identity matrix, and therefore that the polynomial matrix associated with the true representation is also lower triangular. In this case, the innovation to the stationary disturbance does not affect $\Delta \log(CF)_t$, and therefore $\log(CF)$ has no transitory component.

What happens to the representation (4.13) when the financing constraint always binds? To answer this question, first note that when the financing constraint always binds, inequality (4.7) becomes an equation which implicitly defines the optimal investment function. To make this explicit, we assume that the costs-of-adjustment function takes the form, $\psi(i) := \exp(\frac{(i-a)}{\gamma}) - i$. This allows the optimal investment function to be written as,

$$i = a + \gamma \log\left(\frac{\lambda\tilde{\eta}\Pi(K,\theta)}{K}\right). \tag{4.14}$$

Notice that the firms investment policy is now a function of the transitory shocks, $\tilde{\eta}$, even though this shock contains no information regarding the future marginal profitability of capital. Assuming the same Cobb-Douglas functional form for the profit function, and following the same procedure as above, the vector moving-average representation for $(\Delta\log(CF)_t, \log(CF/K)_t)'$ is now given by,

$$\begin{bmatrix} \Delta\log(CF_t) \\ \log(CF/K)_t \end{bmatrix} = \begin{bmatrix} \frac{(1-\alpha+(\alpha-\hat{\rho})L)}{1-\hat{\rho}L} & \frac{(1-(1-\gamma)L)(1-L)}{1-\hat{\rho}L} \\ \frac{1-\alpha}{1-\hat{\rho}L} & \frac{1-L}{1-\hat{\rho}L} \end{bmatrix} \begin{bmatrix} \mu_t \\ \eta_t \end{bmatrix}, \tag{4.15}$$

where $\hat{\rho} = 1 - (1-\alpha)\gamma$.

Let $\epsilon_t = (1-\alpha)\mu_t + (1-L)\eta_t$. Then the above representation can be re-written as,

$$\begin{bmatrix} \Delta\log(CF_t) \\ \log(CF/K)_t \end{bmatrix} = \begin{bmatrix} \frac{(1-\alpha+(\alpha-\hat{\rho})L)}{(1-\alpha)(1-\hat{\rho}L)} & 0 \\ 0 & \frac{1}{1-\hat{\rho}L} \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon_t \end{bmatrix}. \tag{4.16}$$

Two features of the representation in (4.16) are notable. The first is that the error term ϵ_t is a linear combination of the structural errors μ and η. In other words, both errors have the same effect on the observable variables in the system. This is because both variables affect investment. For a given value of (K_{t-1}, θ_{t-1}), we can see that total investment is determined by the realisation of the vector disturbance (μ_t, η_t). However, any realisation of this vector such that $\lambda\tilde{\eta}\Pi(K,\theta) = b$ for some $b > 0$ will lead to the same amount of investment when the firm is constrained. The true structural errors are thus indistinguishable. In particular, since the polynomial in the first row of the matrix in equation (4.16) does not have a unit root, both structural disturbances have a permanent effect on $\log(CF)_t$.

The second feature is that the elements in the vector error of (4.16) are identical. This feature captures the fact that investment is a deterministic function of cash flow as seen in equation (4.14). In the constrained system, cash flow is an exogenous stochastic process. Since the financing constraint is always binding, the capital stock adjusts in a perfectly deterministic way to realisations of cash flow. From the perspective of the test, this is the crucial difference between the constrained and unconstrained systems. In the unconstrained system, both variables are endogenous. The variables in the constrained system, however, are still cointegrated. But only the capital stock adjusts to maintain the long-run equilibrium.[8]

It is straightforward to show that this representation implies that the polynomial matrix of the Wold representation will be diagonal as well. Since the error term ϵ_t is an MA(1), it can be expressed as $\epsilon_t = (1 + \omega L)v_t$, where v_t is serially uncorrelated. Expressing the representation (4.16) in terms of v_t gives the Wold representation (i.e. the errors are serially uncorrelated, and the matrix polynomial is the identity when $L = 0$). Since this transformation involves multiplying the polynomial matrix by the polynomial scalar $1 + \omega L$, the polynomial matrix for the Wold representation is also diagonal. This implies that neither variable Granger causes the other.

Thus, the presence of a financing constraint which is binding in every period places very strong restrictions on the Wold representation for $(\,\Delta \log(CF)_t\,, \log(CF/K)_t\,)'$. In the theory, the term K_t represents the true replacement value of the firm's capital stock in period t (the price of capital was assumed constant and equal to 1). In the data, however, the capital stock is almost certainly measured with error.[9] A very

[8]This can be seen in the error correction representation for $(\Delta \log(CF)_t, \Delta \log(K)_t)'$. In the constrained system, the equilibrium error $\log(CF)_t - \log(K)_t$ enters only in the equation for $\Delta \log(K)_t$. In the unconstrained system, it enters both equations.

[9]Almost all papers in the empirical literature on investment and in the literature on financing constraints have been concerned with measurement error in the capital stock. Many papers use complicated algorithms to construct measurements of the replacement value of the capital stock

pertinent question then is whether these restrictions remain robust to measurement error in the capital stock. Suppose that the capital stock is measured with error, and that as a result the observed cash flow-capital ratio, $\log(\overline{CF/K})_t$, can be written as

$$\log(\overline{CF/K})_t = \log(CF/K)_t + \nu_t.$$

This implies that the moving-average representation for $(\,\Delta\log(CF)_t\,, \log(\overline{CF/K})_t\,)'$ is given by,

$$\begin{bmatrix} \Delta\log(CF_t) \\ \log(\overline{CF/K})_t \end{bmatrix} = \begin{bmatrix} \frac{(1-\alpha+(\alpha-\hat{\rho})L)}{(1-\alpha)(1-\hat{\rho}L} & 0 \\ \frac{1}{1-\hat{\rho}L} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \nu_t \end{bmatrix}. \tag{4.17}$$

From the theorem in Quah (1992) this representation implies that the Wold representation for this process is also lower triangular. Therefore, when the constraint is always binding, the observable cash flow-capital ratio does *not* Granger cause $\Delta\log(CF_t)$. The measurement error does not alter the polynomial matrix in the moving-average representation (4.13) for the unconstrained process. However, the vector error term becomes $(\mu_t,\,(\eta_t+\nu_t)\,)'$.

Thus, the test of financing constraints which emerges from this model of investment is an extremely simple Granger causality test. This can be implemented by estimating the first equation of the VAR representation of the $(\,\Delta\log(CF)_t\,, \log(\overline{CF/K})_t\,)'$ process, then testing the null hypothesis that the coefficients on the $\log(\overline{CF/K})_t$ are all zero. This test is structural insofar as the restriction being tested has been derived from an economic model of investment. Moreover, since the moving-average representation for the process generated by a constrained firm is nested in the representation for the process generated by an unconstrained firm, the resulting test of financing constraints is a direct one. In other words, if the data from a given sample fails to reject the null, then it is consistent with the hypothesis that the firms in this sample are financially constrained. This is an improvement on existing structural tests in the financing constraints literature. These papers only test

from observations on the book value. For a discussion of this in the context of measuring Tobin's q, see Perfect and Wiles (1994).

unconstrained investment models. Rejecting the null hypothesis in such framework implies only that the data is inconsistent with the unconstrained model. The source of the inconsistency, however, remains unidentified.

Several remarks concerning the test must be made. First, the null hypothesis may be rejected if equation (4.14) is not exact. If an error term is introduced into this equation, then it can be shown that $\log(CF/K)_t$ Granger causes $\Delta\log(CF)_t$. Thus, the null hypothesis would be rejected even for firms for which the financing constraint was binding in every period. The exactness of this equation depends upon the assumed functional forms for both the cost-of-adjustment function $\psi(i)$, and the financing constraint (equation (4.7)). This may make the identification of financially constrained firms more difficult. Second, the test is a sufficient condition for the identification of financially constrained firms. As such, rejection of the null hypothesis cannot be interpreted as evidence that financially constrained firms do not exist. Consequently, the test is of limited use in determining the pervasiveness of financing constraints in an economy.

4.4 Data and Empirical Results

The theory developed above places two main requirements on the data. First, the theory is applicable only to firms whose cash flow and capital stock series each have a unit root. Since this will not be true of all firms in a randomly selected sample, these firms must be identified. This was done by running Augmented Dickey-Fuller tests for each firm individually. The power of these tests depend crucially on the number of time series observations for each firm. For this reason, a firm was required to have a minimum of 28 consecutive observations to be included in the sample. With annual data, this requirement would likely overload the sample with unconstrained firms. Therefore, quarterly data was used. Second, the variables in the VAR are expressed in logarithms. Consequently, firms must not have negative observations on cash flow.

This section is divided into four parts. The first two deal with sample selection. The third section deals with specification tests. In the final section the Granger causality tests are run and their results discussed.

4.4.1 Selection Criteria for Samples

Compustat quarterly data for US manufacturing firms were used for this empirical analysis. To an initial population of 4659 firms, three selection criteria were applied. First, firms were required to have at least 28 consecutive observations with the initial observation on the capital stock greater than or equal to $2 million. The number of firms which met these requirements was 666. Second, firms with four or more negative observations on cash flow were deleted from this sample, leaving 509 firms. Finally, all remaining negative observations on cash flow were removed, and the 28 consecutive observation rule re-applied. This left a final sample of 375 firms with observations ranging from 1980:Q2 to 1992:Q1.

Data on the firms' cash flow and captial stock were used. Cash flow was measured as the firms' operating income before depreciation. The capital stock was measured as the book value of the firms' property, plant and equipment. The data was deflated by the US Producer Price Index. To mitigate the effects of seasonality, the first difference of the log of cash flow, and the log of the cash flow-capital ratio were regressed on a set of quarterly dummies for each firm individually. The residuals from these regressions were then used in the econometric analysis.

To identify sub-samples of firms which *a priori* are more likely to face financing constraints, three different criteria were used: size, existence of a bond rating, and existence of a commercial paper rating. These stratification criteria are standard in the empirical literature on financing constraints. For the size criterion, the value of the firms capital stocks in 1985:Q2 was used (this is the first quarter in which the entire cross-section of the panel is observed). The 25th percentile was chosen as the cutoff point between small and large firms. For the commercial paper rating, the criterion was the existence of either a Standard & Poor's or a Moody's rating in at least one time period over the window of observation for the firm. For the bond rating the criterion was the same, except that only Standard & Poor's ratings were available.

4.4.2 Augmented Dickey-Fuller Test Results

One of the implications of the theory developed above is that both the logarithms of cash flow and of the capital stock are non-stationary and co-integrated. Therefore, a necessary precusor to the financing constraints test is to identify the group of firms for which these implications hold.

To do so, Augmented Dickey-Fuller tests were conducted on the log of cash flow, the log of the capital stock, and the log of the cash flow-capital stock ratio for each firm individually.[10] The regression is,

$$\Delta y_t = \alpha + \gamma t + \beta_0 y_{t-1} + \sum_{k=1}^{p} \beta_k \Delta y_{t-k} + \epsilon_t.$$

In the initial specification, p was set equal to 8. The following criteria were used to choose the optimal value of p for each firm: (i) the final lag of the dependent variable was significant at the 5% level (2.048) and (ii), for this specification the existence of either first-order or fourth-order serial correlation in the residuals could be rejected at the 5% level. If the highest lag at which the dependent variable was found to be significant was less than four, and fourth-order serial correlation was detected for this specification, then the specification with $p = 4$ was chosen. If either first or fourth-order serial correlation was detected with eight lags of the dependent variable, that firm was removed from the sample. This removed a further 39 firms, leaving 336 in the sample.

The results of these regressions are summarised in Table 4.1. For cash flow and the capital stock, the diagonal elements show the number of firms for which the null

[10] The first differences of these data were deseasonalised as described above. The levels were formed by recursively adding the residuals.

hypothesis of a unit root failed to be rejected at the 1%, 5%, and 10% levels. The diagonal element for the cash flow-capital stock ratio show the number of firms for which the null hypothesis of a unit root *was rejected* at the 1%, 5%, and 10% levels. The off-diagonal elements show the number of firms satisfying both conditions. For example, for 113 firms the presence of a unit root failed to be rejected in the capital stock series, but was rejected in the cash flow-capital stock series at the 1% level.

Table 4.1 shows that a majority of firms in this sample have a unit root in either the cash flow or the capital stock series. This finding supports the intuition underlying the introduction of stochastic growth into the neo-classical Adjustment Costs model of investment. That is, it is possible that the presence of such firms in other samples have caused reduced-form tests to indicate erroneously that financing constraints exist for certain groups of firms.

The results for the cash flow-capital stock ratio show that the presence of a unit root in this series was not rejected for a large fraction of firms. For only 36 firms did the data reject the null for the cash flow-capital stock ratio series and fail to reject for both the cash flow and captial stock series. If the costs of adjusting the capital stock are high, or if the lags in responding to new information are significant, or if the transitory shock is quite serially persistent, then it may take years for the cash flow-capital stock ratio to return to its equilibrium following a shock. These findings may also be due to the way the variables are measured.

In order to increase the power of the Augmented Dickey-Fuller test for the cash flow-capital ratio, the individual firm t-statistics were aggregated and normalised following the procedures in Im, Pesaran, and Shin (1997). The asymptotic distribution of this test statistic is standard normal. For the group of 164 firms for which the null hypothesis failed to be rejected for both the cash flow and capital stock series, the test statistic had a value of -27.56.[11] Thus, the null hypothesis of a unit

[11]The statistic was computed using the values for the first and second moments of the test statistic provided in table 4 of Im, Pesaran, and Shin (1997). Since the specification of the ADF regressions was allowed to vary across firms, the values of the first and second moments were averaged over different lag lengths for the dependent variable.

root in the cash flow-capital stock series for this group of firms is clearly rejected.

4.4.3 VAR Specification

Based on the results of the Augmented Dickey-Fuller tests, the most appropriate sample for the Granger causality test is the group of 164 firms for which the presence of a unit root was rejected for neither the cash flow nor the capital stock series at the 1% level. At this stage, a panel data estimator was employed. The first equation of the VAR was estimated, namely,

$$\Delta log(CF)_{it} = \alpha_t + \phi_i + \sum_{l=1}^{m} \beta_l \Delta \log(CF)_{t-l} + \sum_{l=1}^{m} \gamma_l \log\left(\overline{\frac{CF}{K}}\right)_{t-l} + \epsilon_{it}, \qquad (4.18)$$

where α_t is a time dummy, and ϕ_i is a firm-specific effect. The Granger causality tests were performed by testing for the joint significance of the $\log(\overline{CF/K})$ terms using a Wald test.

Estimation was performed with the DPD package of Arellano and Bond (1988).[12] To eliminate the firm-specific effect, ϕ_i, equation (4.18) was first differenced. The levels of the regressors from lag 2 to $m+2$ were specified as instruments. Although this estimator is consistent, it is not efficient even amongst linear estimators. Given the relatively large number of observations per firm (ranging from 26 to 46), the efficient linear estimator proposed originally by Chamberlain (1983) and extended to the estimation of VARs by Holtz-Eakin, Newey, and Rosen (1988), requires a very large number of instruments. This was found to give very unstable estimates. The model was over-identified in order to permit the use of an additional specification test, a test on the overidentifying restrictions (referred to as the Sargan statistic (Arellano and Bond (1991)). For all sub-samples, this test statistic clearly rejected the validity of first differences of the regressors as instruments.

Table 4.2 summarises the specification tests for the various samples. The choice of

[12]The one-step GMM estimator was used. This estimator is not efficient because it uses an arbitrary weighting matrix. See Arellano and Bond (1991) for a discussion.

optimal lag length, m, was based on a Wald test of the joint significance of both $\Delta\log(CF)_{t-m}$ and $\log(\overline{CF/K})_{t-m}$. The initial value of m was 16. Even though the data was deseasonalised, some seasonal effects appear to remain in the data. Therefore, the optimal value for m was tested at multiples of 4. These results are presented in the first five columns of Table 4.2. Note that the optimal lag length differs greatly across samples. In particular, for large firms and for firms with bond ratings, the optimal value of m is very high. Indeed, for firms with bond ratings, the regressors are still jointly significant at $m = 32$ (regressions with higher values of m were not run).

With the exception of the group consisting of large firms, the choice of lag length was the first value of m for which the Wald statistic was insignificant at the 5% level. For large firms, $m = 32$ was the chosen specification. Given the optimal choice of lag length, three specification tests were run: (i) the $m1$ statistic which tests for first-order serial correlation, (ii) the $m2$ statistic which tests for second-order serial correlation, and (iii) the Sargan statistic. The asymptotic distribution of both the $m1$ and $m2$ statistics is standard normal (Arellano and Bond (1991)). The Sargan statistic is distributed χ^2, with degrees of freedom equal to the number of identification restrictions less the number of parameters to be estimated. The estimates reported in table 4.2 refer to the regressions with the optimally chosen value of m. Given that equation (4.18) is first differenced, the residuals from the regression should display negative first-order serial correlation and no second-order serial correlation. This is true for all samples. The Sargan statistic could only be computed for two of these regressions. In both cases its value indicates the validity of the instruments.[13]

[13]When there are more instruments than cross-section units DPD uses Generalised Inverses to compute the Sargan statistic. In all cases these were very unstable. Sargan statistics were computed for all lag lengths. In all cases where generalised inverses were not used the test statistic indicated that the instruments were valid.

4.4.4 Financing Constraints: The Granger Causality Test

If financing constraints bind in every period, the model presented above predicts that $\log(\overline{CF/K})$ will not Granger cause $\Delta \log(CF)$. This can be tested using a Wald test of the joint significance of the $\log(\overline{CF/K})$ terms in equation (4.18). The values of this test statistic are reported in table 4.3. Note that the degrees of freedom of the Wald test equals the value of m corresponding to the chosen VAR specification. However, the qualitative results of these Wald tests did not change when different lag lengths were used.

The results show that for two of the three groups of firms more likely to face binding financing constraints, namely small firms and firms with no bond rating, the $\log(\overline{CF/K})$ terms are jointly insignificant. Therefore, the null hypothesis that the financing constraint is binding in every period for these firms cannot be rejected. The $\Delta \log(CF)$ series is exogenous in the system for these firms. As a result, $\log(CF)$ will not admit a Blanchard-Quah decomposition. As argued above, this is because transitory shocks to cash flow are being used to finance investment, which will have a permanent effect on cash flow. Thus, the transitory shock, η, cannot be identified in the data for these groups of firms.

In constrast, for the entire sample and for the samples of large firms and firms with bond ratings, the $\log(\overline{CF/K})$ terms are jointly significant. The null hypothesis is clearly rejected for these firms. This does not necessarily mean that none of these firms are financially constrained. However, it is possible to perform a Blanchard-Quah decomposition on cash flow for firms in these samples, and thereby identify the transitory shock, η.

Unlike the bond rating criterion, the commercial paper rating criterion does not identify constrained firms. The null hypothesis is rejected for the samples of firms both with and without a commercial paper rating. In this sample, 24 firms that have a bond rating do not have a commercial paper rating, whereas only 6 firms which have a commercial paper rating do not have a bond rating. If the 24 firms with bond ratings are removed from the sample of firms without commercial paper ratings, the null hypothesis again fails to be rejected (the Wald test statistic of the joint significance of the $\log(\overline{CF/K})$ terms is 19.94 and $\chi^2_{crit}(16) = 26.3$).

4.5 Conclusions

The main criticism of the existing literature on financing constraints is that the most predominant testing procedure does not directly test a null hypothesis that financially constrained firms exist. As a result, all the evidence compiled in this literature is based on reduced-form regressions. A structural object which can be used to distinguish between constrained and unconstrained firms, the supply shock to a firm's internal finance, remains unidentified by this procedure. Consequently, it is not possible to ascertain the role this shock plays in determining a firm's investment behaviour.

This paper has addressed that criticism by constructing a model of investment from which a direct test of the existence of financially constrained firms can be derived. The model itself is able to explain a high investment-cash flow sensitivity without having to appeal to financing constraints. The test derived from this model is based upon the observation that when a firm is financially constrained, and the constraint is always binding, then the firm's cash flow series is an exogenous process. In contrast, the cash flow series is endogenous when the firm is unconstrained. These results imply that testing for the existence of financially constrained firms can be done via a Granger causality test. Thus not only is this a direct test of financing constraints, it is easily implemented as well.

The test was performed on a panel of U.S. manufacturing firms, sampled at a quarterly frequency between 1980 and 1992. The results show that in both the samples of small firms and of firms without bond ratings, the null hypothesis fails to be rejected. However, the null hypothesis is rejected for large firms and firms with a bond rating. Thus, it can be concluded that firms in the first two samples were financially constrained. This result is identical to the results reported in the existing literature.

However, because the results in this paper are based on a direct, structural test, they are not subject to the same criticisms.

Table 4.1: Presence of Unit Roots

	Cash Flow	Capital Stock	CF/K
	1% 206	-	-
Cash Flow	5% 163	-	-
	10% 134	-	-
	1% 164	1% 265	-
Capital Stock	5% 110	5% 235	-
	10% 89	10% 218	-
	1% 49	1% 113	1% 144
CF/K	5% 55	5% 137	5% 197
	10% 60	10% 148	10% 229

Cash Flow, Capital Stock: Number of firms which fail to reject the null.

Cash Flow/Capital Stock Ratio: Number of firms which reject the null.

Off diagonals: Number of firms satisfying both conditions.

Critical values (t=25): -4.38 (1%), -3.6 (5%), -3.24 (10%).

Table 4.2: VAR Specification Tests

Sample	Lag Length					m1	m2	Sargan	D.F.
	16	20	24	28	32				
Full	5.29	3.31	-	-	-	-6.1**	1.11	1.85	2219
Small	2.95	0.95	-	-	-	-4.39**	-1.19	-	576
Large	4.2	17.06**	12.43**	8.06*	1.37	-2.98**	-0.016	-	490
No Bond	4.71	0.32	-	-	-	-5.1**	-0.08	-	1422
Bond	21.78**	28.83**	15.74**	8.48*	7.58*	-4.42**	0.27	-	283
No CP	1.03	1.11	-	-	-	-5.22**	0.81	0.54	1072
CP	2.78	15.81**	0.87	-	-	-2.68**	0.021	-	581

** indicates significance at the 1% level, * at the 5% level.

Table 4.3: Granger Causality Tests

Sample	Firms	Wald (df)
Full	164	45.58** (20)
Small	40	21.85 (16)
Large	124	141.28** (32)
No Bond	86	22.73 (16)
Bond	78	204.1** (32)
No CP	104	49.54** (20)
CP	60	148.74** (24)

** indicates significance at the 1% level.

4.A Appendix

4.A.1 Proof of Proposition 4.2.1

To prove the existence of a solution to the Bellman equation as well as part (i) of Proposition 4.2.1, it must be shown that the following three conditions are satisfied:

1. The state space $\Re_{++} \times \Re_{++} \subset \Re^2$ is a convex cone with Borel subsets.

2. The correspondence $\Gamma : \Re_{++} \to \Re_{++}$ is nonempty, compact valued, and continuous. For any $K \in \Re_{++}, I \in \Gamma(K)$ implies $\kappa I \in \Gamma(\kappa K)$, all $\kappa \geq 0$. For some $d \in (0, \beta^{-1})$, $|I| \leq d|K|$, all $I \in \Gamma(K)$, all $K \in \Re_{++}$.

3. The reward function $R(I, K, \theta; \tilde{\eta}) := \tilde{\eta}\Pi(K, \theta) - I - \phi(i, K)$ is continuous. $R(\cdot, \cdot, \cdot)$ is homogeneous of degree 1. For some $0 < B < \infty$, $|R(I, K, \theta)| \leq B(|I| + |K| + |\theta|)$, all $(I, K, \theta) \in$ A, where A is the graph of Γ.

The first condition is satisfied by definition of the state space. Given the assumptions about the cost-of-adjustment function, the correspondence describing the set of feasible choices for I can be written as $\Gamma(K) = [-dK, dK]$ without loss of generality. Thus, the second condition follows under the assumption that d lies in the interval $(0, \beta^{-1})$. Finally, since the reward function is homogeneous of degree 1 in all three arguments, it can at most grow linearly, which implies that the boundedness condition in (iii) is met.

Given these assumptions, it is straightforward to show that there exists a unique solution to the Bellman equation (4.4) in the space of functions $f(x, y) : \Re_{++} \times \Re_{++} \to \Re$ that are linearly homogeneous, and that this solution equals the supremum function defined by equation (4.3). It can also be shown that the optimal policy correspondence associated with the solution to the Bellman equation is compact-valued, *u.h.c*, and homogeneous of degree 1, (i.e. $I \in G(K, \theta)$ implies that $\lambda I \in G(\lambda k, \lambda\theta)$, all $\lambda \geq 0$.)

To show part (ii), first note that the reward function is strictly increasing in K and θ. Let $H(X)$ represent the space of functions $f(x, y) : \Re_{++} \times \Re_{++} \to \Re$ that are linearly homogeneous. $H'(X) \subset H(X)$, the subspace of weakly increasing linearly homogeneous functions is closed. Therefore, by a corollary to the Contraction Mapping Theorem it suffices to show that the Bellman operator, T, implicitly defined by equation (4.4), maps elements of $H'(X)$ into $H''(X)$, the set of strictly increasing linearly homogeneous functions. Since the reward function is strictly increasing in K and θ, and the feasibility correspondence is monotone this follows straightforwardly.

The proof of part (iii) is similar to that of part (ii). The subspace of linearly homogeneous functions weakly concave in their first argument, $C'(X) \subset H(X)$ is closed. Since the reward function is strictly concave in K, it follows that T maps elements of $C'(X)$ into $C''(X)$, the set of linearly homogeneous functions which are strictly concave in their first argument.

To prove part (iv), note that the solution to the Bellman equation, $V(\cdot, \theta)$, is strictly concave in its first argument, and $\Gamma(K)$ is a convex set for all K. Thus, it follows that $G(\cdot, \theta)$ is a continuous (single-valued) function.

Finally, since the return function, R, is concave and continuously differentiable in K and θ, and since the assumptions on the adjustment costs function ensure that there is always an interior solution, it follows that the value function is differentiable with respect to K.

4.A.2 Proof of Proposition 4.2.2

The law of motion for k, $g(k, \mu)$, is given by,[14]

$$k' = \frac{(1+i)k}{\mu}. \tag{4.19}$$

[14]The law of motion is a function of next periods realization of the shock μ'. Since μ is i.i.d. the prime is dropped for notational convenience. Also for notational convenience, it is assumed that $\tilde{\mu} = \mu$.

Together with the density function for the exogenous shock, $\phi(\mu)$, this defines a transition function on the state space $\Re_{++} \times [\overline{\mu}, \underline{\mu}]$ with the Borel algebra. Call this transition function P.

The first part of this proof is to show there exists a unique set $S := [\underline{k}, \overline{k}] \subset \Re_{++}$ which is ergodic. As a first step, it can be shown that $g(k, \mu)$ is strictly increasing in k. Suppose that i_0 and k_0 satisfy the first order condition (3.14). Consider an increase in k to k_1. Suppose the firm decreases i so as to entirely offset the increase in k. In other words i_1 is set such that

$$\begin{aligned}
& & i_1 &= \frac{(1+i_0)k_0 - k_1}{k_1}, \\
&\Rightarrow & k_1 i_1 - k_0 i_0 &= k_0 - k_1, \\
&\Rightarrow & k_0 i_0 &> k_1 i_1, \\
&\Rightarrow & i_0 &> i_1, \\
&\Rightarrow & \psi'(i_0) &> \psi'(i_1).
\end{aligned} \tag{4.20}$$

This implies that i_1 is not optimal and that the optimal choice of investment, $i^*(k_1) < i_1$. A similar argument can be made for $k_1 < k_0$. Therefore, $g(k, \mu)$ is strictly increasing in k. Since $i(k)$ is strictly decreasing, $g(k, \mu)$ is strictly concave. To show $g_k(0, \mu) > 1$ note that

$$\begin{aligned}
g_k(0, \mu) &= \lim_{\Delta \downarrow 0} \left\{ \frac{1 + i(\Delta)}{\mu} \right\} \\
&= \frac{1+a}{\mu},
\end{aligned}$$

since

$$\begin{aligned}
\lim_{\Delta \downarrow 0} i(\Delta) &= \lim_{\Delta \downarrow 0} \psi_I^{-1} \left(\beta v' \left(\frac{(1+i)k)}{\mu} \right) - 1 \right) \\
&= \lim_{\Delta \downarrow 0} \psi_I^{-1} \left(\beta(\Pi_K(\Delta, \theta) - \Psi_K(I, \Delta) + \Psi_I(I, \Delta)) \right) \\
&= a,
\end{aligned}$$

where $\psi_I := \psi'(i)$. The first line follows from the first-order condition, and the second from the Euler equation. The third line follows from: (i) $\lim_{\Delta\downarrow 0}\Pi_K(\Delta,\theta) = \infty$, (ii) $\lim_{\Delta\downarrow 0}\Psi_K(I,\Delta) = -\infty$, (iii) $\lim_{\Delta\downarrow 0}\Psi_I(I,\Delta) = \infty$, and (iv) $\lim_{i\uparrow\infty}\psi'(i) = \alpha$. The restrictions on μ ensure that $g_k(0,\mu) > 1$.

This implies that for each $\mu \in [\underline{\mu},\overline{\mu}]$ there exists a unique fixed point $\hat{k}$ satisfying $\hat{k} = g(\hat{k},\mu)$. At that point $1+i(\hat{k}) = \mu$, and the slope $g_k(\hat{k},\mu) = 1+i'(k)k < 1$ since $i'(k) < 0$. Therefore each fixed point is stable.

Let $\underline{k}$ be the fixed point associated with $\overline{\mu}$, and $\overline{k}$ with $\underline{\mu}$. It is straightforward to show that for any $k \in S$, $k' \in S$. Moreover, the same is not true of any proper subset $[c,d] \subset S$. Therefore, S is ergodic. Since any $k < \underline{k}$ implies $k' > k$, and any $k > \overline{k}$ implies $k' < k$, S is unique.

That S is a unique ergodic set has two implications. First it implies that assumption 12.1 of Stokey, Lucas, and Prescott (1989) is satisfied for S. Second, it implies that if there exists a unique invariant measure, λ^* under $\hat{P}$, where $\hat{P}$ is the transition function P restricted to $S \times [\underline{\mu},\overline{\mu}]$, then λ^* is the unique invariant measure under P.

We will show that $\hat{P}$ is monotone and has the Feller property. It then follows from theorems 12.10 and 12.12 of Stokey, Lucas, and Prescott (1989) that λ^* exists and is unique.

First define $\hat{P}$ in terms of the density function ϕ.

$$\begin{aligned}
\hat{P}(q,(-\infty,x)) &= \Pr\{k' \leq x \mid k = q\}, \\
&= \Pr\left\{\frac{(1+i(q))q}{\mu} \leq x\right\}, \\
&= \Pr\left\{\mu \geq \frac{(1+i(q))q}{x}\right\}, \\
&= 1 - \Phi\left(\frac{(1+i(q))q}{x}\right),
\end{aligned}$$

where Φ is the distribution function for μ. Let $\vartheta(x)$ be the density function associated with $\hat{P}$. Then,

$$\vartheta(x) = \left(\frac{(1+i(q))q}{x^2}\right)\phi\left(\frac{(1+i(q))q}{x}\right). \tag{4.21}$$

It can be seen from the above equation that if $\phi(\cdot)$ is continuous then $\hat{P}$ has the Feller property. $\hat{P}$ is monotone if for any $b, c \in [\underline{k}, \overline{k}]$ with $b \geq c$, the probability measure $\hat{P}(b, \cdot)$ dominates $\hat{P}(c, \cdot)$, or equivalently,

$$\int h(x)\hat{P}(b, dx) \geq \int h(x)\hat{P}(c, dx), \tag{4.22}$$

for any bounded, increasing function $h(z)$. Rewriting the above equation using the density function ϕ gives,

$$\int_{\underline{b}}^{\overline{b}} h(x)\left(\frac{(1+i(b))b}{x^2}\right)\phi\left(\frac{(1+i(b))b}{x}\right)dx \geq \int_{\underline{c}}^{\overline{c}} h(x)\left(\frac{(1+i(c))c}{x^2}\right)\phi\left(\frac{(1+i(c))c}{x}\right)dx,$$

where $\underline{x} = (1+i(x))x/\overline{\mu}$ and $\overline{x} = (1+i(x))x/\underline{\mu}$. Defining $z = (1+i(q))q/x$ and changing variables in the integration allows the above equation to be rewritten as.

$$\int_{\underline{\mu}}^{\overline{\mu}} h\left(\frac{(1+i(b))b}{z}\right)\phi(z)dz \geq \int_{\underline{\mu}}^{\overline{\mu}} h\left(\frac{(1+i(c))c}{z}\right)\phi(z)dz.$$

The last inequality follows from the fact that $(1+i(b))b \geq (1+i(c))c$ and that $h(x)$ is increasing. Therefore, $\hat{P}$ is monotone. Note that because the bounds of integration have been switched, this cancels the negative sign arising from the change of variable.

□

4.A.3 Proof of Proposition 4.2.3

Since the only difference between the constrained and unconstrained problems is the feasibility correspondence, extending the proof of proposition 4.2.1 to the constrained problem relies on showing the feasibility correspondence defined by equation (4.7) satisfies the same conditions as the feasibility correspondence for the unconstrained problem.

To define the feasibility correspondence for the constrained problem, first let $\hat{F}(I, K) = I + \Psi(I, K)$. Since the function $\hat{F}$ is homogeneous of degree one let $f(I/K) = \hat{F}(I/K, 1)$. The function f is strictly convex with a global minimum at $b < 0$. Letting $m = f^{-1}$ over $[b, \infty)$ and $n = f^{-1}$ over $(-\infty, b)$, we can write the financing constraint as,

$$n(\eta_t\Pi(1,\theta_t/K_t))K_t \le I_t \le m(\eta_t\Pi(1,\theta_t/K_t))K_t. \tag{4.23}$$

Equation (4.23) defines the feasibility correspondence for the constrained problem, which can be written as,

$$\Gamma(K,\theta,\eta) = [\,n(\eta\Pi(1,\theta/K))K\,,\, m(\eta\Pi(1,\theta/K))K\,].$$

We need to show that $\Gamma(K,\theta,\eta)$ is

1. non-empty and compact valued,
2. continuous in K, θ, and η,
3. increasing in K and θ,
4. homogeneous of degree one in K and θ,
5. convex in K and θ.

Items (i) - (iv) are easily derived from the properties of $m(\cdot)$ and $n(\cdot)$. Note that since $n(x) < 0$ for all $x \in (-\infty, b)$ and $m(x) > 0$ for all $x \in [b,\infty)$ item (iii) implies that $n(\cdot)K$ is decreasing in K and θ, and that $m(\cdot)K$ is increasing in K and θ. To prove item (v) it is sufficient to show that $n(\cdot)K$ is convex and that $m(\cdot)K$ is concave in K and θ. The Hessian of $m(\cdot)K$ is given by

$$\begin{bmatrix} \frac{\partial^2}{\partial K^2} & \frac{\partial^2}{\partial K \partial\theta} \\ \frac{\partial^2}{\partial\theta\partial K} & \frac{\partial^2}{\partial\theta^2} \end{bmatrix} = \begin{bmatrix} \frac{A\theta^2}{K^3} & -\frac{A\theta}{K^2} \\ -\frac{A\theta}{K^2} & \frac{A}{K} \end{bmatrix},$$

where

$$A = \eta\left\{m''(\cdot)\Pi_2(\cdot,\cdot)^2\eta + m'(\cdot)\Pi_{22}(\cdot,\cdot)\right\}.$$

Over the interval $[b,\infty)$, $f' > 0$ and $f'' > 0$ which implies that $m' > 0$ and $m'' < 0$. Since $\Pi_{22} < 0$, A is negative and the Hessian is therefore negative semi-definite. Similar arguments can be made for the function $n(\cdot)K$. $\square$

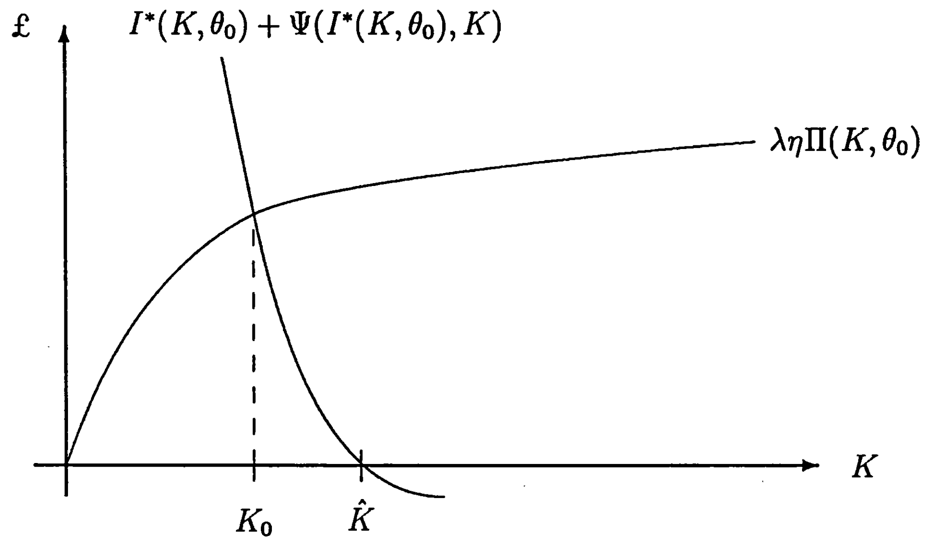

Figure 4.1: Existence of $\tilde{k}(\eta)$

4.A.4 Proof of Lemma 4.2.4

To prove the first part of the lemma, consider figure 4.1. For a given value of θ, θ_0, desired investment $I^\star$ is decreasing in K. There is a value of K, $\hat{K}$, such that $I^\star(\hat{K},\theta_0) = 0$. Since $\Psi(0,K) = 0$, $I^\star(\hat{K},\theta_0) + \Psi(I^\star(\hat{K},\theta_0),\hat{K}) = 0$. For all $K < \hat{K}$, $I^\star > 0$. Moreover, $\Psi(I^\star(K,\theta_0),K)$ is strictly decreasing and convex in K. Since $\Pi(K,\theta)$ is strictly increasing in K, and $\Pi(0,\theta_0) = 0$, there exists a value of K, K_0, such that $I^\star(K_0,\theta_0) + \Psi(I^\star(K_0,\theta_0),K_0) = \lambda\eta\Pi(K_0,\theta_0)$.

The functions $I^\star$, Π, and Ψ are homogeneous of degree one. Consider $\theta_1 = b\,\theta_0$, for any $b > 0$. We know there exists a unique $K_1 \in (0,\hat{K})$ such that $I^\star(K_1,\theta_1) + \Psi(I^\star(K_1,\theta_1),K_1) = \lambda\bar{\eta}\Pi(K_1,\theta_1)$. Due to the linear homogeneity of these functions this equation is satisfied for $K_1 = bK_0$. Therefore, for each $\eta \in [\underline{\eta},\overline{\eta}]$ there exists a unique k, $\tilde{k}(\eta) = K_0/\theta_0$, such that such that $\lambda\eta\pi(\tilde{k}(\eta))/\tilde{k}(\eta) = i(\tilde{k}(\eta)) + \psi(i(\tilde{k}(\eta)))$, and for $k < \tilde{k}(\eta)$, $\lambda\eta\pi(k)/k < i(k) + \psi(i(k))$.

The function $\tilde{k}(\eta)$ is decreasing in η. Its lowest value will therefore be $\tilde{k}(\overline{\eta})$. The financing constraint will always be binding for $k < \tilde{k}(\overline{\eta})$. We now want to show that there exists a λ such that the upper support of the distribution of k in the constrained problem, $\overline{k}_c$ is less than $k < \tilde{k}(\overline{\eta})$. To do so, we assume that the lower bound of the support for the distribution of μ, $\underline{\mu} > 1$. This is an assumption which has be made in the proof of Proposition 4.2.5, the stationarity of the constrained problem. This implies that $\theta_{t+1} > \theta_t$, $\forall t$. Consequently, the firm's demand for new capital will always be positive. So these firms will always be growing. As was discussed earlier in the chapter, growing firms are of particular interest in this analysis. First, growing firms may appear to be constrained because the presence of a non-stationary shock may lead to a rejection of some of the predictions of the standard formulation of the Adjustment Costs model. Second, in those studies that report growth rates, it is usually found that firms classified *a priori* as constrained have a higher average growth rate.

When the firm is constrained, and $\eta = \overline{\eta}$, the investment function is implicitly defined by the equation,

$$\lambda\overline{\eta}\pi(k)/k = i + \psi(i). \tag{4.24}$$

Now consider the fixed point of the law of motion for k corresponding to $\underline{\mu}$,

$$k_{t+1} = \frac{(1+i_t)k_t}{\underline{\mu}}.$$

Since this fixed point corresponds to the lower bound of the support of μ, it will be the upper bound of the support for k, which will be called $\overline{k}_c$. At $\overline{k}_c$, $1 + i_t = \underline{\mu}$. Therefore, we can write equation (4.24) as

$$\lambda\overline{\eta}\pi(k)/k = \underline{\mu} - 1 + \psi(\underline{\mu} - 1).$$

The left-hand side is decreasing in both k and λ, and $\lim_{k\to 0}\pi(k)/k = \infty$. If we decrease λ, k must be increased to satisfy this equation. Therefore, for a given $\underline{\mu}$, there exists a $\lambda \in (0,1)$ such that the k which satisfies the above equation, $\overline{k}_c < \tilde{k}(\overline{\eta})$.

4.A.5 Proof of Proposition 4.2.5

The proof of this Proposition follows the same argument as that of the proof of Proposition 4.2.2. Therefore, we need to show that

1. The law of motion for k' is increasing and strictly concave.

2. The derivative of the law of motion at $k = 0$ is greater than 1.

3. The transition function for the constrained problem is monotone.

Since we are focusing on the case where the firm is growing the law of motion is

$$k' = \frac{(1 + m(\lambda \frac{\bar{\eta}\pi(k)}{k}))k}{\mu'},$$

where $m(\cdot)$ is the function which defines the upper bound of the feasibility correspondence for investment in the constrained problem. The first and second derivatives are given by the expressions,

$$\frac{\partial}{\partial k}k' = \frac{1}{\mu'}\left\{1 + m'(\cdot)\left(\frac{\pi'(k)k - \pi(k)}{k}\right) + m\left(\frac{\eta\pi(k)}{k}\right)\right\},$$

and

$$\frac{\partial^2}{\partial k^2}k' = \frac{1}{\mu'}\left\{m''(\cdot)\left(\frac{\pi'(k)k - \pi(k)}{k}\right)^2 k + m\left(\frac{\eta\pi(k)}{k}\right)\pi''(k)\right\}.$$

Since $m''(\cdot) < 0$ and $\pi''(k) < 0$ the second derivative is negative. Taking limits of the first equation gives,

$$\lim_{k\downarrow 0}\frac{\partial}{\partial k}k' = \frac{1+a}{\mu'},$$

and

$$\lim_{k\uparrow\infty}\frac{\partial}{\partial k}k' = \frac{1}{\mu'}.$$

Since the first derivative is continuous and decreasing, it must be positive everywhere. Note that a stationary solution to this problem requires that $[\underline{\mu}, \overline{\mu}] \in (1, 1+a)$.

These results imply that there is a unique ergodic set for k in the constrained problem. Each element of that set can be represented as a fixed point of the law of motion corresponding to a particular value of μ. Define $\hat{S} := [\underline{k_c}, \overline{k_c}]$, where $\underline{k_c}$ is the fixed point corresponding to $\overline{\mu}$ and $\overline{k_c}$ is the fixed point corresponding to $\underline{\mu}$.

Let $\hat{P}_c$ be the transition function defined on $\hat{S}_c \times [\underline{\mu}, \overline{\mu}]$. To show that $\hat{P}_c$ is monotone, initially fix $\eta = \hat{\eta}$. Then,

$$\begin{aligned} \hat{P}_c((a, \hat{\eta}), (-\infty, x)) &= Pr\{k' \leq x \mid k = a, \eta = \hat{\eta}\}, \\ &= 1 - \Phi\left(\frac{(1 + g(\frac{\hat{\eta}\pi(a)}{a}))a}{x}\right). \end{aligned}$$

Using the same arguments as in the proof for the unconstrained problem, it can be shown that for each $\eta \in [\underline{\eta}, \overline{\eta}]$, $\hat{P}_c((a, \hat{\eta}), \cdot)$ dominates $\hat{P}_c((b, \hat{\eta}), \cdot)$, for any $a, b \in \hat{S}_c$ such that $a \geq b$. Since this is true for every η it follows that $\hat{P}_c(a, \cdot)$ dominates $\hat{P}_c(b, \cdot)$, for any $a, b \in \hat{S}_c$ such that $a \geq b$, (i.e. the transition function *not* conditioned on η). $\square$

Chapter 5

Conclusions

There is a large empirical literature on financing constraints and investment. The papers in this literature have performed several different tests of financing constraints on many different datasets. The impact of financing constraints has been examined not only on physical capital investment, but also on investment in inventories, working capital, and R&D. With few exceptions, the studies in this literature have found very similar results which have been interpreted as evidence that there exist some firms which are financially constrained.

Chapter 2 is an application of several of these tests to a sample of UK firms. The sample was stratified into innovative and non-innovative groups, and several investment regressions were run. The data for both groups of firms rejected the restrictions implied by the structural equations from the Adjustment Costs model. However, cash flow was found to be a more important determinant of the investment behaviour of innovative firms. Moreover, this result was found in all of the investment regressions that were run. In keeping with the existing literature, this result was interpreted as evidence that innovative firms are financially constrained.

In the Introduction some of the weaknesses inherent in the existing tests of the financing constraints literature are discussed. Existing structural tests are based on the necessary conditions of the Adjustment Costs model of physical capital investment. These conditions are derived under several assumptions, one of which is the assumption of perfect capital markets. Consequently, this methodology tests a joint hypothesis which is comprised of all of these assumptions. The main weakness of this methodology is that when the data rejects the joint hypothesis, as, for example, was found in Chapter 2, it is not possible to ascertain which of the assumptions has been violated. Moreover, once the structural equations have been rejected, they can no longer be given a structural interpretation. Therefore, one is left only with reduced-form evidence to support the argument that the source of rejection is the violation of the perfect capital markets assumption. However, inference about the validity of an economic model cannot be based on a comparison of reduced-form coefficients, as has been commonly done in the financing constraints literature. Such a testing methodology is open to several strong criticisms, one of which can be understood as a version of the Lucas critique.

In Chapter 4 a new test of financing constraints which is not subject to these criticisms is developed and implemented. In that chapter, a model of physical capital investment is built by introducing several modifications to the Adjustment Costs model. From this model, a vector moving average representation for two observable variables was derived. It was shown that, when a financing constraint is imposed on this model which is assumed to bind in every period, this places a testable restriction on the moving average representation. As this restriction was explicitly derived from an economic model, the test is structural and therefore is free from the criticisms made of reduced-form tests. Moreover, the test is direct insofar as one of the assumptions in its null hypothesis is that capital markets are imperfect. Therefore, this test is also free from the criticism made of existing structural tests. Finally, the test takes the form of a simple Granger causality test, and thus is easily implemented.

The new test was run on a panel of U.S. manufacturing firms. The sample was stratified using three different criteria: firm size, existence of a bond rating, and existence of a commercial paper rating. The results show that the null hypothesis that firms are financially constrained failed to be rejected for both the sample of small firms, and the sample of firms without bond ratings. For the sample of large firms, and of firms with bond ratings, the null hypothesis was decisively rejected. These are precisely the same results that have been reported in many papers in the financing constraints literature. That these results have been replicated using a test that is both structural and direct greatly strenthens their original interpretation as evidence of financing constraints. For both firms with and without a commercial paper rating, the null hypothesis was rejected. However, 28 firms which did not have a commercial paper rating did have a bond rating. When they were excluded from the sample of firms without commercial paper ratings, the null hypothesis failed to be rejected for this group.

Chapter 3 modeled the impact of financing constraints on inventory investment. The aim of that chapter was to use financing constraints to explain the excess variance of production puzzle. This is the widely observed fact that the variance of production exceeds that of sales. The intuition behind the general model was illustrated with an example. In that example, the financing constraint is binding only after low sales realizations. This forces the firm to produce less than what it has sold and what has depreciated from its stock of unsold goods. After higher sales realizations, the financing constraint is not binding. This permits the firm to produce more than what has been sold and has depreciated. Thus, in this example and in the general model, underproduction is associated with low sales realisations, and overproduction is associated with medium and high sales realisations. In this way the financing constraint causes the production distribution to be a mean-preserving spread of the sales distribution, thereby explaining the excess variance of production puzzle.

In addition to providing a theoretical explanation of this puzzle, the model of Chapter 3 delivers several other predictions. One is that inventory investment and sales are positively correlated. Another is that the sales process is positively serially correlated despite the fact that the underlying demand process is not. Both of these predictions are observed in data. Moreover, the second prediction suggests a way in which financing constraints may act as a propagation mechanism for aggregate output fluctuations. Empirically, inventory disinvestment accounts for a very large proportion of declines in the aggregate output of both the US and UK during recessions.

There are several other explanations of the excess variance of production puzzle. The puzzle has such widespread empirical support that it is difficult to imagine that it can be explained exclusively by one of the competing theories. A prediction of the model in Chapter 3 that distinguishes it from the other explanations is that sales which do not affect the firm's expectation of future demand nevertheless affect its inventory investment. This is a prediction which is common to models of investment under financing constraints. The identification of such sales is precisely the same problem that lies at the center of the controversy surrounding existing tests in the empirical financing constraints literature. A potential avenue of future research is the use of the methodology developed in Chapter 4 to test the inventory investment model developed in the third chapter.

Bibliography

ABEL, A. (1980): "Empirical Investment Equations: An Integrative Framework," in *On the State of Macroeconomics*, vol. 12 of *Journal of Monetary Economics Supplement*, pp. 39–91.

ABEL, A. (1985): "Inventories, Stock-Outs and Production Smoothing," *Review of Economic Studies*, LII, 283–293.

ABEL, A., and O. BLANCHARD (1986): "The Present Value of Profits and Cyclical Movements in Investment," *Econometrica*, 54, 249–273.

ANDERSON, T., and C. HSIAO (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

ARELLANO, M., and S. BOND (1988): "Dynamic Panel Data Estimation using DPD - A Guide for Users," Institute for Fiscal Studies Working Paper no. 88/15.

ARELLANO, M., and S. 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.

BERNANKE, B., and M. GERTLER (1989): "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, 79, 14–31.

BHATTACHARYA, S., and J. R. RITTER (1983): "Innovation and Communication: Signalling with Partial Disclosure," *Review of Economic Studies*, 98, 331–346.

BILS, M., and J. KAHN (1996): "What Inventory Behaviour Tells Us about Business Cycles," mimeo.

BLANCHARD, O., and S. FISCHER (1989): *Lectures on Macroeconomics.* MIT Press, Cambridge, Massachusetts.

BLANCHARD, O., F. LOPEZ-DE SILANES, and A. SCHLEIFER (1994): "What Do Firms Do with Cash Windfalls?," *Journal of Financial Economics.*

BLANCHARD, O., C. RHEE, and L. SUMMERS (1990): "The Stock Market, Profit and Investment," *NBER Working Paper.*

BLANCHARD, O. J. (1983): "The Production and Inventory Behaviour of the American Automobile Industry," *Journal of Political Economy*, 91, 365–400.

BLANCHARD, O. J., and D. QUAH (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79, 655–673.

BLINDER, A. S. (1981): "Real Inventory Behaviour and Business Cycles," *Brookings Papers on Economic Activity*, 2, 443–505.

BLINDER, A. S. (1986): "Can the Production Smoothing Model of Inventory Behaviour be Saved?," *Quarterly Journal of Economics*, CI, 431–453.

BLINDER, A. S., and L. J. MACCINI (1991): "Taking Stock: A Critical Assessment of Recent Research on Inventories," *Journal of Economic Perspectives*, 5, 73–96.

BLUNDELL, R., S. BOND, M. DEVEREUX, and F. SCHIANTARELLI (1992): "Investment and Tobin's Q: Evidence from Company Panel Data," *Journal of Econometrics*, 51, 233–257.

BLUNDELL, R., R. GRIFFITH, and J. VAN REENEN (1993): "Knowledge Stocks, Persistent Innovation and Market Dominance: Evidence from a Panel of British Manufacturing Firms," IFS Working Paper, London.

BOND, S., and C. MEGHIR (1994): "Dynamic Investment Models and the Firm's Financial Policy," *Review of Economic Studies*, 61, 197–222.

BRESNAHAN, T., and V. RAMEY (1994): "Output Fluctuations at the Plant Level," *Quarterly Journal of Economics*, 109, 593–624.

CABALLERO, R. J., and J. V. LEAHY (1996): "Fixed Costs: The Demise of Marginal q," *NBER Working Paper*, 5508.

CALOMIRIS, C., and R. HUBBARD (1990): "Firm Heterogeneity, Internal Finance and Credit Rationing," *Economic Journal*, 100, 90–104.

CAPLIN, A. (1985): "The Variability of Aggregate Demand with (s,S) Inventory Policies," *Econometrica*, 53, 1395–1409.

CARPENTER, R., S. FAZZARI, and B. PETERSEN (1994): "Inventory Investment, Internal-Finance Fluctuations, and the Business Cycle," *Brookings Papers on Economic Activity*, pp. 75–138.

CHAMBERLAIN, G. (1983): "Panel Data," in *Handbook of Econometrics*, ed. by Z. Gilliches, and M. Intriligator, vol. 2. Amsterdam: Elsevier Science Publishers.

CHIRINKO, R. S. (1993): "Business Fixed Investment Spending: A Critical Survey of Modelling Strategies, Empirical Results, and Policy Implications," *Journal of Economic Literature*, 31, 1875–1911.

COCHRANE, J. H. (1997): "Time Series for Macroeconomics and Finance," mimeo.

COGLEY, T., and J. NASON (1995): "Output Dynamics in Real-Business-Cycle Models," *American Economic Review*, 85, 492–511.

CUMMINS, J. G., K. A. HASSETT, and S. D. OLINER (1997): "Investment Behaviour, Observable Expectations, and Internal Funds," mimeo, New York University.

EICHENBAUM, M. S. (1989): "Some Empirical Evidence on the Production Level and Production Cost Smoothing Models of Inventory Investment," *American Economic Review*, 79, 853–864.

EISNER, R., and R. STROTZ (1963): "Determinants of Business Investment," in *Impacts of Monetary Policy*, pp. 60–337, Englewood Cliffs:N.J. Prentice-Hall.

FAIR, R. C. (1989): "The Production Smoothing Model is Alive and Well," *Journal of Monetary Economics*, 24, 353–370.

FAZZARI, S., G. HUBBARD, and B. PETERSEN (1988): "Financing Constraints and Corporate Investment," *Brookings Papers on Economic Activity*, 1, 141–95.

FAZZARI, S., and B. PETERSEN (1993): "Working Capital and Fixed Investment: New Evidence on Financing Constraints," *RAND Journal of Economics*, 24, 328–41.

FREIXAS, X., and J.-C. ROCHET (1997): *Microeconomics of Banking.* MIT Press, Cambridge, Massachusetts.

FROOT, K. A., and J. C. STEIN (1991): "Exchange Rates and Foreign Direct Investment: An Imperfect Capital Markets Approach," *Quarterly Journal of Economics*, 106, 1191–1217.

GERTLER, M. (1992): "Financial Capacity and Output Fluctuations in an Economy with Multiperiod Financial Relationships," *Review of Economic Studies*, 59, 455–72.

GERTLER, M., and S. GILCHRIST (1994): "Monetary Policy, Business Cycles, and the Behaviour of Small Manufacturing Firms," *Quarterly Journal of Economics*, 95, 309–40.

GILCHRIST, S., and C. P. HIMMELBERG (1995): "Evidence on the Role of Cash Flow for Investment," *Journal of Monetary Economics*, 36, 541–572.

GILCHRIST, S., and C. P. HIMMELBERG (1998): "Investment, Fundamentals, and Finance," mimeo, Columbia University.

GOULD, J. P. (1968): "Adjustment Costs in the Theory of Investment of the Firm," *Review of Economic Studies*, 35, 47–55.

GREENWALD, B., and J. STIGLITZ (1993): "Financial Market Imperfections and Business Cycles," *Quarterly Journal of Economics*, 108, 77–114.

GUARIGLA, A. (1996): "The Effects of Financial Constraints on Inventory Investment: Evidence form a Panel of UK Firms," Discussion Paper, University of Essex.

GUARIGLIA, A., and F. SCHIANTARELLI (1995): "Production Smoothing, Firm Heterogeneity and Financial Constraints: Evidence from a Panel of UK Firms," mimeo, University of Essex.

HALL, B. (1992): "Investment and R&D at the Firm Level: Does the Source of Financing Matter?," Department of Economics Working Paper 92-194.

HALL, G. (1996): "Non-convex Costs Capital Utilization: A Study of Production and Inventories at Automobile Plants," mimeo.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica*, 50, 1029–54.

HART, O., and J. MOORE (1994): "A Theory of Debt based on the Inalienability of Human Capital," *Quarterly Journal of Economics*, 109, 841–879.

HAUSMAN, J. A., and D. A. WISE (1979): "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," *Econometrica*, 47, 455–473.

HAYASHI, F. (1982): "Tobin's Average Q and Marginal Q: A Neoclassical Interpretation," *Econometrica*, 50, 213–224.

HAYASHI, F., and T. INOUE (1991): "The Relation Between Firm Growth and a With Multiple Capital Goods: Theory and Evidence from Panel Data on Japanese Firms," *Econometrica*, 59, 731–53.

HIMMELBERG, C., and B. PETERSEN (1994): "R&D and Internal Finance," *Review of Economics and Statistics*, pp. 38–50.

HOLT, C. C., F. MODIGLIANI, J. F. MUTH, and H. SIMON (1960): *Planning Production, Inventories and Work Force*. Englewood Cliffs: Prentice-Hall.

HOLTZ-EAKIN, D., W. NEWEY, and H. S. ROSEN (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371–1395.

HORSTMANN, I., G. M. MACDONALD, and A. SLIVINSKI (1985): "Patents as Information Transfer Mechanisms: To Patent or (Maybe) Not to Patent," *Journal of Political Economy*, 93, 837–858.

HUBBARD, G. R. (1997): "Capital market Imperfections and Investment," NBER Working Paper 5996, Cambridge.

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.

IM, K. S., M. H. PESARAN, and Y. SHIN (1997): "Testing for Unit Roots in Heterogeneous Panels," mimeo.

JENSEN, M. C. (1986): "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers," *American Economic Review*, 76, 323–329.

KAHN, J. A. (1987): "Inventories and the Volatility of Production," *American Economic Review*, 77, 667–679.

KAHN, J. A. (1992): "Why is Production more Volatile than Sales? Theory and Evidence on the Stockout-Avoidance Motive for Inventory Holding," *Quarterly Journal of Economics*, 107, 481–510.

KAMIEN, M. I., and N. L. SCHWARTZ (1982): *Market Structure and Innovation.* Cambridge University Press, Cambridge.

KAPLAN, S. N., and L. ZINGALES (1997): "Do Financing Constraints Explain Why Investment Is Correlated With Cash Flow?," *Quarterly Journal of Economics*, pp. 169–215.

KASHYAP, A. K., J. STEIN, and D. W. WILCOX (1993): "Monetary Policy and Credit Conditions: Evidence from the Composition of External Finance," *American Economic Review*, 83, 78–98.

KASHYAP, A.K. LAMONT, O., and J. STEIN (1994): "Credit Conditions and the Cyclical Behaviour of Inventories," *Quarterly Journal of Economics*, 109, 565–591.

KIYOTAKI, N., and J. MOORE (1995): "Credit Cycles," Discussion Paper.

KRANE, S., and S. BRAUN (1991): "Production Smoothing Evidence from Physical Product Data," *Journal of Political Economy*, 99, 558–81.

LACH, S., and M. SCHANKERMAN (1989): "Dynamics of R&D and Investment in the Scientific Sector," *Journal of Political Economy*, 97, 880–904.

LAMONT, O. (1997): "Cash Flow and Investment: Evidence from Internal Capital Markets," *Journal of Finance*, 52, 83–110.

LEVIN, R. C., A. K. KLEVORICK, R. R. NELSON, and S. G. WINTER (1987): "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity*, 3, 783–831.

LUCAS, R. (1967): "Optimal Investment Policy and the Flexible Accelerator," *International Economic Review*, 8, 78–85.

LUCAS, R., and E. PRESCOTT (1971): "Investment under Uncertainty," *Econometrica*, 39, 659–681.

METZLER, L. A. (1941): "The Nature and Stability of Inventory Cycles," *Review of Economic Statistics*, 23, 113–129.

MEYER, J. R., and E. KUH (1957): *The Investment Decision.* Harvard University Press, Cambridge.

MIRON, J. A., and S. P. ZELDES (1988): "Seasonality, Cost Shocks and the Production Smoothing Model of Inventories," *Econometrica*, 56, 877–908.

MOSSER, P. (1990): "Trade Inventories and (S,s)," *Quarterly Journal of Economics*, pp. 1267–1286.

MYERS, S., and N. MAJLUF (1984): "Corporate Financing and Investment Decisions When Firms Have Information that Investors Do Not Have," *Journal of Financial Economics*, 13, 187–221.

NICKELL, S. (1978): *The Investment Decisions of Firms.* Cambridge University Press, London.

PAKES, A. (1985): "On Patents, R&D, and the Stock Market Rate of Return," *Journal of Political Economy*, 93, 390–409.

PAKES, A., and M. SCHANKERMAN (1984): "An Exploration into the Determinants of Research Intensity," in *R&D, Patents, and Productivity*, ed. by Z. Grilliches, Chicago. Univ. of Chicago Press (for NBER).

PERFECT, S. B., and K. W. WILES (1994): "Alternative Constructions of Tobin's q: An Empirical Comparison," *Journal of Empirical Finance*, 1, 313–341.

QUAH, D. (1992): "The Relative Importance of Permanent and Transitory Components: Identification and some Theoretical Bounds," *Econometrica*, 60(1), 107–118.

RAMEY, V. A. (1991): "Nonconvex Costs and the Behaviour of Inventories," *Journal of Political Economy*, 99, 306–334.

RAMEY, V. A., and K. D. WEST (1998): "Inventories," in *Handbook of Macroeconomics.* forthcoming.

SARGENT, T. J. (1987): *Macroeconomic Theory.* Academic Press, San Diego.

SCHIANTARELLI, F. (1996): "Financial Constraints and Investment: Methodological Issues and International Evidence," *Oxford Review of Economic Policy*, 12, 70–89.

SCHIANTARELLI, F., and D. GEORGOUTSOS (1990): "Monopolistic Competition and the Q Theory of Investment," *European Economic Review*, 34, 1061–1078.

SENSIER, M. (1996): "Inventories and Asymmetric Business Fluctuations," mimeo.

STIGLITZ, J. E., and A. WEISS (1981): "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, 71, 393–410.

STOKEY, N. L., R. E. LUCAS, and E. C. PRESCOTT (1989): *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge.

STONEMAN, P. (1995): "Introduction," in *Handbook of Economics of Innovation and Technological Change*, ed. by P. Stoneman, pp. 1–13, Oxford. Blackwell.

VERBEEK, M., and T. NIJMAN (1992a): "Incomplete Panels and Selection Bias," in *The Econometrics of Panel Data: Handbook of Theory and Applications*, ed. by L. Matyas, and P. Sevestre, pp. 262–302, Dordrecht. Kluwer Academic Publishers.

VERBEEK, M., and T. NIJMAN (1992b): "Testing for Selectivity Bias in Panel Data Models," *International Economic Review*, 33, 681–703.

WEST, K. D. (1986): "A Variance Bounds Test of the Linerar Quadratic Inventory Model," *Journal of Political Economy*, 94, 374–401.

WHITED, T. M. (1992): "Debt, Liquidity Constraints, and Corporate Investment: Evidence from Panel Data," *Journal of Finance*, 47, 1425–60.