The Engineer in Business

Lecture 1

The theory of the firm

Notes

Michael G. Pollitt

So, you want to be in an engineer in business, but what do businesses do?

Introduction

We set out to answer three questions:

1. Why do firms exist (what do they do)?
2. What determines their size and features?
3. What determines the boundary between *the firm* and *the market*?

Think of a firm as a production-distribution unit in the economy.

These are all central questions which theories of the firm attempt to answer in economics. They are important in trying to explain why we observe the forms of economic organization that we see in practice and related questions such as: why are some firms large and some small, why might the relative mix of different sizes of firms be changing over time, what determines the mix of economic activities collected together within one firm, why might firms outsource more or less etc…

What we will emphasise is the huge variety of firms that we observe in practice and the variety of theoretical explanations which attempt to answer some of or all of the questions above.

Some facts about firms in the UK

The population of firms in the UK

Population of UK Businesses

Table 5 (UK sections): Number of businesses in the private sector and their associated employment and turnover [Note 6], by number of employees and industry section, UK, start 2023

This worksheet contains 1 table. Some cells refer to Notes which can be found in the Notes worksheet.
Cells in rows have been left blank where the cell at the start of the row indicates the industrial section to which data in the rows following refer.
Freeze panes are turned on. To turn off, select the 'View' ribbon then 'Freeze panes, then 'Unfreeze panes' or use [Alt,W,F]

Employee size band	Business number [Note 5]	Employment number *(thousands)* [Note 7]	Turnover *(£ millions)* [Note 1][Note 3] [Note 8]	*Businesses percentage*	*Employment percentage*	*Turnover percentage [Note 1][Note 3]*
All Industries						
All businesses	5,555,130	27,524	4,479,552	*100.0*	*100.0*	*100.0*
All employers	1,444,985	23,039	4,148,763	*26.0*	*83.7*	*92.6*
With no employees (unregistered) [Note 2]	2,914,125	3,204	125,547	*52.5*	*11.6*	*2.8*
With no employees (registered) [Note 2] [Note 4]	1,196,020	1,281	205,241	*21.5*	*4.7*	*4.6*
1	131,040	290	27,369	*2.4*	*1.1*	*0.6*
2 to 4	773,070	2,156	303,836	*13.9*	*7.8*	*6.8*
5 to 9	273,225	1,842	260,927	*4.9*	*6.7*	*5.8*
10 to 19	144,320	1,975	257,626	*2.6*	*7.2*	*5.8*
20 to 49	78,465	2,371	411,927	*1.4*	*8.6*	*9.2*
50 to 99	24,455	1,692	340,791	*0.4*	*6.1*	*7.6*
100 to 199	10,340	1,433	294,365	*0.2*	*5.2*	*6.6*
200 to 249	2,110	471	127,792	*0.0*	*1.7*	*2.9*
250 to 499	4,060	1,400	300,418	*0.1*	*5.1*	*6.7*
500 or more	3,900	9,409	1,823,714	*0.1*	*34.2*	*40.7*

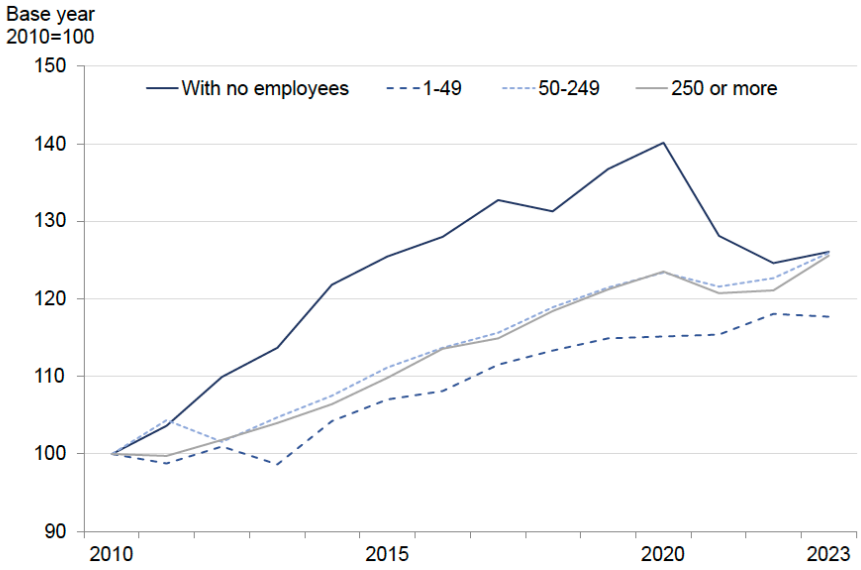
Source: Department Business and Trade
BUSINESS POPULATION ESTIMATES FOR THE UK AND THE REGIONS 2023, Statistical Release October 2023.

UNIVERSITY OF CAMBRIDGE Judge Business School

Note that most firms are actually self-employed one person entities. In terms of employment in ‘private’ businesses most people work in firms with less than 250 employees. Large firms with 250 or more employees are only around 8000, but they employ almost 40% of all ‘business’ employment and produce around half of all turnover. So large firms disproportionately contribute to the economy.

The figure below shows that the number of all firms is growing, but self-employment particularly so. This indicates a move towards smaller firms in the economy in terms of employment and income.

Number of Private Firms in the UK by size class (2010-2023)

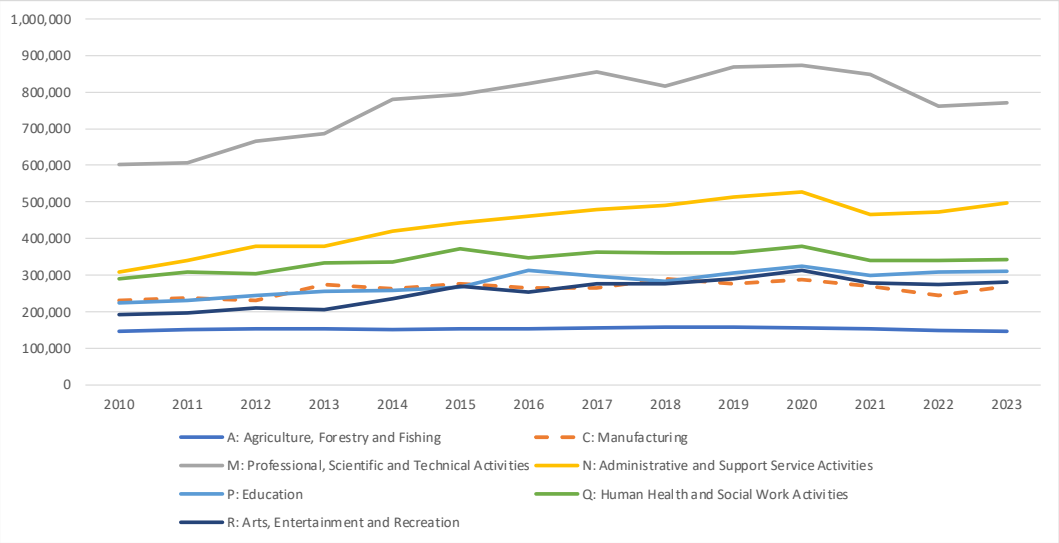


Source: Department Business and Trade
BUSINESS POPULATION ESTIMATES FOR THE UK AND THE REGIONS 2023, Statistical Release October 2023.

UNIVERSITY OF CAMBRIDGE
Judge Business School

The sectors that these firms are in is also substantially shifting. The UK is a service economy and this is reflected in the massive growth of the number of firms in service sectors. In the figure below manufacturing firms have grown relatively slowly in recent years compared to a range of service sectors.

Number of businesses by industry in the UK



Source: Dept Business and Trade
Business Population Estimates 2023, Table 29.

7

UNIVERSITY OF CAMBRIDGE
Judge Business School

There is a large turnover of firms each year, among the smaller firms. For a new business, the chance of still being financially viable is c.42% after 5 years, as shown below. Note this only true of firms that are being measured by the government (usually those registered for VAT payments, with turnover above £85k p.a.), true failure rates might be substantially higher than this.

Dimension: Age of the firm

Source: https://www.merchantsavvy.co.uk/uk-sme-data-stats-charts/

UNIVERSITY OF CAMBRIDGE
Judge Business School

Theories of the firm

A firm is an organisation consisting of one or more individuals working as a decision-making unit to produce goods or services. Firms can take a variety of legal forms. For example, unincorporated businesses such as sole proprietors e.g. hairdressers, corner shops; partnerships e.g. lawyers, accountants. They can be incorporated businesses: private limited company (e.g. Virgin); or public (as in with stock market traded shares) limited company (e.g. ITV, Shell). Public sector firms exist such as the NHS (or indeed individual hospital trusts or individual schools). These are firms in the sense that they are production-distribution units. Charities are also firms (e.g. a Cambridge college) in this sense.

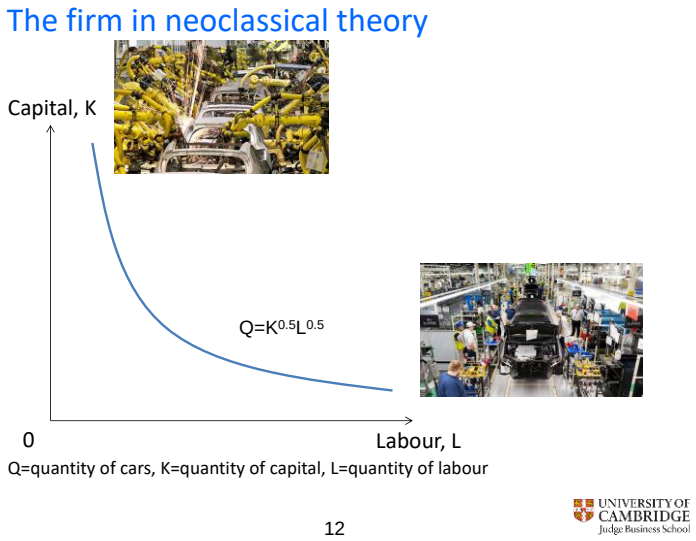
The alternative to the firm is the market.

'The market is an institution for resource allocation which is based on horizontal voluntary exchanges by individual economic agents, who are motivated by preferences and price signals.' Adapted from Pitelis (1999).

Note that markets are about voluntary transactions between buyers and sellers and characterised by voluntary non-heirarchial relationships. Production-distribution can be organised within the firm or via the market. One could be an educational entrepreneur who contracted in the spot market for lecturers and lecture halls (private tutoring firms do do this) or one can have a full-service university which employs lecturers on a salary and owns the lecture halls (this is firm production).

Neoclassical theory of the firm

This is the textbook theory of the firm that you will find in the microeconomics text book (such as the one on the reading list) and which you will study next year. In this the firm can be reduced to a production function. What the firm does is turn inputs into outputs. Its size and scope is determined by profit (or objective if in the public sector) maximisation. The figure below shows how we might represent a car firm which turns labour and capital into cars according to a formula. The curve shows the particular combinations of labour and capital that are required to produce a given quantity of cars (say 10). Note the pictures illustrates what might be going on along an isoquant: one could use more robots (capital) or more workers (labour) on a production line.



Does this view capture the essence of the firm? Does it explain existence of the firm in the first place? Sure, I can probably derive a production function for JLR or BP, but why does this collection of assets/employees exist in the first place in one firm? Nor does the formula explain why a car firm is necessary and why an entrepreneur can't just outsource the optimal combination to third parties.

Contrast this with Dennis Robertson (1923) who said that a firm was:

'An island of conscious power in this ocean of unconscious co-operation, like lumps coagulating in a pail of buttermilk'

This emphasises that the firm differs from the market by being about the fact that in a firm there exists decision making authority rather than decentralised buying and selling.

Thus, in neoclassical theory the firm can be reduced to a mathematical construct. This is 'convenient' because it allows modelling of the determination of behaviour (e.g. output choices) under specific assumptions. It also helps the modelling of firm strategies under different types of market structure and specific assumptions (*equilibrium* framework) – something which will be talked about in the strategy and marketing part of the course.

Back to our questions at the beginning within this theory. It does not explain existence. However, it can explain size of the firm. If technology or input prices change then given the nature of economies of scale

and the rational pursuit of 'profit maximizing' size, the size of the firm will change. If the nature of competition in the market changes (other firms compete less strongly) this will also cause a firm to rationally increase in size in this theory (because it is profitable to do so).

The theory treats the firm as a 'black box'. What happens inside the firm is not explained. There is also a pervasive assumption of rationality (+ perfect information): maximization of profit (or public sector benefit for public firms) is assumed.

Transactions Cost theory

Transactions costs are the cost of discovering prices and the cost of negotiating and concluding prices as well as making and enforcing contracts. This theory treats the firm and market as 'alternative' ways of carrying out transaction processes. The fundamental intuition is that: if markets are efficient resource allocators, why don't entrepreneurs just contract for everything through the market? Answer because there are transaction costs. Why do universities own lecture halls and employ lecturers: because it is costly to have to contract via the market every time you want to run a new lecture course, it is much easier to reallocate owned lecture halls and employed lecturers.

It was Ronald Coase who came up with this theory of the firm. "A firm consists of the system of relationships which comes into existence when the direction of resources is dependent on an entrepreneur." (Coase, 1937)

The firm is characterised by an authority structure (a *hierarchy*) and **incomplete contracts**. An employment contract usually just defines a job title, normal working time and says who you report to (mine only says these three things): it does not detail specific duties. Uncertainty leads to incomplete contracts; uncertain terms may be decided in the future by one party in a long-term contract (e.g. the head of the business school (or their delegated manager) directs me to teach you this course as part of my employment contract). The alternative to this is the *market* which has complete contracts with legal force. Firms have flexibility of response in conditions of uncertainty about the future. Transactions cost theory thus answers all of our questions at the beginning. Firms exist because they offer a cheaper alternative to the market for organising production based on the ability to direct the use of employees and owned capital, due to authority structures within firm. What determines the size of the firm? *The point where the cost of internal organisation of a transaction and external organisation are equal.*

One question is that if transaction costs are lower within the firm is why is all production not in one firm? The answer is that as a single firm gets bigger transactions are spatially dispersed, transactions are dissimilar, costs of intervention rise over time (the costs of disciplining employees who don't do what the top management want rises because of extended chains of command). Large firm size implies market power and hence regulation, so firm size reaches limits in certain markets. The Coasean view is that in capitalism the economically efficient amount of hierarchy emerges, while in command economies (such as the USSR) imposed firm size is not efficient. This can also happen in capitalist economies when state owned firms grow too big, protected by restrictions on private sector competition. A good example of this being the Central Electricity Generating Board in Great Britain, which at its height in 1989 owned all of the electricity generating plants in England and Wales (and was very inefficient).

Critique of the theory includes the following. Authority is not the essence of the firm; loyalty and identification may be more important. Market transactions not just contract based: *trust* matters. Thus it is possible to have flexible contracting via the market – open ended contracts do exist which can mimic the flexibility of within firm production. What generates whatever it is that is transacted is still a somewhat open question (which is what the capabilities view below focuses on)?

The property rights view of the firm

"In a world of transaction costs and incomplete contracts, ex post residual rights of control will be important because, through their influence on asset usage, they will affect ex post bargaining power and the division of ex post surplus in a relationship. This division in turn will affect the incentives of actors to invest in that relationship."

Hart, O. (1989). An economist's perspective on the theory of the firm. Columbia Law Review 89(7), p.1766

The property rights view of the firm is set in a world with transactions costs and incomplete contracts (similarly it can explain existence and size/scope of the firm). Whereas Coase emphasized the flexibility of the employment contract, this view emphasizes the importance of ownership of capital. Ownership matters because it carries *residual control rights* to take decisions in unpredicted/uncontracted for situations (Hart and Moore, 1990). Ownership reduces the possibility of returns to particular investments being partially appropriated by other firms. Hart and Moore have a nice example of a luxury yacht which provides meals and sailing and show how third party ownership of the whole ship is likely to be more efficient than either the chef or the skipper owning their part of the ship. This is because separate ownership results in the partial appropriation of individual investments by the chef and the skipper in their part of the yacht. This is in spite of the fact that separate ownership might be operationally more efficient in terms of sharpening the chef's incentives to reduce food waste or the skipper's incentives to reduce fuel burn.

Thus particular collections of assets (firms) may represent the optimal trade-off between operational and investment incentives. Thus while it may be operationally more efficient to outsource production to a third party, the problem is they would own the assets that the first firm relies on to produce the final product. The classic example of this was the GM acquisition of Fisher Body (in 1926) which illustrated that complementary/specific assets should be under joint ownership. Fisher Body made the bodies of GM cars, as car sales grew rapidly in the 1920s GM wanted to increase purchases from Fisher Body. But because this was in addition to their initial contract Fisher Body started increasing the price they charged, to the point that GM had to purchase Fisher Body to gain control of the strategic assets that it owned and that it was using to 'hold up' GM and demand higher prices. This is an illustration of the problem of separating ownership and control of assets.

The Marxist view of the firm

The rise of the factory system in the 18th and 19th centuries may be seen as an attempt by factory owners to gain control over the production process in order to extract a larger share of the value added in production (Marglin, 1974).

In this sense transactions cost view of the firm as a hierarchy and Marxist view are similar (and explain both the existence and size/scope of firms). BUT purpose is not 'efficiency' but exploitation. It is an

attempt by the owners of capital to increase their share of the value added in production by reducing the share going to wages.

This view suggests that firms may develop in ways designed to shift the distribution of value rather than to improve efficiency (e.g. outsouring, or threatening to outsource, maintaining multiple factories in different countries rather than one in one country etc…). It highlights the potential conflict of interest between owners/workers (capital vs. labour). The rise in self-employment might in this view be seen as either a result of capital owners trying to reduce the power of unionized labour by outsourcing to small competitive firms (or self-employed individuals in the gig-economy) OR the desire of workers to escape from the restrictive practices of capital owners and leverage their own intellectual capital (in the knowledge economy).

Managerial theories of the firm

These were driven by the rise of the 'Modern Corporation', characterised by dispersed shareholding and the observed separation of owners and managers noted in famous writings:

Berle & Means (1932) The Modern Corporation and Private Property

Marris (1963) The Economic Theory of Managerial Capitalism

Thus we might observe the pursuit of conflicting managerial and shareholder objectives. Managers want to maximise their own welfare. Shareholders want to maximise profit. This gives rise to the principal agent problem, where the principal (the shareholders) wants the agent (the management) to maximise profits in conditions of a lack of observability of their actions. Various writers imagined what managers might have been maximizing if not profits: sales (Baumol, 1959); slack (private jets, time on golf course etc.) maximisation (Williamson, 1964); and growth (Marris, 1963).

These theories (which are variants of neoclassical theory) don't explain existence of firms but can explain the size and scope of the firm.

Behavioural theories of the firm

Behavioural theories of the firm suggest that bounded rationality and uncertainty make optimisation impossible. Whereas all the theories we have considered so far assume some sort of optimization is going on with respect to the activities of the firm.

Large firm organisations have multiple objectives reflecting interest groups in organisation (setting of targets). Managers reconcile the competing objectives of interest groups within the firms subject to a 'satisfactory' level of profits (Cyert and March, 63). Thus large firms spend their time playing off the interests of the finance group, the marketing group, the engineering group etc within the firm and periodically indulging them to keep then happy. The need to adjust to an uncertain environment may lead to organisational slack.

One behavioural theory is that firms engage in a gift exchange with workers: workers supply effort in return for fair treatment (Akerlof, 1982). This could be a profitable strategy if the workers respond by raising aggregate productivity. In Akerlof's paper ('Labour Contracts as Partial Gift Exchange', *Quarterly Journal of Economics*) he gives the example of an office where employees had a target level of

envelopes to stuff every day. Everyone exceeded the minimum (because it was quite low), but the average amount was much higher than the minimum, which he attributed to the benign working environment encouraging generous reciprocal behavior by the employees (Cambridge University is certainly a good example of this).

This theory explains what is going on within the firm, but does not explain existence of firms and only explains the size/scope as an outcome of non-optimising behaviour.

The capabilities approach to the firm

In this theory the modern capitalist firm exists because it 'knows' how to do certain things, and can exploit this ability (Chandler, 1992, *Scale and Scope: The Dynamics of Industrial Capitalism*). Those capabilities may reflect/arise from core competencies based on: management skill, history in the industry, specific assets (tangible and intangible) and a particular location. These capabilities generate *competitive advantage.*

A classic example of this is the BASF chemical plant in the picture. It exploited economies of scale and scope at a particular location on the Rhine, ideal for chemical production and distribution. Other examples include Cambridge University itself…

This theory explains existence of firms rather well.

Evolutionary theories of the firm

A good example of this is by Nelson and Winter (1982). This is a dis-equilibrium, as opposed to equilibrium approach… bounded-rationality is assumed (as in behavioural theories): Bill Gates or Steve Jobs did not optimize. They were the lucky winners at an opportune time. There is a firm population dynamics and firm heterogeneity ('fitness'). Some firms are going to get lucky and discover a new technology which is out there or which is now possible to be discovered given other technological advances (e.g. in lasers or silicon). In this view competition is a process which gives rises to variation, selection and retention (Schumpeterian *creative destruction – some firms fail, others take their place*). Innovation drives structural change ('dynamic' capabilities) in the economy – e.g. personal computers massively aid the creation of small firms by reducing admin costs etc. A nice example of this might be to compare the history of Acorn Computers (in the UK) with Apple (from the US). Acorn was founded in 1978, Apple in 1976. Acorn was the 'British Apple', inventing the BBC Microcomputer, but lost its independence in 1985 after failing to capitalize on its early promise. Apple also struggled in 1984, but the rest is history….

We have firms as part of the discovery process and what determines their success (size and scope) is largely their success in the discovery and exploitation of innovations/opportunities.