MPRA
Munich Personal RePEc Archive

A Theory of Return-Seeking Firms

Murray, Cameron and Markey-Towler, Brendan

University of Queensland

20 January 2014

Online at https://mpra.ub.uni-muenchen.de/53051/
MPRA Paper No. 53051, posted 20 Jan 2014 14:06 UTC

A Theory of Return-Seeking Firms

Cameron K. Murray* and Brendan Markey-Towler†

January 14, 2014

Abstract

We introduce a theory of return-seeking firms to study the differences between this and standard profit-maximising models. In a competitive market return-maximising firms minimise average total costs leading to output choices independent of price movements. We investigate the potential for mark-ups over cost under both competitive and non-competitive market structures and characterise output and input choices under both, amongst a series of other interesting results. We also extend the model in the case of discrete output and input space and show what conditions are required of demand shifts for firms to modify their production plan.

1 Objectives and behaviour of firms

Standard models of firm behaviour in economics assume firms maximise profits, taken to be sales revenue minus production costs, when choosing their combinations of inputs and outputs. However, in the financial management and business literature it is taken as somewhat axiomatic that the decisions of the firm are characterised by maximisation of the risk-adjusted rate of return on costs, empirically[1], and theoretically[2]. A similar, and venerable, body of research in empirical economics supports the theory that firms seek to maximise returns, finding that firms typically assess capital investments and decisions in general through internal rate of return or net present value criteria (Kuh, 1963; Jorgenson and Stephenson, 1967; Blinder, Canetti and Lebow, 1998).

In a world of constraints on behaviour it is perfectly logical for firms to seek returns in their production activities rather than profits in the first instance. When a firm is constrained in the amount of resources it can invest, the optimal behaviour is to choose that production plan which returns the greatest amount on the costs incurred in pursuing a revenue stream. This is the standard axiom of the literature on "capital budgeting" which analyses investment decisions where constraints on financial capital availability are present, stating that that investment which returns the greatest for the amount invested should be selected.

*University of Queensland, contact email: ckmurray@gmail.com.
†University of Queensland, contact email: brendan.markeytowler@uqconnect.edu.au.

[1] Graham and Harvey (2001) find that 75% of firms surveyed used net present value, or, somewhat equivalently, internal rate of return decision rules

[2] See any major text on financial management such as Berk and DeMarzo (2013, pp.207-214)

Moreover, the logic of return maximising firms is supported by the real options literature, centered around Dixit and Pindyck (1994) in which firms have the ability to delay incurring irreversible costs of production, and thus face the problem of selecting the optimal point at which exercise their option to invest. The opportunity cost of *not* exercising the real option to invest in a production plan is the rate of return on that production plan (Pindyck, 1991, pp.1119-1120). The value of the option is then directly and monotically dependent on the rate of return (Pindyck, 1991, p.1123), and if the value of the firm is reflected in the value of its real option to invest and produce, the value of the firm also depends directly on the rate of return. Hence it follows that, were the option to be exercised and the firm to engage in production, the logic of optimality implies it would choose that production plan which maximised its returns, and hence the opportunity cost of *not* investing, and hence the value of the real option.

We seek here to analyse, using standard and mathematically rigorous methods what implications this assumption on the objective function, rather than profit maximisation has in the economic sense for the supply of output. In short, we seek here to build a theory concerning the choices of return-seeking firms. We know of no such theory prior to the present work that incorporates standard investment analysis into a model of the firm within a market context.

Although the conceptual difference between profit and rate of return seems and is really rather minor[3], we show that changing this key element in fact radically alters the theory of firm behaviour in a number of ways. In the first instance, a firm within a competitive market (i.e. one where the firm is a price taker) no longer seeks to equate marginal costs with the price prevailing, but rather to minimise costs alone, which implies that the concept of a supply curve, a response of supply to prices, breaks down. In such markets the supply "curve" of each firm is vertical at the cost minimising level of output. Rather strong additional assumptions on demand are required to eliminate the possibility of non-trivial mark-ups over marginal cost, even in competitive markets. Interestingly, this also requires us to incorporate in a significant way economies of scale in production to allow for decreasing costs over some region of output space, which is necessary for a non-trivial minimum output to exist.

In general, without imposing the assumption of perfect competition on our model there remains some general agreement of our theory of firm behaviour with standard models insofar as non-competitive markets are concerned, as output under non-perfect competition will be restricted relative to perfectly competitive output. In this model we can clearly observe how the degree of competitiveness in the market affects the decision making process of the firm, and it is quite transparent that as the market becomes more competitive and as the firm increasingly loses control of the price it sets, it shifts its attention with increasing exclusivity toward minimising costs. However, in this model, we cannot without additional knowledge, or assumptions about the exact specification of the demand and cost curves analyse the response of firm choices to changes in demand under non-competitive markets. Hence, the direction, in general, of the relationship between demand, output choice and price cannot be determined. Indeed, we demonstrate informally that it is entirely possible for an increase in

[3]Profit $\pi(q)$ being the revenue minus costs, and rate of return here taken to be the revenue minus costs divided by costs, or profit divided by costs $\frac{\pi(q)}{c(q)}$

demand to lead to a *decrease* in the price of output. Fascinating though, is the implication of this model that, insofar as non-competitive firms *will* respond to demand while firms in competitive markets will *not*, perfect competition can be, in some sense, less efficient than imperfect competition.

In this paper we also introduce concepts in our model of return-seeking firms to make tractable the discrete rather than continuous nature of many economic processes, which have been demonstrated to be of importance in a range of different economic contexts by Levine and Pesendorfer (1995). Many economic processes are "quantum" for physical or institutional reasons, requiring the involvement of some minimal unit, such that dynamics are discrete rather than continuous. In particular, our model "quantises" (or, makes discrete) both the output and capital input spaces, reflecting the fact that some minimal unit of output is required for consumption and similarly some minimal unit of capital is required before it is productive. This has significant implications under non-competitive markets for the responses of firms to changes of demand. Because of this quantised nature of inputs and outputs, markets do not clear in the traditional sense, and as such, the welfare theorems do not hold in an unaltered form. Demand is not necessarily met fully (under-supply), or firms may have to oversupply simply because of the minimal unit required for output, even in price-taking markets with utterly free entry. Nor are profits necessarily always maximised. In this sense we can say the theory is one of return-seeking, rather than maximising firms. Moreover the capital debates (see Samuelson (1966)) made clear that the lumpiness of capital is an important consideration in firm production, and the current model based on a theory of return-seeking firms makes it possible to explicitly address this point, showing that quantum shifts in output decisions are required for increases in the use of capital inputs.

We proceed as follows. In Section 2 we work first through the implications for output of a theory in which firms maximise returns, deriving a fundamental, single equation, maximising condition on output which allows us to analyse both competitive *and* non-competitive markets. Continuing with our model, in Section 3 we introduce a quantised output space and analyse the dynamics of firm supply in response to changes in demand, before introducing in Section 4 a standard cost-minimisation theory of input choices, but where capital inputs are quantised in a similar manner to output for a number of physical or institutional reasons. In Section 5 we discuss further some implications of our model in an informal manner, before concluding the exposition of our theory of return-seeking firms.

2 Production plan of a return maximising firm

We can express returns on costs from the present up to an horizon $T \in \mathbb{R}$ over continuous time by means of the expression $R(q)$ for returns on costs as a function of output $q = \{q_t\}_{t=0}^{T} \subset Q \subset \mathbb{R}_+$

$$R(q) = \int_0^T e^{-\rho t} \left[\frac{p(q_t)\, q_t - c(q_t)}{c(q_t)} \right] dt \tag{1}$$

Here, $p(q_t) : Q \to P$ is a mapping from the firm's output to the price $p \in P \subset \mathbb{R}_{++}$ of that output, and can thus be thought of as the demand curve *for the firm*. This specification is convenient in allowing us to switch easily between perfect competition and a variety of non-competitive market structures including monopoly and oligopoly. To do so, we simply alter the form of the function to include the variables of interest as parameters *from the point of view of the firm*[4], most importantly, the output of substitute products by competitors. The cost function $c(q_t) : Q \to C$ maps the firm's output into its costs $c \in C \subset \mathbb{R}_+$ of producing that output. Both mappings are taken to be independent of time across the life of the planning problem. The term ρ is a discount rate. In a similar fashion we can define a production function $q : X \to Q$ from a space of inputs $X \subset \mathbb{R}_+^{|X|}$ into the output space $Q \subset \mathbb{R}_+$, independent of time also and implicit in 1 so that $q = q(x)$.

We take the returns maximised by firms to be the returns on cost because this takes account of the fact that the cost $c(q_t)$ is the cost which is required to obtain the profits $p(q_t)\, q_t - c(q_t)$ from production q_t, thus reflecting the total value of all the resources which must be dedicated toward earning that profit. $R(q)$ then reflects the return on the resources committed to production, incorporating the costs of obtaining capital within wider costs incurred. Indeed, if finance is required to pay these costs in the first place (as is the case when costs occur before revenues are realised) then this expression is return on *all* investments. Hence, maximising this function maximises the value of the firm in a quite broad sense, when the firm is valued according to its return on investment.

Indeed, we can show that maximising returns thus defined maximises the rate of change of profit per resource input (i.e. cost) and hence maximises the rate of change of firm value in costs incurred[5].

Our theory of firms then states that the planned sequence $q = \{q_t^*\}_{t=0}^T \subset Q$ of outputs that the firm selects will maximise returns, or, formally that the vector $\{q_t^*\}_{t=0}^T$ satisfies the condition

$$\{q_t^*\}_{t=0}^T = argmax_{\{q_t\}_{t=0}^T \subset Q} R(q) \tag{2}$$

which, assuming that $R(q_t)$ is a continuous and differentiable mapping in Q requires that

$$\frac{\partial}{\partial q_t} R(q) = 0 \quad \forall\, t \in [0, T] \tag{3}$$

Now, it has been hitherto implicitly assumed in the expression of demand and costs at time t that they depend on output at t, q_t alone[6]. If this is indeed taken to be the case we can see that

[4] We follow the mathematical convention of defining the argument of the function as the variables of the function rather than its parameters, which we could disregard at the cost of complicating the notation

[5] Suppose we take returns to be time-independent so that $R(q)=\frac{\pi(q)}{c(q)} = \frac{p(q)q-c(q)}{c(q)}$ then we can say that $\pi(q) = R(q)c(q)$, so

$$\frac{\partial \pi(q)}{\partial c(q)} = R(q) max_{q\in Q} R(q) = max_{q \in Q} \frac{\partial \pi(q)}{\partial c(q)}$$

[6] We emphasise again it is alone in the *variable* space Q of the firm, *not* in its parameter space. Output of other firms at *various* times may well enter the expression, but they, from the firm's perspective, are not variables it can control and are therefore for all intents and purposes parameters.

$$\frac{\partial}{\partial q_{t'}}\left[\frac{p\left(q_t\right)q_t - c\left(q_t\right)}{c\left(q_t\right)}\right] = 0 \tag{4}$$

hence

$$\frac{\partial}{\partial q_t}R\left(q\right) = 0\,\forall\, t \in [0,T]\; \frac{\partial}{\partial q_t}\left[\frac{p\left(q_t\right)q_t - c\left(q_t\right)}{c\left(q_t\right)}\right] = 0\,\forall\, t \in [0,T] \tag{5}$$

Hence, the firm effectively solves a symmetric problem for each and every time period $t \in [0,T]$ and so we can, without loss of generality reduce equation 1 to a non-time-dependent expression of the objective function

$$R\left(q\right) = \frac{p\left(q\right)q - c\left(q\right)}{c\left(q\right)} \tag{6}$$

and our theory (equation 2) to

$$q^* = argmax_{q\in Q}R\left(q\right) \tag{7}$$

$$q^* = q : \frac{d}{dq}R\left(q\right) = 0 \tag{8}$$

The theory that firms maximise their returns according to the first order conditions 8[7], implies that

$$c\left(q\right)\left[q\frac{d}{dq}p\left(q\right) + p\left(q\right) - \frac{d}{dq}c\left(q\right)\right] = \left[p\left(q\right)q - c\left(q\right)\right]\frac{d}{dq}c\left(q\right) \tag{9}$$

It is advantageous at this point to define average total costs as $ATC\left(q\right) \equiv \frac{c(q)}{q}$ and marginal costs as $MC\left(q\right) = \frac{d}{dq}c\left(q\right)$, so that, dividing each side of 9 by q we obtain

$$ATC\left(q\right)\left[d\frac{d}{dq}p\left(q\right) + p\left(q\right) - MC\left(q\right)\right] = \left[p\left(q\right) - ATC\left(q\right)\right]MC\left(q\right) \tag{10}$$

which can be simplified to a condition which the choice of return-maximising firms, q^*, must satisfy:

$$q^* = q : ATC\left(q\right)\left(q\frac{d}{dq}p\left(q\right) + p\left(q\right)\right) = MC\left(q\right)p\left(q\right) \tag{11}$$

We can take this to be the fundamental condition on the production plan q^* of a return-maximising firm, in that the return maximising output of the firm must satisfy this condition[8]. Note that if revenue is taken to be $R_v = p\left(q\right)q$, then we can define marginal revenue to be $MR_v\left(q\right) \equiv q\frac{d}{dq}p\left(q\right) + p\left(q\right)$, and so 11 becomes

---

7

$$\frac{d}{dq}R\left(q\right) = \frac{1}{c\left(q\right)^2}\left\{c\left(q\right)\left[q\frac{d}{dq}p\left(q\right) + p\left(q\right) - \frac{d}{dq}c\left(q\right)\right] - \left[p\left(q\right)q - c\left(q\right)\right]\frac{d}{dq}c\left(q\right)\right\} = 0$$

8 The first order necessary conditions imply the output choice is a maximum provided the second order sufficient condition that the function be concave holds, which follows if as below we assume the cost function is convex and the revenue function is weakly concave.

$$q^* = q : ATC\left(q\right) MR_v\left(q\right) = MC\left(q\right) p\left(q\right) \tag{12}$$

The great beauty of this equation is that it is highly general, and allows us to vary the degree of competition and analyse the implications for the return-maximising output simply by changing the first derivative of the demand function, reflecting inverse of the price elasticity of demand.

2.1 Production plans under special case of competitive markets

Let us take a competitive market as a baseline scenario for the return maximisation problem. In a competitive market firms are price takers as they face an infinitely elastic demand curve, so $\frac{d}{dp}q\left(p\right) \to \infty$, and it follows that, taking the limit,

$$\frac{d}{dq}p\left(q\right) = 0 p\left(q\right) = p \tag{13}$$

i.e. the price the firm receives for its output is independent of its output. If this is indeed the case then applying 13 to the fundamental condition (equation 11) on the choice of a return-maximising firm we find that the production plan under a competitive market, q_c^* must satisfy a special case of 11 where $\frac{d}{dq}p\left(q\right) = 0$:

$$q_c^* = q : ATC\left(q\right) = MC\left(q\right) \tag{14}$$

Hence the return-maximising firm in a competitive market seeks to produce that amount which equates its average and marginal costs, regardless of the price of their output, provided that that price is above costs[9]. This may seem a rather unusual statement, though it makes sense when we consider that firms in a competitive market have no control over their prices, while they do over their costs, so it would make sense for them to focus on costs in their decision making process unless prices are so low as to make their activities unprofitable. In fact, it is fairly straightforward to demonstrate a very intuitive result following from 14 (a special case of the fundamental condition 11) that in competitive markets, firms which have no control over the price of their output seek to minimise average total costs in order to maximise return.

This proposition is in fact the corollary of a well-known result (which will be proved here for the sake of rigour), that under certain conditions on the cost function, average total costs are minimised when average total cost is equal to marginal cost.

Proposition 1 *If* $\frac{d^2}{dq^2}c\left(q\right) > 0 \forall q \in Q$ *(that is, the cost function is globally convex), and the cost function is second-order but not third order differentiable then*

$$q_c^* = q : ATC\left(q\right) = MC\left(q\right) q_c^* = argmin_{q \in Q} ATC\left(q\right)$$

[9]It may seem trivial, but if $p\left(q\right) > ATC\left(q\right)$ then $p\left(q\right)q - c\left(q\right) < 0$, and return maximisation would imply that $q^* = 0$ as any production will lead to a loss

The first order necessary conditions for a minimisation problem require that

$$\frac{d}{dq}ATC\left(q\right)=\frac{c\left(q\right)}{q}=0$$

hence the first order necessary condition for the average total cost minimisation are

$$\frac{q\frac{d}{dq}c\left(q\right)-c\left(q\right)}{q^{2}}=0ATC\left(q\right)=MC\left(q\right)$$

Now, for a second-order but not third order differentiable average total cost function, the Taylor expansion about q is (for some constant κ)

$$ATC\left(q'\right)=ATC\left(q\right)+\left(q'-q\right)\frac{d}{dq}ATC\left(q\right)+\left(q'-q\right)^{2}\frac{1}{2}\frac{d^{2}}{dq^{2}}ATC\left(\kappa\right)$$

the first order conditions imply that

$$ATC\left(q'\right)=ATC\left(q\right)+\left(q'-q\right)^{2}\frac{1}{2}\frac{d^{2}}{dq^{2}}ATC\left(\kappa\right)$$

So we can say that q minimises $ATC\left(q\right)$ if and only if $ATC\left(q\right)<ATC\left(q'\right)\ \forall\, q'\neq q$, which would require that $\frac{d^{2}}{dq^{2}}ATC\left(\kappa\right)>0$ and the average cost function be convex, which is indeed the case, by assumption on the total cost function. Hence since q_{c}^{*} must satisfy the first order necessary conditions, it is indeed a minimiser of average total cost.

Given that the definition of q_{c}^{*} follows from the special case 14 of the fundamental condition 11 on output choice within the theory return-maximisation, we can state an immediate corollary of proposition 1.

Corollary 1 *Under competitive markets, return maximisation in output choice implies that firms seek to minimise costs*[10].

It an interesting though somewhat mathematically *de jure* requirement that in a competitive market there be economies of scale for the average cost minimisation problem to be non-trivial. Mathematically this is a trivial requirement, for if there was no region where average total costs were decreasing, then average total costs in all regions must be increasing, and there would be no turning point in $Q\subset\mathbb{R}_{++}$ where average costs are minimised at zero output. This is the content of another proposition, stating that economies of scale must exist for supply to be non-trivial.

Proposition 2 *If supply is non-trivial then there must be a region where average total costs are non-increasing. That is,* $q_{c}^{*}>0\exists\left[q_{1}\quad q_{2}\right]\subset Q:\frac{d}{dq}ATC\left(q\right)\leq 0$

Suppose that q_{c}^{*} is indeed a minimiser of costs. Then the second order sufficient conditions for the average total cost function imply that the first order necessary conditions are satisfied at a turning point of costs, and in particular

[10] As will be discussed later (see p.15 below), this maximising condition implies that there is in fact no supply curve at a firm level

a nadir (minimum) of the average total cost function. Hence it must be the case about q_c^* that $ATC\left(q_c^* - \varepsilon\right) > ATC\left(q_c^*\right)$ for some range of $\varepsilon > 0$. Let us suppose, by way of contradiction, that $\nexists\left[q_1 \quad q_2\right] \subset Q : \frac{d}{dq}ATC\left(q\right) \leq 0$. This then implies that average total costs must be everywhere increasing in q, that is, $\frac{d}{dq}ATC\left(q\right) > 0 \, \forall \, q \in Q$. But if this then were the case, we would have $ATC\left(q_c^* - \varepsilon\right) < ATC\left(q_c^*\right)$, which contradicts that $q_c^* > 0$ is a minimiser of average total costs, because only for $q_c^* = 0$ is $ATC\left(q_c^* - \varepsilon\right) \not< ATC\left(q_c^*\right)$, and even then this is only true trivially as $-\varepsilon \notin Q$. Hence it must be the case that for a non-trivial minimiser to exist there must be some region of Q for which average total costs are decreasing (strictly speaking, non-increasing).

Hence, for a positive supply to be consistent with return-maximisation under perfect competition, there *must* exist economies of scale. Obviously in an economic sense as opposed to a mathematical one, this requirement is highly important and interesting. Indeed it is crucial, for it implies if there is no region where marginal cost is non-increasing, there will be no point other than zero output at which costs are minimised, and thus in a perfectly competitive market return-seeking firms will produce nothing. Economies of scale then are vital (quite literally, were the output food) for without them there would be no supply in a perfectly competitive market according to the logic of return-maximising firms[11].

It should be clear then that under competitive markets, return maximising firms seek to minimise their costs given that they have no control over their prices. This is an interesting result as it differs in its conditions on choice from the standard models of profit maximisation under competitive markets, which state that firms will choose to maximise profits by producing to a level where their marginal costs are equal to the market price of their output (marginal revenue). Under that model prices and marginal costs will fall to average total costs only under specific conditions, which are taken generally to hold in the "long term", when competitive forces and free entry lead profit maximising firms to minimise costs. Here, firms attempt to minimise costs in the first instance, maximising their return. Like the standard profit maximisation models, we would expect competitive pressures over time to shift the demand for the firm's output toward the minimum of average total costs, but we would have to make additional assumptions within this model for that to be the case of necessity.

Indeed, given that firms have no control over their prices in the logic of this-as in the profit maximisation-model it is entirely possible in the first instance that prices are a mark-up over marginal costs (and average total costs), since firms focus on minimising costs only in striving for maximum returns rather than equating prices to marginal costs in striving for maximum profit. Indeed, in a competitive market with return seeking firms, we must say in the first instance that $p = MC\left(q_c^*\right)$ if and only if the demand curve for the firm's output satisfies that exact property. That is[12], if $q = q_c^*$ then $p = MC\left(q_c^*\right)$ if and only if consumers are willing to pay a price equal to the marginal cost of that product,

[11]This might seem utterly absurd and contradictory, given that the firm would make zero return producing nothing. We suggest that the problem lies not in the logic of return maximisation but rather in the fact that perfect competition is in fact an impossible scenario, and hence absurd to assume as anything other than a limiting case. Taking this view, these results seem not so contradictory given they occur within an absurd scenario.

[12]Given that all these variables are defined over the set of real numbers $\mathbb{R}$ which is continuous and non-reflexive for any two distinct elements of the set

or $p(q_c^*) = MC(q_c^*)$. This is equivalent to assuming that there is *perfectly* free entry of *exactly* symmetric *infinitesmal* firms. Hence we require additional, and rather strong assumptions on demand to eliminate the possibility of mark-ups in competitive markets with return-seeking firms. Over time, we may expect entry of other return-seeking firms if this were not the case, but in the first instance we cannot in a competitive market guarantee that prices equal marginal costs, and in general must say that even in such a market prices are a mark-up over marginal costs.

It may seem that this is a failing of the model. We disagree, and argue that it is actually a strength, given the universally acknowledged ubiquity of mark-ups of prices over costs across a variety of market structures. The model actually even under the strong assumption of perfect competition manages to reflect the reality of pricing strategies.

2.2 Production plans in the general case

In general, we can take *firm-specific* demand to be a non-specified function of the firm's output, $p(q)$. In this case, the firm's production plans must satisfy the general condition implied by return maximisation (equation 11)

$$q^* = q : ATC(q)\left(q\frac{d}{dq}p(q) + p(q)\right) = MC(q)\,p(q) \tag{15}$$

Immediately we can see that in general, when demand is not infinitely elastic, costs are no longer minimised. Simply observe that equation 12, following directly from this condition, implies

$$ATC(q)\frac{MR_v(q)}{p(q)} = MC(q) \tag{16}$$

and hence provided that $\frac{MR_v(q)}{p(q)} \neq 1$ (this occurs when demand is not infinitely elastic and the market is imperfectly competitive), marginal costs are not equated to average total costs when returns are maximised. This is quite simply because when the firm does not face an infinitely elastic demand curve, it is not a price taker and can affect the price of its output in the market. It thus must take prices and their determination into account in its production plans. It is quite elementary to demonstrate that average costs are higher than marginal costs, and the immediate corollary, applying Proposition 1 is that costs are no longer minimised under return-maximising firms.

Proposition 3 *The production plan of a return maximising firm for a normal good is in general such that average total costs are higher than marginal costs. That is, for q^* satisfying 11, and for $\frac{d}{dq}p(q) \leq 0$*

$$ATC(q^*) \geq MC(q^*)$$

This proposition immediately follows from the definition of q^* as satisfying condition 11, which may be rewritten as

$$\frac{\left(q^*\frac{d}{dq}p(q^*) + p(q^*)\right)}{p(q^*)} = \frac{MC(q^*)}{ATC(q^*)}$$

Now, if the output is a normal good the demand curve has a negative slope, so we have

$$\frac{d}{dq}p\left(q^*\right) \leq 0 \frac{\left(q^*\frac{d}{dq}p\left(q^*\right)+p\left(q^*\right)\right)}{p\left(q^*\right)} \leq 1 \frac{MC\left(q^*\right)}{ATC\left(q^*\right)} \leq 1$$

And thus it must be the case for 11 to hold that $ATC\left(q^*\right) \geq MC\left(q^*\right)$

This makes sense when we consider that in any deviation from the competitive market production plan q_c^*, costs will no longer be minimised because q^* must satisfy a condition which no longer guarantees the equality of marginal and average total costs. Intuitively, this non-minimisation of costs is because the firm can in general affect the price of its output, and can increase returns either by increasing price or decreasing costs. Indeed, we can and will show now that the return maximising firm will set the price of its output as a mark-up over its costs.

Proposition 4 *Any non-trivial*[13] *production plan of a return maximising firm for a normal good is in general such that prices are a mark-up over its marginal costs. That is, for $q^* \neq 0$ satisfying 11, and for $\frac{d}{dq}p\left(q\right) \leq 0$*

$$p\left(q^*\right) \geq MC\left(q^*\right)$$

The same conditions here are assumed as for Proposition 3, and so it must be the case that $ATC\left(q^*\right) \geq MC\left(q^*\right)$. Now if we can show that $ATC\left(q^*\right) \leq p\left(q^*\right)$, then by the transitivity of the real numbers $\mathbb{R}$ upon which all these variables are defined, we will have the result.

Now, that $ATC\left(q^*\right) \leq p\left(q^*\right)$ is in fact implied by the definition of q^* in the theory of maximum returns (equation 7), which states that

$$q^* = argmax_{q\in Q}R\left(q\right) = argmax_{q\in Q}\left\{\frac{p\left(q\right)q - c\left(q\right)}{c\left(q\right)}\right\}$$

To see this, suppose by way of contradiction that $ATC\left(q^*\right) > p\left(q^*\right)$ and $q^* \neq 0$. Then $p\left(q^*\right)q^* < c\left(q^*\right)$ and $R\left(q^*\right) < 0\,\forall\,q \in Q\backslash\{0\}$. But this contradicts q^* being a maximiser of returns, since $R\left(0\right) = 0 > R(q^*)$. Hence in all but non-trivial production plans, we must have $ATC\left(q^*\right) \leq p\left(q^*\right)$, while we have from above proposition 9 that $ATC\left(q^*\right) \geq MC\left(q^*\right)$, so by the transitivity of the real numbers $\mathbb{R}$ we can state that

$$p\left(q^*\right) \geq ATC\left(q^*\right) \geq MC\left(q^*\right) p\left(q^*\right) \geq MC\left(q^*\right)$$

Hence when we are in the general case where markets can be competitive or non-competitive in any manner[14], return-seeking firms will supply output such that prices are a mark-up over costs, both marginal and (as was demonstrated in the proposition) average. This is because whereas in the competitive market case, they seek only to minimise costs and have no control over prices, in a non-competitive market they will have control over prices, and can increase their

[13]i.e. any production plan which involves some form of actual production rather than the trivial no-production production plan $q^* = 0$

[14]$p\left(q\right)$ can be taken as a monopolists demand curve, or if we were to take other firms' quantities as parameters it could be the demand curve for a Cournot oligopolist

returns by manipulating prices to be higher and costs lower. We can see quite easily that they do this by restricting supply, producing less than they would in a competitive market, as we would expect a non-competitive firm to act.

Proposition 5 *For a return-maximising firm producing a normal good, the production plan for a non-competitive firm entails restricting supply relative to the competitive supply. That is, for $q^* \neq 0$ satisfying 11, and for $\frac{d}{dq}p(q) \leq 0$*

$$q_c^* \geq q^*$$

From Proposition 1, we have that $q_c^* = argmin_{q \in Q} c(q)$ since $ATC(q_c^*) = MC(q_c^*)$. But in general q^* does not satisfy this, by the fundamental condition on the choice of a return-maximising firm (equation 11), and hence $ATC(q^*) \geq ATC(q_c^*)$.

Suppose by way of contradiction then that $q^* > q_c^*$. Then as above we have $ATC(q^*) \neq min_{q \in Q} ATC(q)$. However, in almost all cases q is a normal good, so $\frac{d}{dq}p(q) \leq 0$, so $p(q^*) < p(q_c^*)$. Hence $q^* > q_c^*$ contradicts the definition of $q^* = argmax_{q \in Q} R(q)$, since prices can be increased and costs decreased, and return thus increased by changing production plans from q^* to q_c^*. Hence it must be the case that $q^* \leq q_c^*$.

These results are on the face of it much in agreement with the results of the standard profit maximising theory of firms. Indeed, as in such models we have firms restricting output in a bid to push prices upward and obtain economic rents[15]. However, this model shows us also how the decisions processes of firms vary with the degree of market competitiveness. We have noted this already (see p.6 above), but to be more specific, notice that the fundamental condition

$$q^* = q : ATC(q)\left(q\frac{d}{dq}p(q) + p(q)\right) = MC(q)\,p(q)$$

varies with the degree of competition as the slope of the demand curve changes, and hence the relative importance of the different variables in the decision process also changes. In general firms which maximise returns according to this condition must consider *both* prices and costs. But as the market becomes more competitive the demand curve becomes more price elastic, and $\frac{d}{dq}p(q)$ tends to zero, implying price-setting matters less in the firm's decision process relative to cost minimisation. As markets become more and more competitive then, return-maximising firms respond less to demand when making decisions, in the limit of perfect competition effectively merely checking to see prices are above or below average total costs.

3 Output choices in discrete alternatives space and the supply curve

We have now studied *ad nauseam* the properties of q^* as defined by the conditions above, but before we apply the theory of return-maximising firms to the

[15] If the firm obtains rents, then it has value. Rents arise from the non-equality of prices with costs, and so if prices equal costs the firm has no rents and therefore no value.

choice of inputs to support output of q^* we will acknowledge and try to incorporate into our theory that in reality it is exceedingly rare for the production space Q to be continuous[16]. In general production is limited to discrete quantities as goods and services tend to be objects which can be only supplied in multiples of some minimal unit. We can only buy multiples of q many *units* of a good in general, be that constraint imposed "artificially" by the firm, through market institutions or physical limits. There exists, in a sense a "quanta", or minimum unit of production.

To that purpose, we define $Q_\Delta \subset Q$ to be a discrete grid[17] contained within the production space Q where production plans are defined, such that we can say Q is dense in Q_Δ and that Q_Δ "quantises" the output space Q. It seems fairly reasonable to us that maximal elements may exist between two discrete units (i.e. a $q^* \in Q \backslash Q_\Delta$), so that, technically speaking, the firm would maximise revenue at a point in Q but not in Q_Δ. The firm may indeed know that this is the case, and that the constraint imposed by the discrete nature of $Q_\Delta \subset Q$ makes such a production plan impossible. It is for this reason that we say our theory is one of return-"seeking", rather than return-maximising firms. In such cases, we say that the firm's production plan becomes $q_\Delta^* \in Q_\Delta$ such that the distance between q_Δ^* and q^* is minimised,

$$q_\Delta^* = argmin_{q_\Delta \in Q_\Delta} \{\lambda = |q_\Delta - q^*|\} \tag{17}$$

In fact, this can be a general specification for the choice of a firm, since if $q^* \in Q_\Delta$ then $min_{q_\Delta \in Q_\Delta} \lambda = 0$ and $q_\Delta^* = q^*$. This may seem a rather arbitrary specification, though it merely specifies that while firms may know that their maximal production plan is q^*, if they cannot produce at this level due to the constraint imposed by the discrete production requirement, they will produce as close to that point as possible. There are nonetheless two drawbacks to such a specification, the first being that in the case where q^* is equidistant from two points in Q_Δ, the set of q_Δ which satisfies this condition is not a singleton, though we would tentatively suggest that such a case is rather rare. A second drawback is that equation 17 is only consistent with the theory of return maximisation provided that $q_\Delta^* = argmax_{q_\Delta \in Q_\Delta} R(q)$. We can show that this is only the case when the returns function is symmetric about the maximal point q^*.

Proposition 6 $q_\Delta^* = argmax_{q_\Delta \in Q_\Delta} R(q)$ *for any quantisation of* Q *if and only if* $R(q)$ *is symmetric about* q^* *(that is,* $R(q^* + \varepsilon) = R(q^* - \varepsilon)\ \forall \varepsilon \in \mathbb{R}$*).*

Notice that $argmax_{q_\Delta \in Q_\Delta} R(q) = argmin_{q_\Delta \in Q_\Delta} R(q_\Delta) - R(q^*)$, since by definition $q^* = argmax_{q \in Q} R(q)$, that is to say, the q_Δ which maximises returns will be the same q_Δ which minimises the difference between the return yielded by it and the maximal return overall. Now, $argmin_{q_\Delta \in Q_\Delta} R(q_\Delta) - R(q^*) = argmin_{q_\Delta \in Q_\Delta} |q_\Delta - q^*|$ for any quantisation of Q if and only if $R(q)$ is symmetric about q^* because if this is indeed the case, then the only criterion determining whether $q_\Delta = q^*$ is a maximal element is its distance from q^* since

[16]If $Q \subset \mathbb{R}$ were convex then for any two $q, q' \in Q$ we would have $q'' = \alpha q + (1-\alpha) q'' \in Q\, \forall \alpha \in [0,1]$ and Q would be an interval contained in $\mathbb{R}$

[17]Δ=delta, d for delta, d for discrete. A grid such as Q_Δ could be taken to be, *inter alia,* the set of natural numbers $\mathbb{Z} \subset \mathbb{R}$, in which the real numbers $\mathbb{R}$ are dense

symmetry implies that $R(q)$ is decreasing in an equivalent manner in both directions around q^*.

A more consistent theory of q^*_Δ would be to assume outright that it is selected to minimise the difference between the return it yields, $R(q^*_\Delta)$, and $R(q^*)$. However an important implication of the theory that is clearer to see as it stands is that when we have a discrete output space, *quantum* jumps in demand of certain magnitudes are required for output of the firms to increase, and hence there is a range of demand curves $p(q)$ for which supply q^*_Δ remains unchanged, and hence there is excess demand or supply, and the market does not "clear" in the traditional sense.

To see this, take a production plan q^* which satisfies 11, and which thus depends on demand $p(q)$. Suppose then that there is an increase in autonomous (i.e. non-price dependent) demand for every output q, denoted $\partial p(q) = p'(q) - p(q) > 0 \, \forall \, q \in Q$,[18]. When this happens, 11 will no longer hold with equality at q^*, and in particular, proposition 3, and a non-competitive market (where demand is not infinitely elastic) implies that

$$ATC(q^*)\left(q^* \frac{d}{dq} p'(q^*) + p'(q^*)\right) > MC(q^*)\, p'(q^*) \tag{18}$$

provided that the increase in demand is uniform across Q (i.e. we have a simple shift of intercept) or, failing this, that the change in the slope of the demand curve is negligible, so that $\frac{d}{dq}p'(q)$ remains effectively constant or at least sufficiently small. Restoring the equality of 11 to maximise returns with the new demand curve then requires a decrease in costs through an increase in output, which we can see by the fact that the re-written inequality

$$\frac{\left(q^* \frac{d}{dq} p'(q^*) + p'(q^*)\right)}{p'(q^*)} > \frac{MC(q^*)}{ATC(q^*)} \tag{19}$$

must be restored to equality through a change of the return maximising output q^*. The right hand side must increase via an increase in the maximal production plan, denoted $\partial q^* > 0$, which is implied by Proposition 1 and Proposition 5 (provided that the cost function has a uniform first derivative throughout the half-space of Q to the right of q^*), while the left hand side must decrease via the same increase in q^* (implied by negatively sloped demand when the second derivative of demand is small) until equality is restored.

This implies that the return-maximising output increases with increases in demand, and in some sense supply curves slope "up" . However, in another sense, supply curves can be of any shape, because additional assumptions on the exact shape of the cost curves and demand curve will be required for the increase in return-maximising output to not lead to a decrease in price below its previous level. Hence, while return-maximising output will increase (under certain conditions), in response to an increase in demand, the increase in demand may nonetheless not *necessarily* lead to an increase in prices.

Moreover, the discrete nature of output space Q_Δ means that it is not necessarily the case that what is *actually* supplied, q^*_Δ, as defined by 17 above *will* respond to an increase in demand, even if q^* does. If there is an increase

[18] The reverse argument will hold for a decrease in demand at every output, $\partial p(q) = p'(q) - p(q) < 0 \, \forall \, q \in Q$

in demand, then there are three possible cases for the response of supply in a discrete output space to demand.

Suppose that initially the maximal production plan lay within a closed interval encompassed by two adjacent points in the grid Q_Δ, that is, $q^* \in \left[q_\Delta^1 \quad q_\Delta^2\right] \subset Q$, then

1. If $q_\Delta^* \leq q^*$ ($q_\Delta^* = q_\Delta^1$) prior to the change in demand then the required change in demand for an increase in maximal output q^* for supply to change *at least* to the next point in the grid $Q_\Delta \subset Q$, and hence $\partial q_\Delta^* > 0$, according to the definition of q_Δ^* is[19]

$$\partial p\left(q\right) : \left(q^* + \partial q^*\right) \geq \frac{q_\Delta^1 + q_\Delta^2}{2} \tag{20}$$

   intuitively, if there were under-production relative to the return-maximising amount, then the quantum shift in demand must push the return-maximising production plan in Q at least past the halfway point between the two adjacent points in Q_Δ for quantity supplied by the firm to increase.

2. If $q_\Delta^* \geq q^*$ ($q_\Delta^* = q_\Delta^2$) prior to the change in demand then the required change in demand for an increase in maximal production q^* for supply to change *at least* to the next point in the grid $Q_\Delta \subset Q$, and hence $\partial q_\Delta^* > 0$ according to the definition of q_Δ^* is

$$\partial p\left(q\right) : \left(q^* + \partial q^*\right) \geq \frac{q_\Delta^3 + q_\Delta^2}{2} \tag{21}$$

   where q_Δ^3 is the grid point right-adjacent to q_Δ^2 and defining the end-point of the closed interval $\left[q_\Delta^2 \quad q_\Delta^3\right]$. Obviously, this is a more difficult requirement for an increase in demand to fulfil, as it requires the return-maximising production to increase beyond the quantity currently supplied *and* the midpoint on the next interval, in that it requires a larger increase in demand *ceteris paribus.*

3. If neither of the preceding conditions are met, then $\partial q_\Delta^* = 0$. This occurs if the shift in demand is not quantum, and does not cause the new return-maximising production plan q^* to exceed at least one critical point for supply to switch from one point in the output grid $Q_\Delta \subset Q$ to an adjacent one.

These conditions taken together specify the dynamics of supply in response to a change in demand in a market with return-maximising firms, where for whatever reason (physical or institutional) only discrete outputs are permitted. We know from above (see p.13) that supply curves will in some sense slope upward under return-maximisation (i.e. with the demand curve, though not necessarily in price), but they will be a discrete and stepwise collection of points $q_\Delta^*\left(p\right)$ rather than a continuous curve $q^*\left(p\right)$. We will require quantum increases in demand (which is to say, an increase in demand of at least some minimal amount) for supply to respond given the constraint imposed by the discrete output space,

[19]The proof of these conditions is quite intuitive, and follows directly from the definition of $q_\Delta^* = argmin_{q_\Delta \in Q_\Delta}\left\{\lambda = \left|q_\Delta - q^*\right|\right\}$. Were this indeed the case, the point at which q_Δ^* switches between two adjacent points on the grid Q_Δ is the midpoint between them.

and hence it is not necessarily the case that supply maximises returns in a "global" sense. While markets will *tend* toward a maximising of returns (and presumably the desires of consumers) we cannot say that they in a traditional sense "clear". It is possible for there to be an under-supply ($q^* > q^*_\Delta$) when consumers would desire output which is not possible for the firm, but for which an over-supply ($q^* < q^*_\Delta$) on the part of the firm is not quite justified.

Again, some may see this as a failing of our theory. We disagree, and suggest that the quantisation of output space allows us to understand better part of the reason why inventories actually exist in reality, and why they can often run "short" relative to consumer demand. If only discrete outputs are possible, then part of the reason inventories exist and why they might suffer a shortfall is exactly that only discrete outputs are possible for whatever physical or institutional reason, while the "market clearing" output (where there would be no inventories, under-supply or over-supply) would actually lie .

Let us here introduce something of a *piece de resistance* of counter-intuitive results regarding the supply of output under the theory of return-maximising firms. Implicit in the foregoing discussion of the response of supply to demand was the assumption that markets are somewhat non-competitive, so that the firms demand curve is not infinitely elastic, such that $\frac{d}{dq}p(q) = 0$, so that q^* was determined by the general condition on return maximising production plans (equation 11). However, if we are in a competitive market, the return-maximising condition is no longer 11 but the special case (equation 14), where $\frac{d}{dq}p(q) = 0$,

$$q^*_c = q : ATC(q) = MC(q)$$

But here demand, $p(q)$ does not enter the consideration of the firm in setting its return maximising output, and hence changes in prices have no effect on changes in output, even without the output space being discrete, such that quantum jumps in demand are required. The response of supply to demand utterly is non-existent and supply curves are vertical at the cost-minimising output.

This rather counter-intuitive result stems from the fact in a competitive market each individual return-seeking firm is merely concerned with minimising costs (Proposition 1), as it has no control over the price level. So *a priori*, or in the short term without the assumption that there will be entry upon observance of positive returns, we cannot say that supply will respond to demand in a competitive market with return seeking firms. An increase in demand in a competitive market justifies not an increase in supply, because an increase in supply would increase what were hitherto minimised costs, and cut away at the increased returns yielded by an increased price.

Hence a competitive market with return-seeking firms will not clear in a traditional sense, *even if the output space were not quantised.* Without the assumption that entry occurs rather unusually rapidly up to the point where firm demand has fallen to average total costs, there will be "cash left on the table" insofar as increases in demand go unmet. In a sense then, while monopolistic markets are somewhat inefficient in that prices do not reflect accurately the cost of the output produced, insofar as supply response to changing demand is concerned they are *more* efficient than competitive markets. Under certain conditions, increased demand will be met with increased output, whereas we cannot, without additional and rather strong assumptions on entry, say the same for competitive markets.

4 Choices of inputs when capital inputs are quantum

A firm's production plan is not characterised alone by a decree that a certain output q^*_Δ be called into existence, the plan must also specify the inputs $x \in X \subset \mathbb{R}^{|X|}_+$ which are required for that output to be produced. The relation between the two is given by the production function $q : X \to Q$, or $q = q(x)$. Let us, as is common, restrict the cardinality of the input space to two, $|X| = 2$ and propose that there are two inputs into production, capital $k \in K$ and labour $l \in L$, so that $q = q(l \quad k)$. It is relatively standard to assume that there are diminishing but positive marginal products, $\frac{\partial}{\partial \cdot} q(\cdot) \geq 0$, $\frac{\partial^2}{\partial \cdot^2} q(\cdot) \leq 0$, and we will also assume that the cross-derivative of the marginal product of both labour and capital are positive, $\frac{\partial^2}{\partial l \partial k} q(l \quad k), \frac{\partial^2}{\partial k \partial l} q(l \quad k) \geq 0$, so that increased use of capital makes labour more productive and vice versa, reflecting complementarity of inputs in production.

The logic of return maximisation would suggest that the selection of inputs to produce output q^*_Δ is guided by the striving to minimise the cost $c(q^*_\Delta)$ of that output in the choice of inputs. Now, if we take $w \in \mathbb{R}$ to be the cost of labour, and $r \in \mathbb{R}$ to be the cost of capital in competitive input markets we can express the cost of producing q^*_Δ as

$$c(q^*_\Delta) = wl + rk \tag{22}$$

and so the theory of return maximisation would dictate that 22 be minimised subject to the constraint imposed by the production function $q(l \quad k)$. Hence our theory of input selection under return maximisation is much the same as that under profit maximisation

$$\{l \quad k\}^* = argmin_{\{l,k\} \in \left\{[l,k]: q^*_\Delta = q(l,k)\right\}} c(q^*_\Delta) \tag{23}$$

This is a simple minimisation problem which can be solved by the application of Lagrangian optimisation[20] to give the first order necessary conditions

$$k^* = k : \frac{\partial}{\partial l} q(l \quad k) = \frac{w}{\Lambda} \tag{24}$$

$$l^* = l : \frac{\partial}{\partial k} q(l \quad k) = \frac{r}{\Lambda} \tag{25}$$

eliminating the shadow price Λ of output[21] we obtain the condition which must be satisfied by input choices for costs to be minimised

$$\{l \quad k\}^* : \frac{\Delta l}{\Delta k} = \frac{\left[\frac{\partial q(l \quad k)}{\partial k}\right]}{\left[\frac{\partial q(l \quad k)}{\partial l}\right]} = \frac{w}{r} \tag{26}$$

[20]The Lagrangian, modified objective function, is here $L = wl + rk - \Lambda\left(q(l \quad k) - q^*_\Delta\right)$

[21]Which allows us to interpret these conditions as setting marginal products to shadow real prices of inputs

this condition implicitly defines the choice of inputs l and k as those which tangent the slopes of the isoquant $q^*_\Delta = q(l, k)$ and iso-cost $c(q^*_\Delta) = wl + rk$. This is a familiar and standard result of cost minimisation exercises under profit maximisation problems, where inputs are selected so that the marginal rate of technical substitution between inputs is equal to their relative costs.

Now let us consider the dynamics of a quantum increase in demand which causes an increase of supply in the quantised output space[22], that is, $\partial p(q)\ \forall\, q : \partial q^*_\Delta > 0$. Now, the return-seeking firm must produce this new level of ouptut, and so it is the case that the constraint in the cost minimisation problem must still hold, that is

$$q^*_\Delta = q(l \quad k) \tag{27}$$

but for 27 to hold, given that $\partial q^*_\Delta > 0$, we must have an increase in the inputs into production, that is, either $\partial k^* \geq 0$ or $\partial l^* \geq 0$ or both. However, it is typically the case that, like output, capital is quantised so that there is a minimum amount of capital that must be employed, for either institutional or physical reasons. If the capital in question were a machine, then increasing the stock involved in production by half a machine is unlikely to have an affect on output, as half a machine is useless. Thus, we think it reasonable to suppose that the space of capital inputs K is, like production, dense in a grid $K_\Delta \subset K$ of discrete capital inputs, each point in K_Δ representing a multiple of one unit of capital. Hence, like with output, though k^* minimises costs, it might be the case that $k^* \notin K_\Delta$ and therefore this level of input is technically impossible, and we can specify as with output that the selected capital input in K_Δ is $k_\Delta = argmin_{k_\Delta \in K_\Delta} \{\lambda = |k_\Delta - k^*|\}$. Obviously, 27 must still hold and so this is only possible by selecting whatever labour input makes it so.

Let us suppose, for the sake of illustrating the dynamics of quantised capital inputs, that initially (i.e. before the increase in demand), $k^* \in K_\Delta$, and so $k^*_\Delta = k^*$. Now, if there is an increase in demand such that there is a quantum increase in the supply that the firm selects, $\partial q^*_\Delta > 0$ but this for reasons specified below does not induce the firm to move from k^*_Δ to the next adjacent capital input in the grid input space K_Δ, then in order to produce this and for 27 to hold, labour input must increase to make it so, that is, $\partial l > 0$.

However, if labour inputs are increased, then given the diminishing marginal product of labour, $\frac{\partial^2}{\partial l^2} q(l \quad k) \leq 0$, we have that $\frac{\partial}{\partial l} q(l \quad k)$ will decrease, while the complementarity of labour and capital, $\frac{\partial^2}{\partial k \partial l} q(l \quad k) \geq 0$ implies that the marginal product of capital $\frac{\partial}{\partial k} q(l \quad k)$ will increase. Together these imply that

$$\frac{\left[\frac{\partial q(l \quad k)}{\partial k}\right]}{\left[\frac{\partial q(l \quad k)}{\partial l}\right]} \geq \frac{w}{r} \tag{28}$$

that is, the cost minimisation condition 26 no longer holds (in non-trivial cases) and costs $c(q^*_\Delta)$ are no longer minimised for the new quantised output. Hence the quantisation of capital inputs, which is imposed physically or institutionally in much the same manner as for output means that in general, a decision for the firm to increase output leads to a non-minimisation of costs as the exact additional capital required will be invested only in specific cases. Indeed, we

[22]Again, a similar argument would apply in reverse for a decrease in demand

can only say that capital inputs will increase at all if the cost of producing an increased level of output with the same level of capital stock, $c_{k_\Delta}(q^*_\Delta)$ is greater than the cost of that output with an increased capital stock k'_Δ in the half-space of K_Δ to the right of k_Δ, $c_{k'_\Delta}(q^*_\Delta)$

$$\partial k^*_\Delta > 0 \iff \partial q^*_\Delta : c_{k_\Delta}(q^*_\Delta) > c_{k'_\Delta}(q^*_\Delta) \tag{29}$$

If this is indeed the case, then a quantum increase in supply responding to a quantum increase in demand will lead to an incremental increase in capital from its previous level in the grid of capital inputs K_Δ to at least the right-adjacent point. Thus we can say that when firms seek to minimise costs[23] for any given output level q, when capital[24] is quantised so that only discrete increases are technically possible, they will not be able to minimise costs in general, and capital inputs increase in increments with quantum increases in decided output.

Nonetheless, we can see that since k^*_Δ is selected so as to minimise costs for a given return maximising (in Q_Δ) level of output q^*_Δ, the criterion guiding capital investment is equivalent to maximising net present value. Investment in an incremental unit of capital is only undertaken if not investing in that unit of capital were to be sub-optimal in providing inputs for producing the return-maximising level of output. Hence, capital investment seeks to maximise returns, by minimising the cost of producing a return-maximising level of output. It is the case then that the input selection theory of our model of return-seeking firms implies that choices concerning capital investment conform to the net present value, or internal rate of return methods typically employed in investment analysis.

5 Discussion

It is at this stage advantageous to elaborate in an informal manner some of the finer points of our theory of return-seeking firms with the aid of Figure 1.

Panel (a) of Figure 1 illustrates optimal firm output q^* consistent with Proposition 1 in the special case of "competitive" price-taking. Three points are of note. First, return maximisation corresponds with cost minimisation as per Corollary 2, implying that firm output will be unresponsive to any shift in the demand curve. Second, without additional assumptions on the position of the firm's demand curve, there is no implication that firm output is socially optimal viz. prices accurately reflecting production costs. Third, a return-seeking firm has at most the same output (likely less) than an analogous profit-maximising firm in all cases, and hence we can also intuit that it will underutilise capital in comparison.

The general case of the model is illustrated in Panel (b). Not only do we observe that $ATC(q)$ is above $MC(q)$ at the optimal firm output as per Proposition 4 and Proposition 6, we can see graphically the corollary that $ATC(q)$ must always be non-increasing up to, and including, the limiting case of a competitive price-taking firm. A broad emprical literature on firm cost structures is

[23] In general, these results hold for any number of output decision rules, not merely return maximisation or profit maximisation. Indeed, these results would hold for a random increase in quantity supplied in response to an increase in demand.

[24] Or indeed, any other input into production

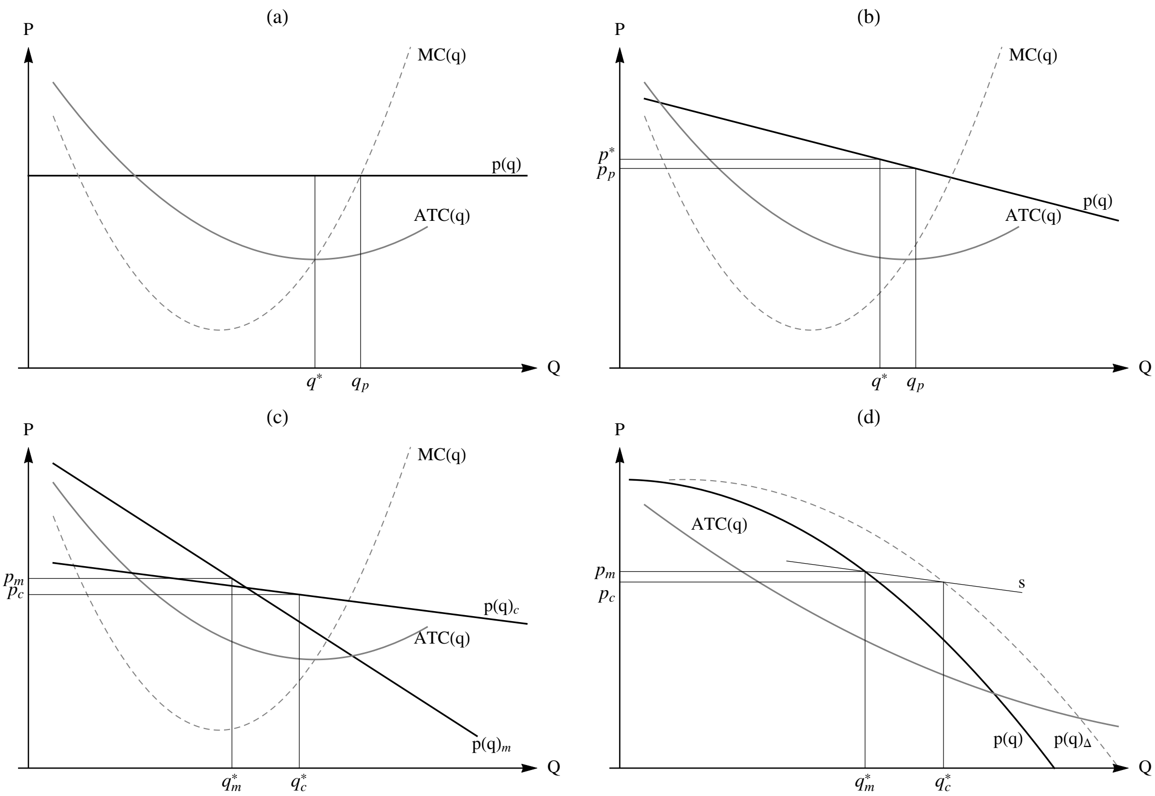

Figure 1: Various market structures with return-seeking firms

highly supportive of this model prediction (Walters, 1963; Blinder, Canetti and Lebow, 1998).

Panel (c) allows us to observe the relationship noted above between the slope of firm demand and its relative focus on cost minimisation compared to price setting. Compared to the cost-minimising amount we see the firm restricting output in response to its degree of price-setting power. As price-setting power decreases (as demand flattens from $p(q)_m$ to $p(q)_c$), the markup over costs decreases *ceteris paribus* while the firm shifts its focus from restricting output and increasing prices to minimising costs. Such a prediction is in keeping with a body of evidence showing that firms set prices by marking-up costs, and that the size of this mark-up is a function of their pricing power (Kalecki, 1952; Sraffa, 1926). As noted above, this is a particular beauty of the model in that we can analyse the dynamics of output choice of return-maximising firms as a response to increase competition. We take competition to manifest as a flattening of the firm demand curve due to the supply of close substitute goods, which we illustrate in Panel (c). Our model prediction on this point is somewhat consistent with profit maximisation models whereby competition reduce price and increases output. However, in our model competition is not solely dependent on the number of firms in a market, rather it can equally rest on consumer access to close substitute goods, regardless of the number of goods or number of firms producing them.

Since the shape of the firm demand curve is a result of the competitiveness of market, it takes little to derive that the model predicts return-maximising firms may also engage in behaviours that reduce competitiveness to augment or steepen their own demand curve. Common examples include product differentiation, loyalty schemes, or other methods of capturing market share. Alternatively, firms may lobby for patent protections or other regulatory barriers that restrict production of substitutes.

In Section 2.3 it was shown that return-maximising output increases in response to a positive shift in the demand curve (under certain fairly innocous assumptions), as can be illustrated by way of the response of q^* in Panel (d). However in Panel (d) we may also observe that it is not necessarily the case that an increase in demand will always result in an increase in price. Indeed, under certain conditions on the relative slopes of the demand and cost curves, demand increases may generate downward-sloping supply curves for an individual firm and for a market broadly defined. Such a result is expected from firms facing relatively steep demand curves, but extensive economies of scale yet to be utilised[25]. Thus supply can be broadly seen to respond positively to demand increases under non-competitive conditions, but while in this sense the supply curve slopes "up", the supply curve (defined as the relationship between output supplied and price) may actually be inverse.

We turn now to the impact of the quantised input and output space, graphically representing our model of quantised output and input space in Figure 2.

The top panel represents the firm's view of the market, while the bottom panel characterises its input choice problem of cost minimisation given a certain return-seeking output. We can project output choices in Panel (a) into the representation of the cost function in Panel (b). In Panel (a) we observe the

[25] A result which is consistent with the findings of Shea (1993)

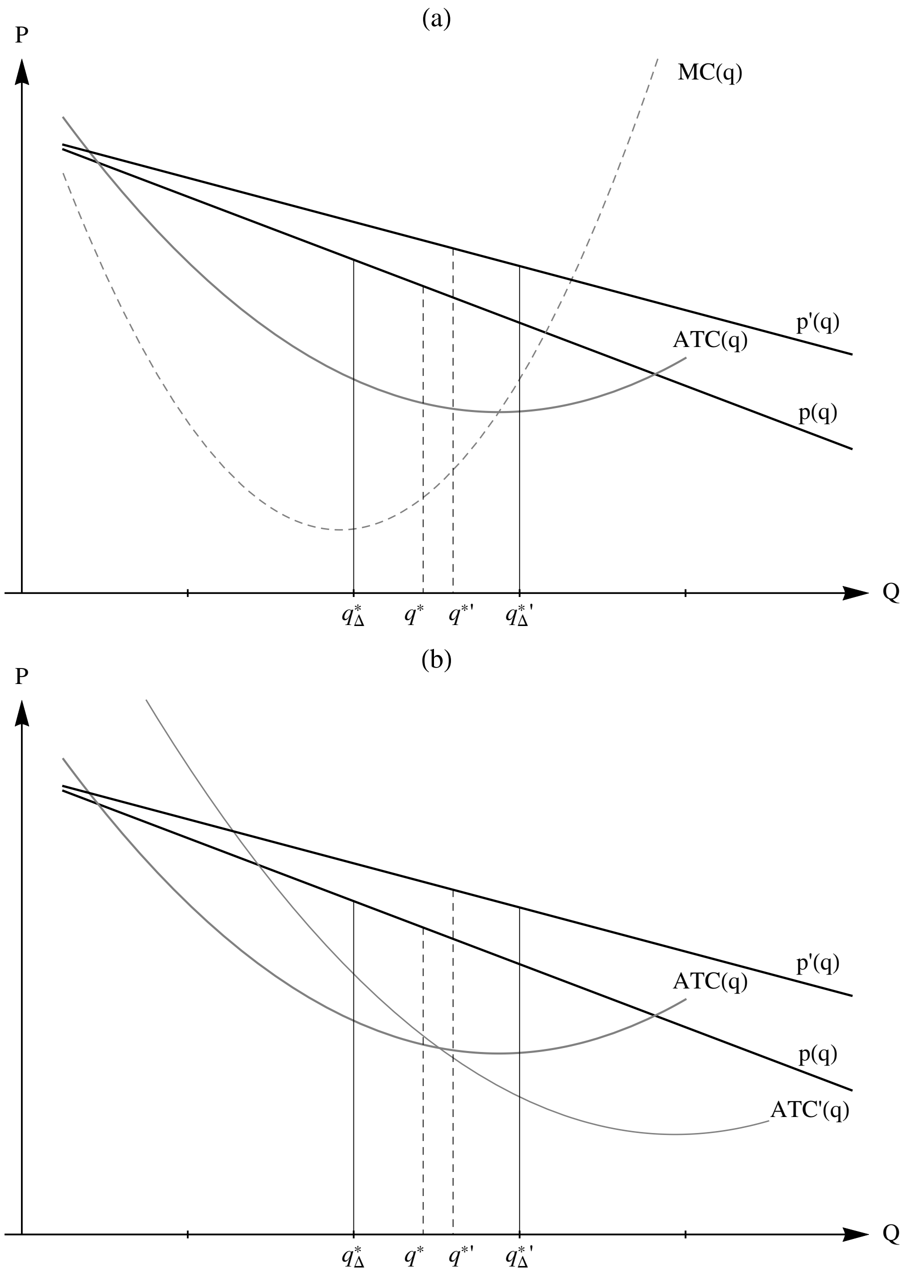

Figure 2: Quantised output and capital input choices facing return-seeking firms

result of quantising output space, where there exists a range a shifts in the demand curve that do not generate change in firm output from one point to another in the output grid. For an output change to occur we require a demand curve shift of a magnitude so as to increase the return-maximising level of output such that a quantum increase in the output supplied is justified. That is, the shift in demand must cause return-maximising output to pass a critical mid-point between to adjacent points in the output grid. This may seem like a trivial matter in many markets, but for firms producing large discrete units of goods, such as large scale manufacturing or construction, these considerations are crucial determinants of whether they commit to their next project or not.

We can also observe that markets do not clear in the traditional sense, and that the optimal level of output may lie between two points in the output grid. Therefore, depending on the exact placement of the optimal output we may have under or over supply relative to the "globally" optimal level, and hence the possibility of inventories in an otherwise perfectly rational market.

In Panel (b) of Figure 2 we examine the impact on firm output choices on choices of quantised capital input. The cost curves plotted here are level curves for quantum levels of the capital input, hence corresponding to costs for particular levels of the capital input in the quantised capital space. Projecting down from Panel (a), the return maximising problem and the cost-minimisation nested within it dictates which of these curves the firm finds itself on as an outcome of its capital input selection in response to any shift demand. We can observe that often a significant quantum increase in output is required to justify on the basis of minimising costs for a particular production level a quantum shift in choice of capital input.

Regarding the effect of interest rates on capital choice in the capital space, we note here that the cost curve corresponding to each discrete capital choice is a function of interest rates. Therefore, a change in interest rates may shift the return-maximising capital choice to another discrete choice, which is consistent with the long debated idea of capital re-switching were we to extend the model to include an alternative capital input with a differential market and price (Samuelson, 1966). We could also see at the market or economy-wide level that at the margin of firm's decision problems we may find a confluence of capital investment decisions, leading to cascades of capital investment as improved prospects for returns (through increased demand) justify quantum increases in capital investment. Hence the quantised nature of capital input spaces could go some way to explaining business cycles through surges in investment as critical points in capital input spaces are reached.

6 Conclusions

In this paper we have built a theory of firm behaviour methodologically consistent with conventional methods in economics on the empirically and theoretically solid assumption that firms maximise returns rather than profits. While this assumption alters ever so slightly the objective function of the theory (literally dividing profits by costs), the implications flowing from this theory are radically different to those of the implications of the profit maximising theory. Particularly in the limiting case of perfect competition, we no longer have a supply curve in any meaningful sense unless rather strong additional assumptions on entry are made, and we no longer can eliminate the possibility of mark-ups over marginal costs, and economies of scale become a necessary condition for a market to even exist. While firms in an imperfectly competitive market price above marginal cost and thus are pricing inefficiently by restricting supply relative to its competitive level, they at least respond to demand to satiate it (under certain conditions), and in this sense imperfectly competitive markets become a *more* efficient market structure than perfectly competitive markets.

Notwithstanding this failure of markets to clear, we also made output and capital input spaces discrete, reflecting the quantised nature of these variables imposed physically or institutionally in reality, and showed that quantum jumps in demand are required for supply, and for capital investment to respond, given that they will only do so when firms are within imperfectly competitive markets. This means that markets do not clear in any traditional sense, since "the" return-maximising level of output may lie within two technically possible quantised outputs, and likewise for capital inputs. The quantum constraint on output means that there can be under and over-supply, "cash left on the table" and inventories even when firms are maximising returns. The similar constraint on capital inputs allows us to understand under what conditions capital investment will in fact respond to changes in output, and captures the "lumpy" nature of capital, and why firms may not even be able to minimise costs in any global sense for a given output.

While some might suggest that these negations of many results from profit-maximising models are a weakness, we in fact regard them as a strength of the model. They explain many aspects of markets that we observe in reality. They explain why firms often seek to minimise average total costs regardless of their pricing strategy (and why this is intuitively what a firm would do). We can observe in the model how increasing competition leads firms to consider controlling prices relatively less to minimising costs in their decision rules. We also can see why markets will almost never clear, with inventories existing, and running "short" due to over-supply and under-supply respectively, even when firms and consumers are maximising. This model may negate many results of profit maximisation that have over the years become canon and perhaps viewed as indispensable, but in doing so it replaces these results with interesting new ones and opens up new lines of theoretical and empirical investigation to flesh out all the intricate subtleties of the return-seeking firm.

References

Berk, Jonathan, and Peter DeMarzo. 2013. *Corporate Finance.* . 3rd ed., Pearson.

Blinder, Alan S., Elie Canetti, and David Lebow. 1998. *Asking about Prices: A New Approach to Understanding Price Stickiness.* Russell Sage Foundation Publications.

Dixit, Avinash K., and Rober S. Pindyck. 1994. *Investment under uncertainty.* Princeton University Press.

Graham, John R, and Campbell R Harvey. 2001. "The theory and practice of corporate finance: evidence from the field." *Journal of Financial Economics*, 60(2–3): 187–243.

Jorgenson, Dale W., and James A. Stephenson. 1967. "Investment Behavior in U.S. Manufacturing, 1947-1960." *Econometrica*, 35(2): 169–220.

Kalecki, Michal. 1952. *Theory of Economic Dynamics: An Essay on Cyclical and Long-Run Changes in Capitalist Economy.* Allen and Unwin.

Kuh, Edwin. 1963. "Theory and Institutions in the Study of Investment Behavior." *The American Economic Review*, 53(2): 260–268.

Levine, David K., and Wolfgang Pesendorfer. 1995. "When Are Agents Negligible?" *The American Economic Review*, 85(5): 1160–1170.

Pindyck, Robert S. 1991. "Irreversibility, Uncertainty, and Investment." *Journal of Economic Literature*, 29(3): 1110–1148.

Samuelson, Paul A. 1966. "A Summing Up." *The Quarterly Journal of Economics*, 80(4): 568–583.

Shea, John. 1993. "Do Supply Curves Slope Up?" *The Quarterly Journal of Economics*, 108(1): 1–32.

Sraffa, Piero. 1926. "The Laws of Returns under Competitive Conditions." *The Economic Journal*, 36(144): 535–550.

Walters, Alan A. 1963. "Production and Cost Functions: An Econometric Survey." *Econometrica*, 31(1/2): 1–66.