Chapter 6
Firms

We have studied the static aspect and the dynamic aspect of the demand side (consumers) of the economy. We now study the supply side (firms) of the economy. As with consumers, we could separate our analysis of firms into distinct static and dynamic aspects. However, having become comfortable with static (single time period) versus dynamic (multiple time periods) analysis, we start right away with a version of firm theory that immediately features both dimensions. The basic lessons we will learn are the same as if we studied the two aspects independently.

We first study a small firm making its profit-maximizing decisions. The firm, existing in perfect competition, acts as a price taker in goods markets, labor markets, and capital markets. Based on its profit function, we can obtain, both intuitively and via formal optimization, its optimality conditions. The profit function is explicitly **dynamic,** though, in a way that is different from the usual microeconomic study of profit functions.

We will then switch our interpretation of the firm to an "aggregate" firm to focus on market prices and market quantities – in this respect, it is the **representative firm,** exactly analogous to the representative consumer. In terms of market outcomes, the representative firm's decisions provide the foundation for the **demand side of the labor market and the demand side of the capital investment market** (an important term yet to be defined). Hence, the term "supply side" of the economy mentioned above regards output markets. Layering the representative firm's optimization choices into a market in which the other side already exists allows us (in subsequent chapters) to finally consider general-equilibrium macroeconomic outcomes.

We adopt a multi-period view of firms because they make fundamentally dynamic (intertemporal) decisions. To keep things as similar as possible to the way we studied intertemporal consumer decisions, we will use the two-period setup that was the basis for the consumption-savings framework. Once we understand the model's setup, its analysis, and the main insights it provides, the entire analysis extends readily beyond two periods.

The discussion proceeds as follows. We begin by introducing some basic concepts regarding firms and production, focusing on how inputs get transformed into outputs. With these basics, we construct the dynamic profit function. Using the profit function, we develop the conditions that both formally and intuitively characterize a small firm's profit-maximizing choices of labor and capital. Studying capital decisions inherently requires details about intertemporal markets. We then switch interpretation to a representative firm, which allows us to consider market demand in the two key input markets of labor and capital. We close by returning to the starting point above of "static" versus "dynamic" aspects of firm-level decisions.

Firm Profit Function

Start with a single, atomistic firm (which will often be referred to as a "small firm") that maximizes its profits. The small firm takes prices as given in labor markets, capital markets, and goods markets. **The firm's profit-maximizing decisions occur at the beginning of period one, and these decisions encompass economic decisions for both period one and period two.** Because of the multi-period (two-period) nature of lifetime profits, the firm must maximize a **dynamic profit function.**

To build up to the dynamic profit function, we establish some basics. First, in each period, the small firm uses **labor and capital goods in order to produce final output goods.**[36] Condensing all of the inputs that a firm uses into the two categories of "labor" and "capital" is a useful and nearly-universal approach in macroeconomic analysis. Labor should be thought of in exactly the same way as in the static consumption-leisure framework.

Capital goods require more discussion. Capital goods are physical goods such as machines, factories, computers, delivery trucks, hair dryers, and so on that are used by firms in the production of other goods and services. A critical aspect of capital goods is that they are **accumulation (stock) quantities,** which means they take time to build up. If these are, for example, machines or factories, "building" is a very natural view. The way we formalize the "building" idea in the model is to take the simplest view of it: capital goods take one period to build. Which should be contrasted with, for example, labor: labor takes "zero periods" to build.[37]

Consider a simple example to illustrate: suppose a firm has zero units of capital at the beginning of period one. If the firm optimally desires a certain strictly positive level of capital for use in production in period two, it must sink resources into purchasing that capital during period one. It does so knowing full well that because of the one-period time-to-build aspect of capital, those purchases will not be ready for use until period two.

The **time-to-build feature of capital goods is the important point of distinction between capital and labor.** The "stock versus flow" differentiation between capital and labor is the main reason behind the nearly-universal adoption of these two distinct inputs in macroeconomic analysis. The timeline in Figure 30 displays capital goods (indicated by "k") on the junctions between time periods due to their accumulation nature.

[36] We include only labor costs and capital purchase costs. Broadly, any "other" costs that firms incur can typically be counted as either labor costs and/or capital purchase costs, so omitting "other" costs does not change any of the results.

[37] In richer applications that feature more than two periods, it is easy to think that the building of certain types of capital itself takes more than one time period. And, once built, the capital will last for multiple time periods. But this is an enrichment of exactly the framework being built here.

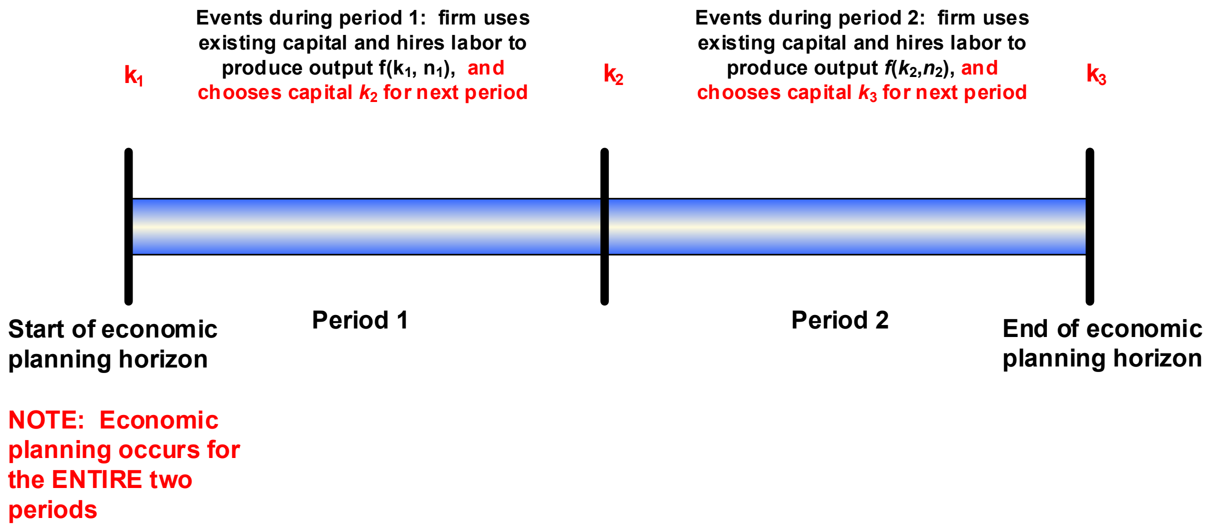

Figure 30. Timing of events for a firm in the two-period framework. Because capital, k, is an accumulation (stock) quantity, it appears on the junctions between time periods.

Second, as is common in microeconomic analysis of firms, the small firm uses labor and capital as inputs to a production function in order to produce output. The production function is represented as $y_1 = f(k_1, n_1)$ for period one and $y_2 = f(k_2, n_2)$ for period two.[38] An everyday example is that the coffee shop on the corner uses both workers (labor) and coffee machines (capital) to create, via some production process $f(k, n)$, coffee for its patrons.

Because we are ultimately adopting a macroeconomic view, we make two very important broad assumptions regarding production functions. First, production is assumed strictly increasing in each of the two input arguments. Second, and separately, production displays an ever-decreasing rate of transformation from each input individually into output. Stated in a different way, the latter property simply means: holding the quantity of capital input fixed, increases in labor input increase total output at an ever-decreasing rate; or (switching inputs), holding the quantity of labor input fixed, increases in capital increase total output at an ever-decreasing rate.

Formally, the two concepts correspond to a strictly **positive marginal product** in each input individually, and a strictly **diminishing marginal product** in each input individually. The term "marginal product" is important in firm analysis (you perhaps

[38] For ease of exposition, both here and elsewhere below, we will sometimes use the subscript notation t, rather than the subscripts "1" and "2," in cases in which there is no ambiguity about the time period that is meant. Hence, $y_t = f(k_t, n_t)$ is a shortcut representation of the production function in any period t.

already know its basic definition), but let's leave intuitive description of it until the analysis of first-order conditions below. But the simple idea is that a marginal product describes how much extra (marginal) output is generated by using a little extra (marginal) input.

To continue the formal description of these statements, we can also think via a strict calculus characterization, which is in terms of first derivatives and second derivatives. The strictly positive marginal product corresponds to restrictions on the first derivatives: $f_n(k_t, n_t) > 0$ and $f_k(k_t, n_t) > 0$. And the strictly diminishing marginal product corresponds to restrictions on the second derivatives: $f_{nn}(k_t, n_t) < 0$ and $f_{kk}(k_t, n_t) < 0$. The key words in all of the above are **positive** and **diminishing,** and these effects are all displayed in Figure 31. Unless stated otherwise (which could be the case in particular examples), this view of how inputs combine to produce output will be standard.

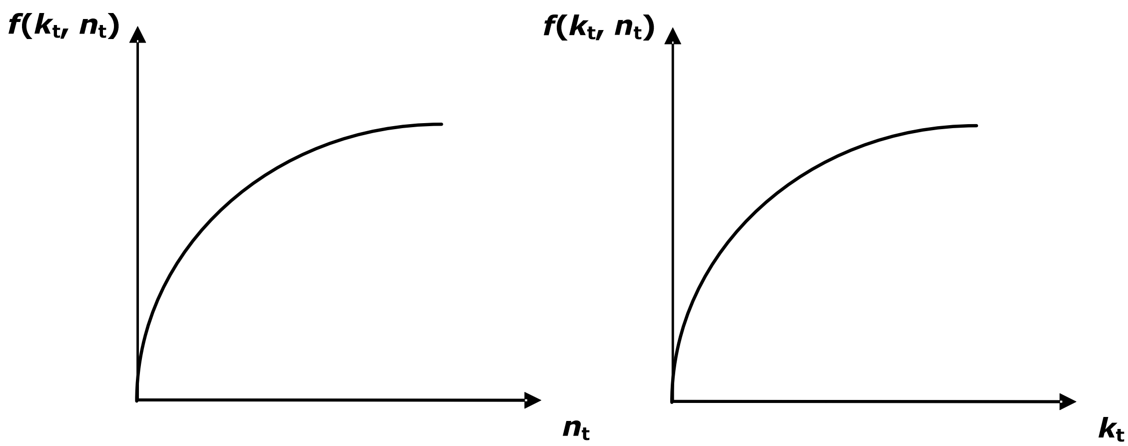

Figure 31. The production function $f(k_t, n_t)$ is strictly increasing in each of its two arguments labor and capital, and displays diminishing marginal returns in each of labor and capital individually. The left panel shows that as capital is held constant, increases in labor increase output at a diminishing (ever-decreasing) rate. The right panel shows that as labor is held constant, increases in capital increase output at a diminishing (ever-decreasing) rate.

With these basics established, the small firm's dynamic profit function (sometimes also referred to as the **lifetime profit function** or the **intertemporal profit function) from the perspective of the beginning of period one** is

$$\underbrace{P_1 f(k_1, n_1)}_{\text{Period-one revenues}} \underbrace{-P_1(k_2 - k_1) - P_1 w_1 n_1}_{\text{Period-one total costs}} + \underbrace{\frac{P_2 f(k_2, n_2)}{1+i}}_{\text{Period-two revenues}} \underbrace{- \frac{P_2(k_3 - k_2)}{1+i} - \frac{P_2 w_2 n_2}{1+i}}_{\text{Period-two total costs}}.$$

The profit function has the following notation: P_1 is the price of final goods during period one, P_2 is the price of final goods during period two, w_1 is the real wage during period one, w_2 is the real wage during period two, i is the nominal interest rate between period one and period two, k_1 is the firm's accumulated capital at the very beginning of period one, k_2 is the firm's accumulated capital at the very beginning of period two, and k_3 is the firm's accumulated capital at the very beginning of period three. The prices P_1, P_2, w_1, w_2, and i are all taken as given by the small firm.

The profit function is written in nominal terms (each term is denominated in units of currency), which makes the nominal interest rate i the appropriate one to use to discount period-two revenue and cost terms back to period one. We could instead represent profits in a purely real form, in which case the real interest rate r would be the relevant discount factor. Regardless of nominal or real, the terms that are discounted by the appropriate interest rate simply adjust for any changes in prices that may occur between period one and period two, all of which is taken as given by the small firm.

Also regardless of nominal or real, we have made the assumption that the market "sticker price" for physical capital in each period is identical to the market "sticker price" for output goods produced and sold by the firm in each period: P_1 in period one and P_2 in period two. Observe that there is **no** price here called, say, P_t^k, which would be the price in period t of a unit of capital goods as distinct from the price, P_t, of a unit of output goods. This assumption makes it appear that the distinction between capital goods and output goods is blurred. To the contrary, as we explain further below when we study the investment demand function, it actually enhances the critical distinction between capital goods and output goods.

The firm's choice variables, which are chosen at the beginning of period one, are the quantities of labor in periods one and two, n_1 and n_2, and the choice of the quantity of **capital investment** purchases in which to engage during the course of period one and during the course of period two. What we have specified "net investment" in the profit function, which are $k_2 - k_1$ and $k_3 - k_2$, respectively, rather than "gross investment." The details of this are described below, but regardless of net versus gross, the accumulation nature of capital is once again critical. As explained further below, **capital investment (net or gross) is the quantity by which the firm desires to change its capital level between consecutive time periods.**

Analysis

Formally, the next set of steps is to construct three first-order conditions.

First-Order Conditions

Based on the dynamic profit function, the first-order conditions with respect to n_1 and n_2 are, respectively,

$$P_1 f_n(k_1, n_1) - P_1 w_1 = 0$$

and

$$\frac{P_2 f_n(k_2, n_2)}{1+i} - \frac{P_2 w_2}{1+i} = 0,$$

in which $f_n(k_t, n_t)$ denotes the **marginal product of labor.** We introduced the term "marginal product" above. The marginal product of labor is simply the extra (marginal) quantity of output that results from the hiring and use of an extra (marginal) unit of labor input, holding all else (including capital input usage) constant.

Cancelling terms appropriately – in particular, cancelling the nominal price of output goods and the nominal interest rate – in each of these expressions gives $w_1 = f_n(k_1, n_1)$ and $w_2 = f_n(k_2, n_2)$.[39] The economic content of this statement is crucial: when the small firm is maximizing profit, it chooses its optimal quantity of labor in such a way that the **(real) marginal product of labor is exactly equal to the market real wage.** Nominal prices and nominal interest rates have nothing to do with this condition, on which we build further below.

Proceeding to the capital purchase decision, the only interesting decision in the two-period framework is with respect to k_2. This is for two reasons. First, as in the two-period framework that was the foundation of consumption-savings decisions, the firm knows that there is no period three, so it is (trivially) optimal for it to "choose" $k_3 = 0$.[40] Second, given the accumulation nature of capital goods, k_1 cannot be chosen at the beginning of period one – k_1 is instead "pre-determined" (intuitively, think of it as

[39] Stated more generally in terms of any period t, the condition that characterizes the optimal choice of labor is $w_t = f_n(k_t, n_t)$.

[40] Just as in the two-period consumption framework, the rational individual (here, the rational firm) will always choose to have zero wealth at the end of the economy. The capital (machines, equipment, etc.) is the wealth (i.e., the assets) that a firm owns.

reflecting choices that occurred in "period zero," which is outside our analysis). Which thus leaves only k_2 as a firm-level decision.

The first-order condition with respect to k_2 is

$$-P_1 + \frac{P_2 f_k(k_2, n_2)}{1+i} + \frac{P_2}{1+i} = 0,$$

in which $f_k(k_t, n_t)$ denotes the **marginal product of capital.**[41] Similar to the marginal product of labor, the marginal product of capital is simply the extra (marginal) quantity of output that results from the use of an extra (marginal) unit of capital input, holding all else (including labor input usage) constant.

Labor Demand

From the first-order conditions above and the general properties of the production function $f(k_t, n_t)$, we can establish the **labor demand function, which is a market-based relationship between the real wage and the optimal choice of labor.** Demand for labor is a **derived demand** because it arises due to (is derived from) market demand for the firm's output good.

The expression $w_t = f_n(k_t, n_t)$ characterizes the labor demand function. To see this graphically in a diagram of real wages and labor, consider the right-hand side of the expression. Because the second derivative is, by assumption, strictly negative ($f_{nn}(k_t, n_t) < 0$), the first derivative ($f_n(k_t, n_t)$, which **is** the marginal product of labor) is strictly decreasing as n increases. This argument is just an application of the properties of the second derivative function to understand something about the behavior of the first derivative function – less technically, this argument simply tells us that the marginal product of labor is becoming strictly smaller as n becomes strictly larger. Hence, **diminishing** marginal product.

Figure 32 plots the downward-sloping marginal product of labor schedule (named "*mpn*" in the diagram, a term that will be used interchangeably with $f_n(k_t, n_t)$) as a function of n. Because it emphasizes the qualitative properties on which we are largely interested, a downward-sloping straight line is sketched for the marginal product schedule. In standard macroeconomic applications, though, marginal product functions themselves

[41] To draw attention to a modification of this expression that is made clearer below: once we properly define net investment versus gross investment, another term appears in the first-order condition on k_2 if we formally consider gross investment. This more complete version appears in the Appendix, but the simpler case being studied here is just a special case of the more complete version, and none of the economic insights to come are changed by it.

typically have some convexity (that is, they are bowed in towards the origin).[42] Figure 32 also plots as a horizontal line the market-determined real wage *w* that the small firm takes as given.

For the small firm, *mpn* represents the marginal benefit of one more worker, and *w* represents the marginal cost of one more worker. Standard microeconomic results allow us to conclude that the optimal quantity of labor is exactly where marginal benefit equals marginal cost. This optimal quantity is *n** in Figure 32.

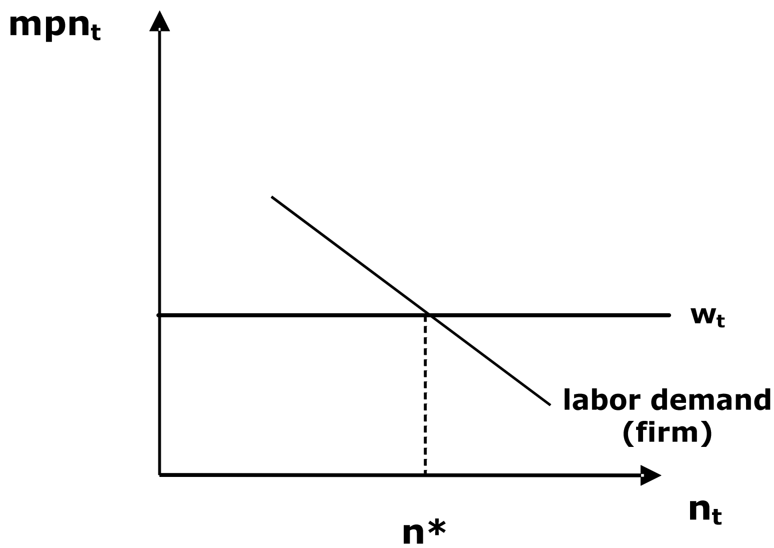

Figure 32. From the perspective of a small firm, the marginal product of labor and market real wage *w* as functions of the firm's own quantity of labor in any period *t*. The point at which *mpn* = *w* is the optimal choice of labor for the small firm, taking all else as given.

The sketch in Figure 32 is for a small firm. Moving to an aggregate view requires asking what happens to optimal labor when the market real wage w_t changes. As is clear from Figure 32, as the market real wage w_t declines, the small firm hires more labor; and, in the opposite direction, hires less labor if w_t increases. This intuitive relationship between movements in the market real wage and optimal labor is the **market labor demand function,** which is shown in Figure 33.

[42] A typical macroeconomic functional form for production functions is the Cobb-Douglas production function, $f(k_t, n_t) = k_t^{\alpha} n_t^{1-\alpha}$, in which $\alpha \in (0,1)$ measures the share (or the "importance") of physical capital in the production process. The parameter $1-\alpha \in (0,1)$ thus measures the share of labor in the production process. The share $\alpha \in (0,1)$ is almost always taken as a parameter, which will also be our view. (For example, in the U.S., a standard value is $\alpha \approx 1/3$, based on econometric evidence.) The associated marginal product of labor function, which is simply the partial derivative with respect to labor, is $f_n(k_t, n_t) = (1-\alpha) k_t^{\alpha} n_t^{-\alpha}$.

The two downward-sloping schedules in both Figure 32 and Figure 33 are termed "labor demand," which may appear confusing. Standard terminology is that the *mpn* schedule for a small firm is usually referred to as "labor demand," regardless of the market wage (which, note once again, is plotted as a horizontal line in Figure 32). For macroeconomic purposes, we are interested in market prices and market quantities, so plotting the relationship between the market wage w_t and the market quantity of optimal labor requires moving from the "small firm" view in Figure 32, to the aggregate view, or **representative firm** view, in Figure 33. Thus, Figure 33 is what we will usually mean when we refer to "the" market labor demand function, and it allows cleaner use of terminology; but many would also consider Figure 32 as showing the same result.

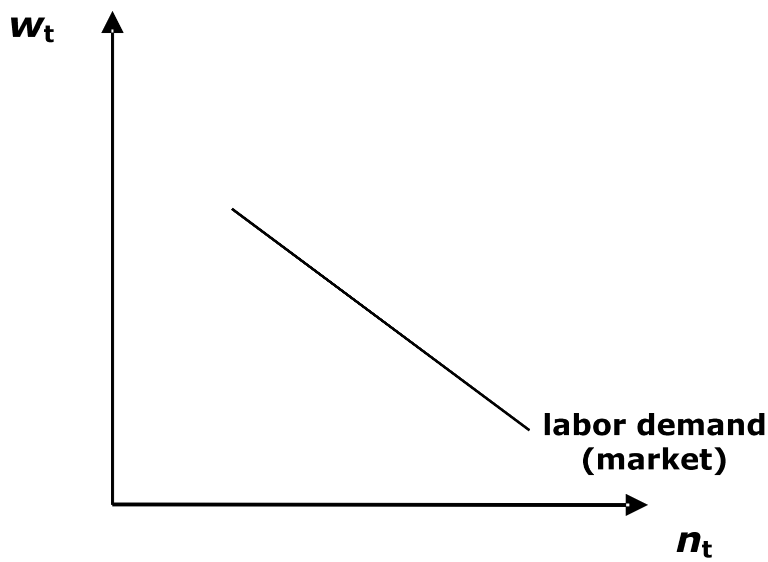

Figure 33. The market labor demand function depends negatively on the real wage.

Building Towards Capital Demand

Studying the small firm's decisions about capital purchases proceeds in a similar way as labor demand. As with labor demand, capital purchase decisions are also a derived demand for the firm, hence many big-picture points are very similar. The point at which we will eventually arrive below is that the **expression** $r = f_k(k_t, n_t)$ **characterizes the capital investment demand function,** which appears very analogous to the $w_t = f_n(k_t, n_t)$ expression that characterizes the labor demand function. However, getting there requires detailing three further issues, which are important.

First, the macroeconomic notion of **investment is that it measures the change in the quantity of capital between the start of one period and the start of the next.**[43] In the context of our two-period model, this change only has interesting meaning between period one and period two because, as noted above, the "optimal choice" regarding k_3 is that it trivially equals zero.

Net Investment and Gross Investment

First, there are two distinct concepts of investment: net investment and gross investment. We have included the former (net investment) in our dynamic profit function, so we start there. Mathematically, define the variable inv_1 (which will also be referred to as simply inv below because there is no ambiguity about time periods in the two-period model) to denote the change in capital between the start of period one and the start of period two,

$$inv = k_2 - k_1.$$

Formally, this is **net investment,** defined as the change in the level of capital between the start of the subsequent period (period two) and the start of the current period (period one). The terms "subsequent" and "current" inherit from the timing of the model: in the two-period model, profit-maximizing decisions, including investment decisions, are made at the beginning of period one.

Net investment does **not** include the wearing out of capital goods. Interpreting capital goods as, say, machinery, it is clear that capital goods naturally wear out due to use. **Economic depreciation** is the wearing out of capital goods due to use in the production process.[44] Rates of economic depreciation (often referred to as just "depreciation") are

[43] The term "investment" should be properly distinguished from a broader, colloquial, usage of the term. Formally, macroeconomic **investment** is the sum of business purchases of **capital goods** (goods that businesses use in the production of other goods and services), new homes built, and addition to firms' inventories. In everyday language, we often use the term "investment" to refer to someone's collection of stocks and bonds, as in "I invested in 100 shares of Microsoft stock last week." In formal economics, this latter type of activity is **not** investment. In fact, this latter type of activity is termed savings, a topic we have already studied. It is very important, however, to keep this terminology straight, as it is often a source of confusion when discussing matters of savings and investment. As we will see later, there is in fact a deep connection between macroeconomic savings and macroeconomic investment. For now, however, we are only considering the topic of investment. Because it is consumers (for the most part) that purchase homes, we see from the above definition of investment that investment encompasses activities of both consumers and firms. However, for convenience of exposition, we will simply speak of investment as being undertaken by firms only.

[44] This idea of wearing out of goods is, as the terminology emphasizes, the **economic** notion of depreciation, and it has nothing to do with any types of depreciation rules you may have learned in an accounting class. Accounting standards and regulations are such that sometimes a company has some control over how to *report* its depreciation of capital goods (i.e., accelerated depreciation, straight-line depreciation). Our economic notion of depreciation has only to do with how quickly goods *actually* wear out over time and nothing to do with how a company may choose to *report* how quickly goods wear out.

generally thought to vary by country, if we are interested in macroeconomic analysis. U.S. data show that roughly eight percent of the nation's capital stock depreciates every year, and this numerical value is fairly stable over long periods of time. Let this constant rate of depreciation ("constant" because the **rate** of depreciation does not vary between one time period and the next) be denoted by δ (the Greek letter "delta"), with the natural restriction that $0 \le \delta \le 1$. In the U.S. example, $\delta = 0.08$. Considering how a U.S. business is affected by depreciation: if it owns k_1 units of capital at the beginning of period one and it purchases zero new capital goods during the course of period one, then it will own $(1-\delta)k_1$ units of capital at the start of period two.

While net investment, *inv*, does not consider economic depreciation, the highly-related concept of **gross investment** does. Define gross investment inv^{gross} as

$$inv^{gross} = inv + \delta k_1,$$

which takes into account the replacement of depreciated capital through new investment purchases. Inserting the definition, $inv = k_2 - k_1$, from above allows us to rewrite gross investment as

$$inv^{gross} = k_2 - (1-\delta)k_1.$$

These two alternative expressions for inv^{gross} measure exactly the same idea: the first casts the relationship between two different notions of investment (gross versus net), and the second casts the relationship in terms of levels of capital.

A simple example illustrates the distinction between net investment and gross investment. Suppose a hair-drying salon (all they do is dry hair!) uses hair dryers as capital goods (part of their "machines and equipment"). The salon begins period one with k_1 hair dryers, and it purchases a total of x new hair dryers during period one. Of the hair dryers with which it began period one, some of them wore out during the period. At the end of the period, the salon thus has $(1-\delta)k_1 + x$ hair dryers, with which it will begin period two. Because it replaces the hair dryers that wore out, but also may want to expand the total number of hair dryers, **the total quantity *x* of new hair dryer purchases is gross investment.**

If the salon did want to expand the total number of usable hair dryers, the **net** addition to the number of hair dryers (net investment in hair dryers) during period one was smaller than x (gross investment in hair dryers). The reason is simply depreciation. In the example, total depreciation is δk_1, which is the number of hair dryers that wore out. The portion of x that just "replaces" the depreciated capital goods δk_1 does not actually expand the level of capital goods, it simply maintains the number of usable hair dryers.

This distinction leads to the definitional relationship between inv and inv^{gross} that appears above.

In quantitative applications, it is gross investment that is most relevant because that is what is measured by the official GDP and GDP-components accounts. However, note that the absolute gap between gross investment and net investment is fairly small, provided that the depreciation rate δ is fairly small, as it is for the U.S. and other advanced economies.[45] If we are mostly interested in qualitative, rather than quantitative, analysis and the insights qualitative analysis provides, a first-pass simplification is to completely ignore the difference. The reason that ignoring depreciation does not change any of the economic reasoning was (implicitly) stated above: the depreciation **rate** is constant over time, and moreover is thought to be outside the control of any firm. Hence, it is only **quantitative** results that could be different; none of the economic insights will be different.

Thus, for most of our qualitative analysis and resulting insights, we typically assume the depreciation rate is $\delta = 0$, in which case there is literally no distinction between inv^{gross} and inv. Indeed, $\delta = 0$ was already assumed in everything that has been thus far mathematically formalized. The Appendix (which will also be referred to below) nevertheless conducts a slightly-richer mathematical analysis including the δ term; to recover the analysis in the main text from the slightly-richer analysis in the Appendix simply requires setting $\delta = 0$.

Capital Goods and Output Goods are Physically Identical

The second issue is that, whether analyzing gross investment or net investment, we make an assumption that at first seems an excessive oversimplification. This assumption was pointed out briefly following the dynamic profit function above: **the market "sticker prices" for both capital goods and output goods are identical in each period.** The underlying economic assumption is that "final goods" (the goods that the firm produces and sells) are exactly identical to "capital goods" (the machines, etc. that firms use to produce final goods). Of course, this cannot literally be true – it is doubtful that the salon in the example above uses hair dryers to produce hair dryers.

But to understand this "identical sticker price" idea, recognize that there are quite a number of goods that have uses both in production of other goods and services, as well as end-user (final good) value in themselves. For example, about 80% of households in the U.S. own a personal computer – these would formally be counted as "consumption

[45] As noted above, $\delta = 0.08$ at an annual frequency for the U.S. is a common-used economy-wide figure. In many macroeconomic applications, quarterly data are studied (simply because many macroeconomic time series are made available at a quarterly frequency), which makes the economic depreciation rate about $\delta = 0.02$, which is also very small.

goods" in the GDP accounts.[46] But if you walk into any place of business, you are certain to see a computer on virtually every employee's desk – these would formally be counted as "capital goods." To make this even more stark, computer companies probably use a lot of computers (capital goods) in producing more computers (final goods)! Even if the computers themselves are exactly identical! So there are clearly some indistinguishable goods that are used as both capital goods and consumption goods.

There are of course also many goods that are usefully thought of partially as final goods and partially as capital goods. And there are also many examples of goods that are either one type or the other, but not both. If we want to admit such distinctions in the model (which is certainly possible), we would require a separate price, call it P_t^k, that would denote capital-goods prices in period *t*, distinct from the price for output goods, P_t.

While obviously realistic, raising such distinctions actually makes it harder to understand the conceptually most important "price" for capital goods, which is described next. Furthermore, while "realism" of a model is important for quantitative studies, sometimes "departing" from realism actually makes it easier to display the **economics** of an idea.[47] Thus, we simply make the assumption that consumer goods and capital goods are exactly identical goods – for concreteness, the computer example is helpful.

The very stark nature of the previous assumption then begs the question: what **is it** that actually is different about capital goods versus output goods?

The Importance of the Real Interest Rate *r*

The answer is the third issue: **the conceptually most important price for capital investment goods is the real interest rate *r*.** This point is critically, critically, critically significant, and it arises because capital investment is a fundamentally **intertemporal decision.** A firm decides how much capital it would like to have in the future, and the time-to-build nature of capital requires resources to be expended now.[48]

To understand this formally, we can see the "*r* price aspect" of capital based directly on the first-order condition on k_2. As shown in the Appendix, the first-order condition on k_2 can be rewritten as

[46] Source: http://www.marketingcharts.com/interactive/home-internet-access-in-us-still-room-for-growth-8280/nielsen-internet-access-household-income-February-2009jpg/.

[47] Which also is the motivation for considering net investment in the main analysis, rather than gross investment. The latter is more "realistic" quantitatively by adding depreciation explicitly into the firm profit function, but doing so does nothing to change the economic insights learned by studying net investment.

[48] As we first learned from studying consumption-savings decisions, the real interest rate *r* is the crucial intertemporal price. Switching to studying firm-level decisions, rather than consumer decisions, does not make the importance of *r* disappear. Rather, *r* is vital for the same reasons as it was when studying consumption-savings decisions: the centrality of intertemporal, or dynamic, decisions.

$$r = f_k(k_2, n_2).$$

This expression appears very different from the one presented above – the Appendix shows that getting to this representation simply requires several algebraic steps.[49] But this way of looking at the first-order condition on k_2 clearly displays a tight link with the analysis above of labor demand, in which $w_t = f_n(k_t, n_t)$ characterized the labor demand function. The simple and powerful economic result here, stated as an analogy with labor, is that **the capital investment decision is governed by a condition that is very similar to the labor demand condition.**

Digging deeper on the economics of this statement: if the firm is making its optimal choice about capital investment, it does so in a way that the marginal product of capital is exactly equal to something which it takes as given. The economics then reduces to: what is it that the firm takes as given?

When applied to labor, the answer is that the real wage was taken as given. In turn, the real wage is the appropriate notion (measured in real units) of the price of labor.

When applied to capital here, the answer is that the real interest rate is taken as given. By analogy with labor, **the real interest rate is thus the appropriate notion (measured in real units) of the price of capital investment.** Hence follows the conclusion that the critical price for capital investment goods is the real interest rate *r*.

If you have followed the arguments this far, it may raise the following chain of logic: interest rates are only relevant components of costs if a firm is borrowing in order to purchase capital (capital can be prohibitively costly, thinking in terms of machines and factories). Instead, if we consider a firm that had enough cash on hand to purchase capital outright without any need for borrowing, then the real interest rate seemingly would not be relevant for purchasing capital.

Unfortunately, this logic is incorrect. In terms of a multi-period (two periods or longer, it does not matter) analysis, the correct logic is **that it does not matter whether or not the firm is "borrowing" to purchase physical capital investment goods.** The real interest rate *r* is the relevant price of capital investment goods regardless of financial market arrangements.[50]

[49] Along with the assumption of $\delta = 0$; as the Appendix shows, for the more general case of $0 \leq \delta \leq 1$, the condition actually reads $r = f_k(k_2, n_2) - \delta$. But this slightly richer condition including δ blurs the main point being made here.

[50] To preview a point of departure later in the text, considering "financing frictions" may alter this result and hence intuition. Even more generally, other market frictions might layer on top of *r* other prices, taxes, regulations, and so on. But the centrality of *r* in the capital investment decision is critical in macroeconomic analysis. Thus, another way to state the result here, before we get to various models of

For the sake of terminology, it is easier to adopt the point of view of a firm that owns its own capital in both periods, which is what is embodied in the dynamic profit function with which we started.

To recap, the real interest rate r is the central price in capital investment decisions. The centrality of r is true regardless of whether we are considering net investment or gross investment. Gross investment is quantitatively richer because it considers depreciation of capital goods as they are used in the production process. But the economic insights generated by studying net investment, which is the main case in the text, are identical to the ones that emerge from studying gross investment.

From Capital Demand…

With all of these basics, we can turn to the small firm's optimal choice of capital. Figure 34 plots the downward-sloping marginal product of capital schedule (named "*mpk*" in the diagram, a term that will be used interchangeably, here and later in the text, with $f_k(k_t, n_t)$) as a function of k_2. As with labor demand, because it emphasizes the qualitative properties on which we are mostly interested, a downward-sloping straight line is sketched for the marginal product schedule. Figure 34 also plots as a horizontal line the market-determined real interest rate r that the small firm takes as given.

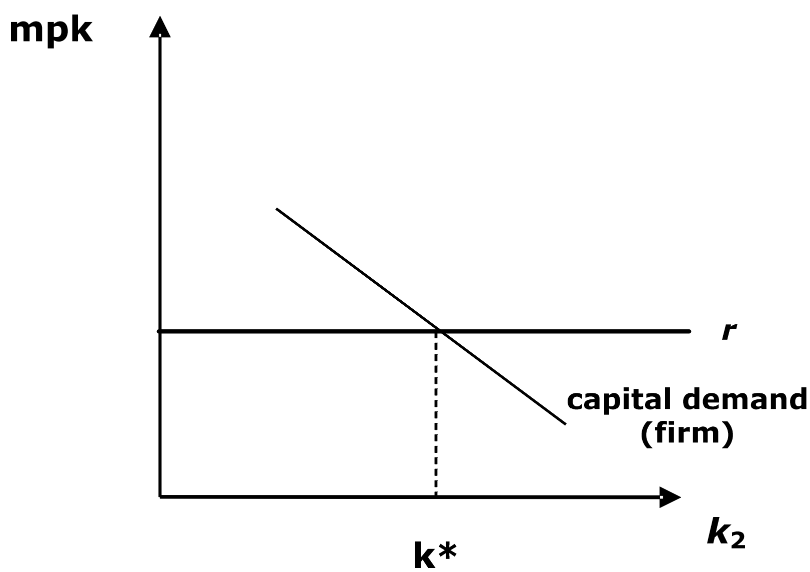

Figure 34. From the perspective of a small firm, the marginal product of capital and market real interest rate r as functions of the firm's own period-two desired capital stock. The point at which $mpk = r$ is the optimal choice of period-two capital for the small firm, taking all else as given.

market imperfections, is that **perfect competition in financial markets** (even though left unstated) is the basis for the result.

For the small firm, *mpk* represents the marginal benefit of one more unit of capital, and *r* represents the marginal (intertemporal) cost of one more unit of capital. Standard microeconomic results allow us to conclude that the optimal quantity of period-two capital is exactly where marginal benefit equals marginal cost. This optimal quantity is *k** in Figure 34.

…to Investment Demand (Part I)

We are virtually, but not completely, finished. Figure 34 displays the optimal desired quantity of capital, which is an accumulation measure; we are technically interested in the optimal quantity of investment, which is a flow measure. It is easy to make this connection.

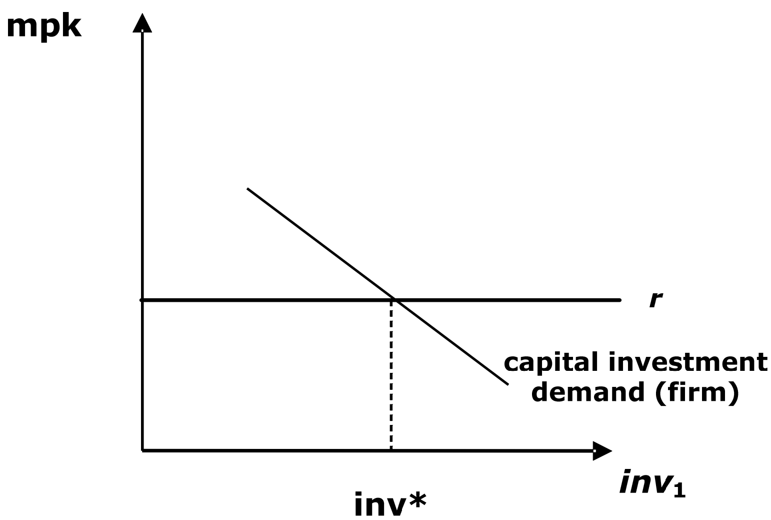

Figure 35. From the perspective of a small firm, the marginal product of capital and market real interest rate *r* as functions of the firm's own quantity of investment between desired period-two capital and period-one pre-existing capital. The point at which *mpk* = *r* is the optimal choice of capital investment for the small firm, taking all else as given

At the start of period one, the firm has k_1 pre-determined units of capital that cannot be changed. The firm's optimal desired quantity of capital at the start of period two is k_2^*. The definition of (net) investment, $inv = k_2^* - k_1$, shows that there is a one-for-one relationship between quantity of desired future capital and the quantity of current investment. With the assumption that $k_1 < k_2^*$ (because investment is nearly always positive at the macroeconomic level), we have the representation in Figure 35. Figure 35, which plots the optimal investment decision (which is a flow quantity during period one) of a firm, is qualitatively identical to Figure 34, which plots the optimal desired quantity of capital (which is an accumulation quantity in period two). In terms of economic intuition, there is nothing different between Figure 35 and Figure 34; but it is Figure 35 that is more immediately relevant for the subsequent analysis.

The sketch in Figure 35 depicts a small firm. Switching to an aggregate view requires asking what happens to optimal investment when the market real interest rate r changes. As is clear from Figure 35, as the market real interest rate r declines, the small firm invests more; and, in the opposite direction, invests less if r increases. This intuitive relationship between movements in the market real interest rate r and optimal investment is the **market capital investment function,** which is shown in Figure 36.[51]

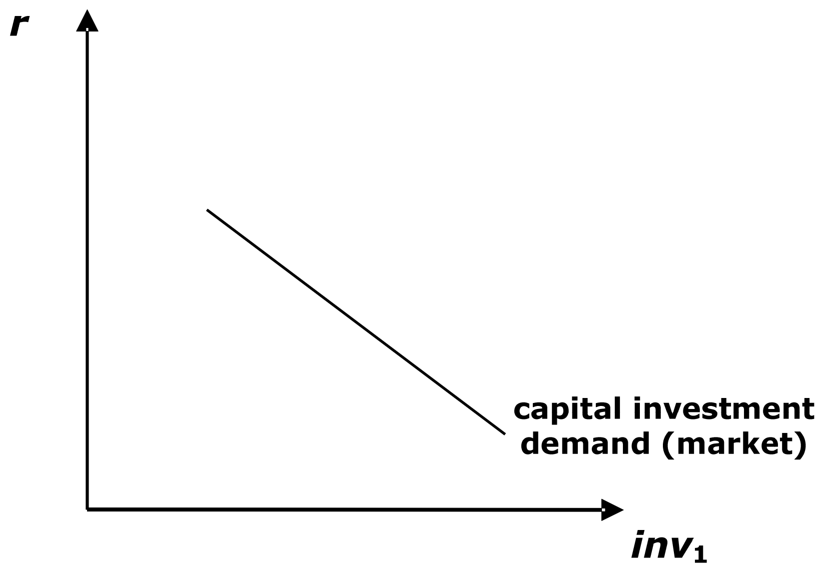

Figure 36. The market capital investment demand function depends negatively on the real interest rate.

[51] Similar to labor demand, the two downward-sloping schedules in Figure 35 and Figure 36 both are termed "investment demand." Standard terminology is that the *mpk* schedule for a small firm is usually referred to as "investment demand" (or even "capital demand") regardless of the market real interest rate. For macroeconomic purposes, we are interested in market prices and market quantities, so plotting the relationship between the market rate r and the market quantity of optimal investment requires moving from the "small firm" view in Figure 35 to the aggregate view, or **representative firm** view, in Figure 36. Thus, Figure 36 is what we will usually mean when we refer to "the" market investment demand function, and it allows cleaner use of terminology; but many would also consider Figure 35 as showing the same result.

Investment Demand (Part II)

The market capital investment demand function plotted in Figure 36 is the properly correct use of the term. Sometimes, though, it is not the Figure 36 representation, but instead a representation that represents P_1 on the vertical axis and optimal investment on the horizontal axis (and taking all other objects as given, most importantly P_2 and i) to which people refer as the investment function.

To consider why, the Fisher relation tells us that there is a direct relationship between P_1 and r,

$$1+r = \frac{1+i}{1+\pi_2} = \left(\frac{P_1}{P_2}\right)(1+i),$$

if we hold P_2 and i fixed. This leads to another graphical view of the investment demand function, shown in Figure 37.

There is nothing formally incorrect about Figure 37. But in terms of economic interpretation, it makes it appear that a "static" price leads to a downward-sloping investment demand function. This is somewhat misleading and misses the essence of the effort spent above establishing that it is the **intertemporal price *r*** that is relevant for investment demand. The Fisher relation does show exactly this – changes in P_1 lead directly to changes in r if P_2 and i are held constant. But it largely mutes the central point above that the intertemporal price r is to be thought of as the most important price when studying changes in investment.

However, the diagram in Figure 37 is useful because it allows horizontal summation with the consumption demand function derived in (either or both) the consumption-leisure model and the consumption-savings model. With a closed-economy view and leaving government decisions aside, what we thus generate is the aggregate demand function.

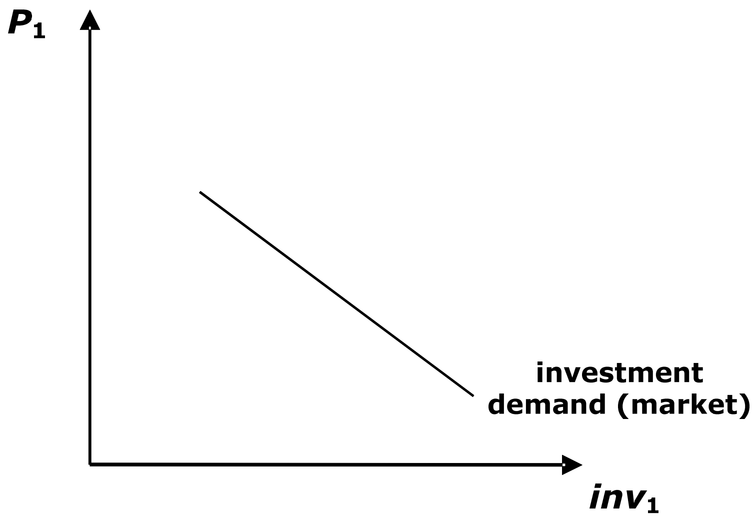

Figure 37. Capital investment demand from the perspective of nominal prices in a given time period, rather than from the perspective of the intertemporal price *r*.

Nonetheless, rarely will we consider Figure 37 what we have in mind when we use the term "investment demand function." Instead, what we mean is Figure 36.

Discussion

There are a number of important definitional and conceptual issues that arose as we considered a firm's profit maximization in a multi-period environment. But let's zoom back out to broad issues. The two main results of firm profit-maximization are the labor demand function in any given period and the capital investment demand function across any pair of periods. The labor demand function is depicted in Figure 33. The capital investment demand function is depicted in Figure 36.[52]

Returning to the very beginning of the chapter, we can think of labor demand and investment demand in terms of, respectively, static aspects of firm profit maximization versus dynamic aspects of firm profit maximization. To do so, let's depart slightly, but intuitively, from the formal mathematical statements presented earlier about $f(k, n)$.

Suppose that a firm's production process required zero capital inputs, and it required only labor input. This production process is "extremely labor intensive." The firm would hire labor in a profit-maximizing way, and the labor demand function would look identical to the one in Figure 33. What about investment demand? Investment demand would simply not exist, because the extremely-labor-intensive firm never needs to own any

[52] Rather than in Figure 37, which technically plots the same function in a different space; to reiterate the point made above, let's consider Figure 36 the nearly-universal way of thinking about the capital investment demand function.

capital. This extreme type of firm allows the "static" aspect of profit-maximization – labor demand – to arise in the same way as analyzed above. But it completely shuts down the "dynamic" aspect of profit-maximization – capital investment.

Instead, consider an "extremely capital intensive" firm that requires capital inputs but zero labor inputs in its production process. The firm would invest in capital in a profit-maximizing way, and the capital investment demand function would look identical to the one in Figure 36. What about labor demand? Labor demand would simply not exist, because the extremely-capital-intensive firm never needs any labor. This extreme type of firm allows the "dynamic" aspect of profit-maximization – capital investment – to arise in the same way as studied above. But it completely shuts down the "static" aspect of profit-maximization – labor demand.

In these extreme examples, labor demand is a "static" phenomenon (at one extreme) because it portrays a relationship between **only** period-t prices and period-t quantities – the real wage and the quantity of labor to hire. There is nothing inherently dynamic about the economic decisions in the model. At the other extreme, investment demand is a "dynamic" phenomenon because it portrays a relationship **only** between the real interest rate **across time periods** and how much investment to purchase **across those time periods.** There is nothing inherently static about the economic decisions in the model.

As we stated at the beginning, our analysis of firms combined both the static and dynamic aspects into one model, and the model generated the same broad results.

Appendix: Construction of capital demand function

The following shows how to convert the first-order condition on k_2 into a relationship between the real interest rate and the marginal product of capital. This derivation uses the slightly-richer version of the model, in which economic depreciation may occur to a firm's capital. This slightly-richer analysis does not affect any of the economic intuition described in the text.

With economic depreciation of a firm's capital at the rate $0 \le \delta \le 1$, the dynamic profit function is

$$P_1 f(k_1, n_1) - P_1(k_2 - (1-\delta)k_1) - P_1 w_1 n_1 + \frac{P_2 f(k_2, n_2)}{1+i} - \frac{P_2(k_3 - (1-\delta)k_2)}{1+i} - \frac{P_2 w_2 n_2}{1+i}.$$

Note where the δ term appears: in the investment term of the period-one component of profits, and in the investment term of the period-two component of profits. The first-order condition on k_2 is

$$-P_1 + \frac{P_2 f_k(k_2, n_2)}{1+i} + \frac{P_2(1-\delta)}{1+i} = 0.$$

The main analysis in the text is recovered by setting $\delta = 0$. The important economic insights remain the same as in the simpler version presented in the main text.

Starting from this richer first-order condition on k_2, divide by P_1, which gives

$$1 = \left(\frac{P_2}{P_1}\right)\left(\frac{1}{1+i}\right)\cdot f_k(k_2, n_2) + \left(\frac{P_2}{P_1}\right)\left(\frac{1-\delta}{1+i}\right),$$

in which we have also pulled the first term on the left-hand side over to the other side of the expression. Regrouping terms,

$$1 = \left(\frac{P_2}{P_1}\right)\left(\frac{1}{1+i}\right)\cdot\left(f_k(k_2, n_2) + 1 - \delta\right).$$

Next, using the definition of the one-period inflation rate, $1+\pi_2 \equiv \frac{P_2}{P_1}$, this can be written as

$$1 = \left(\frac{1+\pi_2}{1+i}\right)\cdot\left(f_k(k_2, n_2) + 1 - \delta\right).$$

Then, applying the Fisher relation, we have

$$1 = \left(\frac{1}{1+r}\right) \cdot \left(f_k(k_2, n_2) + 1 - \delta\right),$$

or, multiplying both sides by 1+*r* gives

$$1 + r = f_k(k_2, n_2) + 1 - \delta .$$

Canceling the "1" terms on each side gives the final expression,

$$r = f_k(k_2, n_2) - \delta ,$$

which can be plotted (provided some specific aspects of the function *f*(.)) in a diagram with r on the vertical axis and k_2 on the horizontal axis as the **capital demand function.** The capital demand function takes as given a constant value for the depreciation rate δ. The depreciation rate of $\delta = 0$ in the main analysis in the text is just a simpler case.

A common specification in macroeconomic applications is the Cobb-Douglas production function, $f(k,n) = k^{\alpha} n^{1-\alpha}$, so let's examine it. The parameter $\alpha \in (0,1)$ measures the capital share of output; hence $1-\alpha$ (which is also $\in (0,1)$) measures the labor share of otutput.

The Cobb-Douglas function has a marginal product of capital function (in period two, in particular, which is the period of interest here) is $f_k(k_2, n_2) = \alpha k_2^{\alpha-1} n_2^{1-\alpha}$. Substituting this in the previous displayed expression (and maintaining, for sake of completeness of the derivations in the Appendix, the parameter $0 \le \delta \le 1$),

$$r = \alpha k_2^{\alpha-1} n_2^{1-\alpha} - \delta .$$

Because $\alpha \in (0,1)$ and recalling the rules of negative exponents (observe that $\alpha - 1 < 0$), we can rewrite this as

$$r = \alpha \left(\frac{n_2}{k_2}\right)^{1-\alpha} - \delta ,$$

which simply characterizes the capital demand function for the Cobb-Douglas case.

Without appealing to the Cobb-Douglas or any other functional form, Figure 38 shows the capital demand function (at the microeconomic level) for the case of $0 < \delta < 1$.

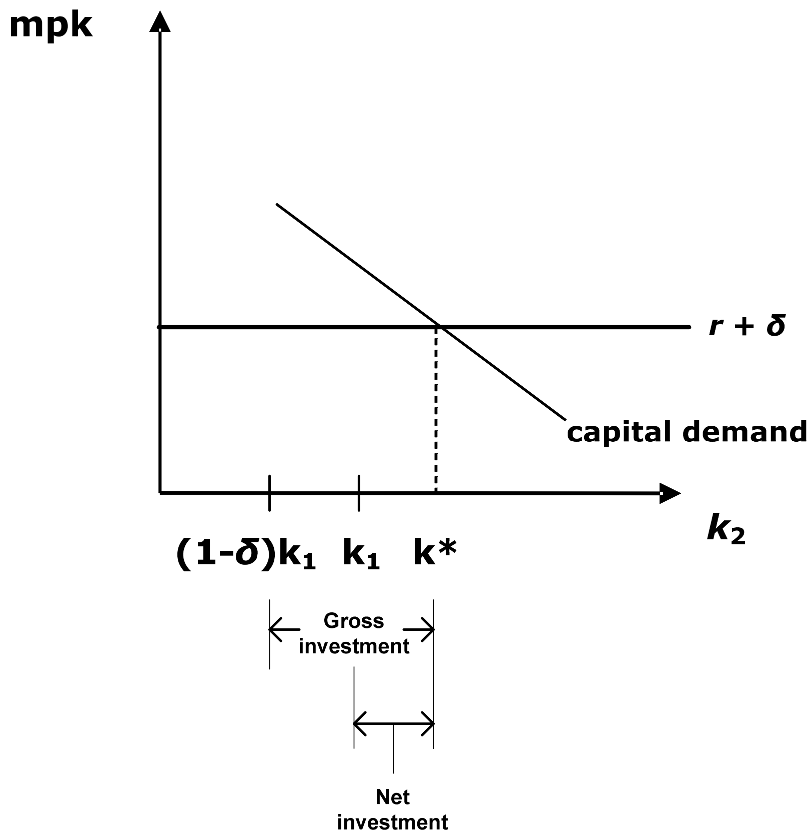

Figure 38. Net investment and gross investment as deviations of optimal period-two capital from, respectively, depreciated period-one capital and undepreciated period-one capital.