Linear Model in Economics

Part 1: Applied Prodcution Analysis

Tapas Kundu

2024-09-06

Section 1

Introduction

The first part of this course focuses on applied production analysis. We study microeconomic production theory and explore empirical applications. For the empirical component, we will be using R.

Supplementary reading materials include *Introduction to Econometric Production Analysis with R* (sixth draft version) by Arne Henningsen (Henningsen 2024, chaps. 1–5), and for the theoretical component, *Microeconomic Analysis* by Hal Varian (Varian 1992, chaps. 1–6).

The core question involves estimating production technology based on observational micro-level data of firms' behavior. This methodology can be applied in various contexts, including profit-motivated firms in the private sector, the performance of non-profit organizations, and across different industries such as agriculture, manufacturing, and services.

There are essentially four different methods used in the applied production analysis.

1. Least square methods for estimation of production functions
2. Total-factor productivity (TFP) indices
3. Data envelopment analysis (DEA)
4. Stochastic frontier analysis (SFA)

In this course, we will focus on the first two methods, which assume that all firms are technically efficient. In contrast, methods three and four provide measures of relative efficiency among a group of firms, which implicitly assume not all firms are technically efficient.

Efficiency and productivity

Efficiency and productivity are distinct technical concepts. The following diagram may help illustrate these distinctions.

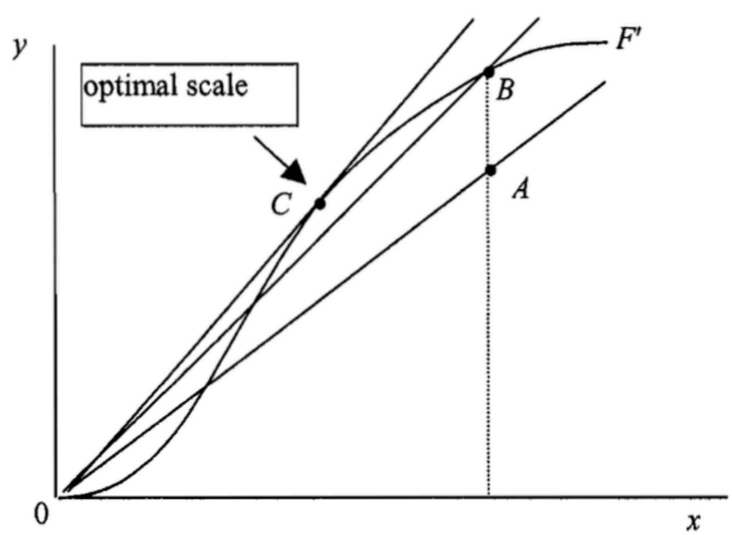


Figure 1: Productivity and technical efficiency

Consider a simple illustration of production method that converts an input x to an output y, and the relationship is given by the curve OF in Figure 1. There are three firms who are engaged in this productive activity, and their input-output combinations are given by A, B, and C, respectively. We can say that firm producing at A is technically inefficient compared to the firm producing at B, since the former is not producing as much output as the other one, even if both are using the same volume of input. By this argument, any production point below the OF curve are technically inefficient. Are all technically efficient points equally productive? Not necessarily! If we measure productivity by output produced per unit of input, then the firm producing at C has the productive use of its input.

Allocative efficiency is another important concept, which requires choosing the optimal combination of inputs to produce a given quantity of output. Allocative and technical efficiency together provide an overall efficiency measure.

Data set

We will be using a simple dataset for the empirical analysis of the theoretical concepts. The data set is shared in the *micEcon* R package (Henningsen 2005). The data set consists of production data of 140 French apple producers from the year 1986. These data are extracted from a panel data studied in (Ivaldi et al. 1996).

```
library(micEcon); library(psych)
```

```
If you have questions, suggestions, or comments regarding one of the 'micEcon' package
https://r-forge.r-project.org/projects/micecon/
```

```
data( "appleProdFr86", package = "micEcon" )
help("appleProdFr86")
dat <- appleProdFr86
rm(appleProdFr86)
head(dat, 5)  # A truncated preview of the data set
```

```
    vCap    vLab    vMat  qApples qOtherOut       qOut      pCap      pLab      pMat
1 219183 323991 303155 1.391803  0.977272 1374063.5 2.607756 0.899811 8.893672
```

```
2 130572 187956 262017 0.863985  1.072186 1122979.2 3.292048 0.752518 6.418985
3  81301 134147  90592 3.316798  0.404785 2158518.1 2.194315 0.956237 3.740465
4  34007 105794  59833 0.438498  0.436311  507389.2 1.602476 1.268100 3.167017
5  38702  83717 104159 1.831252  0.014982 1070815.7 0.866295 0.938286 7.221408
       pOut adv
1 0.6603515   0
2 0.7155490   0
3 0.9367494   1
4 0.5969751   1
5 0.8253856   1
```

```r
describe(dat)
```

```
          vars   n       mean         sd     median   trimmed        mad
vCap         1 140  102576.24   79992.28   84114.50   89202.88   55160.87
vLab         2 140  237199.39  194867.78  175871.00  199077.29   71095.12
vMat         3 140  201250.06  208054.52  136291.50  160457.37   92486.81
qApples      4 140       3.07       5.46       1.37       1.87       1.68
qOtherOut    5 140       1.50       1.32       1.07       1.29       0.95
qOut         6 140 2649825.38 3300778.29 1773989.17 2005998.00 1440430.19
pCap         7 140       1.30       0.79       1.11       1.20       0.56
pLab         8 140       1.01       0.20       0.96       1.00       0.20
pMat         9 140       6.77       2.64       6.25       6.55       2.75
pOut        10 140       1.01       0.53       0.83       0.91       0.31
adv         11 140       0.52       0.50       1.00       0.53       0.00
               min         max       range skew kurtosis        se
vCap       8951.00   452146.00   443195.00 1.90     4.23   6760.58
vLab      79569.00  1682201.00  1602632.00 4.10    23.24  16469.33
vMat      27781.00  1523776.00  1495995.00 3.15    13.50  17583.82
qApples       0.00       37.98       37.98 4.03    19.59      0.46
qOtherOut     0.01        5.96        5.95 1.44     1.59      0.11
qOut      95391.44 24210263.82 24114872.38 3.79    18.36 278966.68
pCap          0.17        4.48        4.32 1.66     3.71      0.07
```

pLab	0.49	1.72	1.23	0.42	0.29	0.02
pMat	2.70	15.71	13.01	0.85	0.64	0.22
pOut	0.46	2.94	2.49	1.66	2.10	0.05
adv	0.00	1.00	1.00	-0.08	-2.01	0.04

The data frame contains the following columns:

vCap: costs of capital (including land).

vLab: costs of labour (including remuneration of unpaid family labour).

vMat: costs of intermediate materials (e.g. seedlings, fertilizer, pesticides, fuel).

qApples: quantity index of produced apples.

qOtherOut: quantity index of all other outputs.

qOut: quantity index of all outputs (not in the original data set, calculated as 580,000 (`qApples` + `qOtherOut`)).

pCap: price index of capital goods pLab: price index of labour.

pMat: price index of materials.

pOut: price index of the aggregate output (not in the original data set, artificially generated).

adv: dummy variable indicating the use of an advisory service (not in the original data set, artificially generated).

Note that the firms were engaged in multi-output production. Analyzing this can be challenging, as firms optimize the mix of both inputs and outputs based on relative returns. As discussed in (Ivaldi et al. 1996), non-specialization was a crucial aspect of the agricultural production process.

Primal versus dual approach

Duality is a fundamental concept in optimization, particularly in linear programming and game theory. The essence of duality is that every optimization problem (referred to as the *primal* problem) can be associated with a corresponding *dual* problem, where the solution to one provides bounds to the solution of the other.

Since the study of economics draws heavily on optimization theory, it is no surprise that duality plays a significant role in economics as well. For example, in the context of producer behavior, the primal approach involves studying how firms can optimally decide on the input mix for a given production technology to achieve an objective, such as minimizing expenses to produce a certain volume of output. The resulting optimal expense is referred to as the cost function, which characterizes the minimum expenditure needed to produce a specific quantity. Duality tells us that the cost function is sufficiently informative, allowing us to confidently trace back the production technology under mild conditions.

Prior to 1970, economists mostly followed Samuelson's classic treatment of profit-maximizing firms, where firms face technological constraints, typically modeled with a smooth production function, and standard optimization techniques are used to infer producer responses to price perturbations. This approach is often referred to as the primal approach. Later, the dual approach gained prominence, where exploring cost, profit, or revenue functions allows us to trace back the technological constraints.

In this course, we will first explore the primal approach, both theoretically and empirically. Later, we will conduct similar exercises using the dual approach.

Production technology

The set of all combinations of inputs and outputs that comprise a technologically feasible way to produce is called a ***production (possibility) set***.[1]

The function describing the boundary of this set is known as the ***production function***. It measures the maximum possible output from a given amount of input.

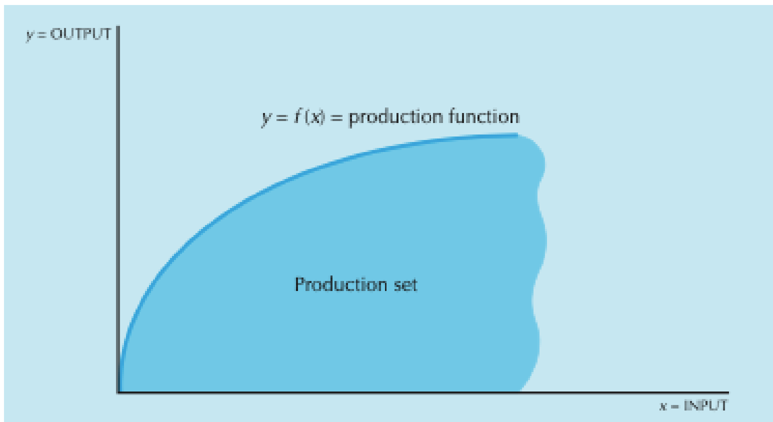


Figure 2: Production set and production function

As discussed before, a point in the interior of the production set represents a case of technically inefficient production. In the two-input case there is a convenient way to depict production relations in the form of an isoquant or indifference curve. An ***isoquant*** is the set of all possible combinations of inputs $(x_1, ..., x_n)$ producing a given amount of output y.

The isoquants move in the top-right direction as y goes up, since we need more inputs to produce more output. The top-right section of the isoquants, and including the points on the isoquants, are often referred to as the ***input requirement set***. Observe that for a given volume of output y, the input requirement set consists of all points on all the isoquants corresponding to the output level y or higher.

[1]You might notice variations in how this set is represented in different books. For example, in Figure 2, we include vectors (y, x), where y represents output and x represents input. In some cases, the production set is defined as the collection of all (y,−x), where the negative sign indicates the use of inputs, and the positive sign indicates the production of output; see, for example,(Varian 1992).

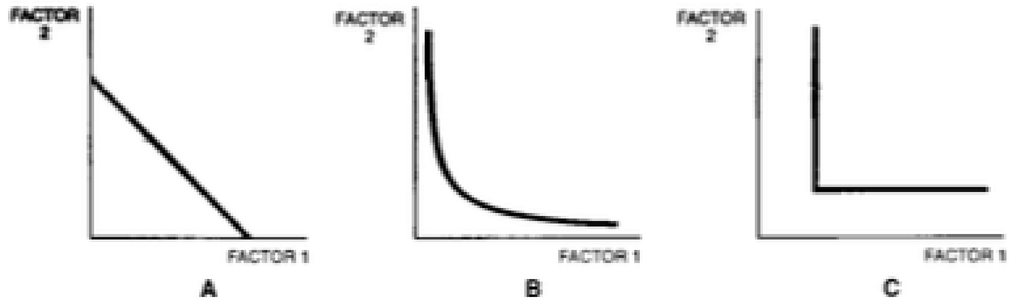


Figure 3: Isoquants—Linear, Cobb-Douglas, and Leontief production technology

A technology is called ***convex*** if the input requirement set is convex. For a convex technology, a convex combination of input choices increases the output volume.

A technology is called ***monotone*** if its input requirement set satisfies the *monotonicity* property, which suggests that for any input vector $\mathbf{x}$ belonging to the input requirement set, all input vectors weakly greater than $\mathbf{x}$ must belong to the input requirement set. The idea is that if we increase the amount of each input beyond what is required to produce a certain volume of output, we can produce an output at least as large as the initial volume.

While a production function is a useful way to characterize the production possibility in one-output case, a general representation of multi-output and multi-input production possibility is given by a ***transformation function*** $T : R^{n+m} \to R$ such that $T(\mathbf{x}, \mathbf{q}) = 0$ represents a relationship where an input vector $\mathbf{x}$ is used to produce an output vector $\mathbf{q}$.

Some examples of useful production functions

Linear: $y = \beta_0 + \sum_{i=1}^{N} \beta_i x_i$

Cobb-Douglas: $y = \beta_0 \prod_{i=1}^{N} {x_i}^{\beta_i}$, or equivalently, $\ln y = \beta_0 + \sum_{i=1}^{N} \beta_i \ln x_i$

Leontief: $y = \min_{i=1}^{N} \{\beta_i x_i\}$

CES: $y = \left[\sum_{i=1}^{N} \beta_i x_i^{\rho}\right]^{\frac{1}{\rho}}$

Quadratic: $y = \beta_0 + \sum_i \beta_i x_i + \frac{1}{2}\sum_i \sum_j \beta_{ij} x_i x_j$

Translog: $\ln y = \beta_0 + \sum_i \beta_i \ln x_i + \frac{1}{2}\sum_i \sum_j \beta_{ij} \ln x_i \ln x_j$

Returns to scale

Consider the following experiment: Let's scale the amount of all inputs up by some constant factor k; what will happen to the output?

If the output goes up by the same factor k, we call it a ***constant returns to scale (CRS)*** technology. Mathematically, a CRS technology exhibits $f(k\mathbf{x}) = k\mathbf{f}(\mathbf{x})$.

If the output increases less than k times, we call it a ***decreasing returns to scale (DRS)*** technology. Mathematically, a DRS technology exhibits $f(k\mathbf{x}) < k\mathbf{f}(\mathbf{x})$.

If the output increases more than k times, we call it an ***increasing returns to scale (IRS)*** technology. Mathematically, an IRS technology exhibits $f(k\mathbf{x}) > k\mathbf{f}(\mathbf{x})$.

Test exercise: Consider the following Cobb-Douglas production function is given by $f(x_1, x_2) = Ax_1^a x_2^b$. Find conditions under which the technology exhibits different kinds of returns to scale.

Productivity

Average and marginal product

Single-input case:

Consider a production relationship given by $y = f(x)$.

The ***average productivity*** of the input x is defined by $AP = f(x)/x$.

The ***marginal productivity*** of the input x is defined by $MP = \partial f(x)/\partial x$.

Multi-input case:

As in the single-input case, we can define the average product or marginal product of the i^{th} input with respect to each inputs. However, these measures then reflect simply partial productivity measures, and they can only be computed for some given values of other inputs.

$$
\begin{aligned}
AP_i &= \frac{y}{x_i} = \frac{f(\mathbf{x})}{x_i} \\
MP_i &= \frac{\partial y}{\partial x_i} = \frac{\partial f(\mathbf{x})}{\partial x_i} = f_i
\end{aligned}
$$

Output elasticity

The ***output elasticity of an input***x_i measures the percentage changes in output because of a percentage change in input x_i.

$$
\varepsilon_i = \frac{\partial f(\mathbf{x})/f(\mathbf{x})}{\partial x_i/x_i} = \frac{MP_i}{AP_i}
$$

Observe that output elasticities are free of the unit of measurement.

The ***elasticity of scale*** is the sum of output elasticities of all input: $\varepsilon = \sum_i \varepsilon_i$.

A technology exhibiting IRS, CRS, and DRS has the elasticity of scale $\varepsilon > 1$, $\varepsilon = 1$, and $\varepsilon < 1$, respectively. Using calculus, it can be derived that if a firm has an elasticity of scale as 1 at its current size of production and if the elasticity of scale only monotonically decreases with further increase in size, then the firm has the most productive scale size at the current level.

Total factor productivity

In multi-input production process, it is often desirable to calculate the ***total factor productivity (TFP)*** by aggregating inputs into an input index:

$$TFP = \frac{y}{X},$$

where X is a quantity aggregating index of all inputs.

Indexing

Indexing is used for measuring changes in a set of related variables. Conceptually, it can be used for comparison over time or space or both. Examples include price indices for measuring changes to consumer price, export or import prices, quantity indices measuring changes in output volume by a firm or industry over time or across firms.

As an illustration, consider a formula for measuring the change of the value of a basket consisting of n goods between the two period t and s can be measured by

$$X = \frac{\sum_{i=1}^{n} x_{it} p_{it}}{\sum_{i=1}^{n} x_{is} p_{is}}.$$

However, as time changes between s and t, it is unclear whether the change in value is driven by the changes in p_i or changes in x_i. To address this issue, we can fix one of the two variables, and look at the value index. For example, if we fix the prices (either to current or old prices), we get a measure due to changes in quantity, and it then reflects a quantity index. Similarly, if we fix the quantity (either to current or old quantity levels), we will get a price index. Although we consider changes with respect to time, we can use the concept for other types of changes, for example, variation across firms.

Various (quantity) indices:

Denoting the good by subscript i, the sample observation by subscript j, and a base observational value (for example, the mean of the sample observations) by 0, we measure

Laspeyres quantity index:

$$X_j^L = \frac{\sum_i x_{ij} p_{i0}}{\sum_i x_{i0} p_{i0}}$$

Paasche quantity index:

$$X_j^P = \frac{\sum_i x_{ij} p_{ij}}{\sum_i x_{i0} p_{ij}}$$

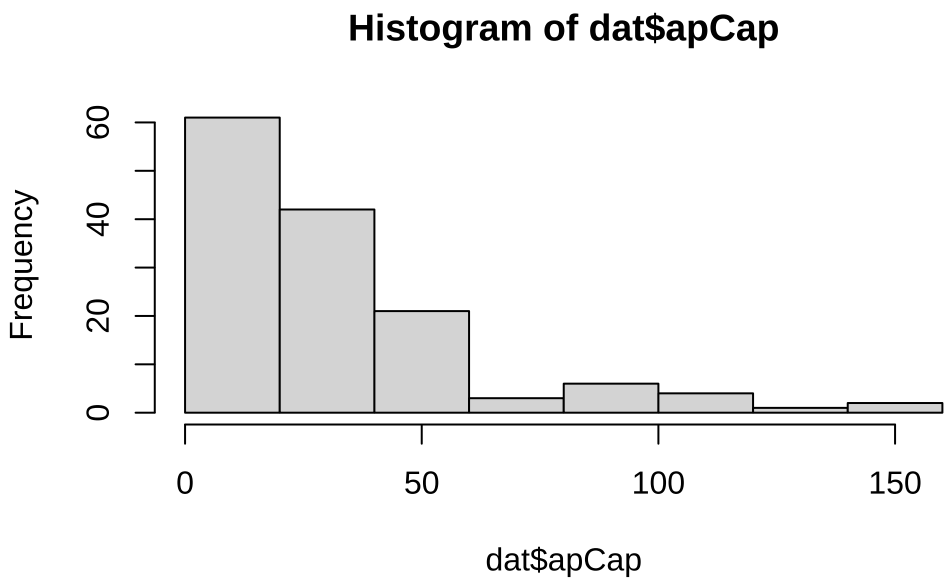
Fisher's quantity index:

$$X_j^F = \sqrt{X_j^L \times X_j^P}$$

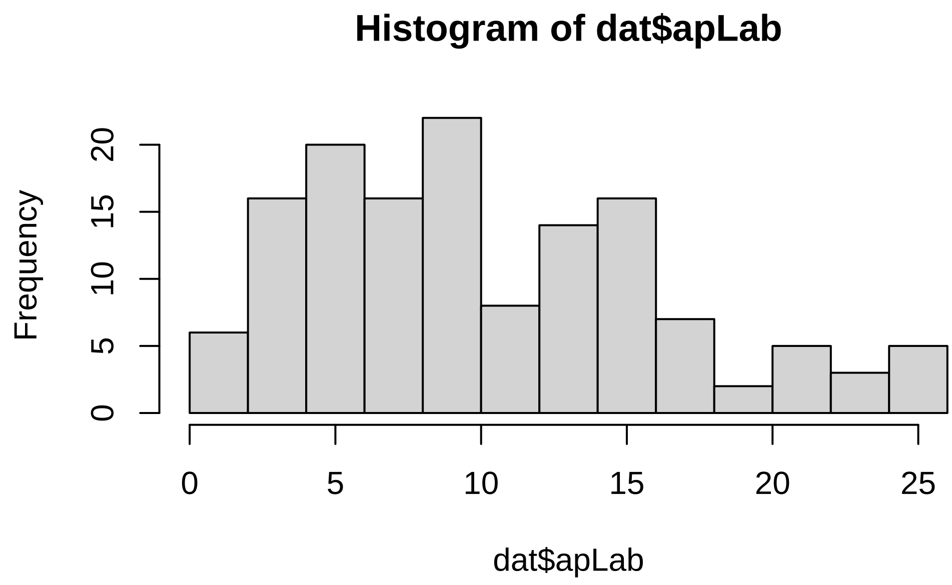
```
# Generate input quantities
dat$qCap <- dat$vCap / dat$pCap
dat$qLab <- dat$vLab / dat$pLab
dat$qMat <- dat$vMat / dat$pMat
#
# Creating quantity indices
dat$XP <- with( dat, ( vCap + vLab + vMat ) / ( mean( qCap ) * pCap + mean( qLab ) *
dat$XL <- with( dat, ( qCap * mean( pCap ) + qLab * mean( pLab ) + qMat * mean( pMat
dat$X <- sqrt( dat$XP * dat$XL ) # Fisher Index
# You can also generate these indices directly using micEconIndex package
```

Data: AP and TFP

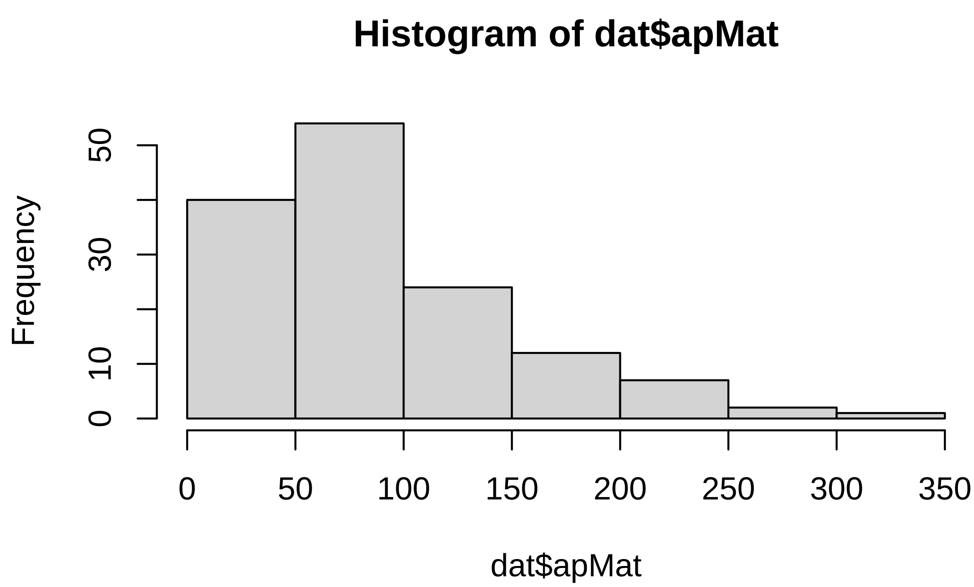
```
# Measuring (partial) average product
dat$apCap <- dat$qOut / dat$qCap
dat$apLab <- dat$qOut / dat$qLab
dat$apMat <- dat$qOut / dat$qMat
hist( dat$apCap )
```




```
hist( dat$apLab )
```

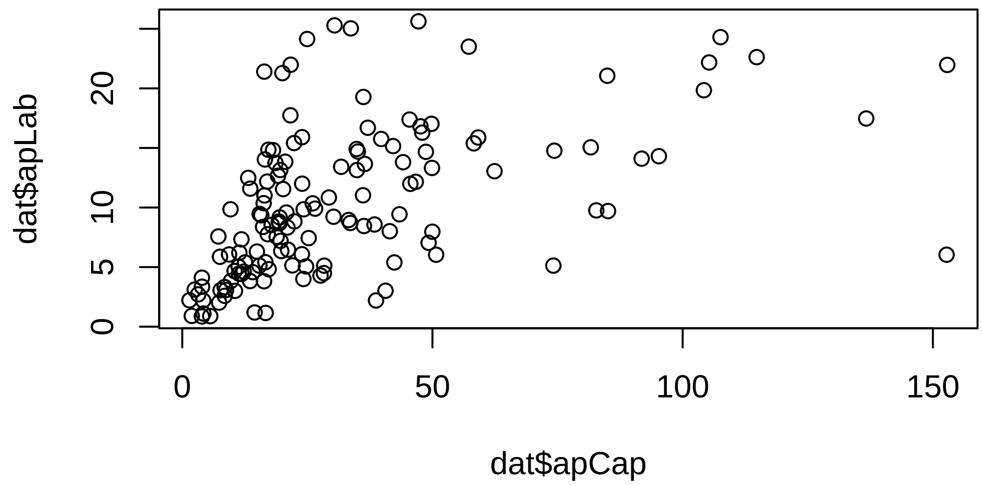



```
hist( dat$apMat )
```

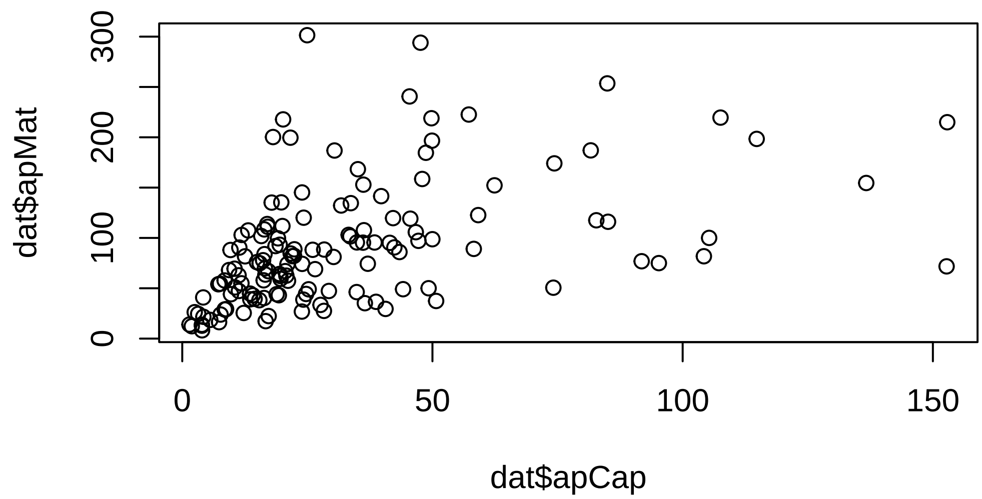


Average product measures vary considerably across firms, with most firms falling into the relatively low-productivity range.

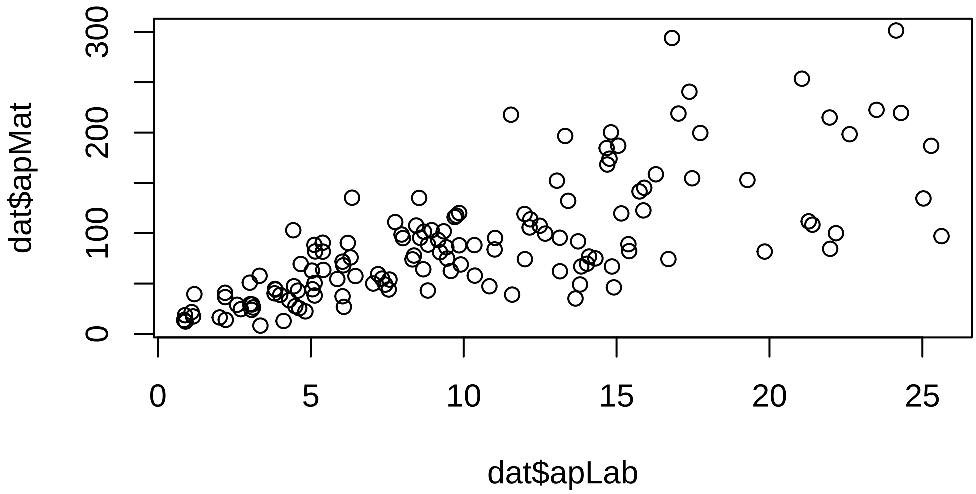
```
# Plotting average partial productivity of one input against another across firms
plot( dat$apCap, dat$apLab )
```




```
plot( dat$apCap, dat$apMat )
```

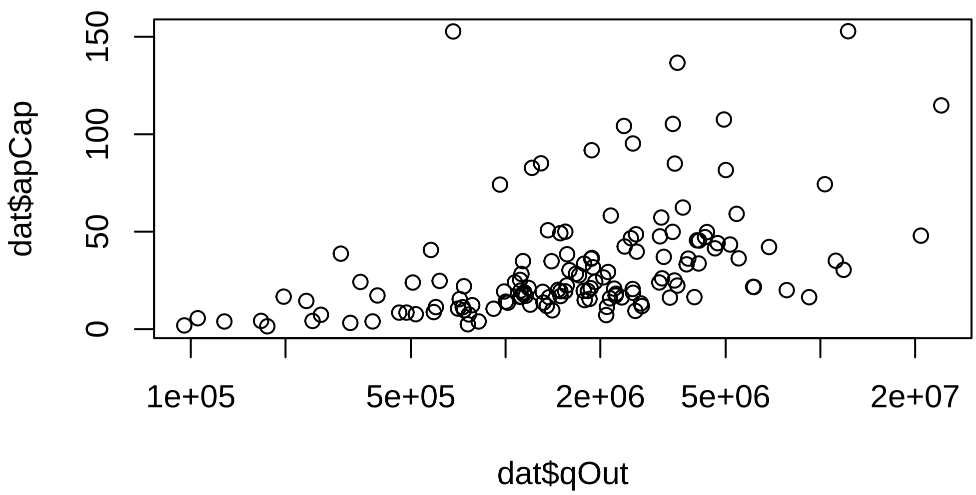



```
plot( dat$apLab, dat$apMat )
```

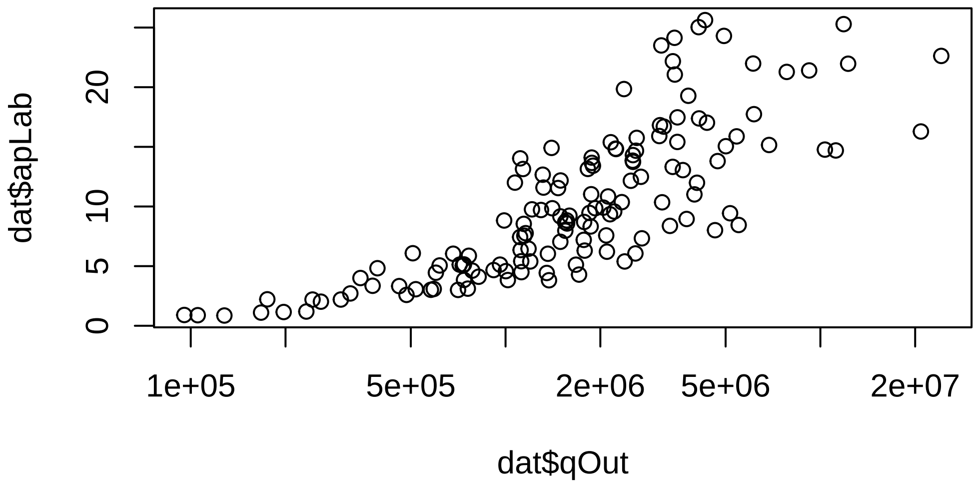


It appears that the average products of the three inputs are positively correlated.

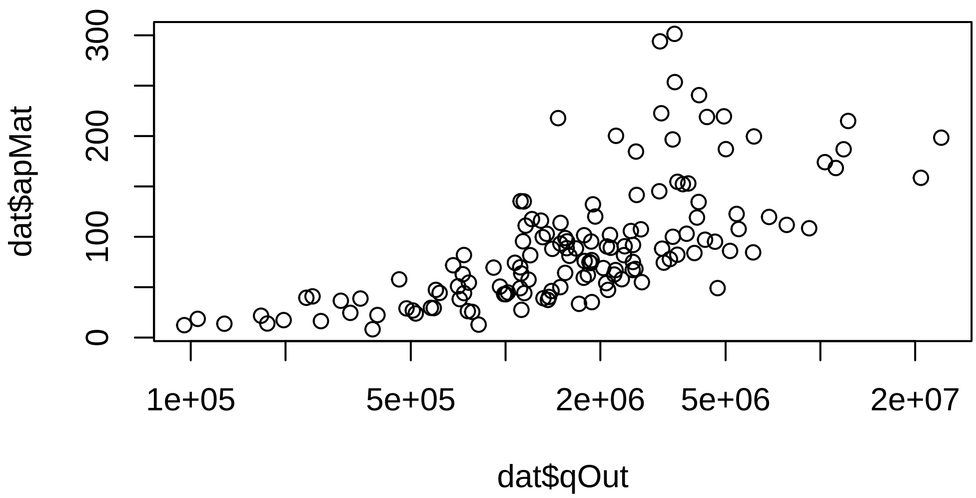
```
# Plotting partial average products against output
plot( dat$qOut, dat$apCap, log = "x" )
```




```
plot( dat$qOut, dat$apLab, log = "x" )
```



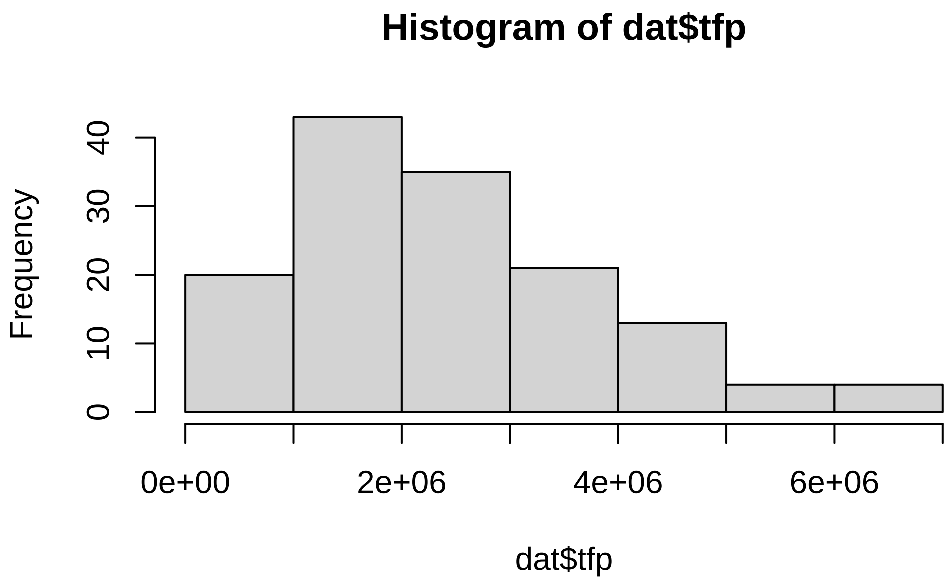

```
plot( dat$qOut, dat$apMat, log = "x" )
```



We did not have data on firm size. Assuming the volume of output as a proxy for firm size, we examined the plot of partial average products of each input against output. It appears that firms producing more also exhibit higher output per unit of input used.

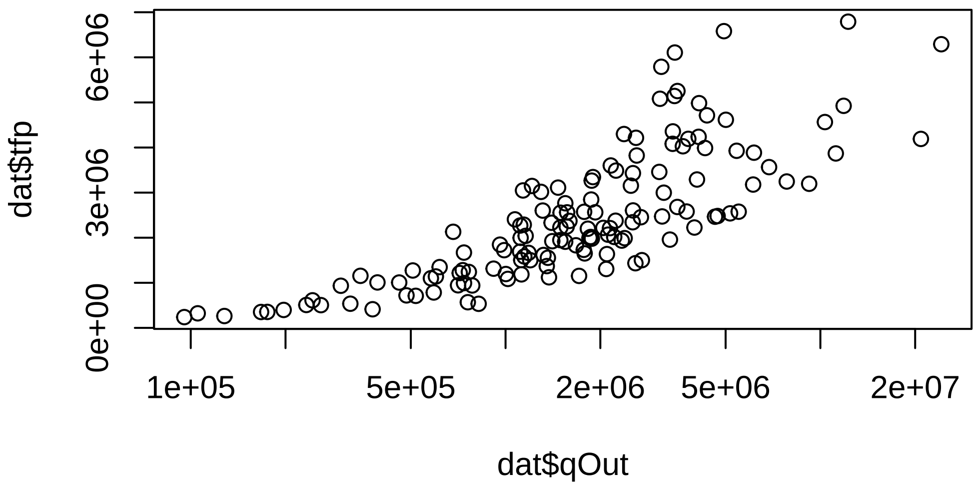
```
# Measuring total factor productivity
dat$tfp <- dat$qOut / dat$X # using Fisher index
dat$tfpP <- dat$qOut / dat$XP # using Paasche Index
dat$tfpL <- dat$qOut / dat$XL # using Laspeyres Index
```

```r
hist( dat$tfp )
```

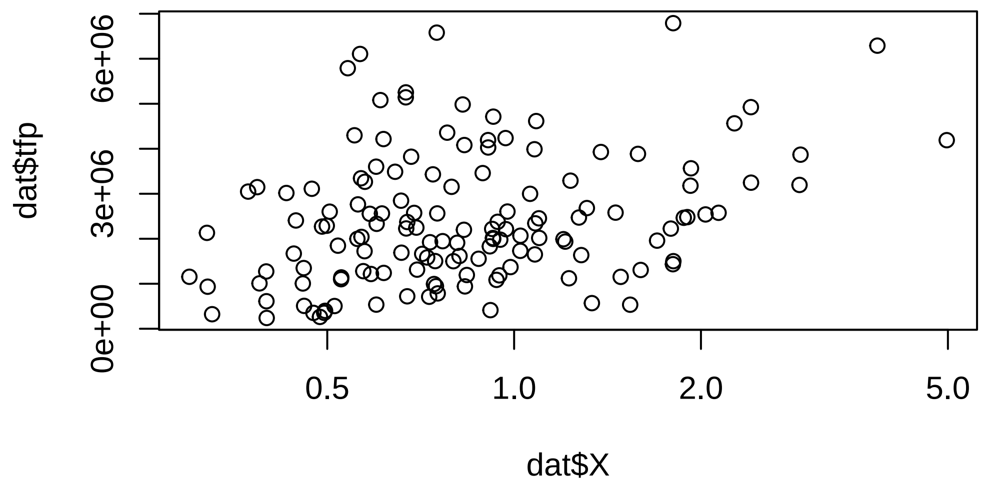


TFP varies considerably across firms, with the majority falling into the relatively low-TFP range.

```r
# Plotting tfp against output and input quantity index
plot( dat$qOut, dat$tfp, log = "x" )
```

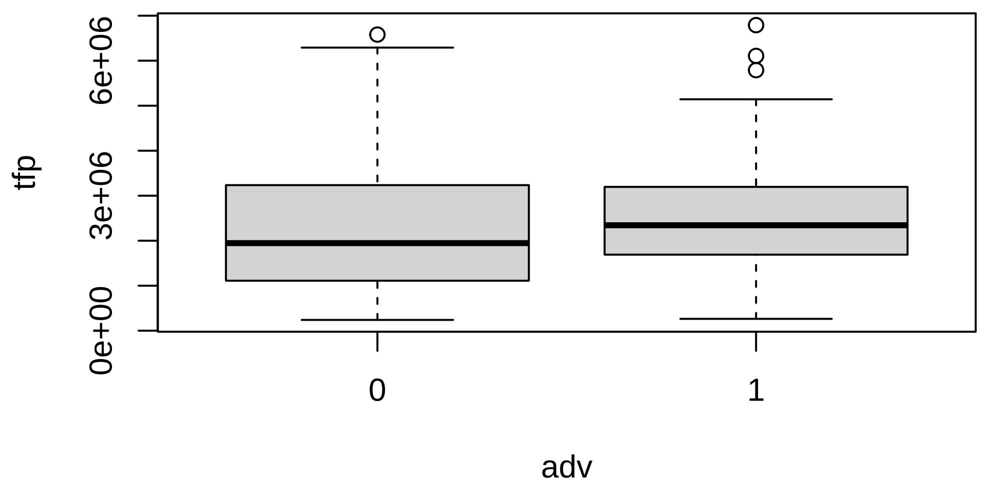


```r
plot( dat$X, dat$tfp, log = "x" )
```

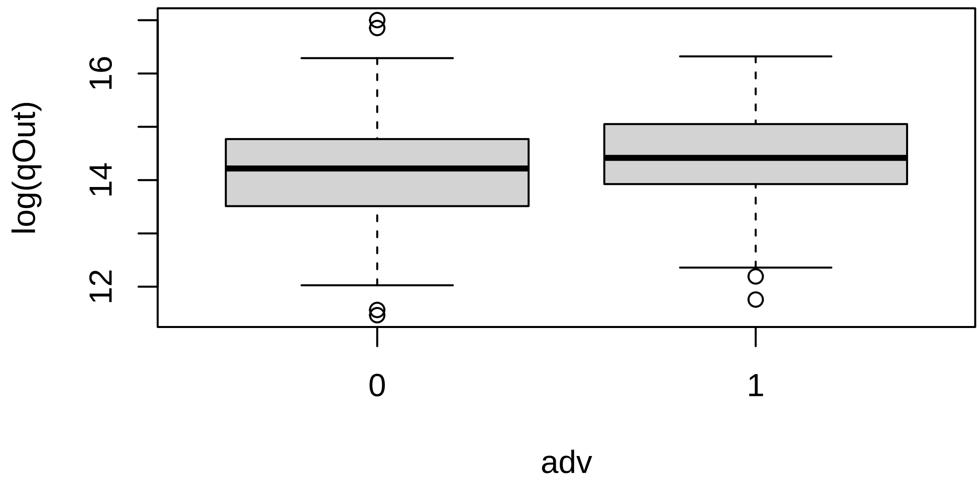


These plots indicate that larger firms, characterized by higher output volumes, are typically associated with greater TFP. However, the plot of TFP against the aggregate input index shows only a mild positive association between the two.

```r
# Does advisory service (a dummy) affects tfp?
boxplot( tfp ~ adv, data = dat )
```




```r
boxplot( log(qOut) ~ adv, data = dat )
```



Some firms used advisory services. It appears that firms with or without advisory services use similar input quantities; however, those with advisory services are associated with a slightly higher TFP (in terms of expected value).

Input substitution

What might cause variation in the input mix chosen by different firms? Are all firms operating with allocative efficiency?

Marginal rate of technical substitution

Suppose that we are operating at an input mix (x_1, x_2) and that we consider substituting a little bit of input 1 with input 2 to produce the same amount of output y. How much extra of input 2 do we need? Mathematically, this is measured by the slope of the isoquant; we refer to it as the ***Marginal Rate of Technical Substitution (MRTS)***.

Setting $dy = f_1 dx_1 + f_2 dx_2 = 0$, we define MRTS as

$$MRTS = \frac{dx_2}{dx_1} = -\frac{f_1}{f_2} = -\frac{MP_1}{MP_2}.$$

Note that in some books, it might be measured as dx_1/dx_2. However, what is more important is how we interpret the formula once it is defined. In the current definition, it is interpreted as the amount of x_2 needed to substitute for one unit of x_1, while keeping the output at a constant level.

Relative marginal rate of technical substitution

The ***relative marginal rate of technical substitution*** is defined as the ratio of MRTS and input ratio:

$$RMRTS = \frac{MRTS}{x_2/x_1} = -\frac{MP_1}{MP_2}\frac{x_1}{x_2}$$

It can be interpreted as relative percentage change in one input (say, capital) needed to compensate for a relative percentage change in another input (say, labour) while maintaining the same level of output. Divide both sides by y and using the definition of output elasticity, we can rewrite the above formula as

$$RMRTS = -\frac{MP_1}{y/x_1}\frac{y/x_2}{MP_2} = -\frac{MP_1}{AP_1}\frac{AP_2}{MP_2} = -\frac{\varepsilon_1}{\varepsilon_2}$$

Elasticity of substitution

The importance of input substitution led to various definition of *elasticities of substitutions*. The elasticity of substitution between two inputs measures how easily one input can be substituted for another in response to changes in their relative prices, holding output constant. It is a measure of the curvature of the isoquant. Hicks (Hicks 1963) offers the following definition of elasticity σ between inputs x_1 and x_2:

$$\sigma = \frac{d(x_2/x_1)}{d(f_1/f_2)} \frac{(f_1/f_2)}{(x_2/x_1)} = \frac{\% \text{ change in input ratio}}{\% \text{ change in MRTS}}.$$

To compute the elasticity, we typically express $MRTS$ in terms of the input ratio, or $\ln(MRTS)$ in terms of $\ln(x_j/x_i)$, to find the corresponding derivative.

$$\sigma = \frac{MRTS}{(x_2/x_1)}(1/\frac{dMRTS}{d(x_2/x_1)}) = 1/\frac{d\ \ln MRTS}{d\ \ln(x_2/x_1)}.$$

Test exercise:

Consider a regular Cobb-Douglas production and show that it has an elasticity of substitution of 1.

An equivalent representation of σ is

$$\sigma = \frac{-f_1 f_2(x_1 f_1 + x_2 f_2)}{x_1 x_2(f_{11} f_2^2 - 2f_{12} f_1 f_2 + f_{22} f_1^2)},$$

where f_i and f_{ii} are the first- and second-order partial derivatives, and f_{ij} is the second-order cross derivative.

Finally, another useful equivalent representation of the above formula, using the matrix notation, is given by

$$\sigma = \frac{x_1 f_1 + x_2 f_2}{x_1 x_2} \frac{F_{12}}{F},$$

where F is the determinant of the bordered Hessian of the production function:

$$F = \begin{vmatrix} 0 & f_1 & f_2 \\ f_1 & f_{11} & f_{12} \\ f_2 & f_{12} & f_{22} \end{vmatrix},$$

and F_{12} is the associated co-factor of f_{12}.

With multiple inputs, we can consider the same formula—replacing 1 by i and 2 by j—to compute the elasticity of substitution σ^D_{ij} for any pair of inputs x_i and x_j.

Two observations to note: (i) This measure then implicitly assumes that we are holding all other inputs constant; (ii) When the production is continuously differentiable, the cross-derivatives are symmetric, implying that $\sigma^D_{ij} = \sigma^D_{ji}$. As we hold other inputs constant, this measure is also referred to as ***short-run elasticity of substitution*** (because of first point) or ***Direct elasticity of substitution***.

A generalization of the above measure of elasticity of substitution is ***Allen partial elasticity of substitution***, which is defined as

$$\sigma_{ij} = \frac{\sum_i x_i f_i}{x_i x_j} \frac{F_{ij}}{F},$$

where F is the determinant of the bordered Hessian matrix:

$$F = \begin{vmatrix} 0 & f_1 & f_2 & \cdots & f_n \\ f_1 & f_{11} & f_{12} & \cdots & f_{1n} \\ f_2 & f_{12} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_n & f_{1n} & f_{2n} & \cdots & f_{nn} \end{vmatrix},$$

and F_{ij} is the co-factor of f_{ij}.

The final elasticity measure is the ***Morishima elasticity of substitution***, which is given by

$$\sigma^M_{ij} = \frac{f_i}{x_i} \frac{F_{ij}}{F} - \frac{f_j}{x_j} \frac{F_{jj}}{F} = \frac{x_j f_j}{\sum_i x_i f_i} (\sigma_{ij} - \sigma_{jj}),$$

where σ_{ij} (without the superscript) denote the Allen elasticity measure.

Observe that unlike Allen elasticity measure, Morishima measure is not symmetric. Further, a pair of goods can be complements in terms of Allen elasticity, whereas the corresponding Morishima measure could class them as substitutes.

We will consider these measures of elasticity for the specific functional forms of the production function estimated for our dataset.

Reading materials:

- Henningsen, chapter 2
- Varian, Chapter 1

Section 2

Supply behavior of a firm facing a competitive market

A fundamental assumption in most economic analyses of firm behavior is that a firm acts to maximize its profits. This means it chooses its productive activities—such as input choices, volume of production, and so on—in a way that maximizes revenue net of costs.

Note that this assumption can be easily relaxed without altering the basic principles of analysis. We can consider other objectives for the firm, such as maximizing consumer surplus, or expand the set of possible activities, such as choosing specific technologies that meet regulatory standards.

Consider a firm that takes prices as given in both its output and input markets. The firm produces a single output y using inputs $\mathbf{x} = (x_1, x_2, \dots, x_n)$, and the technology of the firm is represented by a production function $f(\mathbf{x})$.

- **Revenue**: The total revenue TR is given by: $TR = p \cdot y = p \cdot f(\mathbf{x})$ where p is the output price.
- **Cost**: The total cost TC is given by: $TC = C(\mathbf{x})$ where $C(\mathbf{x})$ represents the cost function, typically assumed to be $C(\mathbf{x}) = \sum_{i=1}^{n} w_i x_i$, with w_i being the price of input x_i .
- **Profit**: The profit π is: $\pi(\mathbf{x}) = p \cdot f.(\mathbf{x}) - C(\mathbf{x})$.

Profit maximization

The firm's problem is to choose $\mathbf{x}$ to maximize profit:

$$\max_{\mathbf{x}} \pi(\mathbf{x}) = p \cdot f(\mathbf{x}) - \sum_{i=1}^{n} w_i x_i$$

First-Order Conditions (FOC)

To find the optimal input levels $\mathbf{x}^*$, we take the partial derivative of the profit function with respect to each input x_i and set it equal to zero:

$\frac{\partial \pi}{\partial x_i} = p \cdot \frac{\partial f(\mathbf{x})}{\partial x_i} - w_i = 0,\ \forall i = 1, 2, \dots, n$

Rearranging the terms, we get the marginal condition:

$p \cdot \frac{\partial f(\mathbf{x}^*)}{\partial x_i} = w_i,\ \forall i = 1, 2, \dots, n$

The condition can be expressed as:

$$w_i = p \cdot MP_i = MVP_i,$$

where MVP_i is the marginal value product of input i.

Second-Order Conditions (SOC) for a Maximum

The second-order conditions ensure that the solution found using the first-order conditions is indeed a maximum (rather than a minimum or saddle point). The SOC requires that the Hessian matrix of second derivatives of the profit function with respect to the inputs is negative semi-definite.

For profit maximization, H must be negative semi-definite at $\mathbf{x}^*$.

Implications of FOC

A direct implication of profit maximizing behavior is that the ratio of prices equals the absolute value of marginal rate of technical substitution:

$$\frac{w_i}{w_j} = \frac{MP_i}{MP_j} = -MRTS_{ij}$$

In addition, denoting the cost share of input i by s_i where $s_i = w_i x_i / \sum_{k=1}^{n} w_k x_k$, we can express the ratio of cost share by a profit maximization firm as

$$\frac{s_i}{s_j} = \frac{w_i x_i}{w_j x_j} = -\frac{MRTS_{ij}}{x_j / x_i} = -RMRTS_{ij}$$

Recall the definition of $RMRTS = -\varepsilon_i / \varepsilon_j$, which gives us

$$\frac{s_i}{s_j} = \frac{\varepsilon_i}{\varepsilon_j}$$

The above relation along with the fact that $\sum_i s_i = 1$ implies that for a profit maximizing firm the cost share of each input must equal the ratio of output elasticity over the elasticity of scale:

$$s_i = \frac{\varepsilon_i}{\varepsilon}$$

Input demand function

If we solve the simultaneous equations represented by the FOCs, we get the ***input demand function*** as a function of the output price and the input/factor prices:

$$x_i = x_i(p, \mathbf{w})$$

Supply function

Replacing the inputs by the input demand functions in the production, we can derive the output ***supply function*** as a function of the output price and the input/factor prices:

$$y = f(x_1(p, \mathbf{w}), \dots, x_n(p, \mathbf{w})) = y(p, \mathbf{w})$$

As the derived input demand and output functions are expressions of prices (output and input prices), we can derive the price elasticities of demand and supply:

$$\varepsilon_{ij}(p, \mathbf{w}) = \frac{\partial x_i(p, \mathbf{w})}{\partial w_j} \frac{w_j}{x_i(p, \mathbf{w})} : \text{elasticity of input } i \text{ w.r.t the price of input } j$$

$$\varepsilon_{yi}(p, \mathbf{w}) = \frac{\partial y(p, \mathbf{w})}{\partial w_i} \frac{w_i}{y(p, \mathbf{w})} : \text{elasticity of output } y \text{ w.r.t the price of input } i$$

$$\varepsilon_{yp}(p, \mathbf{w}) = \frac{\partial y(p, \mathbf{w})}{\partial p} \frac{p}{y(p, \mathbf{w})} : \text{elasticity of output } y \text{ w.r.t the output price } p$$

$$\varepsilon_{ip}(p, \mathbf{w}) = \frac{\partial x_i(p, \mathbf{w})}{\partial p} \frac{p}{x_i(p, \mathbf{w})} : \text{elasticity of input } i \text{ w.r.t the output price } p$$

Profit function

Replacing the inputs by the input demand functions and output by the supply function, we can determine the ***profit function***, which characterizes the maximum profit as a function of the output price and the input/factor prices:

$$\pi(p, \mathbf{w}) = p \cdot y(p, \mathbf{w}) - \sum_{i=1}^{n} w_i x_i(p, \mathbf{w})$$

We will later explore the properties of the profit function, which are useful for examining the dual approach.

Supply behavior with output constraint

While maximizing profit, a firm can freely choose all inputs, which in turn determines the optimal volume of production. Consider a situation where the firm is required to produce a specific volume of output, either due to a contractual obligation or because altering the output level is not feasible in the short run. In this case, the firm must make optimal input choices to maximize profit under these constraints.

Given that the revenue components are fixed, the firm's optimization problem can be reformulated as a cost minimization problem with the constraint $y = f(\mathbf{x})$.

Cost minimization

The firm's problem is to choose $\mathbf{x}$ to minimize costs

$$\min_{\mathbf{x}} \sum_{i=1}^{n} w_i x_i \text{ such that } y = f(\mathbf{x})$$

The constrained optimization problem can be solved by Lagrangian approach:

$$\min_{\mathbf{x},\lambda} \mathcal{L} = \sum_{i=1}^{n} w_i x_i + \lambda(y - f(\mathbf{x}))$$

First-Order Conditions (FOC)

To find the optimal input levels $\mathbf{x}^*$, we take the partial derivative of the profit function with respect to each input x_i and λ and set them equal to zero:

$\frac{\partial \mathcal{L}}{\partial x_i} = w_i - \lambda \cdot \frac{\partial f(\mathbf{x})}{\partial x_i} = 0, \ \forall i = 1, 2, \dots, n$

and $\frac{\partial \mathcal{L}}{\partial \lambda} = y - f(\mathbf{x}) = 0$

Rearranging the terms, we get the marginal condition:

$$w_i = \lambda \cdot MP_i$$

and

$$\frac{w_i}{w_j} = \frac{MP_i}{MP_j} = -MRTS_{ij}$$

As we observed in the profit-maximization analysis, it can also be shown that the ratio of cost shares equals the absolute value of the RMRTS (Rate of Marginal Rate of Technical Substitution). This consistency between the findings from the analyses of profit-maximizing and cost-minimizing behavior is not coincidental; it arises because profit maximization leads to the optimal volume of production at minimum cost.

Conditional input demand function

The solutions to the cost minimization problem are called the ***conditional input demand function***, expressed as a function of the output volume y and the input/factor prices:

$$x_i = x_i(y, \mathbf{w})$$

From these derived demand, we can determine the elasticities with respect to input prices and output volume:

$$\varepsilon_{ij}(y, \mathbf{w}) = \frac{\partial x_i(y, \mathbf{w})}{\partial w_j} \frac{w_j}{x_i(y, \mathbf{w})} : \text{elasticity of conditional input demand } i \text{ w.r.t the price of input } j$$

$$\varepsilon_{iy}(p, \mathbf{w}) = \frac{\partial x_i(y, \mathbf{w})}{\partial y} \frac{y}{x_i(y, \mathbf{w})} : \text{elasticity of conditional input demand } i \text{ w.r.t output } y$$

Cost function

Replacing the inputs by the conditional input demand functions in the firm's cost expression, we can determine the ***cost function***, which characterizes the minimum costs as a function of the output volume and the input/factor prices:

$$c(y, \mathbf{w}) = \sum_{i=1}^{n} w_i x_i(y, \mathbf{w})$$

Later, while examining the dual approach, we will further explore the properties of the cost function.

Emprical estimation of production function

In this course, we will use linear model estimation techniques. Given a set of firm-level observations on production and costs, when can we appropriately estimate a production function using econometric methods such as OLS?

Several conditions must be met:

1. **Firms must operate under similar conditions**. If this is not the case, the differences should be incorporated into the econometric model.

2. **The sample must be free of selection bias**. When studying a specific industry based on a representative sample of firms, it is crucial that the sample is representative and unbiased.

3. **All firms in the data set must produce at the maximum output level given the inputs**. We implicitly assume that any deviation from the production frontier reflects random noise (firm-specific shocks).

4. **No perfect multicollinearity**. There should be no perfect linear relationship between the independent variables.

5. **All input quantities must be uncorrelated with the error terms**. Input quantities should not be correlated with the residuals of the model.

Linear technology

We can estimate the following linear production function with N inputs based on our data:

$$y = \beta_0 + \sum_{i=1}^{N} \beta_i x_i + \epsilon$$

Estimation

```r
# Fitting a linear model
prodlinear <- lm( qOut ~ qCap + qLab + qMat, data = dat )
summary( prodlinear )
```

```
Call:
lm(formula = qOut ~ qCap + qLab + qMat, data = dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-3888955  -773002    86119   769073  7091521 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.616e+06  2.318e+05  -6.972 1.23e-10 ***
qCap         1.788e+00  1.995e+00   0.896    0.372    
qLab         1.183e+01  1.272e+00   9.300 3.15e-16 ***
qMat         4.667e+01  1.123e+01   4.154 5.74e-05 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1541000 on 136 degrees of freedom
Multiple R-squared:  0.7868,    Adjusted R-squared:  0.7821
F-statistic: 167.3 on 3 and 136 DF,  p-value: < 2.2e-16
```

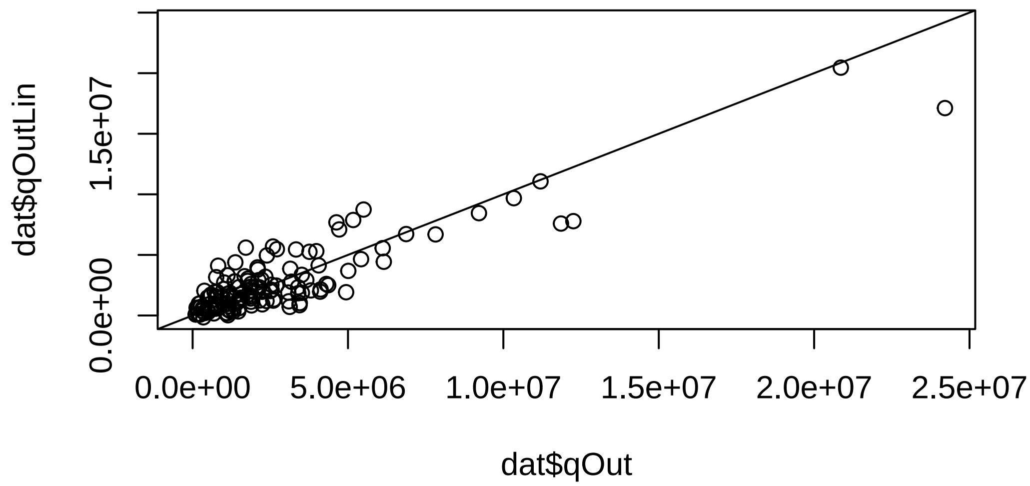
The regression results show that the coefficients for labor and materials are positive and significant, but the coefficient for capital is not. Whether we should drop capital from our estimated model is a complex issue. If capital is an essential input in the true underlying relationship, then dropping it from the model would lead to biased and inconsistent estimates of other input coefficients. On the other hand, including an insignificant variable like capital would result in less efficient estimates (making other estimates less precise), but they would still remain unbiased and consistent (converges to the true coefficient if sample size increases).

Is it a good fit?

In the absence of comparable models, it remains unclear whether a linear production function is a goof fit for our data. However, we can still infer the strength of our model based on some simple observation. For example, the residual/error term, captures the difference between observed and predicted values. This error term arises due to measurement error, omitted variable bias, and other random shocks. A high R^2 value provides some evidence of a better fit.

A plot of predicted values against the observed ones helps visualize how well the predicted values align with the observed values. If the regression model is a good fit, the points should cluster around the 45-degree line (where predicted equals observed). Deviations from the 45-degree line can indicate systematic errors in the model. If the points form a curved pattern instead of clustering around the line, this could indicate that a linear model may not be appropriate. The plot can also help identify outliers—points that are far from the line—which may disproportionately affect the regression results. However, it should be used in conjunction with other tools, such as residual plots, to fully assess the model's performance. Relying solely on this plot may not reveal all potential issues, such as heteroscedasticity or multicollinearity.

```
# Predicted vs. observed plot
library(miscTools)
dat$qOutLin <- fitted( prodlinear )
compPlot( dat$qOut, dat$qOutLin )
```

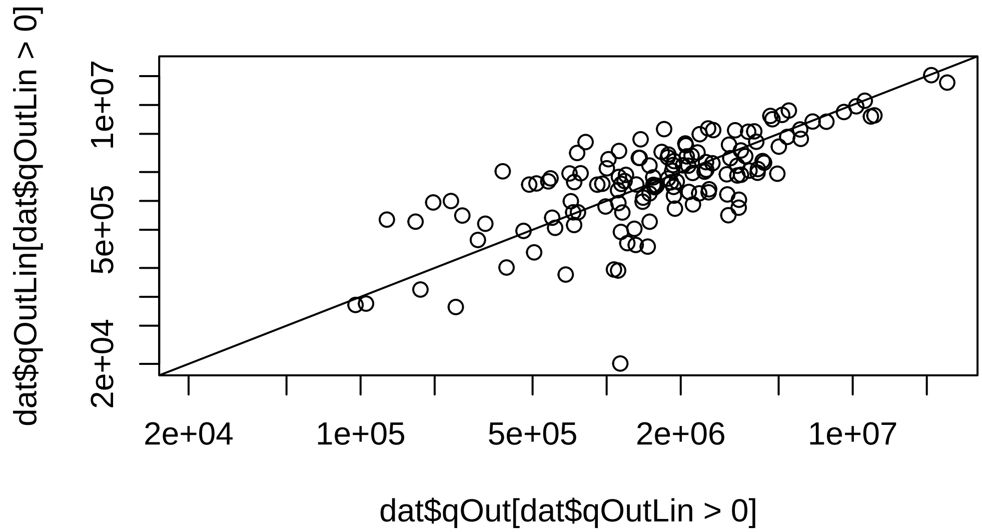


Due to some extreme values, most of the data points appear clustered toward the origin. One way to address this issue is by scaling the axes, for example, using a logarithmic scale. However, since the logarithm of negative values is undefined, we would need to exclude those negative values from the plot.

```
# Predicted vs. observed (logarithmic scaling of axes)
table( dat$qOutLin >= 0 )
```

```
FALSE  TRUE
    1   139
```

```
compPlot( dat$qOut[ dat$qOutLin > 0 ], dat$qOutLin[ dat$qOutLin > 0 ], log = "xy" )
```



The deviations from observed values looks random (okay) in both scatter plots.

Next, we perform consistency checks on our estimated production technology to ensure alignment with theoretical predictions derived from microeconomic principles.

Theoretical consistency

Essentiality: Weak essentiality means that each input in the production function has some non-zero contribution to the output when other inputs are held constant. A *strict essentiality*, on the other hand, is a stronger condition, which requires that the production function collapses to zero output if any one of the inputs is not present. In other words, every input is strictly necessary for production to take place.

Positive input coefficients for labour and materials are evidence of essentiality of these inputs. The negative and significant intercept term violates weak essentiality condition.

Monotonicity: The monotonicity condition refers to the requirement that the output should not decrease as the quantity of any input increases, holding all other inputs constant. The positive and significant coefficients of labour and materials satisfy the monotonicity condition.

Quasi-concavity: A production function is quasi-concave if its isoquants are convex, or equivalently, if the input requirement set is convex. In the case of a linear production function, the isoquants are straight lines, which means they are both concave and convex, and thus inherently convex. Consequently, the input requirement set is also convex. Therefore, quasi-concavity is trivially satisfied due to the linear nature of the production function.

Non-negativity: The production function should yield non-negative output values for any non-negative input choices. However, the negative intercept in our estimated production function violates this non-negativity assumption. Also, we are going to see below that there is one negative predicted output value for one firm, violating non-negativity condition.

Studying properties of the production function

Output elasticity

Note that the estimated linear production function implicitly assumes the marginal productivity of an input is the same across firms, and it is measured by the estimated coefficient. Any variation in marginal productivity across firms, even if it exists, cannot be addressed by these coefficients. To account for this variation, we can calculate the output elasticity by dividing each input coefficient by the corresponding average product and then examine how this elasticity measures may vary across firms.

In contrast, when estimating a Cobb-Douglas production function, the elasticity of output with respect to each input is directly represented by the regression coefficients. Then, plots of marginal products derived from a Cobb-Douglas function will typically show variation across firms.

```
# Caclulate output elasticity, compute mean and meadian, plot histograms
dat$eCap <- coef(prodlinear)["qCap"] / dat$apCap
dat$eLab <- coef(prodlinear)["qLab"] / dat$apLab
dat$eMat <- coef(prodlinear)["qMat"] / dat$apMat
```

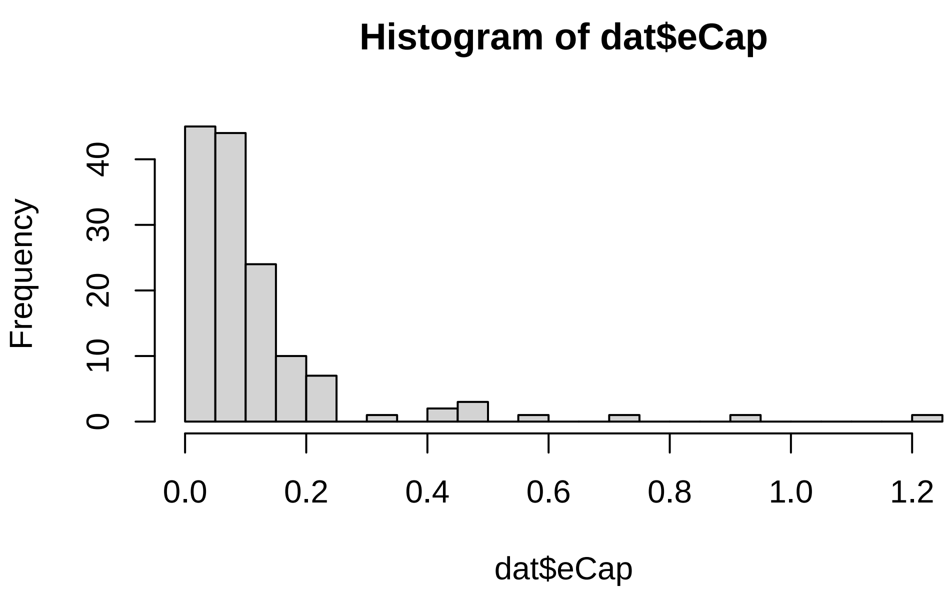
```r
colMeans( subset( dat, , c( "eCap", "eLab", "eMat" ) ) )
```

```
     eCap      eLab      eMat
0.1202721 2.0734793 0.8631936
```

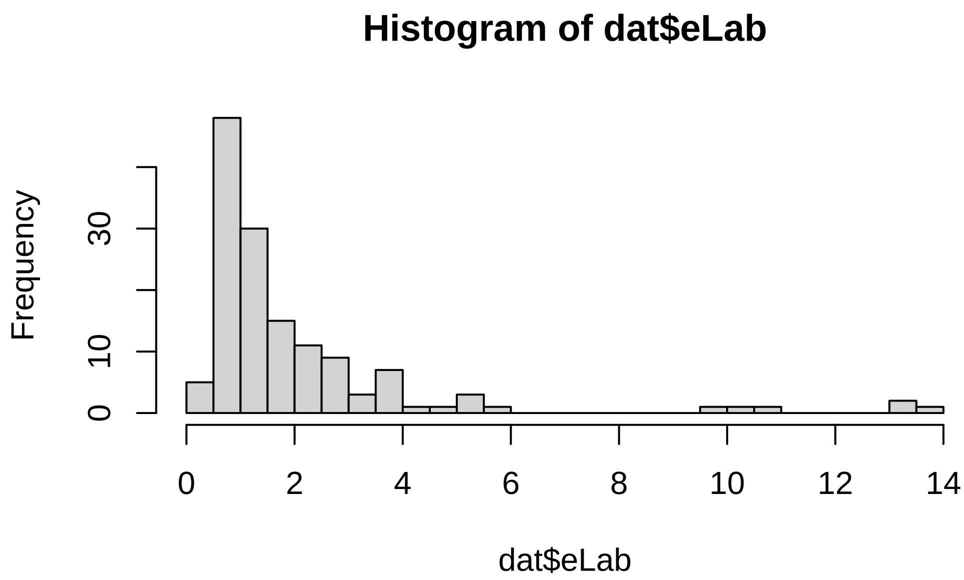
```r
colMedians( subset( dat, , c( "eCap", "eLab", "eMat" ) ) )
```

```
      eCap       eLab       eMat
0.08063406 1.28627208 0.58741460
```

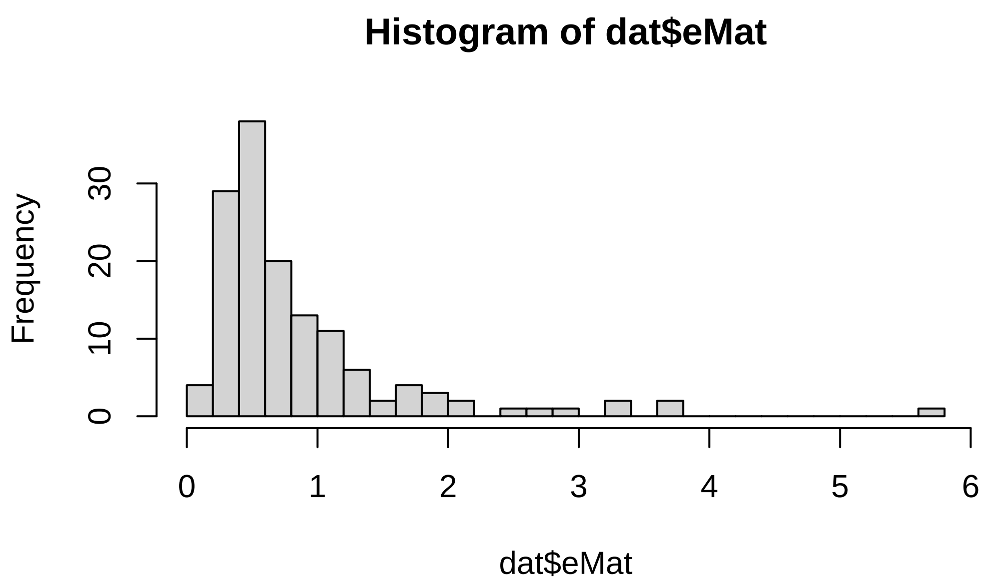
```r
hist( dat$eCap , 20)
```




```r
hist( dat$eLab , 20)
```




```r
hist( dat$eMat , 20)
```



Recall that the elasticity measures percentage change in output due to a percentage change in input. The marginal effect of capital on the output is rather small for most firms (mostly between 0 to 0,2 percent), there are many firms with implausibly high output elasticities of labor (for most, it is between 0,5 to 3) and materials (for most, it is between 0,2 to 1,2).

Returns to scale

Recall that the elasticity of scale reflects the returns to scale of a production technology: $\varepsilon < 1$ implies DRS, $\varepsilon = 1$ implies CRS, and $\varepsilon > 1$ implies IRS. Since firms will have different elasticities, we will see variation of ε values across firms.

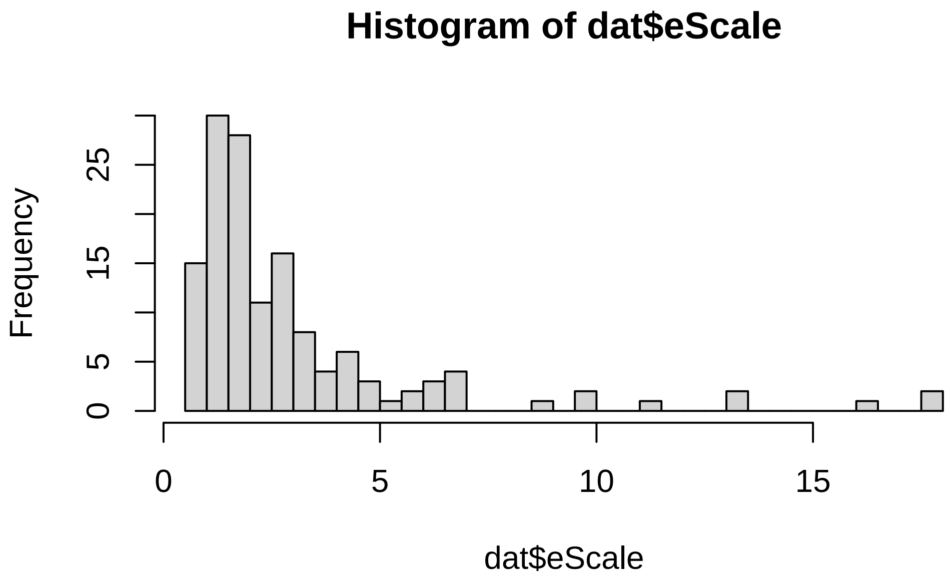
```
# Calculate elasticity of scale, compute mean and median, plot histograms, both base
dat$eScale <- with( dat, eCap + eLab + eMat )
colMeans( subset( dat, , c( "eScale" ) ) )
```

```
  eScale
3.056945
```

```
colMedians( subset( dat, , c( "eScale" ) ) )
```

```
  eScale
1.941536
```

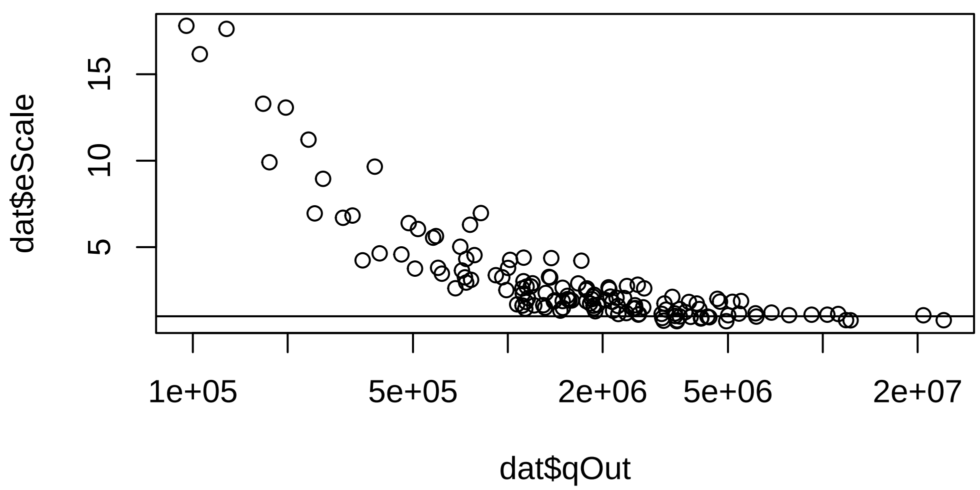
```
hist( dat$eScale, 30)
```



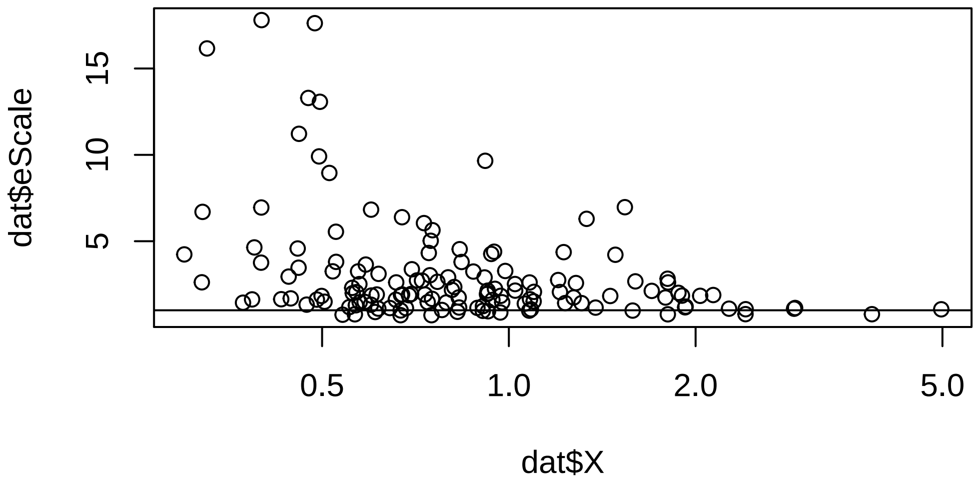
Based on the median values, we can conclude that most firms exhibit increasing returns to scale. The histogram shows that the majority have an elasticity of scale between 1 and 2, with a few showing extremely high returns to scale.

Which firms exhibit high returns to scale? We plot elasticity of scale against size proxies, such as output and an input index.

```
# Plot elasticity of scale against input index and output index
plot( dat$qOut, dat$eScale, log = "x" )
abline( 1, 0 )
```




```
plot( dat$X, dat$eScale, log = "x" )
abline( 1, 0 )
```



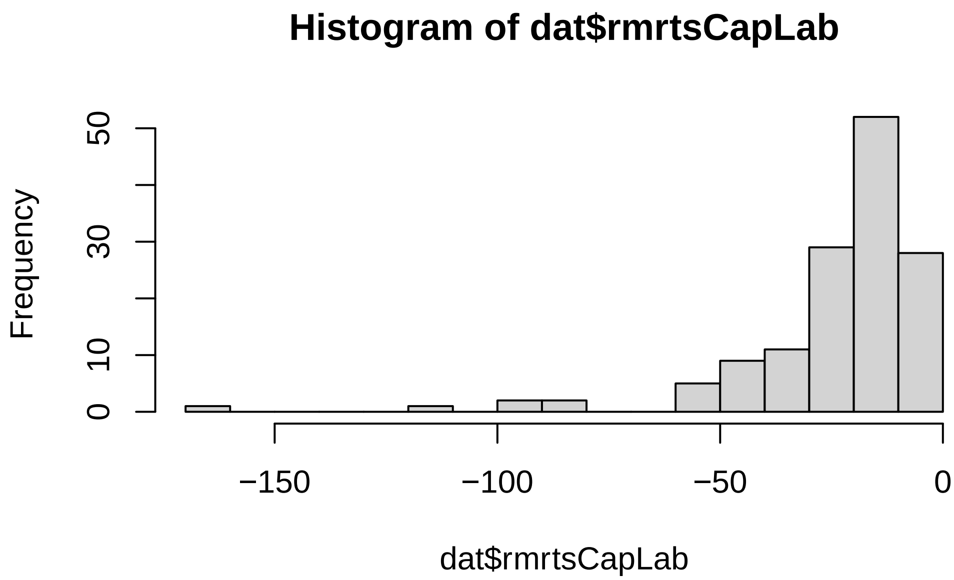
The scatter plots typically show that firms employing smaller input levels exhibit increasing returns to scale, which is expected. A few firms display decreasing returns to scale, and this is observed among firms using both high and low input volumes.

Input substitution–MRTS and RMRTS

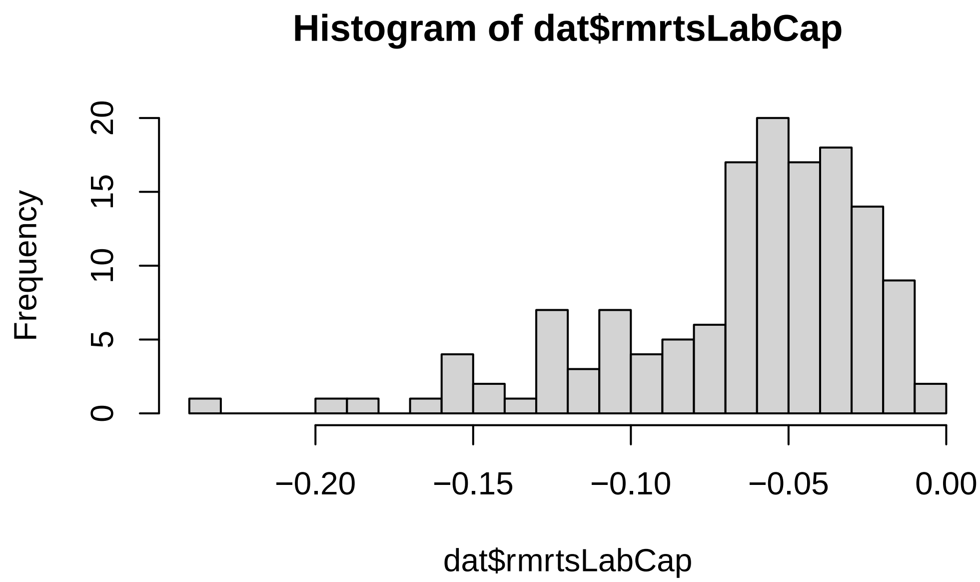
In a linear production function, the Marginal Rate of Technical Substitution (MRTS) is constant and can be measured as the ratio of the coefficients of the inputs. The relative MRTS will however vary across firms, due to the variation in output elasticities.

```
# Calculate MRTS
mrtsCapLab <- - coef(prodlinear)["qLab"] / coef(prodlinear)["qCap"]
mrtsLabCap <- - coef(prodlinear)["qCap"] / coef(prodlinear)["qLab"]
mrtsCapMat <- - coef(prodlinear)["qMat"] / coef(prodlinear)["qCap"]
mrtsMatCap <- - coef(prodlinear)["qCap"] / coef(prodlinear)["qMat"]
mrtsLabMat <- - coef(prodlinear)["qMat"] / coef(prodlinear)["qLab"]
mrtsMatLab <- - coef(prodlinear)["qLab"] / coef(prodlinear)["qMat"]
# Calculate RMRTS
dat$rmrtsCapLab <- - dat$eLab / dat$eCap
dat$rmrtsLabCap <- - dat$eCap / dat$eLab
dat$rmrtsCapMat <- - dat$eMat / dat$eCap
```

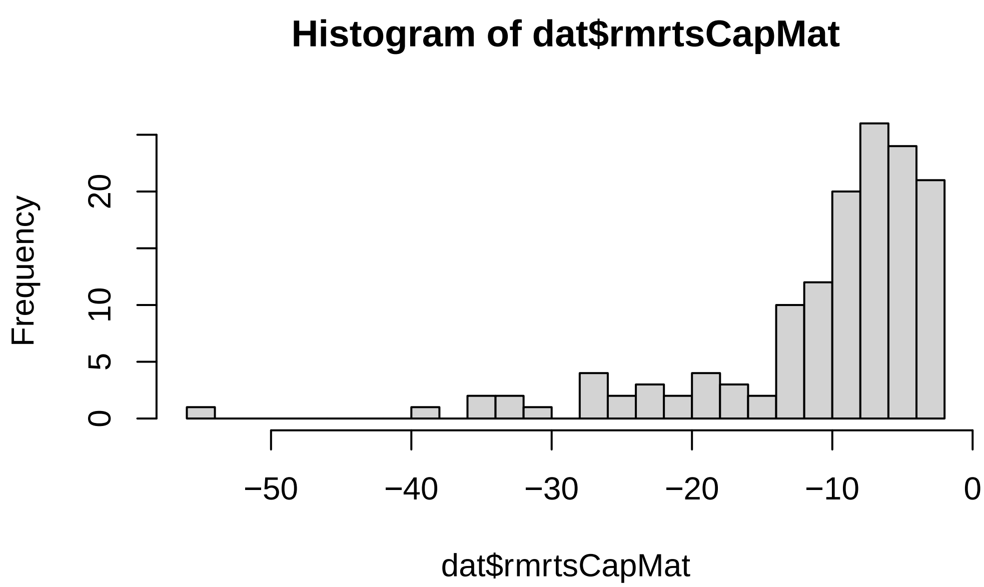
```
dat$rmrtsMatCap <- - dat$eCap / dat$eMat
dat$rmrtsLabMat <- - dat$eMat / dat$eLab
dat$rmrtsMatLab <- - dat$eLab / dat$eMat
# Draw histogram of RMRTS
hist( dat$rmrtsCapLab, 20 )
```



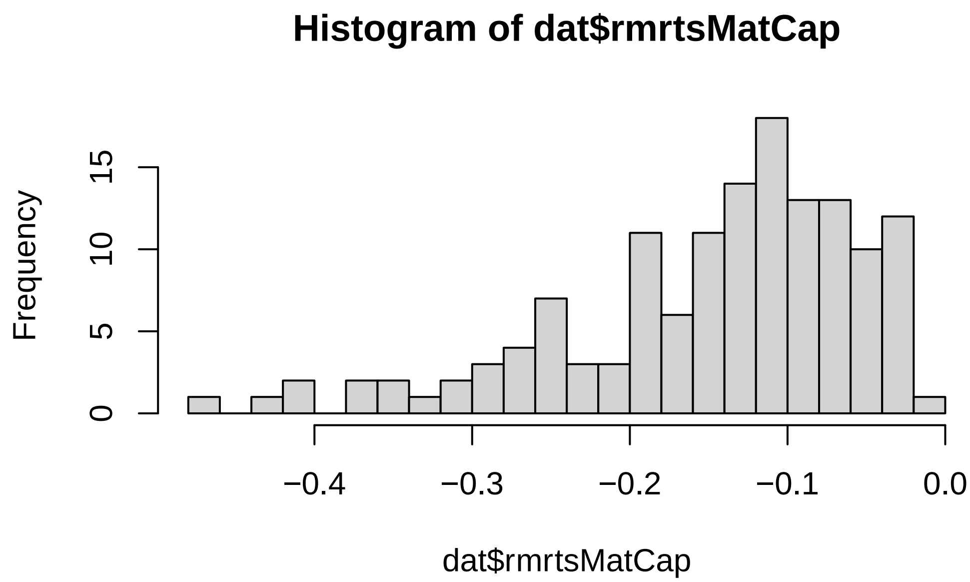

```
hist( dat$rmrtsLabCap, 20 )
```



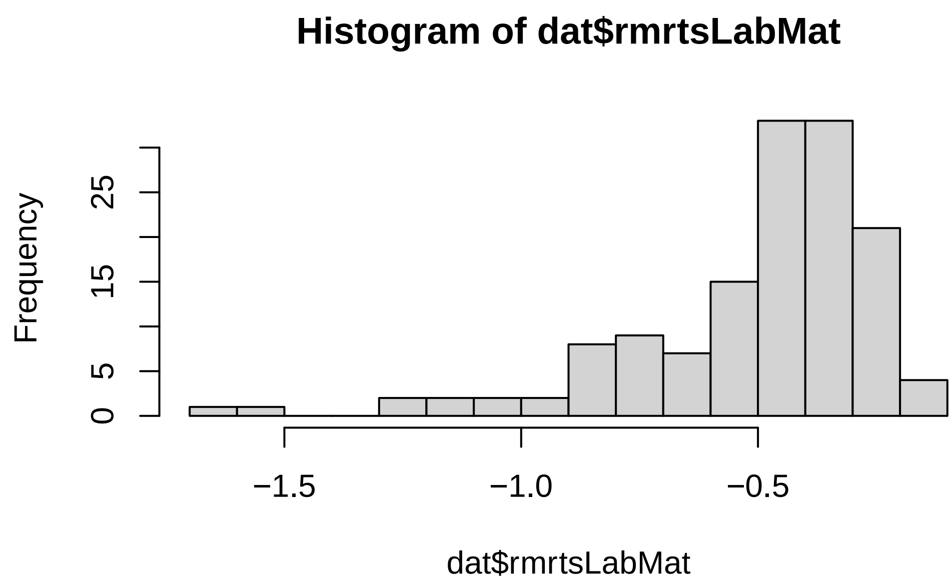

```
hist( dat$rmrtsCapMat, 20 )
```



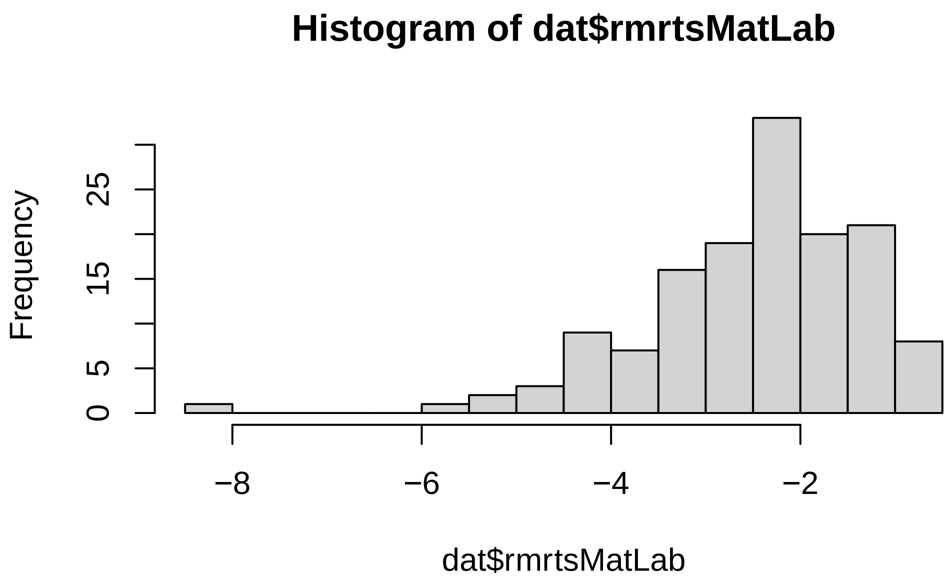

```
hist( dat$rmrtsMatCap, 20 )
```




```
hist( dat$rmrtsLabMat, 20 )
```




```
hist( dat$rmrtsMatLab, 20 )
```

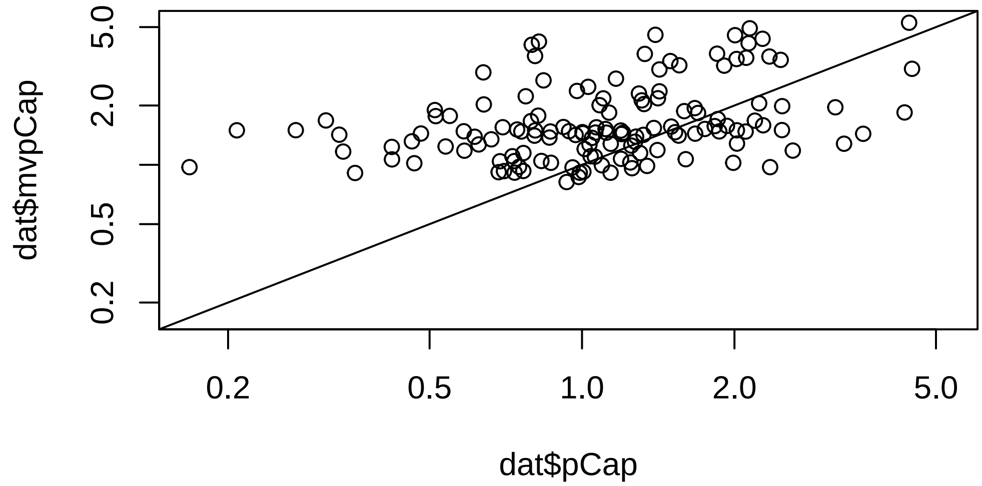


In our data set, most firms require approximately twenty percentage more capital or around two percentage more materials to make up for a one percentage reduction in labor.

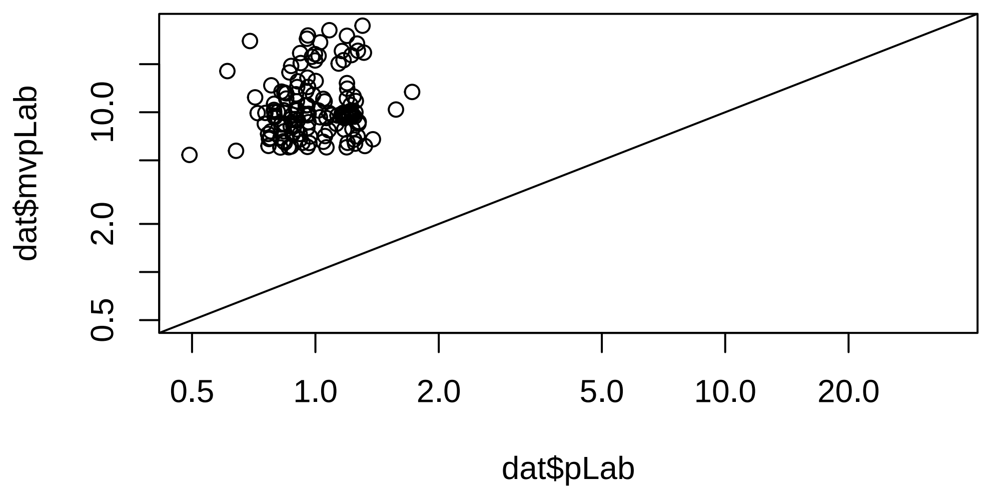
Profit-maximizing behavior

According to the profit-maximizing principle, the marginal value products—calculated as the output price multiplied by the marginal products—must equal the input prices at the optimal input choices. We plot marginal value products against the input prices across firms (we scale the axes for better viewing).

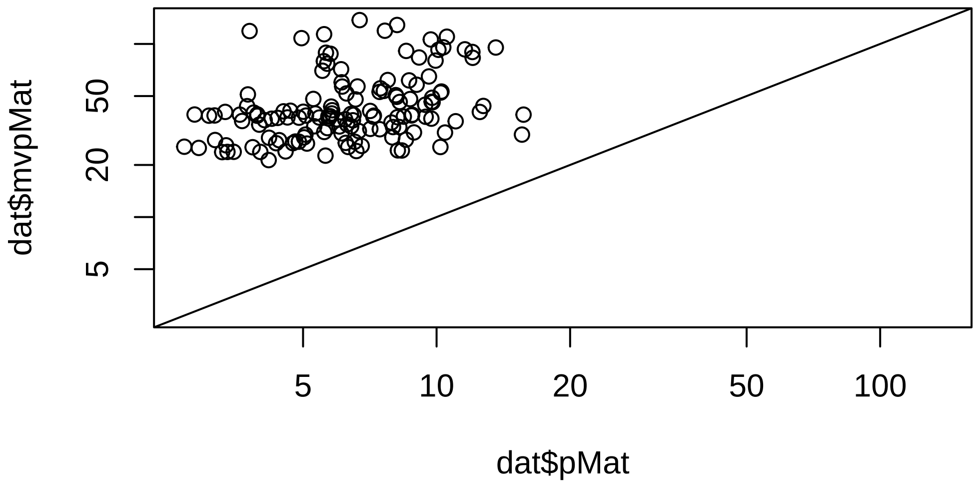
```
# Calculate MVP, plot MVP against input prices
dat$mvpCap <- dat$pOut * coef(prodlinear)["qCap"]
dat$mvpLab <- dat$pOut * coef(prodlinear)["qLab"]
dat$mvpMat <- dat$pOut * coef(prodlinear)["qMat"]
compPlot( dat$pCap, dat$mvpCap, log = "xy" )
```




```
compPlot( dat$pLab, dat$mvpLab, log = "xy" )
```




```
compPlot( dat$pMat, dat$mvpMat, log = "xy" )
```

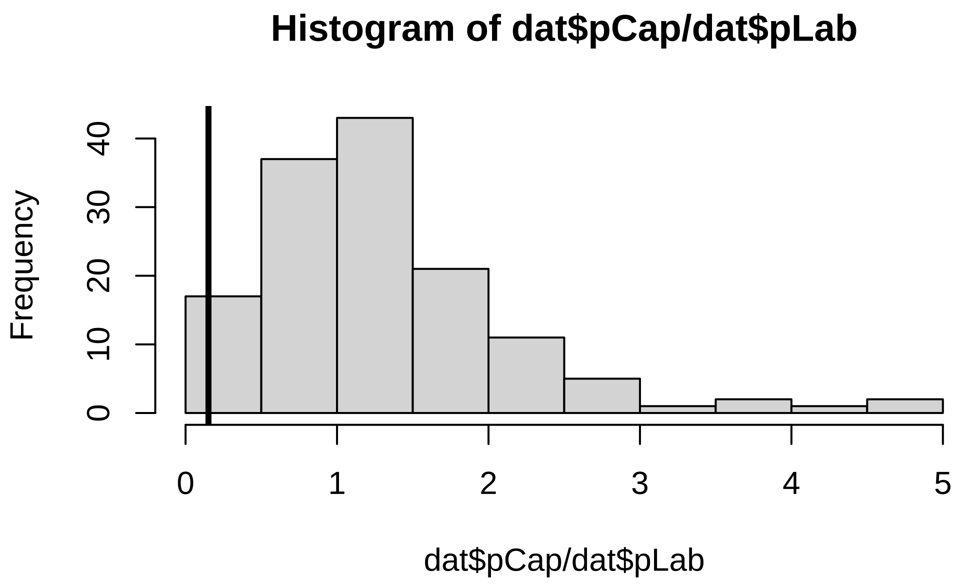


The scatter plot All firms could increase their profits by using more labor and materials, and some could also benefit from using more capital. Since most firms operate under increasing returns to scale, it is not surprising that many would gain from increasing most—or even all—input quantities. Why do not they do so? It might be possible there is imperfections in the input markets so that input prices do not perfectly reflect the marginal contributions.

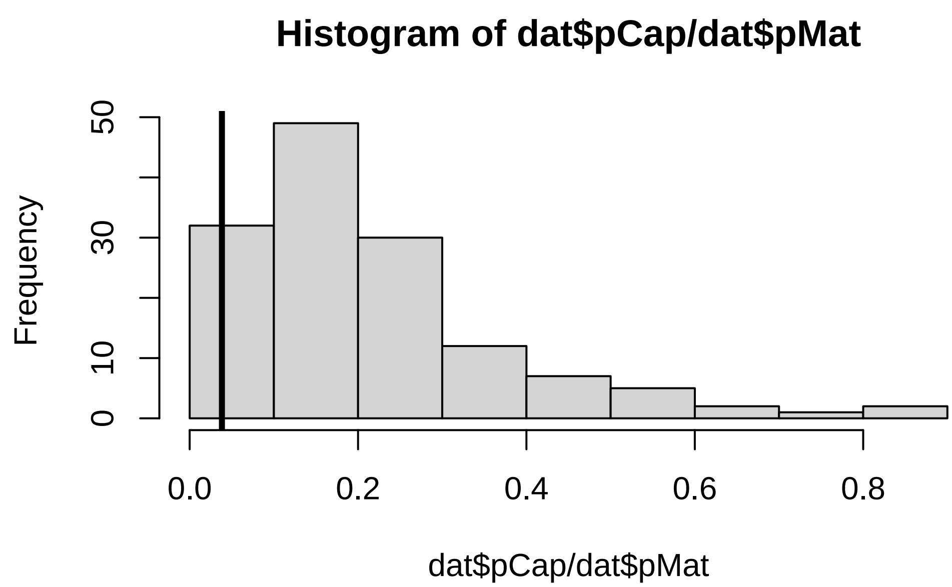
Cost-minimizing behavior

According to the cost-minimizing principle, the ratio of input prices must equal the absolute value of the MRTS between two inputs, which is constant in the case of a linear production function (the ratio of the coefficients). We create a histogram of input price ratios and compare it to the MRTS, represented by a solid vertical line.

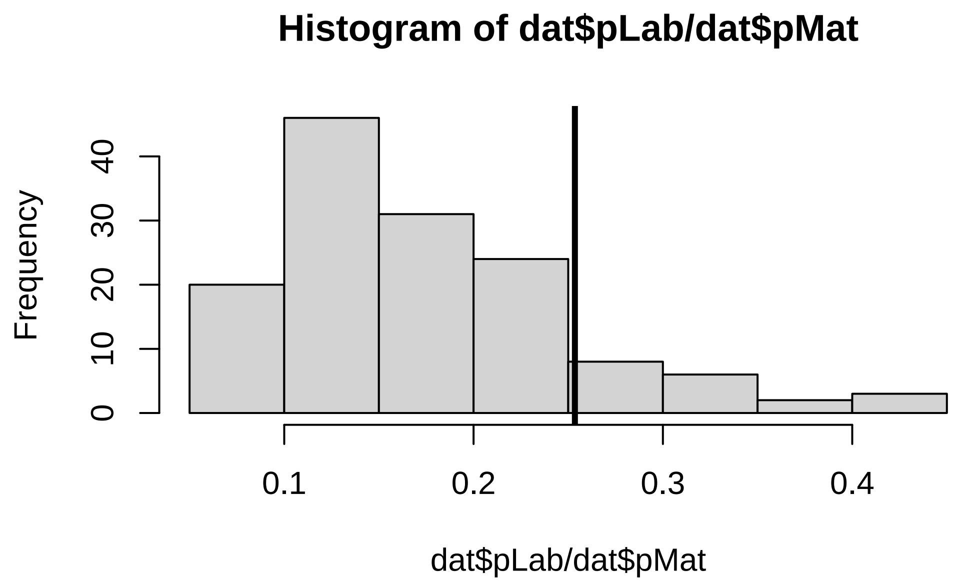
```
# Draw histogram of input price ratios
hist( dat$pCap / dat$pLab )
abline( v = - mrtsLabCap, lwd = 3  )
```



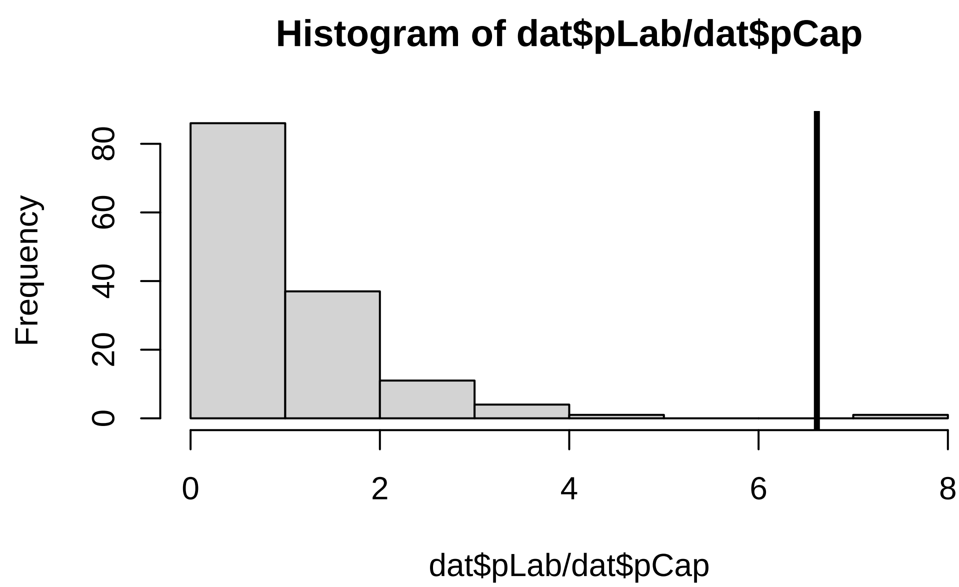

```
hist( dat$pCap / dat$pMat )
abline( v = - mrtsMatCap, lwd = 3  )
```



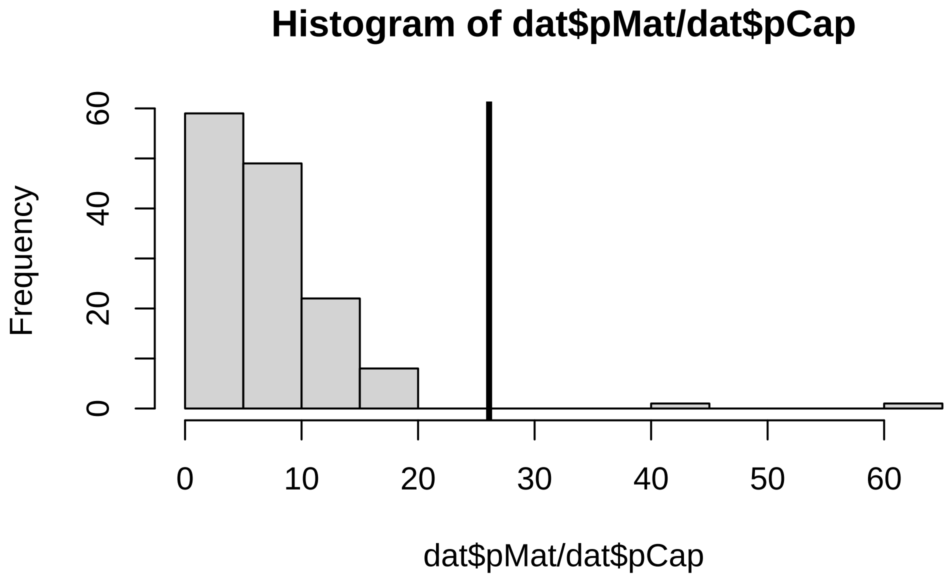

```
hist( dat$pLab / dat$pMat )
abline( v = - mrtsMatLab, lwd = 3  )
```



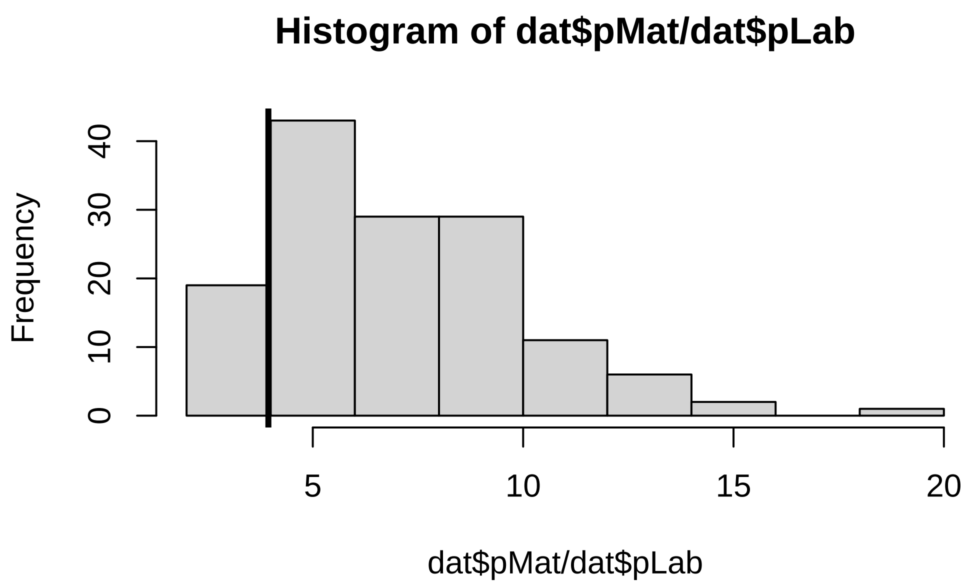

```r
hist( dat$pLab / dat$pCap )
abline( v = - mrtsCapLab, lwd = 3  )
```




```r
hist( dat$pMat / dat$pCap )
abline( v = - mrtsCapMat, lwd = 3  )
```




```
hist( dat$pMat / dat$pLab )
abline( v = - mrtsLabMat, lwd = 3  )
```



All plots provide evidence that most firms could benefit from substituting capital with labor, and capital with materials. When comparing labor and materials, most firms would benefit from substituting materials with labor. These observations align with our findings based on the profit-maximizing principle.

Other production function

Next week, we will discuss estimation of two other production functions based on our data.

We will also compare various fitted models.

Reading materials:

- Henningsen, chapter 2
- Varian, Chapter 2, 4

References

Henningsen, Arne. 2005. "micEcon: Microeconomic Analysis and Modelling." The R Foundation. https://doi.org/10.32614/cran.package.micecon.

———. 2024. "Introduction to Econometric Production Analysis with r." *Department of Food and Resource Economics, University of Copenhagen* Sixth draft version.

Hicks, John. 1963. *The Theory of Wages*. Springer.

Ivaldi, Marc, Norbert Ladoux, Hervé Ossard, and Michel Simioni. 1996. "Comparing Fourier and Translog Specifications of Multiproduct Technology: Evidence from an Incomplete Panel of French Farmers." *Journal of Applied Econometrics* 11 (6): 649–67. https://doi.org/10.1002/(sici)1099-1255(199611)11:6%3C649::aid-jae416%3E3.0.co;2-4.

Varian, Hal R. 1992. "Microeconomic Analysis." WW Norton & Company.