Linear Model in Economics

Part 1: Applied Prodcution Analysis

Tapas Kundu

2024-09-22

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.
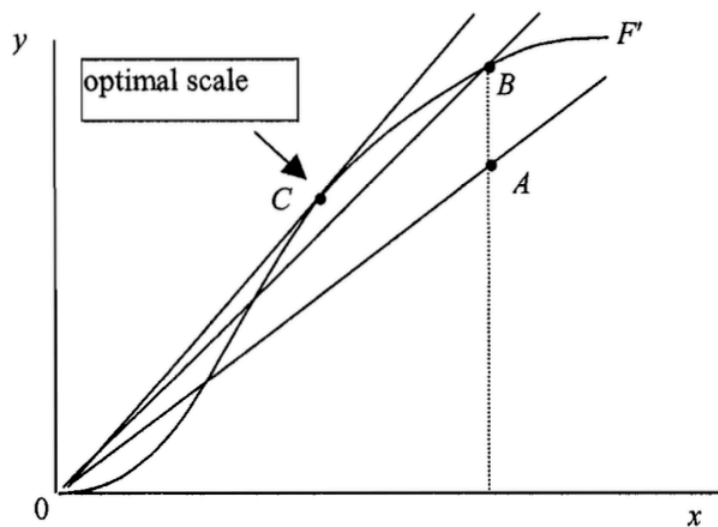

Figure 1: Productivity and technical efficiency

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

2

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); library(lmtest); library(car); library(miscTools)
```

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

Loading required package: zoo

Attaching package: 'zoo'

The following objects are masked from 'package:base':

    as.Date, as.Date.numeric

```
Loading required package: carData


Attaching package: 'car'

The following object is masked from 'package:psych':

    logit
```

```r
options(scipen = 999)
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.
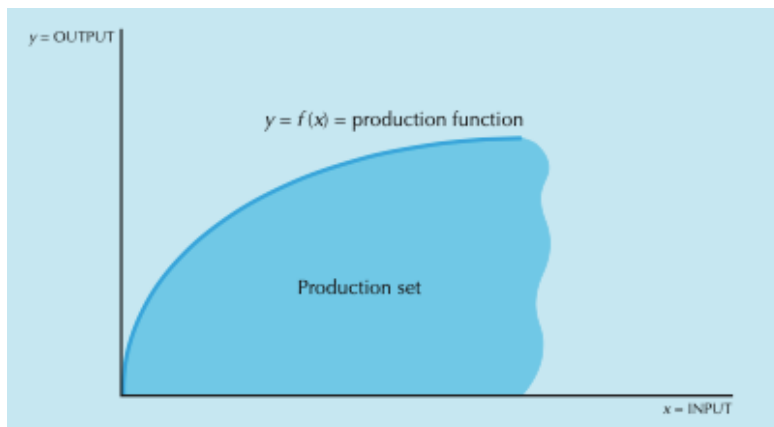
Figure 2: Production set and production function

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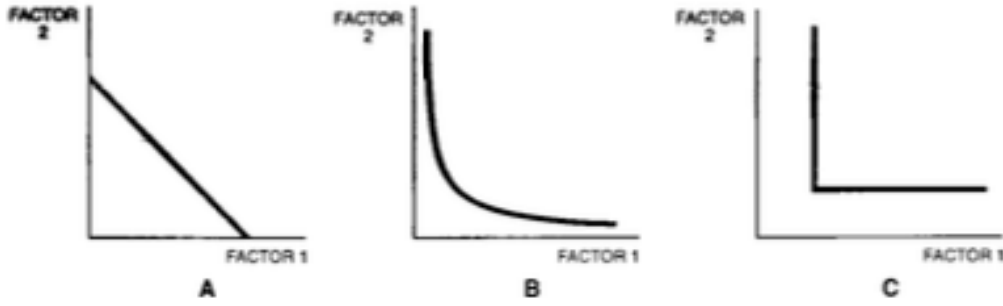
Figure 3: Isoquants—Linear, Cobb-Douglas, and Leontief production technology

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** $T : R^{n+m} \rightarrow R$ 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 ith 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.

$$AP_i = \frac{y}{x_i} = \frac{f(\mathbf{x})}{x_i}$$

$$MP_i = \frac{\partial y}{\partial x_i} = \frac{\partial f(\mathbf{x})}{\partial x_i} = f_i$$

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 )
```

Histogram of dat$apCap
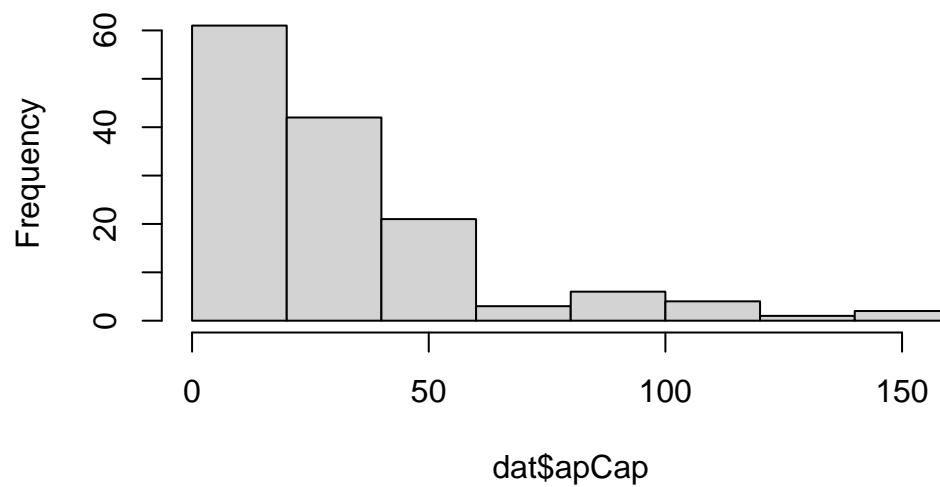

```
hist( dat$apLab )
```

Histogram of dat$apLab
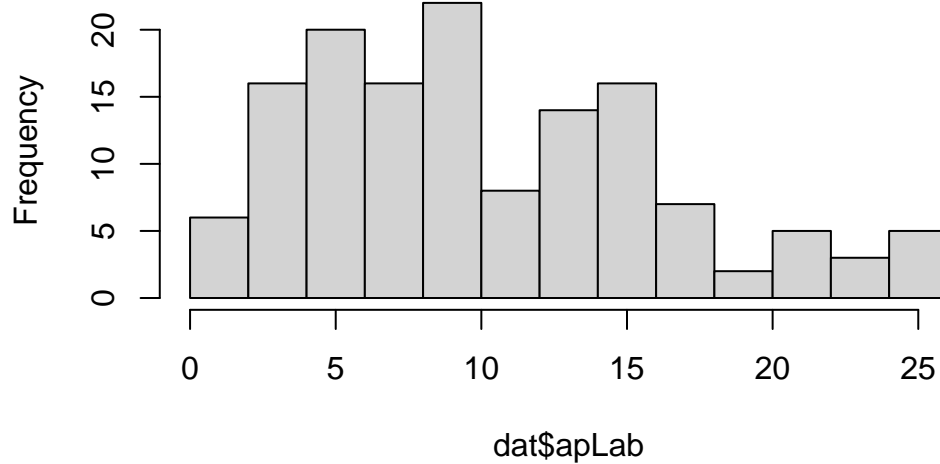

```
hist( dat$apMat )
```

Histogram of dat$apMat
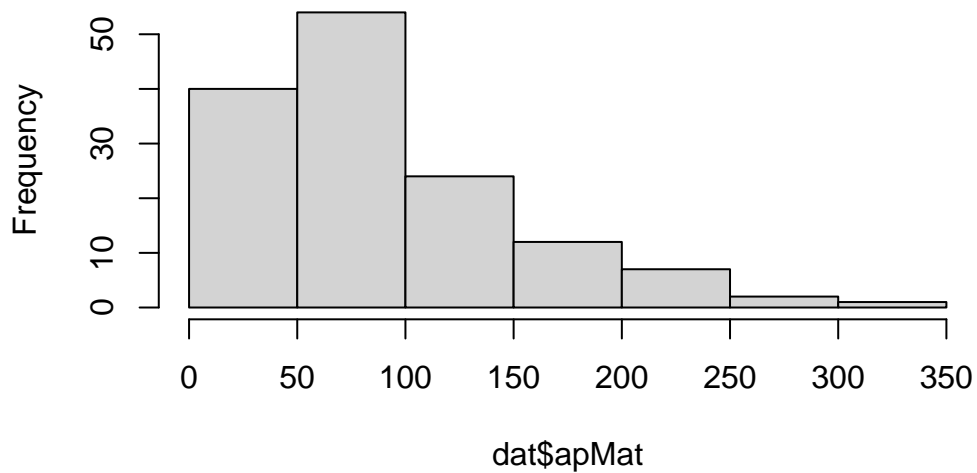
Average product measures vary considerably across firms, with most firms falling into the relatively low-productivity range.
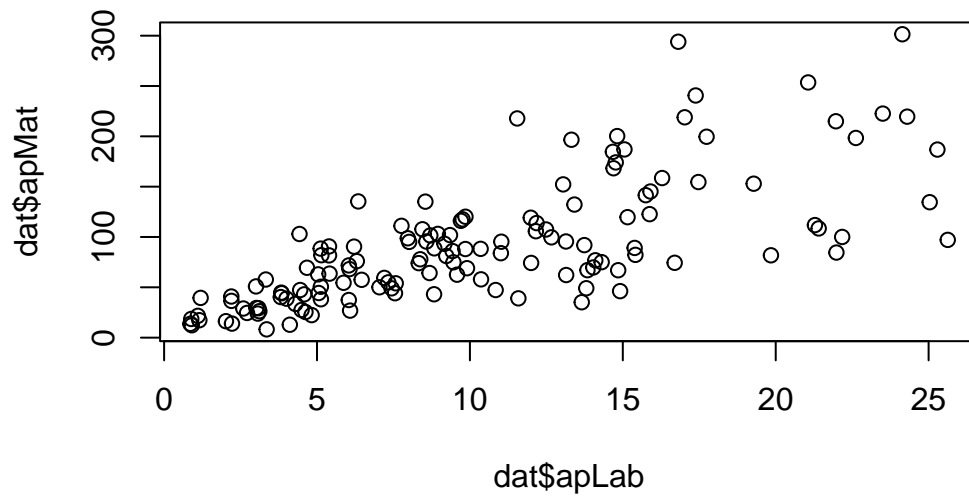
```
# Plotting average partial productivity of one input against another across firms
plot( dat$apCap, dat$apLab )
```
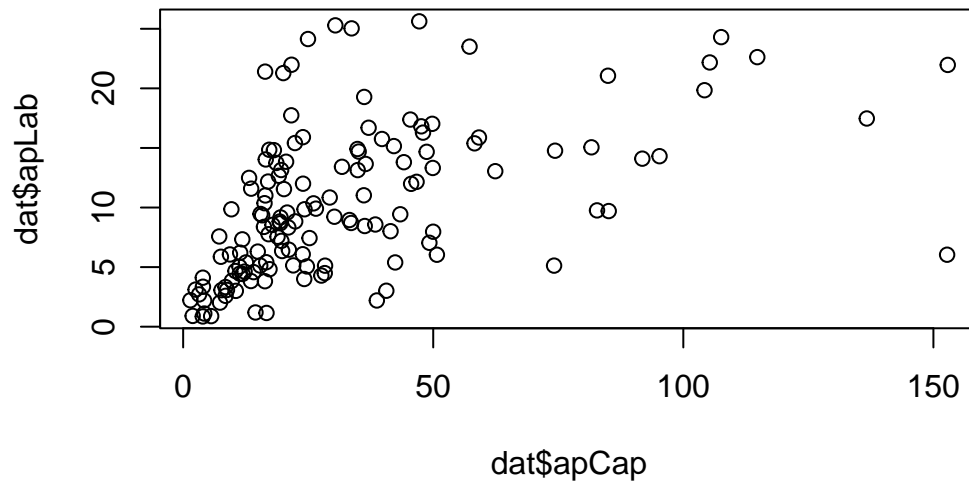


```
plot( dat$apCap, dat$apMat )
```
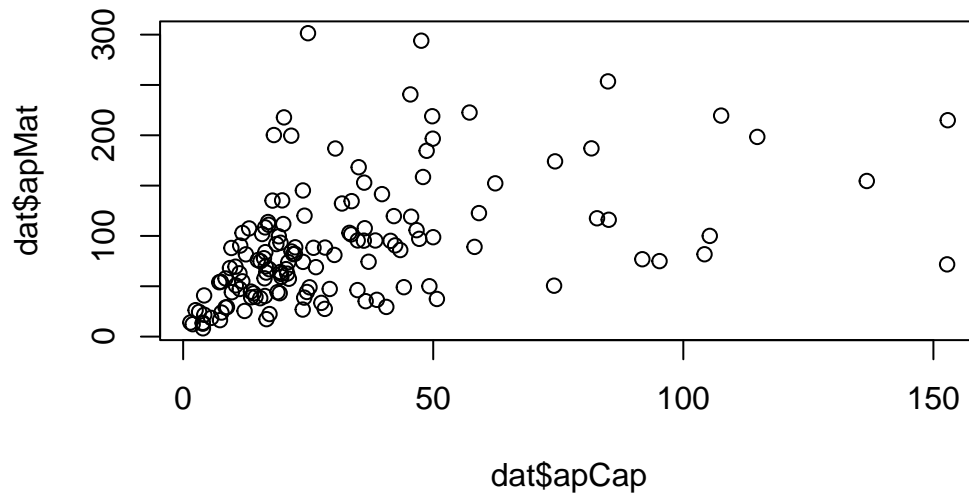


```
plot( dat$apLab, dat$apMat )
```
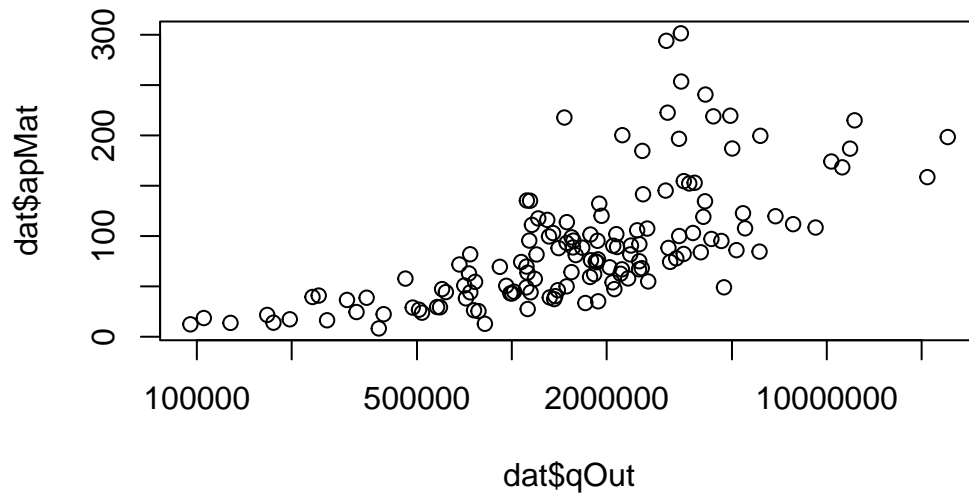
It appears that the average products of the three inputs are positively correlated.

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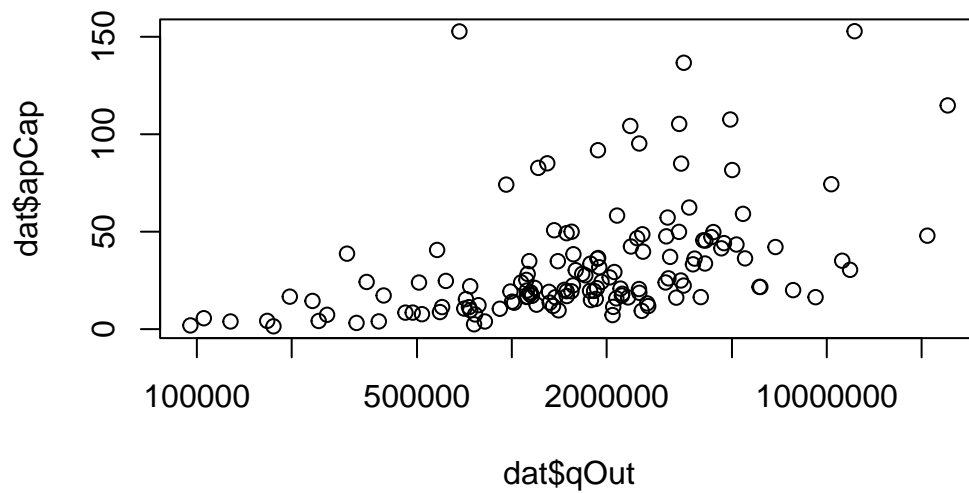


```
plot( dat$qOut, dat$apLab, log = "x" )
```
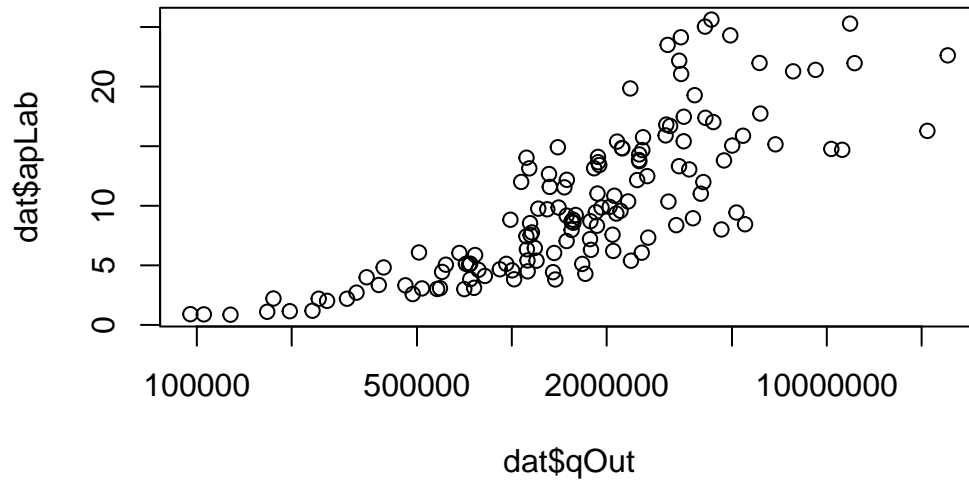
```
plot( dat$qOut, dat$apMat, log = "x" )
```


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
```

```
hist( dat$tfp )
```
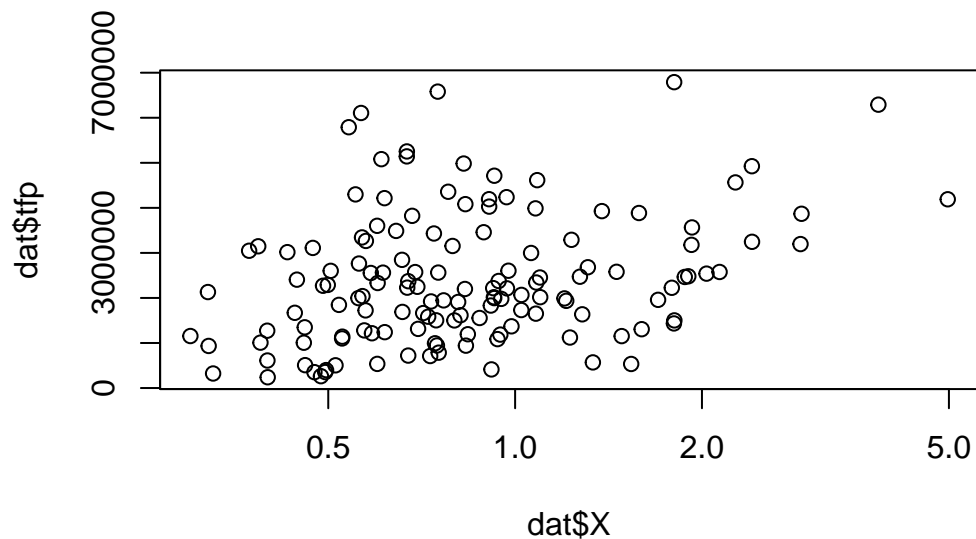
Histogram of dat$tfp
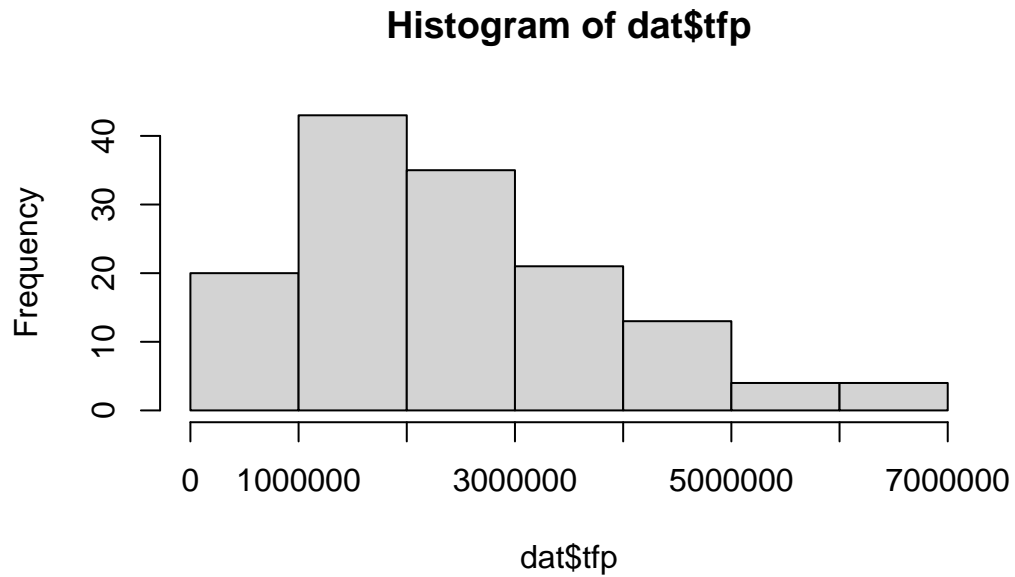
TFP varies considerably across firms, with the majority falling into the relatively low-TFP range.

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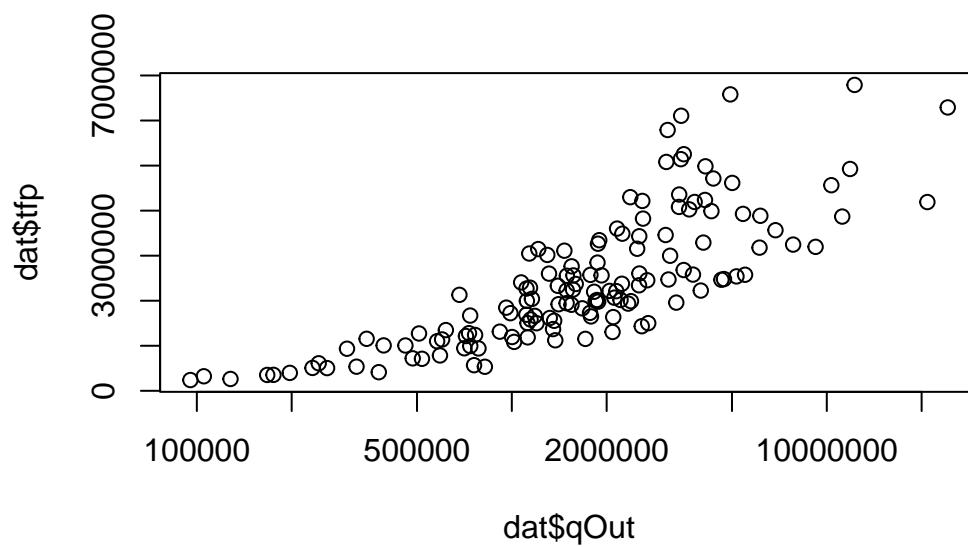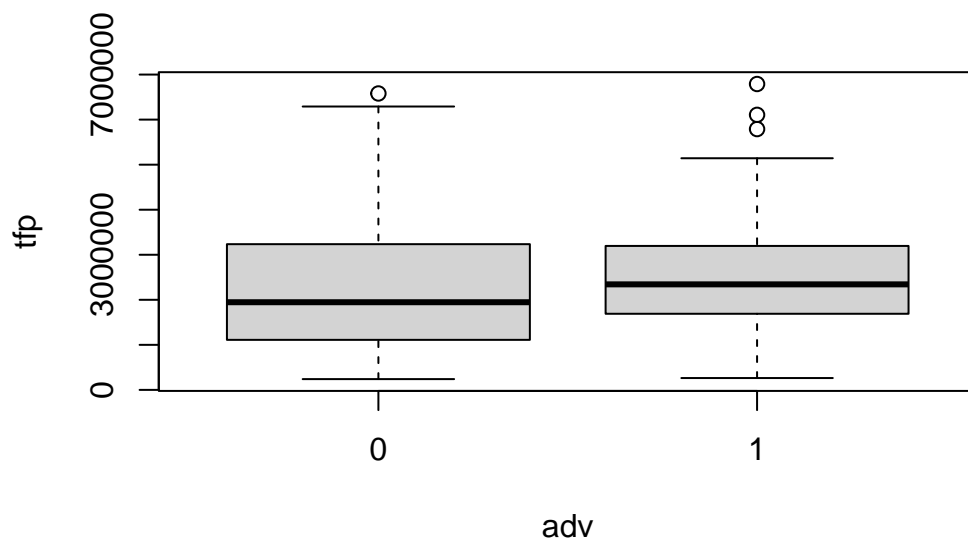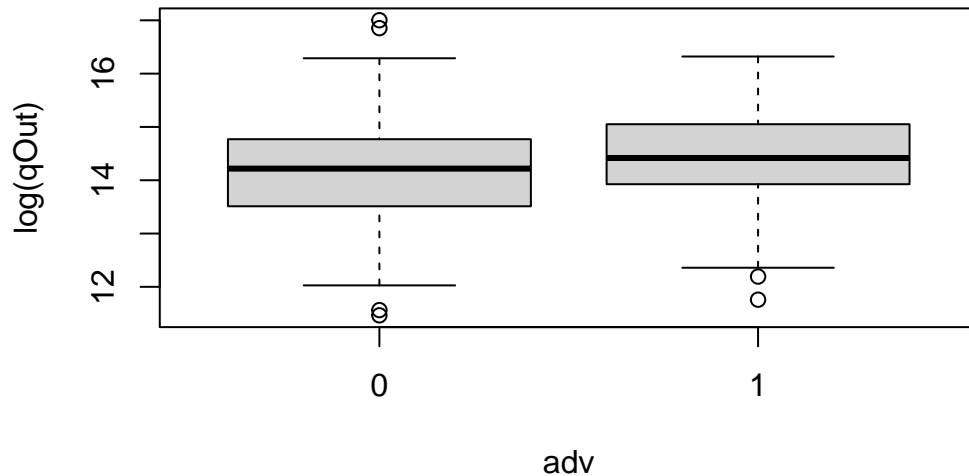
```
plot( dat$X, dat$tfp, log = "x" )
```


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.

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

```
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

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

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

The measure of elasticity provides insight into how easily one input can be substituted for another. When dealing with multiple inputs, we generalize this measure. To do so, the following equivalent representation of the formula, expressed in matrix notation, will be useful.

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

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_{ij}^D 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 $\sigma_{ij}^D = \sigma_{ji}^D$. 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

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

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.

In our analysis below, we will only calculate the direct elasticity of substitution between inputs. For guidance on determining the Allen and Morishima elasticities of substitution, please see (Henningsen 2024).

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 $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, 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:

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

Rearranging the terms, we get the marginal condition:

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

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:

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

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

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.

When estimate a specific form of production function with our data, we will check the properties of the estimated model and evaluate the followings:

- Theoretical consistency

- Productivity and output elasticity

- Returns to scale

- Input substitution

- Profit-maximizing and cost-minimizing behavior

We will start with a production function in linear form.

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

```
# 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) -1615978.639   231771.709  -6.972 0.000000000123189025 ***
qCap               1.788        1.995   0.896                0.372
qLab              11.831        1.272   9.300 0.000000000000000315 ***
qMat              46.668       11.234   4.154 0.000057378294854380 ***
---
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: < 0.00000000000000022
```

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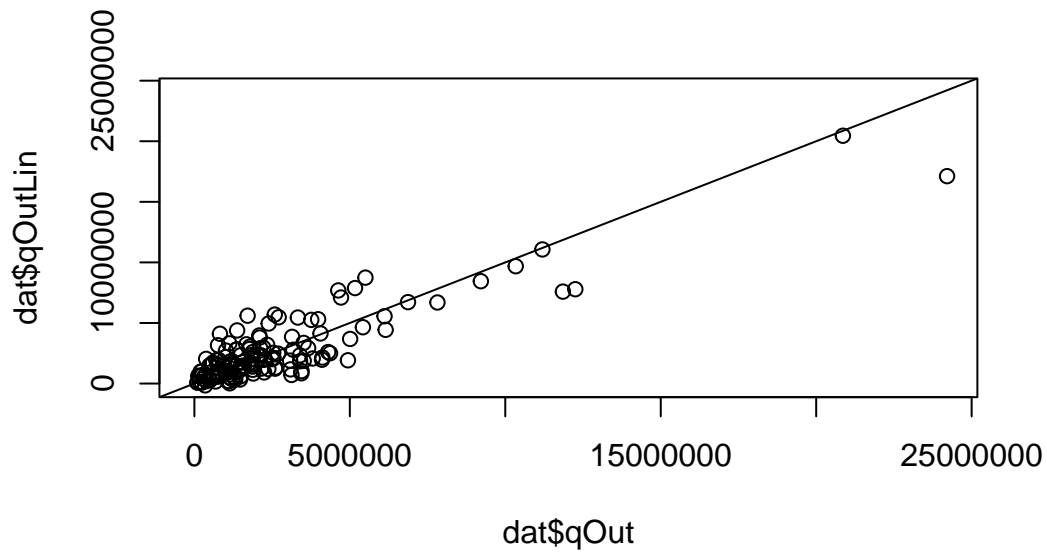
Due to some extreme values, most of the data points appear clustered toward the origin. One way to address this issue is by scaling the axes, for example, using a logarithmic scale. However, since the logarithm of negative values is undefined, we would need to exclude those negative values from the plot.

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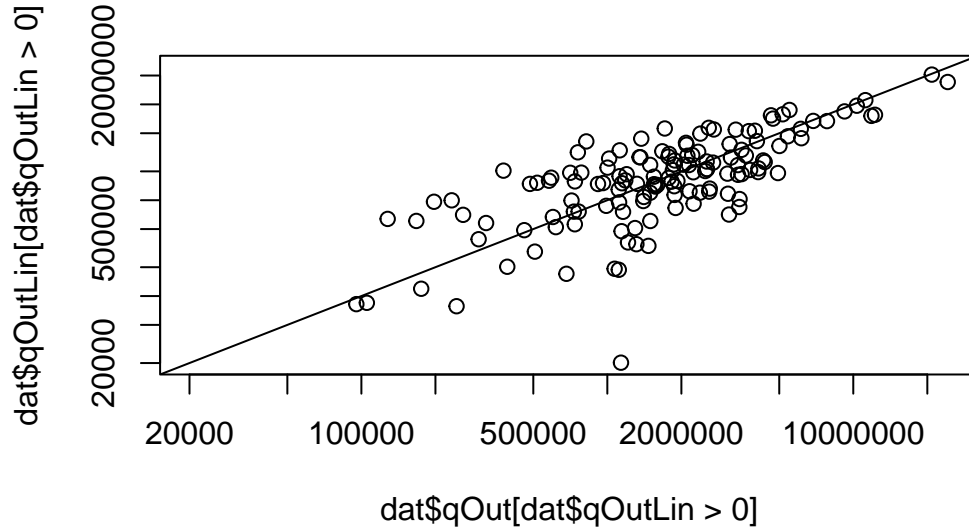
The deviations from observed values looks random (okay) in both scatter plots.

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

Productivity–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
```

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

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

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

```
       eCap          eLab          eMat
0.08063406 1.28627208 0.58741460
```
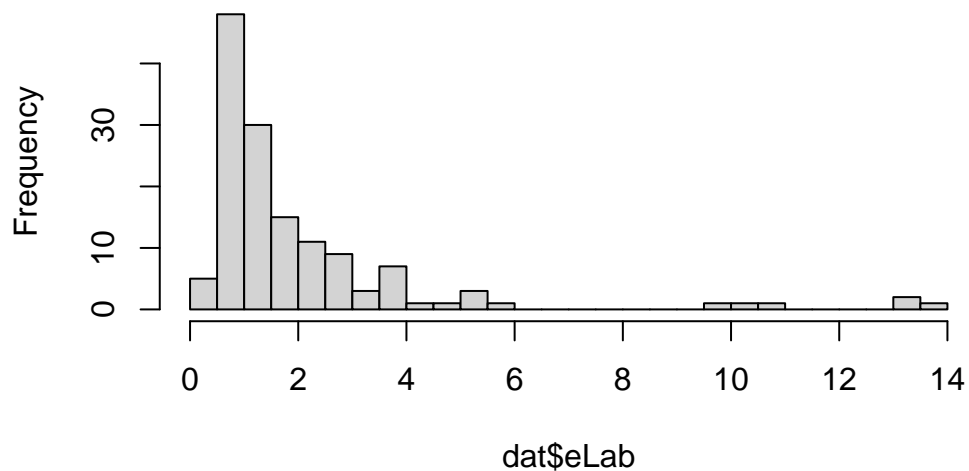
```
hist( dat$eCap , 20)
```

Histogram of dat$eCap


```
sum( dat$eCap > 1)
```

```
[1] 1
```
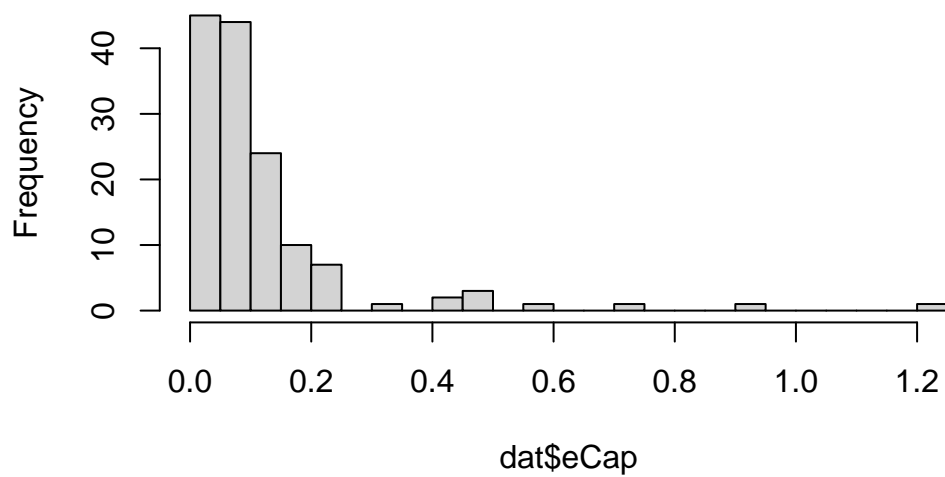
```r
hist( dat$eLab , 20)
```
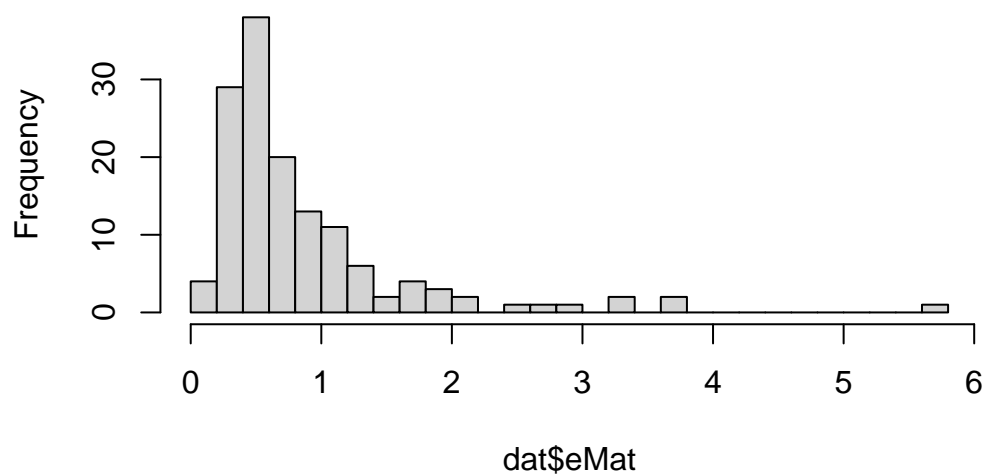
Histogram of dat$eLab


```r
sum( dat$eLab > 1)
```

```
[1] 87
```

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

Histogram of dat$eMat

```
sum( dat$eMat > 1)
```

[1] 36

```
sum( dat$eCap > 1) + sum( dat$eLab > 1) + sum( dat$eMat > 1)
```

[1] 124

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). There are many firms with implausibly high output elasticities ($\varepsilon_i > 1$) of labour and materials. We computed the total number of firms exhibiting an output elasticity of some inputs at a level more than one–the number is quite high, 124 out of 140.

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)
```

Histogram of dat$eScale
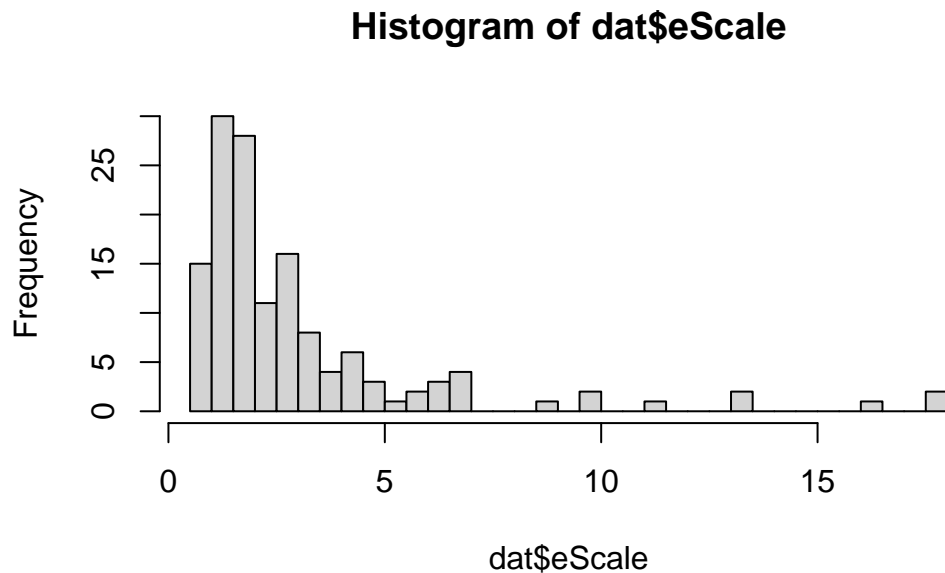

```
sum( dat$eScale > 2)
```

```
[1] 67
```

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 67 firms showing implausibly high returns ($\varepsilon > 2$) 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 )
```
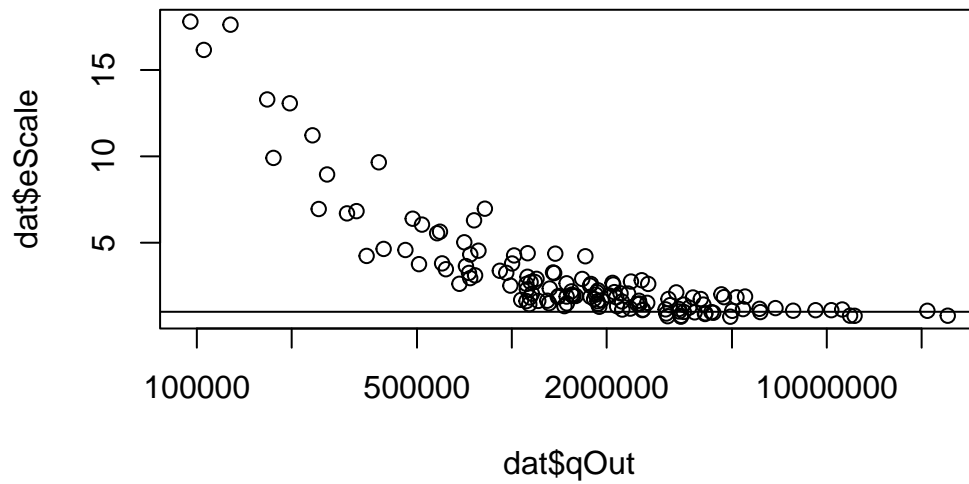
```
plot( dat$X, dat$eScale, log = "x" )
abline( 1, 0 )
```
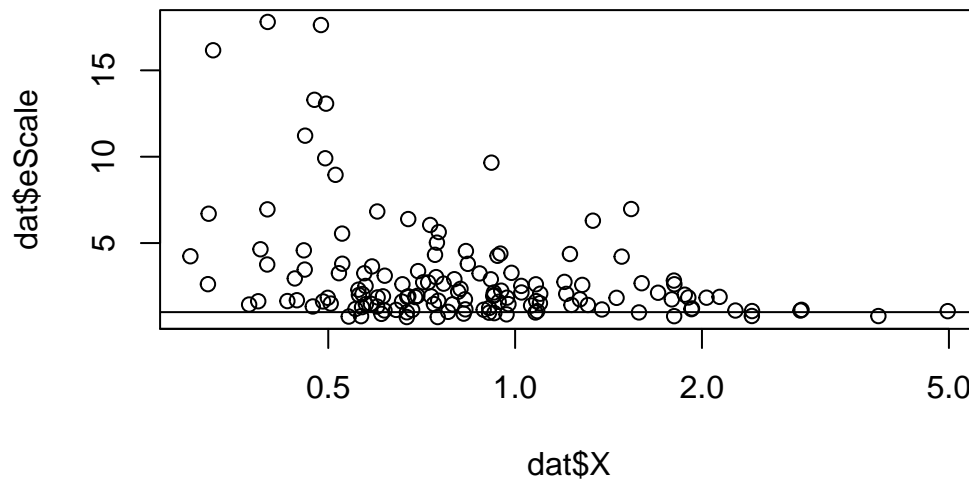

The scatter plots typically show that firms employing smaller input levels exhibit increasing returns to scale, which is expected. A few firms display decreasing returns to scale, and this is observed among firms using both high and low input volumes.

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 (RMRTS) will however vary across firms, due to the variation in output elasticities. Furthermore, since we do not know the units of measurement for the input variables, the interpretation of MRTS is not meaningful. In this context, RMRTS is more useful. As we have shown earlier, it can be expressed as the ratio of output elasticities and is therefore interpreted in a unit-free manner.

```
# 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 )
```

Histogram of dat$rmrtsCapLab
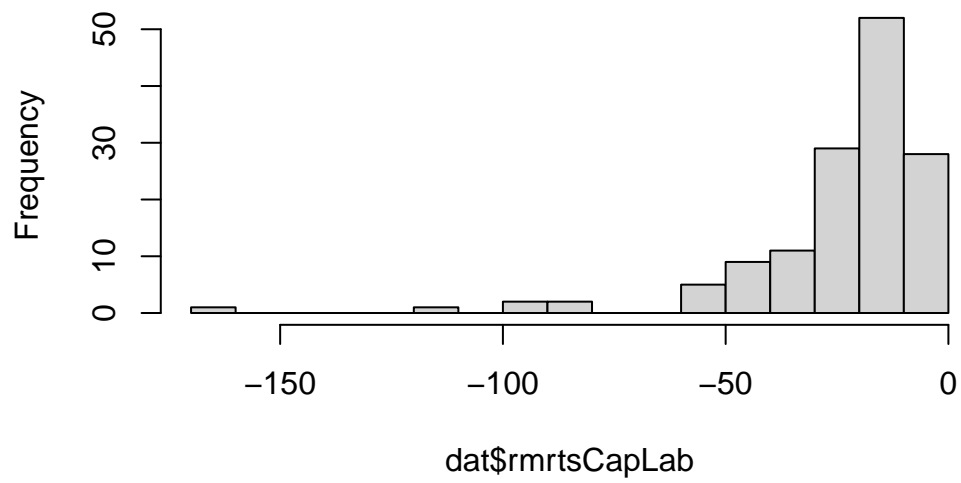

```
hist( dat$rmrtsLabCap, 20 )
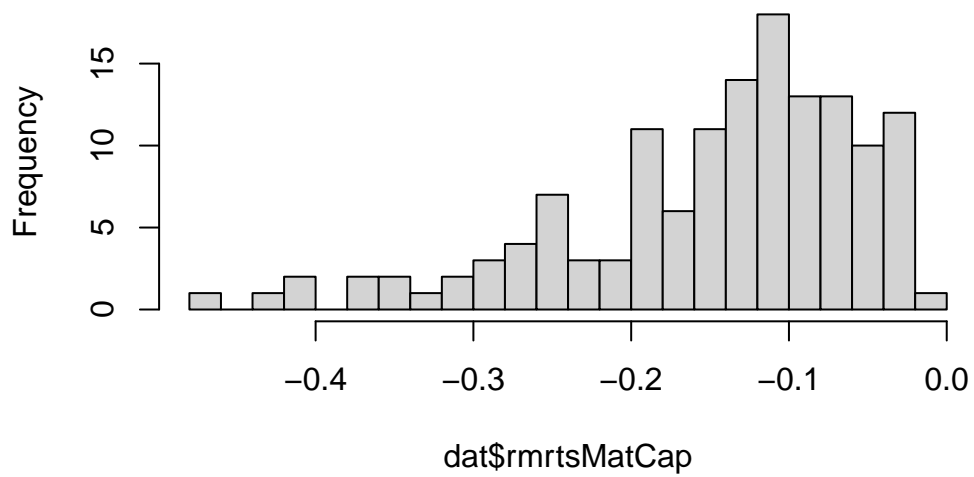```

**Histogram of dat$rmrtsLabCap**
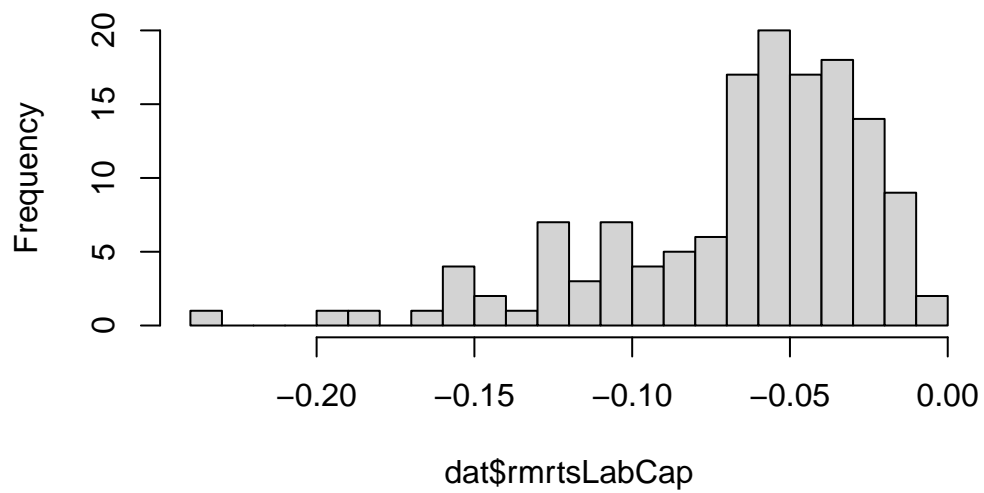


```
hist( dat$rmrtsCapMat, 20 )
```

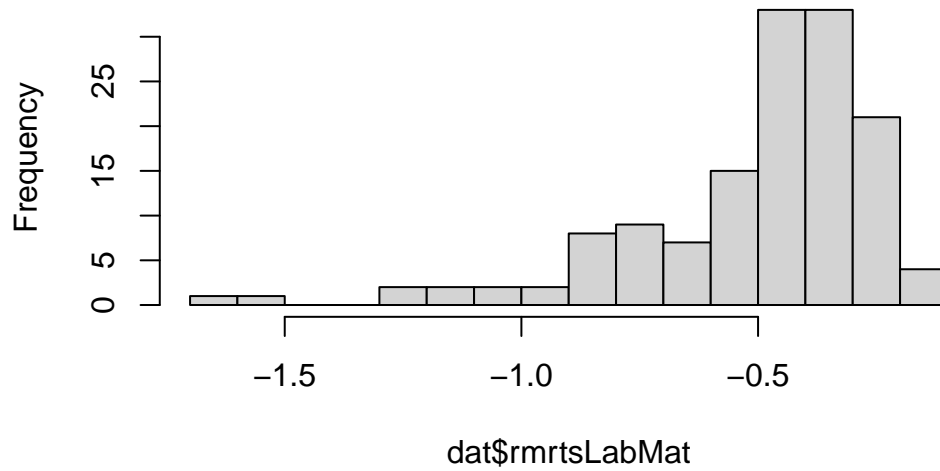## Histogram of dat$rmrtsCapMat
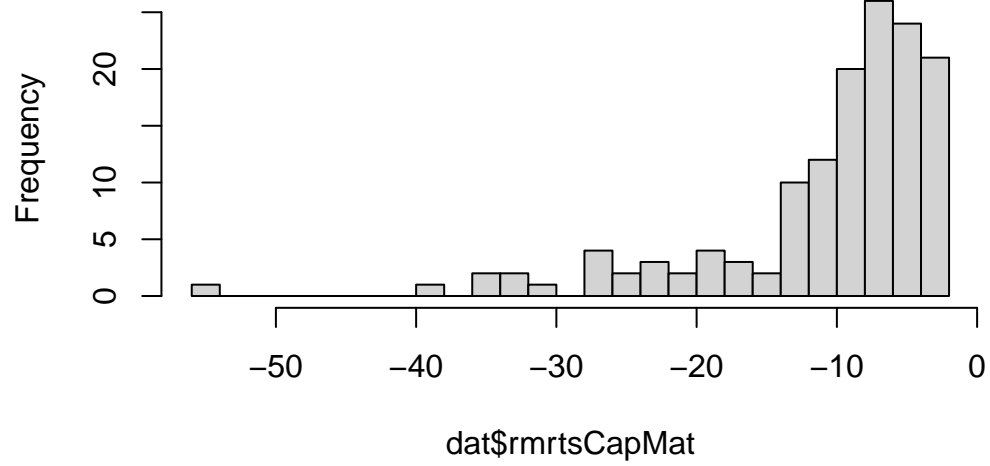

```
hist( dat$rmrtsMatCap, 20 )
```

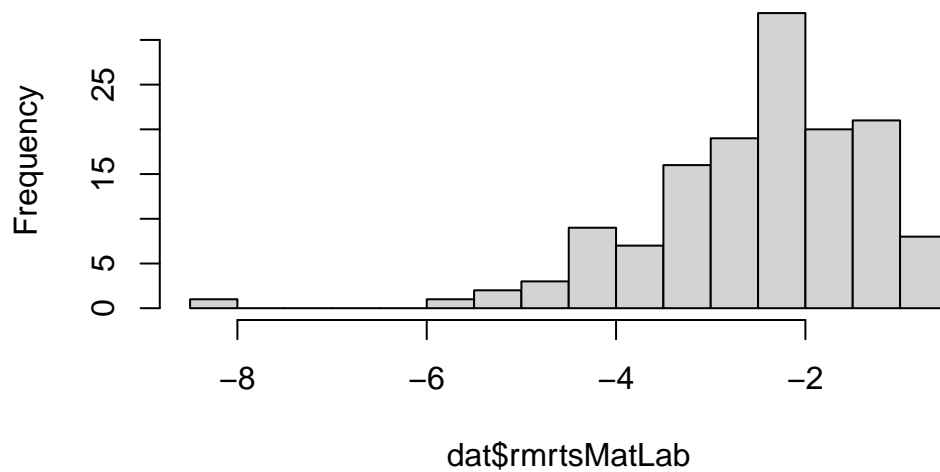## Histogram of dat$rmrtsMatCap


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

Histogram of dat$rmrtsLabMat


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

Histogram of dat$rmrtsMatLab

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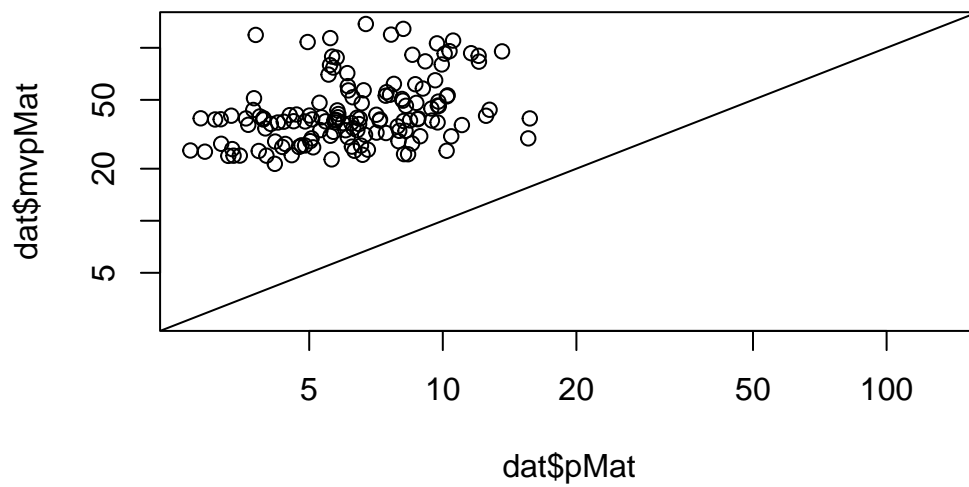
```
# Calculate MVP, plot MVP against input prices
dat$mvpCap <- dat$pOut * coef(prodlinear)["qCap"]
dat$mvpLab <- dat$pOut * coef(prodlinear)["qLab"]
dat$mvpMat <- dat$pOut * coef(prodlinear)["qMat"]
compPlot( dat$pCap, dat$mvpCap, log = "xy" )
```



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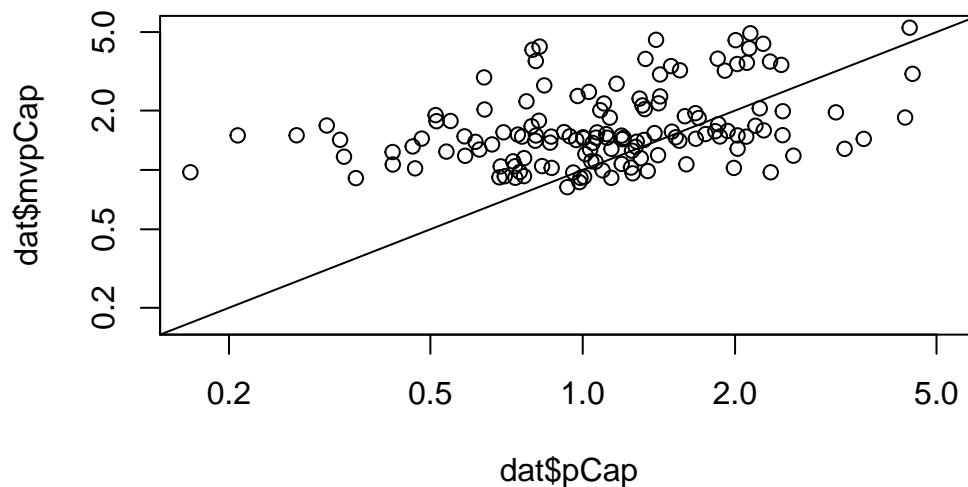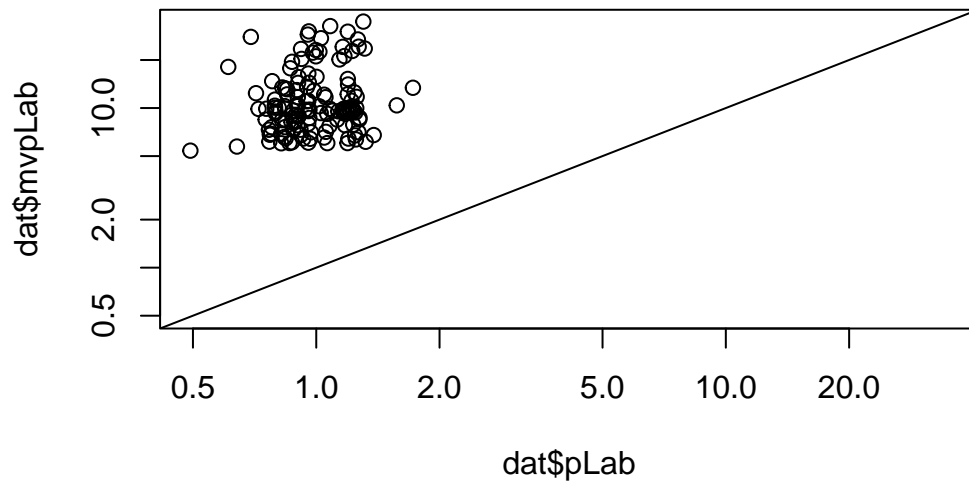
```
compPlot( dat$pMat, dat$mvpMat, log = "xy" )
```


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.

49

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  )
```
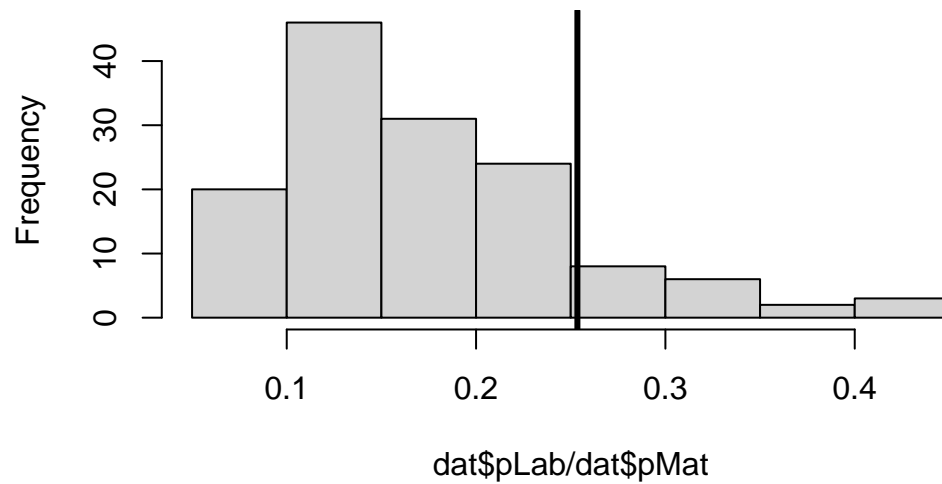

Histogram of dat$pCap/dat$pLab
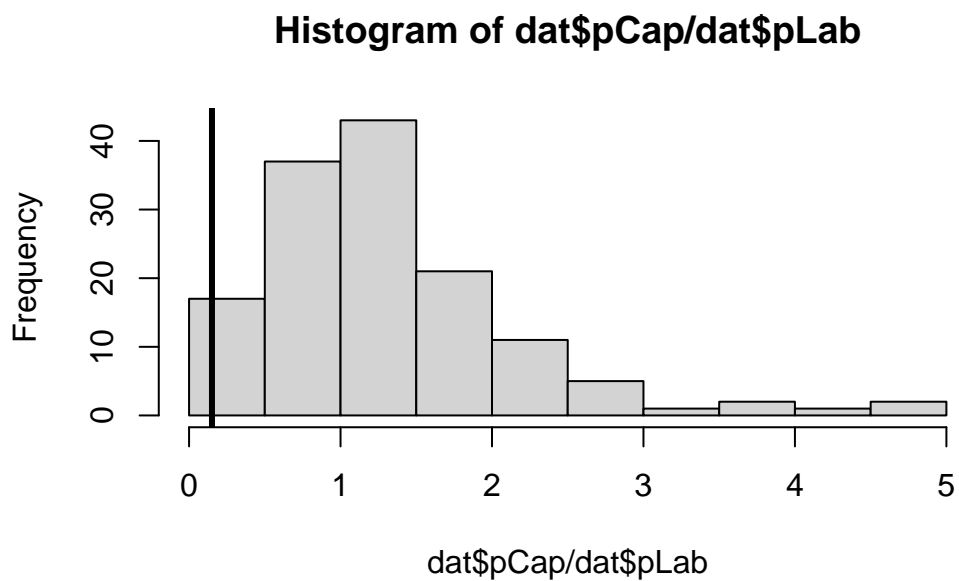
```
hist( dat$pCap / dat$pMat )
abline( v = - mrtsMatCap, lwd = 3  )
```
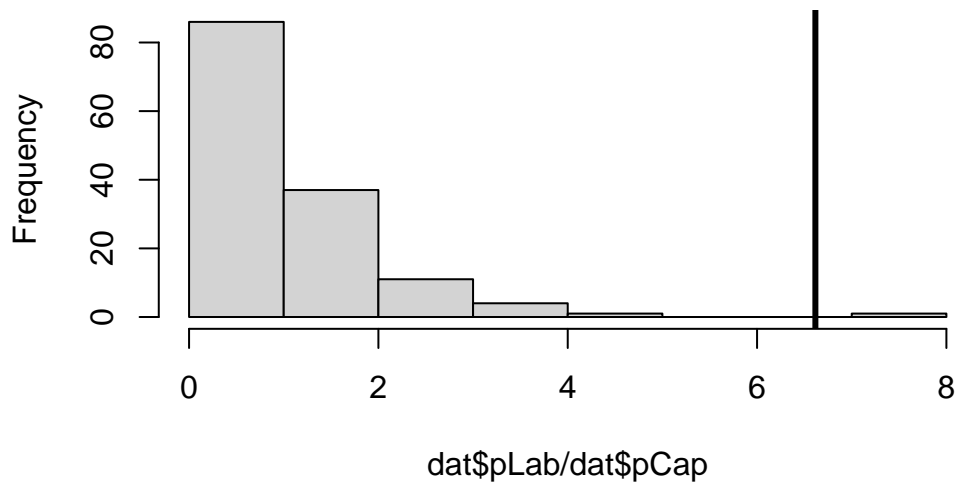
Histogram of dat$pCap/dat$pMat
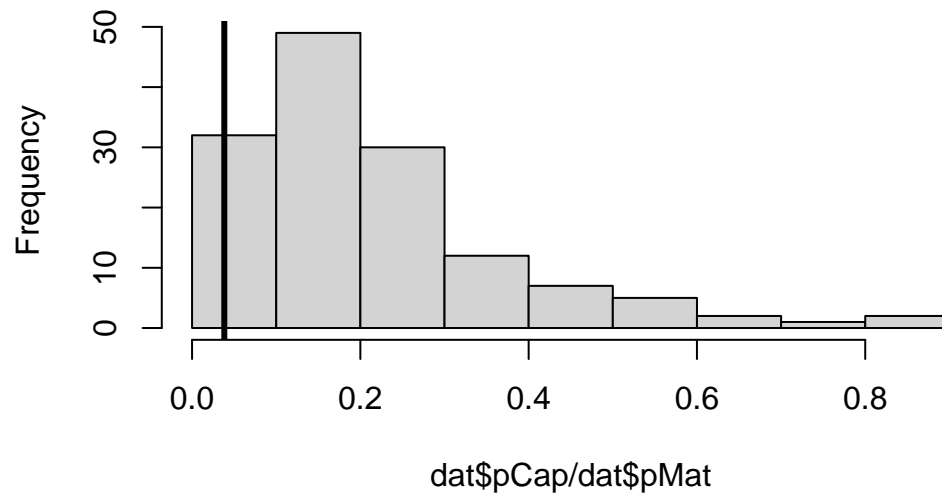

```
hist( dat$pLab / dat$pMat )
abline( v = - mrtsMatLab, lwd = 3  )
```
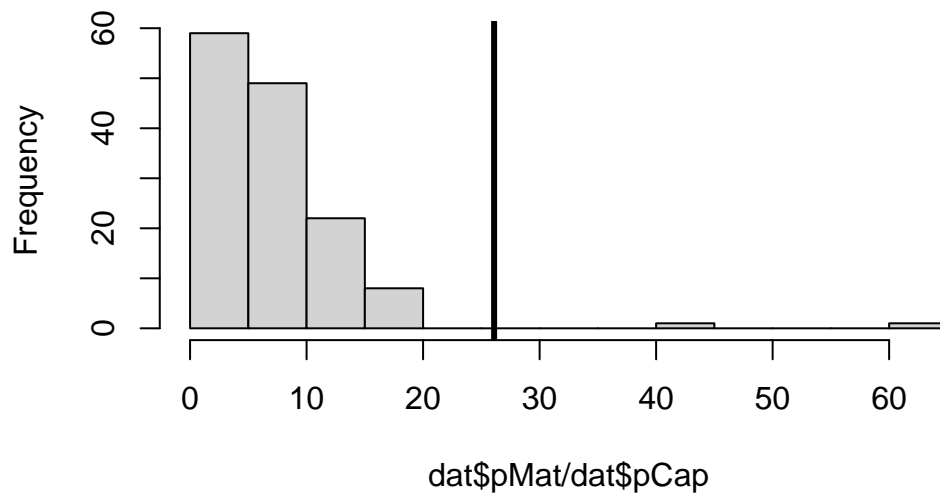
Histogram of dat$pLab/dat$pMat


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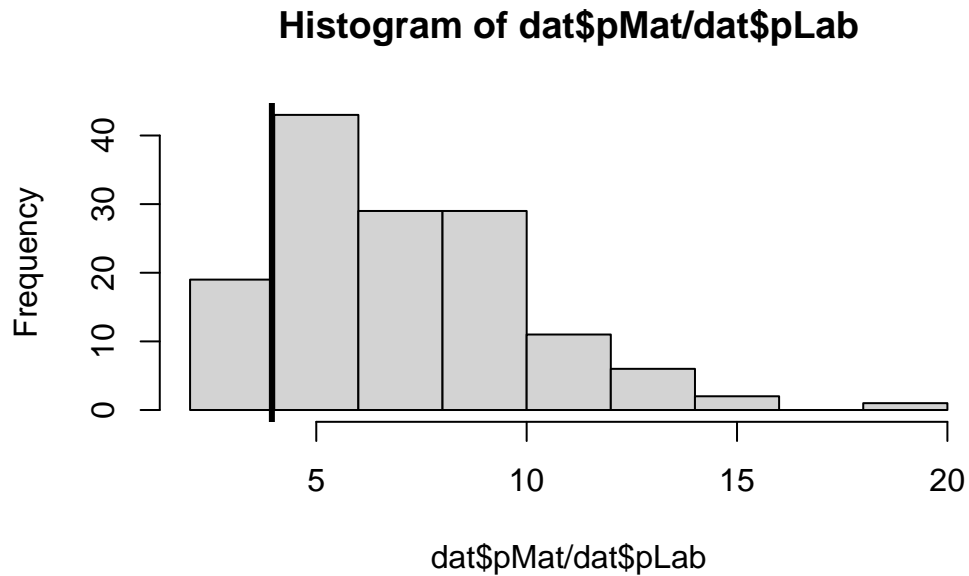
Histogram of dat$pLab/dat$pCap


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

Histogram of dat$pMat/dat$pCap


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

Histogram of dat$pMat/dat$pLab

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

Using similar techniques, we can estimate other production function forms with our data and evaluate the following:

- Theoretical consistency

- Productivity and output elasticity

- Returns to scale

- Input substitution

- Profit-maximizing and cost-minimizing behavior

Below, we will estimate another form: the Cobb-Douglas and compare the two estimated models. Additional specifications (Quadratic form and Translog form) can be found in (Henningsen 2024).

Cobb-Douglas production function

The specification of a Cobb-Douglas production function with N inputs is

$$y = A \prod_{i=1}^{N} x_i^{\alpha_i}$$

Although the specification is not linear, a simple modification allows us to express the relationship in a linear form, enabling us to estimate all parameters using linear regression techniques.

Taking natural log on both sides, we can express the above form (replacing $\ln A$ by α_0) as

$$\ln y = \alpha_0 + \sum_{i=1}^{N} \alpha_i \ln x_i$$

Estimation

```
# Fitting a linearized Cobb-Dougas model
prodCD <- lm( log( qOut ) ~ log( qCap ) + log( qLab ) + log( qMat ), data = dat )
summary( prodCD )
```

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

Residuals:
```

```
     Min       1Q    Median       3Q       Max
-1.67239 -0.28024  0.00667  0.47834  1.30115


Coefficients:
             Estimate Std. Error t value   Pr(>|t|)
(Intercept) -2.06377     1.31259  -1.572     0.1182
log(qCap)    0.16303     0.08721   1.869     0.0637 .
log(qLab)    0.67622     0.15430   4.383 0.00002327 ***
log(qMat)    0.62720     0.12587   4.983 0.00000187 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.656 on 136 degrees of freedom
Multiple R-squared:  0.5943,    Adjusted R-squared:  0.5854
F-statistic: 66.41 on 3 and 136 DF,  p-value: < 0.00000000000000022
```

The regression results show that the coefficients for ln(labor) and ln(materials) are positive and significant, while the coefficient for ln(capital) is not. However, from a theoretical perspective, capital is important, so it would be prudent not to drop it from our estimated model.

Is it a good fit?

Note that the R^2 is 0.59, but it is not directly comparable to the R^2 from a linear model. Later, when comparing various models, we will use a hypothetical R^2 value that measures the correlation between the observed values of the dependent variable and the residuals.

We examine the scatterplot of predicted values against observed values. Note that we must exponentiate the predicted values, as the model was estimated using log-transformed variables. In addition, we scale the axes using a logarithmic scale for finer view; otherwise, we will see a cluster of observation near the origin.

```
# Predicted vs. observed plot
dat$qOutCD <- exp( fitted( prodCD ) )
compPlot( dat$qOut, dat$qOutCD , log = "xy" )
```
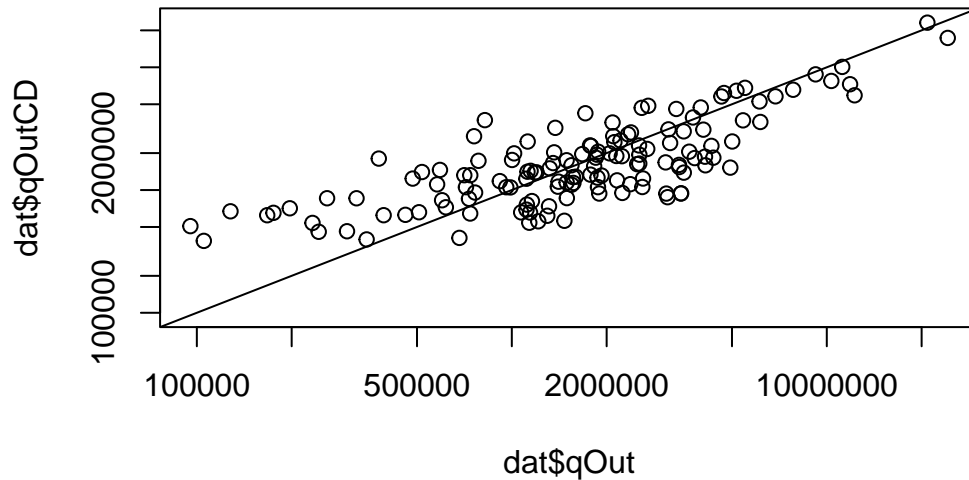

The scatter plot reveals some interesting patterns. For low observed y values, the predicted values are systematically higher than the observed ones, while the opposite is true for large y values. This suggests that the Cobb-Douglas form may not be a good fit, as it systematically overestimates and underestimates the dependent variable within specific ranges of y values.

Theoretical consistency

Essentiality: Theoretically, all inputs are essential in Cobb-Douglas form (if we drop any, the output becomes zero). Since ln(zero) is undefined, the estimated form trivially satisfies essentiality.

Monotonicity: The positive and significant coefficients of labour and materials satisfy the monotonicity condition.

Non-negativity: Although the intercept term is negative, $exp(\alpha_0)$ is positive, and so predicted output remains positive.

Studying properties of the production function

Productivity–Output elasticity

The estimated linear production function implicitly assumes that all firms share the same marginal productivity. However, this is not the case with the Cobb-Douglas production function. In the Cobb-Douglas form, it can be shown that output elasticity is fully determined by the model parameters, implying that the estimated model assumes that all firms have the same output elasticity.

Observe that

$$MP_i = \frac{\partial y}{\partial x_i} = A\frac{\partial \prod_{i=1}^{N} x_i{}^{\alpha_i}}{\partial x_i} = A\alpha_i x_i^{\alpha_i - 1} \prod_{j \neq i}^{N} x_j{}^{\alpha_j} = \frac{\alpha_i y}{x_i}$$

and

$$\varepsilon_i = \frac{\partial y}{\partial x_i} \cdot \frac{x_i}{y} = \frac{\alpha_i y}{x_i} \cdot \frac{x_i}{y} = \alpha_i$$

Therefore, the elasticity of each input is directly measured by the linear model coefficients—0.16 for capital, 0.67 for labor, and 0.62 for materials. This means that, for example, a one-percent change in labor will result in an average 0.67 percent change in output for these firms.

Comparing these numbers with our findings from the linear model, where we observe a distribution of elasticity measures across the firms in our dataset, we see that the elasticity measures for labor and materials are considerably smaller in the Cobb-Douglas form, while the effect is opposite for capital.

Below, we calculate marginal productivity (MP) at the firm level and examine its frequency distribution. Since MP can be calculated at a specific output level, we compute it at the observed output level y for each firm. If our model is a *good fit*, we could also

compute it using the predicted y values. The distribution of MP values from this approach might provide a clearer picture of the true underlying distribution of productivity. See the discussion in (Henningsen 2024).

```
# Compute marginal products
dat$mpCapCD <- coef(prodCD)["log(qCap)"] * dat$qOut / dat$qCap
dat$mpLabCD <- coef(prodCD)["log(qLab)"] * dat$qOut / dat$qLab
dat$mpMatCD <- coef(prodCD)["log(qMat)"] * dat$qOut / dat$qMat
# Draw histograms
hist( dat$mpCapCD )
```
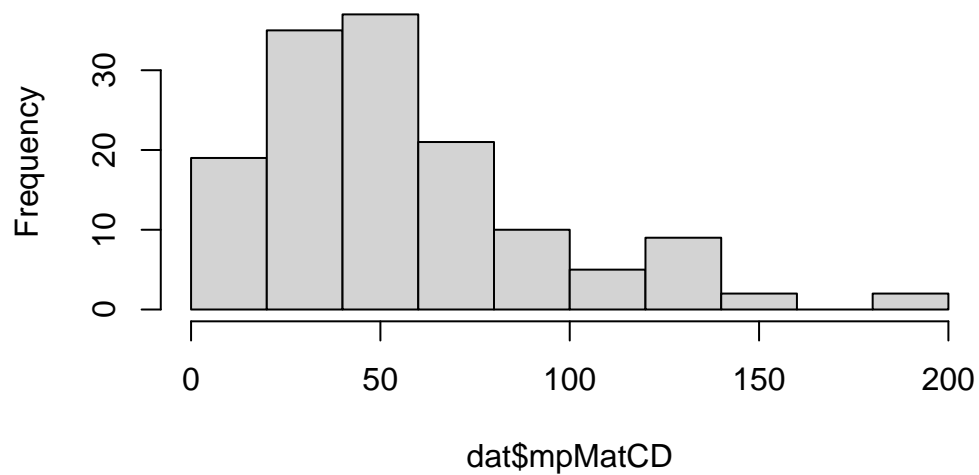

Histogram of dat$mpCapCD
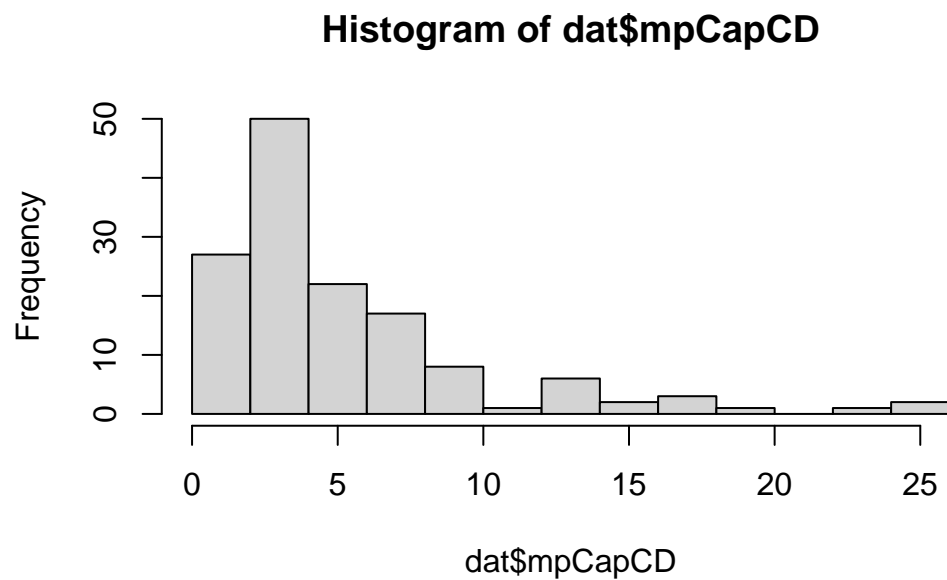
```
hist( dat$mpLabCD )
```

Histogram of dat$mpLabCD
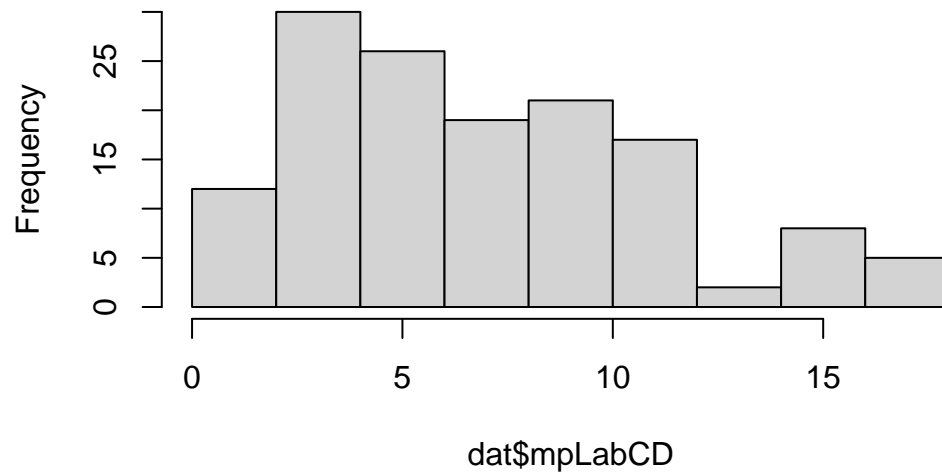

```
hist( dat$mpMatCD )
```

Histogram of dat$mpMatCD

If firms increase capital input by one unit, the output for most firms will rise by between 0 and 8 units. Similarly, a one-unit increase in labor input leads to an output increase of between 2 and 12 units for most firms, while a one-unit increase in materials results

in an output boost of between 20 and 80 units. See the histograms of MPs calculated at the predicted y values in (Henningsen 2024), which show similar results.

Returns to scale

Recall that the elasticity of scale is the sum of output elasticities, which in this case, is measured as $\varepsilon = \sum_i \alpha_i = 1.466$

As with our output elasticity measures, the Cobb-Douglas model implicitly assumes a constant elasticity of scale across all firms. On average, it indicates increasing returns to scale, suggesting that firms could benefit from scaling up production.

Input substitution

As before, we calculate both MRTS and RMRTS, each with its own advantages and disadvantages in this context. Interpreting MRTS is challenging because the units of measurement for the input variables are unknown. However, unlike in the case of a linear production function, MRTS varies across firms, offering a broader perspective on how firms differ in their input substitution choices. RMRTS, on the other hand, provides a unit-free interpretation, but in the case of the Cobb-Douglas production function, it remains globally constant across firms.
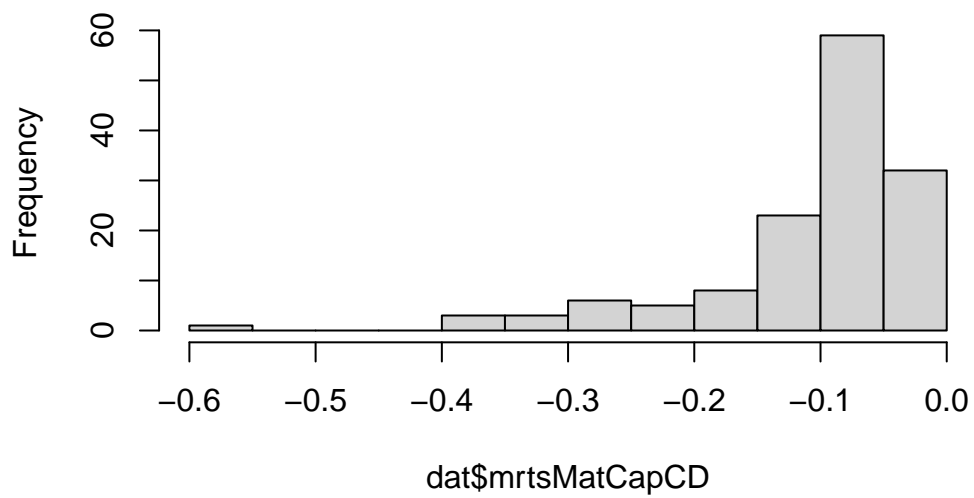
```
# Calculate MRTS and draw histograms
dat$mrtsCapLabCD <- - dat$mpLabCD / dat$mpCapCD
dat$mrtsMatCapCD <- - dat$mpCapCD / dat$mpMatCD
dat$mrtsMatLabCD <- - dat$mpLabCD / dat$mpMatCD
hist( dat$mrtsCapLabCD )
```
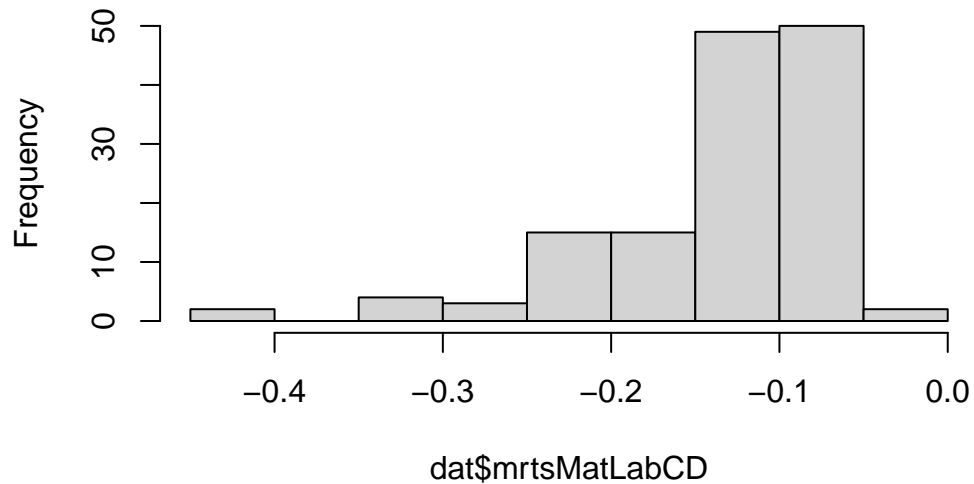
Histogram of dat$mrtsCapLabCD
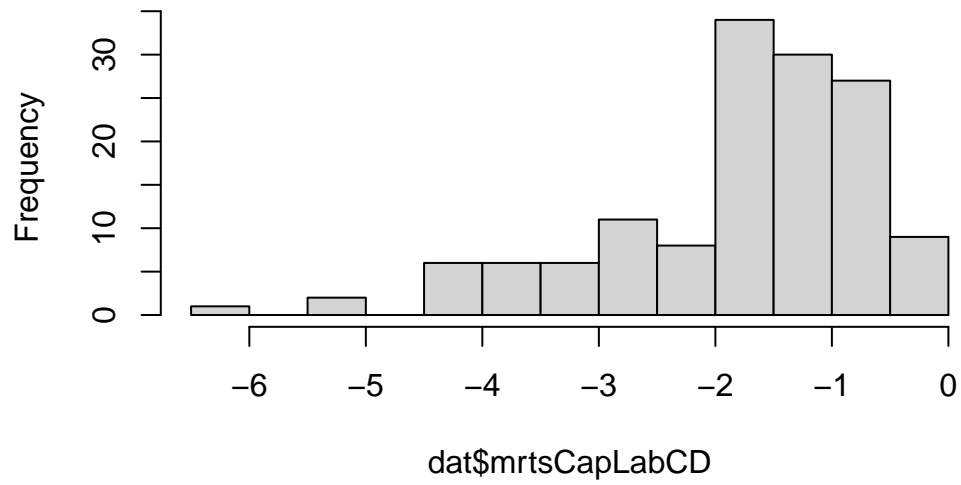

```
hist( dat$mrtsMatCapCD )
```

Histogram of dat$mrtsMatCapCD


```
hist( dat$mrtsMatLabCD )
```

61

Histogram of dat$mrtsMatLabCD

According to the MRTS based on the Cobb-Douglas production function, most firms require between 0.5 and 2 additional units of capital and between 0 and 0.15 additional units of materials to compensate for a one-unit reduction in labor. These findings are very similar to those from the linear production function case.

```
# Calculate RMRTS
rmrtsCapLabCD <- - coef(prodCD)["log(qLab)"] / coef(prodCD)["log(qCap)"]
rmrtsCapLabCD
```

```
log(qLab)
-4.147897
```

```
rmrtsMatCapCD <- - coef(prodCD)["log(qCap)"] / coef(prodCD)["log(qMat)"]
rmrtsMatCapCD
```

```
 log(qCap)
-0.2599291
```

```
rmrtsMatLabCD <- - coef(prodCD)["log(qLab)"] / coef(prodCD)["log(qMat)"]
rmrtsMatLabCD
```

```
log(qLab)
-1.078159
```

RMRTS should be interpreted in percentage terms. If a firm wants to reduce labor use by one percent, it must increase capital use by 4.15 percent or material use by 1.08 percent to maintain the same output level.

An important property of Cobb-Douglas production function is that the direct elasticities of substitution are always equal to one. As a result, the estimated model will not offer any insights into this aspect.

Profit-maximizing behavior

The profit-maximizing principle suggests that the marginal value 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).
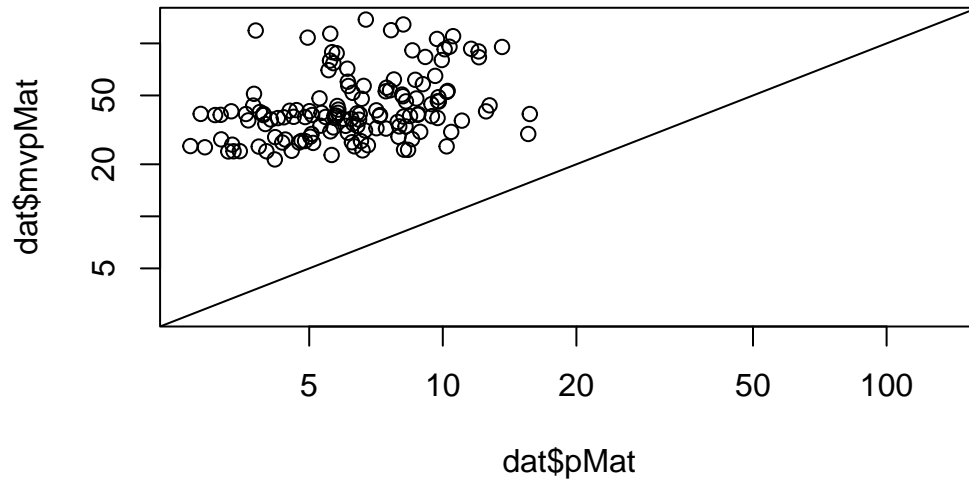
```
# Calculate MVP, plot MVP against input prices
dat$mvpCapCd <- dat$pOut * dat$mpCapCD
dat$mvpLabCd <- dat$pOut * dat$mpLabCD
dat$mvpMatCd <- dat$pOut * dat$mpMatCD
compPlot( dat$pCap, dat$mvpCap, log = "xy" )
```

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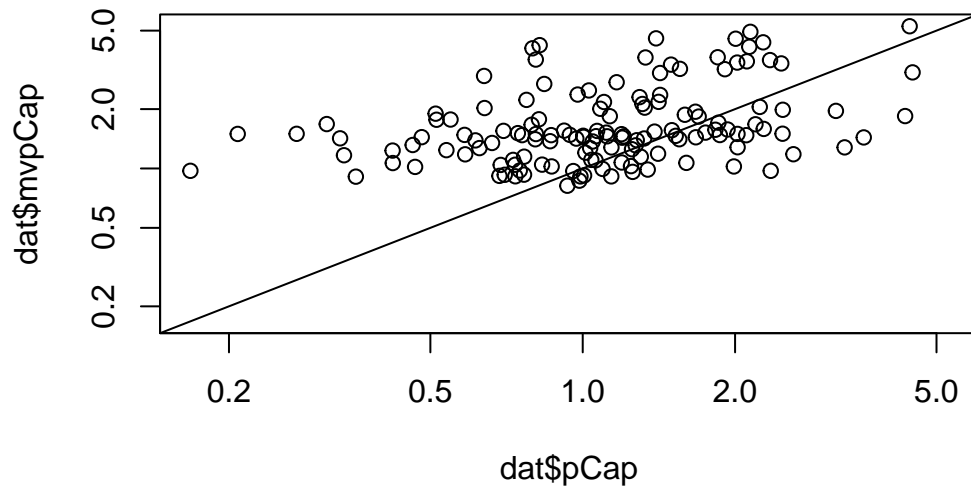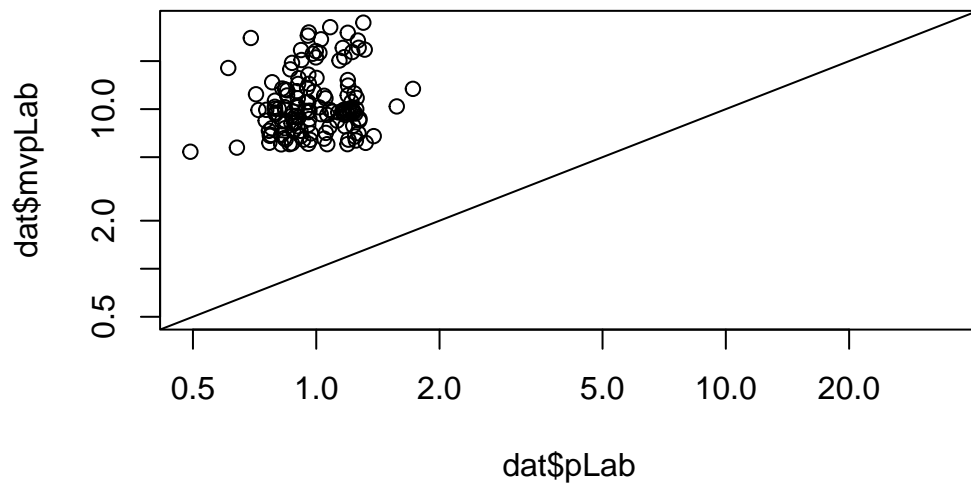


```
compPlot( dat$pMat, dat$mvpMat, log = "xy" )
```

These plots show that the marginal value products are nearly always equal to or higher than the corresponding input prices. Therefore, most firms could increase their profits by using more of all inputs. Given that the estimated Cobb-Douglas technology exhibits an increasing returns to scale, it is not surprising that firms would benefit from increasing input quantities. These observations are also consistent with our findings from the estimated linear production specification. As we pointed out earlier, their decision not to employ more inputs might be related to imperfections in the input markets.

Cost-minimizing behavior
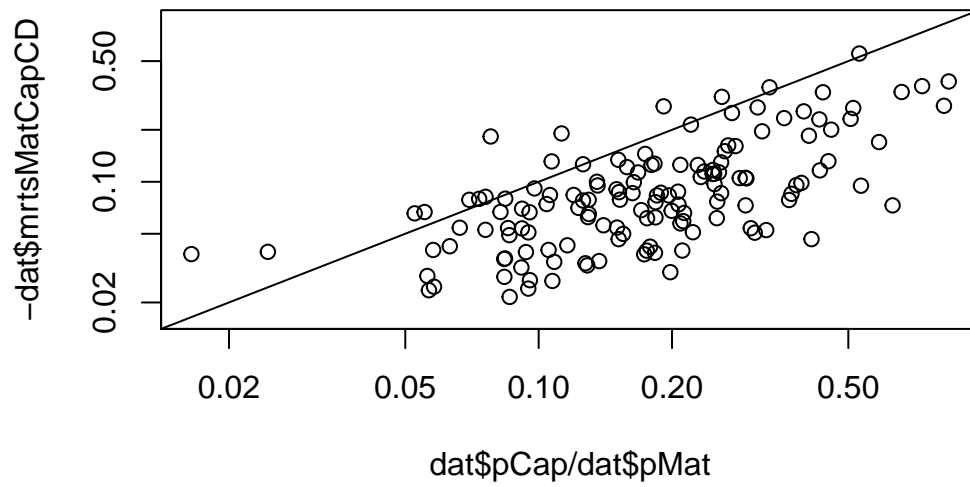
The cost-minimizing principle suggests that the ratio of input prices must equal the absolute value of the MRTS between two inputs.
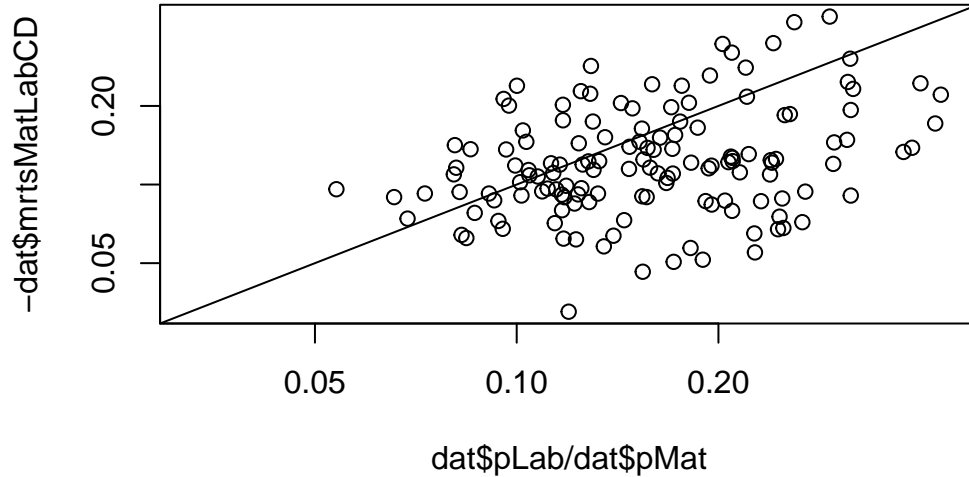
```
# Plot MRTS against input price ratios
compPlot( dat$pLab / dat$pCap, - dat$mrtsCapLabCD, log = "xy" )
```
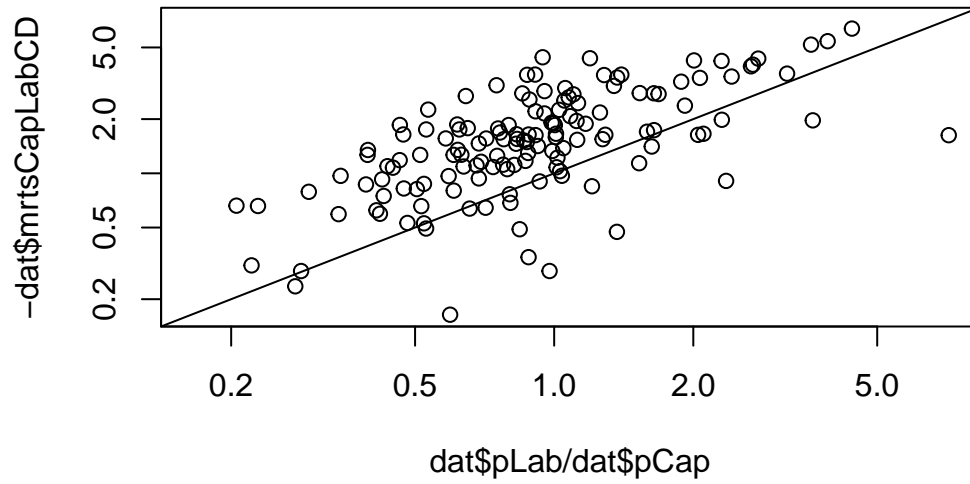
```
compPlot( dat$pCap / dat$pMat, - dat$mrtsMatCapCD, log = "xy" )
```



```
compPlot( dat$pLab / dat$pMat, - dat$mrtsMatLabCD, log = "xy" )
```

Most firms could benefit from substituting capital with labor, and capital with materials. These observations align with our findings based on the profit-maximizing principle.

Comparison of the two models

When one functional form is nested within another (for example, linear vs. quadratic production function, or Cobb-Douglas vs. Translog production function), standard statistical tests can be used to compare the two models; see the sections on estimating quadratic or Translog production functions in(Henningsen 2024). In those cases, the tests rejected the linear production function in favor of the quadratic, while the evidence for rejecting the Cobb-Douglas in favor of the Translog production function was less conclusive.

However, in our discussion, we are comparing two non-nested models—Linear and Cobb-Douglas. Comparing non-nested models is more complex.

We will make a series of observations to compare the pros and cons of these two models based on our data.

Since the two models have different dependent variables, y and ln(y), we cannot directly compare their R-squared values. Instead, we compare the R-squared from the linear model with a hypothetical R-squared based on the estimated Cobb-Douglas model,

which measures the correlation between the observed y values and the residuals. Alternatively, we can reverse the process by comparing the R-squared from the Cobb-Douglas model with a hypothetical R-squared based on the estimated linear model, measuring the correlation between the observed ln(y) values and the residuals.

```
# Linear model: R-square of y
summary(prodlinear)$r.squared
```

[1] 0.786819

```
# Cobb-Douglas: Hypothetical R-square of y
rSquared( dat$qOut, dat$qOut - dat$qOutCD )
```

```
         [,1]
[1,] 0.8067198
```

```
# Linear model: Hypothetical R-square of ln(y)
rSquared( log( dat$qOut[ dat$qOutLin > 0 ] ), log( dat$qOut[ dat$qOutLin > 0 ] ) - 1
```

```
         [,1]
[1,] 0.3796727
```

```
# Cobb-Douglas: R-square of ln(y)
summary(prodCD)$r.squared
```

[1] 0.5943154

While the two models provide similar R-squared values regarding y, the R-squared value regarding ln(y) is considerably higher for the Cobb-Douglas function than for the linear function. However, as we have pointed out before, the scatter plot of observed against predicted values shows a different picture between the two models. It appears that the

Cobb-Douglas form is systematically overestimating or underestimating y values for a broad range of firms.

Based on our analysis above, we have seen that the estimated linear model also failed in some aspects. For example, it predicted a negative output value for one observation and showed implausibly high output elasticities and elasticity of scale for several firms.

We apply Ramsey's Regression Equation Specification Error Test (RESET) on the two functional forms using the resettest command.

```
# RESET test
resettest( prodlinear )
```

```
        RESET test

data:  prodlinear
RESET = 17.639, df1 = 2, df2 = 134, p-value = 0.0000001584
```

```
resettest( prodCD )
```

```
        RESET test

data:  prodCD
RESET = 2.9224, df1 = 2, df2 = 134, p-value = 0.05724
```

The linear form is clearly rejected, while the Cobb-Douglas form is accepted at the 5 percent significance level but rejected at the 10 percent level. Additional specifications—Quadratic and Translog production functions—are estimated in (Henningsen 2024), where the Translog production function appears to provide a better fit among the four specifications. Below we listed various criteria for comparative assessment of the two models.

Assessment criteria	Linear	Cobb-Douglas
R-square of y	0.786819	0.8067198
R-square of ln y	0.3796727	0.5943154
Predicted vs Observed plot	(+)	(-)
Negative output prediction	1	0
Monotonicity violation	0	0
Implausible output elasticities	124	0
Implausible scale elasticity	67	0
RESET (p-value)	0.0000001584	0.05724

Reading materials:

- Henningsen, chapter 2

- Varian, Chapter 2, 4

Section 3

Cost function

So far we have followed the primal approach, which analyses producers' behaviors (profit-maximizing or cost-minimizing) given a technological constraint in the form of a production function. Next, we will follow the dual approach.

Observe that a firm's optimizing behavior, for example, choosing inputs to minimize its expenditure keeping its level of production at a certain given output level, gives us a relationship between the firm's minimum expenditure and the input prices (and output volume). We refer to this relationship as the firm's cost function. The duality in optimization implies that if we begin with a production function and derive its cost function, we can take that cost function and use it to generate a production function. If the original production function is quasiconcave, the derived production function will be identical to it. If the original production function is not quasiconcave, the derived production function is a 'concavication' of it. Moreover, any function with all the properties of a cost function generates some production function for which it is the cost function.

This last observation has important implications for applied work. Applied researchers need no longer begin their study of the firm with detailed knowledge of the technology and with access to relatively obscure engineering data. Instead, they can estimate the firm's cost function by employing observable market input prices and levels of output. They can then 'recover' the underlying production function from the estimated cost function.

Recall that the cost function, defined for all input prices $\mathbf{w}$ and all output levels y is the minimum-value function

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

where $\mathbf{x} = (x_1, ..., x_n)$ is the cost-minimizing demand of input i.

The cost function has the following properties:

Theorem: If f is continuous and strictly increasing, then $c(y, \mathbf{w})$ is

1. Zero when $y = 0$,

2. Continuous on its domain,

3. For all $\mathbf{w} \gg 0$, strictly increasing and unbounded above in y,

4. Increasing in $\mathbf{w}$,

5. Homogeneous of degree one in $\mathbf{w}$,

6. Concave in $\mathbf{w}$.

Moreover, if f is strictly quasiconcave we have

7. **Shephard's lemma** : $c(y, \mathbf{w})$ is differentiable in $\mathbf{w}$ at $(\mathbf{w}, y)$ whenever $\mathbf{w} \gg 0$, and

$$\frac{\partial c(y, \mathbf{w})}{\partial w_i} = x_i(y, \mathbf{w}) \text{ for all } i = 1, 2, ..., n.$$

Proofs of these observations are given in Varian Chapter 5. Shephard's lemma is a straightforward application of the envelope theorem for constrained optimization.

Recovering production function from a cost function

Theorem: Let $c(y, \mathbf{w})$ be a differentiable function that satisfies the properties 1 to 6 of a cost function given in above theorem. Define a function f as follows:

$$f(\mathbf{x}) \equiv \mathbf{max}\{y \geq 0 | \mathbf{w} \cdot \mathbf{x} \geq c(y, \mathbf{w}), \text{ for all } \mathbf{w} \gg 0\}.$$

Then, f is increasing, unbounded above, and quasiconcave. Moreover, the cost function generated by f is c.
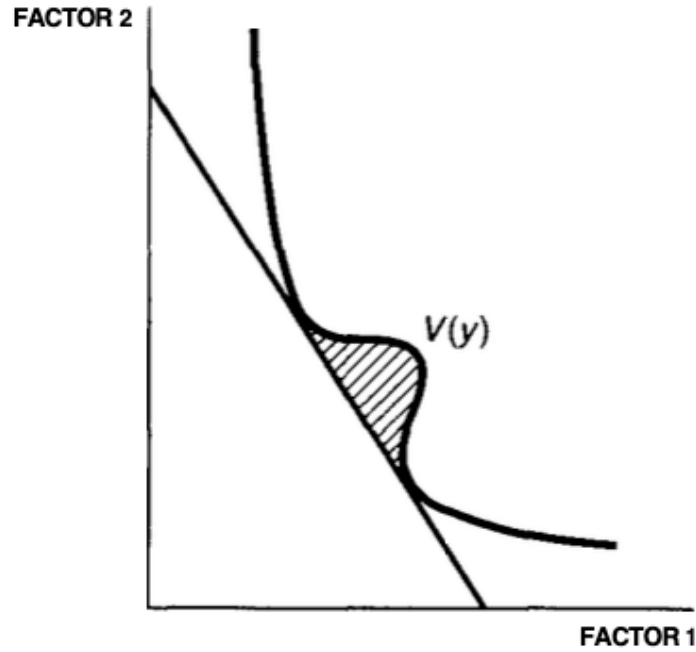

Figure 4: Recovering production function

In a similar manner, we can study the properties of the profit function and use econometric methods for estimating the profit function directly from the data.

Profit function

Recall that the profit function, defined for all input prices $\mathbf{w}$ and output price p is the maximum-value function

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

where $\mathbf{x} = (x_1, ..., x_n)$ is the profit-maximizing demand of input i.

The profit function has the following properties:

Theorem: If f is continuous, strictly increasing, and strictly quasiconcave, then for $p \geq 0$ and $\mathbf{w} \geq 0$, $\pi(p, \mathbf{w})$ is

1. Continuous,

2. Increasing in p,

3. Decreasing in $\mathbf{w}$,

4. Homogeneous of degree one in $(p, \mathbf{w})$,

5. Convex in $(p, \mathbf{w})$.

Moreover, if f is strictly quasiconcave we have

6. **Hotelling's lemma** : if $\pi(y, \mathbf{w})$ is differntiable in $(p, \mathbf{w})$ at $(p, \mathbf{w}) \gg 0$, and

$$\frac{\partial \pi(y, \mathbf{w})}{\partial p} = y(p, \mathbf{w}) \text{ and } -\frac{\partial \pi(y, \mathbf{w})}{\partial w_i} = x_i(p, \mathbf{w}) \text{ for all } i = 1, 2, ..., n.$$

Proofs of these observations are given in Varian Chapter 3. As with Shephard's lemma, Hotelling's lemma is also a straightforward application of the envelope theorem.

Empirical estimations of cost and profit function and their properties

We will estimate a Cobb-Douglas cost function using our data. Note that a Cobb-Douglas production function generates a cost function that exhibits a Cobb-Douglas form in input prices. Furthermore, if the original production technology exhibits constant returns to scale (CRS), the cost function is typically linear in y (see (Varian 1992), chapter 5).

We consider the following specification:

$$c = A\left(\prod_{i=1}^{N} w_i^{\alpha_i}\right) y^{\alpha_y},$$

and in its linearized form:

$$\ln c = \alpha_0 + \sum_{i=1}^{N} \alpha_i \ln w_i + \alpha_y \ln y$$

Estimation of cost function

```
# Estiamting cost function
dat$cost <- with( dat, vCap + vLab + vMat )
costCD <- lm( log( cost ) ~ log( pCap ) + log( pLab ) + log( pMat ) + log( qOut ), 
summary( costCD )
```

```
Call:
lm(formula = log(cost) ~ log(pCap) + log(pLab) + log(pMat) +
    log(qOut), data = dat)

Residuals:
     Min       1Q   Median       3Q      Max
-0.77663 -0.23243 -0.00031  0.24439  0.74339

Coefficients:
            Estimate Std. Error t value          Pr(>|t|)
(Intercept)  6.75383    0.40673  16.605 < 0.0000000000000002 ***
log(pCap)    0.07437    0.04878   1.525             0.12969
log(pLab)    0.46486    0.14694   3.164             0.00193 **
log(pMat)    0.48642    0.08112   5.996        0.0000000174 ***
log(qOut)    0.37341    0.03072  12.154 < 0.0000000000000002 ***
```

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

```
Residual standard error: 0.3395 on 135 degrees of freedom
Multiple R-squared:  0.6884,    Adjusted R-squared:  0.6792
F-statistic: 74.56 on 4 and 135 DF,  p-value: < 0.00000000000000022
```

The coefficients of the log-input prices are all non-negative, indicating that the cost function is non-decreasing in input prices. Additionally, the coefficient of the log-output quantity is non-negative, implying that the cost function is non-decreasing in output quantities. From the functional specification, it is clear that as long as the intercept is positive, the cost function satisfies the non-negativity condition.

Properties of cost function

We can test the homogeneity condition in two ways. First, we can perform a *linear hypothesis test* to check whether the sum of the coefficients is equal to zero. This approach is commonly used in the context of OLS regression to test hypotheses about relationships between variables.

```
# Liner homogeneity condition check
linearHypothesis( costCD, "log(pCap) + log(pLab) + log(pMat) = 1"  )
```

```
Linear hypothesis test

Hypothesis:
log(pCap)  + log(pLab)  + log(pMat) = 1

Model 1: restricted model
Model 2: log(cost) ~ log(pCap) + log(pLab) + log(pMat) + log(qOut)


  Res.Df    RSS Df Sum of Sq      F Pr(>F)
```

```
1      136 15.563
2      135 15.560  1 0.0025751 0.0223 0.8814
```

The F-statistic in a linear hypothesis test is used to evaluate whether the null hypothesis (that the sum of the coefficients equals zero) is valid. It is derived from comparing the fit of two models: the restricted model (which imposes the null hypothesis) and the unrestricted model (without those restrictions). The difference between the sum of squared residuals from the restricted model and the unrestricted model reflects how well the models fit the data when the restrictions are applied. If the null hypothesis is true, the difference should be small, and $\Pr(>F)$ will be large—this is what we find in this case.

However, note that the sum of the coefficients is not exactly one, although it holds true in a statistical sense. We can fit a model with the homogeneity condition (degree one) directly imposed. There is an advantage to doing this, which I will discuss later. Below, we estimate a model where we select one input and assume its coefficient is one minus the sum of all other input coefficients. This modification, however, alters the regression specifications:

$$\ln c = \alpha_0 + \sum_{i=1}^{N-1} \alpha_i \ln w_i + \left(1 - \sum_{i=1}^{N-1} \alpha_i\right) \ln w_N + \alpha_y \ln y$$

or, equivalently,

$$\ln \frac{c}{w_N} = \alpha_0 + \sum_{i=1}^{N-1} \alpha_i \ln \frac{w_i}{w_N} + \alpha_y \ln y$$

Below we estimate the model, considering "materials" as the *N-th* input.

```
# Estimation with linear homogeneity in input prices imposed
costCDHom <- lm( log( cost / pMat ) ~ log( pCap / pMat ) + log( pLab / pMat ) +   log
summary( costCDHom )
```

77

```
Call:
lm(formula = log(cost/pMat) ~ log(pCap/pMat) + log(pLab/pMat) +
    log(qOut), data = dat)

Residuals:
     Min      1Q   Median      3Q      Max
-0.77096 -0.23022 -0.00154  0.24470  0.74688

Coefficients:
                Estimate Std. Error t value          Pr(>|t|)
(Intercept)      6.75288    0.40522  16.665 < 0.0000000000000002 ***
log(pCap/pMat)   0.07241    0.04683   1.546               0.124
log(pLab/pMat)   0.44642    0.07949   5.616        0.000000106 ***
log(qOut)        0.37415    0.03021  12.384 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3383 on 136 degrees of freedom
Multiple R-squared:  0.5456,    Adjusted R-squared:  0.5355
F-statistic: 54.42 on 3 and 136 DF,  p-value: < 0.00000000000000022
```

As you can see, the coefficients have changed slightly, although their interpretations remain the same, as we can still rewrite them in the same Cobb-Douglas form we started with. The coefficient for "materials" can be derived by subtracting the sum of all other input coefficients from one, which gives $\alpha_{MAT} = 0.48$.

We can conduct a likelihood ratio (LR) test, which compares two nested models— typically a more complex, unrestricted model and a simpler, restricted model—to determine whether the simpler model is significantly worse at explaining the data. The test is based on a comparison of the likelihoods—how well the models fit the data—of the two models. The null hypothesis in the LR test typically posits that the restricted model is true. The LR statistic is compared to the critical value from the chi-square

distribution. If the LR statistic exceeds the critical value (indicating that Pr(>Chisq) is small), we reject the null hypothesis. In this case, however, the probability is too large, suggesting that the null hypothesis is true, meaning the restricted model fits well.

```
# Likelihood ratio test
lrtest( costCDHom, costCD )
```

```
Likelihood ratio test

Model 1: log(cost/pMat) ~ log(pCap/pMat) + log(pLab/pMat) + log(qOut)
Model 2: log(cost) ~ log(pCap) + log(pLab) + log(pMat) + log(qOut)
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   5 -44.878
2   6 -44.867  1 0.0232      0.879
```

We will therefore consider the homogeneity-imposed model for further analysis. An advantage of this approach is that we do not need to directly test the concavity condition. A standard result in real analysis states that all Cobb-Douglas cost functions that are non-decreasing and linearly homogeneous in all input prices are always concave.

```
# Predicting total cost based on homogeneity imposed model
dat$costCDHom <- exp( fitted( costCDHom ) ) * dat$pMat
# Storing coeffs from the homogeneity-imposed model
chCap <- coef( costCDHom )[ "log(pCap/pMat)" ]
chLab <- coef( costCDHom )[ "log(pLab/pMat)" ]
chMat <- 1 - chCap - chLab
```

We next test the implication of Shepherd's lemma:

$$\frac{\partial c(y, \mathbf{w})}{\partial w_i} = x_i(y, \mathbf{w})$$

Multiplying both sides by $w_i/c(y, \mathbf{w})$, we can rewrite the above equation to give us

$$\alpha_i = \frac{\partial \ln c(y, \mathbf{w})}{\partial \ln w_i} = \frac{\partial c(y, \mathbf{w})}{\partial w_i} \cdot \frac{w_i}{c(y, \mathbf{w})} = \frac{w_i \cdot x_i(y, \mathbf{w})}{c(y, \mathbf{w})} = \text{cost share of input i}$$

Observe that although the fitted model provides an estimate of α_i, the actual cost share of inputs varies across firms. Furthermore, if the cost share exceeds α_i, it implies that the firm is setting its cost share higher than optimal, which contradicts the cost-minimizing principle. Note that the relationship between α_i and the cost share can also be rewritten as

$$\alpha_i \frac{c(y, \mathbf{w})}{w_i} = x_i(y, \mathbf{w})$$

We can draw a similar conclusion from this relationship: specifically, if $(\alpha_i c)/w_i < x_i$, it implies that the firm is overusing the input compared to what the cost-minimizing principle suggests.

Below, we first present a set of scatter plots of $(\alpha_i c)/w_i$ against x_i for each input across firms, followed by a set of histograms showing the cost shares of inputs across firms. Both sets of figures indicate that most firms are overusing capital and under-using materials.
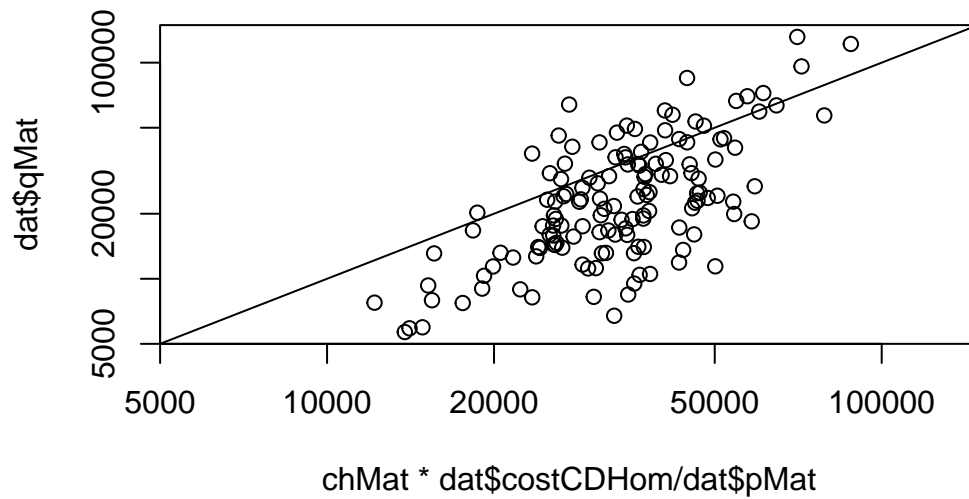
```
# Testing Shepherd's lemma
compPlot( chCap * dat$costCDHom / dat$pCap, dat$qCap, log = "xy" )
```
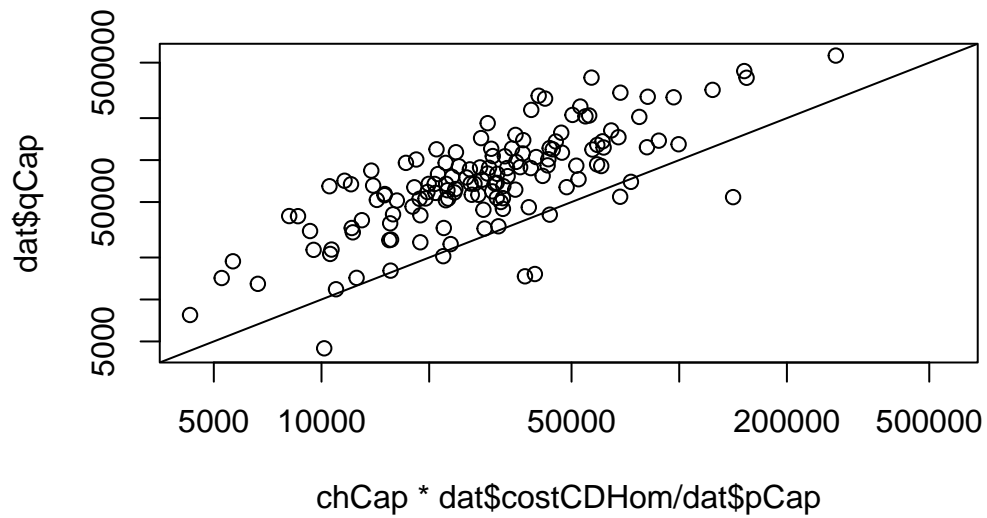
```
compPlot( chLab * dat$costCDHom / dat$pLab, dat$qLab, log = "xy" )
```
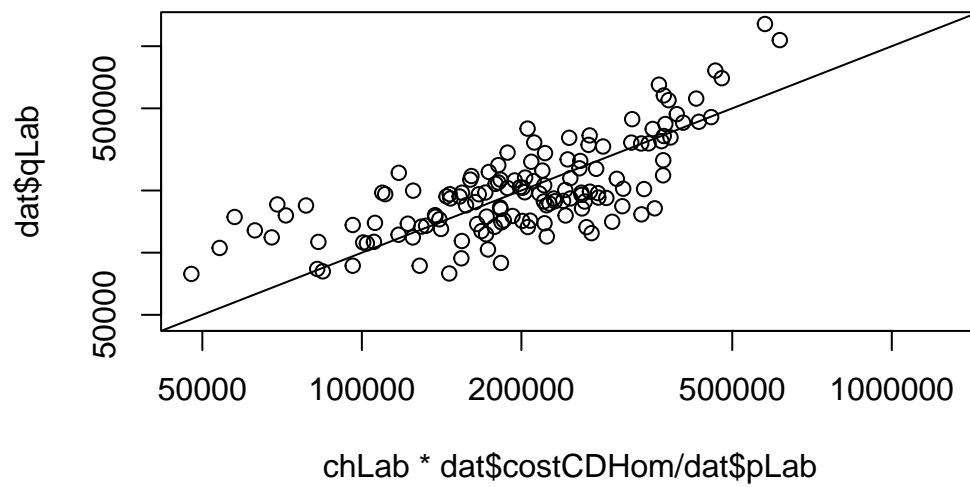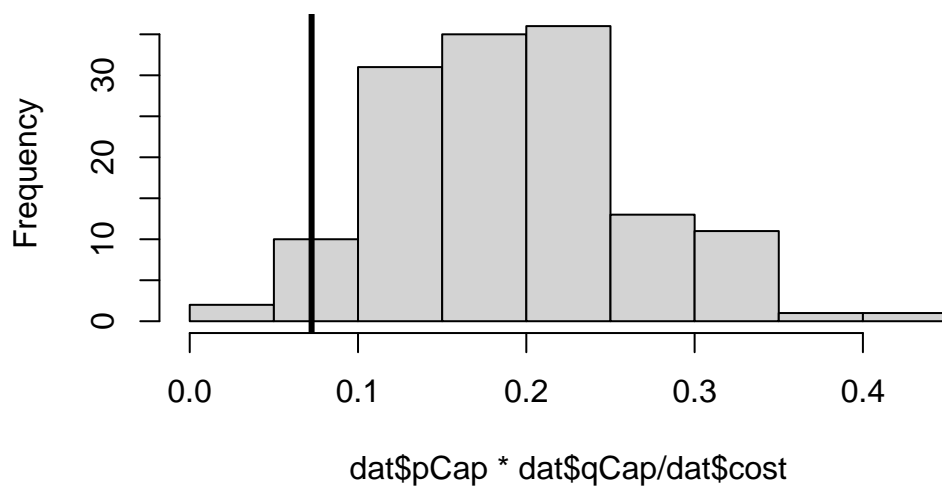


```
compPlot( chMat * dat$costCDHom / dat$pMat, dat$qMat, log = "xy" )
```

81

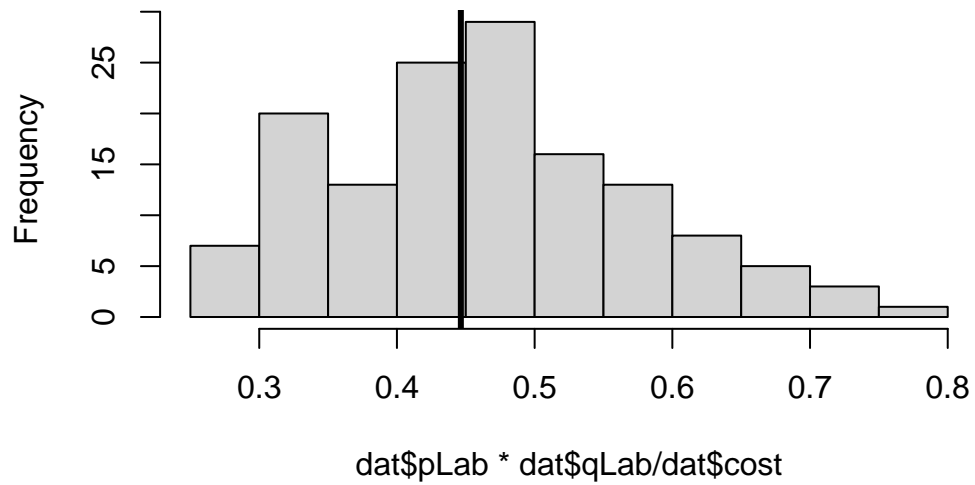Here are the histograms of cost shares for each input.

```r
# Histogram of cost shares
hist( dat$pCap * dat$qCap / dat$cost )
abline(v=chCap,lwd=3 )
```
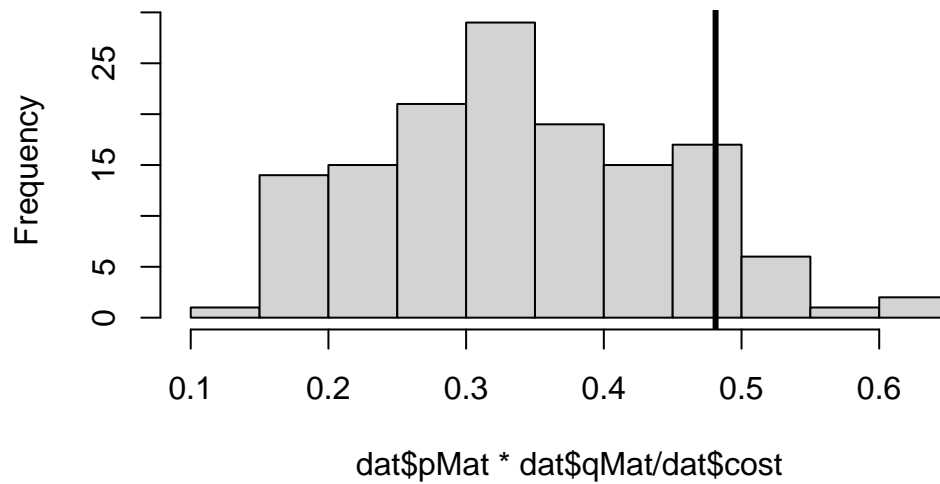
Histogram of dat$pCap * dat$qCap/dat$cost


```r
hist( dat$pLab * dat$qLab / dat$cost )
abline(v=chLab,lwd=3 )
```

Histogram of dat$pLab * dat$qLab/dat$cost

hist(dat$pMat * dat$qMat / dat$cost)
abline(v=chMat,lwd=3)

**Histogram of dat$pMat * dat$qMat/dat$cost**

Estimation of profit function

We define profit as revenue minus costs, assuming we have complete information about both. However, the constructed profit measure results in negative profit realizations for several firms.
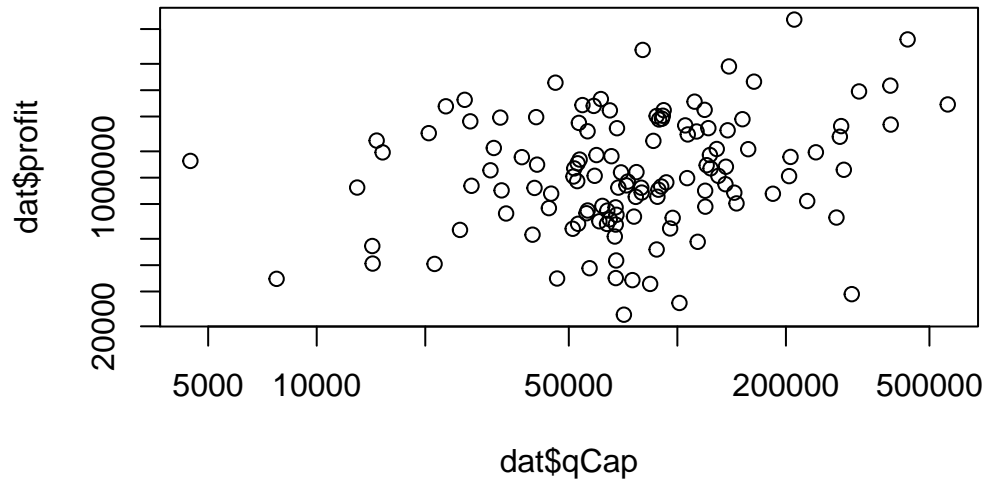
```
# Estimating profit funciton
dat$profit <- with( dat, pOut * qOut - cost )
hist( dat$profit, 30 )
```

Histogram of dat$profit
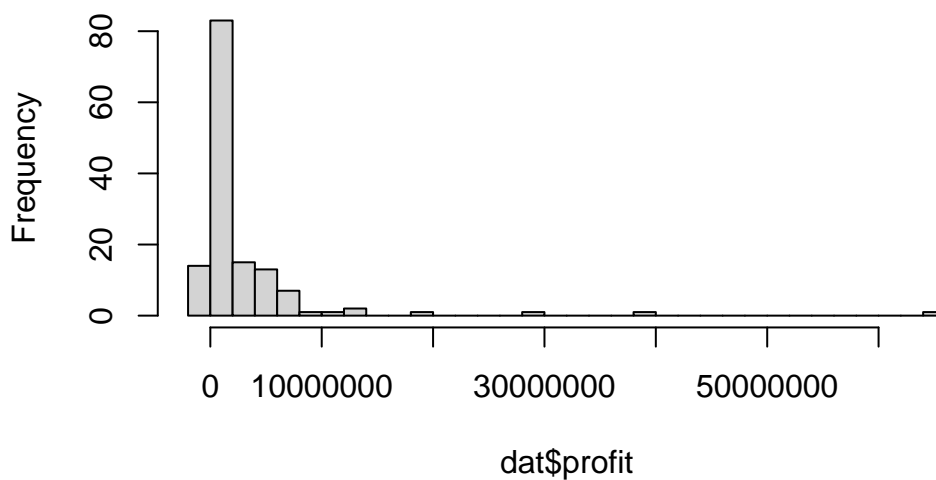

```
plot( dat$qCap, dat$profit, log = "xy" )
```

Warning in xy.coords(x, y, xlabel, ylabel, log): 14 y values <= 0 omitted from logarithmic plot

As the scatter plot shows, there are 14 observations with negative profits. Several factors could explain this (aside from the possibility of misreporting). One reason might be that these figures represent short-run profits, where firms may not be able to adjust all inputs. Below, we assume capital is a fixed input in the short run and derive short-run profit, treating capital costs as fixed. However, we still find 8 observations with negative profits. This issue is discussed in further detail in (Henningsen 2024).

```
# Considering capital as a fixed input, deriving a short-term profit function
dat$vProfit <- with( dat, pOut * qOut - vLab - vMat )
plot( dat$qCap, dat$vProfit, log = "xy" )
```

```
Warning in xy.coords(x, y, xlabel, ylabel, log): 8 y values <= 0 omitted from
logarithmic plot
```
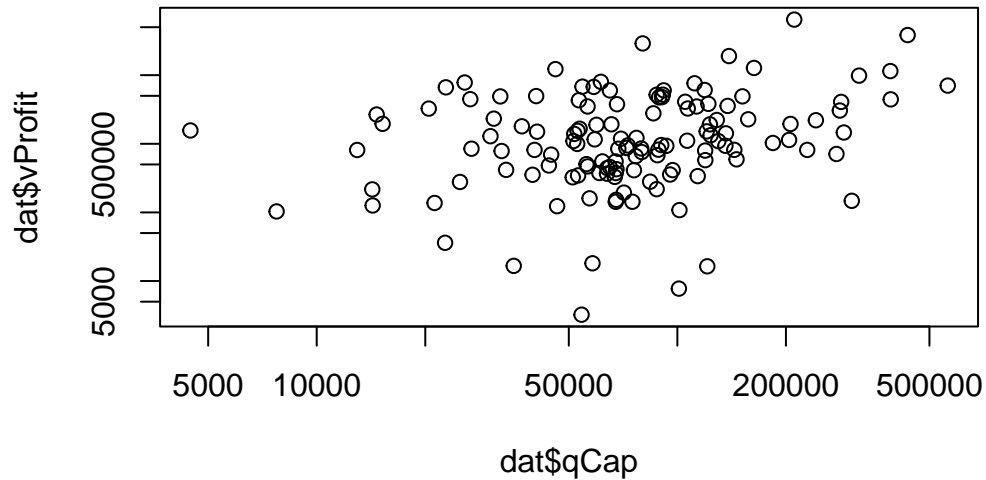
In the analysis below, we exclude the 14 observations with negative profits and proceed with the estimation.

Cobb-Douglas specification

We estimate the following Cobb-Douglas specification:

$$\pi = A p^{\alpha_p} \left(\prod_{i=1}^{N} w_i{}^{\alpha_i} \right)$$

and in its linearized form:

$$\ln \pi = \alpha_0 + \alpha_p \ln p + \sum_{i=1}^{N} \alpha_i \ln w_i$$

```
# Filter out rows with negative or zero values in relevant columns
dat_clean <- subset(dat, profit > 0 & pOut > 0 & pCap > 0 & pLab > 0 & pMat > 0)
# Fitting a Cobb-Douglas form
profitCD <- lm( log( profit ) ~ log( pOut ) + log( pCap ) + log( pLab ) + log( pMat
summary( profitCD )
```

```
Call:
lm(formula = log(profit) ~ log(pOut) + log(pCap) + log(pLab) +
    log(pMat), data = dat_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6183 -0.2778  0.1261  0.5986  2.0442

Coefficients:
            Estimate Std. Error t value          Pr(>|t|)
(Intercept)  13.9380     0.4921  28.321 < 0.0000000000000002 ***
log(pOut)     2.7117     0.2340  11.590 < 0.0000000000000002 ***
log(pCap)    -0.7298     0.1752  -4.165           0.0000586 ***
log(pLab)    -0.1940     0.4623  -0.420               0.676
log(pMat)     0.1612     0.2543   0.634               0.527
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9815 on 121 degrees of freedom
Multiple R-squared:  0.5911,    Adjusted R-squared:  0.5776
F-statistic: 43.73 on 4 and 121 DF,  p-value: < 0.00000000000000022
```

The coefficients of the log-input prices for capital and labour are negative, indicating that the profit function is decreasing in their input prices. The coefficient of log-materials is positive, which contradicts standard microeconomic principle. However, the coefficient is not statistically significant (same as the coefficient of labour). The coefficient of log-output price is positive and significant, confirms that the estimated profit function is increasing in output price. Based on the functional specification, it is obvious that if the intercept is positive, the profit function satisfies the non-negativity condition.

As with the cost function, we run a homogeneity test to check if the estimated profit is homogeneous of degree one in input prices.

```
# Linear hypothesis test
linearHypothesis( profitCD, "log(pOut) + log(pCap) + log(pLab) + log(pMat) = 1" )
```

Linear hypothesis test

Hypothesis:
log(pOut) + log(pCap) + log(pLab) + log(pMat) = 1

Model 1: restricted model
Model 2: log(profit) ~ log(pOut) + log(pCap) + log(pLab) + log(pMat)

```
  Res.Df     RSS Df Sum of Sq       F  Pr(>F)
1    122 119.78
2    121 116.57  1    3.2183 3.3407 0.07005 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F-test indicates that the null hypothesis (homogeneity of degree one) can be rejected at the 10% confidence level but not at the 5% level. This could be a critical factor, especially if we have alternative functional specifications with better properties.

We next fit a model with homogeneity imposed. This modification, however, alters the regression specifications:

$$\ln \pi = \alpha_0 + \left(1 - \sum_{i=1}^{N} \alpha_i\right) \ln p + \sum_{i=1}^{N} \alpha_i \ln w_i$$

or, equivalently,

$$\ln \frac{\pi}{p} = \alpha_0 + \sum_{i=1}^{N} \alpha_i \ln \frac{w_i}{p}$$

```
# Estimating Cobb-Douglas with linear homogeneity imposed
profitCDHom <- lm( log( profit / pOut ) ~ log( pCap / pOut ) + log( pLab / pOut ) +
summary( profitCDHom )
```

Call:
lm(formula = log(profit/pOut) ~ log(pCap/pOut) + log(pLab/pOut) +
    log(pMat/pOut), data = dat_clean)

Residuals:
    Min   1Q Median   3Q   Max
-3.6045 -0.2724 0.0972 0.6013 2.0385

Coefficients:
               Estimate Std. Error t value   Pr(>|t|)
(Intercept)   14.27961   0.45962 31.068 < 0.0000000000000002 ***
log(pCap/pOut) -0.82114   0.16953 -4.844   0.00000378 ***
log(pLab/pOut) -0.90068   0.25591 -3.519   0.000609 ***
log(pMat/pOut) -0.02469   0.23530 -0.105   0.916610

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

Residual standard error: 0.9909 on 122 degrees of freedom
Multiple R-squared: 0.3568,   Adjusted R-squared: 0.341
F-statistic: 22.56 on 3 and 122 DF, p-value: 0.00000000001091

As this model is nested within the original model, we run a likelihood ratio test.

```
# Likelihood ratio test
lrtest( profitCD, profitCDHom )
```

Likelihood ratio test

```
Model 1: log(profit) ~ log(pOut) + log(pCap) + log(pLab) + log(pMat)
Model 2: log(profit/pOut) ~ log(pCap/pOut) + log(pLab/pOut) + log(pMat/pOut)
  #Df  LogLik Df  Chisq Pr(>Chisq)
1    6 -173.88
2    5 -175.60 -1 3.4316    0.06396 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The test shows that the Chi-square statistic is high enough to reject the null hypothesis at the 10% level but not at the 5% level. This result is similar to what we observed in the homogeneity hypothesis test on the original model. However, the homogeneity-imposed model provides estimates that sum to one. We proceed with further analysis based on the homogeneity-imposed model.

Since the dependent variable in the Cobb-Douglas profit function is $\ln \pi$, we must exponentiate it to get the fitted value (and multiply by p in the case of homogeneity-imposed model).

```
# Storing coeffs from the homegeneity-imposed model
ghCap <- coef( profitCDHom )["log(pCap/pOut)"]
ghLab <- coef( profitCDHom )["log(pLab/pOut)"]
ghMat <- coef( profitCDHom )["log(pMat/pOut)"]
ghOut <- 1- ghCap - ghLab - ghMat
```

We next test the implication of Hotelling's lemma. It can be easily shown from Hotelling's lemma that the coefficients represent optimal profit shares of the output and inputs.

$$\alpha_p = \frac{\partial \ln \pi}{\partial \ln p} = \frac{\partial \pi}{\partial p} \cdot \frac{p}{\pi} = \frac{p \cdot y}{\pi} = \text{ profit share of output } (\geq 1)$$

and

$$\alpha_i = \frac{\partial \ln \pi}{\partial \ln w_i} = \frac{\partial \pi}{\partial w_i} \cdot \frac{w_i}{\pi} = -\frac{w_i \cdot x_i(p, \mathbf{w})}{\pi} = \text{ profit share of input i } (\leq 0)$$

Note that, unlike cost shares, these numbers do not lie between 0 and 1 and should be interpreted differently. For example, a profit share of 2 for an output implies that profit maximization would result in revenue that is twice the actual profit realized. A profit share of -1/2 for an input implies that profit maximization would lead to the cost of that input being as high as half of the total realized profit. However, actual profit shares vary across firms. Comparing them to $i\alpha\_i$ i allows us to understand whether a firm is overusing or underusing an input relative to what the profit-maximizing principle suggests, and similarly, whether it is producing output at the profit-maximizing level.

```
# Testing Hotelling's lemma
hist( ( dat$pOut * dat$qOut / dat$profit )[ dat$profit > 0 ], 30 ) >abline(v=ghOut,l
```
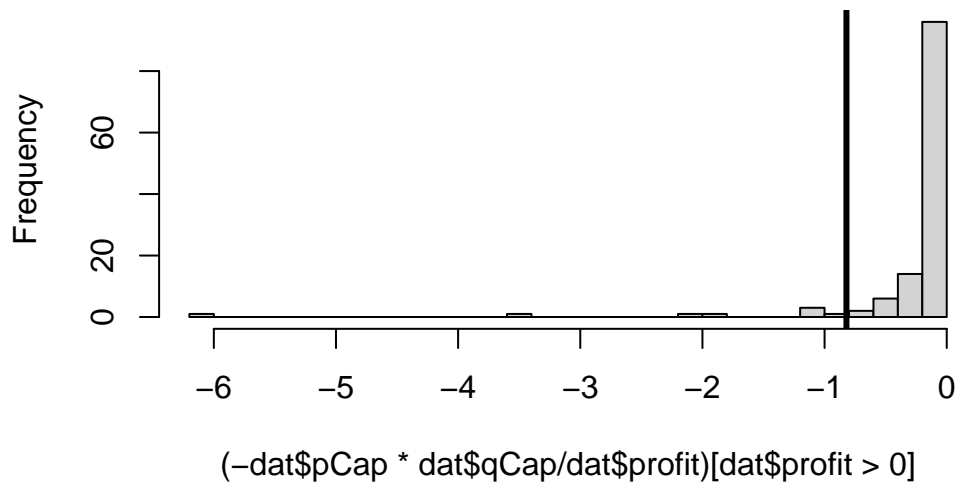
Histogram of (dat$pOut * dat$qOut/dat$profit)[dat$profit >

(dat$pOut * dat$qOut/dat$profit)[dat$profit > 0]
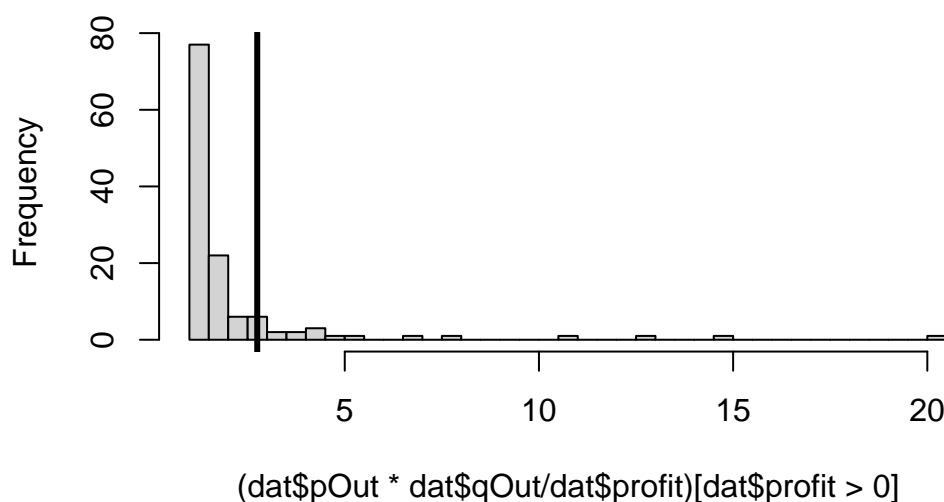
```
logical(0)
```

```
hist( ( - dat$pCap * dat$qCap / dat$profit )[ dat$profit > 0 ], 30 ) >abline(v=ghCap
```
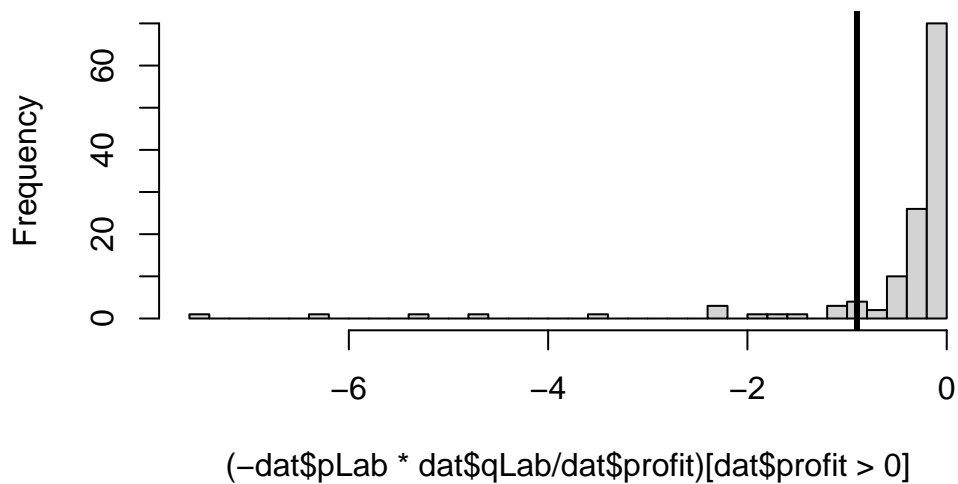
Histogram of (–dat\$pCap * dat\$qCap/dat\$profit)[dat\$profit :

(–dat\$pCap * dat\$qCap/dat\$profit)[dat\$profit > 0]

logical(0)

```r
hist( ( - dat$pLab * dat$qLab / dat$profit )[ dat$profit > 0 ], 30 ) >abline(v=ghLab
```
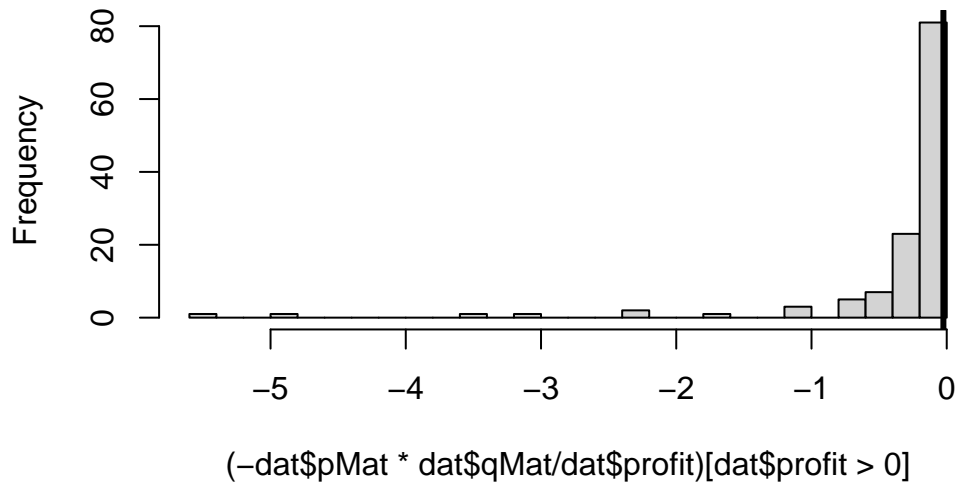
Histogram of (–dat\$pLab * dat\$qLab/dat\$profit)[dat\$profit :

(–dat\$pLab * dat\$qLab/dat\$profit)[dat\$profit > 0]

logical(0)

```
hist( ( - dat$pMat * dat$qMat / dat$profit )[ dat$profit > 0 ], 30 ) >abline(v=ghMat
```

Histogram of (–dat$pMat * dat$qMat/dat$profit)[dat$profit ⊃

(–dat$pMat * dat$qMat/dat$profit)[dat$profit > 0]

```
logical(0)
```

These histograms clearly show that most firms are operating at a level below the optimal profit-maximizing level (recall the similar findings we derived when examining the elasticity of scale based on the estimated production functions). In terms of input usage, the findings are somewhat different and interesting: it appears that most firms could use more capital (and less materials) to increase profits. This observation is inconsistent with previous analysis, but there can be reasons for it. One possibility is that the choice of capital was constrained in the short run, meaning our functional specification of a long-run profit function may not have been appropriate. Additionally, we have not tested other forms of profit functions, which might better explain the data. See [@henningsen2024introduction] for a more detailed discussion on the first point.

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.