

STUDY UNIT 1 – BENCHMARK MODEL OF RESOURCE ALLOCATION: POSITIVE AND NORMATIVE APPROACHES

INTRODUCTION

Public economics should be the concern of all South Africans because it will influence everyone's personal economic position in one way or another. Public economics focuses on government's role in a market economy. Will government's involvement in the economy promote efficiency, equity and economic growth? The benchmark model which is a very valuable tool to determine whether the actions of government will promote efficiency or inefficiency.

1.1 ASSUMPTIONS OF THE BENCHMARK MODEL (2.1)

Our benchmark model is based on a host of patently unrealistic assumptions that are set out below. Any exogenous disturbance will merely set in motion a series of more or less instantaneous adjustments that will automatically return the system to a stable equilibrium.

Basic assumptions of the two sector model:

- 1. There are two individuals, A and B, who are the suppliers of two factors of production, the producers of two commodities, and the consumers of both these commodities - all at the same time.** The two factors of production, capital (K) and labour (L). The two commodities, X and Y, both of which are consumed by the two individuals.
- 2. There are no external effects on consumption and both individuals have fixed tastes** - this is reflected in the existence of smooth and well-behaved individual indifference curves. These curves are convex with respect to the origin, cannot intersect, and exhibit diminishing marginal rates of substitution.
- 3. The two production processes are both characterized by unlimited factor substitutability, diminishing marginal productivities, and constant returns to scale.** The latter assumption rules out internal (dis)economies of scale, while there are also no external costs or benefits in production.
- 4. As consumers, A and B maximize utility and, as producers, they maximize profit.** Both are perfectly informed and are also perfectly mobile.
- 5. The commodity and factor markets are all perfectly competitive,** which implies that each market behaves as if there were a large number of individual demanders and suppliers involved, none of whom can influence price
- 6. These assumptions together ensure the existence, uniqueness, and stability of a general equilibrium**

If one wanted to explain and predict some real-world phenomenon, assumptions are important, as are the functional relations used to make predictions and test the validity of the theory. But if one's aim is merely to develop a normative theory - such as our benchmark model of resource allocation - it is hardly appropriate to judge it only in terms of the realism of its assumptions.

1.2 THE BENCHMARK MODEL AND ALLOCATIVE EFFICIENCY (2.2)

We show that in a perfectly competitive economy, allocative efficiency will occur if three conditions hold, these are:

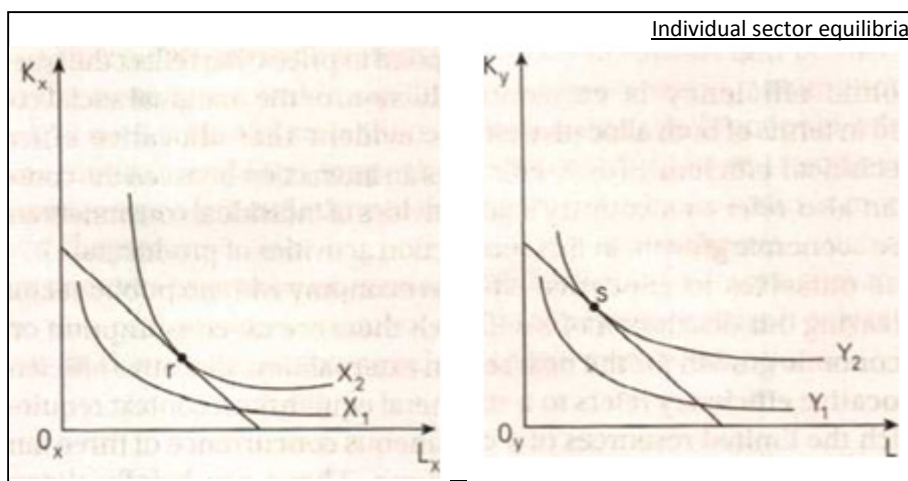
1. **Equilibrium in production**
2. **Equilibrium in consumption**
3. **Simultaneous equilibrium for producers and consumers.**

Allocative efficiency is a situation in which the limited resources of a country are allocated in accordance with the wishes of its consumers. An allocative efficient economy produces an '**optimal mix**' of commodities. Under conditions of perfect competition, the optimal output mix results from the fact that **utility-maximizing consumers** respond to prices that reflect the true **cost of production**, or the **marginal social cost**.

It is thus evident that allocative efficiency involves an interaction between the **consumption activities** of individual consumers and **production activities** of producers.

Condition 1 refers to efficiency in production and can be summarised as follows:

1. **It is a situation where it is impossible to increase the production of one commodity without thereby decreasing the production of another commodity. Pareto optimality** in production means that it should not be possible to increase the output of any one commodity without a decrease in the output of at least one other commodity. This condition requires that each of our two sectors, X and Y, should maximize output subject to its own cost constraint. In the figures, at points *r* and *s* each sector employs a combination of the two inputs, capital (X) and labour (L), for which the marginal rate of technical substitution (given by the slope of the isoquant), equals the corresponding factor price ratio, *f* (the slope of the isocost)



2. **Under perfect competition each firm will try to maximise output** (highest possible *isoquant*) **and minimise costs** (lowest possible *isocost* curve). Equilibrium only possible when firms face the same equilibrium factor prices. This occurs where the isoquants are tangent such as point *f*, *e* or *g*. **At this point labour and capital are used Pareto optimally** – production of good X cannot be increased without reducing production of good Y. When these points are linked, a **contract curve** is obtained. Equation below implies that the economy is operating at some point on its contract curve for production:

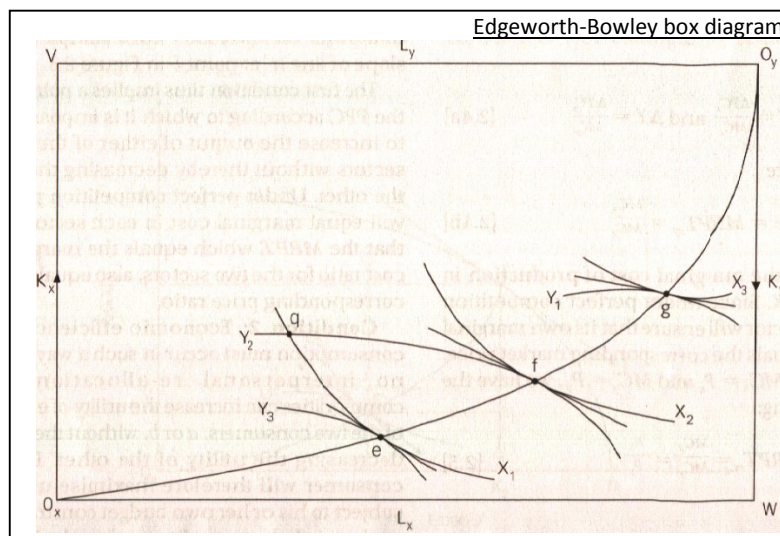
$$MRTS_{lk}^x = \frac{w}{r} = MRTS_{lk}^y$$

This is illustrated in the Edgeworth-Bowley box diagram. $MRTS_{lk}^x$ is the marginal rate of technical substitution.

Each point along the contract curve represents Pareto optimal allocation of the two resources, K and L.

At point *e*, *f* or *g* it is not possible for either sector to increase output with the other sector having to cut back its own output.

Point *q* is not on the contract curve and represents an 'X-inefficient' outcome.



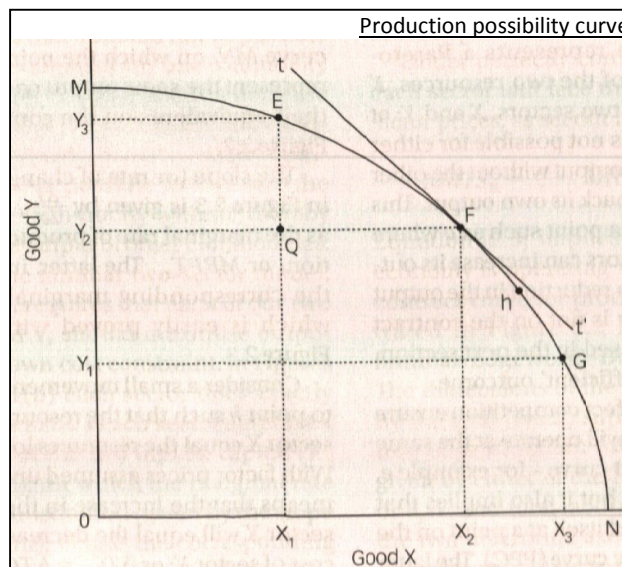
3. **The contract curve can be used to derive the production possibility curve (PPC).** The slope of the tangent to the PPC measures the **marginal rate of product transformation (MRPT)**. The slope of PPC also measures the marginal cost of producing one good (X) relative to producing the other good (Y) and can be expressed as a ratio: MC_x / MC_y .

Consider a small movement from point F to point h such that the resources gained by sector X equal the resources lost by sector Y.

4. **Under perfect competition good X is produced where the $MC_x = P_x$ and good Y is produced where, $MC_y = P_y$.** Because the MRPT is equal to the marginal cost ratio it follows that:

$$MRPT_{xy} = MC_x / MC_y = P_x / P_y$$

This means that the increase in the total cost of sector X will equal the decrease in the total cost of sector Y. Since under perfect competition each sector will ensure that its own marginal cost equals the corresponding market price we therefore have, at all points on the PPC the following condition: $MRPT_{xy} = MC_x / MC_y = P_x / P_y$



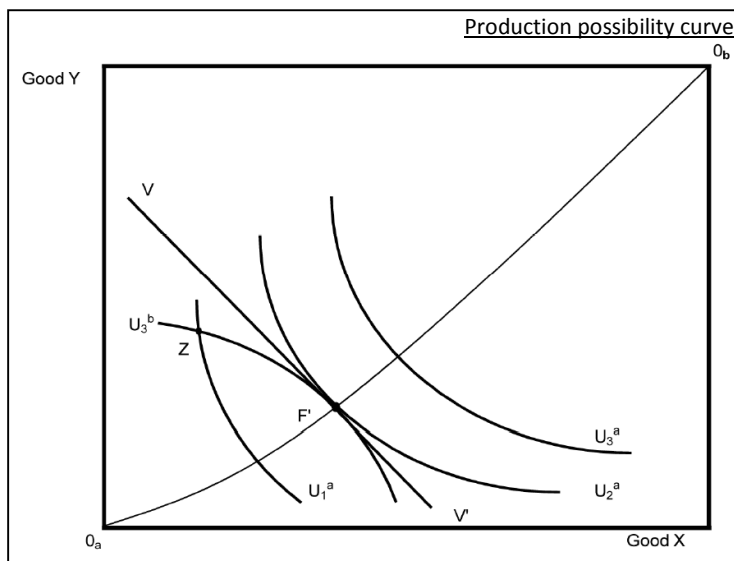
Condition 2 refers to efficiency in consumption (or exchange):

1. It is impossible to increase the utility of one consumer without thereby reducing the utility of another consumer.
2. Under perfect competition consumers will maximise utility subject to their preferences and budget constraint (reach the highest indifference curve in the figure below or budget line in the figure below).

Note: The differences between this figure and page 2. The consumption of good X and good Y by **two individuals (a and b)** are represented. Utility functions (indifference curves) and budget lines are plotted inside the box diagram.

In contrast, page 2 figure represents the production by **two suppliers** combining capital and labour. Isoquants and isocost curves are plotted inside the box diagram.

Consumers face the same relative price ratio (P_x / P_y). If the price ratio differs between consumers they can increase utility through exchange. At point Z the price ratios differ and by exchanging or bargaining consumer (a) can reach a higher indifference curve U_2^a without reducing the utility of person (b) (who remains on indifference curve U_3^b) until point F' is reached. Equilibrium occurs where the budget line (price line vv') is tangent to the indifference curves at a point such as F'. At this point consumption is Pareto efficient since one person (a) cannot be made better off without making the other person (b) worse off.



3. The slope of the budget (price) line at point F' is equal to the relative price ratio P_x/P_y and the slope of the tangent to the indifference curves at point F' is equal to the marginal rates of substitution, that is, $MRS_{axy} = MRS_{bxy}$. Because the slope of the budget line and the slope of the tangent to the indifference curves are equal it implies that

$$MRS_{axy} = P_x/P_y = MRS_{bxy}$$

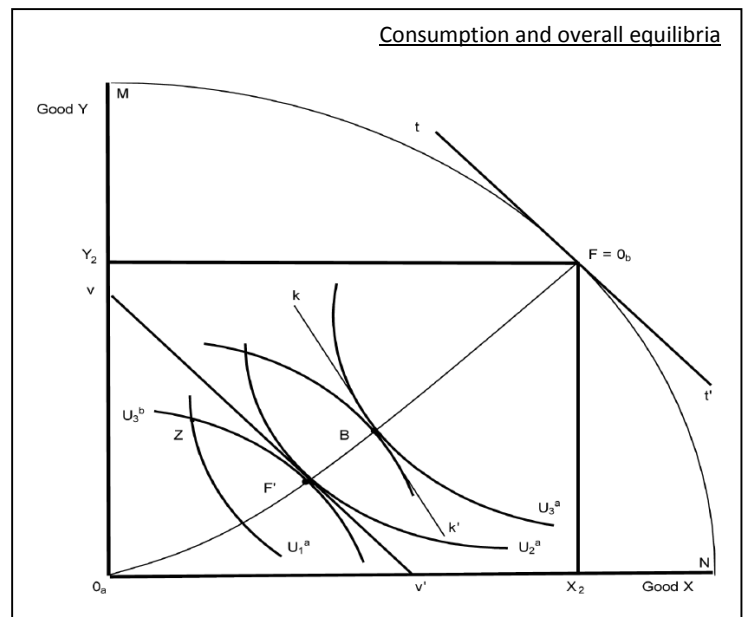
Pareto optimality in consumption implies that it is impossible to increase the utility of one consumer without decreasing the utility of the other.

Condition 3 requires that consumption and production equilibrium is achieved simultaneously to ensure efficiency in the output mix (or market efficiency). We can summarise market efficiency (or **the top-level equilibrium**) as follows:

1. In a competitive market
 - producers will maximise profits where $MRPT_{xy} = MC_x/MC_y = P_x/P_y$ (Condition 1)
 - consumers will use their budgets so that $MRS_{axy} = P_x/P_y = MRS_{bxy}$ (Condition 2)
 - the equilibrium price P_x/P_y is the same for producers and consumers (ie the relative price ratio is the common denominator in both equilibrium conditions) and therefore
 - $MRPT_{xy} = MC_x/MC_y = P_x/P_y = MRS_{axy} = MRS_{bxy}$

2. Assume that the market produces the combination F on the PPC. The indifference curves of the two consumers are drawn within the dimensions of the box – we insert the Edgeworth-Bowley box diagram above within the PPC and obtain the figure right. If the top-level equilibrium condition is to be met, the slopes of vv' and tt' must be same, i.e. parallel. If the lines are not parallel it means that the price ratios for consumption and production differ, implying Pareto inefficiency. It would then be desirable to increase the output of one product and reduce that of the other until the two ratios are the same again.

At a point such as B in the diagram below, we notice that line kk' is not parallel to tt' or the MRS_{xy} differs from the $MRPT_{xy}$. By producing more of good X and less of good Y a Pareto efficient output mix can be obtained.



Point F is a Pareto-optimal top-level equilibrium, in the sense that it is not possible to increase the output of either of the two commodities, or the utility of either of the two consumers, without thereby reducing that of the other.

1.3 X-EFFICIENCY AND ECONOMIC GROWTH (2.3)

X-inefficiency (technical inefficiency) means that firms are not maximising profit or factors of production are not maximising their welfare. A position inside the *PPC* such as *R* in the figure below is indicative of X-inefficiency.

Non-maximising behaviour is found under **conditions of monopoly** and where lack of market information and organisational slack lead to economies not reaching their full potential. Economic growth (dynamic efficiency) can be illustrated as an outward shift of the *PPC*.

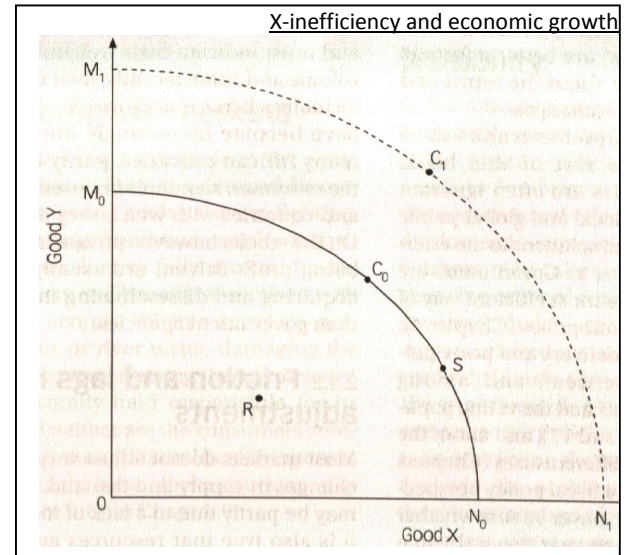
Leibenstein argued that although X-inefficiency derives from:

- A **lack of motivation by production agents**
- Factors such as a **lack of information** about market conditions
- **Incomplete knowledge** of production functions
- The **incomplete specification of labour contracts**

X-efficiency ensures that society is on its *PPC*, but cannot determine where society should be on this curve.

It is possible also to define economic efficiency dynamically (i.e.

dynamic efficiency) in terms of given increases in the quantity and/or productivity of the factors of production. The sources of growth include savings, investment (both physical and human capital), technological inventions and innovations, increases in the availability of labour.



1.4 MARKET FAILURE: AN OVERVIEW (2.4)

These failures relate to the non-applicability of some of the assumptions of perfect competition i.e, the perfectly competitive economic system will not ensure allocative efficiency. Market failure allows for government intervention and thus a role for government in a market economy, these include:

1. **Lack of information**
2. **Friction and lags in adjustments**
3. **Incomplete markets**
4. **Non-competitive markets**
5. **Macroeconomic instability**
6. **Distribution of income**

Lack of information - Producers and consumers do not always have the information to make rational decisions.

- **Producers may be unaware** of certain resources or latest technologies available in their industry.
- **Consumers may be ignorant** of potentially harmful properties inherent in some goods and services they consume, or of the fact that certain goods are available at lower prices.
- **The labour market** where unemployed workers are often unaware of the existence of available job vacancies, or employers are unaware of available job seekers who can fill their vacancies.
- **Asymmetric information** where buyers (sellers) are better informed than sellers (buyers) about the implications of their exchanges
- **Governments are also unsure** about the most efficient way of regulating natural monopolies dealing with service delivery and principal-agent problems between and among politicians, bureaucrats and the voting population and about the economic impact of different taxes
- From a **fiscal policy perspective, governments** can never be sure whether actual tax revenue one year down the line will equal the budgeted equivalent.

Friction and lags in adjustments - Most markets do not adjust rapidly to changes in supply and demand. While this may be partly due to a lack of information, it is also true that resources are not very mobile. In search markets, agents spend time and resources searching for information and incur so-called **friction costs**. Labour may take time to move from one job to another, while physical capital can only move from one location to another at very irregular intervals.

Incomplete markets - Markets are often incomplete in the sense that they cannot meet the demand for certain public goods such as street lighting, defence, or neighbourhood security on their own. Neither do they fully account for the external costs and benefits (externalities).

Non-competitive markets - Non-competitive markets are the rule rather than the exception. Commodity markets are characterised by the presence of monopolies and oligopolies, while labour markets are in turn constrained by minimum wages imposed by trade unions, governments, and by large corporations themselves. Several of the new labour laws may well raise the non-wage costs of employment, thus forcing firms either to reduce output or adopt labour-saving technologies.

Macroeconomic instability - At the macroeconomic level, markets may be slow to react to sudden exogenous shocks. Markets may take too long to adjust to changing external conditions and it is often necessary for domestic policymakers to take appropriate actions aimed at stabilising the currency. The important role played by monetary and exchange rate policies today can be viewed as an attempt on the part of governments to deal with the problem of market failure at the macroeconomic level.

Distribution of income - as reflected by the precise top-level equilibrium on the PPC - is determined to a great extent by the initial distribution of capital and labour between the two individuals. If the initial distribution is highly unequal, then so too will be the final distribution.

1.5 ENTER THE PUBLIC SECTOR: GENERAL APPROACHES (2.5)

A distinction is made between three functions:

1. **Allocative function**
2. **Distributive function**
3. **Stabilisation function**

Allocative function - Market failures distort the allocation of resources in an economy. Market failures due to incomplete and non-competitive markets are particularly important sources of allocative distortions.

There are two manifestations of incomplete markets:

1. Some goods and services have characteristics that prevent competitive markets from supplying them efficiently. In the case of **pure public goods** consumers have a strong incentive not to reveal their demand. This makes it impossible to determine a price or to force users to pay for the benefits they derive.
Mixed good consumers would either not reveal their demand or producers would find it impossible to enforce payment of the price. Mixed goods can be supplied by competitive markets, but neither the quantity supplied nor the price resulting from market provision would be optimal.
2. The **existence of externalities**. The activities of consumers and producers often impact on third parties and failure to account for externalities tend to create a divergence between actual market prices and quantities and their socially optimal equivalents. Externalities can be either negative or positive.

Non-competitive markets may take two forms. 'Artificial' monopolies operate in markets where perfect competition is technically feasible but is prevented by legal restrictions imposed by government or professional bodies. By contrast, 'natural' monopolies develop in industries characterised by large capital outlays that give rise to economies of scale over the entire range of their output. Only one firm can effectively operate in such a market.

Distributive function - A model can be used to determine the Pareto-optimal allocation of resources for a given distribution of income only and suggested that a redistribution of income could improve the general well-being of society, even if it carried a cost in terms of lower levels of productivity or slower economic growth.

Stabilisation function - The stabilisation function of government refers to its **macroeconomic objectives**, which include an acceptable rate of economic growth, full employment, price stability, and a sound and manageable balance of payments.. The notion that governments have an important stabilisation function to fulfill is associated primarily with the Keynesian school of macroeconomic thought. The Keynesian approach to stabilisation rests on three premises:

1. The **market economy is inherently unstable**
2. Macroeconomic instability is a form of **market failure that is highly costly** to an economy
3. Governments are able to stabilise the economy by means of **appropriate macroeconomic policies**.

Keynesians therefore propose active **counter-cyclical policies**. In times of recession, governments should reduce taxes, increase their expenditure, and boost credit expansion in order to raise aggregate demand and stimulate economic activity. Conversely, inflationary overheating of the economy should be addressed by higher taxes and lower levels of state spending and credit expansion, thus moderating aggregate demand.

1.6 DIRECT VERSUS INDIRECT GOVERNMENT INTERVENTION (2.6)

Direct government intervention refers to the **actual participation** of government in the economy. It includes:

- The government's right to **tax individuals and companies**
- **Borrow** on the financial markets
- Execution its **budgeted spending programmes**.

Governments intervene directly when they respond to a market failure by producing or supplying a good or service, such as national defence, waste disposal, or electricity; or by financing production undertaken by the private sector on a contract basis, such as school textbooks and much of the state's infrastructure.

Indirect government intervention refers to the **regulatory function** of government. Regulation entails enacting a law or proclaiming a legally binding rule that gives rise to market outcomes that are different from those that would have been obtained in the absence of the intervention. They include:

- The **new labour laws** that are aimed at improving the working conditions of labour
- The **new anti-tobacco law** through which it is hoped to curb tobacco smoking
- The **new competition policy** that is aimed at preventing abusive behaviour on the part of monopolies
- Several new **environmental control measures**.
- **Indirect taxes and subsidies**, which also change market outcomes, constitute indirect fiscal measures as well.

The distinction between direct and indirect interventions can make an important difference to our estimates of the size of the public sector and its effects. Conventional indicators of the size of the public sector, that are based on the total tax burden, government expenditures, and the budget deficit or surplus, provide a reasonably accurate picture of the size and extent of direct government intervention in the economy.

1.7 CONCLUDING NOTE ON GOVERNMENT FAILURE (2.7)

It is important to realise that governments can also fail. Those involved in the business of government - politicians, bureaucrats, and public employees - often pursue their own self-interest and are not X-efficient. They make mistakes and are even corrupt at times.

Government failure is a natural outcome of the way in which politicians and government officials behave. Like their counterparts in the private sector, they are utility maximisers: politicians want to maximise votes, virtually at all costs, while bureaucrats often strive to maximise the size of their departmental budgets, or 'empires! The net effect is usually an excess supply of public goods and services - or a government that is bigger than its optimal size.

STUDY UNIT 2 – PUBLIC GOODS AND EXTERNALITIES

INTRODUCTION

We determine if there are good arguments for government to supply services such as defence and to intervene when factories pollute the environment. And what kind of policies government could use to promote allocative efficiency.

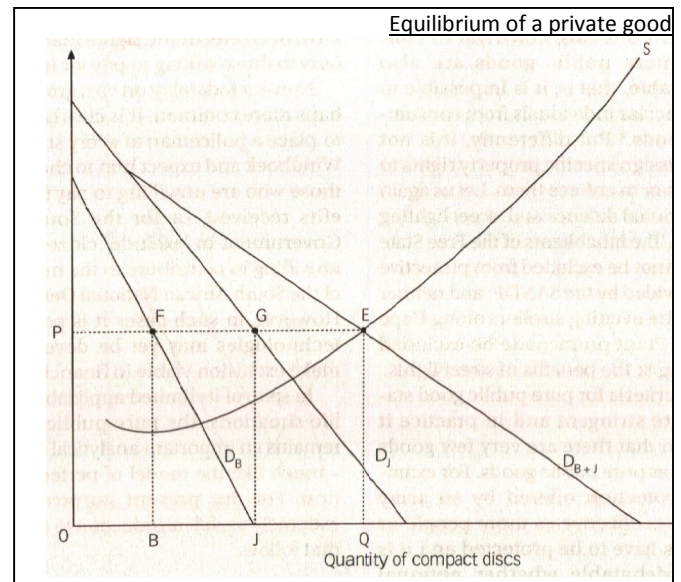
2.1 PRIVATE GOODS AND THE BENCHMARK MODEL (3.1)

Efficient production under **competitive conditions requires that consumers reveal their preferences**. Competition ensures that they do so at minimum cost. Provided that consumer preferences are fully revealed, the market that meets the third or top-level condition for allocative efficiency: **simultaneous achievement of equilibrium by producers and consumers**.

Conversely, competitive markets will fail if consumers cannot reveal their preferences. Whether or not mechanisms exist depends on the nature or characteristics of goods and services. They exist in the case of private goods, which we can define in terms of the following two characteristics:

1. **Rivalry in consumption:** private goods are wholly divisible amongst individuals; this means that one person's consumption of the good reduces its availability to other potential consumers.
2. **Excludability:** the consumption of a private good can be restricted to given individuals, typically those who pay the indicated or negotiated price.

The benefits of consuming private goods are restricted to those who reveal their preferences. The rivalry and excludability force potential consumers to reveal their preferences, setting in motion the competitive processes resulting in allocative efficiency. We can illustrate this point by referring to the market for compact discs. DB and DJ are the **individual demand curves** for two consumers, Bongani and Joan. Each demand curve depicts the quantities of compact discs that the respective consumer would demand at different prices. The market demand curve ($DB + J$) - is simply the **horizontal sum** of the individual quantities demanded at each price. Market equilibrium occurs at point E yielding a single equilibrium price at point P . Joan and Bongani price are **price-takers**. The equilibrium output of compact discs is OQ , with the quantities demanded by Joan and Bongani given by OJ and OB , respectively. Although OJ and OB sum to OQ , there is no reason why the two should be equal. The respective quantities demanded at the equilibrium price may differ according to the tastes, income levels, and other characteristics. They are **quantity-adjusters**, in that each one determines the quantity they demand in accordance with the equilibrium price.



Our compact disc example enables us to highlight *two important characteristics of a private good*:

1. **Marginal utility equals marginal cost for each consumer:** you will recall that the area underneath the demand curve gives the total utility, or the sum of the marginal utilities derived from consuming each compact disc, while the area under the supply curve gives the sum of the marginal costs of producing each compact disc. Therefore, at equilibrium price OP the marginal utilities of Bongani and Joan both equal the marginal cost QE . This is the condition for the efficient supply of a private good.
2. The **price of a private good equals its marginal cost:** this is the efficient pricing rule for private goods, as is evident from the figure.

In equilibrium (where demand intersects supply) $MU_B = MU_J = MC = P$. This is the equilibrium condition for the **optimal provision** of CD's under perfectly competitive conditions. Furthermore, in equilibrium $P = MC$. This is the **optimal pricing rule** for efficient production (maximising profit) under perfect competition. The optimal pricing rule is very important, because if it is violated, allocative inefficiency occurs.

2.2 PURE PUBLIC GOODS: DEFINITION (3.2)

Two characteristics define a pure public good:

1. **Non-Rivalry** - Pure public goods such as street lighting and national defence are indivisible and are therefore **non-rival** in consumption one person's consumption does not reduce the quantity available for consumption by another person.
2. **Non-excludability** - it is impossible to exclude particular individuals from consuming such goods and it is not possible to assign specific property rights to public goods or to enforce them.

Non-rivalry in consumption has two important implications:

1. The fact that one person's consumption does not reduce the quantity available to other consumers implies that the **marginal cost** (i.e. the cost of admitting an additional user) is zero.
2. Excluding anyone from consuming a non-rival good, even if it was feasible to do so, is *Pareto-inefficient*. The reason is straightforward: allowing Ibrahim to use the above street light at zero marginal cost will clearly make him better off than before; yet it will not detract from the enjoyment that Thandi and Roger derive from that same street light.

2.3 THE MARKET FOR PUBLIC GOODS (3.3)

The two music-lovers, Bongani and Joan, live in neighbouring houses. They spend many enjoyable evenings at home listening to their latest purchases of compact discs, often developing a strong demand for snacks in the process. A convenience store is located nearby, but the sidewalks in their neighbourhood are so poorly maintained that street lights are essential.

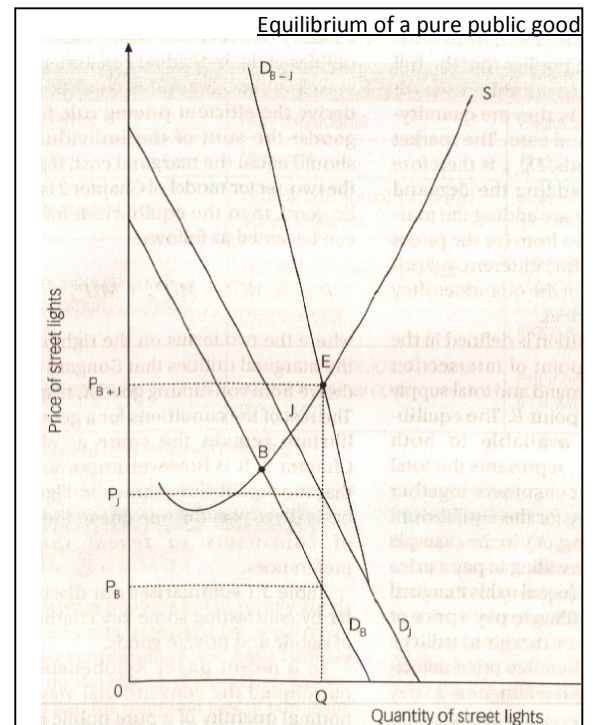
The figure below depicts the market for street lights. We assume that they are the only 'consumers' of the light. Their respective demands for street lighting are given by curves D_J and D_B . These called '**pseudo demand curves**' because they can be drawn only if consumers accurately reveal the quantities that they demand at different prices. Given this assumption, the individual demand curves and the total supply curve S are drawn.

The fundamental difference between the public and private good cases is the manner of **deriving the market demand curve**. For private goods we derived the market demand for compact discs by horizontal summation of the demand curves of Joan and Bongani.

The market demand for public goods ($DB+J$) is therefore derived by **vertically adding the demand schedules**. In effect, we are adding the *marginal utilities* they derive from (prices they are willing to pay) different quantities of street lighting, not the quantities they demand at different prices.

The equilibrium position occurs at point E. The equilibrium output OQ is available to both consumers. Price OP_{B+J} represents the total amount that the two consumers together would be willing to pay for the equilibrium quantity of street lighting, OQ . Bongani is willing to pay a price or equivalent tax of OP_B (equal to his marginal utility), while Joan is willing to pay a price or tax of OP_J (equal to her marginal utility). Bongani and Joan are therefore **price-adjusters** who can adjust their willingness to pay for street lighting.

The rules for the efficient allocation and pricing of public goods are also different from those for private goods. In the figure the areas under the demand and supply curves show the sum of **marginal utilities** and the sum of the **marginal costs**, respectively. The equilibrium position implies that the condition for the efficient provision of a public good is equality' between the sum of the marginal utilities of the individual consumers and the marginal cost. From this condition we derive the efficient **pricing rule for public goods**: the sum of the individual prices should equal the marginal cost. If good X in the two-sector model is the public good, then the equilibrium for sector X can be stated as follows: $P_{B+J}^x = MC^x = MU_B^x + MU_J^x$ where the two terms on the right represent the marginal utilities that Bongani and Joan derive from consuming good X, respectively. It is however important to add that the equilibrium shown is basically a 'pseudo' one due to the inability of consumers to reveal their true preferences.



	Public goods	Private goods
Property rights	Non-excludable	Excludable
Consumption	Non-rival	Rival
Aggregate demand curve	Vertical addition of individual demand curves	Horizontal addition of individual demand curves
Partial equilibrium condition for optimum provision	The sum of marginal utilities equals marginal cost ($\sum MU_i = MC$)	Marginal utility of each consumer equals marginal cost ($MU_i = MC$) with i the individual consumer)
Efficient pricing rule	The sum of individual prices equals marginal cost ($\sum P = MC$)	Price equals marginal cost ($P = MC$)

2.4 WHO SHOULD SUPPLY PUBLIC GOODS? (3.4)

The effects of *non-rivalry* when we stated that the marginal cost of admitting additional users of **non-rival goods is zero**. The condition for efficient pricing by competitive markets ($P = MC$) therefore requires the price to be zero as well. Clearly, profit-maximising producers cannot apply the efficient pricing rule, as charging a zero price would not enable them to cover the costs of providing the good or service

The alternative of setting a cost-covering price would potentially enable a competitive market to supply the good; it would, however, not be **efficient as exclusion cannot occur**.

Any price other than zero exceeds the zero marginal cost of admitting additional users and consequently reduces consumption of a non-rival good. Such Pareto-inefficient. In sum, it is impossible to determine an **equilibrium price for the private provision of a non-rival good**.

The *non-excludability* characteristic of public goods and services creates incentives for '**free riding**' that is, the phenomenon of misrepresenting preferences on the expectation that a benefit may be enjoyed without having to pay for it.

If Bongani reveals his preference for street lights while Joan attempts to 'free ride; a competitive market will under-supply street lighting at the level where Bongani's marginal utility equals the marginal cost of provision. In the extreme case where both Joan and Bongani attempt to 'free ride,' no street lighting would be provided at all.

Government provision of public goods and services can *improve on the inefficient outcomes* of the market; yet it cannot ensure an *optimal provision of public goods*. Compared to the market, the government has the advantage that it can use its **coercive powers** to enforce payment for public goods.

In the figure, the government wishes to apply the efficient pricing rule for public goods, $\sum P = MC$. To do so it would have to know the demand curves of the two consumers so that it can charge each consumer a price that is equivalent to his or her marginal valuation of the benefits of street lighting. In this case, the government would charge Bongani OPB and Joan OPJ thus recovering the full marginal cost (OPB+J) of providing street lighting (OQ). **Optimal provision** of a public good thus requires the application of **price discrimination**, however, the government does not have the required knowledge about people's preferences to enable it to apply perfect price discrimination. This is why governments cover the costs of supplying public goods by collecting a 'tax price' from consumers.

The mandatory nature of tax payments eliminates the 'free rider' option and gives taxpayers a direct stake in revealing their preferences for public goods. Once Bongani and Joan clearly have an incentive to participate in decisions on the use of their tax contributions.

The critical difference between public and private goods therefore lies in the **financing of these goods**. When we refer to public goods, we essentially refer to the need for public financing rather than private financing. In an extreme sense one may say that all public goods could be produced privately as long as they are financed publicly.

Governments often use private goods and services as a means of meeting public demands, with the labour of public employees being the only significant value that is added in the 'production process'!

2.5 MIXED AND MERIT GOODS (3.5)

Mixed goods possess both private and public good characteristics. Two classes of mixed goods and services can be distinguished:

1. **Non-rival, excludable mixed goods and services** - Consider part of the N14. The exclusion principle can be applied by installing a toll gate. There is no rivalry in getting access to the N14 as road users have no need to compete for scarce space. The public good characteristic of non-rival access to the road prevents competitive markets from providing such roads efficiently. The problem is the impossibility of determining a competitive price. The competitive solution of setting the price equal to marginal cost is inappropriate since the marginal cost of access to the road is zero and, similarly, charging a cost-covering price will lead to Pareto-inefficient exclusion.
2. **Rival, non-excludable mixed goods and services** - On weekdays, main thoroughfares in town are an example of the class of mixed goods characterised by rivalry in consumption and non-excludability. Rivalry in the form of competition for the scarce road space is fierce, and the marginal cost of road usage increases as congestion increases. Efficient price determination at the level of the marginal cost becomes theoretically possible. The problem lies in applying the exclusion principle. Imagine the congestion effects of levying toll charges at the entrances and exits to the central business district of a city like Luanda. In this case, market failure arises from the non-excludability characteristic of the mixed good.

Mixed goods as a group represents a '*grey area*' and the question of whether they should be supplied by the public or the private sector remains open. The influence of technology on the application of the non-excludability characteristic is particularly important in this regard.

Mixed goods can be provided either by the government alone (healthcare) or by the private sector (toll roads or subscription television services). Most mixed goods, however, are provided by a combination of the private and public sectors.

Merit goods - In the case of some mixed and even private goods it is possible to apply the exclusion principle, but the goods in question are politically regarded as so meritorious that they are often provided via the national budget. Examples of such merit goods are education and health services. The reason for treating merit goods and services in a special way is that the individual who buys or receives them often confers certain **external benefits** on other people and hence on the broader community

2.6 EXTERNALITIES (3.6)

Externalities, or external effects, can either be:

- **Positive** when the actions of an individual producer or consumer confer a benefit on another party free of charge or
- **Negative** when those actions impose a cost on the other party for which he or she is not compensated. Such actions can be either of a:
 - **Technological** when they have a direct effect on the level of production or consumption of the 'other party'.
 - **Pecuniary** when they change the demand and supply conditions, and hence the market prices, facing the other party.

In either case the beneficiary gets a windfall by not having to pay for the benefit, while the prejudiced party gets no compensation at all.

As far as **pecuniary externalities** are concerned, it can be argued that they do not have a net effect on society - resources are merely transferred from one owner to another, and markets adjust efficiently to changing demand and supply conditions. Consider an area in which crime is rampant and house prices are falling rapidly: current owners and sellers will be disadvantaged but buyers will have the benefit of lower house prices. There is therefore no net loss to society and no real external effect - only a redistribution from one group to another.

External effects drive a wedge between the private (or monetary) and the social costs and benefits associated with everyday market transactions. Social costs (benefits) are the sum of the private costs (benefits) and the external costs (benefits).

Externalities can originate on either the supply side or the demand side of the market, and it is possible to distinguish between four broad categories;

Supply side, the productive activities of a producer can have one of the following effects:

- A **negative external effect** on other producers or consumers, in which case the marginal external cost (MEC) > 0 and marginal social cost (MSC) > marginal private cost (MPC)
- A **positive external effect**, in which case MEC < 0 and MPC > MSC.

Demand side, the consumption activities of an individual consumer can have one of the following effects:

- A **positive external effect** on other consumers or producers, in which case the marginal external benefit (MEB) > 0 and marginal social benefit (MSB) > marginal private benefit (MPB)
- A **negative external effect**, in which case MEB < 0 and MPB > MSB

We shall consider only two cases: a **negative production externality** and a **positive consumption externality**.

2.6.1 Negative Production Externality (3.6.1)

Assume that a coal-fired power station pollutes the air and the water used by livestock and crop farmers. This example of a **negative production externality**.

The diagram shows the normal private (= social) demand curve and the private supply or marginal cost curve ($S_p = MPC$) for the electricity generated by the power station. These curves represent the consumers' benefits from using electricity and the supplier's cost of providing it respectively.

In a typical market situation equilibrium would occur at point E_0 with $0Q_0$ electricity supplied at a unit price of $0P_0$.

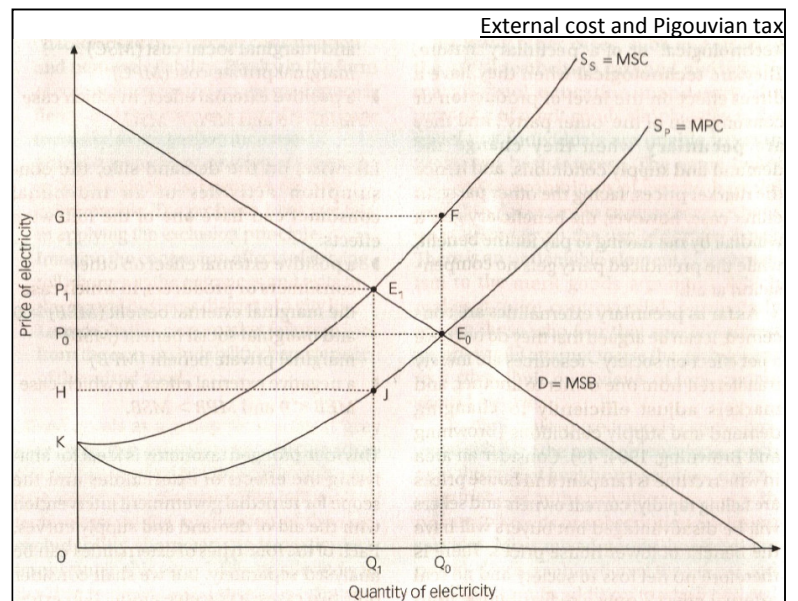
From the perspective of the community the costs incurred by the supplier do not reflect the full cost of providing the electricity. The external costs of pollution to farmers are ignored, yet are in fact part and parcel of the social cost of providing electricity. This is shown by the '**social**' supply curve labelled $S_s = MSC$. This curve indicates that the negative externality raises the social costs of providing electricity above the private costs of the supplier.

By producing $0Q_0$ units of electricity, the supplier incurs a marginal private cost equal to Q_0E_0 and a marginal external cost of Q_0F which together make up the marginal social cost of Q_0F . At the private equilibrium point E_0 total private costs equal $0Q_0E_0K$ and total external costs are, KE_0F .

If the externalities were taken into account, the 'social' equilibrium would be at point E_1 where social supply (or MSC) equals demand (assumed to equal MSB). At point E_1 only $0Q_1$ units of electricity are supplied at a unit price of $0P_1$.

Two points are worth emphasising here:

1. $0Q_1$ represents a lower quantity of output than $0Q_0$ whereas P_1 is a higher price than P_0 . Thus the presence of a negative production externality in a competitive market causes inefficiency in the form of over-provision and under-pricing of the good in question.
2. In moving from point E_0 to point E_1 the externality has not been eliminated. It has merely been reduced - from KE_0F to its optimal level KJE . The latter is an optimal level because our farming community is basically prepared to accept this negative externality from electricity generation in exchange for the value that it adds to their personal comfort and farming activities.



The opposite case - **positive production externality** - implies the presence of negative external costs where a situation where the social supply (MSC) curve lies below and to the right of the private supply (MPC) curve. A good example is the classic case of a bee farmer's bees pollinating the apple blossoms on an adjacent farm.

2.6.3 Positive Production Externality (3.6.2)

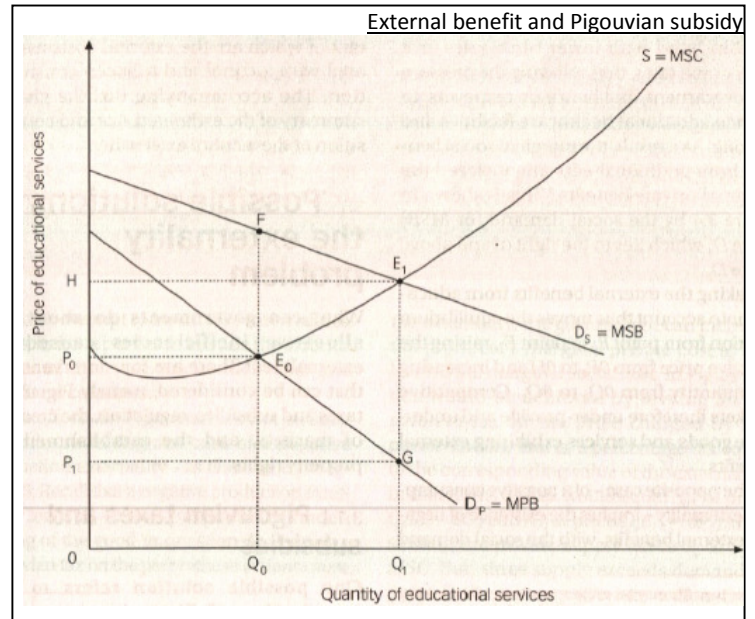
The market for education provides an example of a **positive consumption externality**. In the figure the curve S represents the (social) supply of educational services, that is, the marginal social cost of providing education. Curve D_p depicts the private demand for education, indicating marginal private benefits in the form of skills accumulation, expected higher earnings, and the sheer enjoyment to be had from being more knowledgeable.

The market equilibrium therefore occurs at point E_0 with $0Q_0$ education being supplied at a 'unit price' of $0P_0$. The externality originates on the demand or consumption side of the market. The benefits of education are not restricted to the individual, society as a whole also derives benefits.

As a result, the **marginal social benefits** from additional education exceed the **marginal private benefits**. This is shown by the **social demand (or MSB) curve** D_s , which lies to the right of and above curve D_p .

Taking the external benefits from education into account thus moves the equilibrium position from point E_0 to point E_1 raising the effective price from $0P_0$ to $0H$ and increasing the quantity from $0Q_0$ to $0Q_1$.

Competitive markets therefore under-provide and under-price goods and services exhibiting external benefits.



External effects - whether negative or positive- are an everyday occurrence and affect not only individual consumers and producers but also the fauna and flora via their potentially harmful effects. Air and water pollution or the provision of education are the typical textbook cases.

2.7 POSSIBLE SOLUTIONS TO THE EXTERNALITY PROBLEM (3.7.1)

The figures are important in the analysis of **Pigovian taxes and subsidies**.

- In the case of **negative externalities**, the marginal social cost exceeds the marginal private cost. The difference between the social and private costs is attributable to external costs. Because of external costs, there is overproduction. The socially efficient level of production can be attained by levying a tax equal to the external cost at the social optimum level. The tax generates tax revenue equal to P_1E_1JH . Note that the externality is not eliminated, but reduced to a socially acceptable level. Using the tax option sometimes leads to unintended outcomes and it is also subject to informational constraints.
- When **positive externalities** are considered, the marginal social benefits exceed marginal private benefits and the difference is caused by external benefits. At the private optimum level there is thus under consumption. To obtain the social optimum level, a subsidy which is equal to the external benefit at the social optimum should be offered. The total cost of the subsidy is equal to HE_1GP_1 in figure 2.6.2.

STUDY UNIT 3 – IMPERFECT COMPETITION

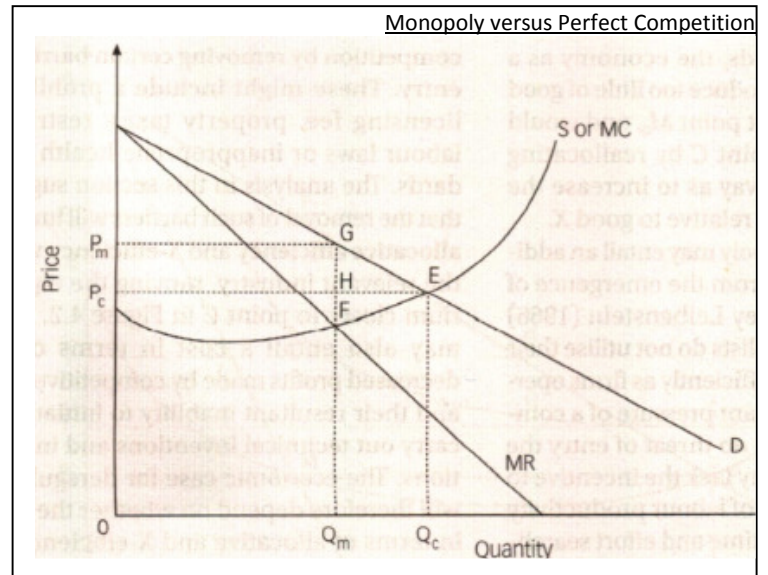
3.1 ON THE SOCIAL COST OF MONOPOLY (4.1)

Here we assume that both the demand function, D , and marginal cost, MC , are the same for the two market forms. The only difference is that under perfect competition MC represents the sum of the marginal cost curves of the individual firms making up the market, whereas under monopoly it represents the marginal cost of the monopolist only.

The perfectly competitive equilibrium occurs at point E where supply equals demand and $0Q_c$ of the good is produced at a price of $0P_c$. Under monopoly, equilibrium occurs at point F where $MC = MR$ and the market produces a smaller quantity, $0Q_m$, at a higher price, $0P_m$, than it does under perfect competition.

The loss in consumer surplus is the area given by $P_m G E P_c$, part of which $P_m G H P_c$ is a straight transfer from consumers to the producer with the remaining triangle, $G E H$, being the net welfare (or 'deadweight') loss.

The value represented by the rectangle labelled $H E Q_c Q_m$ is assumed to be transferred to other sectors in the economy, this is evidently easier said than done.



The difference highlighted above can also be shown in terms of the **two-sector model**. Recall that the marginal rate of product transformation ($MRPT$) equals the marginal cost ratio for the two commodities, that is:

$$MRPT_{xy} = \frac{MC_x}{MC_y} = \frac{P_x}{P_y}$$

If Y is now assumed to be a monopolist a X a perfectly competitive industry, then $P_y > MC_y$ while $P_x = MC_x$. It follows that:

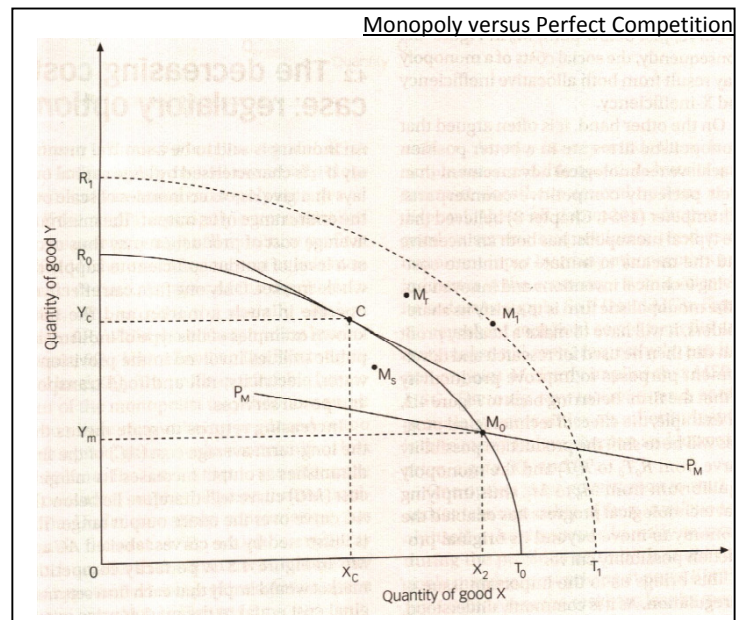
$$MRPT_{xy} > \frac{P_x}{P_y}$$

indicating that the first, and by inference, the third or 'top-level' condition for a Pareto optimum has been violated. This is illustrated by the difference between the slope of the production possibility curve, $R_0 T_0$ and the commodity price line, $P_m P_m$ passing through point M_0 . In contrast to the competitive equilibrium at point C the effect of introducing a monopoly here is to lower the output of good Y and raise its relative price.

The difference between points M_0 and C in is often taken to reflect the degree of allocative inefficiency arising from the presence of a monopoly in one of the two sectors. In other words, the economy as a whole is deemed to produce too little of good Y relative to good X at point M_0 , and would prefer to move to point C by reallocating resources in such a way as to increase the production of good Y relative to good X .

A monopoly may entail an additional cost resulting from the emergence of **X-inefficiency**. Monopolists do not utilise their existing resources as efficiently as firms operating under the constant pressure of a competitive market.

On the other hand, monopolistic firms are in a better position to achieve **technological advancement**. Monopolist has both an incentive and the means to initiate cost-saving technical inventions and innovations to satisfy its shareholders.



What action can government take to improve efficiency?

1. **Deregulation.** Often monopolies are caused by government. The economic case for deregulation will therefore depend on whether the gains in terms of allocative and X-efficiency are sufficient to offset the slower pace of technological advancement amongst competitive firms
2. **Do nothing.** In time, obstacles preventing entry to the market may disappear. In the long run, the demand curve could shift as a result of changing patterns in taste, rising incomes and the development of substitutes. In addition, patent rights lapse after a number of years, which allow other firms to enter the market.
3. **Tax policy.** By imposing taxes the government can tax away the monopolist's excess profit. Three kinds of taxes can be used, namely a unit tax, lump-sum tax and income tax. Taxation, however, does not improve the allocation of resources.
4. **Price control.** Government can reduce the price of the monopolist by pinning it down to where $P = MC$. Excess profit is reduced in this way and the socially efficient output level is achieved. However, if the price is fixed too low, it may lead to black market prices, which in turn require additional measures

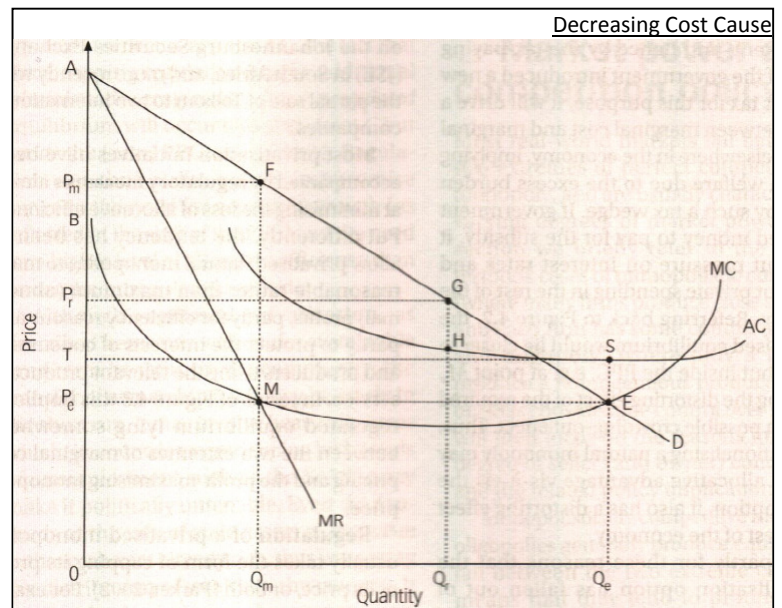
What do we learn:

- Since the market price of a monopolist exceeds the price that would prevail under perfect competition. This leads to **allocative inefficiency** which can be illustrated as the deadweight loss (GEH).
- **The "wrong" combination of good X and Y is produced.**
- It is theoretically possible to have productive efficiency under monopoly, X-inefficiency is the probable outcome.

3.2 THE DECREASING COST CASE: REGULATORY OPTIONS (4.2)

An industry is said to be a **natural monopoly** if it is characterised by large capital outlays that give rise to economies of scale over the entire range of its output. Only one firm can effectively operate in such a market. **Increasing returns to scale means** that the long-term average cost (AC) of the firm diminishes as output increases. Its marginal cost (MC) curve will therefore lie below the AC curve over the entire output range. A perfectly competitive market would imply that each firm sets marginal cost equal to the market price, e.g. at point E where $P_e = MC$. With increasing returns to scale, the industry will make a unit loss equal to ES so that individual firms will **eventually close down until a natural monopoly emerges**.

If the natural monopoly is not controlled by the government, it will maximise profit at point M where its marginal cost equals marginal revenue. At point M the equilibrium price, OP_m exceeds the socially efficient price, OP_e while the corresponding level of output, OQ_m is smaller than the Pareto-optimal level, OQ_e . The profit-maximising behaviour of the monopolist may therefore **result in too little output being produced at too high a price, giving rise to a concomitant loss of welfare**. The latter refers to the difference in consumer surplus between the two equilibria. Under monopoly, consumer surplus equals the area AFP_m which is evidently much smaller than consumer surplus under the hypothetical competitive solution, the area AEP_e .



Government does have several options at its disposal, if the good or service in question is used as an important input - such as electricity and water supply. So if D and MC represent marginal social benefits and marginal social costs, respectively, production at point M implies that $MSB > MSC$. The only way to create and confer **pecuniary externalities** on other industries is then to expand production and lower the price.

One option is for government to start up or take ownership of the natural monopoly itself, it could then apply marginal cost pricing at point E and cover the resultant loss by means of a unit subsidy equal to ES, which would have to be paid for by government, and hence by the tax-paying public:

- A higher tax for this purpose will drive a wedge between marginal cost and marginal revenue elsewhere in the economy, implying a loss in welfare due to the excess burden caused by such a tax wedge.
- If government borrowed money to pay for the subsidy, it could put pressure on interest rates and crowd out private spending in the rest of the economy. The nationalised equilibrium would lie closer to point C but inside the PPC, e.g. at point Ms, indicating the distorting effect of the required tax and a possible crowding-out effect. Thus, while nationalising a natural monopoly may have an allocative advantage vis-a-vis the private option, it also has a distorting effect on the rest of the economy.

Another advantage of privatisation is that the proceeds from the sale of state assets can be used to **redeem the public debt** or **boost investment** in the physical infrastructure of the country.

Most privatisation initiatives have been accompanied by **regulatory measures** aimed at minimising the loss of allocative efficiency. The tendency has been to allow privatised natural monopolies to **make reasonable rather than maximum (abnormal) profits**, partly for efficiency reasons and partly to **protect the interests of consumers and producers**. This implies a regulated equilibrium lying somewhere between the two extremes of marginal cost pricing and the profit-maximising monopoly price.

Regulation of a privatised monopolist usually takes the form of capping its profit or its price. The government could cap profit at point G where the maximum profit allowed would be GH per unit. Price capping regulation can also be **administratively burdensome** because regulators have to make forecasts of the future growth of demand and input costs. There is also a '**sliding scale**' regulation that combines the two capping regulations. Thus if the monopolist's profit should rise to some predetermined level, price is immediately adjusted downwards. The main advantage of this form of regulation is that both the **producer and the consumers benefit from the efficiency gains** secured by the monopolist.

These efficiency gains imply an outward shift in the PPC, where the unregulated monopoly equilibrium occurs at point M. The fact that price capping regulation entails additional administrative costs means that the regulated equilibrium will occur closer to point C but inside the (new) PPC.

Although regulated privatisation may well confer sustainable long-term benefits on the community, the transfer of ownership itself may be **costly in the short run** and may **involve heavy job losses**. Proceeds of the sale of state assets may accrue to those who are already rich, thus worsening the distribution of wealth in the country.

On the whole, it would appear that the case for regulated privatisation is a pretty powerful one. The transfer of state monopolies and other public functions to the private sector is likely to **boost efficiency and economic growth, lessen the burden of the public debt, reduce interest payments and government expenditure, broaden the tax base** and, ultimately, **enable the government to cut taxes** and initiate a **process of sustained economic growth**.

STUDY UNIT 4 – EQUITY AND SOCIAL WELFARE

4.1 INTRODUCTION

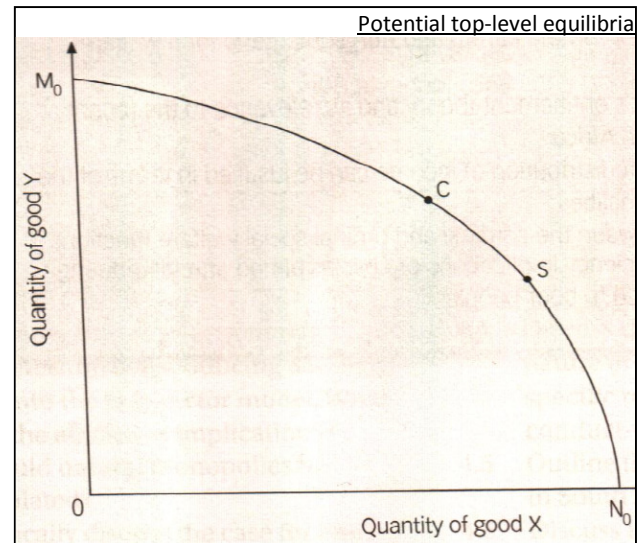
In terms of the two-sector model all points along the **PPC are Pareto-efficient**. This means that the distribution at point S is different from that at point C. In particular, one individual is in a better position relative to the other at point S than he or she is at point C.

Two important implications arising from our familiar two-sector model are relevant here:

- A competitive economy producing the output mix given by point C will not necessarily also yield the most preferred distribution of income; the latter may, for example, occur at point S;
- A policy-induced movement along the PPC, for example, from point C to point S, will necessarily change the distribution of income and thus place one individual in a worse position compared to the other

Economists normally distinguish between two criteria when assessing the welfare effects of public policy:

1. **The Pareto criterion** - implies that a policy-induced change is justified only if it improves the well-being of at least one person without harming any other
2. **The Bergson criterion** - is much broader and allows for a welfare improvement even if one or more individuals are harmed in the process. In this chapter we shall consider both criteria



4.2 NOZICK'S ENTITLEMENT THEORY (5.2)

The **Pareto criterion** is commonly associated with the libertarian approach to public policy, to which individual freedom is viewed as the primary goal. This is usually defined in terms of the **maximisation of 'negative freedom'** or protection of the right not to be coerced by others. The role of government is reduced to that of a caretaker charged with the responsibility of protecting individual freedom. Libertarians are in principle opposed to policies that infringe upon the freedom of individuals.

There is an exception to the libertarian rule that derives from Robert Nozick's entitlement theory. Nozick distinguished between three **'principles of justice'** in which he sets the conditions for just distribution. The first two principles are:

- **Principle 1: Justice In acquisition** - which states that individuals are entitled to acquire things that do not belong to others or do not place others in a worse position than before. Such 'things' refer to property and capital goods only - not to labour income, which Nozick regards as an inalienable individual right.
- **Principle 2: Justice in transfer** - according to which material things can be transferred from one individual to another on a voluntary basis, for example, in the form of gifts, grants, and bequests, or through voluntary exchange.

Violating either of the first two principles gives rise to Nozick's third principle of justice:

- **Principle 3: Rectification of injustice in holdings** - in terms of which a redistribution of wealth is potentially justified only if one or both of the first two principles have been violated. Nozick's third principle provides his only justification for a policy aimed at redistributing resources between individuals. But it is evidently easier said than done.

The Pareto flavour of Nozick's rectification principle is straightforward: if Tom enriched himself at Thandi's expense and did so against her will, the principle demands that Tom should give back to Thandi what rightfully belonged to her so that both parties would be in the same position as they would have been in the absence of the injustice

4.3 OTHER PARETO CRITERIA (5.3)

Policies aimed at redistributing income from rich to poor people can be justified on **Pareto grounds in terms of the theory of externalities**, i.e. the externality argument for redistribution. In communities characterised by a high degree of inequality it is possible that the poor may impose certain **negative externalities on the rich**.

Rich people may be prepared to **transfer part of their income to the poor** in an attempt to reduce poverty and minimise its negative external effects. However, no single rich person can do so alone and it is partly for this reason that the distribution of income is often viewed as a public good: **rich people stand to benefit from a reduction in poverty**, and hence in the level of crime and violence or in the incidence of disease.

Government policy could take the form of direct transfer payments to the poor, or it could be used to provide basic services or strengthen the security system, in which case both poor and rich people stand to benefit from a healthier and more secure environment.

A **related justification for redistribution** derives from the so-called insurance motive. Individuals may view their tax payments as a relatively inexpensive means of insuring themselves against a possible future loss of income or ill-health. On becoming unemployed, they may qualify for support from a state-run unemployment insurance fund. If they should become ill, they could likewise avail themselves of health services provided by the state. These individuals may view tax payments as a superior or cheaper alternative to taking out private insurance.

In all these cases there is no charity involved, but rather a quid pro quo principle: rich people give up part of their income for distribution among the poor because they expect to derive commensurate material benefits from such actions.

By contrast, a redistribution of income can be justified on Pareto grounds if one or more individuals are assumed to be **altruistic**, that is, **both concerned and generous**. Such individuals could experience a net increase in utility from a policy that taxes their own income and redistributes it in favour of another (non-altruistic) individual. In terms of our two-sector model, a movement along the PPC would then improve the welfare of both individuals.

4.4 BERGSON CRITERION (5.4)

Two (**Bergson**) social welfare functions are considered, these are:

1. **The additive social welfare function** which assumes cardinal or measurable utility
2. **The generalised social welfare function** where it is assumed that consumers can order or rank their preferences.

The Additive Social Welfare Function

By making a few assumptions it can be shown that an additive social welfare function will require that government redistributes income completely equally. The assumptions are as follows:

- Individuals have **identical utility functions** which depend only on their income.
- **Utility diminishes as income increases.**
- The total amount of **income is fixed.**

In the figure below income is measured on the horizontal axis. The marginal utility of Jack is measured vertically from O and is a decreasing function of income (utility decreases from left to right as his income increases).

Zini's utility is measured vertically from O' . Since Jack and Zini have identical utility functions, Zini's marginal utility curve is a mirror image of Jack's.

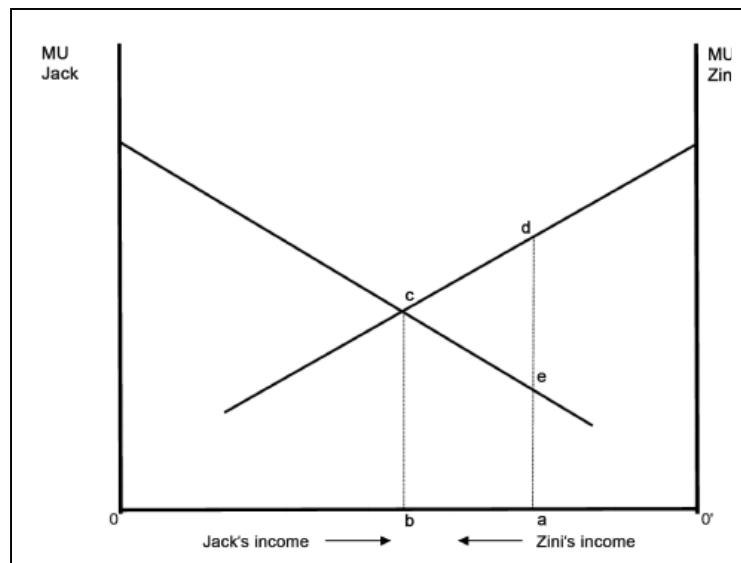
Suppose Jack's income is Oa and Zini's $O'a$ and that ab rand is taken away from Jack and given to Zini. The area under each person's marginal utility of income curve measures the total utility of each person.

If Zini receives additional income of ab , her total utility increases by the area $abcd$. Jack, on the other hand forfeits ab rand income which means that his total utility declines by the area $abce$. The sum of the utilities thus increases by cde .

The figure illustrates that if income is redistributed from (rich) Jack to (poor) Zini, **total welfare (the sum of the individual marginal utilities) increases and will be maximised where the marginal utilities are equal.**

The limitations of the assumptions are;

1. Firstly, it is impossible to determine the **validity of identical utility functions** since utility cannot be measured objectively.
2. Secondly, it is just as **difficult to prove that the marginal utility** from an extra rand income declines.
3. Thirdly, it is **possible that total income may decline as income is redistributed** since taxes and subsidies change people's incentive to work. It may also have negative implications for savings. If work effort and saving is decreased, the size of the cake is reduced.



The additive welfare function is represented by:

$$W = U_a + U_b + \dots \quad \text{Where:}$$

- W represents the level of community welfare, and
- $U_a + U_b$ are individual utilities.

The equation represents a very restrictive welfare function. Apart from the measurability issue, it assumes that individual utility functions are identical and depend only on their incomes. It is also highly debatable whether increases in income engender smaller increases in utility at higher levels of income.

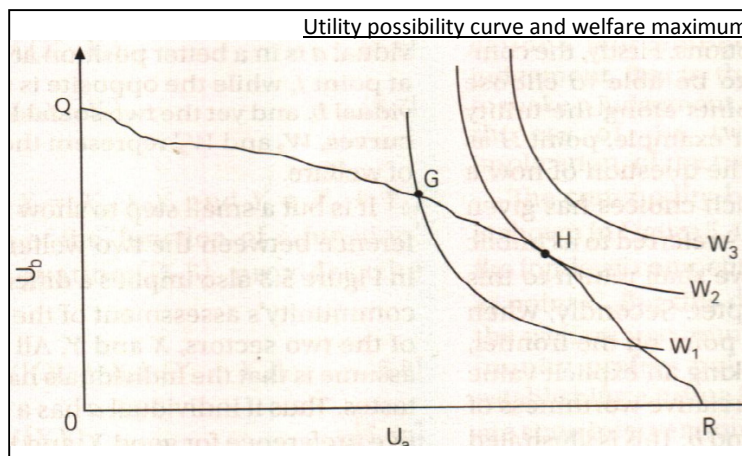
The Ordinal (Generalised) Social Welfare Function

The ordinal social welfare function, $W = W(U_a, U_b)$ does away with the assumption of measurability. By letting W take on different (constant) values, it is possible to derive a set of social or community indifference curves, such as those labelled W_1 , W_2 , and W_3 . These functions have the same properties as individual indifference curves:

- They are **convex** with respect to the origin
- **Cannot intersect**
- And **exhibit diminishing marginal rates of substitution.**

The **utility possibility frontier**, QR , gives the utility combinations associated with all the top-level equilibrium points along a conventional PPC. Here, the community prefers the combination at point H - the welfare maximum. This analysis crucially depends on two closely related assumptions:

1. Firstly, the **community is assumed to be able to choose between different points along the utility possibility frontier.** The question of how a community makes such choices has given rise to a vast literature - referred to as **public choice theory.**
2. Secondly, when choosing a particular point on the frontier, the **community is making an explicit value judgement** about the relative worthiness of the two individuals a and b .

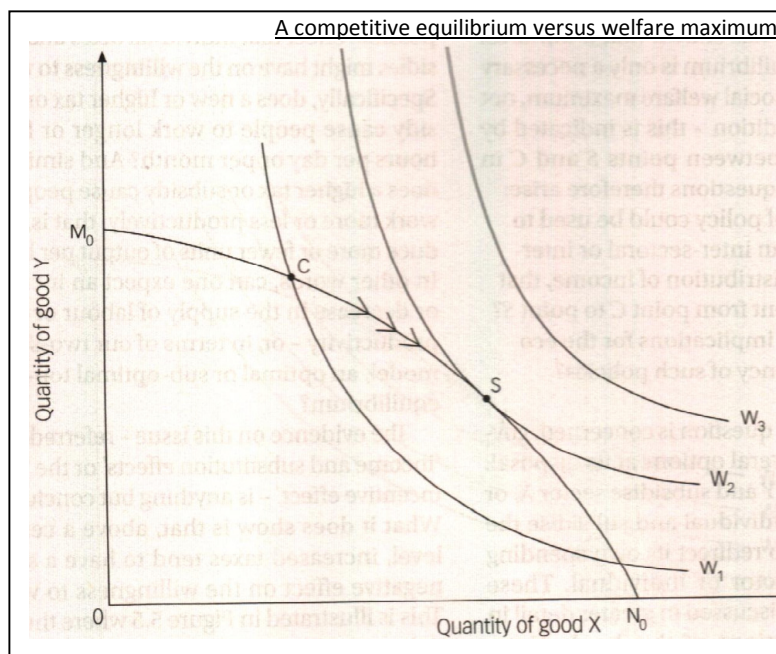


By considering two individual utility functions and the application of the 'function-of-a-function' rule, we get the following:

$$W=V(X,Y)$$

where V indicates a different functional relationship from our earlier function. The equation simply states that the welfare of society depends on the production levels of the two commodities X and Y. It is a very general version of the welfare function and will embody the above **value judgement**, that is, that the community has to make a judgement about the relative worthiness of the two sectors and, by implication, of the two individuals as well.

The commodity-based welfare function is where we assume that the top-level competitive equilibrium occurs at point C. But this does not coincide with the welfare maximum at point S: the community prefers point S to point C, thus establishing a prima facie case for appropriate state intervention to move the economy to point S.



This analysis has shown that a top-level competitive equilibrium is only a necessary condition for a social welfare maximum, not a sufficient condition - this is indicated by the difference between points S and C. Two questions therefore arise:

- What kinds of policy could be used to bring about an inter-sectoral or interpersonal redistribution of income, that is, a movement from point C to point S?
- What are the implications for the economic efficiency of such policies?

Government has several options at its disposal: it can tax sector Y and subsidise sector X, or it can tax one individual and subsidise the other; it can also redirect its own spending towards one sector or individual. Note, however, that a movement from C to S entails an improvement in social welfare even though U_b may be reduced on account of the reduced supply of good Y. This is the difference between the Bergson welfare function and the Pareto criterion discussed earlier.

A number of important observations with regard to figure above can be made:

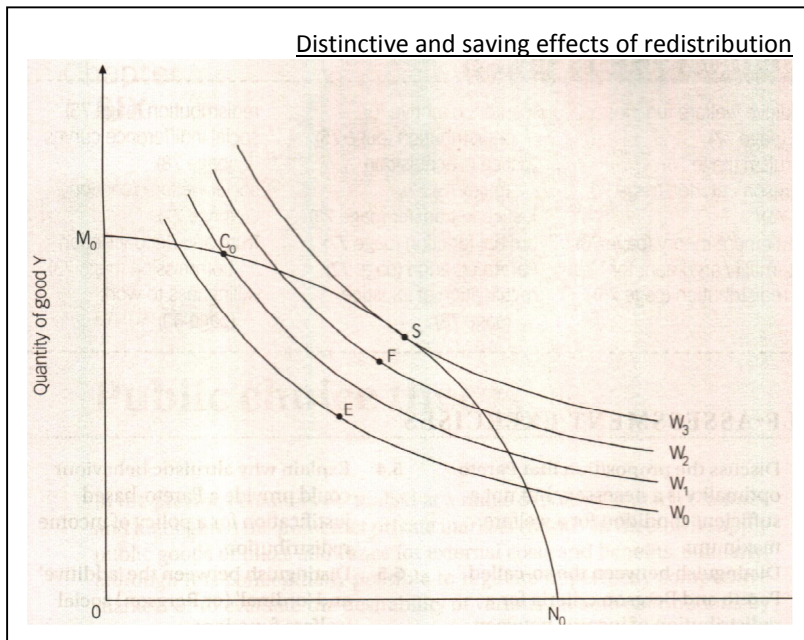
- All points on the PPC represent Pareto **optimal top-level equilibria** (allocative efficiency).
- Each point on the PPC is at a **different initial distribution of resources between person a and b**. If person b owns most of the capital and labour and has a preference for good Y, a competitive equilibrium occurs at point C.
- The **welfare maximum is at S** (the highest social indifference curve is reached here).
- **Government intervention is necessary** to move the economy closer to the social optimum equilibrium. For example, taxing the income of person b results in less of good Y being produced and a reduction in the utility of b.
- **A movement from C to S increases social welfare** even though the utility of person b is reduced (illustrates the Bergson criterion).

4.5 EFFICIENCY CONSIDERATIONS (5.5)

Redistributive policies of government may improve social welfare, but often with costs involved. In this section we can use the figure to illustrate

- The **conflict** between the equity and efficiency objectives
- The **trade-off** between equity and efficiency
- The **effect of distortions** created by redistribution
- **Dynamic consequences** of redistribution policies (or the absence thereof).

Assuming that the initial distribution of resources resulted in combination C_0 being produced. This point corresponds with the Pareto optimal top-level equilibrium, that is, **there is allocative efficiency**. Social welfare is at level W_1 . By redistributing resources it is possible to move to a higher level of welfare such as point F on W_2 . At point F resources are distributed more equally between good X and Y (and thus between person a and b if it is assumed that person a has a preference for good X and person b a preference for good Y), but this point is not an efficient allocation (it is inside the PPC) – there is a conflict between equity and efficiency. But social welfare is at a higher level (compare W_2 to W_1) – there is a trade-off between equity and efficiency.



The impact of redistribution policies on incentives to work and to save and invest is introduced and illustrated as a movement from C_0 to point F . The target was S , but because taxes and subsidies affect the willingness of people to work, an inferior social welfare level is reached (W_2 instead of W_3). Taxes also effect saving and investment decisions, which may result in sub-optimal outcomes within the PPC. In the absence of such redistribution actions, the economy may in time have experienced positive growth (dynamic consequences) resulting in the PPC shifting outwards.

Remember that large **income and wealth inequalities** may have **negative externalities**. Investors (local and foreign) base their decisions also on how stable (social and political) the investment environment is. The total absence of redistribution policies may well fuel the perception of potential instability and could impact negatively on savings and investment. Therefore, doing nothing may result in a point inside the PPC on a lower social indifference curve such as W_0

STUDY UNIT 5 – PUBLIC CHOICE THEORY

5.1 PUBLIC CHOICE THEORY: A BRIEF OVERVIEW (Study Guide)

In the political market there are **suppliers (politicians and bureaucrats)** and **demanders (voters)**. The voters demand public goods and services. In contrast with private markets where prices are used as signals, voters use voting systems to signal their wishes.

Unanimity voting is defined as *a rule where each member or representative group within a community must support a proposal before it becomes the collective decision.*

Its major advantage is that it leads to **Pareto-optimal outcomes**. For example, a Rawlsian social welfare function is postulated, the principles of unanimity voting hold. Assuming that people are risk averse and that they vote behind a “veil of ignorance”, voters will attempt to maximise the utility of the person with the lowest utility (the poorest person). The question is whether voters are really so risk averse that they are unwilling to take any chances. In other words, if they have a good chance of becoming rich, will they not accept some probability of being poor?

A number of shortcomings are also noted. Unanimity is **time consuming** when trying to win the support of all involved. It may also lead to the **tyranny of the minority** – minorities hold the majority at ransom .

Simple majority voting is the most common social choice rule. To determine the preferences of voters in the political marketplace, two types of democratic systems are used namely **direct democracy** and **indirect democracy or representative democracy**. The role of elected politicians is of particular importance in an indirect democracy because they take decisions on behalf of the electorate. Business people are in business because they want to maximise profits. Consumers want to maximise utility. Economists would therefore argue that politicians are in politics to maximise votes. But fortunately some politicians might pursue the public interest rather than trying to maximise voters.

The advantages of the simple majority voting rule are the following:

- Reaching a **majority approval is less costly** and **less timely** than for example the unanimity rule.
- **It is less likely that a minority can prevent a majority from getting their proposals accepted.**

Arrow’s impossibility theorem alludes to a potentially serious shortcoming of the majority-voting rule in that it can lead to logically inconsistent results. The shortcomings include the following:

- The **outcomes are logically inconsistent** when voter preferences are extreme.
- Outcomes **depend on the order of voting** (the voting paradox).
- There is **agenda manipulation**.
- **Intensities of preferences are ignored.**
- The **winner takes all outcomes** that may lead to the minorities being tyrannised.

Voting causes **external cost** (costs which result when voting goes against the interest of a group of voters or an individual) and **decision-making cost** (costs to persuade voters to support a particular cause). To minimise these costs, the optimal voting majority for each issue voted on would differ. This mechanism provides insight into the reasons for having a two-thirds majority voting rule for changing or amending the South African Constitution.

5.2 GOVERNMENT FAILURE: POLITICIANS, BUREAUCRATS AND RENT-SEEKING (AND CORRUPTION) (6.7)

A condition for success of both direct and indirect intervention by government is the presence of institutional framework. Important institutions include the legislative authority, law enforcement, the judiciary, tax collection or revenue services, and regulatory bodies. A sufficient condition for success refers to the **value system of the community**, including behavioural norms and customs, which should ideally entrench high levels of trust between and among consumers, producers and government institutions.

Both these conditions should ultimately ensure that government performs its functions in a transparent, accountable and consistent manner.

Politicians

Politicians engage in vote maximising strategies in order to secure and retain political office. It is important to consider the implications for resource allocation resulting from such behaviour. The likely consequences can be more readily determined given two further characteristics of the majority-voting rule:

- **Voters are rationally ignorant** of much of what politicians stand for, since they usually do not have a sufficient incentive to acquire all the information necessary to determine the desirability of all the relevant public issues.
- Politicians are **elected on the basis of a package of policies** and therefore do not have to please a majority of voters on each separate policy issue.

These characteristics can give rise to **implicit logrolling** favouring special interest legislation. The **trading of votes to ensure a favorable outcome for two or more separate decisions**. Logrolling occurs when each of two people agree to vote for the other's project to ensure that both are passed. Logrolling is commonly used when neither decision is able to obtain the necessary majority of the votes needed for passage on their own accord.

Two important consequences for resource allocation flow from this example:

- We can **anticipate a preponderance of special interest legislation** producing a variety of relatively unpopular public goods
- We can expect an **aggregate oversupply of public goods** in society.

It is clear that vote-maximising behaviour on the part of politicians can lead to outcomes inimitable to the wishes of the majority of voters. This phenomenon is a consequence of constitutional failure and can only be dealt with by constitutional reform by limiting the proportion of scarce resources expended on public goods to some fixed percentage of national income and specifying the distribution of these resources between alternative kinds of public goods.

Bureaucratic failure

Bureaucratic failure results from rational responses on the part of utility-maximising civil servants to the incentives presented to them by politicians and the institutional structures within which they operate.

Individuals in the bureaucracy, like the rest of us, do react to different incentive schemes; they do have various preferences, and have the capacity, will and desire to fulfill these preferences. They prefer more rather than less income, power, prestige, pleasant surroundings, and congenial employees.

The rational behaviour of bureaucrats can be analysed in terms of the demand for and supply of public goods. The demand for public goods in a representative democracy is generated by the decisions of vote-maximising politicians, while the supply of public goods is usually the responsibility of the state bureaucracy.

Since higher salaries, more power, greater prestige are positively related to bureau size, bureaucrats have an incentive to maximise budgets.

The bias towards the excess provision of public goods is illustrated in Figure a where the diagram shows **the total social cost (TSC) and total social benefit (TSB) curves** for a public good.

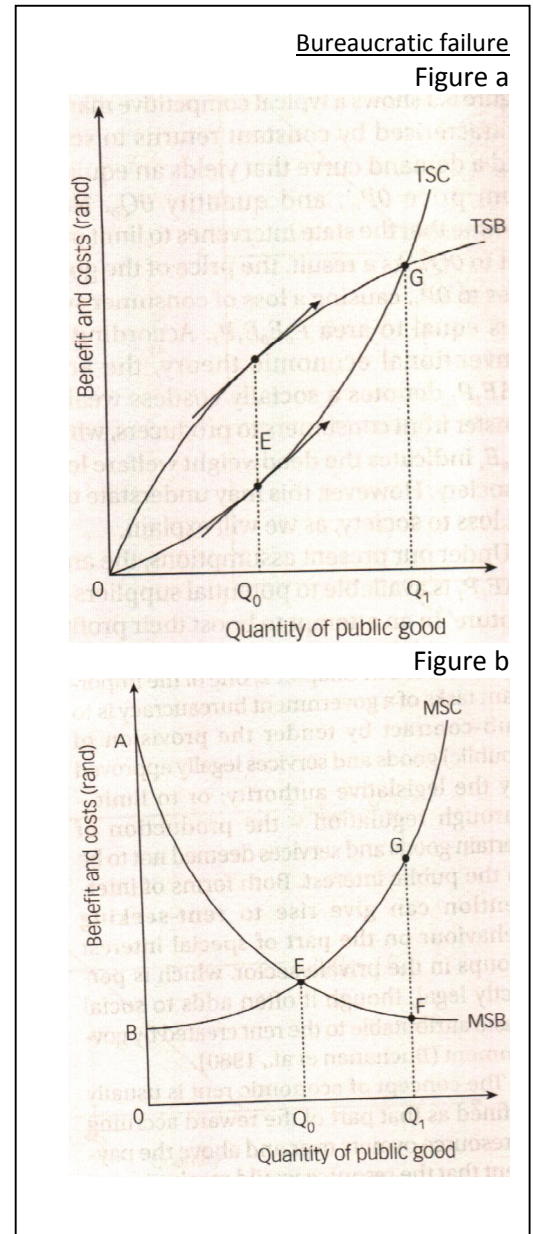
The usual marginal principles apply here: **total cost rises** at an increasing rate as output expands (due to the principle of diminishing marginal productivity), while **total benefits increase** at a decreasing rate as output expands. The rates of change of these curves, or their slopes, determine the shapes of the corresponding marginal curves shown in Figure b.

The socially optimal level of output of the public good is given by $0Q_0$, where marginal social benefit (MSB) equals marginal social cost (MSC) in Figure b, or where the difference between TSB and TSC is maximised in Figure a. But a budget-maximising bureaucrat would attempt to justify output $0Q_1 > 0Q_0$, where TSB equals TSC and where MSC exceeds MSB by the distance FG. The result is that the total value to consumers increases from $0AEQ_0$ to $0AEQ_1$ or by the area Q_0EGQ_1 , while total cost increases from $0BEQ_0$ to $0BGQ_1$ or by the larger area Q_1EGQ_1 ,

The increase in **total cost thus exceeds the increase in total benefits** by the area EGF- the net welfare loss to society.

Given the absence of a profit motive, bureaucrats may **supply public goods inefficiently**. The resultant inefficiency thus stems from an excessive use of inputs rather than from excessive production by the bureau.

In essence, the problem of bureaucratic failure is simply an example of a **principal - agent problem**. In the public sector bureaucrats act as agents for their principals, the taxpaying public, who are in turn represented by elected politicians. Since each individual bureaucrat benefits directly from a large budget, he or she has a strong preference for a high level of expenditure. Individual taxpayers, on the other hand, generally benefit only marginally from that expenditure and therefore remain rationally ignorant about it. Bureaucrats have a greater incentive to increase spending than taxpayers have to reduce taxes. It follows that one should expect spending levels on public goods to exceed their corresponding optimal levels.



Rent Seeking

One of the important tasks of a government bureaucracy is to sub-contract by tender the provision of (public) goods and services legally; or to limit - through regulation - the production of certain goods and services. Both forms of intervention can give rise to rent-seeking behaviour.

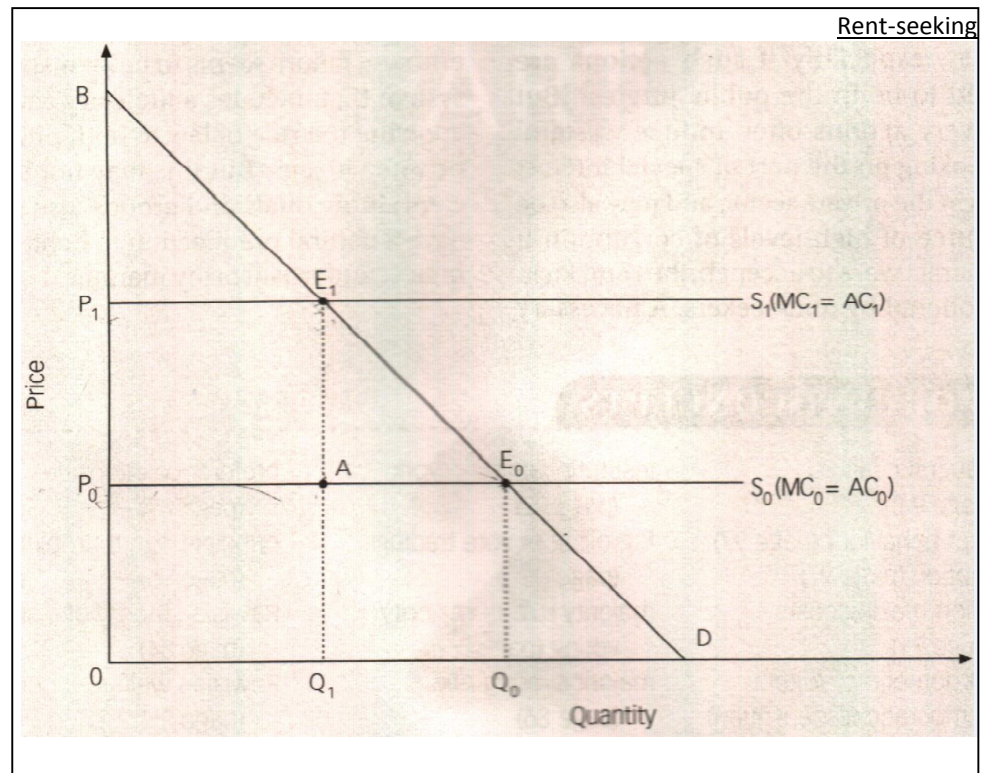
The concept of economic rent is usually defined as **that part of the reward accruing to resource owners over and above the payment that the resource would receive in any alternative employment**. In a perfectly competitive world market forces would ensure the dissipation of rent. The existence of **positive rent in a competitive market** will attract resources in the same way as the existence of potential profits, and consequently result in the erosion of such rent through an efficient re-allocation of resources.

The theory of rent-seeking deals with the origins of, and competition for, artificially created rent. The latter usually results from government protected monopoly power.

The figure shows a typical competitive market characterised by constant returns to scale and a demand curve that yields an equilibrium price OP_0 , and quantity OQ_0 . Assume that the state intervenes to limit output to OQ_1 . As result, the price of the good rises to OP_1 causing a loss of consumer surplus equal to area $P_0E_0E_1P_1$. According to conventional economic theory, the area $P_0AE_1P_1$ denotes a socially costless wealth transfer from consumers to producers, while AE_0E_1 , indicates the deadweight welfare loss to society.

Under our present assumptions, the area $P_0AE_1P_1$ is available to potential suppliers to 'capture' in an attempt to boost their profits. They would thus be prepared to incur additional costs - by lobbying government. Two possibilities exist here.

- **If participating suppliers** - taxi drivers- undertook the lobbying function themselves and the additional costs were internal to them, then marginal cost will increase to the critical level MC. At this point the taxi drivers will have no further incentive to engage in rent-seeking activity since all rent consequent upon state intervention will have been dissipated; and with lobbying being largely a one-off expense, the suppliers could share the future profits among themselves (when marginal cost would have fallen back to MC_0). Taxi drivers will have **transferred part of their own resources away from productive activities in favour of non-productive rent-seeking activities**.
- **If concerned citizens** were to lobby government to reduce the number of taxis on the roads, then the additional costs will be external to the suppliers or taxi drivers, and the conventional theory will hold: the area $P_0AE_1P_1$ will indeed represent a wealth transfer from one group (the consumers) to another (the suppliers). The latter group will have received additional income that it can save or spend in the economy - much like a cartel.



STUDY UNIT 6 – PUBLIC EXPENDITURE AND GROWTH

6.2 REASONS FOR GROWTH OF GOVERNMENT: MACRO MODELS (7.4)

Three **macroeconomic** explanations are discussed:

1. **Wagner, Musgrave and Rostow's stages-of-development approach**
2. **Peacock and Wiseman's displacement effect**
3. **The Meltzer-Richard hypothesis**

1. Wagner, Musgrave and Rostow's Stages-of-Development Approach

This theory explains how **government expenditure tends to increase** when a country develops from a subsistence or traditional economy to an industrialised economy. During the first or early stages of development Wagner emphasised the need for government to create and **maintain internal and external law** and order and set the **legal and administrative institutions** to cut the costs of doing business. According to the stages-of-development approach, it is also important for government to get investment going;

- During the **first stage**, the formal sector of the economy is still relatively small and, as a result, government may have to participate **actively by providing the basic infrastructure** necessary to create an environment conducive to economic development. The implication of the first stage for government expenditure is that **capital expenditures** will feature prominently.
- The **middle stages** of development, government will continue to supply investment goods while **private investment** will also start to take off, partly due to the **positive pecuniary external effects** of government investment undertaken during the first stage. However, the development of the private sector may cause **certain market failures**, including externalities, monopoly pricing and high-density living conditions, that government would have to address, thereby giving rise to further **increases in government expenditure**.
- In the **last stage** of development capital expenditure by government, expressed as a percentage of GDP, usually decreases because most of the necessary infrastructure is already in place. At this stage, however, **expenditure on education, health, welfare programmes, and social security will tend to increase** due to the high-income elasticity of demand for such expenditures. The result is a continuous increase in the share of government in the economy.

2. Peacock and Wiseman's Displacement Effect

Peacock and Wiseman (1967) used a political theory to **explain the influence of political events on public expenditure**. They did acknowledge a point made by Wagner, that '...government expenditure depends broadly on revenues raised by taxation'. Governments would continue increasing expenditures and expanding their role in the economy - provided their economies continued to grow through industrialisation.

- On the other hand, individuals may not be prepared to continue paying higher taxes in order to finance such increased expenditure.
- Under normal circumstances government expenditure would therefore only increase when it is strictly necessary, and it can be expected that governments would take into account the possible resistance of voters against higher tax rates.

Social upheavals or disturbances like a world war or national disaster may change the established conceptions of the public. Peacock and Wiseman called this the **displacement effect**, as certain government expenditures displace private expenditures. After a crisis had subsided, government expenditure could be expected to return to its pre-crisis level. According to Peacock and Wiseman, this was unlikely and government expenditure could even remain at the **new post-crisis level**, the reason being that taxpayers would become **accustomed to the higher levels of taxation**.

The displacement effect may help to explain the growth in government expenditure in SA. The **previous political dispensation** resulted in a **massive military build-up and excursions** into neighbouring countries. This triggered rapid increases in expenditures on protective services and social services. Expenditure for education started to increase after the unrest in 1976. It is therefore, possible to argue that both **social upheaval and the conflict situation** contributed to the growth in government expenditure in South Africa during the period up to 1994. One may also argue that displacement pressures in SA were **dampened by a redirecting of government expenditure** from economic services to social services.

3. The Meltzer-Richard Hypothesis

Redistributive policies have an important impact on the growth of government expenditure. Meltzer and Richard developed a general equilibrium model in which **majority voting determines the magnitude of income distribution** and also the share of government expenditure in the economy. The most important reason for the increase in government expenditure can be attributed to an **extension of the franchise**, which brings about a **change in the median voter**.

If all voters were ordered according to their income with the lowest income on the far left hand side and the individual with the highest income on the far right, the median voter would be the one right in the middle. The median voter is important because in a two-party democracy they determine which party will win the election.

According to Meltzer and Richard, there will be **pressure for redistributing income** if the income of the median voter lies below the average income. Redistribution would benefit the median voter and they will therefore vote for the party that proposes a programme of redistribution. The median voter effectively determines the tax level.

Meltzer and Richard's model does not allow for **unlimited redistribution**, because they assume that voters are aware of the **disincentive effects** associated with high taxes and redistribution. The rational median voter will thus choose the tax rate that **maximises his or her utility**. If the median voter chooses a relatively high tax rate other individuals may decide to work less and consume more leisure with the result that the economic 'pie' may become smaller for redistribution.

This hypothesis would suggest that the **extension of suffrage** should have resulted in a major increase in the share of government expenditure in the South African economy. But this did not happen, despite the fact that the election of 1994 resulted in the appearance of a new median voter with an income well below the average income. It may be argued that the median voter model does not apply to South Africa, or that a substantial degree of redistribution had already occurred earlier in an attempt to counter social upheaval. Expenditure on social services (as a percentage of both total government expenditure and GDP) increased between 1983 and 2004, while defence expenditure declined dramatically both prior to and after the 1994 elections. The overall fiscal restraint imposed by the need for macroeconomic stability, forced the government to meet many of its distributional goals by means of a reallocation of social spending

6.3 MICRO MODELS OF EXPENDITURE GROWTH (7.5)

Three microeconomic explanations are discussed:

1. **Baumol's unbalanced productivity growth model**
2. **Brown and Jackson's microeconomic model**
3. **The role of politicians, bureaucrats and interest groups**

1. Baumol's Unbalanced Productivity Growth Model

Government expenditure may also **increase disproportionately** due to an increase in the prices of inputs used by the public sector relative to those employed in the private sector. William Baumol developed a microeconomic model of unbalanced productivity growth to explain the **growth in government expenditure**. He divides the economy into two broad sectors, a **progressive sector** and a **non-progressive sector**. The progressive sector is characterised by technologically progressive activities, such as innovation, capital formation, and economies of scale.

An important feature of this sector is a **cumulative increase in the productivity** of employees that justifies increases in wages and salaries. The inherent characteristics of the non-progressive sector only permit sporadic changes in productivity.

The **technological structure** of a sector will determine the increase in the productivity of labour inputs used. In the progressive sector **labour is only one of the inputs** in the production process, while in the non-progressive sector **labour is often the end product**. Consumers are usually not interested in the labour used to produce an air conditioner. However, the labour input is of great concern when one has purchased a ticket to attend a one-hour concert by a Beethoven quartet. Any effort to increase the overall productivity of the concert to below four man hours may upset listeners and detract from the end product. In this case there is clearly a limit to productivity increases, which is greatly determined by the labour intensive nature of the service.

The non-progressive sector usually consists large component of the public sector where labour plays an important role. Technological changes do not have such an important effect on productivity in the non-progressive sector as they do in the progressive sector. As a result there are only sporadic improvements in productivity in the non-progressive sector compared to relatively rapid increases in the progressive sector.

Baumol argues further that there cannot be too **big a difference in the wages and salaries between the two sectors**, otherwise employees would be leaving the non-progressive sector to join the progressive sector. This raises the relative costs of the non-progressive sector because salary increases are not accompanied by the same increases in productivity as in the progressive sector.

Baumol thus came to the conclusion that '**the costs of even a constant level of activity on the part of government can be expected to grow constantly higher**'. If production in the non-progressive sector has to be maintained relative to that in the progressive sector, it will imply that a larger share of the labour market will have to be employed in the former sector, which could have negative effects on economic growth. One point of critique, however, is that Baumol may have underestimated the opportunities for technological advancement in the public sector.

The largest share of government expenditure in South Africa goes towards the remuneration of employees. The structure of government expenditure thus corresponds well with Baumol's notion of unbalanced productivity growth. However, the Baumol hypothesis has not been tested empirically in South Africa or in other African countries; no firm conclusions can therefore be drawn about its relevance to government expenditure growth in the African continent.

2. Brown and Jackson's Microeconomic Model

Brown and Jackson developed a microeconomic model to **derive the levels of publicly provided goods and services** by, inter alia, taking the preferences, the income, and the tax rate of the median voter - **all determinants of demand** - into account, as well as the costs of the goods and services in question.

It happens that **government expenditure increases without there being a corresponding change** in the level of quality or service. This can be viewed as a sign of inefficiency, Brown and Jackson argue that it may in fact be the result of changes in the service environment such increased policing and additional funding as a **result of increased crime**.

Changes in the **size and density of the population** and its age structure may also influence the service environment. Population growth may lead to an increase in the demand for **publicly provided goods and services**. In such instances, increases in the population will lead either to **higher levels of expenditure** or **to a drop in standards**.

Population growth and human migration may also lead to changes in the density of the population, which may cause **congestion and add to the costs of government**. The fight against the HIV/Aids pandemic, malaria, and tuberculosis should also put pressure on government expenditure.

Another factor that may influence the level of **government expenditure is the quality of goods** demanded by the median voter. According to Brown and Jackson, a good or service is of superior quality if it requires more inputs in its production process than a good requiring fewer of those inputs, ceteris paribus.

Their model provides a useful starting point for anyone interested in modelling government expenditure in African countries and identifying the explanatory variables that will determine future trends in government expenditure

3. Role of Politicians, Bureaucrats, and Interest Groups

Politicians, bureaucrats, and other interest groups are often powerful enough to pressure government in a direction that is detrimental to the social welfare of the community. This may result in a higher than optimal level of government expenditure, thereby contributing to the growth of government's share in the economy at the expense of the private sector.

There are many other ways in which the vote-maximising behaviour of politicians can bring about an increase in government expenditure. They may grant wage and **salary increases to state employees** in order to gain the support. From a macroeconomic point of view, the ruling party may be tempted to **relax its fiscal and monetary policies** in an attempt to stimulate the economy before an election to please voters. But **populist macroeconomic policies are not sustainable in the long term**, giving rise to inflationary pressures and balance of payments problems, and ultimately calling for even tighter monetary and fiscal policies than had been the case prior to the election.

Bureaucrats tend to maximise the size of their departmental budgets, partly to help build their **own personal 'empires'**. The net result is a bureaucracy that is **larger than the optimal size**. In South Africa we have seen how the so-called homelands policy of the past and the increase in the number of provinces after the 1994 election have contributed to a duplication of government activities which has put additional pressure on government expenditure.

Individuals with shared interests are often able to organise themselves into powerful interest groups and put pressure on government to implement programmes and pass legislation that will meet their own parochial interests. South African farmers, labour unions and organised business put pressure on government to grant them special tax allowances which will erode the revenue base of government.

Government and the Economy: Long-Term Effects

The long-term role of government can be usefully viewed within the context of the '**new growth theory (NGT)**'. Several kinds of public investment and expenditure programmes are reputed to confer significant **positive externalities** (both a pecuniary and technological kind), and it is these programmes that policymakers should turn in their quest to promote economic growth and development.

According to NGT we adopt a much broader definition of 'capital' and focus attention on the role played by each of the components of capital in the growth process. In addition to privately owned physical capital, 'capital' also includes the following three components:

- The **existing physical infrastructure** (e.g. Roads, street lighting, and sewage systems)
- Accumulated **human capital** acquired through education, training, and healthcare
- The **stock of technical expertise** acquired through learning-by-doing and research and development (R&D).

The main thrust of NGT is that **additions to capital may yield increasing returns** as they create positive externalities that benefit a range of sectors and industries. Investment in the physical infrastructure creates externalities by lowering production costs and boosting returns in the private sector. Likewise, recipients of education may transfer their skills free of charge to third parties, healthier citizens will be more productive and limit the spread of disease, etc.

Each of the components of capital can be **influenced by government through appropriate policy intervention**. Since the marginal private benefits from investments in physical infrastructure, human capital, and R&D are lower than the corresponding marginal social benefits, the untrammelled operation of the market will lead to an under-provision of these services. NGT thus provides a strong justification for government intervention in these areas as it will create favourable conditions for private investment and economic growth.

Investment in a country's physical infrastructure usually boosts the productivities of a range of inputs used by private enterprises. It may lower transport and communication costs and facilitating the discovery of new input and output markets. It may also "**crowd in' new private investment**, including foreign direct investment (FDI).

Infrastructure investment is usually financed by means of borrowing on the open market but it is more likely to have a net **crowding-in effect** than a **crowding-out effect**. The reason is that government borrowing usually constitutes a small portion of total borrowing, so that it is unlikely to have an effect on interest rates. But even if it did cause a rise in the interest rate, private investors would not necessarily be deterred by it.

The role of infrastructure in growth is substantial, significant and frequently greater than that of investment in other forms of capital.

Vocational training often produces higher returns than poor quality formal schooling, and so too does good quality primary and secondary education in relation to certain categories of tertiary education. Other categories of tertiary education in turn provide the specialised skills that are necessary for sustained economic growth.

These findings are clearly important in helping governments to prioritise their spending programmes, especially in view of the important **redistributive role** that such programmes play. The prioritisation of state expenditures, based on efficiency criteria, can help governments to do two things - **achieve a more equitable distribution of income** and **create the conditions for sustainable economic growth over the long term**.

STUDY UNIT 7 – POVERTY AND INEQUALITY IN SOUTH AFRICA

7.1 THE DISTRIBUTIONAL CONTEXT IN SOUTH AFRICA (8.1.2)

Primary Income versus Secondary Income

A household's power of disposal over goods and services depends upon the personal incomes, the taxation and spending of government.

- Primary income or personal income of any group is the actual value of income received by individuals and households, including the value of subsistence production activities such as subsistence agriculture.
- Secondary income can then be regarded as primary income minus direct taxation (which leaves disposable income), to which is added the value of government services consumed. Therefore:

$$Y_s = Y_p - T_d + G$$

Where:

- Y_s - represents secondary income
- T_d - represents secondary income
- Y_p - represents primary income
- G - represents government spending

Social spending is the only category of government spending that can be assigned to beneficiary households with any degree of certainty. If other public expenditures are ignored, Equation becomes

$$Y_s = Y_p - T_d + G_s$$

Where:

- G_s - represents social spending

Considering the high degree of racial inequality of primary incomes in South Africa, it is indeed pertinent to ask what the role of the budget is in accentuating or reducing this inequality through its impact on secondary income. This goes beyond investigating only the fairness of the budget itself, and also asks how much inequality remains once the budget has had its effect.

7.2 THE ROLE OF PUBLIC FINANCE IN REDISTRIBUTION (8.2)

Conceptually the factor $(G_s - T_d)$ combines the notions of **tax and expenditure incidence** into what is called **fiscal incidence**, that is, the **net result of the incidence of tax burdens and expenditure benefits in society**.

The **balanced budget incidence**, entails calculating the combined effect of government spending and the taxes levied to finance it, assuming there is no budget deficit or surplus.

In this context, the distributional impact of a tax depends not only on the incidence of the tax itself, but also on how the government spends the proceeds from the tax. However, balanced budget incidence ignores the distributional impact of fiscal policies on primary incomes and hence fails to take into account the overall distributional context. It ignores the fact that primary income has already been influenced by the impact of fiscal policies on market prices. At any point in time, therefore, the primary income distribution already incorporates an element of fiscal incidence. To avoid '**double counting**' balanced budget incidence studies should adjust for this

Government's Distributive Role

The government budget embodies influences income distribution by determining which services are provided to whom and how these expenditures are to be financed (taxes and loans), including their impact on the long-term distribution of human capital that fundamentally determines the distribution of earnings (primary incomes).

From a broad economic perspective it is possible to identify a number of roles that the government can play in affecting distribution (see Bromberger, 1982:168):

- **As a rule-maker** - in particular the rules of competition in markets or rules of access to levels of market opportunity
- **As a controller** of prices and wages in markets
- **As a market operator** - as a major employer of labour and through the size and nature of its purchasing activities
- **As an influence** on the long-term pattern of activities, for example industrial decentralisation measures that affect location decisions of employers, or the impact of taxes on the capital intensity of production
- **As taxpayer, supplier** of public goods and welfare services, and payer of transfer incomes
- **As (potential) redistributor** of assets that carries claims to current and future income.

The first four roles mainly affect primary incomes in the market and are only indirectly related to public economics. The last two activities however influence secondary income through the budget.

Government Taxation

The extent to which taxes can be shifted of course differs, as does the progressivity of different forms of taxes. Just as the perceived equity of a tax system is very important for the legitimacy of taxes and for tax morality, so the **perceived (statutory) incidence**, in contrast to the economic incidence, is often quite crucial.

In terms of the **statutory incidence**, the most progressive taxes are usually income taxes. An income tax may even be extended to include a negative income tax (i.e. a transfer) for low levels of income, thus strengthening its progressive distributional effect. This may be seen as an alternative to means-tested social transfers. Wealth taxes are also progressive, and appropriately selected excise duties too can have a progressive impact.

By contrast, most other indirect taxes are usually **relatively regressive**. This regressivity can in some instances be reduced through certain exclusions, for example zero-rating certain food items from value added tax; however, this process has to be weighed against the increased complexity of administration and the erosion of the tax base that it entails.

Government Spending, Subsidies, Externalities, and Income Transfers

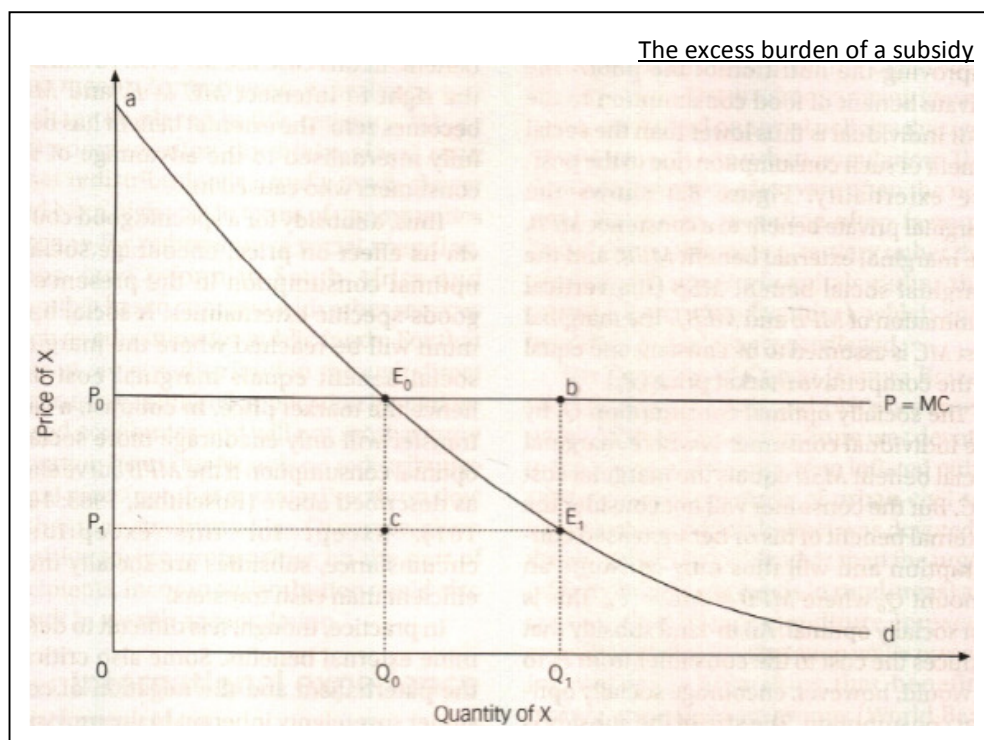
The distributive stance of a government can furthermore be determined from its **spending priorities**. Social spending includes important income transfers as part of social security, education, health, and housing.

In analysing social spending, one should take note that subsidisation or the public provision of goods or services gives rise to economic inefficiencies in the form of the **excess burden of a subsidy**.

The concept of consumer surplus as developed by Marshall and Hicks may be used to show that the **costs of a subsidy are larger than its benefits**, thus leaving an **excess burden**. Assuming constant production costs, if the full benefit of a subsidy were passed on to consumers, it would result in a lowering of the price charged from P_0 to P_1 and an increase in quantity demanded from Q_0 to Q_1 . This would **increase consumer surplus** (the difference between what the consumer is willing to pay and the actual price) from the area aP_0E_0 to the area aP_1E_1 . Thus, the subsidy results in a benefit for consumers measured as area $P_0E_0E_1P_1$ consisting of area $P_0E_0CP_1$, due to the lower price for the original quantity demanded, and area E_0E_1c , originating from the extra quantity purchased due to the lower price.

However, the cost of the subsidy to the state, that is, area $P_1P_0bE_1$, (the subsidy per unit multiplied by the Q_1 units consumed) still has to be accounted for. This cost exceeds the consumer benefit by area E_0bE_1 . This is known as the **excess burden**, or **deadweight** or **welfare loss**, of the subsidy.

As subsidies interfere with consumer choice and therefore lead to a socially sub-optimal outcome, it seems that public subsidies should in principle be avoided, perhaps by rather providing income transfers if the intention is to redistribute resources.



Asset Redistribution

Asset redistribution usually takes place over a period of time through the interaction of **market and fiscal processes**. Direct interventions to force rapid changes in asset distribution usually require a large degree of coercion and are therefore only common in **post-revolution situations**, for example nationalisation or land redistribution without full compensation. Where full compensation does take place, such asset redistribution is usually much slower and has a high cost in terms of opportunities forgone for redistributive social spending. Thus, land reform in South Africa and Namibia has to compete with other services such as education for public funds. Asset redistribution through **direct measures has limited application** in market-based economies. Our focus is on redistributive fiscal measures that operate directly on flow rather than stock variables. Of course, given **positive saving propensities** on the part of recipients, income redistribution could also result in wealth accumulation.

Radical asset redistribution is a policy option that is seldom used in market-based economies, because of the negative economic implications it may have. The distribution of wealth is a product of history and personal achievement. Wealth, for example, can be attributed to accumulated savings and the latter can to some extent be influenced by government actions (eg taxes and government's ability or failure to promote a stable social and economic environment).

International Experience and Practice

Since the nineteenth century, the state's role has gradually shifted from an earlier emphasis on the **indirect satisfaction of needs** to a greater emphasis on the more **direct satisfaction of needs**. There was a particularly strong rise in income transfers in the form of social security spending. Increased expenditures also necessitated increased taxation, which later caused growing tax resistance.

In trying to meet the demands of strong urban pressure groups, many governments embarked on social policies that were biased towards the urban population. Thus the major beneficiaries were often the non-poor. Moreover, spending often favoured higher-level services (e.g. tertiary rather than primary education, hospitals rather than primary healthcare facilities), which again benefited the relatively privileged.

In most developing countries social security has been relatively neglected, with education usually dominating social spending, particularly in Southern Africa which has recorded many of the highest education spending ratios (expressed relative to GDP) in the world.

7.3 Selected Policy Issues (8.4.3; 8.4.5)

Social security protects people against various contingencies, such as income loss from unemployment, disability, injury sustained at work, illness or old age. The South African social security system has two major components:

- **Occupational insurance** is based on contributory insurance to protect those in formal employment and includes the following: **retirement benefits** for a large part of those in formal employment: an inadequate system of workers compensation; a system of **unemployment insurance** which cannot address the major risks associated with structural unemployment; and health or medical insurance for some of the employed.
- **Social assistance** (social grants or social transfers) has three main pillars, namely social **old-age pensions, disability grants, and child support grants**. All are means-tested to ensure targeting at the poor and are funded from the national budget.

Insurance against risk is unlikely to be offered by the private sector due to the problem of **adverse selection**. This means that people most likely to face certain risks (e.g. of unemployment) and to collect benefits from insurance, will have an especially high demand for, and are thus more likely to buy, such insurance. However, the high premiums needed to compensate for these high-risk clients could lower the demand to such an extent that it is unprofitable for a private insurance firm to offer such coverage. This problem of **moral hazard** which is said to exist when economic behaviour adjusts to incentives in such a manner as to increase the likelihood that the event insured against will actually take place. For instance, the presence of lucrative unemployment insurance may make continued employment less attractive and therefore lead to greater unemployment than would have been the case in the absence of such insurance. This phenomenon ties in with the question of **asymmetric information**: the insurer's insufficient knowledge about the insured person makes such an insurer more susceptible to such moral hazard. Thus, adverse selection and moral hazard generally result in market failure and an inefficient allocation of resources: there is a lack of insurance against risks for which a sufficient demand does exist.

Social Assistance

The figure explains the effect of a **cash transfer** or the **free provision of public goods and services** on the work effort of an economically active individual who is therefore also subject to income tax. We measure hours per time unit and the corresponding unit income on the vertical axis. OA hours are available to the worker to allocate between work and leisure. This distance is referred to as the **time endowment**. The vertical line AX signifies that time endowment is fixed at all income levels.

The line AB shows his or her combined budget and time constraint, where sacrificing leisure would allow him or her to earn a bigger wage income and therefore to increase consumption of goods.

Initially the worker is in equilibrium at E_0 on the price line AB , given his or her indifference curve (I_0) between goods (income) and leisure, sacrificing CA hours of leisure in order to consume OD of goods. OD is the value of goods which can be purchased with the revenue earned by working AC hours and thus, assuming no saving, is equal to total income earned.

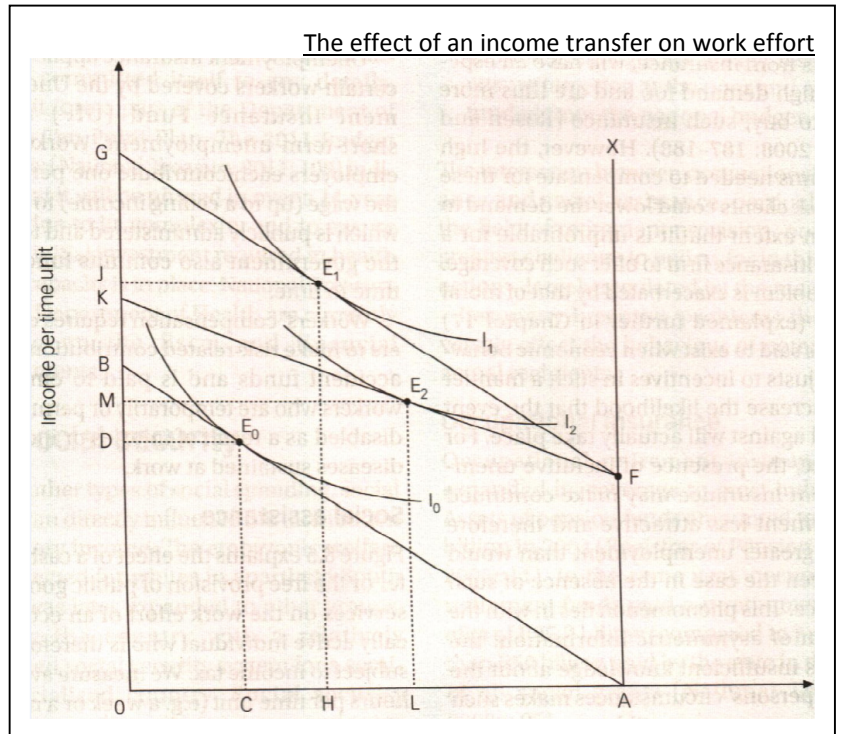
If the **government now provides an income transfer** irrespective of its effect on individual work effort, the budget constraint shifts parallel to the right to FG . Note the relative prices of the items of choice remain unaffected. The worker is now at point E_1 on the new budget line and on the higher indifference curve. Thus, consumption of both leisure and goods is increased (from OC to OH and from OD to OJ). The individual now has a higher level of welfare but works fewer hours [AH as opposed to AC] due to the income effect arising from the cash transfer.

If an **income tax is now imposed** to finance the cash transfer, this would reduce the opportunity cost of leisure, thus swivelling the budget line from FG to FK . The individual moves to E_2 (where FK is tangent to indifference curve I_2) and work effort is further reduced to AL . This is even lower than the work effort arising from the cash transfer alone (AH).

The combined effect of the cash transfer and taxation to finance it is thus to reduce work incentives in two ways, namely;

1. **The negative (income) effect CH of the cash transfer plus**
2. **The negative net (income and substitution) effect HL of the income tax.**

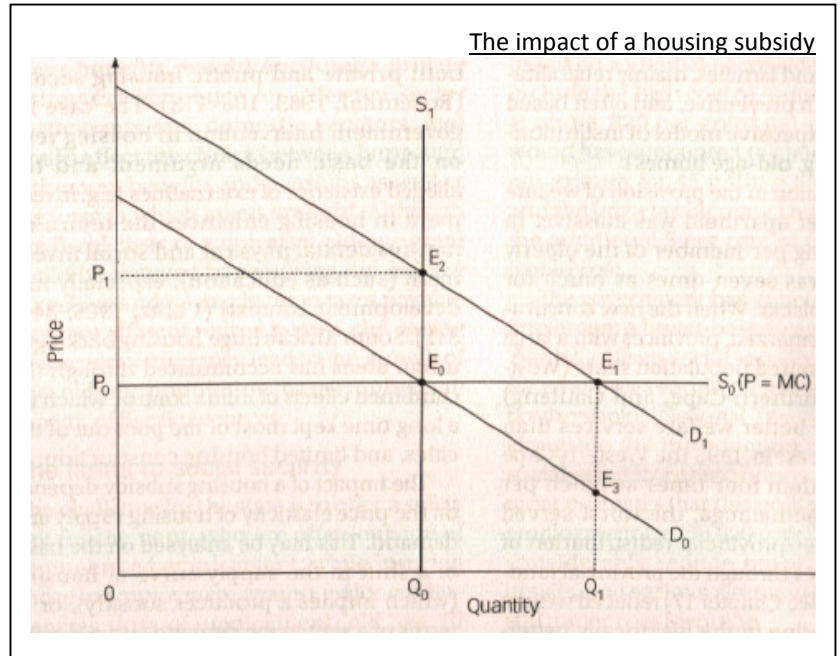
Government needs to be as careful in the design of its cash transfer programmes as in its taxes to ensure that, in combination they lead to as little reduction in work effort as possible. Social old-age pensions are an example where the cash transfer would not provide a disincentive to work as the beneficiaries are not part of the labour force.



Subsidisation of Housing

The case for government intervention in housing rests on the basic needs argument and the alleged existence of externalities (e.g. investment in housing enhances the return on non-residential physical and social investment). South Africa's huge housing backlog in urban areas has accumulated through the combined effects of **influx control**, which kept the poor out of the cities, and limited housing construction.

The impact of a housing subsidy depends on the **price elasticity of housing supply and demand**. This may be analysed on the basis of a shift in the supply curve of housing (which implies a producer subsidy), or in terms of a shift in the demand curve (a consumer subsidy). As in the case of an indirect tax, the result in terms of the shifting of the subsidy benefit is the same. We can illustrate the impact of the subsidy by means of a shift in the market demand curve for housing. Assume firstly a perfectly price-elastic supply of housing S_0 (i.e. $P = MC$). The initial equilibrium is at E_0 , with P_0 and Q_0 the corresponding equilibrium price and quantity. A housing consumer subsidy will increase the purchasing power of prospective homeowners and thus shift the demand curve upwards and to the right, from D_0 to D_1 . The new equilibrium is at E_1 . Since the price of houses has not been affected, buyers enjoy the full benefit of the subsidy. No shifting of the benefit occurs.



At the other extreme, there is a fixed housing supply (a perfectly price-inelastic supply curve such as S_1), a housing subsidy will only result in an increase in house prices (from P_0 to P_1) to the benefit (increased capital gains) of existing homeowners, thus completely cancelling the benefit which the subsidy was supposed to have for potential buyers.

In practice, neither of these extremes is likely, and the net impact is likely to be that some benefit will accrue to existing homeowners through an increased demand for housing and higher prices, and some benefit to the intended beneficiaries through an increase in their purchasing power and an acceleration of housing supply. To the extent that the price of new houses is increased by the subsidy, some of the benefit is shifted to home-builders.

The major problem is that the incomes of a large part of the urban population are too low to allow them to afford even a relatively small house with rudimentary services.

STUDY UNIT 8 – INTRODUCTION TO TAXATION AND TAX EQUITY

8.1 SOURCES OF FINANCE (10.1)

The dominant source of finance for public expenditure is taxation. In addition to taxation there are four other important sources of finance: user charges, administrative fees, borrowing, and 'inflation taxation'!

1. **User charges (benefit taxes)** - are prices charged for the delivery of certain public goods and services- The role these charges play in the allocation and distribution of resources is analogous to the role of prices in the market mechanism. The important difference is that user charges are set in the 'political market'. User charges can only be levied if exclusion is possible, i.e, it should be possible to exclude those who do not pay for the consumption of the public good or service in question. Examples of user charges include toll roads, public swimming pools, ambulance services, and university education.
2. **Administrative fees** - similar to user charges but differ in the sense that the service (or benefit) received in return for the fee is defined rather broadly and imprecisely. Such fees include business licences, television licences, diamond export rights, fishing licences, and motor vehicle licences.
3. **Government can borrow** - from its own citizens and from abroad. Borrowing is often used to finance capital expenditure. Borrowed funds must be repaid at some point and can therefore amount to deferred taxes. Because lenders have to be adequately compensated, borrowed money should be spent on productive activities.
4. **Government-induced inflation** - If public expenditure is financed in such a way that increases in the money supply occur, such financing may eventually raise the price level. Inflation changes the real value of public debt. If government borrows R2 000 from a taxpayer (e.g. if government imposes a loan levy on all taxpayers with incomes in excess of R100 000) and inflation is 10 per cent. If the value of the loan is not linked to a price index, the real value of government debt decreases.

8.2 DEFINITION AND CLASSIFICATION OF TAXES (10.2)

Taxes are transfers of resources from persons or economic units to government and are compulsory. Taxes are compulsory due to the free rider problem. As no one will pay taxes voluntarily, people have to be compelled to do so. The fact that government has legally been granted the power to tax distinguishes government's confiscation of resources through taxation from other involuntary transfers of resources

Tax revenue of SA is grouped into six categories, each of which is then sub-divided further, these are:

1. Taxes on **income and profits** (59,1 per cent)
2. Taxes on **payroll and workforce** (1,3 per cent)
3. Taxes on **property** (1,4 per cent)
4. Domestic taxes on **goods and services** (37,8 per cent)
5. Taxes on **international trade** and transactions (4,1 per cent)
6. **Stamp duties and fees** (0,001 per cent).

Tax base and rates of taxation

Taxes can generally be imposed on three tax bases: income, wealth, and consumption. A tax on people can be added to these three bases. A poll tax (lump sum tax per head) is an example of a tax on people. Tax bases can also be viewed in terms of their **flow and stock** characteristics. Flows are associated **with a time dimension** and are measured over a period. Stocks are measured at a **particular point in time**. Most countries have hybrid systems, exploiting three (or four) bases simultaneously. Once the tax base has been identified, the **tax rate structure** can be set. The tax rate refers to the amount of **tax levied per unit of the tax base**. Three variants can be distinguished: **proportional, progressive, and regressive taxes**. The rate structure can be described in at least two ways. When the average tax rate increases as the tax base increases, the tax is **progressive**. If the average tax rate decreases as the tax base increases, the tax is **regressive**.

Taxes that generate the same proportion of income as income rises are **proportional** (e.g. corporate income tax). A tax with a proportional rate structure is also called a **flat-rate tax**. Taxes that take an increasing proportion of income as income increases are **progressive taxes** (e.g. personal income tax). If a tax generates a decreasing proportion of income as income increases, it is a **regressive tax** (e.g. a value added tax without any zero-ratings).

General and selective taxes

- **A general tax** (also called broad-based tax) is one that taxes the entire tax base and allows for no exemptions. A value added tax (VAT) without any exemptions or zero-ratings is a general tax on consumption. An income tax that taxes all sources of income (including capital income) without any tax deductions would be a general tax.
- **Selective taxes** (also called narrow-based taxes) are imposed on one or a few products or only income, that is, excluding leisure. The whole tax base is therefore not taxed. An excise tax on cigarettes is an example of a selective tax.

The importance of the distinction between general and selective taxes lies in the fact that under certain assumptions' general taxes are similar to head or lump sum taxes that leave relative prices unchanged. In contrast, selective taxes distort relative prices by driving a wedge between the before-tax price and after-tax price of a commodity. This violates the Pareto efficiency condition.

Specific and ad valorem taxes

Taxes can also be specified according to the size or the value of the tax base. The size of the tax base can be measured in terms of weight, quantities, or units. It is called a **unit tax** or **specific tax**. Examples of specific taxes in South Africa in 2011/12 are excise duties on sparkling wine, beer made from malt and cigarettes.

Taxes imposed on the value of products are called **ad valorem taxes**. Such a tax is usually levied as a rate (i.e. a percentage) of the excisable value of a commodity. VAT and excise duties are examples of ad valorem taxes and are often imposed on 'luxuries.

Direct and indirect taxes

Direct taxes are imposed directly on individuals and companies. Indirect taxes are imposed on commodities. This distinction fundamentally revolves around the issue of **tax incidence** (i.e. the question of who really pays the tax).

From the perspective of **tax shifting**, **direct taxes** are defined as taxes that cannot be shifted readily. They are collected from individuals, households, or firms and allow for the possibility of adjusting the tax according to the personal circumstances of the taxpayer (e.g. the marital status, gender, size of household, wealth status). These taxpayers are the intended bearers of the tax burden and it is assumed that they pay the tax over to SARS.

Indirect taxes are taxes that are imposed on commodities or market transactions and are likely to be shifted such as excise duties and fuel levies. It is also more difficult to adjust the tax rate to the personal circumstances of the consumer. In the case of indirect taxes it is often possible to shift the burden of the tax to someone else. VAT is collected from merchants who, in turn, can pass on the tax to consumers by way of a price increase. The consumer then indirectly bears the burden.

8.3 PROPERTIES OF A 'GOOD TAX'

A good tax system must first of all generate sufficient revenue to finance budgeted government expenditure. The other more important criteria or properties of a 'good' tax system can be classified under four headings:

- **Equity:** Taxes should promote an equitable distribution of income. Cognisance has to be taken of the fact that the burden of taxes can be shifted. To determine fairness, the incidence of taxes has to be examined.
- **Economic efficiency:** All taxes impose a burden and most taxes affect the behaviour of taxpayers ('excess burden'). Taxes should be designed so that their distorting effects on the choices made by taxpayers are minimised.
- **Administrative efficiency:** Taxes are levied to yield sufficient revenue but in order to be efficient, administration and compliance costs have to be kept low. This calls for tax simplicity and certainty.
- **Flexibility:** As economic circumstances change, taxes and tax rates need to adjust. Taxes should therefore be flexible enough to facilitate macroeconomic stability and economic development.

8.4 TAXATION AND EQUITY: CONCEPTS OF FAIRNESS

Fairness is a subjective concept and it lies in the eye of the beholder. Although fairness is a value-laden concept, economists can help to make informed value judgements.

Benefit principle

The benefit principle stipulates that the tax burden should be apportioned to taxpayers in accordance with the benefits each receives. It seems unfair to expect non-users also to pay for the construction and maintenance. Here we have an example of **forced carrying**: while free riding refers to someone failing to carry a 'proper' tax burden, forced carrying refers to someone being made to carry a heavier than 'proper' burden.

A major advantage of the benefit principle is that

- It links the **expenditure side of the budget to the revenue side**.
- the allocative procedures of market behaviour are approximated. Individuals can **adjust their consumption** of services.

Disadvantages are:

- the scope for applying the benefit principle to government funding is rather restricted. Governments generally provide goods and services that are public in nature, i.e. the benefits are generally **non-excludable**.
- it takes the existing distribution of **income and wealth for granted**. The effective demand for public goods and services is often determined by this distribution and can be skewed towards the rich.

The benefit approach can ideally only handle a tax-expenditure process that has no redistributive objectives. It cannot handle a tax-expenditure programme designed for redistributive purposes (e.g. taxes levied to finance transfer payments or expenditure programmes designed to benefit the poor rather than the rich) or where it is administratively not feasible to exempt poor people from paying for the benefits of a particular service (e.g. a toll road).

Benefit taxes (often referred to as **user charges**) are nevertheless levied in some cases, for example, tolls for roads and bridges, admission charges to museums and parks, license fees, and (to a certain extent) university tuition fees and school fees.

The benefit approach may therefore be rationalised on the **grounds of equity**. However, a frequent criticism against user charges is that they may contradict the equity objective by preventing the poor from using government-supplied services.

Ability-to-pay principle

The ability-to-pay principle calls for people with equal capacity to pay the same amount of tax (**horizontal equity**), and for people with greater capacity to pay more (**vertical equity**).

The implementation of a tax system based on ability to pay requires public consensus on an appropriate definition of ability to pay. It also calls for consensus on the rate structure. Income is generally regarded as one such measure or criterion. Other possible measures include consumption, wealth, and utility.

Income, however, is by no means a perfect measure of a person's ability to pay. Income measures outcomes and does not necessarily reflect **ability or capacity**.

Suppose income is used to measure ability to pay. This still leaves the question of how to determine the taxes payable by people with different incomes. Most countries try to deal with vertical equity by applying the progressive tax rule to income. To determine whether a tax is truly progressive, proportional, or regressive, one needs to know who really pays the tax. This is what tax incidence is all about

8.5 TAX INCIDENCE: PARTIAL EQUILIBRIUM ANALYSIS

All taxes reduce the real disposable income of taxpayers. All taxes therefore involve a burden.

The **statutory incidence** refers to the legal liability to pay the tax over to the revenue authorities. Since taxes affect behaviour, economists are concerned with who bears the tax burden, that is, the **economic incidence** of the tax. The importer can shift cigarette prices and shift the actual burden forward to retailers who can pass the tax to the consumers. The tax burden can also be shifted backward if the importer cuts back on staff or lowers real wages. The tax can be avoided by cutting back on the taxed activity by reducing imports of cigarettes.

Tax incidence studies may apply either a

1. **Balanced-budget incidence methodology** - the **overall distributional effect of a tax and the spending** financed by the tax is considered. Income taxes lower real disposable income but the income tax revenue is spent on education and other public services that raise real disposable income. The advantage is that it relates the cost of spending programmes to those who pay for it. The disadvantage is that tax revenue is pooled in the national revenue account and linking an expenditure item to a tax source is almost impossible
2. **Differential-incidence methodology** - considers the **distributional impact as one tax is substituted for another**, holding total revenue and expenditure constant. The benchmark tax often used for purposes of comparison is a lump sum tax that does not affect relative prices and hence economic behaviour. For example, suppose government replaces a lump sum tax on all redheads with tin excise tax on beer drinkers that yields the same tax revenue. Assuming that redheads are not beer drinkers, they would gain and beer drinkers would lose. Furthermore, the owners of breweries and their employees will be affected by the tax. By comparing the total impact of the tax change on the incomes of beer drinkers to those of redheads, we are engaging in differential tax incidence analysis.

However, when a tax is imposed on one good it usually means that the prices of other goods change as well. Once the secondary effects of a tax are taken into account, tax incidence is studied in a **general equilibrium framework**.

- **Partial equilibrium studies** are less complex since all the ramifications of a tax change are not considered. It is ideally suited to studying taxes levied on goods and services that are characterised by low degrees of substitutability or complementarity. Where the secondary effects are considered to be small, uncertain, or spread thinly over a number of other markets, economists would argue that these can be ignored and that the conclusions based on partial equilibrium analysis will not differ significantly from those based on general equilibrium analysis.
- The **general equilibrium framework**, is more suitable for studying the incidence of a general sales tax that is levied on a broad base, or a levy on an important product such as fuel that has important ramifications for the economy. Since general equilibrium analysis considers relative price changes in more than one market, it is conceptually superior to partial equilibrium analysis.

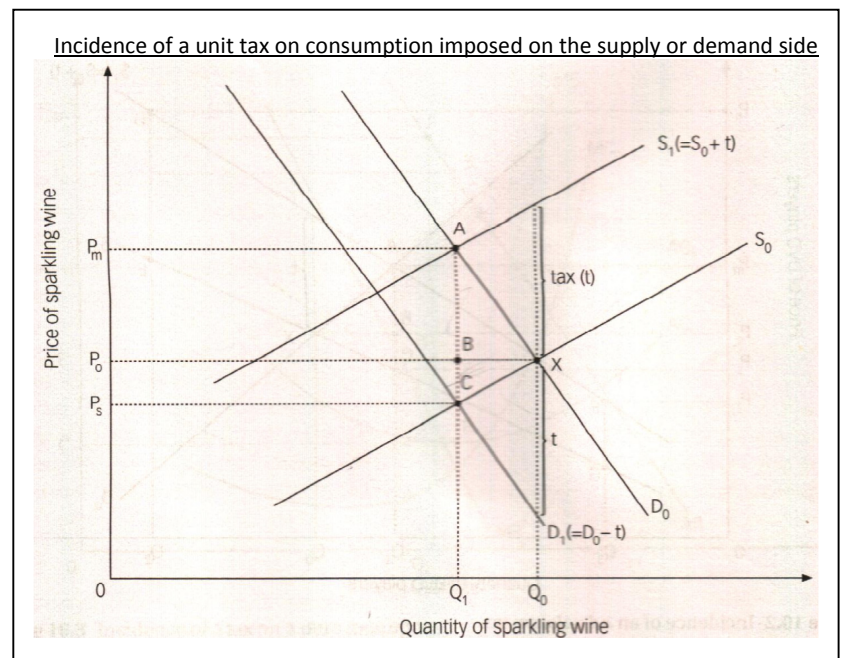
Partial equilibrium analysis of tax incidence

Two types of taxes are considered: a unit tax (i.e. a fixed amount per unit of a good or service sold) and an ad valorem tax.

Incidence of a Unit Tax

In the figure D_0 is the demand curve and S_0 the supply curve for sparkling wine. The before-tax equilibrium price is P_0 and the equilibrium quantity is Q_0 . Suppose that an excise tax (a unit tax) of t per unit is imposed on sparkling wine. The excise tax is a fixed amount per unit of the product.

If the tax is collected from the seller (**statutory burden is on the seller**), it means that the minimum price at which firms will supply the equilibrium quantity Q_0 is $P_0 + t$. The sellers are interested in the after-tax price and will try **recover the full tax amount** at any given quantity - they will charge a higher price. The supply curve therefore shifts up vertically by the tax amount to $S_1 (= S_0 + t)$. Note that because it is a unit tax, **S_1 is parallel to S_0** . The new market equilibrium is at point A with price P_m and quantity, Q_1 . The **market price**, price paid by buyers of sparkling wine, increases from P_0 to P_m . Sellers receive the after-tax price, P_s (i.e. $P_m - t$). What is the **total tax revenue that accrues to government**? Total tax revenue is the tax per unit t multiplied by the number of units sold. Geometrically the total revenue is represented by the rectangle, $P_m A C P_s$.



Before the imposition of the tax, consumers paid a price P_0 for the product. After the tax they pay, P_m . Note that the price that buyers have to pay did not increase by the full amount of the tax ($P_m - P_0$). The total tax amount paid by consumers is equal to the rectangle, P_mABP_0 .

What about the sellers? The before- tax price received was, P_0 . The after-tax price received by sellers is, P_s which is less than P_0 . Producers therefore also pay part of the tax. The total tax amount paid by sellers is equal to the rectangle, P_0BCP_s . From this analysis, **partial equilibrium**, shows the total tax burden is split between buyers and sellers, the tax **incidence falls on both**.

Would the result be any different had the **statutory burden been on buyers**? The answer is no. To show this, one must remember that the demand curve in our example indicates the maximum price consumers are willing to pay for each different quantity of sparkling wine. The before-tax equilibrium price is P_0 and the equilibrium quantity is, Q_0 . At the equilibrium quantity Q_0 the consumer is willing to pay only, P_0 . If a tax of t is now collected from the buyer for this quantity, the buyer will still be willing to pay only, P_0 . The price paid to the seller, however, will be reduced by the tax amount ($P_0 - t$) that is paid over to SARS by the buyer. This will hold for each point on the demand curve. In this case, therefore, the effective demand curve (i.e. as perceived by the seller) shifts downward by the tax amount. The new demand curve is $D_0 (D_0 - t)$. Since a unit tax of t is imposed, D_1 is parallel to D_0 . The after-tax equilibrium is at point C with price P_s and quantity Q_1 . While the original demand curve D_0 shows what price buyers are willing to pay for each quantity, the demand curve D_1 shows the price received by sellers (the after-tax price) for each quantity. For quantity Q_1 buyers are willing to pay, P_m , and sellers receive, P_s (i.e. $P_m - t$). The result is exactly the same - the tax incidence is on both buyers and sellers.

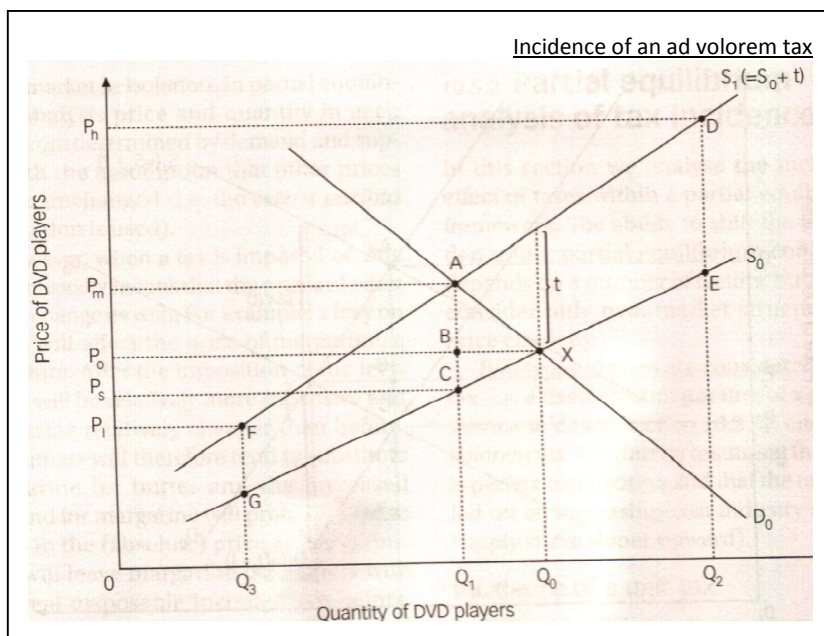
Incidence of an ad valorem tax

An ad valorem tax is levied as a percentage of the price of a good or service. The most important difference between the analysis of an ad valorem tax and a unit tax is that the **after-tax supply curve swivels** in the case of an ad valorem tax (compared to the parallel shift in the case of a unit tax).

Because the tax is proportional, the higher the price, the greater the amount of tax to be paid over by producers will be. For example, at a price of R1 000 per DVD player and an ad valorem rate of 20 per cent, the absolute amount of tax is R200. If the price is R2 000, the absolute amount of tax is R400.

Thus, when an after-tax supply curve such as S_1 is constructed, the vertical distance between S_1 and S_0 (the before-tax supply curve) at a relatively high price such as $P_h (= DH)$ should be greater than at a low price such as, $P_1 (= FG)$. The rest of the incidence analysis of an ad valorem tax is similar to that of a unit tax. The total tax burden of buyers is the rectangle P_mABP_0 and that of sellers is, P_0BCP_s . The ad valorem tax rate, expressed as the ratio of tax to the gross price paid by the

buyer, equals, $\frac{AC}{AQ_1}$.

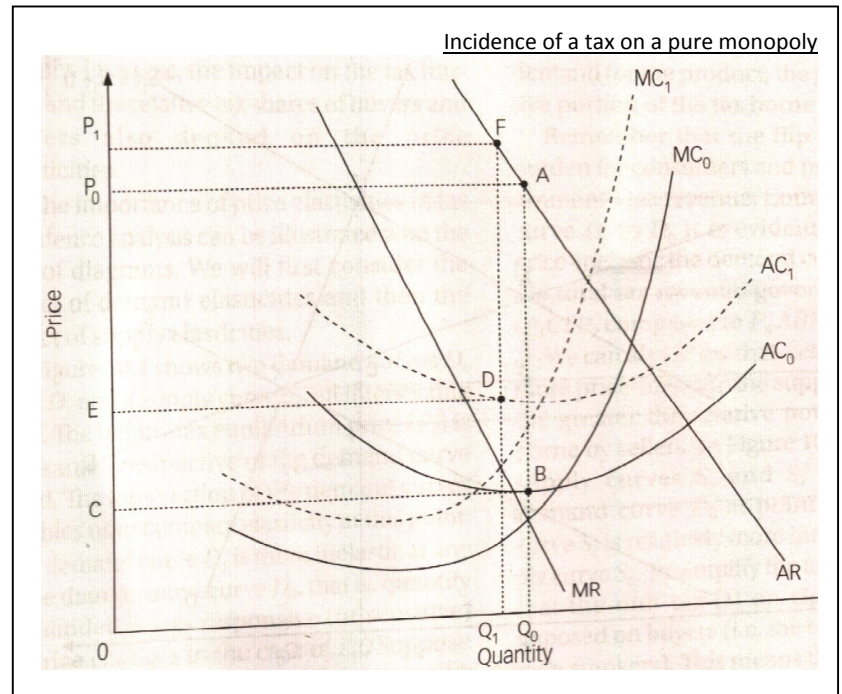


Incidence and Pure Monopoly

Pure monopoly is another extreme form of market structure. It is often taken for granted that a monopolist, being a price maker, can always shift taxes forward in full. Under certain circumstances (e.g. where a monopolist is not maximising profit) this may indeed be the case; however, the conclusion does not necessarily apply. Let's consider the tax shifting capacity of a monopolist-maximising profit.

A monopolist maximises profits where **marginal cost (MC) equals marginal revenue (MR)**. In the figure the before-tax equilibrium of the pure monopolist is at quantity Q_0 and price P_0 . The before-tax profit is the area P_0ABC .

A **unit tax on the output** of the monopolist will raise the average cost (AC) and the marginal cost (MC) of the firm. The reason why both AC and MC increase is that the tax is levied on each unit produced and is therefore viewed by the firm as a variable cost. In the figure the average cost curve shifts from AC_0 to AC_1 , and the marginal cost curve from MC_0 to MC_1 . Profits are now maximised at an output of Q_1 and at price P_1 .



When you compare the before-tax equilibrium to the after-tax equilibrium, you will notice the following:

- The after-tax price is higher and the quantity is lower than before the imposition of the tax. The increase in the price from P_0 to P_1 , indicates the extent of forward shifting.
- The after-tax profit (the rectangle P_1FDE) is less than the before-tax profit (the rectangle P_0ABC). The pure monopolist therefore does not shift the tax burden fully.

Monopolists can also be taxed (at least in theory) on their **economic profits**. Economic profit (or excess profit) is the difference between total revenue and total costs (both explicit and implicit costs). Implicit costs include the opportunity costs of self-owned resources (i.e. normal profit). The cost curves represent economic costs. In other words, if average revenue exceeds average cost, economic (or excess) profits are earned. The pure monopolist's before-tax profit (area P_0ABC) is considered to be economic profit. A tax at a rate of t per cent on the monopolist's profit of P_0ABC will simply reduce its economic profit by the tax amount. Neither the average nor marginal cost curves will be affected. The monopolist will maximise after-tax profit at the same before-tax level of output and price. The tax is therefore borne fully by the owners of the firm and in such circumstances the profits tax cannot be shifted.

Incidence and Price Elasticities of Demand

Tax incidence is affected by **market structure** (e.g. perfect competition or pure monopoly) and **price elasticities**. When a tax is imposed, both the quantity demanded and the quantity supplied at equilibrium decreases. The **magnitude depends on the elasticity** of demand and supply. The impact on the tax burden and the relative tax shares of buyers and sellers also depend on the price elasticities.

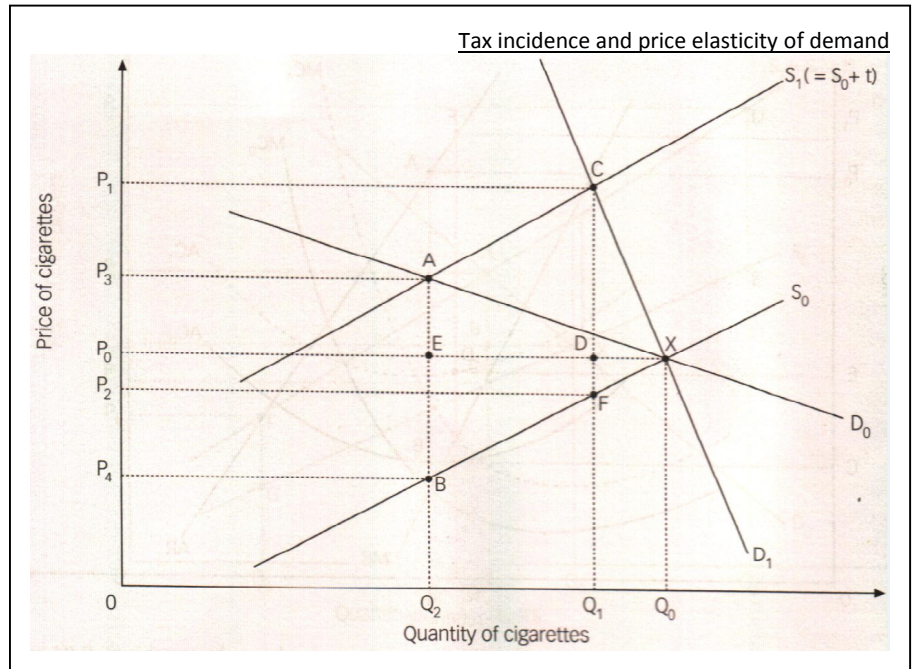
The figure shows two demand curves D_0 and D_1 and a supply curve, S_0 , all intersecting at X . The before-tax equilibrium price (P_0) is the same, irrespective of the demand curve used. The intersection of the demand curves enables us to compare elasticity at this point. The demand curve D_1 is more inelastic at any price than demand curve D_0 , that is, quantity demanded is less responsive (or sensitive) to price changes in the case of D_1 .

Suppose the taxed good is cigarettes. If a unit tax (t) is now imposed on the importer or seller, the effective supply curve shifts upwards and to the left from S_0 to S_1 ($S_1 = S_0 + t$). Let us consider the relatively inelastic demand curve D_1 , first.

If the demand for cigarettes is represented by D_1 the price charged to buyers increases from P_0 to P_1 . The price received by the seller (or importer) decreases from P_0 to P_2 . The proportional price change for the buyers therefore exceeds that of the sellers. The tax burden (P_1CFP_2) is divided between the portion borne by buyers (the area P_1CDP_0) and the portion borne by the sellers (the area P_0DFP_2). In the case of the more elastic demand curve D_0 , the proportional price changes for buyers and sellers are approximately the same in this example. The total tax burden of P_3ABP_4 , is then divided between the buyers (the area P_3AEP_0) and the sellers (the area P_0EBP_4). Thus, the more price-inelastic the demand for a product is, the greater the relative portion of the tax borne by buyers, *ceteris paribus* (sometimes referred to as the **inverse elasticity rule**). Conversely, the more price-elastic the demand for the product, the greater the relative portion of the tax borne by the sellers.

Remember that the flip-side of the tax burden for consumers and producers is government's tax revenue. Comparing demand curve D_0 to D_1 it is evident that the **more price-inelastic the demand curve, the greater the total tax revenue** government collects (P_1CFP_2 compared to P_3ABP_4).

So far we can generalise by stating that **the more inelastic the demand, the easier it is to shift the burden of a tax forward through a higher selling price**



8.6 GENERAL EQUILIBRIUM ANALYSIS OF TAX INCIDENCE (10.6)

We have distinguished between **partial equilibrium analysis** and **general equilibrium analysis**. Partial equilibrium analysis examines a single market in isolation (i.e. on the basis of the ceteris paribus assumption) and ignores the secondary effects of a price change. When the secondary effects of a tax are taken into account, we are studying tax incidence within a general equilibrium framework.

Because general equilibrium analysis considers what happens in all markets simultaneously, modelling tax incidence becomes rather complex. To make it manageable, assumptions have to be made;

- We assume that only **two products** are produced in the economy: shoes and reed baskets.
- There are **two factors of production**, labour and capital, that are perfectly mobile between sectors.
- Both factors of production are **fixed in supply** (i.e. the supply curves are vertical).
- Shoes are produced using a **capital-intensive technique** (i.e. the capital-labour ratio is high) whereas reed baskets are produced using a **labour-intensive method** (i.e. the capital-labour ratio is low).

A Selective Tax on Commodities

Suppose a tax is imposed on shoes. The price of shoes increases relative to that of reed baskets. The after tax price does not necessarily increase by the full amount of the tax. The price increase has two effects - one is on the **consumption (uses) side**, the other is on the **factors of production (sources) side**.

On the uses side, the price increase of shoes causes consumers to demand fewer shoes (illustrated by a movement along the demand curve for shoes) and demand more reed baskets. When the demand for reed baskets increases (illustrated by a shift of the demand curve to the right), the price of reed baskets increases, ceteris paribus. The tax thus causes an increase in the price of both products, that is, the **tax burden (and incidence) is spread to the consumers (and producers)** of the non-taxed product as well.

On the sources side, the fact that the tax on shoes results in fewer shoes being demanded implies that fewer shoes will need to be produced. Some capital and labour used becomes redundant. But since shoes are produced using capital-intensive technology, more capital than labour is released into the market. The redundant labour and capital must now find employment in the reed baskets sector. Since the reed basket sector is a labour-intensive sector (and technology is assumed to remain unchanged), all the redundant capital cannot be absorbed into the sector. But with capital in excess supply, its relative price will decrease, so that both sectors will end up using more capital intensive production techniques. Thus, in addition to the rise in the relative price of the non-taxed commodity (reed baskets), a tax on shoes also causes the **relative price of capital (i.e. the return on capital) to decrease**. The burden of the tax is therefore also spread to the owners of the factor of production used most intensively in the production of the taxed commodity

A General Tax on Commodities

If a tax is imposed on shoes and reed baskets simultaneously at the same rate, the general tax is equivalent to a tax on income (e.g. a tax on capital and labour at the same rate). The tax will be borne in **proportion to the consumption (or income)** of each member of the economy. A general tax on commodities leaves relative prices unchanged (including the relative price of leisure)

8.7 TAX INCIDENCE AND TAX EQUITY REVISITED (10.7)

When we introduced the topic of **tax shifting** and **tax incidence**, we stated that most taxes alter the distribution of income. To make inferences about the equity of a tax, it must be ascertained whether the tax alters the distribution in a **progressive, proportional, or regressive** way.

If the economic **burden of the tax is on buyers** and the **expenditure on this good increases** as individuals move up the income scale (i.e. if the demand is relatively income elastic), the tax is progressive. Luxuries tend to fall in this category of goods and services - as income increases, individuals tend to spend more on luxuries.

Expenditure on necessities tends to fall as total income rises (i.e. low-income earners spend proportionally more on these goods and services than high-income earners). A tax on necessities that is **shifted to buyers is therefore regressive**. A value added tax or a general sales tax levied on a broad base will be regressive. The reason is that consumption as a percentage of income decreases as individuals move up the income ladder - high-income earners tend to save proportionally more, leaving them with a proportionally lower tax burden than low-income earners.

When the **tax is shifted to the seller** it also has **distributional implications**. If the seller has to bear the tax, his or her real factor **income (wage, rent, interest, profit) is reduced**. Whether the tax is progressive or regressive now depends on the factor income shares of the **'sellers' and their relative positions on the income ladder**. If the tax is shifted to unskilled workers who find themselves at the bottom end of the income scale, the tax could be said to be regressive. On the other hand, if the tax is shifted to highly skilled workers earning high incomes, the tax will be progressive.

To determine the final impact of a consumption tax on distribution, the effect on buyers and sellers should be considered simultaneously. There is **no systematic relation between the distribution in consumption of a good and the distribution of the factor income that the production of a good generates**. They conclude that in the absence of evidence to the contrary, the distribution of the tax burden is dominated by what happens to consumption (i.e. the extent of tax shifting to buyers) since the initial impact is on the uses side.

STUDY UNIT 9 – TAX EFFICIENCY, ADMINISTRATIVE AND FLEXIBILITY

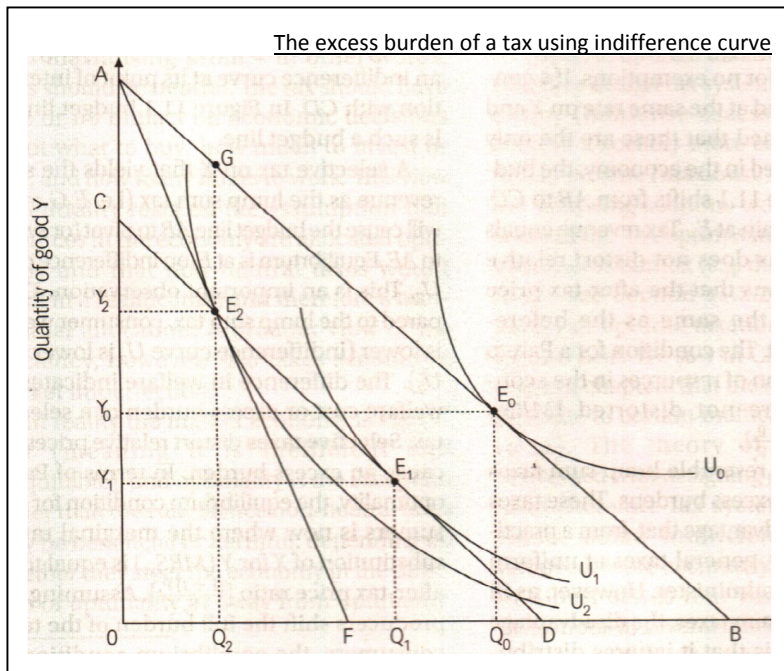
9.1 EXCESS BURDEN OF TAXATION: INDIFFERENCE CURVE ANALYSIS (11.1)

All taxes place a burden on consumers, workers, or producers. In addition to this direct burden, most taxes cause a burden that is greater than what is necessary to generate a certain amount of tax revenue. This additional burden is called the **excess burden, welfare cost, or deadweight loss of a tax**. It measures the loss in benefits (well-being) to consumers and producers that results when prices are distorted by a tax and then inhibits the markets for the taxed goods to achieve efficient levels

Lump Sum Taxes and General Taxes

A lump sum tax is a fixed amount of tax an individual would pay in. These taxes do not distort relative prices and therefore do not affect people's choices. Put differently, **a lump sum tax does not cause a substitution effect**. It reduces the taxpayer's disposable income and thus has an effect on income only. Lump sum taxes have one major disadvantage. Since the tax as a percentage of income decreases as income increases, **lump sum taxes are regressive**. They leave the after-tax distribution of income more unequal than the before-tax distribution.

The effects of a lump sum tax are illustrated. The before-tax budget line is AB . The consumer is initially in equilibrium at point E_0 where the indifference curve U_0 is tangent to the budget line. A lump sum tax is introduced that lowers the income of the consumer and causes the budget line to shift from AB to CD . The lump sum tax yields revenue equal to AC if measured in terms of Y , or DB in terms of X . Note that the after-tax budget line, CD , is parallel to AB since relative prices are unchanged. The consumer is now in equilibrium at E , where fewer of X and fewer of Y are consumed than before.



The consumer is in a worse position after the tax, since consumption is on a lower indifference curve, U_1 . This is due to the normal burden of the tax. The condition for Pareto optimality has not been disturbed and resources are allocated efficiently at the new after-tax income level. A lump sum tax thus causes a normal burden but it has no excess burden.

A general tax is defined as one that taxes the **entire tax base and allows for no exemptions**. If a general tax is imposed at the same rate on Y and X and it is assumed that these are the only products produced in the economy, the budget line shifts from AB to CD . Equilibrium is again at E , where Tax revenue equals AC . A general tax does not distort relative prices. This means that the after-tax price ratio is the same as the before-tax price ratio. The condition for a Pareto efficient allocation of resources in the economy is therefore not distorted.

General taxes resemble lump sum taxes and do not have excess burdens. These taxes have the added advantage that general taxes at **uniform rates are easy to administer**. However, as in the case of lump sum taxes, the disadvantage of general taxes is that it **ignores distributional implications**.

Selective Taxes

Suppose a selective tax is now imposed on X in the figure above. To analyse its welfare implications, we compare the impact of the selective tax to that of a lump sum tax that generates the same tax revenue (i.e. AC). We know that a selective tax on X will increase the price of A and leave the price of Y unchanged. If the consumer spends his or her entire budget on X , fewer of X can be obtained after the tax than before the tax. If the consumer spends his or her entire budget on Y , the same amount of Y can still be purchased (the price of Y has unchanged). The budget line showing combinations of Y and X will, therefore, swivel inward (or pivot around point A). To obtain the after-tax budget line, which yields the same tax revenue as the lump sum tax, we must find an equilibrium point which is also on budget line CD . A budget line therefore has to be drawn through A in such a way that it is tangent to an indifference curve at its point of intersection with CD . This is budget line AF .

A selective tax on X that yields the same revenue as the lump sum tax ($E_2G = AC$) will cause the budget line AB to swivel to AF . Equilibrium is at E_2 on indifference curve U_2 . Compared to the lump sum tax, consumer welfare is lower (indifference curve U_2 is lower than U_1). **The difference in welfare indicates the welfare cost or excess burden of a selective tax.** Selective taxes distort relative prices and cause an excess burden.

Tax Neutrality

The traditional, narrow neutrality concept is mainly concerned with allocation, the idea being that taxes should not prevent consumers from maximising utility or producers from maximising profit - taxes should be neutral. **The tax should have little or no impact on economic decisions.** This view of neutrality rests on the assumption that resources are allocated optimally and that non-neutral taxes would result in a reallocation and therefore a non-optimal allocation. A broader view of tax efficiency, however, also takes account of market imperfections.

In reality the market economy is imperfect and optimality is the exception rather than the rule. Hence **non-neutral taxes may be beneficial or harmful**, depending on whether they steer the economy in the **direction of optimality or away from optimality**. These two possibilities are called **positive non-neutral** and **negative non-neutral effects** respectively.

A **selective tax is clearly non-neutral** as it disturbs relative prices. Economic choices are affected and, according to the traditional view, selective taxes are therefore inefficient. According to the modern approach, however, a selective tax can move the economy in the direction of optimality, which amounts to positive non-neutral action.

In contrast, a **general tax such** as a head tax, is a neutral tax and does not influence allocation decisions. From the modern perspective, a general tax may well perpetuate existing distortions in the economy or perpetuate a socially unacceptable income distribution, which implies that **tax non-neutralities can sometimes improve allocative efficiency**. Examples include levying taxes to correct for negative externalities and taxing sumptuary consumption (e.g. excises on cigarettes and liquor).

The term **optimal taxation** is derived from efforts to design tax systems to improve efficiency and to achieve a socially more equitable distribution of income (maximise social welfare). Related to the theory of optimal taxation is the theory of second best. The theory of second best is concerned with **designing government policy (and tax systems) in situations where some inefficiencies cannot be removed**. More accurately, the theory of second best states that whenever there are distortions in several markets, removing one may not necessarily improve matters. In fact, it may introduce other distortions and, in general, may result in a lower level of welfare for consumers

9.2 EXCESS BURDEN: CONSUMER SURPLUS APPROACH (11.2)

The Magnitude of Excess Burden

Selective taxes cause an excess burden, whereas general taxes do not. Using the indifference curve approach, we have shown that the same tax revenue can be obtained by levying a general tax or lump sum tax and at the same time leave the consumer better off.

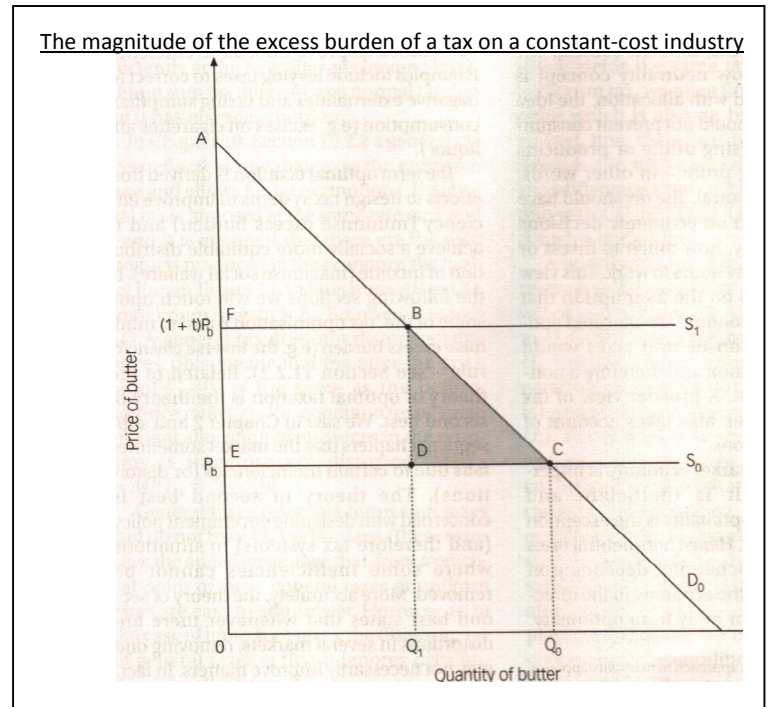
We follow the **consumer surplus approach** in order to measure the excess burden of a tax. To enable measurement, we employ a **partial equilibrium framework** and focus on the burden of a unit tax. We can simplify the example by assuming that supply is produced under constant-cost conditions (i.e. the supply curve is horizontal). Furthermore we use standard demand curves.

The demand curve for butter is shown as D_0 and the supply curve is S_0 . The equilibrium price is P_b and equilibrium quantity is Q_0 . We now have to introduce the consumer surplus. The demand curve indicates what consumers are willing to pay for different quantities. The difference between what consumers are willing to pay for a good and what they actually pay is the **consumer surplus**. This is represented by the area under the demand curve and above the price line. The consumer surplus measures the rand value of consumer welfare at different quantities. **The consumer surplus is ACE.**

Suppose a selective tax (t) is levied on the producers of butter that increases the price of butter to $(1 + t)P_b$. The supply curve S_0 shifts parallel upwards to S_1 . The equilibrium quantity of butter decreases from Q_0 to Q_1 and the consumer surplus is reduced from ACE to ABF . The loss in consumer surplus is the trapezoid $FBCE$. But this is not the total welfare cost to the consumer. The tax revenue is the tax per unit (tP_b) multiplied by the quantity of butter purchased (Q_1) (i.e. $FE \times 0Q_1 = FBDE$). The tax on butter therefore yields revenue equal to the area $FBDE$. If government were to return this amount of tax to consumers as a lump sum,

the consumers are worse off by the triangle BCD ($FBCE - FBDE$). The **triangle is the welfare loss or excess burden of the selective tax** on butter. In other words, the tax causes a reallocation of resources (less butter and more other goods are produced than without the tax). The triangle measures the welfare loss caused by this misallocation of resources.

The size of the triangle (excess burden) is determined by the price elasticities of demand and supply and the tax rate.



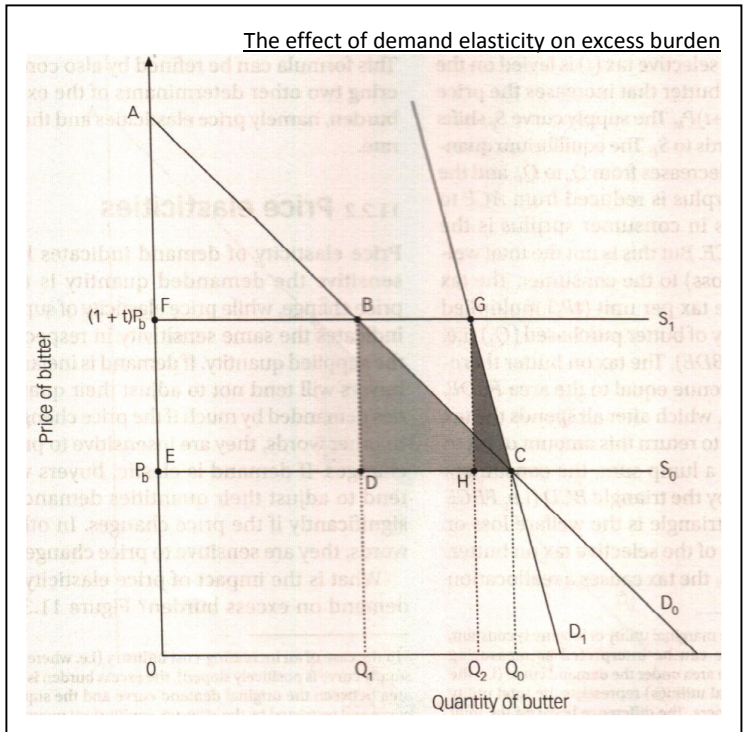
Price Elasticities

Price elasticity of demand indicates **how sensitive the demanded quantity is to a price change**, while price elasticity of supply indicates the same **sensitivity in respect of the supplied quantity**.

If **demand is inelastic** - buyers will tend not to adjust their quantities demanded by much if the price changes. i.e. they are insensitive to price changes.

If **demand is elastic** - buyers will tend to adjust their quantities demanded significantly if the price changes. i.e. they are sensitive to price changes.

What is the impact of price elasticity of demand on excess burden? The only difference is that in we compare the impact of a selective tax on butter for two cases of demand. In the one case, the demand curve (D_1) is **relatively inelastic** and in the other case the demand curve (D_0) is **relatively elastic**. The before-tax quantity (Q_0) and price (P_b) are the same for both. The selective tax (t) causes the supply curve to shift upwards from S_0 to S_1 . In the case of demand curve D_0 , the equilibrium quantity decreases to Q_1 . With demand curve D_1 the equilibrium quantity decreases to Q_2 indicating that the quantity demanded is less sensitive to the imposition of the tax than in the case of D_0 . Comparing excess burdens we notice that for demand curve D_0 the excess burden is BCD (the same as in Figure 11.2) and for demand curve D_1 the excess burden is GCH. This illustrates that the welfare cost of a given tax is less where the demand for a good is relatively inelastic. In other words, the more elastic the demand, the greater the excess burden (or welfare cost).

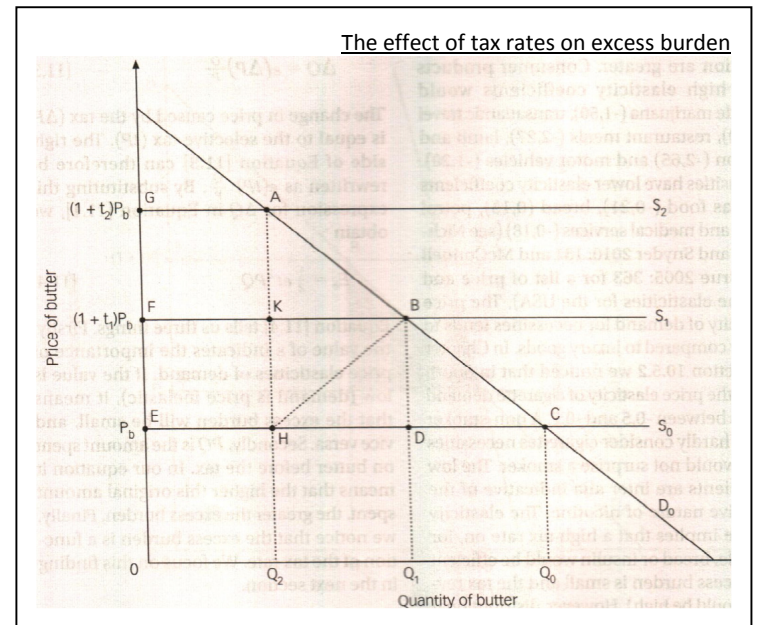


The figure also illustrates **how price elasticities impact on the amount of tax revenue** that can be collected from a tax. Consider first the relatively elastic demand curve (D_0). The tax revenue is the tax per unit (tP_b) multiplied by the quantity of butter purchased (Q_1) (i.e. $FE \times OQ_1 = FBDE$). With the more inelastic demand curve D_1 the tax revenue is clearly much greater (i.e. $FE \times OQ_2 = FGHE$). From a revenue perspective it would make perfect sense for tax authorities to levy taxes on commodities for which demand is relatively inelastic. The implications are that it is **more efficient from an economic efficiency point of view** (i.e. low excess burden) and **tax revenue perspective** (i.e. high tax revenue) to levy taxes on price inelastic commodities than on price elastic commodities. Commodity taxes should thus be high on inelastic goods and services and low on goods and services with high demand elasticities. This tax rule, commonly referred to as the **inverse elasticity rule**, is attributed to Frank Ramsey. A further implication of this tax rule is that **uniform tax rates are not necessarily efficient**, since the higher the elasticity of demand of good X relative to that of good Y, the lower the tax rate on X should be relative to that on Y.

The Tax Rate

The magnitude of the excess burden also depends on the tax rate. As the tax rate increases, the excess burden increases by a multiple of the tax rate.

In the figure, the initial equilibrium is at price P_b and quantity Q_0 . The commodity, butter, is produced under constant-cost conditions. If a selective tax of t_2 is levied on butter, the after-tax price is $(1 + t_2) P_b$ and the equilibrium quantity decreases to Q_2 . The excess burden is the triangle ACH . The tax revenue is the area $GAHE$. Suppose that the tax rate is halved to t_1 . The new equilibrium price is $(1 + t_1) P_b$ and the equilibrium quantity is Q_1 . Tax revenue is now equal to the area $FBDE$. The excess burden is now the triangle BCD . We notice that although the tax rate was halved, tax revenue did not halve. More importantly, the excess burden fell by about three-quarters. This can easily be confirmed geometrically by decomposing the triangle ACH into four smaller triangles of equal size: ABK , KBH , BDH , and BCD .



We can conclude that **low tax rates on a large number of commodities will produce smaller excess burdens** (and more tax revenue) than **high tax rates on a few commodities that yield the same total revenue**. This analysis therefore suggests that broad-based taxes such as VAT and income taxes are more efficient than narrow-based selective taxes.

Excess burdens are real and quite significant. Knowing the excess burdens of different taxes are useful in designing taxes. If, for example, an ad valorem tax of 30% is levied on a good of which the price elasticity of demand is equal to one and total expenditure on the good is R200 million, it can be shown that the excess burden will be R9 million (i.e. 15 per cent of the tax revenue).

$$E_b = \frac{1}{2} \epsilon t^2 P Q$$

Put differently, if the excess burden is expressed per rand of tax revenue the efficiency-loss ratio of the tax can be calculated (i.e. excess burden ÷ tax revenue). Given tax revenue of R60 million from the ad valorem tax, the efficiency-loss ratio is 0,15. This means that for each rand raised in taxes, the deadweight loss (excess burden) is 15 cents. By comparing the efficiency-loss ratios of different taxes, tax designers can attempt to minimise the total excess burden of the tax system. Overall economic efficiency can be improved by reducing taxes on goods and services with high efficiency-loss ratios and increasing taxes on goods and services with low efficiency-loss ratios.

9.3 ADMINISTRATIVE EFFICIENCY (11.3)

The excess burden is not the only cost of a tax. In addition to administration costs borne by government, and ultimately taxpayers, individual taxpayers incur costs in order to meet their tax obligations, called **compliance costs**. These include the cost of time spent filling in tax returns, as well as the cost of employing tax specialists (accountants and lawyers). When taxes are designed, these costs must also be considered.

Administrative efficiency entails minimising both administration costs and compliance costs. Two phenomena are related to these costs:

1. Tax avoidance is perfectly legal and includes the actions by taxpayers to take advantage of special provisions (tax loopholes) in the tax code so that their tax liability is reduced.
2. Tax evasion is illegal and consists of actions that contravene tax laws. The most common forms of evasion are not registering as a taxpayer, under-reporting of income and claiming more deductions than warranted.

To reduce the level of underreporting taxes the revenue authorities have two clear policy options: **increase the marginal cost for underreporting** and/or **decrease the marginal benefits taxpayers derive from tax cheating**.

When taxes are evaluated according to the criterion of administrative efficiency, a number of issues should consequently be considered:

- **The golden rule in tax design is simplification.** Simple tax laws are easy to understand and comply with. High marginal rates should be avoided and the poor should not be taxed. High penalties and a high probability of being detected increase the marginal cost of cheating.
- An important consideration is the **community's level of literacy**. For example, income taxes require high levels of skill, whereas a head tax is fairly easy to understand.
- A further consideration is the **efficiency and expertise of the tax administration**.
- To **improve tax collection**, taxes should be withheld at source. Most taxpayers are subject to the pay-as-you-earn (PAYE) system, tax is withheld at the source (e.g. by the employer) and paid over directly to the tax authorities.
- A further consideration is **the tax morality of the community**. Taxpayers' willingness to part with money is linked to perceptions about the vertical and horizontal equity of taxes as well as the way in which the tax revenue is spent.
- Tax efficiency is also affected by the **political will to enforce tax laws**.
- For **taxes should be certain and transparent**. The tax to be paid should be certain (predictable) to ensure rational decision-making on the part of both taxpayers and the government. Both the tax collector and the taxpayer should thus be given as little discretion as possible.
- **Transparency** means that the government should not take advantage of people's ignorance. Transparency also means that the government has a responsibility to subject tax decisions to the political decision-making process tax decisions should be embodied in legislation and be actively debated.

9.4 FLEXIBILITY (11.5)

Economic activity is characterised by recurrent recessions and booms. Taxes should be flexible enough to provide for changing economic conditions. Taxes can influence economic activity from both the supply and the demand side.

On the supply side, economic growth can be influenced by changing the incentive to work and spend (or save). For example, if the price elasticity of the supply of female labour is greater (i.e. labour supply is more sensitive) than that of males, female labour should be taxed at a lower marginal rate (according to the inverse elasticity rule).

In macroeconomics the use of demand-side measures to smooth out business cycles. This is known as stabilisation policy and a distinction is made between automatic and discretionary stabilisers. The timing of discretionary fiscal action is decisive. The problem of timing, however, is not too serious in the case of automatic stabilisers. Automatic stabilisers are characterised by built-in flexibility. An example of an automatic stabiliser is a progressive income tax system. When the economy is entering a recession, for example, the average income tax rate will automatically begin to decrease. As individuals incomes decline, they are automatically assessed at lower rates. By the same token, tax revenue for the government declines.

STUDY UNIT 10 – INCOME TAXATION

10.1 THE PERSONAL INCOME TAX RATE STRUCTURE (12.3)

The tax base which probably lends itself best to the ability-to-pay principle of taxation is personal income tax. Personal income tax can be adjusted for **personal circumstances** and the rate structure can be designed to comply with the values of society. The circumstances are the taxpayer's marital status, age, gender and whether the unit of taxation should be the household or the individual.

Personal income tax is often subject to **progressive rates**. Progressive rates can be achieved by providing tax rebates and graduated brackets, that is taxing extra amounts of income at successively higher marginal rates.

Although progressive personal income tax rates are still the rule, recently a **flat (proportional) rate** structure has been the topic of serious discussion. An alternative tax reform option is the **dual income tax**.

10.2 ECONOMIC EFFECTS ON PERSONAL INCOME TAX (12.4.1)

Economic Efficiency

When the economic efficiency of any tax is considered, economists try to establish if the tax has an excess burden. The excess burden is a burden in addition to the normal burden of a tax that reduces the taxpayer's welfare (i.e. leaves the taxpayer on a lower indifference curve). To determine whether an income tax has an excess burden, we will consider a general tax on income and then a selective tax on labour income.

General Tax on Income

In the impact of a general tax we concluded that such a tax has **no excess burden since relative prices are not distorted** (head tax). If the entire income base (personal income and company income) is taxed and leisure is ignored (or it is assumed that leisure can be taxed at the same rate as income), an income tax is also a general tax. The tax on income will simply shift the after-tax budget line (CD) parallel and downwards to the origin. Relative prices remain unchanged and there is no excess burden. The tax has a normal burden only, which is illustrated by consumption occurring on a lower indifference curve (U_1) than before. An income tax that taxes the entire tax base (excluding leisure) is therefore an efficient tax. Unfortunately, leisure cannot be ignored and neither can leisure be taxed that easily.

Selective Tax on Labour Income

Once we include leisure, a personal income tax **does have an excess burden**, i.e. choice between income and leisure. Since leisure cannot be taxed easily, a tax on income only is in fact a selective tax. People may decide to work more or less as a result of the tax (i.e. **the supply of labour is affected**). The net result will depend on the **relative strengths** of the income and substitution effects of the price change.

We show that it is the substitution effect that has an **adverse impact on the incentive to work** and that the substitution effect is determined by the **marginal income tax rate**.

Say, Peter, has 18 hours a day which can be used for two activities: work and leisure. In the figure the daily **time endowment** is measured on the horizontal axis as the distance OL . We measure two things;

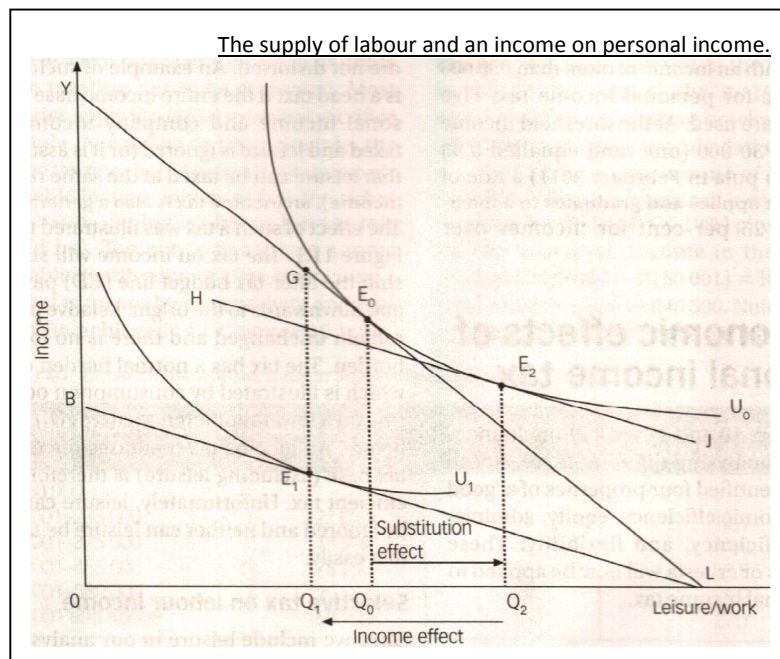
1. From left to right (i.e. from point 0 to point L) we measure the number of **hours spent on leisure activities**.
2. From right to left (i.e. from point L to point 0) we measure the **number of hours worked**.

Suppose Peter earns a wage of R10 per hour, using the entire daily time endowment (18 hours) to work, he can earn R180 per day. Peter's income is plotted on the vertical axis. If the full-time endowment is used to earn income, one combination of income and leisure or work is obtained (i.e. point Y). If, instead, he wastes all his time doing nothing, zero income is earned and we have another combination of income and leisure/work which we can plot (i.e. point L). We can continue in this way and trace out the different combinations of income and leisure or work for each number of hours worked. This is shown as line YL , that is, Peter's budget line. **The two 'goods' in this case are income and leisure**. By adding indifference curves we can determine the combination of income and leisure or work most preferred by Peter. Suppose that the highest indifference curve that Peter can attain is U_0 . He is then in equilibrium at E_0 where OQ_0 hours are spent on leisure and LQ_0 hours are used or supplied for work.

Suppose a proportional tax or flat rate tax (14%) is now levied on income only. A proportional tax lowers the after-tax wage. This means that Peter's income is now less at each number of hours worked. The after-tax budget line pivots (or swivels) from YL to BL . The after-tax equilibrium is at E_1 where indifference curve U_1 is tangent to the new budget line BL . Peter's welfare has declined as indicated by the fact that he finds himself on a lower indifference curve. At the new equilibrium Peter has a tax liability of E_1G (i.e. the difference between his before-tax and after-tax income earned by working LQ_1 hours). The quantity of labour supplied has increased from LQ_0 to LQ_1 hours per day. This is not always be the case. The movement from E_0 to E_1 is due to the combined effect of the income and substitution effects of the tax change. Whether the number of labour hours supplied increases or decreases, will depend on which effect is the stronger. The budget line DF and indifference curve U_2 are discussed later in this section.

What is the income effect? The tax reduces Peter's after-tax income. Peter is worse off and will tend to work more to partly offset this loss in income and cannot afford the same amount of leisure than before - **less leisure means more hours of work**. The income effect will therefore cause the quantity of labour supplied to increase.

What is the substitution effect? It is important to note that **leisure has a price**, i.e. the hourly wage sacrificed. Thus, the **opportunity cost of leisure is the wage**. Since the introduction of an income tax reduces the after-tax wage, the opportunity cost of leisure therefore decreases - leisure becomes cheaper. In other words, consuming leisure involves a smaller sacrifice in income than before the introduction of the tax. Because leisure is now relatively cheaper, it is substituted for work. The substitution effect increases leisure, which means that it reduces the quantity of labour supplied.



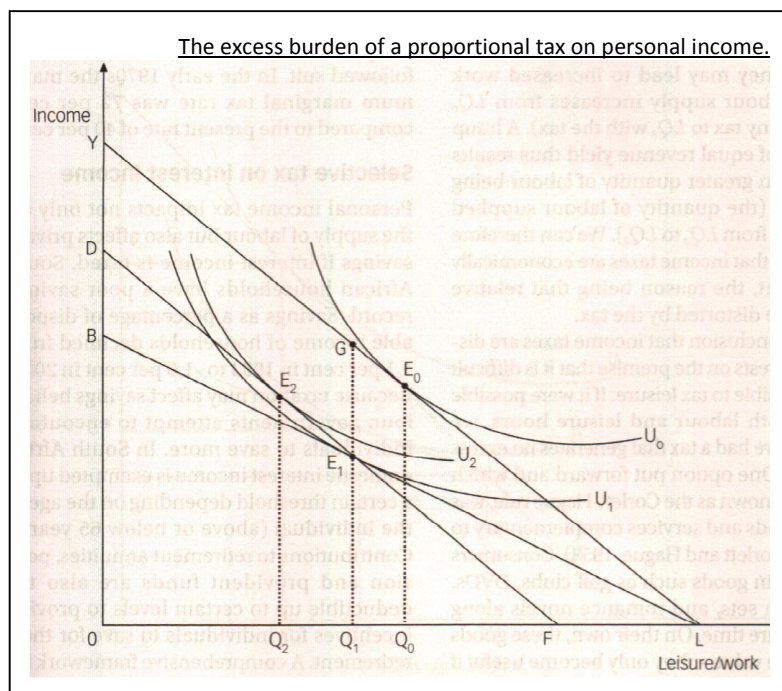
After the introduction of the proportional tax, equilibrium changes from E_0 to E_1 . The movement from E_0 to E_1 , can be decomposed into the **income and substitution effects**. By hypothetically compensating the individual with an amount just enough to make him as well off as before the tax, the budget line BL shifts parallel outwards to HJ which is tangent to indifference curve U_0 at E_2 . The movement from E_2 to E_1 is therefore, the **income effect of the tax**. The income effect shows that the individual increases work effort in response to the imposition of the proportional tax. The movement from E_0 to E_2 , depicts the **substitution effect**, which is a consequence of the change in the relative price of labour alone. The movement shows how the individual substitutes more leisure for less work if a proportional tax on labour is imposed or increased.

If the income effect dominates, the after-tax quantity of labour supplied will increase. If the substitution effect dominates, the number of hours worked will decrease and the quantity of leisure will increase. It is impossible to say what the outcome will be. Empirical evidence suggests that the labour supply **elasticities for prime-age men are generally close to zero**. This means that the supply curve is almost vertical or that the quantity of **labour supplied is very insensitive to changes in the net wage**. On the other hand, the estimated **elasticities are positive and high** for married women. In their case, the quantity of **labour supplied is thus quite sensitive to changes in the net wage**.

The analysis used for a proportional tax can be repeated for a progressive tax on income. One difference would be the shape of the after-tax budget line. In the case of a progressive tax, the budget lines will be kinked or curved. This does not fundamentally change the conclusion with regard to the impact of the tax on the supply of labour, but there is one important difference. For a proportional tax the average tax rate is equal to the marginal tax rate, but for a progressive income tax the marginal tax rate is greater than the average tax rate. This means that the adverse incentive effects due to the substitution effect are likely to be more important for a progressive tax than for a proportional tax.

We know that a tax which selectively taxes income has income and substitution effects which may cause the quantity of labour supplied to increase or decrease. We also know that the taxpayer's welfare is reduced by the tax. However, we still have to determine the **economic efficiency of the tax** (excess burden). We have to compare the results of the proportional tax to a lump sum tax of equal revenue. Remember that a **lump sum tax does not distort the relative prices of leisure/work** and income and therefore has no excess burden. The impact of the personal income tax on economic efficiency (the excess burden) is explained in Figure 12.2. This figure is similar to Figure 12.1 except that the income and substitution effects are not illustrated.

If a **lump sum tax (e.g. head tax)** is introduced to raise the same tax revenue as the proportional tax and (i.e. E_1G at Q_1 hours), the after-tax budget line shifts parallel to the original budget line (YL) to DF , which intersects BL at point E_1 . Peter will now be in equilibrium at E_2 where the highest indifference curve U_2 is tangent to the (new) budget line. Peter is working LQ_2 hours a day which is more than the LQ_1 hours worked under the proportional tax. Peter's welfare is also higher under the lump sum tax than under the equal-yield proportional income tax as illustrated by the attainment of a higher indifference curve (U_2) than before (U_1). The difference between U_2 and U_1 can be ascribed to the **excess burden of a proportional income tax** that selectively taxes income. Note that, compared to lump sum taxes, selective income taxes lower the incentive to work even though they may lead to increased work effort (labour supply increases from LQ_0 without any tax to LQ_1 with the tax). A lump sum tax of equal revenue yield thus results in an even **greater quantity of labour being supplied**. We can therefore conclude that income taxes are economically inefficient, the reason being that relative prices are distorted by the tax.



Our conclusion that income taxes are distortional rests on the premise that it is difficult or impossible to tax leisure. If it were possible to tax both labour and leisure hours, we would have had a tax that generates no excess burden. One option put forward and which became known as the **Corlett-Hague rule**, was to **tax goods and services complementary to leisure**. Consumers use certain goods such as golf clubs, DVDs, television sets, and romance novels along with leisure time. The policy solution would be to **tax goods complementary to leisure** at a rate that would reduce demand for leisure indirectly. In this way the excess burden of the income tax system could possibly be reduced.

Equity (12.4.2)

Personal income taxation lends itself to the application of the ability- to-pay principle. Through a system of exemptions, deductions, tax rebates, and marginal tax rates, the rate structure can be made to conform to society's notion of fairness. Even though the **statutory burden of the tax is on the individual**, it may be possible for individuals to **shift the burden**. The critical issue is how sensitive the supply of labour is to price and tax changes. As already noted, the net effect on work effort will be determined by the **relative strengths of the income and substitution effects**. Empirical evidence points to an **insensitive or inelastic supply of labour for men** (i.e. the supply curve is almost vertical). If the supply curve is relatively inelastic, the burden is on the supplier, that is, the employee in this case. If the **supply curve is less inelastic**, the employer and the employee will share the burden. This scenario is possible in the case of married women and high-income professionals who are internationally mobile.

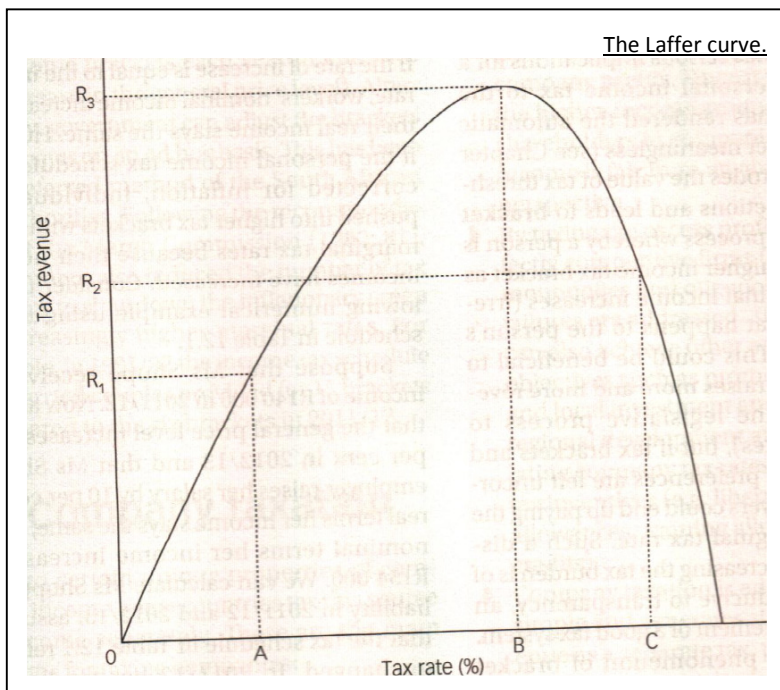
Administrative efficiency and tax revenue (12.4.3)

Income taxes are complex and administrative efficiency requires relatively sophisticated taxpayers and administrators.

Efforts could be made to capture those outside the tax net such as those in the informal sector and other groups that are difficult to tax. These groups can be taxed using **presumptive taxes**. Presumptive taxation involves the use of certain indicators (e.g. ownership of certain assets, personal servants, average profit margins, or average gross turnover) to determine tax liability.

Another method that has been proposed to **increase tax revenue** is to reduce tax rates. This recommendation is based on the alleged tax rate-tax revenue relationship that has become known as the **Laffer curve**. The logic of his argument is that higher tax rates will not necessarily produce more tax revenue, since the tax base will shrink as taxpayers reduce their work effort in response to the higher rates. The debate is again about how sensitive labour supply is to tax changes. Tax revenue is the **product of the tax rate and the tax base**, where the tax base depends on work effort (the number of hours worked). At low tax rates, an increase in the tax rate will tend to increase tax revenues (at low rates people will still work more, i.e. the income effect dominates).

Total tax revenue is plotted on the vertical axis and the tax rate on the horizontal axis. If the tax rate is at A_1 government can still increase tax revenue by raising the rate. Once the tax rate is at B_1 however, a further increase in the rate will cause tax revenue to decline (R_1 to R_2).



The supporters of the Laffer hypothesis use the mechanism to show that if the tax rate is at C_1 the authorities should lower the rate since this will result in higher tax revenue as illustrated by a movement from R_2 to R_3 . The critical question, of course, is to determine where countries are on the curve. The proponents argue that some countries may well be beyond point B. The debate can only be settled empirically but it does appear that labour supply elasticities are such that it is unlikely that rate reductions would be fully counteracted by increased work effort

Flexibility (12.4.4)

One of the 'good' characteristics of a personal income tax is its built-in flexibility to counter cyclical economic behaviour. Personal income tax is thus considered to be an **automatic stabiliser**. On the negative side, inflation has serious implications for a progressive personal income tax to the extent that it has rendered the automatic stabilising effect meaningless. Inflation erodes the value of tax thresholds and deductions and leads to **bracket creep**; this is a process whereby a person is pushed into a higher income tax bracket as his or her nominal income increases (irrespective of what happens to the person's real income). This could be beneficial to government (it raises more and more revenue without the legislative process to increase tax rates), but if tax brackets and the value of tax preferences are left uncorrected, all taxpayers could end up paying the maximum marginal tax rate.

The personal income tax schedule taxes a person on his or her **nominal income**. In inflationary times employers compensate workers for the decline in purchasing power by increasing their wages or salaries. If the rate of increase is equal to the inflation rate, workers' nominal income increases, but their real income stays the same. However, if the personal income tax schedule is not corrected for inflation, individuals are pushed into higher tax brackets with higher marginal tax rates because their nominal incomes have increased

10.3 CAPITAL GAINS TAX (12.8)

Capital gains can be defined as increases in the net value of assets over a period (e.g. an accounting period or fiscal year). anything that makes consumption possible without diminishing wealth at the beginning of a fiscal year is considered to be income. Capital gains are, accordingly, often classified as a form of income and are taxed as such. Capital gains can be taxed as they accrue (an unrealised gain) or when they are realized.

The most important reasons for capital gains taxation are:

- **To protect the integrity of the personal income tax base.** If capital gains are not taxed, taxpayers have an incentive to convert income into capital gains in order to avoid taxation.
- **To ensure horizontal equity.** A capital gain represents an increase in economic power and increases the individual's ability to earn income and to be taxed.
- **To ensure vertical equity.** Capital gains accrue mostly to higher-income taxpayers. If they are not taxed on these gains, the vertical ability-to-pay principle is jeopardised.
- **To improve economic efficiency.** If investments are chosen on the basis of tax considerations, the allocation of investment funds is distorted and an excess burden results.

There are also a number of arguments against capital gains tax of which the most important are:

- **Capital gains taxation is subject to numerous administrative problems.** Assets have to be valued and there is a need for accurate and up-to-date deeds registers in the case of, for example, works of art and real property.
- **If nominal profits (instead of real profits) are taxed, equity is at risk.** Inflation causes imaginary capital gains (i.e. increases in the nominal value of assets) and it may be unfair to tax someone just because inflation has increased the nominal value of an asset.
- **Capital gains are usually once-off events and to avoid the tax, taxpayers tend to lock in rather than realise investments.** This lock in effect can affect investment negatively. Concessions are therefore usually made either in the form of lower personal income tax rates on capital gains, or by not taxing capital gains once a certain period has elapsed.

STUDY UNIT 11 – FISCAL FEDERALISM

11.1 THE ECONOMIC RATIONALE FOR FISCAL DECENTRALISATION (17.1)

The Tiebout Model

Tiebout asserted that if there were a large enough number of local government jurisdictions and each of these local governments offered a different mix of local public goods and taxes, individuals would reveal their true preferences for local public goods by choosing a particular local government jurisdiction in which to live. In this model, citizens are mobile and choose to settle in the local government jurisdictions that produce a mix of tax and public good outputs, which correspond most closely to their preferences.

The greater the number of communities and the greater the variation in taxes and public services offered, the closer consumers will be to satisfying their preferences. Under these conditions, local public goods can be decentralised in a way that is immune to the free-rider problem.

The Tiebout model thus describes a theoretical solution for the problem of preference revelation, a phenomenon that inhibits the achievement of allocative efficiency.

It must be noted that the Tiebout model is based on a number of restrictive assumptions, namely:

- **All citizens are fully mobile**
- Individuals have **full information about the local public goods** offered by each jurisdiction
- there is a **large number of jurisdictions to choose from**, spanning the full range of public good combinations desired by citizens
- There are **no geographic employment restrictions**: people receive income from capital only and are not tied to a particular location through job or family ties
- There are **no spillovers across jurisdictions**
- There are **no economies of scale in the production of public goods**.

Although the Tiebout model is based on a number of stringent assumptions, it does clearly demonstrate that a decentralised fiscal system - which can accommodate a diversity of preferences for public goods - can be welfare-increasing in relation to a centralised system which imposes a standardised public good-tax mix on people. Fiscal decentralisation can in principle contribute to a more efficient provision of local public goods by aligning expenditures more closely with local priorities.

Public Choice Perspective On Fiscal Federalism

From a public good perspective we have already discussed Arrow's impossibility theory and the maximising behaviour of politicians, bureaucrats and special interest groups within a centralised democratic system. Both **explicit and implicit logrolling** on the part of political parties and individual politicians, and the 'empire-building' motives of bureaucrats, coupled with rational ignorance among the broad electorate, can give rise to an over-supply of public goods in the economy. The net effect is clearly sub-optimal, with goods and services not being allocated efficiently.

This Leviathan hypothesis, first proposed by Brennan and Buchanan, views government as a revenue-maximising monopolist which **seeks systematically to exploit its citizens by maximising the tax revenue**. According to this perspective, fiscal decentralisation would place a powerful restraint on the government's Leviathan tendencies. Devolution of taxing and spending powers to sub-national governments would act as a disciplinary force on the size of government by forging a closer link between raising funds and spending funds. Centralised fiscal systems break this link, encouraging the growth of government. In centralised fiscal systems, local residents have more opportunities to lobby for spending programmes that are financed out of nationally collected revenues or national loans.

Other Reasons for Fiscal Decentralisation

Competition between sub-national jurisdictions may **enhance innovation**. Successful local government experiments may be replicated elsewhere and the failures discarded.

There may be a high cost associated with decision-making if it is completely centralised. Due to the smaller groups involved, the devolution of spending and taxing powers may **reduce the cost of decision-making**. Fiscal decentralisation could also **encourage public participation** in decision-making since local and provincial governments may be closer to the communities they serve.

11.2 REASONS FOR FISCAL CENTRALIZATION (17.2)

Although there are advantages associated with fiscal federalism, there are also factors that favour centralization, these are:

1. Firstly, **spatial externalities may arise when the benefit or costs of a public service 'spill over'** to non-residents of a particular jurisdiction. Goods with external benefits are likely to be under-produced, since each sub national government is concerned primarily with the welfare of its own residents. Similarly, public goods with significant external costs may be over-produced since the residents of a jurisdiction do not bear the full social cost. Under these circumstances it may be preferable to have a centralised provision to 'internalise' these costs and benefits.
2. Secondly, **centralised provision of public services may be justified by economies of scale**, that is, certain services (such as transport systems and water) may require areas larger than a single sub-national jurisdiction for cost-effective provision.
3. Thirdly, **centralised provision may lead to lower administration and compliance costs** in the financing of public services. For example, using one computer system for the whole country or one revenue collection system serving national and provincial governments may prevent the cost of duplication.

In practice, no country has a completely decentralised or completely centralised system. While the provision of certain goods is preferable at national level, others are best provided at sub-national level.

11.3 TAXING AND SPENDING AT SUB- NATIONAL LEVEL: THE ASSIGNMENT ISSUE (17.3)

The assignment problem is concerned with how spending and taxation responsibilities should be distributed among national and sub-national governments. Fiscal federalism literature provides broad guidelines on this fundamental issue.

Stabilisation Function

There is general consensus that **macroeconomic policy should be assigned to central government**. Sub-national governments should not conduct monetary policy. If the power to create money were decentralised to regional entities, there would be strong incentives for sub-national governments to print money to finance public service provision, rather than raising sub-national taxes or imposing user charges. Such behaviour would clearly lead to inflationary pressures

The conventional wisdom is that fiscal **stabilisation policy would be ineffective at sub-national level**. Provincial and local economies tend to be 'open' (i.e. they 'import' and export' large shares of what they produce or consume from other provinces or local jurisdictions). If a single sub-national government were to pursue an expansionary fiscal policy, for example, much of the increase in demand would be lost to outside jurisdictions due to the openness of such economies. If, for example, a provincial government were to cut taxes substantially in order to stimulate the provincial economy, most of the newly generated spending would flow out of the provincial economy in payment for goods and services produced elsewhere. The ultimate impact on employment levels in the province would be very small. Fiscal policy by sub-national governments is thus likely to prove impotent since the extent of import leakages would substantially reduce any multiplier effects.

Distribution Function

In the fiscal federalism literature it is generally argued that only a centralised redistribution policy by central government is likely to be effective. The argument is that any effort to **redistribute income by a single sub-national government** (e.g. by increasing taxes on high- income earners and firms and spending the proceeds on the poor) **would ultimately be self-defeating**. There would be an influx of poor migrants into the jurisdiction, attracted by the fiscal benefits. This would be accompanied by an exodus of high-income earners and businesses from that jurisdiction. It then becomes more difficult for the jurisdiction to attain its distributional goals, given the dwindling tax base. Sub-national governments may therefore end up in a worse distributional position.

Allocation Function

Probably the most compelling economic case for **fiscal decentralisation is its potential to secure efficiency gains**. The static arguments linking fiscal decentralisation with improved efficiency include the following:

- **Uniform centralised policy forces every region to consume the same mix of taxes and public spending**, even though tastes and attitudes may vary widely across regions in a large country with many cultural and ethnic groups. Each decentralised jurisdiction could more closely tailor its service and tax package to the preferences of its citizenry.
- **Different public goods have different spatial characteristics**. Some benefit the entire country (e.g. defence) whereas others benefit only a province (e.g. forestry services) or a locality (street lighting). Public services are provided most efficiently by a jurisdiction that has control over the minimum geographic area.
- **Lower-tier governments may have more information about the needs and priorities of their citizens** as well as region specific conditions and prices than national governments, which could improve programme design and service delivery.
- **Diseconomies of scale and increasing bureaucratic inefficiency** arise when spending programmes become too large, that is, when they serve too large a geographical area.

The **dynamic efficiency arguments** point out that fiscal decentralisation can stimulate innovation. Variety in policy design and application at sub-national level is seen as desirable as it diversifies the country's exposure to disastrous policy experiments. Successful policy experiments at sub-national level can be replicated by other tiers of government as best practice and the failures can be discarded.

Improved allocative efficiency in a decentralised system depends heavily on the political and institutional mechanisms through which sub-national governments can be made aware of their electorates' preferences and are held accountable for their actions. However:

- In many developing countries these **democratic structures** are not in place, or if they are nominally in place, de facto do not function.
- Furthermore, **sub national governments may lack capacity** and may be prone to corruption.

Tax Assignment

Tax assignment refers to the assignment of tax sources to different tiers of government. Expenditure assignment refers to the assignment of expenditure functions to different tiers of government.

Two important factors for consideration are

1. **Equity** (ensuring vertical and horizontal equity among individual tax payers as well as across regions)
2. **Efficiency** (minimising the cost of collection and compliance, as well as minimising any market distortions).

In the light of equity and efficiency efficiency, the following assignment guidelines:

- **Progressive redistributive taxes** should be assigned to the national government
- Taxes appropriate for **macroeconomic stabilisation** should likewise be centralised (e.g. value added tax and personal income tax). As its corollary, taxes assigned to the sub-national governments should be less sensitive to economic and business fluctuations, that is, it should be cyclically stable (e.g. motor vehicle taxes).
- **Unequal tax bases among jurisdictions** should be assigned to the national government (e.g. mining tax).
- Taxes on **mobile factors of production** should be centralised (e.g. corporate income tax or value added tax where companies are able to shift the accounting base of the tax to lower-tax jurisdictions).
- **Residence-based taxes** such as excise taxes should be assigned to the provinces.
- The local authorities should **levy taxes on immobile factors of production**, such as property taxes.
- All levels of government may **charge user charges and benefit taxes**

11.4 TAX COMPETITION VERSUS TAX HARMONISATION (17.4)

When different sub-national governments impose different tax rates, citizens and businesses may react by moving to jurisdictions with lower tax rates.

- **Tax competition** occurs when sub-national governments adjust (lower) their tax rates to attract mobile factors of production (notably capital) from other jurisdictions.
 - Uncoordinated tax policies could therefore lead to market distortions with regards to mobile factors of production as well as tradable goods and services.
 - Tax competition as a positive influence and efficiency enhancing. Decentralised tax powers could promote innovation, as sub-national governments would be able to experiment with various fiscal packages
 - It could also permit sub-national governments to tailor tax mixes to their citizens' preferences and furthermore encourage accountability
- **Tax harmonisation** occurs when sub-national governments coordinate their tax policies (for instance, by limiting the degree of variation in tax rates levied, or by defining the tax bases in a uniform way).

11.5 INTERGOVERNMENTAL GRANTS (17.6)

Intergovernmental grants are transfer payments from one sphere of government (e.g. national) to another sphere of government (e.g. a provincial or local government). Intergovernmental grants may be unconditional or conditional.

- **Unconditional grants** may be spent by recipient governments as they see fit.
- **Conditional grants** must be spent on the specific service stipulated by the grantor (i.e. the sphere of government which is making the grant).

Grants may also be matching or non- matching.

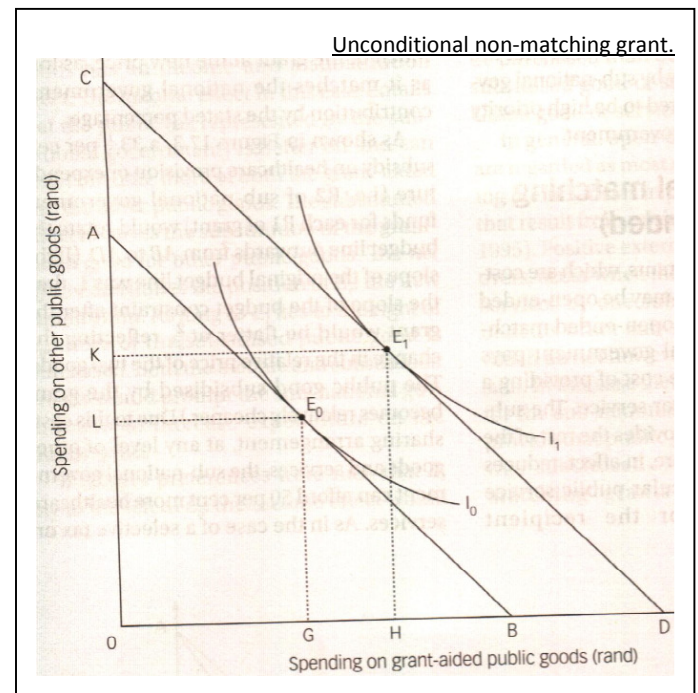
- **A matching grant**, the grantor government will match a certain percentage of each rand of spending by the sub-national government on the same activity.
- **A non- matching grant** is just a lump-sum allocation that does not depend on the level of sub- national expenditure.

Unconditional Non-matching Grants

An unconditional non-matching grant is a **lump-sum transfer to sub-national government which no constraints** are placed as to how it is to be spent. The national government could recommend that the grant be spent on certain public goods - referred to below as 'grant-aided public goods' - but the choice ultimately lies with the recipient government. A non-matching grant is in effect an **income supplement**. A **block grant** in the South African context refers to a type of unconditional non-matching grant where a global lump sum is transferred to a sub-national government to be spent at its discretion. This is also referred to as revenue sharing.

The effect of introducing such a grant is illustrated. We measure spending on the grant-aided public good in rand, on the horizontal axis. Spending on all other public goods is measured on the vertical axis. The line AB shows the government's budget constraint before receiving a grant. The line CD shows the government's new budget constraint after receiving the grant. I_0 and I_1 , are the indifference curves of the median voter, indicating society's relative preferences.

Point E_0 shows the initial equilibrium, which could be thought of as reflecting the median voter's preferences (i.e. preferred combination of grant-aided and other public goods). An unconditional grant (of BD rand) shifts the recipient government's budget constraint from AB to CD. The new equilibrium is at E_1 signifying a higher level of social welfare. There is an increase in grant-aided public good expenditure by the sub-national government (GH). This increase is, however, less than the amount of the grant (BD). Unconditional grants, because they may be spent on any public good or to finance tax breaks, have the least stimulatory effect on the recipient government's consumption of the grant-aided public good.



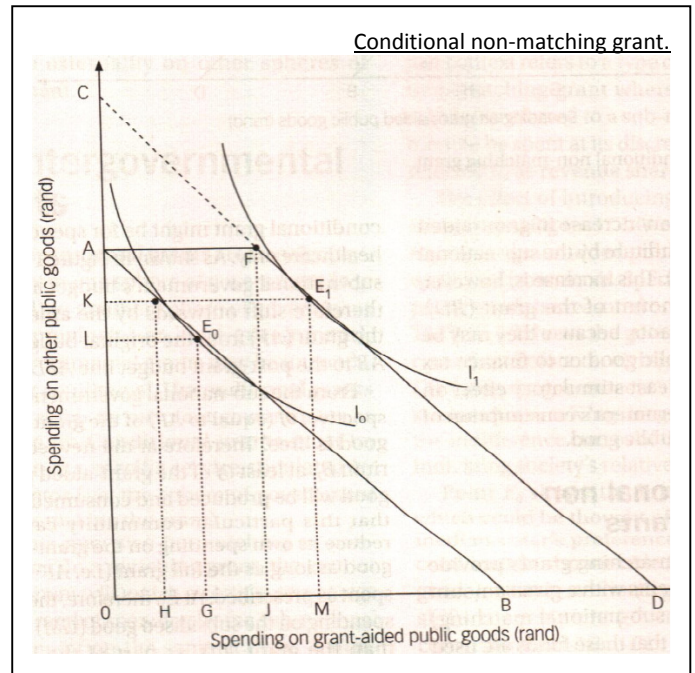
Conditional Non-matching Grants

Conditional non-matching grants provide recipient governments with funds on the condition it's used for a **particular purpose**. The sub-national government's budget line will therefore **shift outwards** by the amount of the grant (AF) from the original budget line AB to the post-grant budget line AFD .

From the sub-national government's perspective, OJ (equal to AF) of the grant-aided good is 'free'. Therefore at the new equilibrium E_1 at least OJ of the grant-aided public good will be produced and consumed. Note that this particular community can still reduce its own spending on the grant-aided good as long as the full grant (i.e. $AF= OJ$) is spent as prescribed.

At E_1 therefore, the extra spending on the subsidised good (GM) is less than the grant (OJ) as part of the initial spending on the subsidised good by the sub-national government was diverted to the other goods (LK).

Such grants are most appropriate to subsidise activities that are considered low priority by sub-national governments but considered to be high priority activities by national government.



Conditional Matching Grants (open-ended)

Matching conditional grants, are **cost-sharing arrangements**, may be **open-ended** or **closed-ended**.

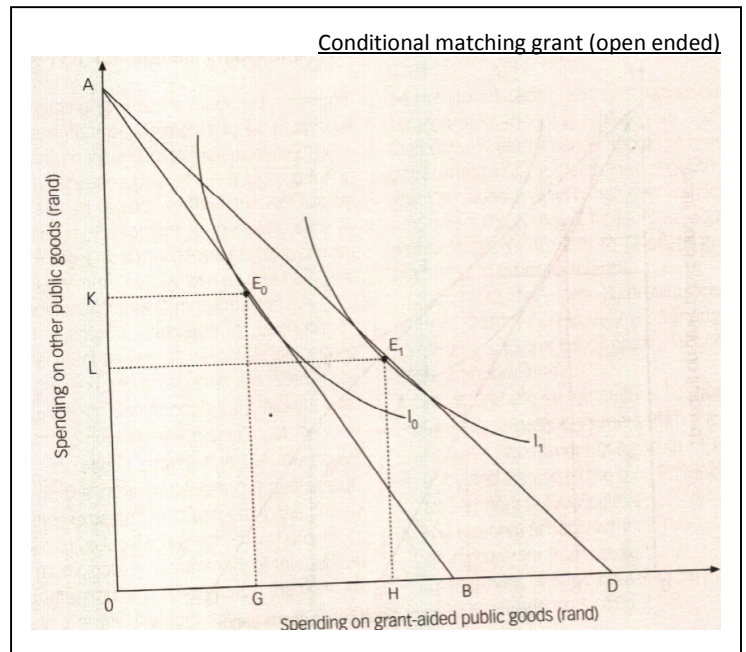
With open-ended matching grants the **national government pays some proportion of the cost** of providing a particular public good or service. The sub-national government provides the rest of the funds needed. In effect, reduces the price of that particular public service for the recipient government.

Since the grant is open-ended, the sub-national government can use as much of the grant at the new price, as long as it matches the national government's contribution by the stated percentage.

As shown, a 33 ⅓ per cent subsidy on healthcare provision or expenditure (i.e. R2 of sub-national government funds for each R1 of grant) would rotate the budget line outwards from AB to AD . If the slope of the original budget line was 1, then the slope of the budget constraint after the grant would be flatter at ⅔, reflecting the change in the relative price of the two goods.

The public good subsidised by the grant becomes relatively cheaper. Due to this cost-sharing arrangement, at any level of other goods and services, the sub-national government can afford 50 per cent more healthcare services. A grant that changes the relative price of public goods has an **income and a substitution effect**.

- The income effect in this case entails that the public is better off and can thus consume more of both the grant-aided and the other public goods.
- The substitution effect involves the substitution of the grant-aided good for other public goods.



The **net effect** determines the position of E_1 the new equilibrium. As long as E_1 lies to the right of E_0 more of the subsidised public good is purchased. Both the income and substitution effects would prompt the sub-national government to increase expenditure on the public good.

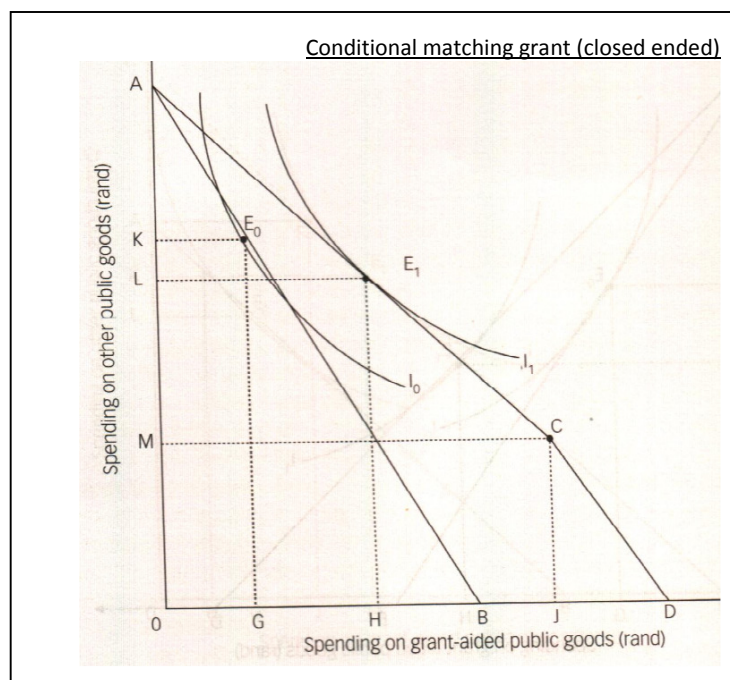
If relative preferences were such that E_1 lay to the left of E_0 the income effect would dominate the substitution effect to such an extent that less of the subsidised good will be purchased than before the grant (i.e. the subsidised good or service is an inferior or Giffen good or service).

In general, open-ended matching grants are regarded as most appropriate for correcting inefficiencies in public good production that result from **positive externalities**. Positive externalities occur when the provision of goods and services by one sub-national government benefits other sub-national governments, which do not however bear the cost of provision. Note that open-ended matching grants may benefit richer sub-national governments more than poorer ones who might not be able to match national government expenditure. Geometrically it can be shown that if E_1 were to lie directly above E_0 , the cost to the sub-national government of the new bundle of public goods would be the same as the pre-grant combination. The response of a poor community to a conditional grant may well be to seek a combination of goods that does not increase or even decrease the total cost in respect of all public goods, that is, E_1 will be directly above or even to the left of E_0 .

Conditional Matching Grants (closed-ended)

There are also closed-ended matching grants where the national government **pays some proportion of the cost** of providing a particular public good or service, up to a certain limit.

When there is a 33 ½ per cent subsidy on, for instance, healthcare up to a limit, the budget line will move from AB to ACD. Costs of healthcare provision will be shared along AC until the subsidy limit (at spending level OJ) is reached. Beyond the subsidy limit, healthcare is unsubsidised and the sub-national government faces the full price of provision; hence the steeper slope of the section CD of the new budget line (the slope of CD is the same as that of AB). At the new equilibrium, E_1 more healthcare will be provided than would have been the case without the grant. Grantor governments generally prefer closed-ended matching transfers as these allow them to retain control over their budgets.



The Rationale for intergovernmental Grants

The main arguments for intergovernmental transfers are summarised below. The design of the grant should be appropriate to the objective it seeks to attain:

- **Fiscal imbalances between expenditure needs and revenue generation** capacities of sub-national government can be addressed. Sometimes revenue is collected at national level and then transferred to sub-national governments as block grants to address fiscal imbalances. This is known as **revenue sharing**.
- To **ensure minimum standards in the provision of public goods** and services across the nation, conditional non-matching grants are appropriate.
- To **compensate for benefit spillovers**, conditional matching transfers (open-ended) are suitable. The rate of subsidisation should reflect the degree of benefit spillover.

A conditional matching grant (closed-ended) may be considered to assist sub-national governments financially while promoting expenditures on an activity considered by the national government to be of a high priority, but at the same time affording the national government better control over its own budget.